UNIVERSIDAD AUTÓNOMA CHAPINGO

DEPARTAMENTO DE FITOTECNIA

INSTITUTO DE HORTICULTURA

IDENTIFICACIÓN DE QTL Y SELECCIÓN GENÓMICA PARA RESISTENCIA AL COMPLEJO MANCHA DE ASFALTO EN VARIEDADES NATIVAS DE MAÍZ

T E S I S

QUE COMO REQUISITO PARCIAL

PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS EN BIOTECNOLOGÍA AGRÍCOLA

PRESENTA:

DAVID OMAR GONZÁLEZ DIÉGUEZ

DIRECTOR DE TESIS: DR. JOSÉ DE JESÚS LÓPEZ REYNOSO

Chapingo, Estado de México, diciembre de 2015

UNIVERSIDAD AUTÓNOMA CHAPINGO

"ENSEÑAR LA EXPLOTACIÓN DE LA TIERRA, NO LA DEL HOMBRE"

IDENTIFICACIÓN DE QTL Y SELECCIÓN GENÓMICA PARA RESISTENCIA AL COMPLEJO MANCHA DE ASFALTO EN VARIEDADES NATIVAS DE MAÍZ

Tesis realizada por **DAVID OMAR GONZÁLE DIÉGUEZ** bajo la dirección del Comité Asesor indicado, aprobada por el mismos y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS EN BIOTECNOLOGÍA AGRÍCOLA

COMITÉ ASESOR

DIRECTOR: _____
DR. JOSÉ DE JESÚS LÓPEZ REYNOSO

CO-DIRECTOR: _____
DRA. SARAH HEARNE

ASESOR: _____
DR. LUIS FERNANDO CONTRERAS CRUZ

ASESOR: _____
DR. TERENCE MOLNAR

ASESOR: _____
DR. JAIME SAHAGÚN CASTELLANOS

ASESOR: _____
DR. SANTOS LEYVA MIR

Chapingo, Estado de México, diciembre de 2015

DEDICATORIA

Al Elohim: Por regalarme el don de la vida, por ser el protector de mi alma y mi fortaleza. Porque Él es quien da la sabiduría al hombre y es el camino de la verdad. A ti doy la Honra y Gloria.

Mis padres: Ermides González y Gladis Diéguez, por su amor incondicional y guiarme por el sendero del bien. Por sus inagotables esfuerzos y oraciones constantes. Los Amo!

Mis hermanos: Obed, Eliú, Astrid y Jamileth, por estar siempre pendientes y dispuestos a apoyarme en todo momento.

Mi familia: Especialmente a mi tía Olinda González, por sus atenciones y cuidados desde niño, y a mis demás tíos, tías, primos y sobrinos por permitirme vivir la dicha de crecer dentro de una gran familia.

Mis abuelos: A mi abuelo Ignacio González y a la memoria de mi abuelita Olimpia Recinos (Q.E.P.D), por sus sabios consejos y enseñanzas compartidas. Al abuelito Julio (Q.E.P.D) por los buenos momentos compartidos.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología e Instituto Interamericano de Cooperación para la Agricultura, por la beca otorgada mediante el convenio IICA-CONACyT para la manutención durante los estudios de maestría y a la Universidad Autónoma Chapingo, especialmente al Departamento de Fitotecnia e Instituto de Horticultura, por haberme hecho sentir en casa durante este tiempo.

Al Centro Internacional de Mejoramiento de Maíz y Trigo –CIMMYT-, especialmente a MasAgro Biodiversidad, por el estipendio económico, por los datos aportados y la valiosa asesoría del grupo de prestigiosos científicos que laboran en esta institución, para la realización de esta investigación.

A mi comité asesor: Dr. José de Jesús López Reynoso, Dra. Sarah Hearne, Dr. Luis Fernando Contreras Cruz, Dr. Terry Molnar, Dr. Jaime Sahagún Castellanos y al Dr. Santos Leyva Mir, por su confianza y asesoría durante la realización de esta investigación.

A todos mis maestros del Posgrado en Biotecnología Agrícola, quienes compartieron sus conocimientos para mi formación científica y profesional durante los estudios de maestría.

Al Dr. Paulino Pérez del Colegio de Postgraduados, al Dr. Fernando Toledo, Dr. José Crossa, Dr. Juan Burgueño y MSc. Enrique Rodríguez del CIMMYT, por los conocimientos compartidos durante la realización de esta investigación.

Al personal técnico de laboratorio y de campo del CIMMYT, especialmente a Aracely Balderas y Gustavo Martínez, por su confianza y valiosa colaboración.

DATOS BIOGRÁFICOS

DAVID OMAR GONZÁLEZ DIÉGUEZ

El autor de la presente tesis nació el 25 de febrero de 1988, en el municipio de Nueva Concepción, Escuintla, Guatemala. En el año 2006 entró a la Facultad de Agronomía de la Tricentenaria Universidad de San Carlos de Guatemala, donde realizó sus estudios de licenciatura, obteniendo el título de Ingeniero Agrónomo en el año 2011. En enero del 2014 ingresó a la Maestría en Ciencias en Biotecnología Agrícola en el Departamento de Fitotecnia de la Universidad Autónoma Chapingo. En diciembre de 2015 obtuvo el título de Maestro en Ciencias en Biotecnología Agrícola con la tesis titulada "IDENTIFICACIÓN DE QTL Y SELECCIÓN GENÓMICA PARA RESISTENCIA AL COMPLEJO MANCHA DE ASFALTO EN VARIEDADES NATIVAS DE MAÍZ"

CONTENIDO

PÁGINA

LISTA DE CUADROS

PÁGINA

LISTA DE FIGURAS

PÁGINA

IDENTIFICACIÓN DE QTL Y SELECCIÓN GENÓMICA PARA RESISTENCIA AL COMPLEJO MANCHA DE ASFALTO EN VARIEDADES NATIVAS DE MAÍZ

RESUMEN

El complejo mancha de asfalto (CMA) en maíz, causado por los hongos *Phyllachora maydis* y *Monographella maydis,* es una enfermedad de importancia económica en zonas tropicales y subtropicales de México, Centroamérica y parte de Sudamérica. El objetivo de este estudio fue identificar fuentes de resistencia al CMA en poblaciones nativas de maíz mediante un Estudio de Asociación del Genoma Completo (GWAS, en inglés) y Selección Genómica (SG), a fin de ampliar la base genética de resistencia a esta enfermedad. Un conjunto de 669 accesiones del Banco de Germoplasma del CIMMYT fueron cruzadas como progenitores masculinos para la obtención de igual número de mestizos, lo cuales fueron evaluados bajo condiciones de infección natural. Las accesiones *per se* fueron genotipadas mediante el método de genotipificación por secuenciación (GBS), obteniéndose 56,092 SNPs de alta calidad. Para el análisis GWAS se evaluaron dos métodos, Single Marker y BayesB. De estos dos, los marcadores identificados con BayesB explican la mayor proporción de varianza genética (49 %). A partir de estos SNPs se logró identificar genes candidatos potencialmente involucrados en el mecanismo de defensa. Siete diferentes enfoques y modelos fueron evaluados en SG; la mayor precisión de predicción se obtuvo utilizando únicamente los marcadores significativos identificados con BayesB ($r_{prueba} = 0.61$). Los resultados demuestran que la resistencia al CMA es de herencia poligénica y comprende QTL de efectos relativamente pequeños que en conjunto explican la mitad de la varianza genética. Los mestizos derivados de las accesiones Guat153 y Oaxa280 fueron los más tolerantes, por lo que éstas podrían ser utilizadas como fuentes de resistencia al CMA.

Palabras clave: *Zea mays;* resistencia genética; estudio de asociación del genoma completo; genotipificación por secuenciación; SNP.

QTL IDENTIFICATION AND GENOMIC SELECTION FOR TAR SPOT DISEASE COMPLEX RESISTANCE IN NATIVE MAIZE LANDRACES

ABSTRACT

Tar Spot Complex (TSC) in maize, caused by the fungi *Phyllachora maydis* and *Monographella maydis*, is a disease with economic importance in tropical and sub-tropical zones of Mexico, Central America and parts of South America. The objective of this study was to identify sources of resistance to TSC in native maize populations through a genome-wide association study (GWAS) and genomic selection (GS), in order to widen the genetic base of resistance to this disease. A set of 669 accessions of CIMMYT's Maize Germplasm Bank were crossed as male parents to obtain the same number of testcrosses that were evaluated under natural infection conditions. The accessions per se were genotyped using the genotype by sequencing (GBS) method, using a total of 56,092 high-quality SNPs. For GWAS, two methods, Single Marker and BayesB approaches, were evaluated. Of these, the markers identified with BayesB explain the greatest proportion of genetic variance (49%). Through these SNPs it was possible to identify candidate genes potentially involved in the defense mechanisms of resistant materials. Seven different approaches and models for GS were evaluated; the greatest prediction accuracy was obtained using only significantly associated markers identified with BayesB ($r_{test} = 0.61$). The results demonstrate that TSC resistance is inherited polygenically, comprising many QTL of relatively small effects that together explain half of the genetic variance. The testcrosses derived from the accessions Guat153 and Oaxa280 were the most tolerant, meaning that these could be used as sources of resistance to TSC.

Key words: *Zea mays,* genetic resistance, genome-wide association study, genotyping by sequencing, SNP.

1 INTRODUCCIÓN

1.1 Antecedentes

El maíz (*Zea mays* L.) es uno de los cereales de mayor importancia mundial por sus diversos usos, principalmente para la alimentación humana y animal, ocupando un lugar preponderante en la agricultura mundial. México figura entre los principales productores de maíz y a la vez es el principal consumidor *per cápita* (González-Rojas *et al.*, 2011). Uno de los factores limitantes en la producción agrícola son las enfermedades, provocando pérdidas de 25 % a nivel nacional. En México se han identificado alrededor de 86 patógenos que atacan al maíz, de los cuales 71 son especies de hongos, 6 especies de bacterias, 7 virus, un espiroplasma y un fitoplasma (Hernández, 1998). Entre las enfermedades foliares, el complejo mancha de asfalto (CMA), inducida por los hongos *Phyllachora maydis* Maubl., *Monographella maydis* Müller y Samuels y *Coniothyrium phyllachorae* Maubl., ha cobrado gran importancia en la última década por su amplia distribución y severidad debida a la interacción sinérgica de los patógenos involucrados, el incremento de la temperatura y la estrecha base genética de los materiales cultivados. Su distribución comprende principalmente zonas tropicales, subtropicales y zonas de transición, y cada vez son más frecuente en zonas de clima templado en México, Centro América y Sudamérica. La reducción en el rendimiento de grano y forraje varía entre 30 y 50 %, e inclusive del 100 %, dependiendo si la enfermedad ataca después o antes de la floración y las condiciones ambientales son favorable para producir una epidemia (Hock *et al.,* 1989).

En México el área afectada supera las 800 mil ha en los estados de Jalisco, Michoacán, Nayarit, Veracruz, Oaxaca, Hidalgo, Chiapas y Guerrero, algunos de los cuales son considerados graneros de México por su volumen de producción. En México uno de los casos más severos ocurrió en el Estado de Oaxaca, en octubre de 2012, donde la enfermedad causó pérdidas de rendimiento entre 70 y 90 % y destruyó en su totalidad algunos cultivos (CIMMYT, 2013). En Centroamérica y el Caribe tiene importancia económica en Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panamá, Puerto Rico, Haití, República Dominicana y Cuba (Hock *et al.*, 1989; CIMMYT, 2013). En América del Sur, se ha reportado en Perú, Ecuador, Colombia, Venezuela y Bolivia,

(McGuire y Crandall, 1967; Arnold, 1986; Hock *et al*., 1989; Bajet *et al.*, 1994). El caso más reciente e inesperado ocurrió en septiembre de 2015, fecha en la que se confirmó la presencia de *Phyllachora maydis* en el Estado de Indiana y en los condados de DeKalb, LaSalle y Bureau, del Estado de Illinos, EUA. Esto fue confirmado por el centro Nacional de Patología de Plantas y por el Departamento de Agricultura de Estados Unidos (USDA). La infección por *P. maydis* no se considera significativa en este momento, sin embargo, la aparición de *Monographella maydis* puede causar daño económico (The Bulletin, 2015).

El desarrollo de materiales resistentes genéticamente a enfermedades es la mejor estrategia para controlar las enfermedades en cultivos de importancia agrícola (White, 1999), motivo por el cual instituciones estatales e internacionales están desarrollando investigaciones en esa línea. Como antecedente, el Instituto Nacional de investigaciones Forestales y Agropecuarias (INIFAP) liberó en 2005 el híbrido H-563, el cual mostró ser el germoplasma más tolerante al complejo mancha de asfalto durante su validación en diferentes localidades, sin embargo, en el ciclo primavera-verano del 2007 mostro severos síntomas de la enfermedad y merma en el rendimiento (González *et al*., 2008). Esto sugiere que el mecanismo de resistencia era tipo vertical (basada en uno o muy pocos genes), la cual tiende a ser poco duradera por su alta especificidad (Allard, 1980; Agrios, 2005).

Una estrategia para lograr una resistencia más durable es ampliar la base genética (Strange y Scott, 2005; Ali y Yan, 2012), lo cual es posible considerando la evidencia aportada por los estudios realizados por Ceballos y Deutsch (1992) y Hernández (2014), quienes identificaron la presencia de efectos dominantes así como un fuerte componente aditivo (55 veces mayor que el efecto dominante), lo que implica que diversos genes están involucrados pudiendo ser acumulados y con el desarrollo de la biotecnología es posible la aplicación de herramientas genómicas para acortar el tiempo de obtención de nuevas variedades de maíz resistentes.

Mediante estudios más avanzados a nivel moleculares, tal como estudios de asociación del genoma completo (GWAS, en inglés) se ha logrado identificar con éxito regiones del genoma o QTL (*loci* de un carácter cuantitativo) asociadas

significativamente a características de importancia agronómica en cultivos de demanda mundial. En maíz se han identificado QTL asociados con la resistencia a la necrosis letal (Gowda *et al*., 2015), carbón de la espiga (Wang *et al*., 2012), enanismo rugoso (Liu *et al*., 2014), tizón sureño y norteño (Kump *et al*., 2011) y otros caracteres como tiempo de floración (Ali y Yan, 2012), arquitectura de hoja (Tian *et al*., 2011) y altura de planta (Weng *et al*., 2011). También se han realizado estudios de selección genómica para enfermedades complejas, tales como el tizón norteño del maíz (Technow *et al.*, 2013), necrosis letal en germoplasma tropical de maíz (Gowda *et al*., 2015), pudrición de la mazorca (Zila, 2014), y la roya en trigo (Ornella *et al*., 2012; Daetwyler *et al.*, 2014), los cuales han demostrado el potencial para ser aplicadas en el mejoramiento de resistencia a enfermedades. Sin embargo, a la fecha ninguno de estos estudios ha sido realizado para investigar la base genética de la resistencia al CMA.

En esta línea, actualmente se desarrolla el proyecto Seeds of Discovery (MasAgro Biodiversidad) que forma parte de la iniciativa Modernización Sustentable de la Agricultura Tradicional (MasAgro), fundada por la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación de México (SAGARPA) y conducida en colaboración por el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), en el cual se están empleando tecnologías de punta para revelar el potencial genético de las colecciones de maíz y trigo como materia prima para el mejoramiento genético. En maíz, uno de los objetivos de Seeds of Discovery es evaluar fenotípica y genotípicamente más de 4,000 accesiones de la colección núcleo del banco de germoplasma del CIMMYT, a fin de identificar alelos favorables para caracteres con base genética cuantitativa y compleja, entre ellos, alelos asociados a la resistencia al CMA y otros caracteres fisiológicos de importancia como tolerancia al calor y a la sequía (Sood *et al.*, 2014).

Es por ello que en convenio entre la Universidad Autónoma Chapingo y el CIMMYT, se plantean los siguientes objetivos para el presente proyecto de investigación:

1.2 Objetivo general

Identificar fuentes de resistencia al CMA en poblaciones nativas de maíz mediante un Estudio de Asociación del Genoma Completo y Selección Genómica, a fin de ampliar la base genética de resistencia a esta enfermedad.

1.3 Objetivos específicos

Identificación de QTL que contribuyan a la resistencia al complejo mancha de asfalto en maíz mediante un estudio de asociación del genoma completo.

Identificar genes candidatos potencialmente involucrados en el mecanismo de defensa de los materiales resistentes al complejo mancha de asfalto en maíz.

Evaluar la precisión de predicción de diferentes enfoques y modelos de selección genómica.

1.4 Hipótesis

La variabilidad genética de las poblaciones nativas de maíz de la Colección Núcleo de Maíz del Banco de Germoplasma del CIMMYT es una alternativa importante para la identificación de genotipos resistentes al complejo mancha de asfalto.

Mediante la realización de un estudio de asociación del genoma completo en un conjunto de variedades nativas de maíz es posible identificar regiones del genoma o QTL (*loci* de un carácter cuantitativo*)* asociados con la resistencia al complejo mancha de asfalto.

Utilizando el enfoque de selección genómica es posible acumular alelos favorables de pequeños efectos que contribuyan a obtener una resistencia más durable contra el complejo mancha de asfalto.

2 REVISIÓN DE LITERATURA

2.1 Complejo mancha de asfalto

El complejo mancha de asfalto (CMA) en maíz se reportó por primera vez en 1904 en México (Maublanc, 1904). Los agentes causales de la enfermedad son tres hongos que interaccionan de forma sinérgica *Phyllachora maydis* Maubl.*, Monographella maydis* Müller y Samuels *y Coniothyrium phyllachorae* Maubl. (Hock *et al*., 1989). La clasificación taxonómica de los hongos asociados al CMA se presenta en el Cuadro 1.

Cuadro 1. Clasificación taxonómica de los hongos asociados al CMA.

Clasificación	*P. maydis*	*M. maydis*	*C. phyllachorae*
Dominio	*Eukaryota*	*Eukaryota*	*Eukaryota*
Reino	*Fungi*	*Fungi*	*Fungi*
Filo	*Ascomycota*	*Ascomycota*	*Ascomycota*
Subfilo	*Pezizomycotina*	*Pezizomycotina*	*Pezizomycotina*
Clase	*Sordariomycetes*	*Sordariomycetes*	*Dothideomycetes*
Subclase	*Sordariomycetidae*	*Xylariomycetidae*	*Pleosporomycetidae*
Orden	*Phyllachorales*	*Xylariales*	*Pleosporales*
Familia	*Phyllachoraceae*	*Not assigned*	*Leptosphaeriaceae*
Genero	*Phyllachora*	*Monographella*	*Coniothyrium*
Especie	*P. maydis*	*M. maydis*	*C. phyllachorae*

Fuente: (MycoBank, 2014; “Species 2000 y ITIS Catalogue of Life”, 2013).

Cada uno de los agentes causales provoca una lesión particular y en conjunto producen la sintomatología característica de la enfermedad. Los síntomas comienzan con la infección de *P. maydis*, cuando la planta tiene entre 8 y 10 hojas, produciendo lesiones pequeñas (2.0 a 5.0 mm de diámetro) de forma oval o circular que se ven como puntos negros abultados de aspecto liso y brillante, distribuidos aleatoriamente en la superficie foliar. Después de dos o tres días, la lesión provocada por *P. maydis* da lugar a la invasión de un segundo patógeno, *M. maydis,* el cual es responsable de la formación de un halo color verde claro de forma oval (1-4 mm) alrededor de la lesión producida por *P. maydis*; la asociación de estos dos hongos resulta en el desarrollo de tejido necrótico, produciendo el síntoma conocido como “ojo de pescado” (Varón y Sarria, 2007; CIMMYT, 2013). Hock *et al.* (1992) indican que el daño más significativo es causado por *M. maydis,* produciendo muerte excesiva del tejido foliar, mientras que el área foliar que corresponde a lesiones de *P. maydis* se mantiene por debajo de 1 %

durante el ciclo de cultivo. Con frecuencia suele observarse en el tejido necrótico un tercer patógeno, *Coniothyrium phyllachorae* Maubl., el cual confiere una textura áspera a la lesión (Müller y Samuels, 1984). Este último hongo es considerado un micoparásito de *P. maydis* y *M. maydis* (Hock *et al*., 1992), por lo que puede considerarse un enemigo natural.

En ausencia de *Phyllachora maydis*, el hongo *Monographella maydis* puede estar en forma epifita colonizando la superficie del tejido foliar en maíz (Ceballos y Deutsch, 1992) o posiblemente endófita (Müller y Samuels, 1984), en ambos casos sin causar ninguna lesión o efecto negativo visible por lo que se considera un patógeno latente (Mostert *et al*., 2000; Maciá-Vicente *et al.*, 2011). Es únicamente en asociación con *P. maydis* que *M. maydis* se torna patogénica y altamente virulenta. La enfermedad generalmente se desarrolla al inicio de la floración en las hojas inferiores, pasando rápidamente a las superiores y a otras plantas. Si la enfermedad aparece previo a la floración, antes del llenado de la mazorca, las pérdidas en el rendimiento pueden ser mayores que el 50 % (CIMMYT, 2013).

Se sabe que *Phyllachora maydis* es un parasito obligado específico del maíz y no se ha encontrado en otras gramíneas u hospederos, incluso en especies del mismo género (*Zea spp.*) (Dittrich *et al.*, 1991), sin embargo, Malaguti y Subero (1972) indican que en los trópicos de Venezuela es sumamente común sobre innumerables huéspedes, especialmente gramíneas. Las ascosporas de *P. maydis* pueden sobrevivir en los restos del cultivo hasta por tres meses (Hock *et al.,* 1995), constituyéndose en fuente de inoculo en las regiones donde es posible realizar más de un ciclo de cultivo al año, sin embargo, se desconoce cómo puede sobrevivir en ausencia de maíz en las regiones donde solo se obtiene una cosecha al año, por lo que es probable que la vegetación espontanea u otras gramíneas sean fuente de ascosporas sin causar daños en estas.

Dada la condición de parásito obligado de *P. maydis, a*ún no se conoce un medio para el cultivo *in vitro* que permita la inoculación artificial (Hock *et al*., 1995), lo cual implica que los estudios de campo dependen de la ocurrencia de la infección natural, lo cual es impredecible y esporádico (Ceballos y Deutsch, 1992).

2.2 Distribución geográfica

El rango de distribución del CMA comprende zonas tropicales, subtropicales y zonas de transición productoras de maíz, con un rango de altitud entre 1,300 a 2,000 metros sobre el nivel del mar (msnm). En México el área afectada supera las 800 mil ha en los estados de Jalisco, Michoacán, Nayarit, Veracruz, Oaxaca, Hidalgo, Chiapas y Guerrero (Hock *et al.*, 1989; CIMMYT, 2013).

En Centroamérica y el Caribe, tiene importancia económica en Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panamá, Puerto Rico, Haití, República Dominicana y Cuba. En América del Sur, se ha reportado en Perú, Ecuador, Colombia, Venezuela y Bolivia (McGuire y Crandall, 1967; Arnold, 1986; Hock *et al*., 1989; Bajet *et al.*, 1994).

El caso más reciente e inesperado ocurrió en septiembre de 2015 en los condados de DeKalb, LaSalle y Bureau del Estado de Illinois y el Estado de Indiana, EUA. El Centro Nacional de Patología de Plantas y el Departamento de Agricultura de Estados Unidos (USDA) confirmaron la presencia de *Phyllachora maydis*, pero en esta temporada no se reportó ninguna pérdida considerable, probablemente por la ausencia de *Monographella maydis.*

2.3 Epidemiología del complejo mancha de asfalto

Se desarrolla en las zonas montañosas, entre 1,300 a 2,300 msnm, donde los ambientes son moderadamente fríos pero bastante húmedos de las regiones tropicales y subtropicales, especialmente en áreas cercanas a las riberas de los ríos, o en suelos con nivel freático alto, pesados o con tendencia al encharcamiento (Malaguti y Subero, 1972; Varón y Sarria, 2007). La epidemia es favorecida por temperaturas promedio mensuales entre los 17 y 22 grados centígrados, con una humedad relativa superior al 75 %. Un mínimo de 7 horas de humedad sobre las hojas durante la noche y en la mañana facilita la infección y el establecimiento de los patógenos (Varón y Sarria, 2007). Si además se presentan de 10 a 20 días de niebla durante el mes, una precipitación mensual mínima de 150 mm y entre 1800 y 1900 horas de luz solar al año, los hongos *P. maydis* y *M. maydis* actúan de forma sinérgica y en menos de ocho días el follaje puede marchitarse por completo (CIMMYT, 2013).

Otros factores asociados que favorecen el desarrollo del complejo mancha de asfalto son altos niveles de fertilización nitrogenada, la siembra de variedades e híbridos de maíz susceptibles; cultivo consecutivo de maíz que proporciona una fuente constante de inóculo, la poca luminosidad y la virulencia de los patógenos causantes de la enfermedad (Pereyda-Hernández *et al*., 2009).

Respecto a su medio de dispersión, Hock *et al.* (1995) consideran que no se dispersa a través de la semilla de maíz, ya que a las ascosporas de *P. maydis* y conidios de *M .maydis* se les dificulta penetrar la testa de la semilla y no se logró detectar ninguno de los patógenos a partir de semillas. Lo más probable es que sean transportados por el viento o la salpicadura de las gotas de lluvia.

2.4 Base genética y molecular del maíz

El maíz se originó en las Américas antes del año 5000 a.C. El gran número de variedades diferentes que existe en la actualidad es el resultado de varios miles de años de selección por los pueblos indígenas y más recientemente por los fitomejoradores. Estas variedades, junto con sus parientes no domesticados constituyen una base genética muy amplia. Como parte de los programas de fitomejoramiento de las compañías privadas aproximadamente entre 500 mil y un millón de híbridos potenciales de maíz son evaluados cada año en Estados Unidos en busca de características agronómicas deseables, tales como resistencia a enfermedades y rendimiento (White, 1999).

La base genética del maíz descrita por White (1999) fue esquematizada en forma de pirámide, donde la base está representada por los parientes silvestres del maíz, ascendiendo por las razas locales, germoplasma exótico/no adaptado, líneas puras y variedades de polinización libre adaptadas, líneas puras élite y sintéticas comerciales y en la cima los híbridos experimentales.

A nivel citogenético, el maíz es una especie diploide con un juego básico de diez cromosomas, sin embargo, ciertas razas de maíz adicionalmente poseen uno o más cromosomas accesorios, pero no son esenciales para el crecimiento y desarrollo normal de la planta (Paliwal *et al*., 2001). También se ha identificado la formación de

nudos cromosómicos, lo cual puede ser una herramienta útil para la clasificación racial y geográfica del maíz y un indicador de diversidad racial (McClintock *et al.,* 1981).

A nivel molecular, el genoma del maíz posee alrededor de 32,000 genes y la mayoría del genoma comprende elementos repetitivos y transponibles (85 %) (McClintock, 1956). La gran diversidad genética del maíz hace posible obtener mapas de alta resolución, pero para ello se requiere aplicar un conjunto denso de marcadores en el genoma completo y análisis sistemático. En cuanto a la estructura del Desequilibrio de Ligamiento (DL) en las especies de polinización cruzada existe una mayor diversidad y rápida reducción del DL en comparación con especies autógamas, y por consiguiente un mayor número de marcadores es requerido para analizar el genoma completo (Flint-García, *et al.* 2003; Romay *et al*., 2013). En este contexto, el maíz es considerado una planta modelo ideal para estudios a nivel molecular por su rápida reducción del DL, especialmente para estudios de asociación del genoma completo (Liu *et al*., 2014).

2.5 Recursos genéticos y pre-mejoramiento

En los últimos 50 años el mejoramiento de los cultivos ha dado lugar a la reducción de la base genética de los materiales cultivados, observándose que en la medida que se obtienen nuevos cultivares con mayor rendimiento mediante sofisticados esquemas de mejoramiento, mayores niveles de vulnerabilidad genética se observa, entendida como una reducción en la capacidad de adaptarse a nuevos cambios ambientales o mayor susceptibilidad a enfermedades, lo cual amenaza la estabilidad y sostenibles del rendimiento potencial (Pritsch, 2001; Sharma *et al.,* 2013).

Respecto a las enfermedades de las plantas, Strange y Scott (2005) proponen que una de las estrategias para reducir la incidencia de enfermedades en los cultivos es promover la diversidad genética. Los genotipos silvestres o nativos constituyen una importante fuente de genes portadores de resistencia útiles para ampliar la base genética de los cultivos, ya que han evolucionado por selección natural. Lamentablemente, los esfuerzos en colecta, identificación, conservación y caracterización de recursos genéticos no han sido aprovechados eficientemente en los programas de mejoramiento, debido principalmente a la dificultad de evaluar colecciones de gran tamaño, por lo que se desconoce el uso potencial de los

materiales colectados. En este sentido es necesario realizar una caracterización de las colectas para identificar su utilidad en los programas de mejoramiento (Pritsch, 2001; Sharma *et al.,* 2013).

Además, existe una importante brecha genética entre los materiales mejorados y los materiales nativos, lo cual hace que la incorporación directa de variabilidad genética desde los materiales nativos a los materiales élite se dificulte, por lo que resulta necesario que previo al mejoramiento se realice un proceso que viabilice la introgresión a los materiales élite, sin que se pierdan los caracteres que lo hacen superior en cuanto a caracteres agronómicos (Pritsch, 2001). Este proceso intermedio se denomina pre-mejoramiento, el cual es el puente entre los recursos genéticos y mejoramiento de plantas, y consiste en viabilizar la incorporación de genes favorables, desde variedades nativas hacia materiales élites, obteniéndose germoplasma puente o intermedios genéticamente valorizados que pueden ser utilizados en los programas de mejoramiento (Pritsch, 2001; Sharma *et al*., 2013). Pritsch (2001) define el objetivo de los programas de pre-mejoramiento como la búsqueda de variabilidad genética en materiales nativos para ampliar la base genética de materiales adaptados para disminuir su vulnerabilidad, además de aportar resistencia genética a enfermedades, acceder a caracteres de interés (calidad y tolerancia a estrés, etc.) y aumento del rendimiento y estabilidad. En este sentido, uno de los factores de éxito de los programas de pre-mejoramiento es entonces la identificación de germoplasma donador de caracteres favorables (Sharma *et al*., 2013).

Basados en el enfoque de ampliación de la base genética y pre-mejoramiento, el Gobierno de México junto con el CIMMYT, han impulsado la iniciativa Modernización Sustentable de la Agricultura Tradicional (MasAgro) que maneja un portafolio de proyectos de gran impacto, entre los cuales está “Seeds of Discovery”*,* cuyo objetivo principal es activar el potencial genético latente en los recursos genéticos de maíz y trigo, para proporcionar a los fitomejoradores un compendio de datos fenotípicos y genotípicos, herramientas y servicios para hacer uso más eficiente del germoplasma nativo, con el fin de desarrollar nuevas variedades de maíz y trigo preparados para enfrentar los efectos del cambio climático (Sood *et al.*, 2014).

En el caso del maíz, uno de los objetivos de Seeds of Discovery es evaluar fenotípica y genotípicamente más de 4,000 accesiones de la colección núcleo del CIMMYT, a fin de identificar alelos de efecto pequeño para caracteres con base genética cuantitativa y compleja, entre ellos, alelos asociados a la resistencia al complejo mancha de asfalto y otros caracteres fisiológicos como tolerancia al calor y a la sequía. De esta manera el proyecto contribuye en cierta medida a incrementar el uso de los recursos genéticos por parte de los mejoradores (Sood *et al.*, 2014).

2.6 Resistencia genética del maíz al complejo mancha de asfalto

En general, la resistencia genética a enfermedades es sin duda el medio más seguro y ampliamente utilizado para controlar las enfermedades del maíz (White, 1999), ya que el control químico encarece los costos de producción y tiene un impacto ambiental negativo (Castaño, 1989; Bajet *et al*., 1994).

La resistencia es la capacidad de la planta para reducir completa o parcialmente el crecimiento y desarrollo del patógeno después que se ha establecido contacto entre el patógeno y el hospedante (Niks *et al.*, 1993). En general se dispone de resistencia vertical y resistencia horizontal (White, 1999). La resistencia vertical o monogénica (atribuida a muy pocos genes) puede manifestarse casi inmune por su alta especificidad a una raza del patógeno. Sin embargo, esta alta especificidad ejerce gran presión de selección sobre el patógeno, forzándolo a mutar, por lo que un simple cambio de base a nivel de nucleótido puede hacer que el patógeno sea más virulento, superando completamente la resistencia (Agrios, 2005; Michelmore *et al.*, 2013). Por otra parte, la resistencia horizontal o poligénica, determinada por varios o muchos genes de efectos aditivos relativamente pequeños o un QTL individual que contenga un grupo de genes de menor efecto (Michelmore *et al*., 2013). Este tipo de resistencia tiende a ser un poco más estable por largo tiempo, ya que es efectiva contra múltiples razas del patógeno, mostrando diferentes niveles de resistencia desde parcial a completa*,* sin embargo*,* también puede ser sensitiva a las condiciones ambientales. Lo ideal sería combinar ambos tipos de resistencia en los cultivares de maíz para mayor efectividad y durabilidad de la resistencia. Sin embargo, la resistencia durable no tiene bases genéticas o moleculares particulares debido a la naturaleza de la resistencia en

la planta y el potencial de evolución del patógeno, haciendo más difícil utilizar la resistencia poligénica en programas de mejoramiento (Michelmore *et al*., 2013). Otra limitante es la identificación de germoplasma portador de genes de resistencia, ya que se requiere evaluar un gran número de accesiones para incrementar la probabilidad de encontrar al menos un material resistente, lo cual incrementa significativamente el costo de evaluación (Pritsch, 2001).

Respecto al mecanismo de herencia de la resistencia al CMA, un primer estudios fue realizado por Ceballos y Deutsch (1992), en el cual identificaron la presencia de un solo gen dominante controlando la resistencia a la enfermedad (resistencia monogénica), sin descartar los efectos aditivos. Con base en este estudio se desarrolló el híbrido H-563 liberado por INIFAP en 2005, el cual desafortunadamente mostró síntomas severos de la enfermedad en ciclos posteriores (González *et al*., 2008), poniendo de ejemplo el inconveniente de la resistencia vertical. Casos como estos han ocurrido a lo largo de la historia del mejoramiento genético (Allard, 1980). Esto también puede atribuirse al hecho de que algunos genes de resistencia de efecto mayor pueden ser sensitivos a las condiciones ambientales, principalmente a la temperatura, lo que cada vez se hace un factor importante a considerar por los efectos del calentamiento global (ej. el gen N en tabaco) (Michelmore *et al*., 2013).

Un segundo estudio más reciente, realizado por Hernández (2014), confirmó los efectos génicos de dominancia y aditivos, pero resalta que los segundos son de mayor importancia (55 veces mayor que el efecto dominante), lo que implica que diversos genes podrían estar involucrados en la herencia de la resistencia y da la posibilidad de que los genes puedan ser acumulados mediante selección recurrente.

Con el desarrollo de la tecnología de marcadores moleculares y herramientas bioinformáticas se espera poder conocer más a detalle la base genética subyacente de muchos caracteres de importancia agronómica, especialmente los mecanismos genéticos de resistencia a enfermedades. Estas tecnologías prometen reducir la dificultad para incorporar resistencia poligénica combinada con genes de efecto mayor en programas de mejoramiento (Visendi *et al.*, 2014).

2.7 Marcadores moleculares y su aplicación en plantas

Los marcadores de ADN son segmentos de ADN con una ubicación física identificable en un cromosoma (*locus*), cuya herencia genética se puede rastrear y presenta polimorfismo dentro de una población (Collard *et al.*, 2005). Tienen la gran ventaja de ser prácticamente ilimitados en número y no ser afectados por factores ambientales o estado de desarrollo del organismo analizado.

Dado que los segmentos del ADN que se encuentran contiguos en un cromosoma tienden a heredarse juntos, los marcadores se utilizan a menudo como formas indirectas de rastrear el patrón hereditario de un gen que todavía no ha sido identificado, pero cuya ubicación aproximada se conoce, razón por la cual han sido ampliamente utilizados en genética humana, animal, vegetal y microbiana. Los marcadores de ADN no representan a los genes causales de un fenotipo por sí mismos, pero actúan como señales de estos (Collard *et al.*, 2005).

En plantas, los marcadores moleculares tienen muchas utilidades, entre ellas: Huella genética como apoyo a registro de variedades, estudio de diversidad genética, relaciones filogenéticas, determinación de niveles de estabilidad genética en procesos *in vitro*, apoyo para selección asistida por marcadores (SAM), caracterización de bancos de germoplasma, registro de parentales, mapeo genético, mapeo por asociación y selección genómica. Numerosos tipos de marcadores moleculares basados en ADN han sido desarrollados y agrupados en tres clases de acuerdo a su método de detección: 1) Detección por hibridación, 2) Detección basada en reacción en cadena de la polimerasa (PCR, por sus siglas en inglés) y 3) basados en secuenciación del ADN (Collard *et al.*, 2005). Los marcadores de ADN más comúnmente utilizados para análisis de QTL son RFLPs, RAPDs, AFLPs, SSR, ISSR y más recientemente los marcadores SNPs.

Un marcador de polimorfismo en un solo nucleótido (SNP, por sus siglas en inglés), basado en secuenciación del ADN, es una variación puntual en la secuencia de ADN que afecta a un solo nucleótido en una posición específica del genoma y son responsables de las variaciones fenotípicas (Kitts y Sherry, 2011).

Los SNPs representan el 90 % de la variación genética total en cualquier organismo y se originan principalmente por mutaciones puntuales o pequeñas inserciones o deleciones (*indels)*, las cuales se han fijado en una parte significativa de la población de una especie (frecuencia ≥1 %), siendo las mutaciones más comunes las transiciones, donde ocurre un cambio de una purina por otra purina (A/G) o una pirimidina a otra pirimidina (C/T) o con menor frecuencia las mutaciones transversas, un intercambio de una purina por una pirimidina o viceversa (A/C, A/T, G/C, G/T) (Gupta *et al.,* 2008). Los SNPs pueden ocurrir en regiones codificantes o en regiones con función reguladora, pero son más frecuentes en regiones no codificantes (Borém y Fritsche-Neto, 2014).

Los SNPs son codominante y bialélicos (dos alelos por locus), por lo que son menos informativos que los SSR multialélicos, pero esta desventaja es compensada por su abundancia y su capacidad de ser utilizado en la genotipificación a gran escala (*ultra-high-throughput genotyping,* en inglés) (Borém y Fritsche-Neto, 2014). La Figura 1 representa una variación en la secuencia de ADN comparada entre tres individuos.

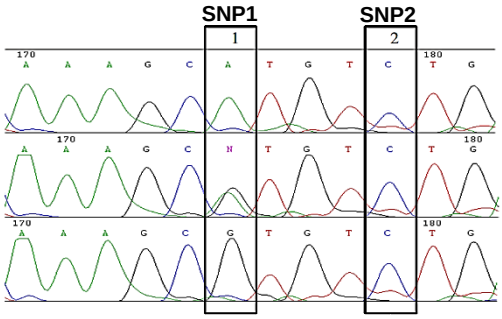

Figura 1. Identificación de SNPs en tres individuos por alineación de las secuencias de ADN. Modificado de Vignal *et al*. (2002).

La ventaja de estos marcadores SNP es que son muy abundantes (1 SNP cada 1000 pares de bases en humanos y 1 cada 300 en maíz) distribuidos a lo largo del genoma por lo que se puede obtener una alta representatividad del genoma (Gupta *et al.*, 2001). Además, se ubican muy cercanos o en el locus de interés, por lo tanto, pueden asociarse precisamente a caracteres como enfermedades. Son muy útiles para la construcción de mapas de ligamiento más densos que los actuales, más estables y lo más importante, pueden ser sujetos a automatización en plataformas de genotipificación (Stram, 2014).

Entre las aplicaciones más comunes están: Construcción de mapas genéticos de ultra alta densidad, selección asistida por marcadores, análisis de la estructura de la población, análisis funcional para identificar posibles implicaciones funcionales de variaciones de la secuencia de ADN (SNP) en la alteración de los transcritos de ARNm; diagnóstico de enfermedades y farmacogenómica; son útiles para estudios de evolución, ya que los SNPs no cambian mucho de una generación a otra, y por ello es sencillo seguir su evolución en estudios de poblaciones; estudios de asociación del genoma completo especialmente para detección de genes de resistencia a una plaga o enfermedad y actualmente son utilizados en selección genómica (Gupta, *et al.*, 2008; Borém y Fritsche-Neto 2014; Stram, 2014).

Con el desarrollo de las tecnologías de secuenciación de alto rendimiento, ha sido posible la detección de miles de SNPs en el genoma completo de varias especies de plantas con una alta eficiencia en costo. Un gran número de plataformas para genotipificación están disponibles y algunas aún se encuentra en fase de desarrollo, capaces de secuenciar millones de fragmentos de ADN por ensayo, superior al método convencional Sanger (Gupta *et al.*, 2008; Ganal *et al*., 2014). Algunas plataformas comerciales son: GoldenGate (Illumina), Infinium (Illumina), SNPstream (Beckman Coulter), GeneChip (Affymetrix), Perlegen Wafers y Molecular Inversion Probe – MIP (Affymetrix), descritas ampliamente por Fan *et al.* (2006) y Syvänen (2005). Estas plataformas varían en la reacción para discriminación de alelos, en la cantidad de muestras y SNPs que se pueden analizar por ensayo (multiplexación), y eficiencia costo. La Figura 2 muestra el nivel de multiplexación de algunas plataforma, es decir, la relación entre número de muestras y SNPs por ensayo.

Según Borém y Fritsche-Neto (2014) algunos factores que deben ser considerados al momento de elegir la plataforma de genotipificación para un proyecto, entre ellos: sensibilidad, reproducibilidad, precisión, capacidad de multiplexación a gran escala, costo de la inversión inicial en equipo y por “data point” y flexibilidad de la técnica para poder ser utilizada en otras aplicaciones.

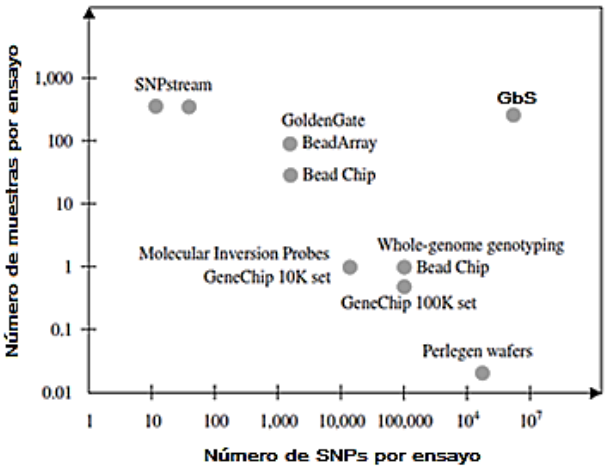

Figura 2. Comparación de niveles de multiplexación en sistemas de genotipificación de SNPs basadas en micro-arreglos. Modificado de Syvänen (2005).

Recientemente un método robusto y simple ha sido desarrollado para alta diversidad en plantas, capaz de analizar un gran número de muestras por ensayo y descubrir millones de SNPs. Este método se conoce como Genotipificación por Secuenciación (GBS, por sus siglas en inglés) y se describe a continuación.

2.8 Genotipificación por secuenciación (GBS)

El método de Genotipificación por Secuenciación (GBS, en inglés), desarrollado por Elshire *et al*. (2011) especialmente para plantas, es un sistema altamente multiplexado que genera cientos a cientos de miles de marcadores tipo SNPs en un gran número de muestras (Figura 2) a partir de representaciones genómicas reducidas y basado en la plataforma de secuenciación de última generación. La técnica de GBS se basa en el uso de enzimas de restricción tipo II para reducir la complejidad del genoma y de esta forma producir la biblioteca de fragmentos que será secuenciada. Sin embargo, es importante recalcar que el éxito de la reducción de complejidad depende en gran medida de la combinación de enzimas de restricción que se seleccionan para cada organismo.

GBS es simple, rápida, extremadamente específica, altamente reproducible, y puede captar importantes regiones del genoma que son inaccesibles para captura de secuencias. Mediante el uso de enzimas de restricción (ERs) sensibles a la metilación, las regiones repetitivas de genomas pueden ser evitadas y regiones menos repetidas

pueden ser secuenciadas con dos a tres veces mayor eficiencia, pudiendo alcanzar regiones del genoma que no son fáciles de estudiar utilizando otros métodos (Elshire *et al*., 2011). A continuación se describe en términos muy generales la metodología empleada, la cual es ampliamente detallada y discutida por Elshire *et al*. (2011).

2.8.1 Extracción del ADN genómico

Parte de la recolección del tejido con el cual se va a trabajar, preferentemente fresco y libre de enfermedades y plagas. Luego prosigue la extracción del ADN genómico, para lo cual existen varios protocolos, entre ellos el protocolo estándar CTAB (Doyle, 1987). En general se divide en tres etapas: Ruptura de tejido, eliminación de contaminantes y precipitación del ADN. Finalmente debe realizarse un punto de control para verificar la calidad del ADN genómico obtenido, mediante la medición de la actividad de las nucleasas presentes en la muestra, así como la verificación en gel de agarosa.

2.8.2 Preparación del ADN previo a secuenciar

Reducción de complejidad genómica (digestión/ligación)

Esto se logra mediante la digestión del ADN genómico con enzimas de restricción tipo II (sensibles a la metilación) para recortar el genoma y la ligación de adaptadores para identificación de las muestras y generar la biblioteca de fragmentos que será secuenciada. El uso de este tipo de enzimas hace que el método sea rápido, sencillo, altamente específico, reproducible, y la sensibilidad a la metilación permite alcanzar regiones del genoma difíciles de estudiar con otros métodos (Elshire *et al*., 2011).

Generalmente se utiliza una enzima de corte frecuente y una de corte especifico. La enzima de corte frecuente más utilizada es *PstI* (C_TGCA^G), la cual se combina con otra(s) enzimas de corte especifico dependiendo de la especie y el objetivo del estudio (Ej. *NspI* para estudios de diversidad y *HpaII* para selección genómica). La selección de la enzima es importante para lograr una mejor reducción de complejidad. En el caso del maíz, una de las enzimas más adecuada es la *Ape*KI, un tipo de endonucleasa de restricción de corte frecuente que reconoce una secuencia degenerada de 5 pares de bases (pb) (GCWGC, donde W puede ser una A o T) y es parcialmente sensitiva a la

metilación y poco reconocimiento para la mayoría de retrotransposones en maíz (Elshire *et al*., 2011). Junto con las enzimas de restricción debe agregarse una ligasa, la cual se encarga de unir un par de adaptadores los cuales permitirán el manejo de las muestras a lo largo del proceso (identificación) y amplificación (común) (Elshire *et al*., 2011).

Dos tipos de adaptadores se utilizan en este protocolo (Elshire *et al*., 2011): los identificadores o "barcode" (secuencias de 4 – 9 pb), que tiene la función de identificar las muestras y diferenciarlas al momento de realizar la mezcla de todas las muestras y debe contener la secuencia afín a la enzima de corte frecuente. Esto es necesario cuando se tienen diferentes especies o genotipos, en un mismo carril de la celda de secuenciación. El segundo tipo de adaptador llamado "común" contiene una secuencia afín al sitio de corte de la enzima secundaria (específica para la especie). Los dos tipos de adaptadores tienen secuencias complementarias para los iniciadores que se utilizan en la amplificación selectiva. La ligación de los adaptadores es facilitada por una ligasa en la misma reacción de digestión del ADN. Finalmente los adaptadores permitirán el manejo de las muestras a lo largo del proceso de identificación y amplificación (Elshire *et al*., 2011).

Amplificación selectiva

La amplificación de estos fragmentos (targets, en inglés) ocurrirá gracias al uso de cebadores (primers, en inglés) complementarios a los adaptadores utilizados en la digestión/ligación. Sólo los fragmentos que tengan ambos adaptadores serán amplificados. Además estos cebadores también contienen secuencias adicionales en su extremo 5' para hacerlos compatibles con fragmentos únicos de secuenciación (Elshire *et al*., 2011).

Revisión de la calidad de fragmentos amplificados

Para revisar la calidad de los fragmentos amplificados se realiza un gel de agarosa, en el cual se busca obtener un pequeño barrido de ADN. Los barridos deben ser uniformes en intensidad, tamaño y migración en el gel si son de la misma especie y si se utiliza el mismo método de reducción de complejidad. Si se observan bandas

definidas y dímeros de cebadores (de aproximadamente 200 pb de longitud) pondrá en duda la calidad obtenida (Elshire *et al*., 2011).

Mezclado, purificación y cuantificación

Se prosigue con la mezcla de todos los fragmentos o *targets* en un solo tubo para simplificar su manejo y finalmente se purifican con un kit comercial. La mezcla es posible gracias al uso de los adaptadores "barcode". Luego se vuelve a verificar en un gel de agarosa esperando observar barridos uniformes. Una vez combinados y purificados, se realiza la cuantificación para medir la cantidad de ADN necesario para la secuenciación, mediante equipo especializado como nanodrop, lectores de placas o espectrofotómetros (Elshire *et al*., 2011).

2.8.3 Secuenciación en plataforma Illumina

Previo a la secuenciación en el *HiSeq 2500* debe prepararse la celda de secuenciación (flowcell, en inglés*),* el cual es un portaobjetos con un canal dividido en 8 carriles, los cuales contienen cebadores específicos a los que se unen los fragmentos de ADN durante la reacción de amplificación de puente (bridge-PCR). Estos fragmentos formarán cúmulos concentrados llamados "clusters" con una misma secuencia, lo cuales son el material de lectura para el secuenciador (Elshire *et al*., 2011).

Generación de clusters en cBot Illumina

El cBot es el equipo especializado en preparar la celda de secuenciación que contiene las muestras que serán secuenciadas. Previo a este paso las muestras de ADN purificado deben pasar por tres pasos: (a) Dilución: la concentración de ADN determina si se forman muchos o pocos clusters, por lo cual se diluyen las muestras para ajustarlas a una misma concentración que forma una cantidad intermedia de clusters; (b) Desnaturalización, consiste en la separación de las cadenas de ADN, formando ADN de cadena sencilla (ssDNA en inglés). Luego de desnaturalizar el ADN es necesario volver a diluir las muestras utilizando un buffer de hibridación. Finalmente se produce la (c) hibridación o amplificación de puente (bridge-PCR) (Elshire *et al*., 2011).

Secuenciación por síntesis

Luego de la generación de los clusters en el cBot, la celda de secuenciación esta lista para ser introducida en el secuenciador *HiSeq 2500* de Illumina, utilizando una serie de reactivos provistos por el mismo fabricante. Este equipo tiene la capacidad de correr entre 2 a 8 placas de 96 muestras por celda de secuenciación. El HiSeq 2500 se basa en tecnología de secuenciación por síntesis (SBS, por sus siglas en inglés) de Illumina, mediante la reacción química denominada Terminación Reversible Cíclica (CTR por sus siglas en inglés), la cual incorpora nucleótidos marcados fluorescentemente y luego toma fotografías de los clusters que posteriormente son procesadas para determinar la secuencia de los fragmentos (Elshire *et al*., 2011).

2.8.4 Obtención y análisis de los datos

Verificación de la calidad de datos

La calidad técnica de los datos obtenidos se revisa utilizando el software Illumina Sequencer Analysis Viewer (SAV). Este software permite ver las métricas de calidad generadas por el Analizador en Tiempo Real del secuenciador (RTA por sus siglas en inglés) (Elshire *et al*., 2011).

Selección de datos e identificación de SNPs

Para la identificación de los marcadores SNPs se utiliza un Sistema de Gestión de Información de Laboratorio (LIMS: Laboratory Information Management System en inglés). Este sistema permite conservar los marcadores de buena calidad y desechar las lecturas que no cumplen con los estándares mínimos. Posteriormente, segrega las secuencias con base a los barcode para que sea posible el análisis de cada muestra por separado. Una vez producidos los marcadores, un investigador capacitado debe revisar manualmente los resultados para seleccionar únicamente los mejores marcadores con base en su alineamiento a un genoma de referencia así como la cantidad de repeticiones del mismo marcador que se obtuvo a lo largo del experimento (Elshire *et al*., 2011).

2.8.5 Ventajas, limitaciones y aplicaciones del método GBS

Entre las principales ventajas del GBS están: Los polimorfismos de ADN son identificados y clasificados al mismo tiempo, lo cual elimina completamente la necesidad de un costoso desarrollo y validación de marcadores. Es escalable, lo cual permite ajustar la densidad de marcadores, el contenido de información por marcador (capacidad de calificar heterocigotos) y el nivel de multiplexación de las muestras. GBS es un enfoque de genotipificación compatible con futuras plataformas y el costo por muestra varía entre 30 y 45 $USD y se prevé que seguirá bajando (Comunicación personal[3]). La principal limitante del método GBS es que los fragmentos (etiquetas) individuales de ADN en una representación genómica se amplifican de forma desigual, lo que tiene como resultado una distribución en "forma de L" con una larga cola de fragmentos de baja frecuencia que dan lugar a un número elevado de datos faltantes (Elshire *et al*., 2011).

Entre las principales aplicaciones del GBS están: caracterización exhaustiva de germoplasma y estudios de diversidad, identificación genética, pureza de semillas/test de calidad de productos, conservación biológica para determinar la estructura de la población, mapeo genético y de QTL, mapeo de asociación, introgresión acelerada de germoplasma salvaje, selección genómica (SG) y Estudios de Asociación de Genomas Completos (GWAS, por sus siglas en inglés) (Elshire *et al*., 2011).

2.9 Aplicación de marcadores moleculares en pre-mejoramiento

2.9.1 Identificación y aplicación de QTL

El mapeo de QTL es un proceso que consiste en la identificación de marcadores moleculares o *loci* genómico que influencian la variación de un carácter cuantitativo (Yi y Xu, 2008). El método clásico para la identificación de QTL ha sido el mapeo de ligamiento, el cual se basa en el hecho de que los marcadores cercanos o fuertemente ligados a los genes son segregados y heredados juntos a la descendencia más frecuentemente que los marcadores que están más alejados o débilmente ligados a los genes (Collard *et al.*, 2005). El mapeo de ligamiento se basa en la formación de

[3] Sarah Hearne, comunicación personal (2014). En reunión sobre mejoramiento molecular para resistencia al complejo mancha de asfalto en el proyecto MasAgro Biodiversidad, CIMMYT. México, D.F.

poblaciones biparentales segregantes (ej. F_2 o Retrocruza), a partir de las cuales se calculan las frecuencias de recombinación para estimar la distancia genética entre marcadores (Collard *et al.*, 2005).

Aunque se dice que encontrar un gen o QTL dentro del genoma de una planta es como buscar una aguja en un pajar, esto es posible partiendo de la detección de una asociación entre el fenotipo y el genotipo de un marcador y determinar si existe diferencia significativa entre las medias fenotípicas de los grupos (agrupados con base a su genotipo para un marcador determinado) (Collard *et al.*, 2005).

Existen varios métodos de detección de QTL, entre ellos: “Single Marker”, intervalo de mapeo y modelos bayesianos. De estos, el método Single Marker es el más sencillo, el cual incluye pruebas de significancia, análisis de la varianza y regresión lineal, a partir de la cual es posible obtener el coeficiente de determinación r^2 del marcador, el cual explica la variación fenotípica derivada del QTL fuertemente ligado al marcador (menos de 20 centiMorgan) (Collard *et al.*, 2005). La principal desventaja de este método es que entre más separado este el QTL del marcador (>20 centiMorgan), es menos probable que sea detectado debido a que puede ocurrir recombinación entre el marcador y el QTL (Tanksley, 1993). Sin embargo, con el desarrollo de tecnologías de genotipificación de alto rendimiento se ha logrado incrementar el número de marcadores, con lo cual es posible explorar mayor parte del genoma y es posible estar más cerca de los genes asociados a un carácter. Sin embargo, existe el inconveniente de que varios marcadores pueden estar en desequilibrio de ligamiento y por consiguiente estar correlacionados, por lo que se requiere mayor precisión para la detección de QTL (Berg *et al*., 2013). Por otra parte, considerando que los caracteres cuantitativos son influenciados por varios QTL a la vez, se esperaría que los modelos bayesianos que analizan simultáneamente todos los marcadores fueran más precisos en la detección de QTL y en la estimación de los efectos en comparación con los modelos que analizan sólo uno o pocos marcadores a la vez (Berg *et al.*, 2013).

Los estudios de asociación a nivel genómico, denominados Estudios de Asociación del Genoma Completo (GWAS, por sus siglas en inglés), han cobrado gran importancia

en estudios de enfermedades en humanos (Corvin *et al.*, 2010). Este tipo de estudios también ha sido aplicado en plantas para el estudio de caracteres de complejidad intermedia. Más recientemente se ha desarrollado el enfoque de Selección Genómica (ver inciso 2.9.2) para predicciones de los valores genéticos de caracteres de mayor complejidad (Hamblin *et al.,* 2011). Una vez identificado un conjunto de QTL ligados fuertemente a un marcador que explica una proporción significativa de varianza genética, y luego de ser validado es posible utilizarlo para selección asistida por marcadores (SAM) o en selección recurrente asistida por marcadores (SRAM) para conseguir resistencias más durable mediante la acumulación de múltiples QTL de pequeño y mediano efecto (Collard *et al*., 2005). La principal ventaja de la SAM es que no se requiere de la evaluación fenotípica durante el ciclo de selección (Heffner *et al.*, 2009).

Estudio de asociación del genoma completo (GWAS)

El primer análisis GWAS se publicó en el año 2005 y se realizó en humanos (Klein *et al*., 2005). Posteriormente se realizaron muchos análisis de este tipo para el estudio de la predisposición a enfermedades en humanos, pero también se ha ampliado su uso a otros organismos, incluyendo las plantas, en las cuales se ha aplicado con éxito. Este tipo de estudio consiste en analizar toda la variabilidad existente a lo largo del genoma en función de los polimorfismos que existen a nivel de nucleótido (SNP) y determinar su asociación a un carácter de interés. En este sentido el análisis GWAS constituye una poderosa herramienta para identificar *loci* de un carácter cuantitativo responsables de la variación fenotípica de tal carácter (Ali y Yan, 2012).

Un requisito fundamental para este tipo de estudios es el uso de un conjunto denso de marcadores moleculares (desde cientos de miles hasta millones de SNPs) para explorar la máxima cantidad de polimorfismos en el genoma completo (Waugh *et al*., 2014), lo cual es ahora posible gracias a las nuevas tecnologías de genotipificación de alto rendimiento (ej. GBS). Por otra parte, ya que la detección se basa en las frecuencias de los alelos, se requiere de grandes tamaños de muestra que van desde cientos a decenas de miles para poder detectar aquellos alelos de baja frecuencia ligados a un QTL (Zhang *et al*., 2010; Stram, 2014). Es importante resaltar que se

requiere realizar un control de calidad de los marcadores SNP antes de realizar el análisis estadístico, ya que sí no se hace se podría incrementar la tasa de falsos positivos (Corvin *et al.*, 2010).

Dada la disponibilidad de un gran número de marcadores, el método de identificación de QTL "Single Marker" ha cobrado mayor importancia y está siendo ampliamente utilizado en los estudios de asociación del genoma completo (GWAS). Este método consiste en realizar un análisis de regresión por cada marcador para detectar si tiene efecto significativo sobre el carácter de interés (Buntjer *et al.,* 2005), de tal forma que se realizan tantas regresiones y pruebas de hipótesis como marcadores existan. Últimamente se han implementado modelos lineales mixtos que incluyen control por estructura de población y relaciones de parentesco para reducir la tasa de falsos positivos o error tipo I (Yu *et al*., 2006), esto es rechazar Ho siendo verdadera, o en otras palabras, concluir que un marcador si tiene efecto significativo sobre el carácter cuando en realidad no lo tiene.

Modelos bayesianos con capacidad para selección de variables, que fueron desarrollados originalmente para selección genómica (Meuwissen *et al.*, 2001), han mostrado potencial para la estimación de los efectos de los marcadores y están siendo utilizados para detección de QTL (Berg *et al*., 2013). Estos modelos implementan una regresión múltiple con distribuciones *a priori* mixtas, permitiendo que los marcadores con efectos relevantes tomen valores diferentes de cero, mientras que los marcadores con efectos espurios (falsos positivos debido a covariables redundantes) sean contraídos hacia cero (Mutshinda y Sillanpää, 2010). Entre ellos se puede mencionar los modelos LASSO, BayesB y BayesC. En varios estudio de simulación, estos modelos han demostraron mayor precisión para detección de QTL y con menos falsos positivos que el método Single Marker con enfoque de modelo mixto, especialmente para caracteres con alta heredabilidad y QTL con grandes efectos (Yi y Xu, 2008; Sahana *et al.*, 2010; Zeng *et al*., 2012; Berg *et al.*, 2013; Gondro *et al.*, 2013; Pérez-Rodríguez y de los Campos, 2014). Sin embargo, ninguno de estos métodos ha sido evaluado con datos reales.

Una ventaja de los análisis GWAS comparado con el método tradicional de mapeo de ligamiento (basado en frecuencia de recombinación) es que no es necesario formar poblaciones bi-parentales de individuos relacionados directamente y tampoco se requiere de información de pedigree o realizar cruzamientos, ya que los individuos siempre están relacionados en alguna distancia genéticamente, por lo cual es posible buscar asociación para un carácter fenotípico en una población diversa de organismos no relacionados directamente (Hamblin *et al*., 2011; Wang *et al*., 2012). Esto permite explorar mayor diversidad y encontrar diversos alelos favorables a través de un conjunto de poblaciones altamente diversas entre sí (ej. variedades nativas de maíz). Otra ventaja del GWAS es que se pueden identificar alelos favorables que controlan caracteres de interés con gran resolución y sensibilidad (Yu y Buckler, 2006). Sin embargo, hay casos en que los QTL detectados en las poblaciones de mapeo bi-parentales no son detectados por GWAS, especialmente aquellos alelos de baja frecuencia y es por ello que se recomienda tamaños de muestra grandes (Weng *et al*., 2011).

Inicialmente el potencial del GWAS fue demostrado en plantas por Aranzana *et al*. (2005) en *Arabidopsis,* quienes lograron identificar genes que controlan la variación natural en el tiempo de floración y resistencia a patógenos en una muestra de 95 accesiones, los cuales ya habían sido identificados previamente por el tradicional de mapeo de ligamiento. El poder del GWAS también fue demostrado usando 250,000 SNPs para analizar 107 fenotipos en *Arabidopsis* (Atwell *et al*., 2010). Una vez aplicado con éxito en *Arabidopsis,* se aplicó a otros cultivos de importancia económica, tales como cebada, maíz, arroz, sorgo y soya (Sukumaran y Yu, 2014).

En maíz, varios estudios de asociación de genoma completo han sido realizados con éxito. Weng *et al*. (2011) identificaron 105 loci genómicos que controlan la altura de planta utilizando un conjunto de 284 líneas puras y un juego de 41,101 SNPs. Wang *et al*. (2012) identificaron QTL asociados con la resistencia al carbón de la espiga y posteriormente Weng *et al*. (2012) identificaron y mapearon finamente el principal QTL *aHS2.09* de resistencia al carbón de la espiga a un intervalo de 1.0 Mb mediante la combinación de mapeo de ligamiento y asociación. Liu *et al*. (2014) identificaron un

total de 73 SNPs asociados a la resistencia al enanismo rugoso utilizando un total de 296 líneas puras y 41,101 marcadores SNPs. De igual forma se identificaron genes de resistencia al tizón sureño y norteño del maíz (Kump *et al*., 2011; Poland *et al.*, 2011) y otros caracteres como tiempo de floración (Ali y Yan, 2012) y arquitectura de hoja (Tian *et al*., 2011).

Paralelo al desarrollo de las tecnologías de próxima generación de genotipificación de alto rendimiento, el desarrollo de herramientas bioinformáticas para el análisis de la información ha ido avanzando. Actualmente están disponibles un gran número de programas especializados para realizar este tipo de análisis, entre ellos TASSEL, GAPIT, PLINK, GEMMA (Bradbury *et al*., 2007; Purcell *et al*., 2007; Lipka *et al*., 2012; Zhou y Stephens, 2012), los cuales implementan modelos mixtos para controlar la estructura de población y las relaciones de parentesco. También puede utilizarse programas estadísticos como R (R-Core Team, 2015) para la implementación de otros modelos con enfoque bayesiano con inducción de selección de variables.

2.9.2 Selección genómica

La selección genómica (SG) permite superar la limitante de la SAM basada en la identificación de QTL, con la que solo es posible capturar una limitada proporción de la variación genética y su impacto práctico en programas de mejoramiento ha sido menor a lo esperado (de los Campos *et al.*, 2013). El enfoque de SG, desarrollado originalmente por Meuwissen *et al*. (2001), es ahora posible gracias a la disponibilidad de tecnologías de genotipificación de alto rendimiento, y se constituye en una alternativa prometedora que permite capturar mayor variación genética mediante la acumulación de alelos favorables a lo largo de todo el genoma completo utilizando todos los marcadores disponibles (Ornella *et al*., 2012).

Mediante SG es posible estimar los Valores Genéticos (Genomic Estimated Breeding Values -GEBVs-, en inglés) para caracteres cuantitativos utilizando todos los marcadores como predictores y sumando el efecto de cada uno de ellos mediante una regresión del fenotipo sobre todos los marcadores disponible (Meuwissen *et al*., 2001). La ganancia genética en SG es linealmente proporcional a la precisión de la predicción del modelo (Daetwyler *et al*., 2014), por lo tanto, es importante evaluar distintos

modelos y enfoques para determinar con cuál se obtiene mayor precisión. Numerosos modelos han sido propuestos para la obtención de los GEBVs (Meuwissen *et al*., 2001; Gianola *et al.*, 2006; Yi y Xu, 2008; Zhe Zhang *et al*., 2010).

Para el caso del maíz, ya se han realizado algunos ensayos para evaluar la viabilidad de la selección genómica. Diferentes modelos han sido evaluados para SG en maíz para caracteres con diferente base genética (Riedelsheimer *et al.*, 2012). Crossa *et al*. (2013) evaluaron distintos modelos utilizando datos genotípicos generados con la tecnología de Genotipificación por Secuenciación (GBS) para comparar la precisión de las predicciones en rendimiento de grano, días a floración masculina y femenina en maíz. También se ha aplicado SG para enfermedades complejas, tales como el tizón norteño del maíz (Technow *et al.*, 2013), necrosis letal en germoplasma tropical de maíz (Gowda *et al*., 2015), pudrición de la mazorca (Zila, 2014), y la roya en trigo (Ornella *et al*., 2012; Daetwyler *et al.*, 2014), los cuales han demostrado su uso potencial en mejoramiento de resistencia cuantitativa a enfermedades.

El procedimiento para realizar selección genómica fue originalmente descrito por Meuwissen *et al.* (2001). En general consiste en lo siguiente: 1) obtención de la población de entrenamiento, 2) entrenamiento del modelo, 3) validación del modelo, 4) obtención de los GEBVs para toda la población de mejoramiento a partir de su información genotípica y finalmente se realiza el 5) “ranking” para seleccionar los individuos con mayor GEBVs (Figura 3).

Figura 3. Etapas en la selección genómica. Modificado de Meuwissen *et al.* (2001).

Obtención de la población de entrenamiento

Se parte de una población de mejoramiento, la cual debe ser genotipada completamente y a partir de la cual se obtiene una muestra representativa de individuos que sirve para el entrenamiento del modelo. A esta muestra se le llama

población de entrenamiento. Únicamente los individuos muestreados deben ser fenotipados para los diferentes caracteres de interés (Meuwissen *et al*., 2001; Borém y Fritsche-Neto, 2014).

El propósito de la población de entrenamiento es utilizar tanto los datos fenotípicos y genotípicos para entrenar un modelo para obtener los GEBVs del resto de la población de mejoramiento. En este sentido, es importante resaltar que la población de entrenamiento debe ser representativa de la población de mejoramiento. Para lograrlo se puede realizar un muestreo simple aleatorio. De lo contrario, el modelo estará sesgado y sus resultados favorecerán a los organismos más parecidos a la población de entrenamiento (Heffner *et al.,* 2009). En cuanto al tamaño de la población de entrenamiento, al menos se requiere una población entre 500 y 1000 individuos seleccionadas (Borém y Fritsche-Neto, 2014).

Entrenamiento del modelo de selección genómica

En los modelos genéticos estándar los valores fenotípicos y_i $(i = 1, \dots, n)$ son vistos como la suma de los valores genéticos g_i y un residuo ε_i, por lo que el modelo puede ser escrito como $y_i = g_i + \varepsilon_i$. En los modelos de SG g_i es descrito como la suma de todos los efectos de los marcadores (covariables) x_{ij} $(j = 1, \dots, p)$, donde p es igual al número de marcadores moleculares, de tal manera que $g_i = \sum_{j=1}^{p} x_{ij}\beta_j$ (Crossa *et al*., 2010). El modelo lineal puede expresarse de la siguiente manera:

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i \qquad (1)$$

En notación matricial sería $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$. Donde $\boldsymbol{y}$ es el vector de fenotipos, $\boldsymbol{X}$ la matriz de marcadores, $\boldsymbol{\beta}$ es el vector de efectos de los marcadores y $\boldsymbol{\varepsilon}$ es el vector de residuales aleatorios. La estimación de $\boldsymbol{\beta}$ por mínimos cuadrados ordinarios produce el estimador: $\widehat{\boldsymbol{\beta}} = [\boldsymbol{X}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{y}$.

Sin embargo, con el desarrollo de las nuevas tecnologías de genotipificación de alto rendimiento, el número de marcadores (p) utilizados en modelos de predicción genómica excede el tamaño de muestra (n), esto es $p \gg n$, lo cual tiene las siguientes implicaciones:

i. $\boldsymbol{X'X}$ es singular y no es posible obtener la matriz inversa $[\boldsymbol{X'X}]^{-1}$.
ii. Es posible obtener una matriz inversa generalizada $[\boldsymbol{X'X}]^{-}$, sin embargo, como la varianza del estimador es muy alta, se producen predicciones poco precisas.
iii. Se produce un sobre ajuste del modelo (overfitting, en inglés).

Por consiguiente, otros métodos de estimación son requeridos en selección genómica. Las alternativas más comúnmente utilizados son métodos penalizados y de selección de variables tales como regresión ridge y Least Absolute Shrinkage and Selection Operator (LASSO, por sus siglas en inglés) o su contraparte bayesiana (Crossa *et al*., 2010).

La idea principal de los modelos penalizados es reducir el cuadrado medio del error mediante la reducción de la varianza del estimador, aun a expensas de introducir sesgo. En regresión ridge (RR) esto se logra simplemente agregando una constante a la diagonal de la matriz de coeficientes, $\widehat{\boldsymbol{\beta}}_{RR} = [\boldsymbol{X'X} + \lambda \boldsymbol{D}]^{-1}\boldsymbol{X'y}$, donde $\lambda > 0$ es un parámetro de regularización y $\boldsymbol{D}$ es una matriz diagonal con cero en su primer entrada y unos en el resto de las entradas (es decir, $d_1 = 0$, esto para evitar sesgar la estimación del intercepto) (Hoerl y Kennard, 1970). Esto vuelve a $[\boldsymbol{X'X}]$ una matriz no singular y es posible obtener la inversa. Esto induce sesgo pero reduce la varianza del estimador y mejora la precisión de las predicciones y reduce el sobre ajuste.

El grado de contracción en la regresión ridge es homogéneo para todos los marcadores, lo cual podría no ser apropiado si algunos marcadores no están asociados a varianza genética o por el contrario, si están ligados a QTL. Este problema puede ser resuelto mediante enfoque bayesiano utilizando distribuciones mixtas como *a priori* para los efectos de los marcadores, tales como BayesA, BayesB y LASSO (Meuwissen *et al*., 2001). Otra alternativa es el uso de métodos semiparamétricos como reproducing kernel Hilbert spaces (RKHS, por sus siglas en inglés) (de los Campos y Gianola, 2010).

Población de validación o simulación

Esta población se obtiene igualmente a partir de la población de mejoramiento, tomando una pequeña muestra de entre 100 a 200 individuos diferentes a los tomados

en la población de entrenamiento. Estos individuos deben ser evaluados en campo para obtener los datos fenotípicos. A partir de los efectos de los marcadores estimados de la población de entrenamiento se obtienen los GEBVs y se predicen los fenotipos de la población de validación. Finalmente, la precisión de las predicciones se evalúa mediante la correlación entre los datos fenotípicos observados y los predichos con el modelo (Meuwissen *et al*., 2001; Borém y Fritsche-Neto, 2014).

Otra manera de validar los modelos es mediante simulación. Varios métodos han sido desarrollados para este fin, entre ellos: una sola partición entrenamiento-prueba, múltiples particiones entrenamiento-prueba y validación cruzada (Pérez-Rodríguez y de los Campos, 2014).

Población de selección o de mejoramiento

Una vez que se han estimado los efectos de los marcadores mediante el entrenamiento del modelo y validado, se procede a estimar los valores genéticos del resto de la población de mejoramiento a partir únicamente de su información genotípica, por lo que no es necesario realizar el fenotipado. Se procede a realizar un "ranking" de los GEBVs para seleccionar los individuos más sobresalientes de la población (Meuwissen *et al*., 2001; Borém y Fritsche-Neto, 2014).

Factores a considerar en la selección genómica

Los principales factores que afectan la efectividad de la selección genómica en función de la precisión de las predicciones son: la base genética del carácter, el desequilibrio de ligamiento existente en la población de mejoramiento, el tamaño de la población de entrenamiento, el tipo de marcador, la densidad de marcadores y la relación de parentesco entre la población de entrenamiento y la población de mejoramiento (Nakaya y Isobe, 2012; de los Campos *et al.*, 2013).

En cuanto a la base genética del carácter, la precisión será mayor para el modelo que mejor se ajuste a la base genética del carácter (de los Campos *et al*., 2013); los modelos de selección de variable (ej. BayesB, BayesC, LASSO) se ajustan mejor a caracteres controlados por pocos QTL, mientras que los modelos penalizados (ej. GBLUP) se ajustan mejor a caracteres complejos con gran número de QTL de efectos

muy pequeños (de los Campos *et al*., 2013). Además, existe una asociación positiva entre la heredabilidad y la precisión de predicción; a mayor heredabilidad, mayor precisión de predicción (Ornella *et al*., 2012; Gowda *et al*., 2015).

La densidad de marcadores está en función del desequilibrio de ligamiento existente en la población de entrenamiento; A menor desequilibrio de ligamiento, mayor densidad de marcadores es necesario para cubrir el genoma y por consiguiente un mayor número de marcadores es requerido (Nakaya y Isobe, 2012). Las especies de polinización cruzada tienden a tener menor desequilibrio de ligamiento que las autógamas completo, y en el caso del maíz se observa una rápida reducción en el desequilibrio de ligamiento, especialmente en poblaciones nativas (Flint-García, *et al.* 2003; Romay *et al*., 2013). Además, un estudio de simulación sugiere que un gran número de marcadores mejora la precisión de la predicción (Solberg *et al.*, 2008), sin embargo, Nakaya y Isobe (2012) indican que demasiados marcadores usualmente conducen a una disminución de la precisión.

En cuanto al tipo de marcador, los marcadores codominantes (ej. SNPs) producen predicciones más precisas que los marcadores dominantes (ej. ISSR) (Nakaya y Isobe, 2012). Además, los marcadores pueden ser categorizados como bialélicos (ej. SNPs) o multialélicos (ej. ISSR), requiriéndose mayor densidad de marcadores bialélicos comparado con los multialélicos (Solberg *et al.*, 2008). Aunque los SNPs son bialélicos, esta desventaja es compensada por su abundancia y la disponibilidad de plataformas para la genotipificación a gran escala (*ultra-high-throughput genotyping,* en inglés) (Borém y Fritsche-Neto, 2014).

Respecto al tamaño de la población de entrenamiento, en general al incrementarse el tamaño de la población de entrenamiento, se incrementa la precisión de predicción (Nakaya y Isobe, 2012). Por consiguiente, predicciones dentro de una población biparental es la situación más favorable para selección genómica, ya que los individuos están altamente emparentados al provenir de los mismos progenitores (Bernardo y Yu, 2007).

En cuanto a la relación entre la población de entrenamiento y la de mejoramiento, si la estructura genética entre ambas poblaciones es diferente, la precisión en las predicciones se puede ver afectada seriamente (Nakaya y Isobe, 2012).

3 MATERIALES Y MÉTODOS

3.1 Fuente de germoplasma

Accesiones: Se utilizó un conjunto de 669 variedades nativas de la Colección Núcleo del Banco de Germoplasma de Maíz del CIMMYT (ver listado en el apéndice 8.1), adaptadas a condiciones tropicales y subtropicales y que representan diferentes grupos raciales. La mayoría de estas accesiones fueron colectadas en México, Brasil, Venezuela y Guatemala, y en menor proporción provenientes de otros países de centro y Sudamérica.

Probadores: Se utilizaron híbridos de cruza simple de líneas del CIMMYT adaptados a condiciones tropicales y subtropicales: CML269/CML264, CML373/CML311, CML451/CML486 y CML495/CML494.

***Testigos*:** Como testigos y borde se utilizaron los híbridos comerciales de PIONEER P3055W, P30F32, P4063W y P4082W, los cuales están ampliamente adaptados a ambientes tropicales y subtropicales. El híbrido P4063W fue utilizado como testigo resistente (tolerante), ya que es el menos susceptible de la región; mientras que el resto fueron considerados como testigos extremadamente susceptibles.

3.2 Generación de mestizos

En el ciclo B (verano) del 2011, en Agua Fría, Puebla, se realizaron los cruzamientos entre las variedades nativas y probadores de la siguiente manera: se cruzó una sola planta por accesión con un probador, utilizando la accesión como progenitor masculino y el probador como hembra (Sood *et al*., 2014). A partir de esta cruza se obtuvieron los mestizos, los cuales fueron sujetos de la evaluación fenotípica de la enfermedad. La cruza de una sola planta por accesión se sustenta en el supuesto de que los haplotipos se replican a través de las accesiones (comunicación personal[4]). En términos de costo-beneficio, con la utilización de una sola planta por accesión fue posible explorar mayor cantidad de germoplasma nativo, mientras que si se hubiera evaluado varias réplicas de una misma accesión el número de poblaciones nativas

[4] Sarah Hearne, comunicación personal (2014). En reunión sobre mejoramiento molecular para resistencia al complejo mancha de asfalto en el proyecto MasAgro Biodiversidad, CIMMYT. México, D.F.

evaluadas hubiera sido considerablemente menor. La Figura 4 muestra el esquema para la obtención de los mestizos.

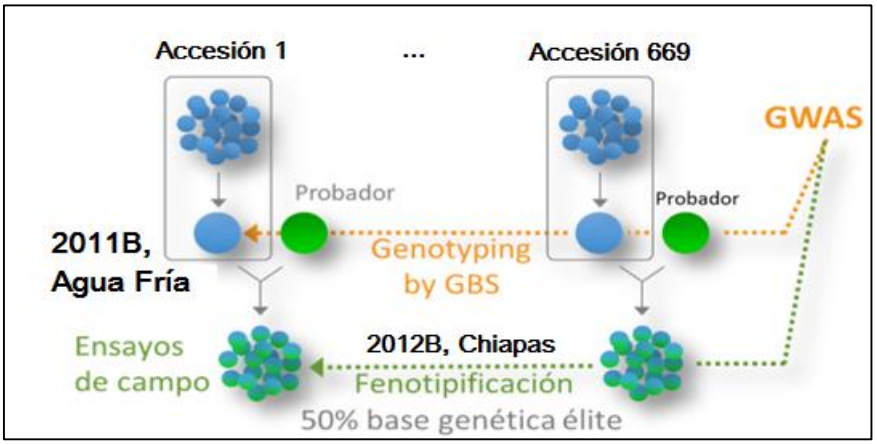

Figura 4. Formación de población y obtención de mestizos. Facilitada por Hearne[5].

3.3 Evaluación de campo

La evaluación de los mestizos se realizó en el ciclo B (verano) del año 2012, en la comunidad Guadalupe Victoria, del estado de Chiapas (coordenadas geográficas 16°26'48.35" N, 93°7'30.75" W; y 793 metros sobre el nivel de mar), bajo condiciones naturales de infección. Dado que no es posible reproducir los agentes causales en cultivo *in vitro* (Dittrich *et al*., 1991), y por consiguiente tampoco es posible realizar inoculación artificial, se eligió esta localidad por el registro de fuertes epidemias en años anteriores y consecutivos. En esta evaluación también se incluyeron cuatro accesiones para ser evaluadas *per se*. Estas accesiones fueron Oaxa-280, Guat-153, Chis-474 y Guer-208, la cuales mostraron los mayores niveles de resistencia al CMA en una evaluación previa realizada por CIMMYT en el año 2011 en la misma localidad (Rodríguez *et al*., 2013). La evaluación y la toma de datos fueron realizadas por la Dra. Martha Willcox y su equipo de trabajo como parte del proyecto MasAgro Biodiversidad (comunicación personal[5]).

Se utilizó la escala de Ceballos y Deutsch (1992) para el registro de la severidad de la enfermedad en campo, en una escala ordinal de 0 a 5, donde el 0 indica el más alto

[5] Sarah Hearne, comunicación personal (2014). En reunión sobre mejoramiento molecular para resistencia al complejo mancha de asfalto en el proyecto MasAgro Biodiversidad, CIMMYT. México, D.F.

nivel de resistencia y el 5 indica el nivel más susceptible (Cuadro 2). Considerando el alto nivel de heterocigosis de las variedades nativas y por consiguiente manifestación de segregación en los mestizos, la medición se realizó individualmente a las 6 plantas centrales de cada parcela y luego se obtuvo el dato promedio por parcela. Se realizaron dos evaluaciones, una en pre-floración y la otra en post-floración. También se evaluó el rendimiento medido en peso de grano por parcela y otras características agronómicas como criterios secundarios de selección, entre ellas: altura de planta y mazorca, acame de raíz y tallo y floración.

Cuadro 2. Escala de severidad del complejo mancha de asfalto en maíz con base en la propuesta de Ceballos y Deutsch (1992).

Clase	Categoría	Observaciones	Área afectada (%)
0	Inmune o muy resistente	Sin o con muy pocas lesiones, todas en hojas inferiores a la mazorca.	0 – 2 %
1	Resistente	Varias lesiones en hojas inferiores a la mazorca.	3 – 10 %
2	Moderadamente resistente	Muchas lesiones en hojas inferiores a la mazorca, algunas áreas necróticas, pero la mayor parte del área todavía verde. Algunas lesiones en hojas superiores.	11 – 25 %
3	Moderadamente susceptible	Mayoría del área de las hojas inferiores a la mazorca necrosada. Muchas lesiones ocurriendo en hojas superiores.	26 – 50 %
4	Susceptible	Sin tejido verde en las hojas inferiores. Considerable área foliar necrosada en hojas superiores.	51 – 80 %
5	Extremadamente susceptible	Planta muerta o con muy poca área verde.	> 80 %

3.4 Diseño experimental y manejo agronómico

Se utilizó un diseño experimental aumentado en hileras y columnas, sin repeticiones de mestizos, a excepción de las accesión que fueron evaluadas *per se* y los mestizos derivados de éstas. Las accesiones *per se* fueron replicadas cuatro veces y los mestizos derivados de estas 5 veces. En el experimento se evaluaron un total de 779 parcelas, las cuales se desglosan de la siguiente manera: 665 mestizos sin réplicas, cinco réplicas de los cuatro mestizos derivados de las accesiones *per se* (20 parcelas), cuatro réplicas de las cuatro accesiones *per se* (16 parcelas) y 78 parcelas sembradas con testigos (10% del total de parcelas). El arreglo fue semialeatorizado, lo cual

consistió en ubicar primero a los testigos de forma sistemática en el campo, incluyendo la presencia de todos los testigos en las filas y columnas y luego se ubicaron los mestizos de forma aleatoria en las parcelas restantes. Cada parcela consistió de un surco de 2 m con una densidad final de 12 plantas, un espaciamiento de 80 cm entre surcos y 1 m de calle. Los testigos incluían híbridos considerados resistentes y susceptibles. El arreglo semialeatorizado permitió realizar un ajuste por variación espacial dentro del área experimental (Burgueño *et al*., 2000).

En cuanto al manejo agronómico, se aplicó una dosis de 41-46-60 Kg de nitrógeno, fósforo y potasio por hectárea al momento de la siembra. A los 30 días después de la siembra se aplicaron 138 unidades de nitrógeno (urea). Se aplicaron herbicidas e insecticidas al momento de la siembra y durante el ciclo del cultivo de acuerdo a las prácticas agronómicas generalmente utilizadas por los agricultores de la región.

3.5 Análisis de datos fenotípicos

Debido a que la información genotípica se obtuvo a partir de las accesiones *per se* (ver siguiente inciso 3.6) y la evaluación fenotípica se realizó a los mestizos, fue necesario estimar las mejores predicciones lineales insesgadas (BLUPs, por sus siglas en inglés) de las accesiones *per se,* es decir, estimar el efecto de la accesión anidado en el efecto del mestizo.

Para este análisis se utilizaron los datos obtenidos en la evaluación en post-floración ya que estos presentaron mayor varianza y por consiguiente mayor contraste entre individuos resistentes y susceptibles. Se ajustó un modelo lineal mixto mediante el método de Máxima Verosimilitud Restringida (Ruppert *et al.*, 2003) utilizando el software ASReml V 3.0 (Gilmour *et al.*, 2009), donde los efectos de los testigos y probadores se consideraron como efectos fijos y los efectos de las accesiones como efectos aleatorios en un modelo completamente anidado. Además, la heterogeneidad del suelo dentro del área experimental se controló modelando los efectos de las hileras y las columnas como efectos aleatorios. También se realizó un ajuste espacial utilizando un modelo autoregresivo para el término del error residual, suponiendo que existe una correlación entre las parcelas que depende de la distancia física entre ellas (Burgueño *et al*., 2000). Con base en lo anterior se definió el siguiente modelo mixto:

$$y_{ijklm} = \mu + \gamma_i + \lambda_j + \alpha_k + \beta_{l(k)} + \delta_{m(kl)} + \varepsilon_{ij} \ , \qquad (2)$$

donde:

y_{ijklm}: es la variable respuesta de la *m*-ésima accesión en el probador *k* en el grupo *K*+1 en la *i*-ésima hilera y *j*-ésima columna.

μ: es la media general,

γ_i: es el efecto de la *i*-ésima hilera, $\gamma_i \sim N\left(0, \sigma_\gamma^2\right)$,

λ_j: es el efecto de la *j*-ésima columna, $\lambda_j \sim N\left(0, \sigma_\lambda^2\right)$,

α_k: es el efecto del *k*-ésimo grupo, *k=1,…,K, K*+1; si *k*≤*K*, el grupo es un testigo, el grupo *K+1* es el promedio de los probadores.

$\beta(\alpha)_{kl}$ es el efecto del l-ésimo probador en el *k*-ésimo testigo,

$\delta_{m(kl)}$ es el efecto de la *m*-ésima accesión en el probador *k* en el grupo *K*+1 y $\delta_{m(kl)} \sim N(0, \sigma_{kl}^2)$. De esta manera se estiman los efectos de las accesiones *per se.*

ε_{ij} es el error experimental en la *i*-ésima hilera y *j*-ésima columna, para el cual se asumió la siguiente distribución $\varepsilon_{ij} \sim N(0, \Sigma)$, donde Σ es el producto de Kronecker de Σ_r y Σ_c ($\Sigma = \Sigma_r \otimes \Sigma_c$), donde Σ_r y Σ_c son las matrices de correlación entre parcelas en el sentido de las hileras y columnas, respectivamente, las cuales se definen a continuación:

$$\Sigma_r = \begin{bmatrix} 1 & \rho_r^1 & \rho_r^2 & \cdots & \rho_r^{d-2} & \rho_r^{d-1} \\ \rho_r^1 & 1 & \rho_r^1 & \cdots & \rho_r^{d-3} & \rho_r^{d-2} \\ \rho_r^2 & \rho_r^1 & 1 & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_r^{d-2} & \rho_r^{d-3} & \rho_r^{d-4} & \cdots & 1 & \rho_r^1 \\ \rho_r^{d-1} & \rho_r^{d-2} & \rho_r^{d-3} & \cdots & \rho_r^1 & 1 \end{bmatrix} \quad \Sigma_c = \begin{bmatrix} 1 & \rho_c^1 & \rho_c^2 & \cdots & \rho_c^{d-2} & \rho_c^{d-1} \\ \rho_c^1 & 1 & \rho_c^1 & \cdots & \rho_c^{d-3} & \rho_c^{d-2} \\ \rho_c^2 & \rho_c^1 & 1 & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_c^{d-2} & \rho_c^{d-3} & \rho_c^{d-4} & \cdots & 1 & \rho_c^1 \\ \rho_c^{d-1} & \rho_c^{d-2} & \rho_c^{d-3} & \cdots & \rho_c^1 & 1 \end{bmatrix},$$

donde ρ_r y ρ_c son las correlación entre parcelas en el sentido de las hileras y columnas, respectivamente y d es la distancia entre parcelas. Esta correlación es como una ponderación de las parcelas vecinas y permite ajustar los BLUPs por la ubicación de la parcela (Burgueño *et al*., 2000). Es importante señalar que al calcular los BLUPs los datos de severidad se transforman en una escala continua, por lo que el estudio de asociación se realizó suponiendo una distribución normal.

A partir de los BLUPs se realizó una gráfica de dispersión y un histograma de frecuencias para observar el comportamiento de los datos fenotípicos. También se realizó una gráfica para observar la correlación entre el nivel de severidad de mancha de asfalto y el rendimiento (peso de grano por mazorca corregido por % humedad) de los mestizos y accesiones *per se.*

3.6 Genotipificación y control de calidad

Las muestras de tejido foliar se tomaron en campo durante el ciclo de cruzamientos (Agua Fría, 2011 B). Se muestreó una porción foliar de 12 cm de longitud de cada una de las accesiones *per se* que fueron utilizadas como progenitores de los mestizos. La extracción de ADN se realizó en el Laboratorio de Análisis Molecular del CIMMYT utilizando como base el protocolo estándar CTAB (Doyle, 1987). La calidad del ADN se verificó mediante un gel de agarosa al 0.8 % y la cuantificación mediante un espectrofotómetro NanoDrop 8000 (Thermo Scientific). Las muestras fueron enviadas al Instituto de Biotecnología de la Universidad de Cornell, Ithaca, EUA, para ser analizadas con la tecnología de GBS desarrollada por Elshire *et al*. (2011) y descrito en el inciso 2.8. Las muestras de ADN fueron digeridas con la enzima de restricción *Ape*KI y analizadas en el secuenciador HiSeq2500 Ilumina.

Control de calidad: marcadores con frecuencia del alelo menor (MAF, por sus siglas en inglés) fue menor de 0.05 fueron descartados, ya que usualmente producen resultados inestables (Corvin *et al*., 2010; Weng *et al*., 2011). Con este primer filtrado se eliminaron también aquellos marcadores monomórficos, cuyo genotipo en todos los mestizos es el mismo (MAF = 0) y por consiguiente no explican la variabilidad del nivel de severidad del CMA. Un segundo filtrado se realizó para descartar aquellos marcadores con más de 20 % de datos faltantes. El control de calidad se realizó utilizando el software TASSEL V 5.2.12 (Bradbury *et al*., 2007).

La información genotípica se manejó en dos formatos dependiendo el software en que se analizó la información: formato estándar HapMap y el Numérico. El formato HapMap permite contener en un solo archivo la información de los SNPs (cromosoma y posición) y su genotipo codificado según la Unión Internacional de Química Pura y Aplicada (IUPAC, por sus siglas en inglés). En este caso la información de los

marcadores se almacena en las filas y la información de los taxa en las columnas. Este formato fue utilizado en el análisis GWAS con el método single marker implementado con el programa TASSEL.

En el formato numérico, las filas corresponden a los *taxa* y las columnas a los marcadores SNPs. Este formato no contiene la información del cromosoma y la posición de cada SNP, por lo que se requiere otro archivo que contenga dicha información. En este caso se utilizó el programa PLINK (Purcell *et al*., 2007) para convertir la matriz en formato HapMap a formato numérico mediante la recodificación de los genotipos de los marcadores para efectos aditivos siguiendo la guía en la web (http://pngu.mgh.harvard.edu/~purcell/plink/dataman.shtml): Homocigoto dominante *AA* → "0", Heterocigoto *Aa* → "1" y Homocigoto recesivo *aa* →"2". En otras palabras, la recodificación se realiza contando el número de alelos menores (ej. *a*). Este formato se utilizó para el análisis GWAS con el método BayesB y en selección genómica.

Previo a la utilización de las matrices de datos genotípicos en ambos formatos, se le dio un tratamiento a los datos faltantes mediante la técnica estadística de imputación, la cual consiste en la sustitución de los datos faltantes por valores obtenidos a partir de los datos disponibles (Rosas y Verdejo, 2009). Para el caso del formato HapMap, la imputación se realizó mediante la sustitución de la media muestral de los valores disponibles para cada SNP en cada uno de los valores perdidos. Para el caso del formato numérico, la imputación se realizó mediante el método Hot Deck (Rosas y Verdejo, 2009), que consiste en seleccionar mediante muestreo aleatorio simple con reemplazo *m* valores a partir de los *n* valores disponibles de la variable a imputar.

3.7 **Estimación de la heredabilidad genómica**

La heredabilidad en sentido estricto fue definida por Falconer y Mackay (1996) como $h^2 = V_A/V_F$, esto es la proporción de varianza genética aditiva (V_A) sobre la varianza fenotípica total (V_F).

La heredabilidad en sentido estricto puede ser calculada a partir de los marcadores moleculares en selección genómica mediante la descomposición de la V_A en la suma de la varianza explicada por múltiples marcadores moleculares así: $V_A = V_{A1} + V_{A2} +$

$V_{A3} + \cdots + V_{Ap}$, donde p es el número de marcadores, bajo el supuesto de que los marcadores no están correlacionados entre sí (Meuwissen *et al*., 2001; Nakaya y Isobe, 2012). En este sentido, la heredabilidad es una estimación de la proporción de varianza genética aditiva capturada por los marcadores, la cual fue denominada como "heredabilidad genómica" por Daetwyler *et al*. (2014), quienes a su vez propusieron el siguiente estimador:

$$\hat{h}^2 = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2},$$

donde $\hat{\sigma}_u^2$ es una estimación de la varianza genética aditiva capturada por los marcadores moleculares, equivalente a V_A; $\hat{\sigma}_e^2$ es una estimación de la varianza residual, la cual incluye la varianza no aditiva (dominancia y epistasis), la variación ambiental, la variación debida a la interacción genotipo ambiente y la varianza del error (Nakaya y Isobe, 2012). $\hat{\sigma}_u^2$ y $\hat{\sigma}_e^2$ se obtienen a partir de la media posterior de los componentes de varianza de los modelo lineales de selección genómica, GBLUP y RKHS (Ecuaciones (8) y (9) respectivamente) (Pérez-Rodríguez y de los Campos, 2014).

3.8 Estudio de asociación del genoma completo para resistencia al CMA

Los BLUPs de cada accesión *per se* fueron utilizados como fenotipos para el análisis de asociación. Para la detección de QTL asociados a la resistencia a CMA se utilizaron dos métodos, "Single Marker" y "BayesB". El primero ha sido ampliamente utilizado en análisis GWAS en plantas, mientras que el modelo BayesB ha demostrado potencial para detección de QTL en estudios de simulación y aún no ha sido probado con datos reales (Sahana *et al*., 2010; Berg *et al*., 2013). Estos marcadores fueron evaluados posteriormente (ver inciso 3.9) para estimar la proporción de varianza genética explicada por los QTL a los que se encuentran ligados y también se evaluó su potencial predictivo con enfoque de selección genómica (ver inciso 3.11.3) (Collard *et al*., 2005).

3.8.1 Método Single Marker

El método single marker consiste en evaluar un solo marcador a la vez, de tal forma que se realizan tantas regresiones y pruebas de hipótesis como marcadores existan.

Recientemente se ha utilizado el enfoque de modelo lineal mixto para incluir ajuste por factores de confusión, tales como estructura de población y relaciones de parentesco entre individuos con la finalidad de reducir la tasa de falsos positivos (reducción del error tipo I) (Yu *et al*., 2006).

Para este análisis se utilizó el enfoque de modelo lineal mixto propuesto por Zhang *et al*. (2010), que incluye los primeros tres componentes principales como covariables para el control de la estructura de población y la matriz de parentesco $\boldsymbol{K}$ (definida abajo) para las relaciones de parentesco. El modelo propuesto por Zhang *et al*. (2010) es el siguiente:

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{v} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}\ , \tag{4}$$

donde $\boldsymbol{y}$ es el vector de fenotipos (BLUPs), $\boldsymbol{v}$ corresponde a los efectos fijos desconocidos de los primeros tres componentes principales, $\boldsymbol{\beta}$ corresponde a los efectos de los marcadores considerados fijos; $\boldsymbol{u}$ son los efectos aleatorios poligenéticos desconocidos, con distribución $\boldsymbol{u} \sim \boldsymbol{N_n}(\boldsymbol{0}, \boldsymbol{G})$. Además, la matriz de covarianza $\boldsymbol{G} = 2\boldsymbol{K}\sigma_a^2$, donde $\boldsymbol{K}$ es la matriz de parentesco (coancestria) de dimensiones $n \times n$, con elementos $\boldsymbol{K}_{ij}$ $(i, j = 1, \ldots, n)$ calculada a partir del conjunto de marcadores genéticos y σ_a^2 es una varianza genética. Por otra parte, $\boldsymbol{W}$, $\boldsymbol{X}$ y $\boldsymbol{Z}$ son las matrices de incidencia para $\boldsymbol{v}$, $\boldsymbol{\beta}$ y $\boldsymbol{u}$ respectivamente y $\boldsymbol{\varepsilon}$ es el vector de residuales aleatorios, el cual se distribuye $\boldsymbol{\varepsilon} \sim \boldsymbol{N_n}(\boldsymbol{0}, \boldsymbol{R})$, donde $\boldsymbol{R} = \boldsymbol{I}\sigma_\varepsilon^2$ e $\boldsymbol{I}$ es una matriz identidad y σ_ε^2 es la varianza residual desconocida.

Con la solución de las ecuaciones del modelo lineal mixto se obtuvieron las mejores estimaciones lineales insesgadas BLUEs de los efectos fijos ($\boldsymbol{v}$ y $\boldsymbol{\beta}$) y las mejores predicciones lineales insesgadas (BLUPs, por sus siglas en inglés) de los efectos aleatorios ($\boldsymbol{u}$). Dado que se evalúa un solo marcador a la vez, $\boldsymbol{\beta} = \beta_i$, y las hipótesis para la prueba de asociación de cada marcador son: hipótesis nula $\beta_i = 0$ e hipótesis alternativa $\beta_i \neq 0$.

Dado el gran número de pruebas estadísticas que se realizan, los valores de p esperados son mucho más pequeños que los umbrales comunes (ej. $\alpha = 0.01$), por lo

que en los análisis GWAS los umbrales son más rigurosos (Corvin *et al.*, 2010). Para este análisis se definió un umbral de significancia $\alpha = 1 \times 10^{-4}$ para seleccionar los marcadores fuertemente asociados a la resistencia al CMA. Por consiguiente, solo los marcadores con valor de p menor que este umbral (o superior a 4.0 en la escala $-log_{10}(p\ \text{valor})$) serán considerados como significativos.

A partir de los resultados del análisis se realizaran dos gráficas: una gráfica cuantil-cuantil, la cual muestra la dispersión de los valores de p estimados en el análisis contra los valores esperados bajo la hipótesis nula (no asociación a la resistencia), para evaluar la efectividad del modelo para controlar los falsos positivos (Pearson y Manolio, 2008; Corvin *et al*., 2010). Bajo el supuesto que la mayoría de marcadores SNPs no están asociados con la resistencia al CMA, se espera que los valores de p estimados se ajusten estrechamente a los valores esperados, a excepción de aquellos marcadores que tengan una asociación verdadera. Por el contrario, fuertes desviaciones de los valores esperados sugeriría una alta tasa de asociaciones espurias, tal como lo interpretan Pearson y Manolio (2008). Una segunda gráfica tipo "Manhattan" (por su semejanza a la ciudad de Manhattan, New York, vista desde el horizonte) muestra los valores de p de acuerdo a su posición genómica y agrupados por cromosomas diferenciados por colores. En ambas gráficas los valores de p son transformados usando $-log_{10}(p\ \text{valor})$ para una mejor visualización. Por ejemplo, un valore de p igual a 0.0001 equivale a 4.0 en la escala $-log_{10}(p\ \text{valor})$.

Software: En este estudio se utilizó el programa TASSEL V 5.2.12 para realizar el análisis GWAS basado en el modelo lineal mixto (Zhang *et al*., 2010), el cual se encuentra disponible en http://www.maizegenetics.net. Se utilizaron los datos genotípicos en formato HapMap (ver Cuadro 4)

3.8.2 Método BayesB

Este método analiza todos los marcadores moleculares simultáneamente mediante la implementación de una regresión múltiple bajo enfoque bayesiano con capacidad para selección de variable, originalmente desarrollado y utilizado en selección genómica (Meuwissen *et al.* 2001). Este método se basa en el hecho de que en realidad hay

muchos *loci* sin varianza genética (no segregantes) y unos pocos *loci* con varianza genética (Meuwissen *et al.* 2001). Para abordar esto, BayesB utiliza una distribución *a priori* mixta para estimar los efectos de los marcadores, la cual consta de un punto de masa fijado en cero y una distribución marginal *t* escalada. De esta manera la mayoría de marcadores con varianza genética nula son llevados al punto de masa en cero, mientras que los pocos *loci* con varianza genética toman un valor de la distribución marginal *t*-escalada (Pérez-Rodríguez y de los Campos, 2014). Bajo la definición de estos supuestos *a priori*, se utilizó el siguiente modelo lineal:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\ , \tag{5}$$

donde $\boldsymbol{y}$ es el vector de fenotipos (BLUPs), $\mathbf{1}$ es un vector cuyos elementos son unos asociado al intercepto μ, $\boldsymbol{X}$ es la matriz de marcadores moleculares, $\boldsymbol{\varepsilon}$ es el vector de residuos aleatorios, el cual se distribuye $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, R)$, donde $R = \boldsymbol{I}\sigma_\varepsilon^2$ e $\boldsymbol{I}$ es una matriz identidad y σ_ε^2 es la varianza de residuos desconocida. $\boldsymbol{\beta}$ son los efectos de los marcadores, a los cuales se les asigna una distribución *a priori* mixta (Pérez-Rodríguez y de los Campos, 2014):

$$\beta_j|\sigma_j^2, \pi = \begin{cases} 0 & con\ probabilidad\ \pi \\ N(0, \sigma_j^2) & con\ probabilidad\ (1-\pi) \end{cases}$$

La distribución *a priori* de la varianza para los marcadores con varianza genética es $\sigma_j^2 \sim \chi^{-2}(v, S)$ para $\mathrm{j} = 1, \ldots, p$ y donde v son los grados de libertad y S es la escala del parámetro. A diferencia del método Single Marker, en este modelo no se necesita incluir ninguna corrección por estructura de población y relaciones entre individuos, ya que al analizarse todos los marcadores simultáneamente se controlan de manera implícita estos factores de confusión (Pérez-Rodríguez y de los Campos, 2014). La proporción de marcadores con efecto diferente de cero fue definida *a priori* con un valor de 0.001, lo que significa que al menos $56{,}092 \times 0.001 \approx 56$ marcadores tienen efecto diferente de cero (Pérez-Rodríguez y de los Campos, 2014).

Como criterio para selección de marcadores se utilizó la probabilidad posterior de inclusión, que es la probabilidad posterior de que un marcador tenga un efecto diferente de cero (Berg *et al*., 2013; Pérez-Rodríguez y de los Campos, 2014). Se

seleccionaron igual número de marcadores que los identificados con Single Marker para poder compararlos.

Software: Este método fue implementado con el paquete BGLR (Bayesian Generalized Linear Regression, en inglés), el cual implementa una colección de regresiones lineales bayesianas, incluyendo el método BayesB (Pérez-rodríguez y de los Campos, 2014). El paquete y manual de referencia están disponibles en http://cran.at.r-project.org/web/packages/BGLR/index.html. Para este caso se utilizó la información genotípica en formato numérico (Cuadro 5).

3.9 Proporción de la varianza genética explicada por los QTL

Se estimó la proporción de varianza genética explicada por los QTL ligados a los marcadores identificados de forma individual y en conjunto. Para el caso del método BayesB sólo se estimó la varianza genética explicada en conjunto ya que no es posible obtener el r^2 de cada marcador a partir de la regresión múltiple. Se utilizó el estimador utilizado por Gowda *et al*. (2015):

$$P_G = {}^{P_f}/_{h^2} \quad , \tag{6}$$

donde P_G es la proporción de varianza genética, P_f es la proporción de variabilidad fenotípica explicada por los QTL en conjunto obtenida a partir del coeficiente de determinación múltiple R^2_{aj} (Collard *et al.*, 2005; Gowda *et al*., 2015) y h^2 es la heredabilidad genómica. Para obtener el R^2_{aj} se ajustaron dos modelos de regresión lineal múltiple, uno incluyendo los marcadores significativos identificados con single marker y el otro incluyendo los marcadores identificados con BayesB. Para estimar la proporción de varianza genética explicada por cada marcador de manera individual se utilizó el r^2 de cada marcador como numerador de la Ecuación (6).

3.10 Análisis de genes candidatos

A partir de los SNPs significativamente asociados a la resistencia al CMA se procedió a realizar el análisis del gen candidato. Para esto se utilizó la posición de los marcadores (ej. la posición del SNP S4_206663710 es **206663710** en el cromosoma

4) para localizar su posición en el genoma mediante la utilización de la base de datos genética y genómica del maíz, conocida como MAIZEGDB, basado en el genoma de referencia "B73" RefGen_v2(MGSC) (http://www.maizegdb.org/gbrowse). Los genes filtrados en la base MAIZEGDB que contenían los SNPs asociados o adyacentes fueron considerados genes candidatos potencialmente involucrados en el mecanismo de defensa de los materiales resistentes al CMA. Para identificar la posible función del gen candidato se utilizaron las bases de datos MaizeCyc (http://maizecyc.maizegdb.org/) y la base de datos del Centro Nacional de Información Biotecnológica (NCBI, por sus siglas en inglés) (http://www.ncbi.nlm.nih.gov/).

3.11 Selección genómica para resistencia al CMA

Se evaluó el potencial de predicción de dos enfoques de selección genómica: 1) El primer enfoque es ampliamente utilizado y consiste en utilizar todos los marcadores disponibles (~ 56 mil SNPs) como variables predictoras (Meuwissen *et al*., 2001). Bajo este enfoque se evaluaron cuatro modelos bayesianos: Regresión Ridge Bayesiana, GBLUP, Bayes*B* y el modelo semiparamétrico RKHS (Pérez-Rodríguez y de los Campos, 2014). 2) El segundo enfoque es una nueva propuesta que consistió en utilizar únicamente los marcadores significativamente asociados a la resistencia al CMA mediante el análisis GWAS. Para ello se ajustaron 3 modelos de regresión lineal múltiple por mínimos cuadrados ordinarios. El primer modelo incluyó únicamente los marcadores significativos identificados con el método single marker; el segundo modelo incluyó los marcadores identificados con BayesB y el tercer modelo incluyó la combinación de los marcadores identificados con ambas metodologías, incluyendo únicamente los SNPs diferentes. Se utilizó la información genotípica en formato numérico (Cuadro 5). Para evaluar la precisión de las predicciones de los modelos se utilizó el método de múltiples particiones aleatorias de prueba y entrenamiento (Pérez-Rodríguez y de los Campos, 2014) (ver inciso 3.11.3).

3.11.1 Modelos bayesianos utilizando todos los marcadores

Para implementar los modelos bayesianos utilizando todos los marcadores como variables predictoras se utilizó el paquete BGLR en R (Pérez-Rodríguez y de los Campos, 2014).

Regresión Ridge Bayesiana (RRB)

La Regresión Ridge fue uno de los primeros métodos para selección genómica propuesto por Meuwissen *et al*. (2001), equivalente a los BLUPs en el contexto de los modelos mixtos, que se caracteriza por realizar un grado de penalización aplicado homogéneamente a todos los marcadores, donde los marcadores tienen efectos pequeños y son considerados variables aleatorias normales independientes e idénticamente distribuidas. El modelo de Regresión Ridge en el contexto Bayesiano (RRB) es el siguiente (Pérez-Rodríguez y de los Campos, 2014):

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{X\beta} + \boldsymbol{\varepsilon}\ , \tag{7}$$

donde $\boldsymbol{y}$ es el vector de fenotipos, $\mathbf{1}$ es un vector cuyos elementos son unos asociado al intercepto μ, $\boldsymbol{\beta}$ es un vector cuyos elementos son los efectos de los marcadores, tal que $\boldsymbol{\beta} \sim N_p(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I})$, donde $\boldsymbol{I}$ es una matriz identidad; $\boldsymbol{X}$ es la matriz de incidencia para los efectos de los marcadores que contiene la información genotípica, $\boldsymbol{\varepsilon}$ es el vector de residuales aleatorios, el cual se distribuye $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I})$, donde σ_ε^2 es la varianza residual desconocida. La varianza residual es desconocida, asignándole una *a priori* Chi-cuadrada inversa escalada $\sigma_\varepsilon^2 \sim X^{-2}(\sigma_\varepsilon^2 | df_\varepsilon, S_\varepsilon)$ (Pérez-Rodríguez y de los Campos, 2014).

Modelo GBLUP

Una parametrización alternativa del modelo RRB se obtiene al sustituir $\boldsymbol{u} = \boldsymbol{X\beta}$**,** con lo cual se logra un modelo con efectos aditivos infinitesimales al incluir la matriz de relaciones genómicas, calculada con base a los marcadores moleculares y equivalente la matriz de relaciones de pedigree (Pérez-Rodríguez y de los Campos, 2014). El BGLR realiza internamente el análisis de componentes principales a partir de la matriz de relaciones genómica, sobre los cuales realiza la regresión. De esta manera se aprovecha la equivalencia entre los procesos Gaussianos y componentes principales (de los Campos *et al.,* 2010). El modelo se define y describe a continuación:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{u} + \boldsymbol{\varepsilon}\ , \tag{8}$$

donde μ es el intercepto, el vector $\boldsymbol{u}$ se distribuye $\boldsymbol{u} \sim N(\mathbf{0}, \sigma_\beta^2 \boldsymbol{XX}')$, donde $\boldsymbol{XX}'$ es proporcional a la matriz de relaciones genómica $\boldsymbol{G}$, por lo que se asume que $\boldsymbol{u} \sim N(\mathbf{0}, \sigma_u^2 \boldsymbol{G})$, donde σ_u^2 equivale a la varianza genética aditiva capturada por los marcadores; $\boldsymbol{\varepsilon}$ el vector de error independiente e idénticamente distribuido $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I})$, donde $\boldsymbol{I}$ es una matriz identidad y la varianza residual σ_ε^2 es desconocida, asignándole una *a priori* Chi-cuadrada inversa y escalada $\sigma_\varepsilon^2 \sim \chi^{-2}(\sigma_\varepsilon^2 | df_\varepsilon, S_\varepsilon)$. Una manera de calcular la matriz $\boldsymbol{G}$ es la propuesta por VanRaden (2008), $\boldsymbol{G} = t^{-1}\boldsymbol{XX}'$, donde $t = \sum_{k=1}^{m} p_k q_k$, donde p_k es la frecuencia del alelo mayor (ej. A) y q_k es la frecuencia del alelo menor (ej. a) del k-ésimo marcador. $\boldsymbol{X}$ es la matriz $(n \times k)$ de datos de marcadores donde n es el número de individuos y k es el número de marcadores (Ornella *et al*., 2012).

Modelo BayesB

El modelo es el mismo que se describe en la Ecuación (5), $\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$.

Pero en este caso se utilizó para la obtención de los GEBVs y predicción de los fenotipos.

Modelo RKHS

El método Reproducing Kernel Hilbert Spaces Regressions (RKHS por sus siglas en inglés) es un procedimiento semiparamétrico basado en un kernel Gaussiano (Gianola y Kaam, 2008). Gianola *et al.* (2006) propusieron la aplicación de este método en SG para estimar los valores genéticos totales.

El modelo utilizado es el descrito por Pérez-Rodríguez y de los Campos (2014), bajo enfoque bayesiano:

$$y = \mu\mathbf{1} + \boldsymbol{u} + \boldsymbol{\varepsilon} \tag{9}$$

Con $p(\mu, \boldsymbol{u}, \boldsymbol{\varepsilon}) \propto N_n(\boldsymbol{u}|\mathbf{0}, \boldsymbol{K}\sigma_u^2) N_n(\boldsymbol{\varepsilon}|\mathbf{0}, \boldsymbol{I}\sigma_\varepsilon^2)$, donde $\boldsymbol{K} = \{K(x_i, x_i')\}$ es la matriz del promedio de distancias euclidianas entre genotipos al cuadrado, y la función del kernel Gaussiano puede definirse como (Gianola y Kaam, 2008):

$$K(x_i, x_i') = exp\left\{-h \times \frac{\sum_{k=1}^{p}(x_{ik} - x_{i'k})}{p}\right\},$$

donde el parámetro h controla que tan rápido la función de covarianza baja a medida que la distancia entre los pares de genotipos de vectores aumenta y además juega un papel importante en inferencias y predicciones, y a la cual se le asigna un valor por defecto $h = 0.5$ (Pérez-Rodríguez y de los Campos, 2014). σ_u^2 es la varianza genética capturada por los marcadores y σ_ε^2 la varianza residual desconocida.

3.11.2 Modelos de regresión lineal múltiple utilizando los SNPs significativos

Se ajustaron 3 modelos de regresión lineal múltiple por mínimos cuadrados ordinarios. El primero incluyó los marcadores significativos identificados con el método single marker, el segundo incluyó los marcadores identificados con BayesB y el tercero incluyó una combinación de los marcadores identificados con ambas metodologías. Para ajustar los modelos se utilizó la función `lm()` en R (R-Core Team, 2015).

3.11.3 Evaluación de la precisión de las predicciones

La validación se realizó por el método de múltiples particiones aleatorias de prueba y entrenamiento propuesto por Pérez-Rodríguez y de los Campos (2014), el cual consiste en generar múltiples particiones de entrenamiento-prueba mediante asignación aleatoria de los individuos. Cada partición se considera una réplica, la cual produce una estimación de la precisión de las predicciones, medida con base en la correlación entre los fenotipos observados y los predichos por el modelo, y el cuadrado medio del error (CME). Para la partición de entrenamiento se asignó aleatoriamente 80 % de individuos (para ajuste del modelo) y 20 % para prueba (predicción). Se realizaron 50 réplicas o particiones a partir de las cuales se obtuvo el promedio del CME y las correlaciones (r) (correlación de Pearson) entre datos observados y predichos para ambas particiones. El criterio de precisión fue el siguiente: A mayor correlación y menor CME en la partición de prueba, mayor precisión del modelo (Pérez-Rodríguez y de los Campos, 2014). Además, para analizar el efecto de sobreajuste de los modelos bayesianos se analizaron diferentes densidades de marcadores (50, 100, 500, 1000, 5000, 15000 y 56000 SNPs) seleccionados de forma aleatoria, cada una

evaluada con y sin los 17 SNPs identificados con el modelo BayesB en el análisis GWAS. Para ajustar las diferentes densidades de marcadores se utilizó el modelo GBLUP como representativo de los modelos bayesianos.

4 RESULTADOS

4.1 Evaluación fenotípica

En la Figura 5 se presenta la dispersión y el histograma de frecuencias de los BLUPs del nivel de severidad de las accesiones. Se observa una distribución continua con tendencia a concentrarse alrededor de un nivel de severidad entre 4.0 y 5.0, lo cual indica que la mayoría de mestizos evaluados (98 %) son susceptibles y extremadamente susceptibles de acuerdo a la escala de Ceballos y Deutsch (1992). En el histograma puede observarse también una distribución que se ajusta bastante a la normal. El coeficiente de variación observado fue de 6.64 %, lo cual indica poca variación en el nivel de severidad de las accesiones, lo cual puede atribuirse en parte al control de la variabilidad ambiental (hileras y columnas) y espacial incluidos en el modelo mixto para la obtención de los BLUPs (Ecuación (2)).

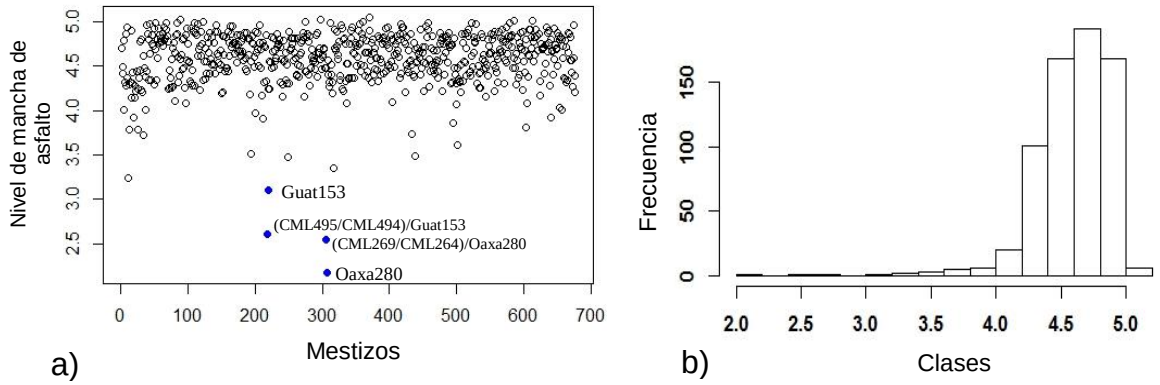

Figura 5. Nivel de severidad del CMA. a) Dispersión, b) Histograma de frecuencias.

En el Cuadro 3 se presenta de forma resumida el listado de las 10 plantas más resistentes y las 10 más susceptibles al CMA, así como el rendimiento medido en peso de grano por parcela corregido por humedad. A partir de la evaluación fenotípica se observó que los mestizos derivados de las accesiones Guat153 y Oaxa280 fueron los más tolerantes y los que produjeron mayor rendimiento (Figura 8). Las accesiones *per se* que actuaron como progenitores de estos mestizos también mostraron un alto nivel de tolerancia, siendo OAXA280 la más tolerante de las dos, sin embargo, GUAT153 presentó mayor rendimiento.

Por otra parte, puede apreciarse un grado de correlación negativa $(r = -0.5)$ entre el nivel de severidad y rendimiento, ya que los mestizos más susceptibles presentan rendimiento mucho más bajo que los más resistentes. Esta correlación puede observarse más claramente en la Figura 6.

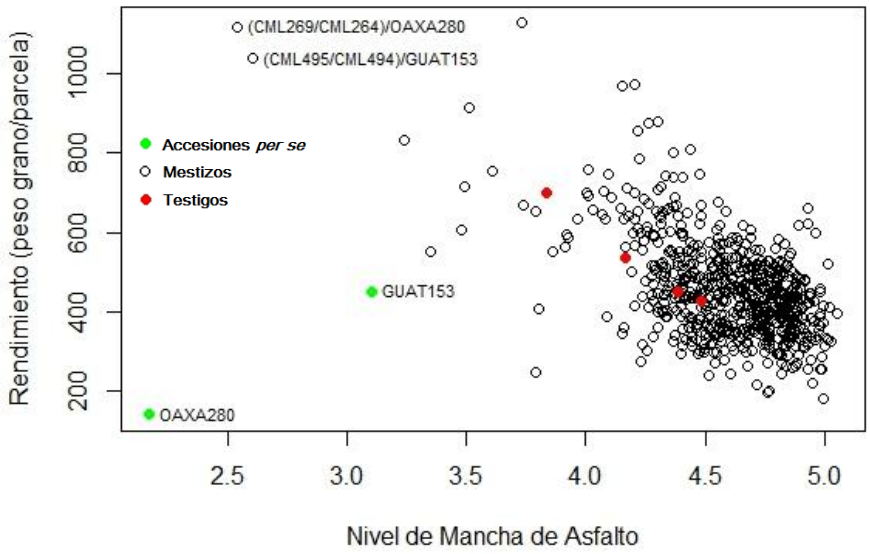

Figura 6. Relación entre rendimiento y nivel de severidad del CMA.

Cuadro 3. Listado de las 10 plantas más resistentes y 10 más susceptibles y su rendimiento respectivo.

	Mestizo (Pedigree) o Accesión	Nivel de severidad mancha de asfalto (BLUPs)	Rendimiento (peso grano/parcela)*
10 más resistentes	OAXA280 (*per se*)	2.17	140.32
	(CML269/CML264)/OAXA280	2.54	1118.72
	(CML495/CML494)/GUAT153	2.60	1038.73
	GUAT153_1861 (*per se*)	3.10	451.15
	(CML451/CML486)/GUER208	3.23	832.69
	(CML269/CML264)/GUAT823	3.34	552.37
	(CML269/CML264)/VERA115	3.48	604.83
	(CML269/CML264)/PUEB814	3.49	717.55
	(CML495/CML494)/VERAGP24	3.51	914.02
	(CML495/CML494)/SNLP328	3.60	753.26
10 más susceptibles	(CML373/CML311)/BRAZSC009	4.99	181.53
	(CML269/CML264)/RDOM249	4.99	372.79
	(CML373/CML311)/JALI183	4.99	416.47
	(CML269/CML264)/SINA176	4.99	312.29
	(CML269/CML264)/BRAZ90AR	5.00	334.05
	(CML269/CML264)/GUAT83	5.01	520.59
	(CML269/CML264)/GUYA812	5.01	328.36
	(CML269/CML264)/VENE541	5.02	410.36
	(CML269/CML264)/GUAT1094	5.02	326.18
	(CML269/CML264)/ARZM10072	5.05	394.51

*corregido por humedad del grano.

4.2 Genotipificación

Se identificaron un total de 2.2 millones de SNPs. Después del filtrado se obtuvo un total de 56,092 SNPs de alta calidad. Las bases de datos genotípicas originales están disponibles en http://hdl.handle.net/11529/10034 (Hearne *et al.*, 2014). La Figura 7 muestra la distribución de los SNPs para cada cromosoma, en la cual se puede observar una mayor concentración en las regiones teloméricas que en la región centromérica. El número de SNPs de alta calidad obtenidos fue similar al número utilizado en otros análisis GWAS (Weng *et al*., 2011; Wang *et al*., 2012; Liu *et al*., 2014; Shi *et al*., 2014). Los Cuadros 4 y 5 son ejemplos de los formatos HapMap y Numérico.

4.3 Heredabilidad genómica

A partir del ajuste de los modelos bayesianos GBLUP y RKHS (Ecuaciones 8 y 9), utilizando toda la distribución de datos fenotípicos, se obtuvieron los componentes de varianza. Para el modelo GBLUP se obtuvo $\hat{\sigma}_u^2 = 0.0572$ y $\hat{\sigma}_e^2 = 0.0188$, mientras que para el modelo RKHS se obtuvo $\hat{\sigma}_u^2 = 0.0910$ y $\hat{\sigma}_e^2 = 0.0190$. La heredabilidad se estimó utilizando la Ecuación (3):

$$\hat{h}^2_{GBLUP} = \frac{0.0572}{0.0572 \ + \ 0.0188} = 0.75 \qquad \hat{h}^2_{RKHS} = \frac{0.0910}{0.0910 \ + \ 0.0190} = 0.83$$

Ambas estimaciones de la heredabilidad son considerablemente altas (Carena *et al.*, 2010; Ornella *et al*., 2012), lo cual sugiere que la resistencia al CMA podría ser un carácter poligénico controlado probablemente por pocos QTL, por lo cual podría responder bien a selección recurrente convencional (Falconer y Mackay, 1996) y aún más si se desarrollan marcadores funcionales para la SAM y SG. Además, se espera que los resultados sean reproducibles en diferentes condiciones (Falconer y Mackay, 1996). Las estimaciones de la varianza residual en ambos modelos fueron considerablemente pequeñas, lo cual es coherente con el bajo coeficiente de variación de los BLUPs (C.V.= 6.64 %). La diferencia en las estimaciones de ambos modelos indica que el modelo RKHS capturó mayor variabilidad genética que el GBLUP, lo cual puede atribuirse al supuesto de que el modelo RKHS captura efectos epistáticos de forma implícita (Gianola y Kaam, 2008), por lo que podríamos referirnos al sentido amplio de la heredabilidad (H^2) (Falconer y Mackay, 1996).

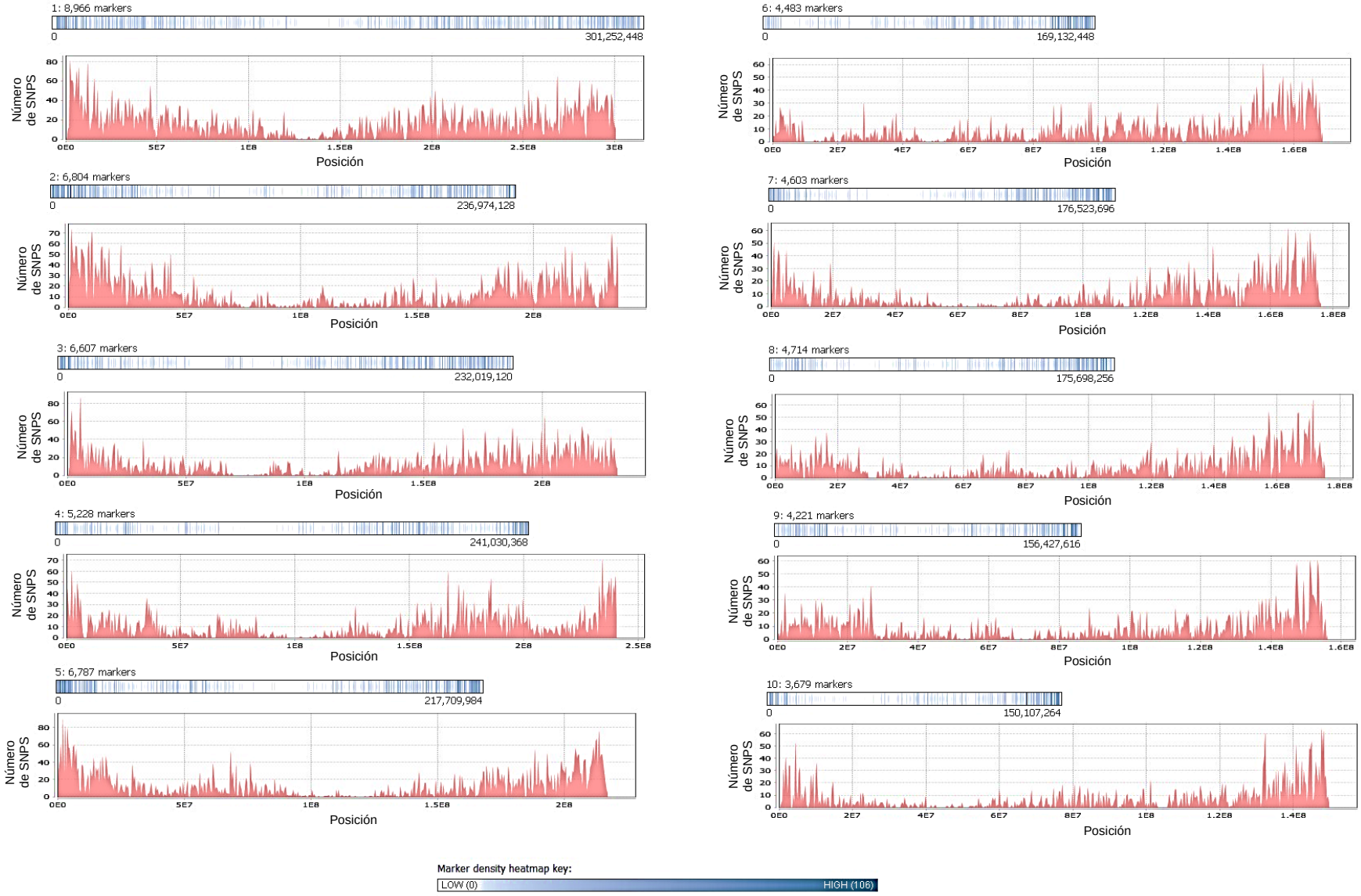

Figura 7. Distribución de los marcadores SNPs a lo largo de los 10 cromosomas del maíz.

Cuadro 4. Ejemplo de información genotípica en formato HapMap.

rs#	alleles	chrom	pos	VERA240_20401	VERA240_20401	VERA240_20401	VERA240_20401	VERA240_20401	VERA240_20401
S1_222473	T/C	1	222473	T	T	N	T	T	T
S1_222588	T/C	1	222588	T	T	T	T	N	T
S1_267769	T/C	1	267769	C	T	T	C	T	T
S1_1665404	G/C	1	1665404	G	N	S	G	N	G
S1_1665405	C/G	1	1665405	C	N	S	C	N	C
S1_1665408	A/C	1	1665408	A	N	M	A	N	A
S1_1763397	A/G	1	1763397	A	A	A	A	A	A
S1_1780419	A/G	1	1780419	A	G	A	A	A	G
S1_1787870	C/T	1	1787870	C	T	C	C	N	T
S1_1787871	C/T	1	1787871	N	T	C	C	N	T
S1_1787874	G/T	1	1787874	G	T	G	G	N	T
S1_1787875	C/T	1	1787875	C	T	C	C	N	T
S1_1787887	C/T	1	1787887	C	C	C	C	N	C
S1_1787892	C/T	1	1787892	C	T	C	C	N	T
S1_1860005	T/C	1	1860005	T	T	N	T	N	T
S1_2039864	T/G	1	2039864	T	N	T	T	N	T
S1_2087599	C/T	1	2087599	T	C	C	C	T	C
S1_2089585	C/G	1	2089585	S	S	S	S	C	G
S1_2091448	G/C	1	2091448	S	G	G	G	C	G
S1_2091497	G/A	1	2091497	G	G	G	G	G	G
S1_2306715	C/T	1	2306715	C	C	C	C	C	C
S1_2306727	C/G	1	2306727	C	C	C	C	C	C
S1_2307029	G/T	1	2307029	G	K	K	K	G	T
S1_2307053	T/G	1	2307053	K	K	T	K	T	T
S1_2307071	T/G	1	2307071	N	G	K	K	T	T
S1_2307134	T/G	1	2307134	K	T	K	T	T	K
S1_2469048	G/A	1	2469048	G	G	G	G	N	R
S1_2469056	C/T	1	2469056	Y	C	Y	C	N	C

Cuadro 5. Ejemplo de información genotípica en formato numérico.

Accesiones	S1_222473	S1_222588	S1_267769	S1_1665404	S1_1665405	S1_1665408	S1_1763397	S1_1780419	S1_1787870	S1_1787871
VERA240_20401	0	0	0	0	0	0	0	0	0	0
VENEM_212_22081	0	0	2	0	1	0	0	2	2	0
CHIS423_23203	0	0	2	1	1	1	0	0	0	2
CHIS434_23208	0	0	0	0	0	0	0	0	0	2
OAXA889_23585	0	0	2	0	2	0	0	0	2	2
CHIS471_24307	0	0	2	0	0	0	0	2	2	0
CHIS526_24953	0	0	1	2	0	0	0	0	2	0
QROO84_25055	0	0	0	0	0	0	1	1	0	2
GUER208_25081	0	0	2	0	0	0	0	0	0	2
VERA577_25093	1	1	2	0	0	0	0	0	2	0
ARZM07143_26757	2	0	2	2	2	2	0	0	0	2
CHIS675_26869	0	0	1	2	2	2	0	2	2	0
CHIS677_26873	1	1	2	0	0	0	0	0	0	2
HIDA240_29309	0	0	0	0	0	0	0	0	2	0
CAMP5_1767	0	0	2	0	0	0	0	0	0	2
HIDA293_29362	2	0	0	0	0	0	0	1	0	0
HIDA295_29364	0	0	1	0	0	0	0	0	2	0
HIDA298_29367	0	0	0	0	1	0	0	2	2	0
HIDA331_29400	0	0	1	0	0	0	1	1	0	2
SNLP299_29426	0	0	1	1	1	1	0	1	0	2
SNLP319_29445	0	0	0	1	1	1	0	0	2	0
SNLP323_29449	0	0	2	0	0	0	0	0	0	2
SNLP337_29463	0	0	2	1	1	1	0	0	0	2
GUAT97_1841	1	1	2	0	0	0	2	1	0	2
SNLP340_29466	0	0	0	0	0	0	0	2	2	0
SNLP371_29497	0	0	0	2	2	2	0	1	1	1
SNLP372_29498	1	0	0	1	0	1	0	2	0	2
VERA749_29505	0	0	0	0	0	0	0	0	0	0

4.4 Identificación de QTL asociados a la resistencia/susceptibilidad al CMA

A partir del análisis de componentes principales se graficó la dispersión de los primeros tres componentes principales, donde puede observarse un arreglo espacial que puede ser atribuido a la estructura de población (Yu *et al*., 2006; Gowda *et al*., 2015) (Figura 8), lo cual justifica su inclusión en el modelo mixto (ver Ecuación (4)) del método Single Marker para reducir la tasa de falsos-positivos.

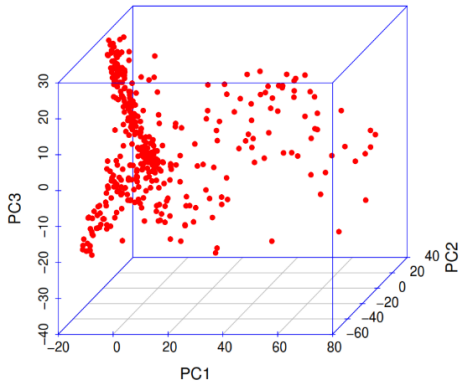

Figura 8. Estructura de población basada en los primeros tres componentes principales.

La gráfica cuantil-cuantil muestra los valores de p obtenidos en el análisis GWAS-Single Marker contra los valores esperados teóricamente bajo la hipótesis nula o no asociación a la resistencia al CMA en una escala de $-log_{10}(p\ \text{valor})$. Se puede observar que los valores estimados se ajustan casi perfectamente sobre la línea de los valores esperados. Esto sugiere que el modelo de asociación ha controlado de manera efectiva las asociaciones espurias (falsos-positivos) mediante la inclusión de la estructura de población y relaciones de parentesco y que solo aquellos pocos marcadores verdaderamente asociados a la resistencia se han desviado de los valores esperados (Figura 9).

Con el método Single Marker, implementado con TASSEL V.5.2.12, un total de 17 SNPs fueron significativamente asociados a la resistencia/susceptibilidad al CMA a un nivel de significancia $\alpha = 1 \times 10^{-4}$, equivalente a 4.0 en la escala de $-log_{10}(p\ \text{valor})$, distribuidos en los cromosomas 1, 2, 3, 7, 8 y 9, los cuales contienen 2, 3, 5, 2, 4 y 1 SNPs respectivamente (Figura 10 y Cuadro 6). Por su parte, con el método BayesB se seleccionaron los primeros 17 SNPs con mayor probabilidad posterior de inclusión,

que es la probabilidad que un marcador tenga efecto diferente de cero (Berg *et al*., 2013; Pérez-Rodríguez y de los Campos, 2014) (ver Figura 11), los cuales se encuentran distribuidos a lo largo de los 10 cromosomas (Cuadro 7) excepto en el 9, distribuidos de la siguiente manera: 6, 2, 3, 1, 1, 1, 1, 1, 1 SNPs respectivamente del cromosoma 1 al 10, lo cual sugiere que el cromosoma 1 está fuertemente asociado a la resistencia al CMA por el número de SNPs identificados.

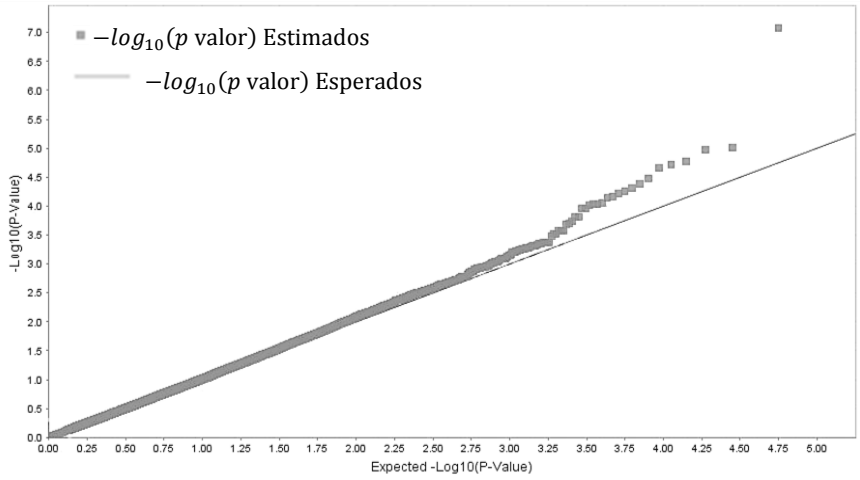

Figura 9. Gráfica cuantil-cuantil mostrando los $-log_{10}(p\ \text{valor})$ estimados contra los esperados.

Al comparar los marcadores identificados entre metodologías, resulta que los SNPs S1_52921129, S2_188057674, S3_124181320 y S7_175205315 fueron detectados por ambos métodos. Esto sugiere que dichos marcadores están fuertemente asociados a la resistencia al CMA. De estos cuatro SNPs, el S7_175205315 mostró el mayor nivel de significancia (valor de p= 8.25×10^{-8}) con el método Single Marker, mientras que con el método BayesB los SNPs S7_175205315 y S1_52921129 obtuvieron los valores más altos de probabilidad posterior de inclusión, 0.983 y 1.0, respectivamente.

Para explicar el efecto de los marcadores es necesario recordar que en la escala utilizada para evaluar la enfermedad el 0 indica el nivel más alto de resistencia y el 5 indica el nivel más susceptible. Además los efectos de los marcadores fueron estimados como una desviación de la media, por lo que un efecto con signo negativo debe considerarse como favorable ya que reduce el nivel de susceptibilidad en dicha

escala, mientras que un efecto con signo positivo debe considerarse desfavorable, ya que aumenta el nivel de susceptibilidad.

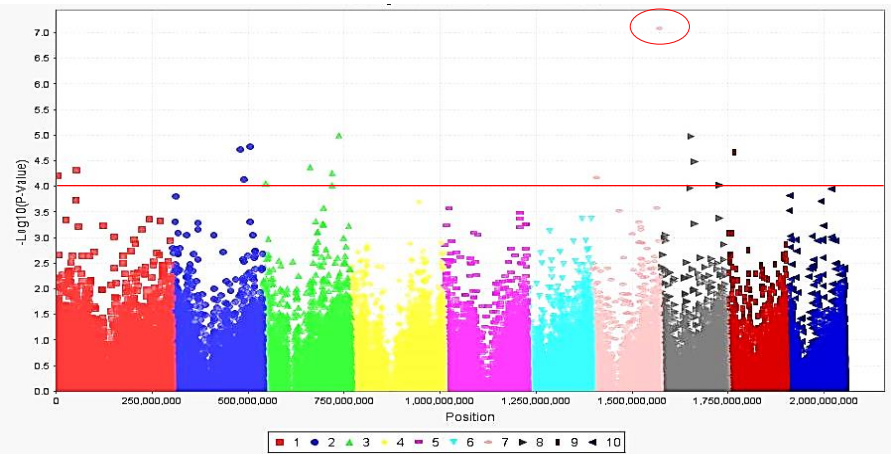

Figura 10. Gráfico Manhattan del modelo lineal mixto Single Marker (TASSEL) para resistencia al CMA. Los diferentes colores y símbolos indican a que cromosoma pertenece cada SNP.

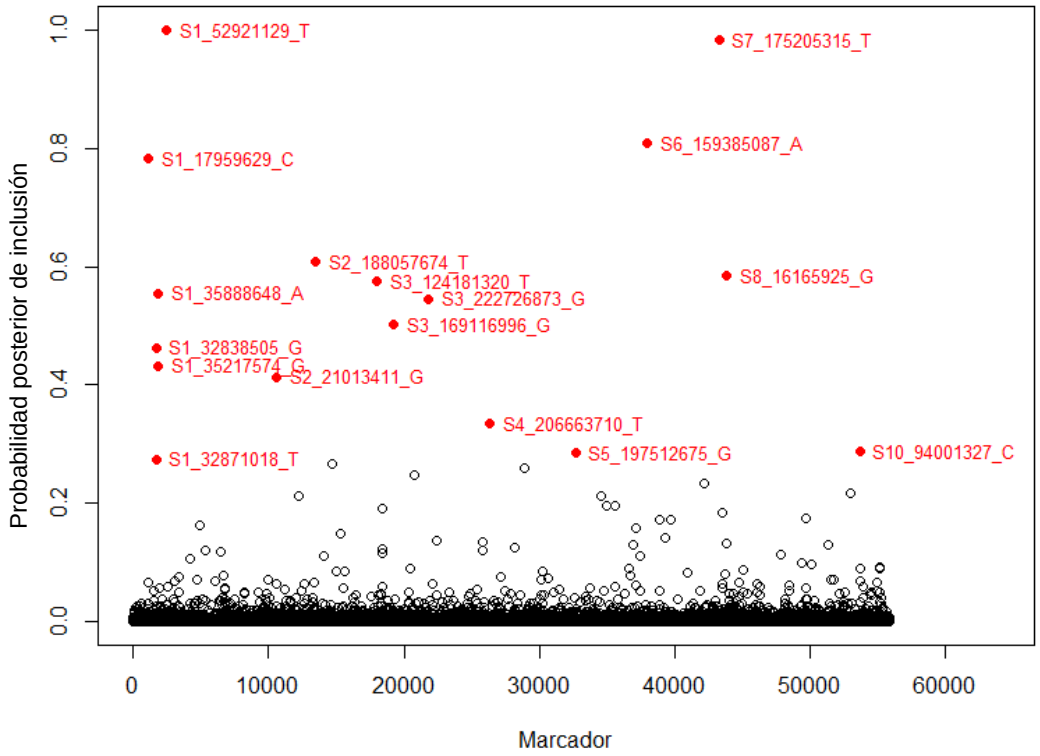

Figura 11. Probabilidad posterior de inclusión para los marcadores identificados con BayesB.

En general, puede observarse que los marcadores más significativos también son los que tienen mayor efecto en ambas metodologías. En el caso de Single Marker el mayor efecto estimado fue de -0.41 para resistencia y 0.24 para susceptibilidad. En el caso de BayesB, el rango de los efectos estimados fue de -0.1083 para resistencia y 0.1009 para susceptibilidad. Es interesante notar como el efecto de la mayoría de marcadores se contrae hacia cero (Figura 12) debido al punto de masa en cero que fue definido como parte de la distribución mixta *a priori* en el modelo BayesB, mientras que muy poco marcadores con efecto diferente de cero siguen una distribución condicional *t*-escalada (Meuwissen *et al*., 2001; Pérez-Rodríguez y de los Campos, 2014). Es notorio que los efectos estimados mediante BayesB en general son más pequeños que los estimados por Single Marker, lo cual se discute posteriormente (ver inciso 5.4).

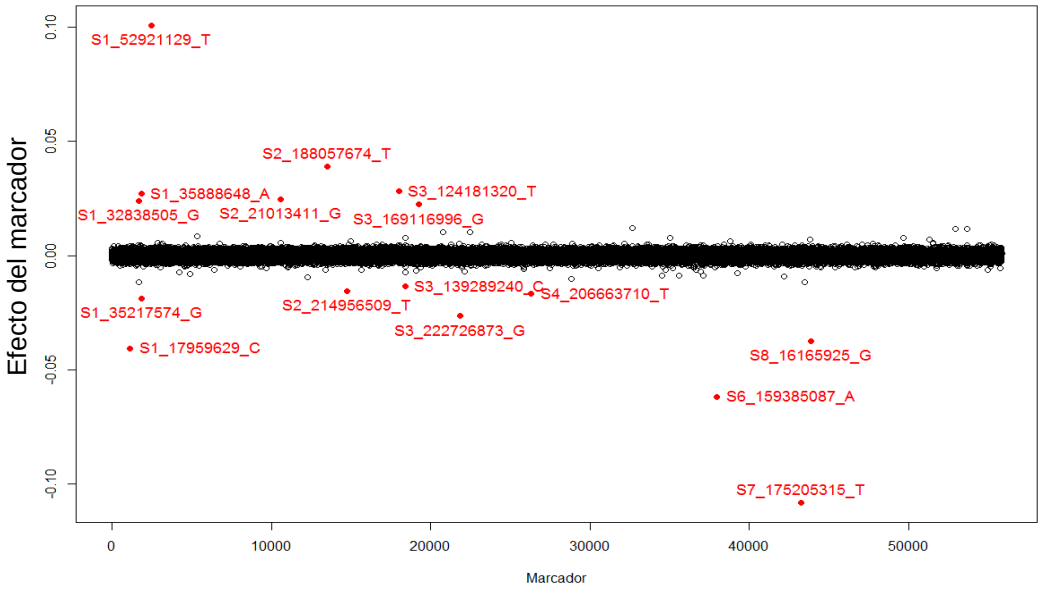

Figura 12. Efecto de marcadores ajustado con el modelo BayesB.

La proporción de variabilidad fenotípica explicada de manera individual por cada marcador identificado con el método Single Marker varía entre 2.81 y 5.2 % (Cuadro 6). En el caso del método BayesB, por tratarse de una regresión múltiple no es posible obtener el coeficiente de determinación r^2 para cada marcador de manera individual. Para poder comparar ambos métodos se estimó la proporción de varianza genética total explicada en conjunto por los QTL ligados a los SNPs significativos de forma simultánea para ambas metodologías (ver inciso 4.5).

Cuadro 6. Marcadores asociados a la resistencia/susceptibilidad al CMA identificados con la metodología Single Marker.

Marcador	Crom.	Posición	Valor de p	r^2 (%)*	P_G (%)**	Alelo	Efecto del Alelo
S7_175205315	7	175205315	8.25E-08	5.20	6.27	T	-0.4128
S3_199023102	3	199023102	9.93E-06	3.83	4.61	T	0.2350
S8_81247760	8	81247760	1.07E-05	4.17	5.02	T	0.2294
S2_204082490	2	204082490	1.68E-05	3.35	4.04	T	-0.1362
S2_178491089	2	178491089	1.91E-05	3.89	4.69	G	0.2245
S9_18784058	9	18784058	2.21E-05	3.43	4.13	T	0.2763
S8_90189319	8	90189319	3.28E-05	3.47	4.18	C	-0.1478
S3_124181320	3	124181320	4.18E-05	3.01	3.63	T	0.0443
S1_52921129	1	52921129	4.91E-05	2.94	3.54	T	0.0648
S3_181818060	3	181818060	5.52E-05	2.98	3.59	G	-0.1502
S1_4446829	1	4446829	6.13E-05	3.26	3.93	G	0.1100
S7_11442611	7	11442611	6.77E-05	2.92	3.52	G	-0.1324
S2_188057674	2	188057674	7.27E-05	2.87	3.46	T	0.0282
S3_9042597	3	9042597	8.71E-05	3.06	3.69	T	-0.1753
S8_155905444	8	155905444	9.35E-05	3.30	3.98	G	0.1377
S8_155905445	8	155905445	9.35E-05	3.30	3.98	T	-0.4128
S3_181818056	3	181818056	9.70E-05	2.81	3.39	T	0.2350
Total r^2 y P_G (%)				20	24		

* r^2 (%), proporción de varianza fenotípica explicada por el QTL ligado al marcador de manera individual.
** P_G (%), proporción de varianza genética explicada por el QTL ligado al marcador de manera individual.

Cuadro 7. Marcadores asociados a la resistencia/susceptibilidad al CMA identificados con la metodología BayesB.

Marcador	Cromosoma	Posición	Probabilidad *a posteriori*	Alelo	Efecto del Alelo
S1_52921129	1	52921129	1	T	0.1009
S7_175205315	7	175205315	0.983	T	-0.1083
S6_159385087	6	159385087	0.808	A	-0.0618
S1_17959629	1	17959629	0.782	C	-0.0406
S2_188057674	2	188057674	0.608	T	0.0390
S8_16165925	8	16165925	0.584	G	-0.0375
S3_124181320	3	124181320	0.575	T	0.0281
S1_35888648	1	35888648	0.554	A	0.0270
S3_222726873	3	222726873	0.545	G	-0.0265
S3_169116996	3	169116996	0.501	G	0.0223
S1_32838505	1	32838505	0.462	G	0.0238
S1_35217574	1	35217574	0.432	G	-0.0187
S2_21013411	2	21013411	0.413	G	0.0246
S4_206663710	4	206663710	0.334	T	-0.0168
S10_94001327	10	94001327	0.287	C	0.0117
S5_197512675	5	197512675	0.284	G	0.0122
S1_32871018	1	32871018	0.273	T	-0.0118
P_G= 49 %					

4.5 Proporción de la varianza genética explicada por los QTL

Los coeficientes de determinación r^2, obtenidos mediante el ajuste de los modelos de regresión lineal múltiple, uno para cada metodología, en los cuales se incluyeron únicamente los marcadores significativamente asociados a la resistencia al CMA, fueron $r^2_{S.M.} = 0.20$ y $r^2_{BayesB} = 0.41$, para Single Marker y BayesB, respectivamente. Se utilizó la heredabilidad genómica estimada con el modelo RKHS, ya que capturó mayor varianza genética, incluyendo efectos aditivos y no aditivos. La estimación de la proporción de varianza genética explicada en conjunto por los QTL ligados a los marcadores identificados con cada método se realizó con base a la Ecuación (6):

$$P_{G(S.M.)} = {}^{0.20}/_{0.83} = 0.24 \qquad P_{G(\,BayesB)} = {}^{0.41}/_{0.83} = 0.49$$

Los resultados indican que los QTL ligados a los marcadores identificados con el método Bayes B explican una mayor proporción de la varianza genética (25 % más que el método Single Marker), lo cual indica que BayesB fue más preciso en detectar QTL que influencian directamente la variación del nivel de severidad (Sahana *et al*., 2010), aún bajo estructura de población (Ornella *et al*., 2012). Posteriormente se discute esta diferencia en la precisión de detección de QTL en función de los modelos utilizados (ver inciso 5.5). El 51 % restante de la varianza genética no explicada sugiere que hay muchos más QTL de efectos aditivos muy pequeños que también contribuyen a la resistencia al CMA pero que no fueron detectados con ninguno de los dos métodos utilizados para el análisis GWAS.

4.6 Análisis de genes candidatos

El Cuadro 8 contiene los genes candidatos asociados a los SNPs identificados en el análisis GWAS y su posible función de acuerdo a las base de datos NCBI, MaizeCyc y CornCyc. La mayor parte de estos SNPs fueron encontrados dentro de la secuencia de los genes candidatos y muy pocos se encontraron adyacentes a ellos.

Uno de los hallazgos más importantes de este estudio fue la identificación del gen AC194670.2_FG001 asociado al SNP S4_206663710, identificado con el método BayesB en el análisis GWAS. Este gen se ubica en el cromosoma 4 (Bin 4.09) y se le atribuye la producción de la enzyma 3-hydroxybutyryl-CoA dehydratase según la base

de datos MaizeCyc, la cual está involucrada directamente en la respuesta a resistencia a enfermedades en plantas según la base de datos NCBI. Otro importante hallazgo fue la identificación del SNP S1_52921129 (Bin 1.04), el cual fue identificado con ambos métodos, pero con BayesB obtuvo la mayor probabilidad posterior de inclusión. Este SNP se encontró dentro del gen candidato GRMZM2G030272, asociado a la producción de un factor de transcripción conocido como *WRKY*, el cual es un mediador de la regulación positiva de la expresión de genes de resistencia en general (Mohr *et al*., 2010; Michelmore *et al*., 2013). Otro factor de transcripción importante fue el *MAD-box* (Lee *et al*., 2008; Wang *et al*., 2012), asociado al gen candidato GRMZM2G059102 y al SNP S1_17959629 (Bin 1.02). Además, se identificó el factor de transcripción *Leucine-zipper* en el gen candidato GRMZM2G041127, asociado al SNP S2_188057674 (Bin 2.07), identificado como responsable de desencadenar una respuesta de defensa sistémica en la planta para una variedad de patógenos (Després *et al.*, 2000; Liu *et al*., 2014). Dado que el nivel de expresión de los genes de resistencia es crítico para el desarrollo de la enfermedad, los factores de transcripción juegan un papel importante, ya que si la expresión es demasiado lenta la resistencia y la planta se encuentra más vulnerable (Mohr *et al*., 2010; Michelmore *et al*., 2013).

Otros genes candidatos asociados a la producción de proteínas receptoras de señal involucrados en mecanismos de resistencia a enfermedades fueron encontrados también, entre ellos: GRMZM2G109360, GRMZM2G061602, AC212835.3_FG008 y AC205982.3, a los cuales se les atribuye la producción de proteínas de la familia kinasa o proteínas de actividad serine/threonine kinasa, agrupadas como RPK (receptor protein kinases, en inglés) o STK (Serine/threonine kinasa) según la base de datos del NCBI. Estas juegan un papel importante en el complejo de interacción de señalización durante la percepción de patógenos y subsecuente activación de respuesta de defensa (Romeis, 2001). Otro grupo de genes candidatos que producen proteínas con dominios LRR (Leucine-rich repeat, e inglés) fueron encontrados, tales como GRMZM2G073884 y GRMZM2G151738, los cuales están involucradas en la detección de patógenos y respuesta de defensa, especialmente en enfermedades causada por hongos (Michelmore *et al*., 2013; Shi *et al*., 2014). Otro gen candidato encontrado fue el GRMZM2G136859 asociado al SNP S1_4446829 en el cromosoma 1 (Bin 1.01), el

cual produce una proteína con función transportador de metales pesados y permite la detoxificación de algunas toxinas producidas por patógenos (Manara, 2012; Michelmore *et al.,* 2013). Otros genes putativos, GRMZM2G049878, GRMZM2G323622, GRMZM2G401308, GRMZM2G055629 y GRMZM2G087625, codificando para NBS (nucleotide-binding site, en inglés), ATP binding, ice binding (anticongelante), GTPase y Cysteine-rich receptor-like protein kinase (RLK, IRAK) respectivamente, todos relativos a resistencia a enfermedades en plantas (Griffith y Yaish, 2004; Michelmore *et al*., 2013; Liu *et al*., 2014).

La Figura 13 muestra un ejemplo de la identificación del gen candidato GRMZM2G030272 asociado al SNP S1_52921129 en el buscador del genoma de referencia "B73" RefGen_v2 en la base de datos MiazeGDB.

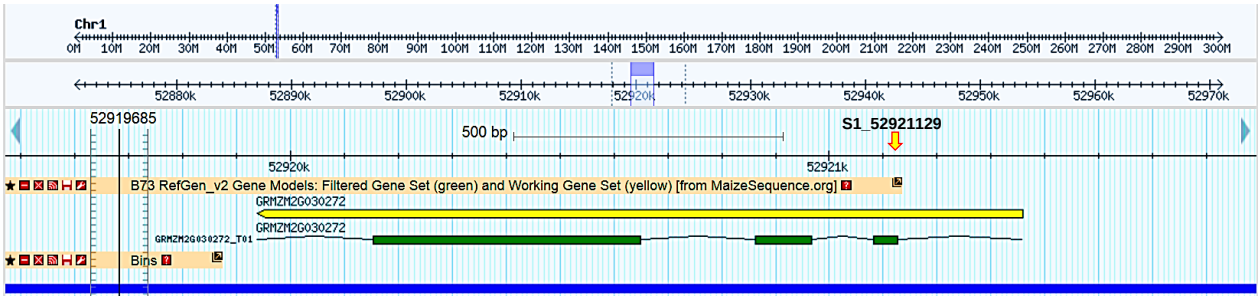

Figura 13. Ubicación de gen candidato GRMZM2G030272 asociado al SNP S1_52921129 en el genoma de referencia "B73" RefGen_v2.

4.7 Selección Genómica

La mayor precisión en predicción en la partición en prueba se obtuvo utilizando únicamente los 17 SNPs identificados con el método BayesB como variables predictoras, ajustados mediante mínimos cuadrados ordinarios ($r_{prueba} = 0.61$ y $CME_{prueba} = 0.0525$), obteniendo una correlación 18 % mayor que el resto de modelos y el menor cuadrado medio del error, en la partición de prueba (Cuadro 9). La correlación obtenida con el modelo ajustado con los 17 SNPs identificados con single marker fue muy cercana a la de los modelos bayesianos que utilizan todos los marcadores simultáneamente ($r_{prueba} = 0.40$ y $CME_{prueba} = 0.0703$).

Cuadro 8. SNPs significativamente asociados al a resistencia al CMA, genes candidatos y su posible función.

SNP	Método	Gen candidato*	Bin***	Posible función**
S1_52921129	BayesB y S.M.	GRMZM2G030272	1.04	Superfamily of TFs having WRKY and zinc finger domains
S1_17959629	BayesB	GRMZM2G059102	1.02	MADS-box transcription regulation factor family protein
S1_32838505	BayesB	GRMZM2G147942	1.03	Negative regulation of catalytic activity
S1_32871018	BayesB	GRMZM2G077942	1.03	Actin depolymerizing factor 5
S1_35217574	BayesB	GRMZM2G030223	1.03	Lipid metabolic process
S1_35888648	BayesB	GRMZM2G174615	1.03	Putative uncharacterized protein
S10_94001327	BayesB	GRMZM2G066044	10.04	G-protein coupled receptor signaling pathway, DNA binding, histamine receptor activity
S2_188057674	BayesB y S.M.	GRMZM2G041127	2.07	Leucine zipper protein trasncription factor
S2_21013411	BayesB	AC212835.3_FG008	2.03	Receptor-like protein kinase At5g47070
S3_124181320	BayesB y S.M.	GRMZM2G063688	3.04	Galactosyltransferase family protein
S3_169116996	BayesB	GRMZM2G328795	3.06	Tetratricopeptide repeat (TRP)
S3_222726873	BayesB	GRMZM2G055629	3.09	G-protein GTPase
S4_206663710	BayesB	AC194670.2_FG001	4.09	Disease resistance response protein
S5_197512675	BayesB	GRMZM2G109360	5.06	Protein kinase superfamily protein
S6_159385087	BayesB	AC205982.3	6.06	Serine/threonine-protein kinasa (STK)
S7_175205315	BayesB y S.M.	GRMZM2G440866	7.06	RING zinc finger domain superfamily protein
S8_16165925	BayesB	AC212565.3_FG001	8.02	Core histone H2A/H2B/H3/H4 domain
S1_4446829	Single Marker	GRMZM2G136859	1.01	Heavy metal transport/detoxification superfamily protein
S2_178491089	Single Marker	GRMZM2G101600	2.06	Formate dehydrogenase
S2_204082490	Single Marker	GRMZM2G087625	2.07	Cysteine-rich receptor-like protein kinase, RLK,IRAK
S3_181818056	Single Marker	GRMZM2G049877	3.06	Enzima Transporter: Nucleoside-triphosphatase ATP binding
S3_181818060	Single Marker	GRMZM2G049878	3.06	Nucleotide binding site (NBS), Nucleoside-triphosphatase activity
S3_199023102	Single Marker	GRMZM2G323622	3.07	ATP binding
S3_9042597	Single Marker	GRMZM2G158194	3.03	PHD-transcription factor 34
S7_11442611	Single Marker	GRMZM2G401308	7.01	Histone H1-like protein, ice binding, DNA binding
S8_155905444	Single Marker	GRMZM2G061602	8.06	Protein serine/threonine kinase activity (STK), ATP binding
S8_155905445	Single Marker	GRMZM2G061603	8.06	Protein serine/threonine kinase activity (STK), ATP binding, protein phosphorylation
S8_81247760	Single Marker	GRMZM2G073884	8.03	Leucine-rich repeat (LRR) receptor-like protein kinase
S8_90189319	Single Marker	GRMZM2G151738	8.03	Leucine-rich repeat (LRR) receptor-like protein kinase
S9_18784058	Single Marker	GRMZM2G105909	9.02	Hipotetical Protein uncharacterized

*Gen candidato o putativo: Basado en el genoma de referencia “B73” RefGen_v2 (MGSC) (http://www.maizegdb.org/gbrowse).
** Posible Función: Se utilizaron las bases de datos MaizeCyc, CornCyc y NCBI.
***El mapa genético del maíz fue dividido en 100 segmentos llamados Bins, cada uno de los cuales fue definido por dos marcadores centrales.

Se esperaba que al combinar los marcadores identificados con single marker y BayesB la precisión de las predicciones incrementara considerablemente, sin embargo, la ganancia en precisión no fue relevante ($r_{prueba} = 0.62$ y $CME_{prueba} = 0.0518$).

Por otra parte, los modelos bayesianos mostraron similar nivel de precisión entre ellos, excepto el modelo BayesB. El promedio de las predicciones en la partición de prueba fue de 0.39, muy cercana a la obtenida con el modelo ajustado con los 17 marcadores identificados con single marker (0.40), pero mucho menor que el modelo que incluye los 17 marcadores del BayesB (0.61). Entre los modelos bayesianos, el GBLUP produjo la mayor correlación en prueba (0.45), mientras que con el modelo BayesB se obtuvo la más baja (0.26) y mayor CME, lo cual puede atribuirse a la baja proporción de marcadores con efecto diferente de cero definida *a priori* (0.001) (ver inciso 3.8.2), por lo que la mayor parte de marcadores fue considerado con efecto nulo (Pérez-Rodríguez y de los Campos, 2014).

Cuadro 9. Comparación de la precisión de predicción de los modelos de selección genómica.

Modelo de Selección Genómica	No. SNPs	CME. Entren.	CME. Prueba	*r* Entren.	*r* Prueba
Regresión Ridge Bayesiana	~56 mil	0.0045	0.0731	0.9958	0.4248
Modelo GBLUP	~56 mil	0.0043	0.0674	0.9963	0.4499
Modelo BayesB	~56 mil	0.0224	0.078	0.9186	0.2591
RKHS – Semiparamétrico	~56 mil	0.0045	0.0694	0.9976	0.4323
QTL Single Marker	17	0.0640	0.0703	0.4733	0.4022
QTL BayesB	17	0.0476	0.0525	0.6500	0.6119
QTL BayesB+SingleMarker	30	0.0446	0.0518	0.6773	0.6158

En general, puede observarse que las correlaciones en la partición de entrenamiento tienden a ser más altas que en la partición de prueba. Este comportamiento es conocido como sobreajuste (overfiting, en inglés) e implica que los modelos son buenos para predecir los fenotipos de los individuos que se incluyeron en la población de entrenamiento para ajuste del modelo, pero al predecir fenotipos de individuos no observados la correlación se reduce considerablemente (Nakaya y Isobe, 2012). El sobreajuste puede considerarse normal (Dietterich, 1995), tal como ocurre en los modelos ajustados solo con los SNPs identificados en el análisis GWAS (Figura 14-a). Sin embargo, en el caso de los modelos bayesianos que ajustan todos los marcadores

simultáneamente se observa un exceso de sobreajuste ($r_{entr.} = 0.99$) (Figura 14-b), a pesar de que incluyen un término de penalización para evitarlo (Crossa *et al*., 2010). Un comportamiento similar de sobreajuste ha sido observado utilizando ~ 3000 SNPs (Endelman, 2011). El sobreajuste se produce porque al agregar un exceso de marcadores que no están asociadas a la resistencia al CMA al modelo, el nivel de complejidad aumenta, lo cual produce imprecisión en las predicciones (Nakaya y Isobe, 2012).

En la Figura 14-a puede apreciarse un nivel de sobreajuste normal (Dietterich, 1995) para el caso del modelo ajustado con los 17 SNPs identificados con BayesB en el análisis GWAS, ya que la dispersión de los datos en la partición de entrenamiento es mayor comparado con el alto grado de correlación obtenida con modelo GBLUP (Figura 14-b).

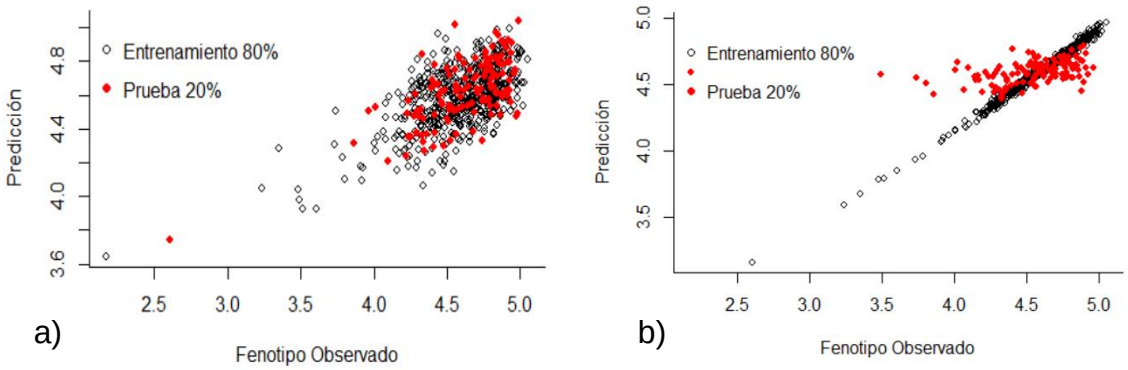

Figura 14. Correlación entre datos observados y predichos para las particiones de entrenamiento y prueba. a) Modelo ajustado con los 17 SNPs identificados con el método BayesB en el análisis GWAS. b) Modelo bayesiano GBLUP que incluye todos los ~56 mil SNPs.

Para analizar más a detalle el efecto de sobreajuste de los modelos bayesianos se evaluaron diferentes densidades de marcadores seleccionados al azar, cada densidad evaluada con y sin los 17 SNPs identificados con el modelo BayesB. Para el ajuste de las diferentes densidades se utilizó el modelo GBLUP. En la Figura 15 puede apreciarse un incremento de la brecha entre las correlaciones en las particiones de entrenamiento y prueba, lo cual representa el grado de sobreajuste que se produce a medida que se incrementa la densidad de marcadores (Dietterich, 1995). Además,

puede apreciarse como las correlaciones en prueba de la densidad de marcadores que incluyen los 17 SNPs significativos va decreciendo a medida que se van agregando marcadores. Este mismo comportamiento también ha sido observado en otros estudios (Daetwyler *et al*., 2014; Gowda *et al*., 2015), lo cual indica que el potencial predictivo de los 17 SNPs se va diluyendo en la medida que se agregan más marcadores. Además se observa que a partir de una densidad de 15,000 SNPs, la adición de más marcadores al modelo ya no produce una ganancia significativa en la precisión de predicción, lo cual concuerda con los resultados de Nakaya y Isobe (2012), quienes indican que el uso de demasiado marcadores generalmente no mejoran significativamente la ganancia en precisión, inclusive podría conducir a una pérdida en la precisión de las predicciones.

La Figura 15 también permite explicar el resto de la proporción de varianza genética no explicada por los 17 QTL identificados con BayesB en el análisis GWAS. Si observamos la correlación en prueba sin incluir los 17 QTL (línea r-prueba), se alcanza una correlación de ~0.48 a una densidad de 56 mil SNPs, lo que es equivalente a una proporción de variabilidad fenotípica $r^2 = 0.23$, que corresponde a un 28 % de la varianza, debida a la acumulación de los efectos aditivos infinitesimales de los *loci* ligados al conjunto denso de marcadores a lo largo del genoma. Esto explica un poco más de la mitad de la proporción de varianza genética no explicada por los 17 QTL.

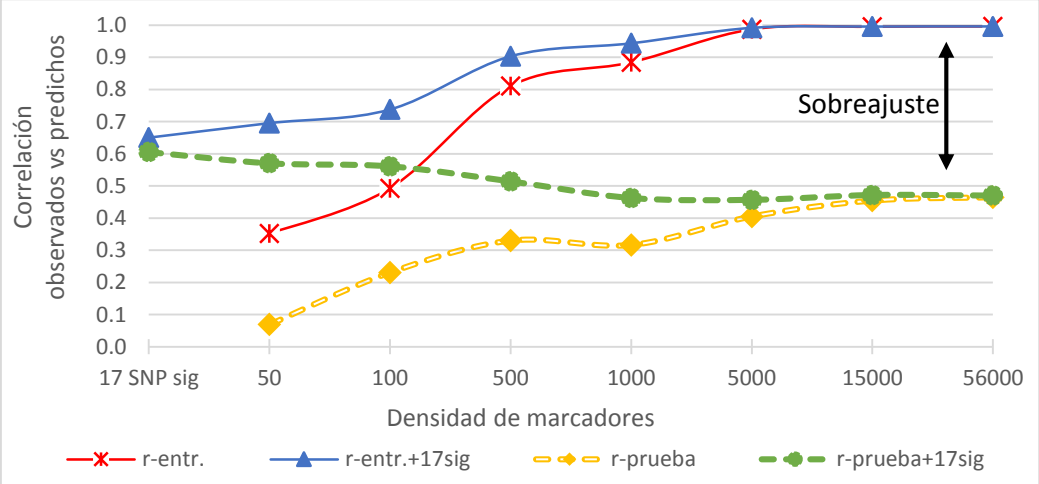

Figura 15. Nivel de sobreajuste a diferente densidad de marcadores ajustados con el modelo bayesiano GBLUP.

5 DISCUSIÓN

5.1 Evaluación fenotípica

La distribución cuantitativa del nivel de severidad (BLUPs) sugiere la resistencia al CMA es de herencia poligénica (Shi *et al*., 2014) y probablemente está controlado por pocos QTL. Además, esto explica porque ningún genotipo inmune fue identificado. La baja variabilidad observada en los datos (C.V.= 6.64 %) puede ser atribuida en alguna medida al control de la variabilidad ambiental y espacial incluidos en el modelo lineal mixto para la obtención de los BLUPs (Ecuación (2)). Los BLUPs también han sido utilizados en otros estudios (Wang *et al*., 2012; Daetwyler *et al*., 2014). A partir de la evaluación fenotípica sólo dos mestizos fueron identificados con alta tolerancia al CMA y potencial de rendimiento, los cuales fueron derivados de las accesiones Guat153 y Oaxa280, por lo que estas dos accesiones *per se* pueden ser utilizadas como potenciales fuentes de resistencia. La gran mayoría de mestizos evaluados fueron susceptibles y extremadamente susceptible, esto debido posiblemente a que se trata de un complejo de hongos los que interactúan, sin embargo, el hecho de que *M. maydis* se vuelve virulento únicamente en presencia de *P. maydis* facilita la identificación de fuentes de resistencia a la enfermedad, ya que cualquier resistencia que impida la invasión de *P. maydis*, también será efectivo contra *M. maydis* de forma indirecta (Ceballos y Deutsch, 1992).

5.2 Genotipificación

En cuanto a la distribución de los SNPs a lo largo del genoma, existe una tendencia a concentrarse más en las regiones cercanas a los telómeros que en las regiones cercanas al centrómero. Esta tendencia es normal debido a la ubicación física del centrómero en el cromosoma y por estar constituido principalmente por heterocromatina, por lo que la probabilidad de que ocurra un entrecruzamiento en las regiones cercanas al centrómero es muy baja y por lo tanto el número de polimorfismos que se producen es muy bajo (Talbert y Henikoff, 2010). Este comportamiento también fue observado por Romay *et al*. (2013), quienes utilizaron 620,279 SNPs y 4,351 muestras.

5.3 Heredabilidad genómica

La habilidad del modelo RKHS para capturar mayor variabilidad genética que el GBLUP se debe al hecho de que el kernel Gaussiano utilizado por el RKHS captura implícitamente varianza genética que incluye epistasis (Gianola y Kaam, 2008; Morota y Gianola, 2014), lo cual fue demostrado recientemente de forma explícita por Jiang y Reif (2015). Por consiguiente, es posible que la diferencia con respecto a la heredabilidad estimada con el GBLUP se deba a los efectos no aditivos (dominancia y epistasis), los cuales también fueron identificados por Ceballos y Deutsch (1992), por lo que la heredabilidad estimada con el modelo RKHS podría corresponder al sentido amplio de la heredabilidad, la cual considera la varianza aditiva y no aditiva (Falconer y Mackay, 1996). Por lo tanto, si se considera la heredabilidad estimada con el modelo RKHS como la varianza genética total (incluyendo efectos aditivos y no aditivos), y la heredabilidad estimada con el modelo GBLUP como la varianza genética estrictamente aditiva, entonces la varianza genética total se podría desglosar de la siguiente manera: 90 % corresponde a efectos aditivos y el 10 % restante a efectos no aditivos (epistáticos y dominante). Esto es congruente con los resultados obtenidos por Hernández (2014), quien encontró que la resistencia al CMA está controlada principalmente por los efectos aditivos. En otros estudios de resistencia a enfermedades en maíz también se han estimado valores altos de heredabilidad en sentido amplio, tal como carbón de la espiga $H^2 = 0.88$ (Wang *et al*., 2012), mancha gris de la hoja $H^2 = 0.88$ (Shi *et al*., 2014), enanismo rugoso $H^2 = 0.83$ (Liu *et al*., 2014), necrosis letal $H^2 = 0.73$ (Gowda *et al*., 2015) y tizón norteño $H^2 = 0.71$ (Poland *et al.*, 2011), lo cual marca un patrón de heredabilidad para enfermedades causadas por hongos en maíz.

La alta heredabilidad observada también refleja la alta calidad de los datos obtenidos de la evaluación fenotípica, lo cual es un factor a favor para aumentar el poder de detección de QTL (Gowda *et al*., 2015). Subsecuentes evaluaciones a través de diferentes ambientes deberían realizarse en el futuro para validar la heredabilidad estimada a partir de la información de marcadores (Weng *et al*., 2011; Wang *et al*., 2012; Liu *et al*., 2014).

5.4 Identificación de QTL asociados a la resistencia/susceptibilidad al CMA

El estudio de asociación del genoma completo o GWAS, resultó ser una potente herramienta para la identificación de marcadores moleculares asociados a la resistencia al CMA en ambas metodologías utilizadas. En este estudio se identificaron 17 SNPs altamente significativos ($\alpha = 0.0001$) con el método Single Marker y otros 17 SNPs con el método BayesB. El hecho de que los SNPs S1_52921129, S2_188057674, S3_124181320 y S7_175205315 fueron detectados por ambas metodologías sugiere que estos marcadores están fuertemente asociados al CMA. Por otra parte, aunque ambas metodologías de identificación de QTL tuvieron cuatro marcadores en común, la diferencia entre los métodos se evidenció al momento de estimar la proporción de varianza genética explicada por los QTL en conjunto y en la precisión de las predicciones, donde los marcadores identificados con BayesB fueron más sobresalientes. Esto se discute con más detalle en el inciso 5.5.

Por otra parte, considerado que el número de SNPs significativos identificados en este estudio mediante el análisis GWAS fue bajo, la resistencia al CMA es un carácter simple en cuanto a su base genética comparada con otros análisis GWAS realizados en otras enfermedades fungosas, tal como el tizón norteño del maíz, donde se identificaron más de 200 SNPs (Poland *et al.*, 2011), u otros caracteres, tal como la altura de planta, donde se identificaron un total de 204 SNPs (Weng *et al*., 2011). Sin embargo, se requieren más evaluaciones a través de diferentes ambientes para confirmar la consistencia de los QTL identificados (Weng *et al*., 2011; Wang *et al*., 2012; Liu *et al*., 2014).

En cuanto al efecto de los marcadores, con el método Single Marker el mayor efecto aditivo estimado fue de -0.41 para resistencia, mientras que con BayesB los efectos fueron más pequeños, con un rango de -0.1083 para resistencia y 0.1009 para susceptibilidad, esto resalta una diferencia en la estimación de los efectos por ambas metodologías. La identificación de QTL con efectos pequeños o moderados tiende a ser una característica propia del análisis GWAS, ya que compara variantes genéticas comunes al conjunto de individuos analizados, mientras que el tradicional mapeo de ligamiento a partir de poblaciones bi-parentales es menos sensible para detectar

variantes genéticas comunes, además la estimación de los efectos puede variar entre poblaciones y no ser consistentes (Melchinger *et al.*, 1998; Witte, 2010; Liu *et al*., 2014). En este sentido, la utilización de un conjunto diverso de poblaciones nativas de maíz favorece que la estimación de los efectos en general sea más consistente a través de las poblaciones en comparación si el análisis se hubiera realizado a partir de una población bi-parental.

Respecto a la diferencia entre ambos métodos en la estimación de los efectos, considerando que los caracteres cuantitativos son influenciados probablemente por un gran número de QTL simultáneamente, los modelos que analizan todos los marcadores simultáneamente (multiple-QTL, en inglés) deberían ser más precisos en la detección de QTL que los modelos que analizan un solo marcador a la vez (single-QTL, en inglés) (Berg *et al.,* 2013). Además, Yi y Xu (2008) indican que el método Single Marker (single-QTL) no provee estimaciones precisas de los efectos de los QTL ya que pocas veces es el modelo correcto para la mayoría de caracteres cuantitativos. Por consiguiente los modelos multiple-QTL (ej. BayesB, BayesC y LASSO) son una opción más razonable (Kao *et al.,* 1999). Varios estudios de simulación han demostrado que los modelos que analizan todos los marcadores de manera simultánea son más precisos en la deteción de QTL y menos falsos positivos que los modelos single-QTL (Kao *et al.,* 1999; Sahana *et al.,* 2010; Berg *et al.,* 2013). En el caso particular del modelo BayesB, este se caracteriza por tener una distribución *a priori* mixta para estimar los efectos de los marcadores, la cual consta de un punto de masa fijado en cero para la mayoría de marcadores con varianza genética nula y una distribución marginal *t* escalada para los pocos *loci* con varianza genética, obteniendo estimaciones más precisas al considerarse todos los marcadores simultáneamente (Zeng *et al.,* 2012; Pérez-Rodriguez y de los Campos, 2014).

Algunos de los factores importantes que favorecieron la detección de QTL que explican buena proporción de la varianza genética fueron el tamaño de muestra, la heredabilidad y la calidad de los datos fenotípicos (Corvin *et al*., 2010). En esta investigación se evaluaron 669 mestizos derivados de variedades nativas con una alta diversidad genética, mientras que en otros estudios el tamaño fue de 95, 144, 161, 236

y 615 (Zhao *et al*., 2007; Wang *et al*., 2012; Liu *et al*., 2014; Shi *et al*., 2014; Gowda *et al*., 2015). Además, la heredabilidad alta y la calidad de los datos fenotípicos también favorecieron la detección de QTL. Con la inclusión de los tres primeros componentes principales como covariables y la matriz de relaciones de parentesco en el modelo lineal mixto del método Single Marker se obtuvo un buen control de las asociaciones espurias, reduciendo la tasa de falsos positivos o FDR (false discovery rate, en inglés), ya que los valores de p observados se ajustaron estrechamente a los valores esperados bajo la hipótesis de que la mayoría de marcadores tiene efecto igual a cero (Pearson y Manolio, 2008).

5.5 Proporción de varianza genética explicada por los QTL

Por otra parte, la proporción de varianza genética explicada por los marcadores de manera individual (P_G) con el método Single Marker fue menor del 10 %, por lo cual, los QTL asociados a tales marcadores pueden ser considerados de efectos menores según la clasificación propuesta por Collard *et al*. (2005). Esto además respalda la idea de que la resistencia al CMA es un carácter cuantitativamente heredado y condicionado por múltiples *loci*. La proporción de varianza genética total explicada por los QTL en conjunto fue de 24 y 49 % para Single Marker y BayesB, respectivamente, lo cual indica que BayesB realizó una mejor selección de marcadores. Esta diferencia en la precisión de detección de QTL está en función del modelo empleado por cada metodología. Los modelos bayesianos con capacidad para selección de variable, tal como BayesB, implementan una regresión múltiple con todos los marcadores disponibles simultáneamente, utilizando una distribución *a priori* mixta, permitiendo que los marcadores con efectos relevantes tomen valores diferentes de cero, mientras que los marcadores con efectos espurios (falsos positivos debido a covariables redundantes) sean contraídos hacia cero (Mutshinda y Sillanpää, 2010). Además, en varios estudio de simulación, los modelos bayesianos con capacidad para selección de variable han demostraron mayor precisión para detección de QTL y con menos falsos positivos que con el método Single Marker con enfoque de modelo mixto, especialmente para caracteres con alta heredabilidad y QTL con grandes efectos (Yi y Xu, 2008; Sahana *et al.*, 2010; Zeng *et al*., 2012; Berg *et al.*, 2013; Gondro *et al.*, 2013; Pérez-Rodríguez y de los Campos, 2014)

Por otra parte, el 51 % restante de la varianza genética total que no fue explicada por los QTL identificados con el método BayesB, sugiere que existen más QTL que contribuyen a la resistencia al CMA pero que no fueron detectados con el análisis GWAS debido posiblemente a las siguientes razones: 1) baja frecuencia de alelos específicos de algunas poblaciones (Weng *et al*., 2011), los cuales fueron eliminados durante el control de calidad (MAF < 5 %); 2) algunos alelos podrían estar en regiones próximas al centrómero, donde el número de polimorfismos naturalmente es muy bajo (Talbert y Henikoff, 2010) y por consiguiente no hay SNPs en desequilibrio de ligamiento que permitan su identificación (Daetwyler *et al*., 2014; Pérez-Rodríguez y de los Campos, 2014) y 3) hay muchos alelos con efectos demasiado pequeños, pero que pueden ser acumulados mediante el enfoque de selección genómica (Ornella *et al*., 2012; Daetwyler *et al*., 2014). Esta última parece ser la más probable, ya que parte de la proporción de varianza genética no explicada sí fue capturada en el análisis de selección genómica, lo cual se discute posteriormente (ver Figura 15 e inciso 4.7). Tampoco hay que descartar otras posibles causas como el rápido decaimiento del desequilibrio de ligamiento en maíz y que con la densidad de marcadores utilizada no fue suficiente para capturar todas las variaciones genéticas presentes en la diversas variedades nativas (Weng *et al*., 2011). En este sentido, las subsecuentes investigaciones deben enfocarse en la densidad de marcadores y su distribución en el genoma; también se recomienda formar poblaciones bi-parentales para mapeo de ligamiento a partir de las dos accesiones más resistentes, Guat153 y Oaxa280, a fin de detectar otros alelos específicos de estas poblaciones, y que debido a su baja frecuencia no fue posible detectarlos en el presente análisis GWAS.

5.6 Análisis de genes candidatos

En general todos los genes candidatos identificados mediante el análisis GWAS son codificadores de proteínas asociadas a procesos enzimáticos de transporte, factores de regulación de la transcripción, fosforilación por nucleósidos trifosfatos, receptores de señal involucrados en rutas metabólicas de mecanismos de resistencia como la familia de proteínas kinasas, y actividad catalítica en general, y están involucradas en mecanismos de defensa contra enfermedades (Després *et al.*, 2000; Romeis, 2001; Lee *et al*., 2008; Mohr *et al*., 2010; Michelmore *et al*., 2013). La mayoría de las

proteínas asociadas a los genes candidatos identificados en este estudio también han sido identificadas en otros estudios de enfermedades causada por hongos, tales como Carbón de la espiga (*Sphacelotheca reiliana*) (Wang *et al*., 2012), mancha gris de la hoja (*Cercospora zeae maydis*) (Shi *et al*., 2014) tizón sureño (*Bipolaris maydis*) y norteño (*Setosphaeria turcica*) del maíz (Kump *et al*., 2011; Poland *et al.*, 2011). Más estudios serán requeridos para investigar la función de estos genes candidatos para una mejor comprensión de cómo éstos actúan en la resistencia al CMA y enfermedades del maíz en general (Weng *et al*., 2011; Liu *et al*., 2014).

Si bien, en este estudio se identificaron tres genes candidatos de gran importancia (AC194670.2_FG001, GRMZM2G030272 y GRMZM2G059102) con el método BayesB, otros genes muy importantes fueron identificados con el método Single Marker, tales como GRMZM2G061602, GRMZM2G061603, GRMZM2G073884 y GRMZM2G151738, los cuales están involucrados en la recepción y señalización en los mecanismos de defensa de las plantas. En este sentido, es conveniente combinar ambos métodos de identificación de QTL para una mejor exploración la base genética.

5.7 Selección genómica

Los resultados de la evaluación de la precisión de los modelos indican que el nuevo enfoque propuesto en esta investigación, que consiste utilizar únicamente los marcadores identificados en el análisis GWAS, es más eficiente para predicción genómica que el enfoque original de utilizar una alta densidad de marcadores (Meuwissen *et al*., 2001), con la limitante de que no se capturaría el resto de variabilidad genética que no fue explicada por los marcadores asociados a la resistencia al CMA en el análisis GWAS. Dentro del nuevo enfoque, el método BayesB demostró ser más eficiente para detectar SNPs fuertemente ligados a QTL que explican mayor proporción de la varianza genética de la resistencia al CMA y a la vez tienen potencial para ser utilizados como variables predictoras en SG.

Estudios similares donde se utilicen únicamente los marcadores asociados a un carácter como variables predictoras no se encontraron, sin embargo, Zhang *et al*. (2015) evaluaron una densidad relativamente baja de marcadores (~200 SNPs) y

obtuvieron buenas predicciones para caracteres simples y con moderada a alta heredabilidad.

La ventaja de utilizar únicamente los marcadores identificados en el análisis GWAS es que no tiene el problema de la relación *p* >> *n*, por lo que se puede utilizar el método clásico de mínimos cuadrados ordinarios para estimar $\boldsymbol{\beta}$, el cual es el método más eficiente entre todos los estimadores lineales de acuerdo al teorema de Gauss-Markov (Rencher y Schaalje, 2008), lo cual hace que sea simple y práctico en términos de tiempo y menos demandante computacionalmente. Se espera que éste nuevo enfoque sea igualmente eficiente para otros caracteres con base genética similar a la resistencia al CMA.

Por otra parte, las correlaciones obtenidas con los modelos bayesianos están dentro del rango de correlaciones obtenidas en otros estudios de SG para enfermedades, tales como en la necrosis letal en maíz (0.36 y 0.56) (Gowda *et al*., 2015), roya del trigo (0.38, 0.27 y 0.44 para roya de la hoja, roya del tallo y roya amarilla respectivamente). Como referencia, en general las correlaciones obtenidas en trigo varían entre 0.3 a 0.8 (Daetwyler *et al*., 2014). Con este enfoque es posible acumular alelos favorables con efectos aditivos muy pequeños a lo largo de todo el genoma utilizando todos los marcadores disponible, lo cual permitiría seleccionar materiales con una amplia base genética de resistencia, lo cual representa es una fuente de resistencia genética diferente a la explicada por los QTL identificados en el análisis GWAS.

Otro aspecto importante a resaltar es el impacto de las relaciones de parentesco entre la población de entrenamiento y la de prueba. A pesar de que en este estudio se evaluó un conjunto diverso de poblaciones nativas no relacionadas en forma directa o genealógica, la relación existente a nivel genómico fue suficiente para obtener buena precisión de predicción con tan solo 17 SNPs significativos. En estudios similares donde se evaluaron materiales no relacionados directamente también obtuvieron buena precisión en las predicciones (Daetwyler *et al*., 2014; Gowda *et al*., 2015). Para analizar el impacto del tamaño de la población de entrenamiento, se evaluó la precisión

de predicción de los 17 SNPs identificados con BayesB utilizando únicamente el 10 % de la población para entrenar el modelo y predecir el 80 % restante. La correlación en la partición de prueba fue de 0.49. Esto indica que las predicciones siguen siendo considerablemente buenas aun utilizando un tamaño de población de entrenamiento bastante bajo (10 %). Esto demuestra que los 17 marcadores identificados con BayesB tienen buen potencial para predicción genómica de la resistencia al CMA en maíz.

5.8 Implicaciones en el mejoramiento genético de la resistencia al CMA

Los resultados indican que la resistencia al CMA es controlada por múltiples QTL con efectos aditivos relativamente pequeños. Desarrollar materiales con resistencia más durable es posible mediante la acumulación de estos QTL de pequeño efecto, ya que el rompimiento de un solo gen con pequeño efecto no haría al hospedero completamente susceptible al CMA. A diferencia de los QTL de grandes efectos que son relativamente fácil de identificar y mantener en poblaciones de mejoramiento mediante selección fenotípica, los QTL de pequeño efecto son fácilmente perdidos sin el uso de selección asistida por marcadores, ya que son enmascarados por los QTL de grandes efectos (Paliwal *et al*., 2001; Shi *et al*., 2014). Aquí es donde cobra importancia la selección asistida por marcadores (SAM), ya que partir de los marcadores significativamente asociados y fuertemente ligados a los genes o QTL de resistencia al CMA, es posible desarrollar marcadores funcionales, los cuales son marcadores de ADN obtenidos de motivos de secuencia funcionalmente caracterizados y se diseñan a partir de la secuencia de los genes candidatos identificados en las bases de datos del genoma del maíz (ej. MaizeGDB) (Andersen y Lübberstedt, 2003). Los marcadores identificados en este estudio mediante análisis GWAS, especialmente aquellos que se encuentran dentro de los genes candidatos y fueron consistentes en ambas metodologías, tales como S1_52921129, S1_17959629, S2_188057674, S3_124181320, S4_206663710, S7_175205315, S8_155905444 y S8_81247760, pueden ser una referencia importante para el desarrollo de marcadores funcionales, lo cual sería una herramienta valiosa para el mejoramiento de la resistencia al CMA. En cuanto a la relación beneficio costo de la SAM para resistencia a enfermedades, algunos estudios han mostrado que una vez

desarrollado los marcadores, la SAM es mucho más económica que el mejoramiento convencional (Yu *et al.*, 2000).

Desde la perspectiva de la selección genómica, los 17 SNPs identificados con el método BayesB muestran potencial para ser utilizados para predicción genómica para mejoramiento de la resistencia a CMA en maíz. Además, el enfoque de utilizar todos los marcadores disponibles puede ser utilizado para seleccionar parentales con una amplia base genética de resistencia (Heffner *et al.,* 2009).

Si bien, a la fecha la SG no es considerada una sustitución perfecta de la selección fenotípica, esta ha sido considerada un método para acelerar parte de los programas de mejoramiento (Nakaya y Isobe, 2012). En general, el uso de selección genómica como herramienta práctica en mejoramiento para resistencia dependerá de la ventaja relativa sobre la selección fenotípica en cuanto a ganancia genética (Gowda *et al*., 2015) y la relación de costos de fenotipado versus genotipado (Ornella *et al*., 2012). Desde el punto de vista teórico de la respuesta de la selección (Falconer y Mackay, 1996), la SG puede tomar ventaja en la reducción del número de ciclos por año y en la precisión del modelo. Con SG es posible realizar hasta tres ciclos de selección por año (Lorenzana y Bernardo, 2009), por lo que la SG podría ser más eficiente en términos de ganancia genética por año comparado con la selección fenotípica (Gowda *et al*., 2015). Además, considerando que la ganancia genética en SG es linealmente proporcional a la precisión de la predicción (Daetwyler *et al*., 2014), en la medida que se desarrollen modelos eficientes en la modelación de la epistasis se espera obtener mayor precisión en las predicciones (Hu *et al*., 2011; Jiang y Reif, 2015), haciendo que la ganancia genética por ciclo se incremente. En cuanto a costos, en el CIMMYT la evaluación fenotípica más sencilla con dos repeticiones por localidad cuesta alrededor de 30 a 40 $USD por genotipo (Ornella *et al*., 2012), sin embargo, si se requiere del uso de invernadero el costo incrementa significativamente. Mientras que el costo de genotipificación por el método GBS varía entre 30 y 45 $USD por genotipo con una duración de aproximadamente dos a tres semanas dependiendo de la demanda del servicio, y la tendencia es que los costos sigan bajando en la medida que se

incremente la capacidad de multiplexación de las plataformas de genotipificación (comunicación personal[6]).

Finalmente, se recomienda aprovechar los beneficios de ambos enfoques, SG y SAM. La SG serviría para realizar una preselección de los parentales con una amplia base genética de resistencia al CMA y la SAM podría utilizarse para realizar una segunda selección de los individuos que contengan algunos de los QTL de efectos significativos identificados en el análisis GWAS o para la introgresión de los QTL desde germoplasma nativo a materiales élite mediante el desarrollando de marcadores funcionales.

[6] Sarah Hearne, comunicación personal (2014). En reunión sobre mejoramiento molecular para resistencia al complejo mancha de asfalto en el proyecto MasAgro Biodiversidad, CIMMYT. México, D.F.

6 CONCLUSIONES

A partir de la evaluación fenotípica para resistencia al CMA se identificaron dos mestizos derivados de las accesiones Guat153 y Oaxa280, los cuales fueron los que mostraron mayor tolerancia, por lo que las accesiones *per se* pueden ser utilizadas como fuentes de resistencia al CMA.

Con el análisis GWAS se identificó un total de 17 SNPs asociados a la resistencia al CMA utilizando el método BayesB, los cuales están ligados a QTL que explican 49 % de la varianza genética.

Los SNPs S1_52921129, S1_17959629, S2_188057674, S3_124181320, S4_206663710, S7_175205315, S8_155905444 y S8_81247760 fueron encontrados dentro o muy cercanos a los genes candidatos GRMZM2G030272, GRMZM2G059102, GRMZM2G041127, GRMZM2G063688, AC194670.2_FG001, GRMZM2G440866, GRMZM2G061602 y GRMZM2G073884, respectivamente, los cuales están involucradas en mecanismos de defensa de los materiales resistentes.

La mayor precisión en predicción se obtuvo con el enfoque de utilizar únicamente los 17 SNPs identificados con el método BayesB en el análisis GWAS ($r_{prueba} = 0.61$ y $CME_{prueba} = 0.0525$) y además tiene la ventaja de utilizar el método clásico de mínimos cuadrados ordinarios para estimar $\boldsymbol{\beta}$, lo cual hace que sea simple y práctico en términos de tiempo y menos demandante computacionalmente. Con el enfoque de utilizar todos los marcadores disponibles es posible seleccionar materiales con una amplia base genética de resistencia al CMA.

Los resultados obtenidos en este estudio indican que la resistencia al CMA es de herencia poligénica, incluyendo 17 QTL de efectos relativamente pequeño que explican aproximadamente la mitad de la varianza genética; el resto es explicado por la acumulación de alelos favorables de efectos muy pequeños a lo largo del genoma y un 10% de la varianza genética total corresponde a efectos epistáticos.

7 BIBLIOGRAFÍA

Agrios, G. N. 2005. Plant Pathology. 5a ed. Amsterdam: Elsevier Academic Press. 952 p.

Ali, F., & Yan, J. 2012. Disease Resistance in Maize and the Role of Molecular Breeding in Defending Against Global Threat. Journal of Integrative Plant Biology, 54(3): 134–151.

Allard R., W. 1980. Principios de la mejora genética de las plantas. 4a ed. España: OMEGA.

Andersen, J. R., & Lübberstedt, T. 2003. Functional markers in plants. Trends in Plant Science, 8(11): 554–560.

Aranzana, M.J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P., Nordborg, M. 2005. Genome-Wide Association Mapping in *Arabidopsis* Identifies Previously Known Flowering Time and Pathogen Resistance Genes. PLoS Genet, 1(5): 531-539.

Arnold, G. R. W. 1986. List of the plant pathogenic fungi of Cuba., 207 p.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., … Nordborg, M. 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature, 465(7298): 627–631.

Bajet, N. B., Renfro, B. L., & Valdez C, J. M. 1994. Control of tar spot of maize and its effect on yield. International Journal of Pest Management, 40(2): 121–125.

Berg, I. van den, Fritz, S., & Boichard, D. 2013. QTL fine mapping with Bayes C(π): a simulation study. Genetics Selection Evolution, 45(1): 11 p.

Bernardo, R., & Yu, J. 2007. Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Science, 47(3): 1082.

Borém, A., & Fritsche-Neto, R. 2014. Biotechnology and Plant Breeding: Applications and Approaches for Developing Improved Cultivars. Elsevier. 270 p.

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.

Buntjer, J. B., Sørensen, A. P., & Peleman, J. D. 2005. Haplotype diversity: the link between statistical and biological association. Trends in Plant Science, 10(10): 466-471.

Burgueño, J., Cadena, A., Crossa, J., Banziger, M., Gilmour, A. R., & Cullis, B. 2000. User's guide for spatial analysis of field variety trials using ASREML. CIMMYT México, DF (México). 60 p.

Carena, M. J., Hallauer, A. R., & Miranda Filho, J. B. 2010. Quantitative Genetics in Maize Breeding. New York, NY: Springer New York. 663 p.

Castaño, J. J. 1989. Enfermedades del maíz cusadas por hongos. *in:* IX Seminario Manejo de Enfermedades y Plagas del Maíz. 28 de noviembre al 1 de diciembre. IICA-BID-PROCIANDINO, Quito, Ecuador. p. 159.

Ceballos, H., & Deutsch, J. A. 1992. Inheritance of resistance to tar spot complex in maize. Phytopathology, 82(5): 505–512.

CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo). 2013. Complejo mancha de asfalto del maíz: Hechos y acciones. (Folleto Técnico). México. 6 p.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica, 142(1-2): 169–196.

Corvin, A., Craddock, N., & Sullivan, P. F. 2010. Genome-wide association studies: a primer. Psychological Medicine, 40(07): 1063–1077.

Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., … Babu, R. 2013. Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 (Bethesda, Md.), 3(11): 1903–1926.

Crossa, J., de los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., … Braun, H.-J. 2010. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. Genetics, 186(2): 713–724.

Daetwyler, H. D., Bansal, U. K., Bariana, H. S., Hayden, M. J., & Hayes, B. J. 2014. Genomic prediction for rust resistance in diverse wheat landraces. TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, 127(8): 1795–1803.

de los Campos, G., & Gianola, D. 2010. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genetics research, 92(4): 295–308.

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. 2013. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics, 193(2): 327–345.

Després, C., DeLong, C., Glaze, S., Liu, E., & Fobert, P. R. 2000. The Arabidopsis NPR1/NIM1 protein enhances the DNA binding activity of a subgroup of the TGA family of bZIP transcription factors. The Plant Cell, 12(2): 279–290.

Dietterich, T. 1995. Overfitting and undercomputing in machine learning. Acm Computing Surveys, 27(3): 326–327.

Dittrich, U., Hock, J., Kranz, J., & Renfro, B. L. 1991. Germination of Phyllachora maydis ascospores and conidia of Monographella maydis. Cryptogamic Botany, 2(2-3): 214–218.

Doyle, J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull, 19: 11–15.

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): 10 p.

Endelman, J. B. 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome Journal, 4(3): 250 p.

Falconer, D. S., y Mackay, T. F. C. 1996. Introducción a la genética cuantitativa. Zaragoza España: Acribia. 365 p.

Fan, J.-B., Chee, M. S., & Gunderson, K. L. 2006. Highly parallel genomic assays. Nature Reviews Genetics, 7(8): 632–644.

Flint-García, S. A., Thornsberry, J. M., S, E., & IV, B. 2003. Structure of Linkage Disequilibrium in Plants. Annual Review of Plant Biology, 54(1): 357–374.

Ganal, M. W., Wieseke, R., Luerssen, H., Durstewitz, G., Graner, E.-M., Plieske, J., & Polley, A. 2014. High-throughput SNP Profiling of Genetic Resources in Crop Plants Using Genotyping Arrays en R. Tuberosa, A. Graner, y E. Frison (eds.), Genomics of Plant Genetic Resources (pp. 113–130). Springer Netherlands.

Garrick, D. J., Fernando, R. L. 2013. Implementing a QTL Detection Study (GWAS) Using Genomic Prediction Methodology. Gondro, C., Werf, J. van der, & Hayes, B. (eds). Springer. Humana Press. 1019: 275-298.

Gianola, D., Fernando, R. L., & Stella, A. 2006. Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. Genetics, 173(3): 1761–1776.

Gianola, D., & Kaam, J. B. C. H. M. 2008. Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. Genetics, 178(4): 2289–2303.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., & Thompson, R. 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, UK. 54 p.

González Camarillo, M., Jesús, M. E., Juan, P. H., y Noé, G. M. 2008. Híbrido de maíz elotero tolerante al complejo “mancha de asfalto” en el estado de Guerrero. LITOCASA S.A. Cuernavaca, Mor. México. 36 p.

González-Rojas, K., García-Salazar, J. A., Matus-Gardea, J. A., & Martínez-Saldaña, T. 2011. Vulnerabilidad del mercado nacional de maíz (*Zea mays* L.) ante cambios exógenos internacionales. Agrociencia. 46: 733-744.

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M. S., Bright, J. M., Beyene, Y., Prasanna, B. M. 2015. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theoretical and Applied Genetics. 128: 1957-1968.

Griffith, M., & Yaish, M. W. F. 2004. Antifreeze proteins in overwintering plants: a tale of two activities. Trends in Plant Science, 9(8): 399–405.

Gupta, P. K., Roy, J. K., & Prasad, M. 2001. Single nucleotide polymorphisms (SNPs): a new paradigm in molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Current Science, 80(4): 524–535.

Gupta, P. K., Rustgi, S., & Mir, R. R. 2008. Array-based high-throughput DNA markers for crop improvement. Heredity, 101(1): 5–18.

Hamblin, M. T., Buckler, E. S., & Jannink, J.-L. 2011. Population genetics of genomics-based crop improvement methods. Trends in Genetics: TIG, 27(3): 98–106.

Hearne, S., Chen, C., Buckler, E., & Mitchell, S. 2014. Unimputed GbS derived SNPs for maize landrace accessions represented in the SeeD-maize GWAS panel. International Maize and Wheat Improvement Center. http://hdl.handle.net/11529/10034. Consultado el 14 de mayo de2015.

Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. 2009. Genomic Selection for Crop Improvement. Crop Science, 49(1): 1-12.

Hernández, D. 1998. Enfermedades de maíz (*Zea mays* L.), trigo (*Triticum aestivum* L.) y cebada (*Hordeum vulgare* L.) presentes en México. Tesis de maestría en ciencias. Universidad Autónoma Chapingo, Chapingo, México. 126 p.

Hernández Ramos, L. 2014. Genética de la resistencia al complejo Phyllachora maydis Maubl., Monographella maydis Müller y Samuels y Coniothyrium phyllachorae Maubl., en diversos genotipos de maíz (*Zea mays* L.). Tesis de Maestría. Colegio de Postgraduados. Montecillo, Texcoco. México. 60 p.

Hock, J., Dittrich, U., Renfro, B. L., & Kranz, J. 1992. Sequential development of pathogens in the maize tarspot disease complex. Mycopathologia, 117(3): 157–161.

Hock, J., Kranz, J., y Renfro, B. L. 1989. El "complejo mancha de asfalto" de maíz, su distribución geográfica, requisitos ambientales e importancia económica en Mexico. Revista Mexicana de Fitopatologia, 7: 129–135.

Hock, J., Kranz, J., & Renfro, B. L. 1995. Studies on the epidemiology of the tar spot disease complex of maize in Mexico. Plant Pathology, 44(3): 490–502.

Hoerl, A. E., & Kennard, R. W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1): 55–67.

Hu, Z., Li, Y., Song, X., Han, Y., Cai, X., Xu, S., & Li, W. 2011. Genomic value prediction for quantitative traits under the epistatic model. BMC Genetics, 12(1): 15 p.

Jiang, Y., & Reif, J. C. 2015. Modelling Epistasis in Genomic Selection. Genetics. 201(2): 759-768.

Kao, C. H., Zeng, Z. B., & Teasdale, R. D. 1999. Multiple interval mapping for quantitative trait loci. Genetics, 152(3): 1203–1216.

Kitts, A., & Sherry, S. 2011. The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation. *In:* The NCBI Handbook, 2ª. ed. Jo McEntyre,

Jim Ostell (eds). National Center for Biotechnology Information (US), Bethesda (MD). http://www.ncbi.nlm.nih.gov/books/NBK21088/. Consultado el 6 julio.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., Hoh., J. 2005. Complement Factor H Polymorphism in Age-Related Macular Degeneration. Science (New York, N.Y.), 308(5720): 385–389.

Kump, K. L., Bradbury, P. J., Wisser, R. J., Buckler, E. S., Belcher, A. R., Oropeza-Rosas, M. A., … Holland, J. B. 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nature Genetics, 43(2): 163–168.

Lee, S., Woo, Y. M., Ryu, S. I., Shin, Y. D., Kim, W. T., Park, K. Y., Lee, I. J., An, G. 2008. Further Characterization of a Rice AGL12 Group MADS-Box Gene, OsMADS26. Plant Physiology, 147(1): 156–168.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., … Zhang, Z. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics, 28(18): 2397–2399.

Liu, C., Weng, J., Zhang, D., Zhang, X., Yang, X., Shi, L., … Zhang, S. 2014. Genome-wide association study of resistance to rough dwarf disease in maize. European Journal of Plant Pathology, 139(1): 205–216.

Lorenzana, R. E., & Bernardo, R. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, 120(1): 151–161.

Maciá-Vicente, J. G., Palma-Guerrero, J., Gómez-Vidal, S., & Lopez-Llorca, L. V. 2011. New Insights on the Mode of Action of Fungal Pathogens of Invertebrates for Improving Their Biocontrol Performance. *In:* Biological Control of Plant-Parasitic Nematodes. En K. Davies y. Spiegel (eds.). Springer Netherlands. 11: 203–225.

Malaguti, G., & Subero, L. J. 1972. Tar spot of maize. Agronomía Tropical 22(4): 443–445.

Manara, A. 2012. Plant Responses to Heavy Metal Toxicity. *In:* Plants and Heavy Metals. A. Furini (ed.). Springer, Netherlands. pp: 27–53.

Maublanc, A. 1904. Espéces nouvelles de Champignons inferius. Bull. Soc. Myc. Fr. 20: 72.

McClintock, B., Yamakake, T. A. K., & Blumenschein, A. (Mexico) C. de. 1981. Chromosome constitution of races of maize: its significance in the interpretation of relationships between races and varieties in the Americas. Colegio de Postgraduados. 517 p.

McGuire, J. U., & Crandall, B. S. 1967. Survey of insect pests and plant diseases of selected food crops of Mexico, central America and Panama. USDA Int. agric. Development Service, 157 p.

Melchinger, A. E., Utz, H. F., & Schön, C. C. 1998. Quantitative Trait Locus (QTL) Mapping Using Different Testers and Independent Population Samples in Maize Reveals Low Power of QTL Detection and Large Bias in Estimates of QTL Effects. Genetics, 149(1): 383–403.

Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4): 1819–1829.

Michelmore, R. W., Christopoulou, M., & Caldwell, K. S. 2013. Impacts of Resistance Gene Genetics, Function, and Evolution on a Durable Future. Annual Review of Phytopathology, 51(1): 291–319.

Mohr, T. J., Mammarella, N. D., Hoff, T., Woffenden, B. J., Jelesko, J. G., & McDowell, J. M. 2010. The Arabidopsis downy mildew resistance gene RPP8 is induced by pathogens and salicylic acid and is regulated by W box cis elements. Molecular Plant-Microbe Interactions, 23(10): 1303–1315.

Morota, G., & Gianola, D. 2014. Kernel-based whole-genome prediction of complex traits: a review. Front. Genet. 5: 363.

Mostert, L., Crous, P. W., & Petrini, O. 2000. Endophytic fungi associated with shoots and leaves of Vitis vinifera, with specific reference to the Phomopsis viticola complex. Sydowia, 52(1): 46–58.

Müller, E., & Samuels, G. J. 1984. Monographella maydis sp. nov. and its connection to the tar-spot disease of Zea mays. Nova Hedwigia, 40(1-4): 113–121.

Mutshinda, C. M., & Sillanpää, M. J. 2010. Extended Bayesian LASSO for Multiple Quantitative Trait Loci Mapping and Unobserved Phenotype Prediction. Genetics, 186(3): 1067–1075.

MycoBank. 2014. MycoBank: Fungal Databases Nomenclature and Species Banks. Disponible en http://www.mycobank.org. Consultado el 25 de agosto.

Nakaya, A., & Isobe, S. N. 2012. Will genomic selection be a practical method for plant breeding? Annals of Botany, London. 110: 1303–1316.

Niks, R. E., Ellis, P. R., & Parlevliet, J. E. 1993. Resistance to parasites. *In:* Plant Breeding. M. D. Hayward, N. O. Bosemark, I. Romagosa, y M. Cerezo (eds.). Springer Netherlands. pp: 422–447.

Ornella, L., Singh, S., Perez, P., Burgueño, J., Singh, R., Tapia, E., Bhavani, S., Dreisigacker, S., Braun, H. J., Mathews, K., Crossa, J. 2012. Genomic Prediction of Genetic Values for Resistance to Wheat Rusts. The Plant Genome Journal, 5(3): 136 p.

Paliwal, R. L., Granados, G., Lafitte, H. R., Violic, A. D., y Marathée, J.-P. 2001. El Maíz en los trópicos: mejoramiento y producción. Food y Agriculture Org. 392 p.

Pearson, T. A., & Manolio, T. A. 2008. How to interpret a genome-wide association study. The journal of the American Medical Association, 299(11): 1335–1344.

Pereyda-Hernández, J., Hernández-Morales, J., Sandoval-Islas, J. S., Aranda-Ocampo, S., León, C. de, y Gómez-Montiel, N. 2009. Etiología y manejo de la mancha de asfalto (Phyllachora maydis Maubl.) del maíz en Guerrero, México. Agrociencia, 43(5): 511–519.

Pérez-Rodríguez, P., & de los Campos, G. 2014. Genome-Wide Regression and Prediction with the BGLR Statistical Package. Genetics, 198(2): 483–495.

Poland, J. A., Bradbury, P. J., Buckler, E. S., & Nelson, R. J. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proceedings of the National Academy of Sciences, 108(17): 6893–6898.

Pritsch, C. 2001. El pre-mejoramiento y la utilización de los recursos fitogenéticos. En Estrategias en Recursos Fitogenéticos para los Países del Cono Sur. Montevideo, Uruguay. 10 p.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., Sham, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3): 559–575.

Rencher, A. C., Schaalje, G. B. 2008. Linear model in statistic. 2ª. ed. John Wiley & Sons, Inc., New Jersey. 688 p.

R-Core Team. 2015. R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing. https://www.r-project.org/. Consultado el 12 de marzo.

Riedelsheimer, C., Technow, F., & Melchinger, A. E. 2012. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. BMC Genomics, 13(1): 452.

Rodríguez, E., Burgueño, J., Mahuku, G., Shrestha, R., Guadarrama, A., Chepetla, D., y Willcox, M. 2013. Caracterización Fenotípica de Maíces Nativos para Resistencia a Mancha de Asfalto. In: Reunión Nacional para el Mejoramiento, Conservación y Uso de los Maíces Criollos. 25 al 27 Septiembre. Chiapas, México. 146 p.

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., … Gardner, C. A. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biology, 14(6): 55.

Romeis, T. 2001. Protein kinases in the plant defence response. Current Opinion in Plant Biology, 4(5): 407–414.

Rosas, J. F. M., y Verdejo, E. Á. 2009. Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus. Disponible en http://www.redalyc.org/articulo.oa?id=233117228001. Consultado el 25 de octubre.

Ruppert, D., Wand, M. P., & Carroll, R. J. 2003. Semiparametric Regression. Cambridge University Press. 416 p.

Sahana, G., Guldbrandtsen, B., Janss, L., & Lund, M. S. 2010. Comparison of association mapping methods in a complex pedigreed population. Genetic Epidemiology, 34(5): 455–462.

Sharma, S., Upadhyaya, H. D., Varshney, R. K., & Gowda, C. L. L. 2013. Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. Frontiers in Plant Science, 4: 309.

Shi, L., Lv, X., Weng, J., Zhu, H., Liu, C., Hao, Z., … Zhang, S. 2014. Genetic characterization and linkage disequilibrium mapping of resistance to gray leaf spot in maize (Zea mays L.). The Crop Journal, 2(2–3): 132–143.

Solberg, T. R., Sonesson, A. K., Woolliams, J. A., & Meuwissen, T. H. E. 2008. Genomic selection using different marker types and densities. Journal of Animal Science, 86(10): 2447–2454.

Sood, S., Flint-Garcia, S., Willcox, M. C., & Holland, J. B. 2014. Mining Natural Variation for Maize Improvement: Selection on Phenotypes and Genes. *In:* Genomics of Plant Genetic Resources. R. Tuberosa, A. Graner, y E. Frison (eds.). Springer Netherlands. pp: 615–649.

Species 2000 & ITIS Catalogue of Life. 2013. Recuperado el 26 de enero de 2015, a partir de http://eol.org/collections/54215. Consultado el 25 de julio.

Stram, D. O. 2014. Design, Analysis, and Interpretation of Genome-Wide Association Scans. Springer, New York. 15: 334 p.

Strange, R. N., & Scott, P. R. 2005. Plant disease: a threat to global food security. Annual Review of Phytopathology, 43: 83–116.

Sukumaran, S., & Yu, J. 2014. Association Mapping of Genetic Resources: Achievements and Future Perspectives. *In:* Genomics of Plant Genetic Resources. R. Tuberosa, A. Graner, y E. Frison (eds.) Springer Netherlands. pp: 207–235.

Syvänen, A. C. 2005. Toward genome-wide SNP genotyping. Nature Genetics, 37: 5–10.

Talbert, P. B., & Henikoff, S. 2010. Centromeres Convert but Don't Cross. PLoS Biol, 8(3): 5.

Tanksley, S. D. 1993. Mapping Polygenes. Annual Review of Genetics, 27(1): 205–233.

Technow, F., Bürger, A., & Melchinger, A. E. 2013. Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. G3 (Bethesda, Md.), 3(2): 197–203.

The Bulletin. 2015. Corn disease alert: New Fungal Leaf disease "Tar spot" Phyllachora maydis identified in 3 northern Illinois counties. Illinois, EUA. Disponible en http://bulletin.ipm.illinois.edu/?p=3423. Consultado el 23 de septiembre.

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T. R., McMullen, M. D., Holland, J. B., Buckler, E. S. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics, 43(2): 159–162.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science, 91(11): 4414–4423.

Varón, F., & Sarria, G. 2007. Enfermedades del maíz y su manejo. Colombia. 56 p.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics, Selection, Evolution: GSE, 34(3): 275–305.

Visendi, P., Batley, J., & Edwards, D. 2014. Next Generation Sequencing and Germplasm Resources. *In:* Genomics of Plant Genetic Resources. R. Tuberosa, A. Graner, y E. Frison (eds.). Springer Netherlands. pp: 369–390.

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y., & Zheng, Y. 2012. Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Science, 196: 125–131.

Waugh, R., Flavell, A. J., Russell, J., Thomas, W. (Bill), Ramsay, L., & Comadran, J. 2014. Exploiting Barley Genetic Resources for Genome Wide Association Scans (GWAS). *In:* Genomics of Plant Genetic Resources. R. Tuberosa, A. Graner, y E. Frison (eds.). Springer Netherlands. pp: 237–254.

Weng, J., Liu, X., Wang, Z., Wang, J., Zhang, L., Hao, Z., Xie, C., Li, M., Zhang, D., Bai, L., Liu, C., Zhang, S., Li, X. 2012. Molecular mapping of the major resistance quantitative trait locus qHS2.09 with simple sequence repeat and single nucleotide polymorphism markers in maize. Phytopathology, 102(7): 692–699.

Weng, J., Xie, C., Hao, Z., Wang, J., Liu, C., Li, M., … Li, X. 2011. Genome-Wide Association Study Identifies Candidate Genes That Affect Plant Height in Chinese Elite Maize (*Zea mays* L.) Inbred Lines. PLoS ONE, 6(12): 8 p.

White, D. G. 1999. Compendium of corn diseases, 3rd edition. 128 p.

Witte, J. S. 2010. Genome-Wide Association Studies and Beyond. Annual Review of Public Health, 31(1): 9–20.

Yi, N., & Xu, S. 2008. Bayesian LASSO for Quantitative Trait Loci Mapping. Genetics, 179(2): 1045–1055.

Yu, J., & Buckler, E. S. 2006. Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology, 17(2): 155–160.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., Buckler, E. S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38(2): 203–208.

Yu, K., Park, S. J., & Poysa, V. 2000. Marker-assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. Plant Breeding, 119(5): 411–415.

Zeng, J., Pszczola, M., Wolc, A., Strabel, T., Fernando, R. L., Garrick, D. J., & Dekkers, J. C. 2012. Genomic breeding value prediction and QTL mapping of QTLMAS2011 data using Bayesian and GBLUP methods. BMC Proceedings, 6: 5 p.

Zhang, X., Pérez-Rodríguez, P., Semagn, K., Beyene, Y., Babu, R., López-Cruz, M. A., San Vicente, F., Olsen, M., Buckler, E., Jannink, J.-L., Prasanna, B. M., Crossa, J. 2015. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity, 114(3): 291–299.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., … Buckler, E. S. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genetics, 42(4): 355–360.

Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D.-J., & Zhang, Q. 2010. Best Linear Unbiased Prediction of Genomic Breeding Values Using a Trait-Specific Marker-Derived Relationship Matrix. PLoS ONE, 5(9): 8 p.

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., Nordborg, M. 2007. An *Arabidopsis* Example of Association Mapping in Structured Samples. PLoS Genetics, 3(1): 71-82

Zhou, X., & Stephens, M. 2012. Genome-wide Efficient Mixed Model Analysis for Association Studies. Nature genetics, 44(7): 821–824.

Zila, C. T. 2014. Traditional and Genomic Methods for Improving Fusarium Ear Rot Resistance in Maize. Doctoral tesis. North Carolina State University. Disponible en http://www.lib.ncsu.edu/resolver/1840.16/9195. Consultado el 25 de octubre.

8 APÉNDICE

8.1 Listado de variedades nativas evaluadas en este estudio

1 OAXA280_5888
2 GUAT153_1861
3 GUER208_25081
4 GUAT823_5314
5 VERA115_18924
6 PUEB814_19987
7 VERAGP24_484
8 SNLP328_29454
9 CAMP3_2107
10 CHIS44_15874
11 SNLP323_29449
12 CHIS202_5906
13 HIDA291_29360
14 MICH166_142
15 HIDA293_29362
16 HAIT13_13797
17 OAXA257_5961
18 CHIS423_23203
19 CHIS474_25003
20 HAIT6_16141
21 SCRO4_1334
22 SNLP287_29414
23 SNLP110_5477
24 OAXA243_5949
25 VERA159_3242
26 VENE832_9965
27 VERA146_2554
28 CHIS675_26869
29 HIDA298_29367
30 VERA646_24643
31 COLI14_2134
32 BRAZBA084_15531
33 OAXA58_101
34 VENE809_9953
35 PUEBGP27_96
36 GUAT470_5271
37 VERA148_3233
38 VERA773_29529
39 VERA788_29543
40 CRIC131_3353
41 VENE820_11241
42 BRVI167_5437
43 PUEB177_24345
44 HIDA331_29400
45 VERA216_3275
46 CHIS621_25788
47 VERA749_29505
48 BRAZ1652_5482
49 VERA203_3268
50 SNLP319_29445
51 GREN11_2458
52 BRVI157_5433
53 VERA10_2323
54 CHIS159_2161
55 TRIN19_3993
56 VERA156_3239
57 TRINGP1_1237
58 HAIT1_2460
59 RDOMGP5_1262
60 BRAZ1844_4817
61 SNLP117_24795
62 VENE883_19817
63 TRIN47_4014
64 VERAGP19_480
65 RDOM270_1315
66 HIDA295_29364
67 BRAZ3923LC_27186
68 VERA195_3263
69 CHIS471_24307
70 BRAZRN001_15488
71 BRAZ1834_4810
72 VERA577_25093
73 SNLP37_2288
74 VERA178_3256
75 HIDA240_29309
76 HIDA324_29393
77 GUAT1178_27669
78 GUATGP23-2A_1190
79 HAIT45_3908
80 CUBAGP6_1250
81 CHIS788_18343
82 VERAGP6_476
83 SNLP351_29477
84 GUAT51_5158
85 JALI285_7023
86 CRIC83_3314
87 CHIS434_23208
88 BRAZ2829_2849
89 HIDA278_29347
90 BRAZMG081_21947
91 HAIT27_3899
92 CAMP5_1767
93 VENEM_325_17788
94 OAXAGP5_102
95 ECUA861_8311
96 VERA197_697
97 SNLP353_29479
98 OAXA6_241
99 VERA175_3253
100 VENE920_9997
101 ARZM06044_26665
102 CHIS4_2140
103 CHIS168_2164
104 SNLP128_2299
105 HIDA31_17835
106 BRVI123_5411
107 GUER110_16260
108 VERAGP20_481
109 CUBA64_5656
110 VENEM_42_16732
111 SNLP348_29474
112 GUAT61_5162
113 CHIS432_16284
114 CHIS112_10460
115 BRVI100_5659
116 SNLP337_29463
117 COAHGP7_454
118 YUCA64_835
119 CHIS233_507
120 BRAZ3047_2997
121 QROO84_25055
122 JAMA10_3910
123 TOBAGP2_1289
124 CHIS49_17806
125 BRAZPR053_22013
126 GUAT1083_27589
127 CUBA86_1108
128 BRAZ3000_2952
129 BRVI127_5414
130 CUBA115_2379
131 BRAZBA186_21880
132 BRAZPA071_15502
133 YUCA81_830
134 BRAZ2237_7802
135 ECUA947_8335

136 BRAZCE001_21886
137 RDOM218_13998
138 PUEB51_16495
139 GUAT295_5248
140 ARZM07041_26714
141 BRAZ1740_4762
142 HIDA313_29382
143 BRAZ2104_4958
144 PUEB777_23520
145 TOBA15_3991
146 SNLP56_2540
147 VERA240_20401
148 COMPCENTROAMERICANO3_1224
149 VENE730_11198
150 CRIC197_3405
151 BRAZ3026_2978
152 TAMA94_24979
153 VENE502_9828
154 ARZM05055_25539
155 CRIC78_3309
156 CHIS677_26873
157 HAITGP5_1256
158 VERA68_682
159 CUBA67_5657
160 TRIN42_1369
161 VERA642_14099
162 SNLP201_25299
163 BOZM1607_24808
164 COAH15_1754
165 SNLP299_29426
166 GUER214_24758
167 CAMP20_729
168 OAXA76_5939
169 GUAT97_1841
170 CRIC297_3485
171 BRAZ2062_4948
172 ARZM01083_19153
173 BRAZ429_4223
174 VERA572_25091
175 CAMP99_2118
176 SNLP371_29497
177 RDOMGP4_2482
178 ECUA365_8161
179 TAMA98_24957
180 VERA193_3261
181 SURI799_4326
182 GUAT829_13788
183 SAOPGP5_3091
184 BRAZ1188_9622
185 GREN13_1362
186 BRAZ1899_4854
187 VERA151_3234
188 GUAT1019_27527
189 SNLP334_29460
190 HIDA297_29366
191 BRAZ358_4198
192 BRAZ362_4200
193 BRAZ1056_9134
194 CAMP45_2110
195 FRGU784_4320
196 TRIN40_4008
197 BOLI437_10218
198 VERA91_2337
199 BRAZ945_4406
200 BRAZ1644_4736
201 CRIC55_3294
202 SNLP116_23627
203 VERAGP8A_478
204 VENE785_9932
205 BRAZ1731_4755
206 VENE861_9976
207 YUCAGP16_851
208 VERA133_459
209 BRAZ3985LC_10774
210 PANA93_3801
211 COAH59_2130
212 GUATGP18-2A_1181
213 CUBAT_25_15680
214 FRGU792_5871
215 PAZM2035_22746
216 BRAZ1826_6661
217 PARA148_4143
218 HAIT10_16143
219 VENE859_11263
220 RIGSGP7_3097
221 HAIT44_3907
222 YUCA125_2354
223 BRAZ30_2724
224 VENE777_9924
225 QROO79_24980
226 BRAZ207_2685
227 OAXA506_23419
228 FRGU794_5832
229 VERA786_29541
230 BOZM1702_24813
231 VERA184_3258
232 BRAZMS020_21966
233 CHIS66_504
234 VENE750_11213
235 BRAZ9_2714
236 ARZM06043_24438
237 CUBA36_5652
238 QROO80_25063
239 GUAT1015_27523
240 GUAT897_2526
241 RIGSGP12_3088
242 HAIT24_3896
243 BRAZ1305_4615
244 ECUA951_8338
245 QROO40_20378
246 TAMA55_18917
247 GUAT1161_27652
248 GUAT793_1085
249 QROO59_24385
250 GUAT597_1866
251 ARZM05003_15579
252 PUEB745_20817
253 SCRO11_5561
254 BRAZ2293_5057
255 VENE800_9944
256 ECUA336_8149
257 CRIC79_3310
258 CRIC128_3351
259 YUCA29_20434
260 CHIS160_2162
261 VENE1011_14432
262 GUAT1010_27518
263 GUAD13_3889
264 ARZM06050_25561
265 COMPUECENT6_1061
266 GUAT1014_27522
267 GUAT28_5156
268 VERA234_18926
269 PAZM4033_21448
270 CUBAT_9_15668
271 BRAZ1766_4781
272 QROO49_25087
273 QROO27_18899
274 BRAZ1951_6685
275 VERA164_3245
276 HAIT19_3894
277 DURA131_16257
278 SALV82_3628
279 VENE1014_14435
280 CRIC91_3321
281 BRVI114_5405
282 SURI797_4324
283 VERA648_24644
284 ECUA939_8333
285 HAIT16_1306

286 PARA108_2675
287 BRAZMG040_15485
288 SNLP34_2539
289 MINGGP1_3085
290 CHIS404_23213
291 ECUA629_7915
292 RDOM7_16218
293 YUCA164_2362
294 GUAT1011_27519
295 SURI803_4329
296 SALV90_3634
297 BRAZBA096_21865
298 BRAZ1259_9693
299 BRAZ51AR_10859
300 NAYA168_7044
301 BRAZ1522_5802
302 GUYA813_4339
303 ARZM07020_26701
304 BRAZ827_4347
305 GUAT1006_27514
306 VENE587_10133
307 ARZM06015_25555
308 CRIC81_3312
309 BRAZ1720_4749
310 SAOPGP7_3060
311 CAMP2_2106
312 VERA202_3267
313 JALI283_7021
314 URUG602_4258
315 YUCAGP12_848
316 PERU1336_9033
317 VENE581_9845
318 PUER2_1331
319 BRVI115_5406
320 BOLI1063_6268
321 OAXA908_23604
322 BRAZ2150_4992
323 COMPUESTOHONDURENO1_4023
324 GUYA810_4336
325 VENE844_9971
326 CUBA82_2450
327 BRAZ2148_4990
328 HAIT40_3905
329 BRAZ3056_3005
330 SCROGP1_1231
331 VENE892_9985
332 TOLI401_23029
333 JALI187_601
334 BRAZ1260_4588
335 BRAZ3048_2998
336 BRAZ1943_6004
337 GUAD11_1056
338 PAZM14026_19083
339 BRAZ2460_2773
340 CRIC393_3563
341 BRAZ1789_5879
342 BRAZ2221_7792
343 BRAZ2903_2862
344 BRAZ64_2644
345 PAZM8076_21584
346 TAMA11_17930
347 GUAT303_2500
348 CHIS108_761
349 ARZM07042_26715
350 TAMA135_24023
351 GUAT1086_27592
352 BRAZ1681_6128
353 GUAT1087_27593
354 TRIN21_1364
355 PANA66_3783
356 RDOM112_13994
357 BRAZ1439_4651
358 PUEB779_23409
359 BRAZ1721_4750
360 BRAZ3929LC_10743
361 BRAZ1486_4669
362 VENE823_11242
363 YUCA98_5865
364 SAOPGP2_3099
365 RDOM147_13997
366 VENE680_9902
367 BRAZ4041LC_10796
368 BRAZ1446_4657
369 CUBA18_5372
370 TAMA12_17931
371 QROO61_25084
372 BOZM1168_14379
373 BOZM1634_24835
374 BRAZPA099_19107
375 OAXA556_25312
376 GUAT1009_27517
377 PANA139_3826
378 GUAT262_1214
379 RDOMGP8_1235
380 GUYA817_4341
381 QROO20_10507
382 CUBA7_5363
383 BRAZRN005_15441
384 ARZM07107_19180
385 BRAZ1714_6138
386 GUYA816_4340
387 BRAZ2283_7813
388 CUBA9_5365
389 BOZM1633_24821
390 BRAZ1845_4818
391 TRIN50_4017
392 ARZM01031_15552
393 CHIS631_10469
394 GUERGP6_28
395 ARZM12241_19202
396 HAIT38_1312
397 MICH408_21369
398 CRIC85_3316
399 MICHGP10_11
400 COAH61_5456
401 NAYA175_7989
402 GUAT1167_27658
403 VENE386_11066
404 BRAZBA177_15477
405 TRIN315_17098
406 VENE705_11180
407 GUER195_24284
408 RDOMGP14_1270
409 ARZM04097_24422
410 BRAZ1190_4551
411 JALIGP42_613
412 TRIN35_4003
413 BRAZ1023_4456
414 BRVI112_5403
415 PANA153_3837
416 ARZM06013_25554
417 CUBA137_2399
418 CUBA11_2438
419 SVIN1_1341
420 GUAT765_1063
421 BRAZ1442_4654
422 PUER3_3911
423 GUAD16_3891
424 GUAT1175_27666
425 GUAT1165_27656
426 VENE732_11200
427 VENE682_9904
428 GUAT1180_27671
429 PANA72_3786
430 BRAZ1287_4604
431 BOZM710_14999
432 BRVI103_5661
433 ARZM07119_25581
434 ARZM07039_26712
435 QROO28_18900

436 URUG298_10615
437 BRAZ2501_5467
438 BRAZ1353_5519
439 HONDGP18_1006
440 VENE839_11250
441 VENE363_10049
442 VENE337_9797
443 BRAZ1251_4583
444 BRAZ1042_4470
445 ARZM13106_19203
446 NAYA142_25808
447 BOZM260_14289
448 BOZM1427_17763
449 RDOM301_3967
450 BRAZ1955_6687
451 GUAT704_1087
452 BRAZ4039LC_27189
453 GUAT286_5244
454 BRAZ1954_4893
455 BRAZ2500_5147
456 BRAZ21_2721
457 BRAZ39_2642
458 CRIC336_3515
459 PANAGP99_1045
460 PANA120_3814
461 GUATGP7-1A_2527
462 URUG601_4257
463 VENE971_11337
464 BRAZ1283_4600
465 OAXAGP9_106
466 ANTI5_3859
467 ARZM14071_25667
468 SNLP340_29466
469 VENE609_9865
470 CUBA165_2427
471 CUBA155_2417
472 BRAZ1557_4711
473 CUBA80_2448
474 BRVI120_5409
475 GUADGP2_1292
476 PANA79_3791
477 VENE648_11160
478 GUAT314_1074
479 VENEM_212_22081
480 ARZM07123_25582
481 ARZM07118_26749
482 SALV49_3607
483 BRAZ312_16951
484 ARZM07049_15585
485 BRAZ1772_5945
486 SONO74_7254
487 ARZM01094_24396
488 GREN313_17025
489 CUBA119_2382
490 NAYA275_7101
491 CUBA114_2378
492 SNLP146_15974
493 GUER176_262
494 BRAZ1193_7492
495 RDOM114_13362
496 GUER376_13796
497 BRAZPE024_15411
498 CUBAI_72_15717
499 CAMP43_5449
500 BRAZ2773_2796
501 PANA78_3790
502 SNLP372_29498
503 OAXA242_5948
504 CUBA158_2420
505 BRAZ2144_4987
506 SALV60_3613
507 ECUA926_8327
508 CUBA71_1132
509 RDOM276_3951
510 CRIC249_962
511 BRAZ2140_4983
512 ARZM16045_19217
513 BRAZ1643_5463
514 CRISTALINOAMARILLOMEZCL_5696
515 CUBA77_2447
516 BRAZ3978LC_10767
517 YUCA69_5863
518 ARZM03056_14599
519 ARZM08178_19189
520 PAZM8078_21586
521 BRAZ1729_4753
522 SLUCGP4_9049
523 ARZM13050_25621
524 PARA119_4123
525 MICH190_18872
526 MICH335_8071
527 PARA100_2625
528 VENE961_11327
529 RDOM289_3959
530 SCROGP3_1280
531 HAIT304_15755
532 NAYA139_25814
533 GUAT123_1071
534 VENE444_9817
535 ARZM07094_24462
536 ARZM07102_26738
537 GUAT58_5160
538 BRAZ51_9194
539 CHIS223_763
540 BRAZ4_2709
541 BRAZ1645_9141
542 RDOM250_3935
543 PARA107_4121
544 PUER22_1126
545 VENE668_9895
546 VENE575_9841
547 CUBAI_42_15692
548 BOZM1631_24823
549 BRAZ2314_5061
550 ARZM07124_26750
551 VENE534_10056
552 ECUA858_8309
553 SALV102_3642
554 RDOM243_3929
555 ARZM17015_25704
556 VERA204_3269
557 ATLA328_3194
558 BRAZ2084_5541
559 HONDGP5_1001
560 JALI280_7019
561 GUER38_17832
562 TRIN24_3996
563 CUBA95_3876
564 BRAZ1462_4660
565 GUAT1020_27528
566 BRAZ1303_4613
567 JALI443_25226
568 NICA33_2616
569 GUAT209_1099
570 BRAZ133AR_10901
571 MORE71_18538
572 OAXA889_23585
573 RDOMGP10_1266
574 URUG745_4311
575 BRAZMG107_15482
576 PAZM14106_21714
577 VENE391_11071
578 OAXA245_5950
579 PANA121_996
580 SNLPGP15_444
581 CUBA79_1303
582 COLI2_2528
583 ARZM14068_25664
584 OAXA158_261
585 JALI276_7015

586 CRIC221_3422
587 BRAZBA057_18315
588 ARZM07061_25577
589 GUAT1079_27585
590 BRAZ35_2674
591 ARZM01053_15558
592 BRAZ1827_4805
593 GUAT1166_27657
594 BRAZ2132_4976
595 BRAZ1712_6137
596 JALI442_25225
597 ARZM06088_24444
598 CUBA6_5362
599 VENE824_11243
600 GUER55_36
601 BRAZ1689_5907
602 PAZM10087_19064
603 NAYA245_24962
604 CUBA31_5380
605 VERA220_3278
606 CRIC285_3475
607 CRIC341_3520
608 ARZM07125_25583
609 BRAZ756_2735
610 CUBA27_5378
611 CRIC303_3489
612 TUXPENOCREMA1_5700
613 ARZM17014_24529
614 ARZM07143_26757
615 BRAZ1028_4460
616 RDOM190_13345
617 OAXA130_245
618 SALV66_3617
619 SALV62_3615
620 VENE564_9836
621 CUBA17_5371
622 SINA101_7207
623 OAXA856_23552
624 ARZM07098_24464
625 SINA95_7201
626 SNLP35_17898
627 GUAT253_1065
628 VENE650_9887
629 BRVI145_5425
630 CUBA142_2404
631 OAXA541_25217
632 VENE1000_14421
633 OAXA907_23603
634 URUG601_6531
635 CHIS526_24953
636 ARZM18025_25728
637 SALV50_3608
638 HAIT25_3897
639 GUATGP6-1A_1158
640 VENE751_11214
641 VENE683_9905
642 BRAZ1212_4563
643 ARZM07043_24453
644 URUG729_4304
645 BRAZ3028_2980
646 GUER181_205
647 BRAZ1273_7521
648 BRVI144_5424
649 CHIS440_10464
650 JALI398_25228
651 HOND66_880
652 GUER48_5719
653 ANTI8_1130
654 OAXA239_5845
655 YUCA62_833
656 GUER216_24974
657 SALV72_3621
658 ARZM07073_24456
659 GUAT92_1096
660 BRAZSC009_22019
661 RDOM249_3934
662 JALI183_577
663 SINA176_25021
664 BRAZ90AR_10876
665 GUAT83_8115
666 GUYA812_4338
667 VENE541_9831
668 GUAT1094_27599
669 ARZM10072_25590