



**Universidad de San Carlos de Guatemala
Dirección General de Investigación
Programa Universitario de Investigación en Alimentación y Nutrición**

INFORME FINAL

Identificación de marcadores moleculares (SNP) asociados a la tolerancia genética al complejo mancha de asfalto en variedades nativas de maíz (*Zea mays* L.)

Equipo de investigación:

Carlos Alfonso Orozco Castillo PhD

Ing. Agr. MSc. David Omar González Diéguez

Ing. Agr. Carlos Ovidio Salazar Quisquinay

Br. Astrid Catherine Mahoví Mérida de la Roca

Noviembre de 2017

**UNIDAD DE INVESTIGACIÓN AVALADORA
FACULTAD DE AGRONOMÍA**

M.Sc. Gerardo Arroyo Catalán
Director General de Investigación

Ing. Agr. MARN Julio Rufino Salazar
Coordinador General de Programas

Inga. Liuba María Cabrera

Dr. Carlos Alfonso Orozco Castillo

Ing. Agr. MSc. David Omar González Diéguez

Ing. Agr. Carlos Ovidio Salazar Quisquinay

Br. Astrid Catherine Mahoví Mérida de la Roca

Partida Presupuestaria
4.8.63.7.02
Año de ejecución: 2017

Índice

1	Introducción	7
2	Marco teórico y estado del arte	10
2.1	Complejo mancha de asfalto en maíz	10
2.2	Distribución geográfica.....	10
2.3	Epidemiología del complejo mancha de asfalto en maíz.....	10
2.4	Base genética y molecular del maíz.....	11
2.5	Importancia de los recursos genéticos en la búsqueda de resistencia	11
2.6	Resistencia genética del maíz al complejo mancha de asfalto.....	11
2.7	Genes de resistencia a enfermedades en plantas.....	12
2.8	Marcadores moleculares	13
2.9	Estimación de la heredabilidad de la tolerancia al CMA.....	14
2.10	Identificación de QTL y Estudio de Asociación de Genoma Completo	15
3	Materiales y métodos	17
3.1	Materiales.....	17
3.1.1	Base de datos fenotípicos	17
3.1.2	Base de datos genotípicos.....	18
3.1.3	Material vegetal para la evaluación preliminar en campo del nivel de tolerancia genética al CMA.....	19
3.2	Métodos.....	20
3.2.1	Estimación de la heredabilidad de la tolerancia al CMA	20
3.2.2	Identificación de marcadores asociados a la tolerancia genética al CMA	21
3.2.3	Proporción de variación fenotípica total explicada por los marcadores	21
3.2.4	Identificación de genes candidatos	22
3.2.5	Evaluación preliminar en campo del nivel de tolerancia genética al CMA	22
3.3	Operacionalización de las variables o unidades de análisis.....	24
3.4	Procesamiento de datos y plan de análisis	24
3.4.1	Manejo de las bases de datos y control de calidad	24
3.4.2	Análisis de datos fenotípicos	25
3.4.3	Distribución general de los marcadores en el genoma del maíz.....	25
3.4.4	Estimación de la heredabilidad de la resistencia al CMA	26

3.4.5	Estudio de asociación del genoma completo	26
3.4.6	Proporción de variación fenotípica total explicada por marcadores.....	28
3.4.7	Análisis de los datos de la evaluación de campo.....	28
3.5	Coherencia de la propuesta de investigación.	28
4	Resultados	29
4.9	Análisis descriptivo de la base de datos fenotípicos	29
4.10	Distribución y estadísticas generales de la información genotípica.....	31
4.11	Estimación de la heredabilidad de la resistencia al CMA basada en SNPs	34
4.12	Identificación de marcadores SNP asociados a tolerancia a mancha de asfalto	34
4.13	Identificación de genes candidatos.....	38
4.14	Evaluación preliminar en campo del nivel de tolerancia genética al CMA	44
5	Discusión.....	52
5.9	Evaluación fenotípica, heredabilidad y donadores	52
5.10	Identificación de marcadores SNP asociados a tolerancia a mancha de asfalto	52
5.11	Identificación de genes candidatos.....	53
5.12	Evaluación preliminar en campo del nivel de tolerancia genética al Complejo Mancha de Asfalto.....	56
6	Impacto esperado	57
7	Conclusiones	57
8	Vinculación	58
9	Agradecimientos	58
10	Referencias.....	59
11	Apéndice	63

Identificación de marcadores moleculares (SNP) y genes candidatos asociados a la tolerancia genética al complejo mancha de asfalto en variedades nativas de maíz

David O. González Diéguez^{1*}, Carlos A. Orozco Castillo¹

¹ Facultad de Agronomía, Universidad de San Carlos de Guatemala.

*correo electrónico: dogrdc@gmail.com

Resumen

El complejo mancha de asfalto (CMA) en maíz (*Zea mays* L.), causado por los hongos *Phyllachora maydis* y *Monographella maydis*, es una enfermedad de importancia económica en Guatemala, causando pérdida en el rendimiento entre 30 a 50 %, inclusive del 100 % si las condiciones son favorables. El objetivo de esta investigación fue identificar marcadores SNP y genes candidatos asociados a la tolerancia genética al CMA. Un total de 463 poblaciones nativas y 329 mil SNPs fueron analizadas con dos modelos, single marker y BayesB, para la identificación de regiones asociadas a la tolerancia genética al CMA. Un total de 40 marcadores SNPs fueron asociados significativamente a la tolerancia genética a mancha de asfalto con ambos modelos. La proporción de variación fenotípica total explicada (PVE) por los 40 SNPs fue de 56%, atribuida a efectos genéticos aditivos. Múltiples genes de resistencia fueron identificados en las regiones señaladas por los marcadores SNP, siendo sus principales funciones receptores y transductores de señal, factores de transcripción que regulan positivamente la expresión de genes de resistencia y genes de la familia kinasa, por lo que potencialmente están involucrados en el mecanismo de defensa a mancha de asfalto. Los resultados preliminares obtenidos en campo mostraron genotipos tolerantes al complejo CMA. Los resultados de esta investigación pueden servir de base para subsecuentes investigaciones orientadas principalmente al mejoramiento genético del maíz asistido por marcadores moleculares.

+

Palabras Clave: *Zea mays* L.; estudio de asociación del genoma completo; genotipificación por secuenciación; resistencia genética; *Loci* de un carácter cuantitativo (QTL).

Identification of molecular markers (SNP) and candidate genes associated to genetic tolerance to tar spot complex disease in maize landraces

David O. González Diéguez^{1*}, Carlos A. Orozco Castillo¹

¹ Facultad de Agronomía, Universidad de San Carlos de Guatemala.

*e-mail: dogrdc@gmail.com

Abstract

The tar spot complex (TSC) disease in maize (*Zea mays* L.), caused by the fungi *Phyllachora maydis* and *Monographella maydis*, is a economic important disease in Guatemala, producing loss in yield between 30 to 50%, inclusive of 100 % if the conditions are favorable. The objective of this research was to identify SNP markers and candidate genes associated with genetic tolerance to TSC. A set of 463 native populations and 329 thousand SNPs were analyzed with two models, single marker and BayesB, for the identification of regions associated with genetic tolerance to the TSC. A total of 40 SNP markers were significantly associated with the genetic tolerance to TSC with both models. The proportion of total phenotypic variation explained (PVE) by the 40 SNPs was 56%, attributed to additive genetic effects. Multiple resistance candidate genes were identified in the regions indicated by the SNP markers, their main functions being signal transducers and receptors, transcription factors that positively regulate the expression of resistance genes and family kinase genes, therefore, they are potentially involved in the defense mechanism to TSC.

The preliminary results obtained in the field showed tolerant genotypes to the TSC. The results of this research can be used as a basis for subsequent studies oriented mainly to genetic improvement of maize assisted by molecular markers.

Keywords: *Zea mays* L.; genome-wide association study; genotyping by sequencing; genetic resistance; quantitative trait loci (QTL).

1 Introducción

El maíz forma parte del grupo de los granos básicos que constituyen la base de la dieta de la población guatemalteca, que para el 2015 se estimaba en 15.9 millones de personas, y para el 2021 se proyectan más de 17.7 millones (Instituto Nacional de Estadística, 2015). Sin embargo, el área de producción de maíz se mantiene con una tendencia ligeramente a la baja, por la presión de urbanización, incremento de cultivos extensivos (caña de azúcar, banano y palma africana) y la interacción entre los factores abióticos como la sequía y factores bióticos como las enfermedades, entre ellas, el complejo mancha de asfalto (CMA), inducida por los hongos *Phyllachora maydis* Maubl., *Monographella maydis* Müller y Samuels y *Coniothyrium phyllachorae* Maubl., la cual ha cobrado gran importancia en la última década en Latino América 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. 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, Kranz, & Renfro, 1989). En Guatemala el primer reporte del CMA data del año 1974, sin causar pérdida económica al cultivo (Monterroso, Gallardo, & Zuñiga, 1974). Fue hasta el año 2007 cuando se reportó la presencia de mancha de asfalto en el municipio de Ixcán, Quiché, con pérdidas de producción de grano de alrededor del 80%; en años subsiguientes aumentó la severidad de la enfermedad (ICTA/Red SICTA/IICA/Cooperación Suiza, 2011).

El desarrollo de materiales genéticamente resistentes a enfermedades es considerado la mejor estrategia para el control de enfermedades (Crute & Pink, 1996; White, 1999). En esta línea de acción, varias empresas productoras y comercializadoras de semilla de maíz en Guatemala han promovido algunos materiales tolerantes a mancha de asfalto, sin embargo, a la fecha no existe ningún estudio serio que demuestre su nivel de resistencia comprobada y estable en las regiones más afectadas por esta enfermedad, por lo que sigue siendo un problema para los productores de maíz. Como antecedente, el Instituto Nacional de investigaciones Forestales y Agropecuarias (INIFAP) liberó en 2005 el híbrido H-563 basado en resistencia vertical, 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 alcanzar una resistencia más durable consiste en ampliar la base genética de los cultivos (Ali & Yan, 2012; Strange & Scott, 2005), 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, lo que implica que diversos genes están involucrados y pueden ser acumulados para la obtención de germoplasma con una resistencia más durable. Con el avance de la biotecnología y el desarrollo de plataformas de genotipificación de alto rendimiento es posible rastrear regiones del genoma de plantas asociadas a determinados fenotipos mediante mapeo de asociación del genoma completo (GWAS, por sus siglas en inglés Genome-wide Association Study). El primer análisis GWAS fue realizado en humanos con gran éxito (Klein *et al.*, 2005), posteriormente se extendió su uso a animales y plantas. Un gran número de análisis GWAS han sido realizado en los principales cultivos de importancia mundial asociados a características agronómicas de importancia, así como resistencia a enfermedades y tolerancia a estrés biótico y abiótico, en los cuales se han logrado identificar regiones del genoma y genes inclusive con mayor éxito que en humanos (Brachi, Morris, & Borevitz, 2011).

Recientemente el Centro Internacional de Mejoramiento de Maíz y Trigo (Cymmit) han realizado un trabajo de genotipificación y fenotipificación de un conjunto de poblaciones nativas de maíz de la región de Latinoamérica, las cuales constituyen una gran diversidad y fuente de variabilidad genética natural para la identificación de regiones genómicas asociadas a características de tolerancia a estrés biótico y abiótico, y caracteres de importancia agronómica. Es por ello que el objetivo general de este proyecto fue identificar fuentes de tolerancia genética al complejo mancha de asfalto en poblaciones nativas de maíz (*Zea mays* L.), utilizando bases de datos genotípicas y fenotípicas generadas y publicadas por el Cimmyt. Para ello se plantearon los siguientes objetivos específicos: a) identificar marcadores moleculares SNPs asociados a la tolerancia genética al complejo mancha de asfalto en poblaciones nativas de maíz mediante un estudios de asociación de genoma completo b)

identificar genes candidatos potencialmente involucrados en mecanismos de defensa al complejo mancha de asfalto, y c) evaluar en campo de manera preliminar el nivel de tolerancia genética al CMA de poblaciones nativas, líneas endogámicas e híbridos experimentales y comerciales, en condiciones del municipio de Ixcán, Quiché, durante el año 2017. Los marcadores identificados y genes candidatos servirán de base para realizar mapeo fino y generar marcadores funcionales en subsecuentes investigaciones; así también, los materiales que presentan mayor tolerancia servirán como potenciales donadores de resistencia para ser utilizados en investigaciones posteriores.

2 Marco teórico y estado del arte

2.1 Complejo mancha de asfalto en maíz

Los agentes causales de la enfermedad son tres hongos: *Phyllachora maydis* Maubl., *Monographella maydis* Müller y Samuels y *Coniothyrium phyllachorae* Maubl. (Hock *et al.*, 1989). Los síntomas comienzan con la infección de *P. maydis*. Luego, la lesión provocada por *P. maydis* da lugar a la invasión de *M. maydis*; la asociación de estos dos hongos resulta en el desarrollo de tejido necrótico, produciendo el síntoma denominado comúnmente como “ojo de pescado” (Cimmyt, 2013; Varón & Sarria, 2007), produciendo muerte excesiva del tejido foliar (Hock, Diettrich, Renfro & Kranz., 1992; Pereyda-Hernández *et al.*, 2009).

2.2 Distribución geográfica

La enfermedad ha sido reportada en México, Centroamérica, el Caribe y Sudamérica (Hock *et al.*, 1989; Cimmyt, 2013). En Centroamérica 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 (Cimmyt, 2013). En Guatemala las regiones más afectadas por la enfermedad son el municipio de Ixcán en Quiché, el departamento del Petén y la región del Valle del Polochic en Alta Verapaz. También se reportó una expansión de la enfermedad en Jalapa e Izabal, donde nunca se había reportado antes (Larios, 2012). Recientemente, en septiembre de 2015 el Centro Nacional de Patología de Plantas y el Departamento de Agricultura de Estados Unidos (USDA) reportaron la presencia de *Phyllachora maydis* en los condados de DeKalb, LaSalle y Bureau del Estado de Illinois y el Estado de Indiana, EUA, sin embargo, no se registraron pérdidas significativas, posiblemente debido a la ausencia de *Monographella maydis*.

2.3 Epidemiología del complejo mancha de asfalto en maíz

Se desarrolla en las zonas tropicales, subtropicales y zonas de transición, con un nivel de altitud entre 1,300 a 2,300 msnm, donde los ambientes son moderadamente fríos, pero bastante húmedos, especialmente en riberas con suelos con nivel freático alto y con tendencia al anegamiento (Malaguti & Subero, 1972; Varón & Sarria, 2007). La enfermedad se desarrolla en un rango de temperaturas promedio mensuales entre 17 y 22 °C, con una humedad relativa superior al 75 % (Cimmyt, 2013).

2.4 Base genética y molecular del maíz

El maíz es una especie diploide con un total de 20 cromosomas (Paliwal, Granados, Lafitte, Violic & Marathée, 2001). Se han identificado alrededor de 32,000 genes y un 85 % del genoma comprende elementos repetitivos y transponibles (McClintock, Yamakake & Blumenschein, 1981). Obtener mapas de alta resolución en maíz es posible gracias a la gran diversidad genética, pero para ello se requiere aplicar un conjunto denso de marcadores y análisis sistemático. El maíz es considerado una planta ideal para estudios a nivel molecular por su rápida reducción del Desequilibrio de ligamiento, especialmente para estudios de asociación del genoma completo (Liu *et al.*, 2014). En promedio, el desequilibrio de ligamiento en germoplasma tropical de maíz decae entre 5 y 10 Kpb (Lu *et al.*, 2011).

2.5 Importancia de los recursos genéticos en la búsqueda de resistencia

Una de las estrategias para reducir la incidencia de enfermedades en los cultivos es promover la diversidad genética (Strange & Scott, 2005). Pritsch (2001) define el objetivo de los programas de pre-mejoramiento como la búsqueda de variabilidad genética, por lo que las variedades nativas 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. 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, Upadhyaya, Varshney, & Gowda, 2013).

2.6 Resistencia genética del maíz al complejo mancha de asfalto

El primer estudio publicado sobre el mecanismo de herencia de la resistencia a mancha de asfalto lo realizaron Ceballos y Deutsch (1992), quienes identificaron la presencia de un solo gen dominante controlando la resistencia a la enfermedad, sin descartar los efectos aditivos. Dos décadas después, en el año 2014, un segundo estudio confirmó los efectos génicos de dominancia y aditivos, pero resalta que los segundos son de mayor importancia (Hernández, 2014), lo que implica que diversos genes podrían estar involucrados en la herencia de la resistencia y existe la posibilidad de que los genes puedan ser acumulados mediante selección recurrente. Los dos estudios anteriores se basaron en análisis genéticos convencionales, es decir, basados en análisis fenotípicos, aportando un valioso conocimiento sobre las bases de la herencia de resistencia al CMA.

Recientemente, dos estudios de asociación del genoma completo para resistencia a mancha de asfalto han sido publicados, utilizando marcadores SNP de última generación, los cuales se caracterizan por analizar conjuntos de líneas endogámicas y poblaciones de mapeo de ligamiento muy similares. Ambos presentan resultados similares, con ligeras diferencias, pero coinciden en la identificación de un QTL (Quantitative trait loci, por sus siglas en inglés) de efecto mayor (*qRtcs8-1*) confiriendo resistencia al CMA, el cual explica entre 13 y 43% de la variación fenotípica observada; además identificaron la presencia de algunos pocos QTL de efecto menor. Éste QTL de efecto mayor se encuentra en el cromosoma 8, bin 8.03; los QTL de efecto menor se encuentran en los cromosomas 2 (bins 2.05 y 2.07), 3 (bins 3.04 y 3.09) y 7 (bin 7.02) (Cao et al., 2017; Li et al., 2016). Estos resultados son coherentes con hallazgos de Ceballos y Deutsch (1992) y Hernández (2014).

2.7 Genes de resistencia a enfermedades en plantas

Las plantas, como cualquier otro ser vivo, han tenido que defenderse de otros organismos patógenos, tales como virus bacterias y hongos, por lo cual han desarrollado un sistema de defensa que consiste básicamente en una red de células y moléculas que interactúan entre sí para activar el sistema inmune (De Schutter & Van Damme, 2015; Hammond-Kosack & Jones, 1997). Desde la perspectiva genética, cualquier gen que modifica el nivel de susceptibilidad de una planta a un patógeno, se considera un gen de resistencia. En general existen dos tipos de resistencia, vertical y horizontal. La resistencia vertical o monogénica confiere altos niveles de resistencia (completa o parcial) y es controlada con mucha frecuencia por un solo gen de gran efecto, por lo cual ha sido ampliamente utilizado para el mejoramiento genético. Sin embargo, al aplicar una alta presión de selección, se acelera el proceso de adaptación y surgimiento de una nueva raza virulenta (McDowell & Simon, 2006). La mayoría de genes de resistencia de grande efecto se encuentra agrupados en el genoma y están codificados por proteínas denominadas NB-LRR (del inglés, nucleotide-binding-leucine-rich-repeat). Este tipo de genes parecen ser los más comunes en poblaciones naturales (Hulbert, Webb, Smith, & Sun, 2001). En estudios recientes se ha encontrado a los genes de efecto mayor ligados fuertemente a otros genes que también contribuyen a la resistencia, por lo que es más apropiado pensar que múltiples genes intervienen en la resistencia a enfermedades (Cook et al., 2012; Gassmann & Bhattacharjee, 2012). Por su parte, la resistencia horizontal o poligénica, es determinada por varios genes que pueden estar agrupados y cuyos efectos son relativamente pequeños, pero en

conjunto pueden aportar altos niveles de resistencia. Tiene la ventaja de ser más estable en el tiempo, ya que es efectiva contra múltiples razas del patógeno. Finalmente lo que se recomienda buscar en los programas de mejoramiento es la combinación de ambos tipos de resistencia, genes de efecto mayor y otros de efectos pequeños (Michelmore, Christopoulou, & Caldwell, 2013)

A nivel molecular se han determinado dos niveles de inmunidad en plantas, en base al tipo de receptor, el lugar donde ocurre el reconocimiento y las moléculas detectadas. El primer nivel consiste básicamente de receptores de reconocimientos de patrones (PPRs, del inglés pattern recognition receptors) localizados en la membrana celular, que detectan patrones moleculares asociados a patógenos (PAMPs, del inglés pathogen-associated molecular patterns) o patrones moleculares asociados a daños endógenos (damage-associated molecular patterns). Los PRRs (ej. RLK, del inglés receptor-like kinase) activan el sistema de defensa basal contra una amplia gama de patógenos. El segundo nivel de inmunidad se localiza al interior de la célula y consiste en receptores intracelulares NB-LRR que directamente e indirectamente reconocen proteínas producidas por patógenos; pueden conferir resistencia contra una amplia gama de patógenos, mediante la activación del sistema de defensa inmune que resulta en la muerte celular programada localizada, reacción conocida como respuesta de hipersensibilidad (Michelmore et al., 2013). Recientemente nuevos genes de resistencia están siendo identificados, entre ellos algunos análogos a NB-LRR, otros cuyo mecanismo de acción es la detoxificación y otros cuyo mecanismo de acción aun es desconocido (Gururani et al., 2012; Poland, Balint-Kurti, Wisser, Pratt, & Nelson, 2009).

2.8 Marcadores moleculares

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, Jahufer, Brouwer & Pang, 2005). Los marcadores de ADN más utilizados para análisis de QTL son RFLPs, RAPDs, AFLPs, SSR, ISSR y los más recientes son los marcadores SNP (polimorfismo en un solo nucleótido). Los SNPs representan el 90% de la variación genética total en cualquier organismo, y se origina principalmente por mutaciones puntuales, inserciones o deleciones, las cuales se han fijado en la población de una especie y pudiendo ocurrir en regiones codificantes y/o en regiones con función reguladora (P. K. Gupta, Rustgi, & Mir,

2008). Los SNPs son codominantes, bialélicos y son muy abundantes, con una tasa de 1 cada 300 pares de bases en el maíz distribuidos a lo largo del genoma, por lo que permite obtener una alta representatividad del genoma (P. K. Gupta, Roy, & Prasad, 2001). Por sus características, los SNPs son preferidos para la generación de mapas de asociación y de ligamiento con gran resolución, y la gran ventaja es que actualmente existe varias plataformas de genotipificación a gran escala que permiten la obtención de millones de marcadores SNP a un costo cada vez más accesibles.

Elshire y colaboradores (2011) desarrollaron el método Genotipificación por Secuenciación (GBS, por sus siglas en inglés), el cuál es robusto y simple para alta diversidad en plantas, capaz de analizar un gran número de muestras por ensayo y descubrir millones de SNPs. Es un sistema altamente multiplexado que genera cientos a cientos de miles de marcadores tipo SNPs en un gran número de muestras a partir de representaciones genómicas reducidas y basado en la plataforma de secuenciación de última generación. Se basa en el uso de enzimas de restricción tipo II sensibles a la metilación para reducir la complejidad del genoma; es simple, rápida, extremadamente específica, altamente reproducible, y puede captar importantes regiones del genoma que son inaccesibles para captura de secuencias (Elshire *et al.*, 2011).

2.9 Estimación de la heredabilidad de la tolerancia al CMA

La heredabilidad en sentido estricto fue definida por Falconer y Mackay (1996) como $h^2 = V_A/V_F$, donde V_A es la proporción de varianza genética aditiva y V_F la varianza fenotípica total, por lo que la heredabilidad se entiende como la proporción de varianza fenotípica explicada. Recientemente, con el surgimiento y aplicación de marcadores moleculares, nuevas formas de estimar la heredabilidad han sido desarrollados. En general, la heredabilidad en sentido estricto h^2 puede ser calculada a partir de los marcadores moleculares mediante la descomposición de la V_A en la suma de la varianza explicada por múltiples marcadores moleculares así: $V_A = V_{A1} + V_{A2} + V_{A3} + \dots + V_{Ap}$, donde p es el número de marcadores, bajo el supuesto de que los marcadores no están correlacionados entre sí (Meuwissen, Hayes, & Goddard, 2001; Nakaya & Isobe, 2012). En este sentido, la heredabilidad es una estimación de la proporción de varianza fenotípica explicada por la varianza genética aditiva capturada por

los marcadores, la cual fue denominada como “heredabilidad genómica” por Daetwyler, Bansal, Bariana, Hayden, & Hayes (2014).

2.10 Identificación de QTL y Estudio de Asociación de Genoma Completo

El mapeo de QTL consiste en localizar o identificar *loci* genómicos a través de marcadores moleculares, los cuales estén influenciado la variación de un carácter o rasgo de interés (ej. una enfermedad, altura, resistencia a una enfermedad, etc.) (Yi & Xu, 2008). El método convencional para mapeo de QTL, es el mapeo de ligamiento o mapeo genético, el cual se basa en el principio mendeliano de segregación, debido al fenómeno natural de recombinación de cromosomas (entrecruzamiento) durante la fase de meiosis. Por lo tanto, consiste en analizar la segregación de un rasgo fenotípico con respecto a los marcadores moleculares, e identificar que marcadores están más asociados al carácter de interés (Doerge, 2002). Dado el fenómeno de ligamiento, los marcadores que estén más cercanos o fuertemente ligados a genes causales son segregados y heredados juntos a la descendencia con mayor probabilidad que los marcadores que están más alejados o débilmente ligados a los genes (Collard, Jahufer, Brouwer, & Pang, 2005). Para llevar a cabo un mapeo genético es necesario formar una población de mapeo (población estructurada, con relación familiar conocida), la cual se caracteriza por ser una población segregante para el carácter o rasgo de interés. Generalmente se forman poblaciones biparentales (provenientes del cruce de dos parentales con rasgos contrastantes (ej. susceptible-resistente), siendo las más comunes y simples las poblaciones F2 y los retrocruzamientos, estas últimas obtenidas mediante el cruce de la generación F1 por una de las líneas parentales. El mapeo en poblaciones biparentales ha mostrado ser un método efectivo para identificar *loci* de grandes efectos y alelos raros (Chen et al., 2012), sin embargo, con la resolución que se obtiene en este tipo de poblaciones se dificulta la detección de QTL de efecto pequeño (Holland, 2007). Por su parte, el mapeo de asociación se basa en el uso de poblaciones no estructuradas, es decir, con relación familiar desconocida (ej. conjunto de poblaciones de origen diferente, paneles de asociación, colecciones de germoplasma, líneas o variedades élite), aprovecha el desequilibrio de ligamiento existente en los conjuntos de poblaciones, generado por los eventos de recombinación históricos, para detectar asociación entre QTL y marcadores, siendo efectivo para detectar nuevos genes o confirmar genes previamente identificados (Oraguzie & Wilcox, 2007).

Recientemente los estudios de asociación del genoma completo o GWAS (del inglés, *genome-wide association studies*), han mostrado gran resolución para identificar QTL de efectos pequeños asociados a la variación de un carácter o rasgo en diferentes paneles de asociación (Myles et al., 2009). Los análisis GWAS son básicamente un método de mapeo de asociación para identificar o mapear *loci* de caracteres cuantitativos o QTL y consisten en analizar toda la variabilidad existente a lo largo del genoma en función de los polimorfismos que existen a nivel de nucleótido simple (SNP) y determinar su asociación estadística a un carácter o rasgo de interés (Ali & Yan, 2012). En este sentido el estudio de asociación del genoma completo constituye una poderosa herramienta para identificar *loci* de un carácter cuantitativo responsables de la variación fenotípica (Sukumaran & Yu, 2014). Sin embargo, uno de los principales problemas de los análisis GWAS es que frecuentemente ocurre una alta tasa de falsos positivos, es decir, asociaciones espurias debido a otros factores como estructura de población y relación familiar o de parentesco (Pearson & Manolio, 2008). Las estimaciones de las asociaciones pueden realizarse mediante simples pruebas de t, modelos lineales generalizados (GLM, por sus siglas en inglés) con análisis de varianza, hasta modelos más complejos como los modelos lineales mixtos (MLM, por sus siglas en inglés) y regresiones lineales bayesianas. Varios modelos tipo MLM han sido desarrollados, los cuales toman en cuenta corrección por estructura de población y relación de parentesco con mayor efectividad en el control de falsos positivos (Yu et al., 2006; Zhao et al., 2007).

Varios estudios de asociación de genoma completo han sido realizados con éxito para caracteres de importancia agronómica en maíz. Weng *et al.* (2011) identificaron QTL que controlan la altura de planta utilizando un conjunto de 284 líneas puras y un total de 41,101 marcadores 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 el QTL aHS2.09 de resistencia al carbón de la espiga. Liu y colaboradores (2014) identificaron 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 para otros caracteres como tiempo de floración y arquitectura de hoja (Ali & Yan, 2012; Tian *et al.*, 2011). Para mancha de asfalto se ha identificado un QTL (qRtsc8-1) de efecto mayor y dos de efecto menor confirmando resistencia al CMA mediante análisis GWAS en un panel de asociación de líneas endogámicas de Cimmyt

combinado con mapeo de ligamiento en poblaciones biparentales (Cao et al., 2017; Li et al., 2016). A la fecha ningún análisis GWAS ha sido publicado utilizando poblaciones nativas de maíz de la región tropical y subtropical de Latinoamérica para explorar la gran variabilidad genética natural existente para la identificación de regiones genómicas y QTL asociados a la tolerancia genética al complejo mancha de asfalto.

3 Materiales y métodos

La investigación realizada fue de carácter cuantitativa para estimar qué marcadores están significativamente asociados a la tolerancia genética al CMA e identificar genes candidatos involucrados en el mecanismo de resistencia. Los alcances fueron, inicialmente descriptivo, luego asociativo o correlacional, predictivo y explicativo.

3.1 Materiales

3.1.1 Base de datos fenotípicos

Se utilizó una base de datos fenotípica de evaluación de tolerancia a mancha de asfalto desarrollada por el proyecto MasAgro-Biodiversidad-Cimmyt y la base de datos está disponible de forma gratuita para profesores, investigadores y mejoradores en <http://germinate.seedsofdiscovery.org/maize/> (CIMMYT.SeeD Maize Germinate, 2017). A continuación, se presenta una breve descripción de la base de datos: Contiene información fenotípica de mestizos derivados del cruce de poblaciones nativas con probadores (híbridos de cruce simple) evaluados para tolerancia a mancha de asfalto bajo condiciones naturales de infección, en el año 2011, en la localidad Guadalupe Victoria, Estado de Chiapas, México. La mayoría de las poblaciones nativas utilizadas como parentales para obtención de los mestizos fueron colectadas entre 1952 y 1960 en México, Brasil y Guatemala, y en menor proporción provenientes de otros países de Centro y Sudamérica. Estas poblaciones forman parte de la Colección Núcleo del Banco de Germoplasma de Maíz del Cimmyt y están adaptadas a condiciones tropicales y subtropicales y representan diferentes grupos raciales. En la evaluación también se incluyeron híbridos comerciales como testigos. La base de datos contenía el registro de dos lecturas del nivel de tolerancia al CMA, TSC1 y TSC2 (del inglés, Tar Spot Complex), que corresponden a lecturas en pre-floración y post-floración, respectivamente, además de otros caracteres agronómicos de importación como el rendimiento de grano por parcela en gramos. Es importante resaltar que los datos de las variables corresponden a los valores BLUPs (por sus

siglas en inglés Best Linear Unbiased Prediction) obtenidos por los investigadores del Cimmyt a partir del análisis del modelo estadístico del ensayo, razón por la cual los datos del nivel de tolerancia que fueron tomados en campo bajo una escala ordinal (Ceballos & Deutsch, 1992), en la base de datos se encuentran en escala continua. Para los fines de esta investigación, considerando que las epidemias que ocurren después de floración son mucho más severas dado el estado avanzado de la enfermedad (Hock et al., 1989), se tomó únicamente la lectura TSC2 como variable a analizar. Previo al análisis estadístico de los datos fenotípicos, se realizó un control de calidad, el cual se detalla en el inciso 3.4.1.1.

3.1.2 Base de datos genotípicos

Se utilizó la base de datos genotípicos generada como parte del proyecto MasAgro-Biodiversidad-Cimmyt y está disponible de forma gratuita para investigadores y mejoradores en <http://germinate.seedsofdiscovery.org/maize/> (CIMMYT.Seed Maize Germinate, 2017). A continuación, se presenta una breve descripción de la base de datos: contiene información genotípica de un panel de asociación de más de cuatro mil accesiones del Banco de Germoplasma de Maíz del Cimmyt, incluyendo las poblaciones nativas que se analizaron en este estudio para tolerancia a mancha de asfalto. Contiene 946,072 marcadores SNP distribuidos en los diez cromosomas del maíz, y los datos faltantes fueron imputados utilizando el software Beagle v.4 (Browning & Yu, 2009). La información genotípica fue generada utilizando tecnología de última generación denominada Genotipificación por Secuenciación (GBS, por sus siglas en inglés), desarrollado por Elshire y colaboradores (2011). La base de datos viene en formato HapMap, el cual consiste en un solo archivo que contiene el identificador de SNP, cromosoma, posición y genotipo codificado según la Unión Internacional de Química Pura y Aplicada (IUPAC, por sus siglas en inglés). En este formato la información de los marcadores se almacena en las filas y los taxa (individuos) en las columnas.

Previo a realizar el análisis preliminar de la información genotípica se realizó un control de calidad de marcadores para los fines de esta investigación, el cual se detalla en el inciso 3.4.1.

3.1.3 Material vegetal para la evaluación preliminar en campo del nivel de tolerancia genética al CMA

Se utilizó germoplasma de maíz de líneas endogámicas, poblaciones nativas, híbridos experimentales e híbridos comerciales como testigos. Catorce líneas endogámicas de maíz de Cimmyt o CMLs (del inglés Cimmyt Maize Lines) y ocho de las poblaciones nativas fueron obtenidas del Banco de Germoplasma de Maíz del Cimmyt, México, proveídas por esta institución en términos de colaboración científica. Las poblaciones nativas fueron colectadas en Guatemala entre 1952 y 1953, y una en Puebla, México. Adicionalmente, cuatro poblaciones nativas cultivadas por los agricultores de la región de Ixcán, Quiché, fueron incluidas en el ensayo, las cuales han mostrado algún nivel de tolerancia a la macha de asfalto. Las líneas endogámicas son originarias de diferentes programas de mejoramiento del Cimmyt, las cuales han mostrado tolerancia o resistencia a mancha de asfalto en otras evaluaciones (CIMMYT, 2013). Se incluyeron además cinco híbridos experimentales del Instituto de Ciencia y Tecnología Agrícola de Guatemala (ICTA) y seis híbridos comerciales de la región. El material vegetal utilizado se muestra en el cuadro 1. El total de materiales fue de 37.

Cuadro 1. Lista de germoplasma de maíz utilizado en la evaluación preliminar de tolerancia a mancha de asfalto en Ixcán, Quiché.

Líneas endogámicas	Poblaciones nativas	Híbridos experimentales	Testigos comerciales
CML264	Procedentes del Cimmyt:	EXP.1 ICTA	HB-83 (ANAGRAB)
CML269	GUAT100	EXP.2 ICTA	DK-390
CML451	GUAT103	EXP.3 ICTA	HB-83(VALLE VERDE)
CML492	GUAT1094	EXP.4 ICTA	DOW ELANCO
CML495	GUAT153	EXP.5 ICTA	JC-24
CML497	GUAT303		HS-5 (CB/MONSANTO)
CML498	GUAT823		
CML500	GUAT83		
CML549	PUEB GP27		
CML552	Procedentes de la región		
CML553	de Ixcán:		
CML574	Santa Ana		
CML575	Ariquin		
CML576	Arrocillo Rivera		

3.2 Métodos

3.2.1 Estimación de la heredabilidad de la tolerancia al CMA

Para obtener una estimación de la heredabilidad genómica (\hat{h}^2), definida como la proporción de varianza fenotípica explicada por la contribución aditiva cada SNP, se empleó el método LDAK (linkage disequilibrium adjusted kinship) propuesto por Speed, Hemani, Johnson, & Balding (2012), el cual ha sido recientemente validado en caracteres humanos complejos (Speed et al., 2017). El método estima la proporción de varianza fenotípica explicada (\hat{h}^2) por la matriz de relaciones genómicas ajustada por la correlación existente entre SNP-SNP que están muy cercanos debido al desequilibrio de ligamiento (LD). Para ello, el método evalúa el patrón de LD local entre pares de SNP mediante el cálculo de una matriz de correlaciones entre pares de SNP, la cual contiene básicamente la ponderación para cada SNP. Esta matriz es utilizada para calcular la matriz de relaciones ajustada, donde la contribución de cada SNP está determinada por su ponderación. A partir de la matriz de relaciones genómica ajustada se realiza un análisis de componentes principales (ACP) para corregir por estructura de población. A partir de la kinship ajustada se estiman los componentes de varianza (varianza genética y varianza residual) y heredabilidad mediante un análisis de máxima verosimilitud restringida (REML, del inglés restricted maximum likelihood), incluyendo los ACP como covariables. Es importante resaltar que, dado que la matriz de relaciones genómica (obtenida a partir de marcadores SNPs) se basa en relaciones genéticas aditivas, la varianza genética que se estima corresponde a los efectos aditivos, por consiguiente, la heredabilidad obtenida corresponde al sentido estricto (Endelman & Jannink, 2012). La estimación de la heredabilidad se realiza a partir de los componentes de varianza mediante la siguiente ecuación propuesta por (Daetwyler et al., 2014):

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

donde $\hat{\sigma}_g^2$ es una estimación de la varianza genética aditiva capturada por los marcadores moleculares, $\hat{\sigma}_e^2$ es una estimación de la varianza residual, la cual incluye la varianza genética no aditiva (dominancia y epistasis), la variación ambiental, la variación debida a la interacción genotipo ambiente y la varianza del error.

3.2.2 Identificación de marcadores asociados a la tolerancia genética al CMA

Para la identificación de marcadores SNP asociados a la tolerancia genética al CMA se realizó un análisis de asociación de genoma completo. Se evaluaron dos modelos estadísticos, “Single Marker” y “BayesB”. El primero ha sido ampliamente utilizado en análisis GWAS en plantas y consiste en evaluar un solo marcador a la vez, de tal forma que se realizan tantas regresiones y pruebas de hipótesis como marcadores existan. El modelo BayesB ha demostrado potencial para detección de QTL en estudios de simulación y aún no ha sido probado con datos reales (Berg, Fritz & Boichard, 2013; Sahana *et al.*, 2010) y se basa en el hecho de que en realidad hay muchos *loci* sin varianza genética (no segregantes) y unos pocos *loci* con varianza genética. Para lograr 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 normal. 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 normal (Meuwissen, Hayes & Goddard, 2001). La descripción de cada modelo de se presenta en el inciso 3.4.4.

3.2.3 Proporción de variación fenotípica total explicada por los marcadores

La estimación de la proporción de variación fenotípica total explicada (PVE) por los marcadores se obtuvo del coeficiente de determinación r^2 de los modelos de regresión lineal múltiple ajustados. En el caso del modelo single marker, dado que se prueba un marcador a la vez, se obtuvo el coeficiente de determinación r^2 de cada marcador derivado del ajuste del modelo lineal mixto, interpretado como la cantidad de varianza fenotípica explicada por un marcador puntual. Además, se ajustó un modelo de regresión lineal múltiple incluyendo los marcadores asociados significativamente identificados con modelo lineal mixto single marker, a partir del cual se obtuvo el coeficiente de determinación múltiple ajustado por el número de marcadores r_{aj}^2 , el cual es equivalente a la PVE por los marcadores en conjunto (Collard *et al.*, 2005). En el caso del modelo BayesB, al ser una regresión simultánea sobre todos los marcadores a la vez, no es posible estimar la PVE por cada marcador puntual. Por lo cual solo se realizó el ajuste de un modelo de regresión múltiple para estimar la PVE por los marcadores en conjunto. Además, se ajustaron dos modelos regresión múltiple, uno incluyendo los 40 SNPs identificados con ambos modelos (single marker y BayesB) y el otro incluyendo únicamente los 11 SNPs identificados en común con ambos modelos. Dado que los modelos de regresión

son lineales, los efectos son aditivos, por lo que la variación explicada puede ser atribuida a efectos genéticos aditivos.

3.2.4 Identificación de genes candidatos

Una vez identificados los marcadores SNPs significativamente asociados a la tolerancia genética al CMA, se realizó el análisis del gen candidato para identificar potenciales genes involucrados en el mecanismo de defensa al CMA de los materiales tolerantes. El procedimiento consiste en utilizar la posición física de los marcadores para ubicarlos en el genoma mediante la utilización de la base de datos genética y genómica del maíz (MaizeGDB database, por sus siglas en inglés) (Andorf et al., 2016), basado en el genoma de referencia “B73” RefGen_v2(MGSC), disponible en <http://www.maizegdb.org/>. Para identificar la posible función del gen candidato se utilizó la base de datos del National Center for Biotechnology Information (NCBI, por sus siglas en inglés), disponible en <http://www.ncbi.nih.gov/>.

Se propusieron como genes candidatos a aquellos genes que contienen al SNP dentro de su secuencia y/o aquellos que están cerca (± 100 Kpb corriente abajo y corriente arriba) de cada SNP identificado en el análisis GWAS, tomando en cuenta las referencias previamente publicada sobre genes de resistencia y las principales proteínas codificadas. Se tomó como gen candidato el gen más cercano al SNP, y genes aledaños potencialmente involucrados se anotaron como genes de interés. En el caso de proteínas no caracterizadas aún para maíz, se tomó como referencia genes análogos en otros cultivos como arroz y *arabidopsis*.

3.2.5 Evaluación preliminar en campo del nivel de tolerancia genética al CMA

3.2.5.1 Aspectos generales del ensayo

En esta evaluación preliminar para tolerancia genética a mancha de asfalto se incluyeron un total de 37 materiales de maíz: 14 líneas CMLs, 12 poblaciones nativas, cinco híbridos experimentales del ICTA y seis híbridos comerciales como testigos susceptibles. La evaluación se realizó en la aldea La Nueva Máquina, del municipio de Ixcán, Quiché, en las coordenadas 16°00'23.0" N y 90°33'14.2" O, al 145 m sobre el nivel del mar. Dado que la reproducción in vitro de *P. maydis* para inoculación artificial no es posible debido a que es un parásito obligado (Dittrich, Hock, Kranz, & Renfro, 1991), el experimento se realizó bajo condiciones de infección natural, por lo que el desarrollo de la epidemia estuvo en función de las condiciones

climáticas de la localidad. En el municipio de Ixcán se han reportado fuertes epidemias de infección natural en los meses de noviembre a febrero (ICTA/Red SICTA/IICA/Cooperación Suiza, 2011), sin embargo, la evaluación de tolerancia se realizó en los meses de agosto a noviembre de 2017 para cumplir con el límite de tiempo del proyecto. Previo a la evaluación se realizó un incremento de semilla en los meses de abril a julio del mismo año, ya que la cantidad de semilla de las líneas endogámicas proveídas por el Banco de Germoplasma de Maíz de CIMMYT era muy limitada (50 semillas por línea), y así garantizar suficiente disponibilidad para la evaluación en curso y una reserva para investigaciones subsecuentes. El método utilizado para multiplicar semillas de las líneas endogámicas fue por autofecundación, y en situaciones cuando se presentó asincronía floral, se realizó incrementos por cruza fraternales. El vivero se estableció en el municipio de Nueva Concepción, Escuintla, ya que reunía condiciones favorables de tipo agronómico, además de facilidad en el aspecto logístico.

3.2.5.2 Diseño experimental y evaluación de la enfermedad

Considerando que el número de tratamientos es relativamente alto (37 genotipos), el ensayo se estableció bajo un diseño alfa lattice, por su mayor eficiencia que un diseño en bloques completos al azar el número de tratamientos es grande (Barreto, Edmeades y Crossa, 1994). Los parámetros del diseño fueron: ocho parcelas o unidades experimentales por bloque, cinco bloques por repetición, dos repeticiones y 40 tratamientos (incluyendo 3 réplicas internas). Los bloques se definieron en función de la variabilidad del suelo (pendiente, fertilidad, textura, etc.). La unidad experimental consistió en un surco de 5 m de largo, 0.20 m entre planta y 26 plantas por surco. El distanciamiento entre surco fue de 0.80 m. Considerando que tanto las líneas puras y los híbridos son materiales altamente uniformes por su condición genética, se espera tener buena precisión para detectar diferencias entre los genotipos bajo los parámetros utilizados.

Se realizaron dos mediciones del nivel de tolerancia al CMA, una en pre-floración y en post-floración, utilizando la escala de Ceballos y Deutsch (1992). En cada unidad experimental, se midió el nivel de tolerancia en las seis plantas centrales de manera individualmente. Posteriormente se obtuvo el promedio de las seis plantas para realizar los respectivos análisis estadísticos.

3.3 Operacionalización de las variables o unidades de análisis

Tabla 1
Operacionalización de variables o unidades de análisis

Objetivo específicos	Variables	Técnicas	Instrumentos	Medición o cualificación
a)	Marcadores moleculares SNP	Cuantitativa: Análisis de asociación del genoma completo	Equipo de cómputo, software TASSEL, y R statistics.	Nivel de significancia (p-valor)
b)	Genes candidatos	Cualitativa: Exploración en base de datos del genoma del maíz.	Equipo de cómputo e internet. Base de datos MaizeGDB	Posición física en pares de bases (pb) de los marcadores en el genoma y función putativa del gen.
c)	Nivel de tolerancia genética al CMA	Cuantitativa: Observación visual.	Escala de 0 a 5 de Ceballos y Deutsch (1992).	0 (resistente) a 5 (muy susceptible)

3.4 Procesamiento de datos y plan de análisis

3.4.1 Manejo de las bases de datos y control de calidad

3.4.1.1 Base de datos fenotípica

El manejo y control de calidad consistió básicamente en eliminar las observaciones con datos fenotípicos faltantes o sin información genotípica. Para ello se utilizó Microsoft Excel 2010. De un total de 535 observaciones, únicamente 463 poblaciones nativas contaban con la información completa del registro del nivel de tolerancia a mancha de asfalto e información de marcadores moleculares, el resto no se incluyó en el análisis. Adicionalmente se incluyeron en el análisis fenotípico dos testigos comerciales que contaban con las lecturas de tolerancia y rendimiento, sin embargo, no se incluyeron en el análisis molecular por falta de información de marcadores.

3.4.1.2 Base de datos genotípica

La base de datos genotípica viene agrupada en un archivo por cromosoma, del 1 al 10. Se realizó la unificación de los 10 cromosomas en un solo archivo utilizando el software de licencia libre TASSEL V5.2 (Bradbury *et al.*, 2007). A partir del archivo unificado que contenía las 4000 poblaciones y 946 mil SNP de los 10 cromosoma, se procedió a extraer la información genotípica de las 535 poblaciones nativas utilizando el software de licencia libre PLINK (Purcell *et al.*, 2007). De las 535 poblaciones nativas, únicamente 463 contaba con la

información de marcadores SNP, las cuales representa una gran diversidad y variabilidad genética.

El control de calidad consistió en filtrar aquellos marcadores con frecuencia del alelo menor (MAF, por sus siglas en inglés) mayor de 5 %. Marcadores con $MAF < 0.05$ fueron descartados, ya que usualmente producen resultados inestables (Corvin, Craddock & Sullivan, 2010; Weng *et al.*, 2011). Con este primer filtrado se eliminaron también aquellos marcadores monomórficos, cuyo genotipo es el mismo en todas las muestras y por consiguiente no explican la variabilidad del nivel de severidad del CMA. Un segundo filtrado se realizó para seleccionar únicamente los marcadores con la información genotípica completa (0 % datos faltantes) y se descartaron aquellos que tenían datos faltantes. Después del filtrado se obtuvo un total de 329,692 SNPs polimórficos para el panel de las 463 poblaciones nativas, sin datos faltantes. Se realizó además el cálculo de las frecuencias alélicas de los marcadores asociados significativamente. El control de calidad y cálculo de frecuencias alélicas de marcadores asociados significativamente se realizó con el software TASSEL.

3.4.2 Análisis de datos fenotípicos

Se realizó un análisis de estadísticas descriptivas (ej. media, mediana, cuantiles, curtosis, etc.) de los BLUPs del nivel de tolerancia genética (TCS2) de los 463 mestizos más dos testigos. Para observar el comportamiento de la distribución de la variable TCS2 se realizó un histograma de frecuencias combinado con un gráfico box-plot utilizando la función *histBxp()* del paquete *sfsmisc* en R. Para verificar la aproximación de los datos a una distribución normal se realizó un gráfico cuantil-cuantil utilizando la función *qqnorm()*. Para observar la correlación entre el nivel de severidad de mancha de asfalto y el rendimiento de grano por parcela (g) se realizó un gráfico de dispersión (x,y) utilizando la función *plot()*. Los análisis anteriores fueron realizados con el software de licencia libre R Statistics (R-Core Team, 2015).

3.4.3 Distribución general de los marcadores en el genoma del maíz

Para verificar que los SNPs estuvieran distribuidos uniformemente a lo largo de los 10 cromosomas, se realizó la visualización con el software de licencia libre Flapjack desarrollado por Cimmyt (Milne *et al.*, 2010). Además, se generó el sumario de parámetros generales de la información genotípica (ej. frecuencias alélicas, proporción de heterocigotos, datos faltantes, etc.). Para esto se utilizará el software TASSEL.

3.4.4 Estimación de la heredabilidad de la resistencia al CMA

Para implementar este método se utilizó el software de licencia libre LDAK v.5. (Speed et al., 2017), el cual corre en plataforma UNIX. Primero se calculó la ponderación de los SNPs que reflejan los patrones locales de correlación entre SNPs. Además, se consideró en el efecto del decaimiento del desequilibrio de ligamiento con la distancia, sugerido por los desarrolladores del software para el caso de plantas. Posteriormente se calculó la matriz de relaciones genómicas de las 463 muestras (poblaciones nativas). A partir de la matriz de relaciones ajustada, se realizó el análisis de componentes principales (ACP) para tomar en cuenta la estructura de población en el panel de asociación de poblaciones nativas. Los eigen-vectores obtenidos del ACP fueron incluidos en el análisis REML como covariables. Finalmente, se estimaron los componentes de varianza mediante el análisis REML, a partir de los cuales se estimó la heredabilidad genómica utilizando el estimador de Daetwyler y colaboradores (2014).

3.4.5 Estudio de asociación del genoma completo

3.4.5.1 Modelo Single Marker

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

$$\mathbf{y} = \mathbf{W}\mathbf{v} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad (1)$$

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

una prueba de F. El umbral de significancia para seleccionar los marcadores fue un valor de $\alpha = 0.0001$. Para implementar este modelo se utilizó el software TASSEL V.5.2. Se realizó un gráfico tipo manhattan para visualizar de manera global la distribución de los marcadores en los cromosomas y el respectivo puntaje en una escala de $-\log_{10}(p)$, donde p es el valor de probabilidad de la prueba de asociación de cada marcador; además se generó otro gráfico tipo cuantil-cuantil para observar la eficiencia del modelo en el control de falsos positivos. Ambos gráficos fueron realizados con el paquete qqman en R (Turner, 2014).

3.4.5.2 Modelo BayesB

Se utilizó el modelo lineal descrito por Pérez-Rodríguez y de los Campos (2014):

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

donde \mathbf{y} es el vector de fenotipos, $\mathbf{1}$ es un vector cuyos elementos son unos, asociado al intercepto μ , \mathbf{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 $R = \mathbf{I}\sigma_\varepsilon^2$ e \mathbf{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 & de los Campos, 2014):

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

La distribución *a priori* de la varianza para los marcadores con varianza genética es $\sigma_j^2 \sim \chi^{-2}(v, S)$ para $j = 1, \dots, p$ y donde v son los grados de libertad y S es la escala del parámetro. 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 0.1% de marcadores tienen efecto diferente de cero (Pérez-Rodríguez & de los Campos, 2014).

Como criterio para selección de marcadores se utilizó la probabilidad posterior de inclusión (PIP, del inglés Posterior Inclusion Probability), que es la probabilidad de que un marcador tenga un efecto diferente de cero (Berg *et al.*, 2013; Pérez-Rodríguez & de los Campos, 2014). El umbral de selección fue una probabilidad mayor o igual que 0.05 (PIP > 0.05). Para implementar este modelo se utilizó el paquete BGLR (del inglés, Bayesian Generalized Linear Regression, en inglés) en R (Pérez-rodríguez & de los Campos, 2014). Se

realizó un gráfico tipo manhattan con el paquete qqman en R utilizando los valores de PIP obtenidos del modelo.

3.4.6 Proporción de variación fenotípica total explicada por marcadores

Para realizar el ajuste de los modelos de regresión lineal múltiple con los marcadores asociados significativamente se utilizó la función $lm()$ en R statistic, a partir del cual se obtuvo el r_{aj}^2 ajustado por número de marcadores. Se ajustó un modelo incluyendo los marcadores significativos identificados con TASSEL, otro con los marcadores identificados con el modelo BayesB y otro incluyendo los marcadores asociados con ambas metodologías (40 SNPs) para determinar la proporción de variación explicada en conjunto. Adicionalmente, se ajustó un modelo incluyendo los 11 SNPs que fueron identificados en común por ambos modelos.

3.4.7 Análisis de los datos de la evaluación de campo

Se realizó un análisis gráfico descriptivo para identificar los materiales con mayor nivel de tolerancia a partir del dato promedio del nivel de tolerancia genética al CMA obtenido por unidad experimental.

3.5 Coherencia de la propuesta de investigación.

Tabla 2 Coherencia de la propuesta de investigación

Objetivos específicos	Actividad	Productos o hallazgos esperados	Responsables
a)	Identificar marcadores moleculares SNPs asociados a la tolerancia genética al complejo mancha de asfalto en poblaciones nativas de maíz	Marcadores SNP asociadas a la resistencia al CMA identificados.	Coordinador e investigador I.
b)	identificar genes candidatos potencialmente involucrados en mecanismos de defensa al complejo mancha de asfalto	Genes candidatos potencialmente involucrados en el mecanismo de defensa al CMA identificados.	Coordinador e investigador I.
c)	Evaluar de manera preliminar en campo el nivel de tolerancia genética al CMA de poblaciones nativas, líneas endogámicas e híbridos experimentales y comerciales, en condiciones del municipio de Ixcán, Quiché	Se espera identificar germoplasma resistencia al CMA con la finalidad de que puedan ser utilizadas como potenciales donadores en programas de mejoramiento genético de maíz.	Coordinador e investigador II.

4 Resultados

4.9 Análisis descriptivo de la base de datos fenotípicos

Las estadísticas descriptivas para los BLUPs del nivel de tolerancia al CMA de la variable *TSC2* de la base de datos fenotípica se presentan en el Cuadro 2. Se observa suficiente diferencia entre valores máximos y mínimos, que van desde 2.428 hasta 4.531, lo cual permite diferenciar claramente entre materiales tolerantes y susceptibles, confirmando así la presencia de la enfermedad por infección natural en el área de investigación. El valor positivo de la curtosis (0.157) indica que el nivel de tolerancia tiende a concentrarse alrededor de la media (3.60). El coeficiente de variación (CV) refleja niveles de variación aceptable (11.2%) considerando el elevado número de mestizos evaluados y la alta variabilidad genética intrínseca de los mismos. Los mestizos derivados de las poblaciones nativas CHIS474, OAXA280, GUAT153, HAIT13 y TABA26 fueron los que mostraron mayores niveles de tolerancia 2.428, 2.430, 2.432, 2.450 y 2.463, respectivamente. Los mestizos más susceptibles son los derivados de las poblaciones OAXA130, BRAZ1 y QROO53, con valores de 4.467, 4.503 y 4.531, respectivamente.

Cuadro 2 Estadísticas descriptivas del nivel de tolerancia al CMA (*TSC2*).

Lectura	Media	Min	Max	C1	Mediana	C3	DE	CV	Curtosis
<i>TSC2</i> (Post-floración)	3.609	2.428	4.531	3.360	3.660	3.890	0.403	0.112	0.157

DE: desviación estándar

CV: coeficiente de variación

C1, C3: primer y tercer cuartil respectivamente.

La Figura 1 muestra un histograma de frecuencias combinado con un gráfico tipo box-plot, y un gráfico cuantil-cuantil para observar la dispersión de los datos y el ajuste a una distribución normal. Se puede observar una asimetría ligeramente negativa, pero en general los datos se concentran alrededor de la media y se aproximan a una distribución normal. En el box-plot se puede observar que el 50% de las poblaciones posee valores de tolerancia entre 3.36 y 3.89, correspondientes a la categoría de moderadamente susceptibles a extremadamente susceptibles. Únicamente el 9.1% de las poblaciones (42) se encuentran en la categoría moderadamente resistente (>2 y < 3) según la escala de Ceballos & Deutsch (1992).

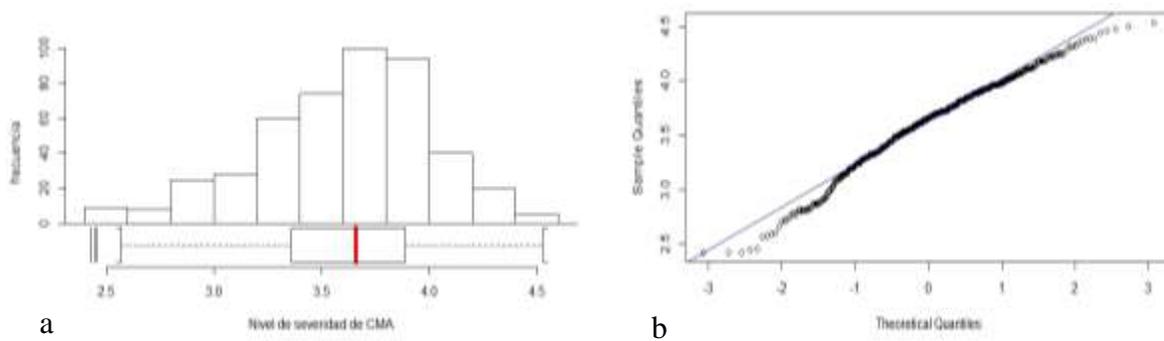


Figura 1. Gráficos de distribución del nivel de tolerancia al CMA (TSC2). a) Histogramas de frecuencias combinado con box-plot de la variable. b) Gráfico tipo cuantil-cuantil de probabilidad normal. Sample Quantiles se refiere a los cuantiles de los valores observados de la variable TSC2, mientras que Theoretical Quantiles corresponde a los cuantiles teóricos esperados bajo una distribución normal.

La Figura 2 muestra la correlación entre el nivel de tolerancia y rendimiento. Se aprecia una correlación negativa (-0.398), por lo que las poblaciones más susceptibles tienden a presentar rendimientos más bajos. Llama la atención el mestizo derivado de la población VERAGP24_484, sobresaliente en rendimiento, a pesar de que su nivel de tolerancia no es de los mejores.

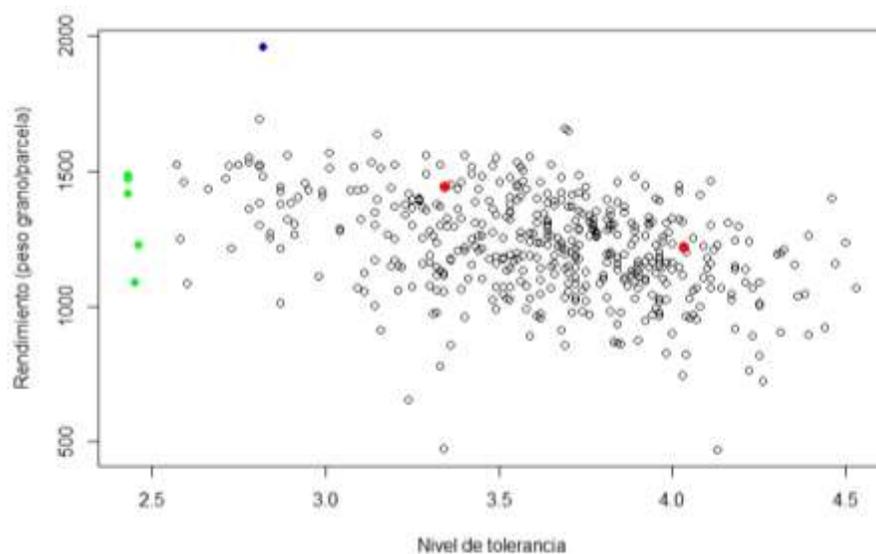


Figura 2. Gráfico de correlación entre nivel de tolerancia a mancha de asfalto y rendimiento. Puntos color verde representan los mestizos más tolerantes; en color rojo se representan los testigos comerciales. Punto azul representa al mestizo derivado de la población VERAGP24_484.

4.10 Distribución y estadísticas generales de la información genotípica

Después del filtrado, un total de 329,692 SNPs de alta calidad fueron obtenidos para el panel de asociación de 463 poblaciones nativas, lo que refleja una gran diversidad genética en el panel de asociación. El número total de marcadores SNP por cromosomas se presentan en el Cuadro 3. El cromosoma uno contiene la mayor cantidad de marcadores SNPs (51,322), lo que resulta en una alta representatividad de este cromosoma. El cromosoma diez contiene el menor número de marcadores SNPs (23,145). Las Figura 4 y Figura 5 muestran la distribución de los SNP a lo largo del genoma del maíz. En general se observa una alta cobertura de SNPs en todos los cromosomas, y por consiguiente, una alta representatividad del genoma del maíz, con tendencia a concentrarse más en las regiones teloméricas que en la región del centrómero. Antes del filtrado la frecuencia promedio del alelo menor (MAF) era de 0.0838 y la proporción de SNPs con $MAF < 0.05$ era de 65.07 % (615,661). Después del control de calidad, la frecuencia promedio del alelo menor (MAF) fue de 0.2269, con un rango de 0.05 a 0.5. En la Figura 3 pueden observarse los histogramas de frecuencia de MAF antes y después del filtrado.

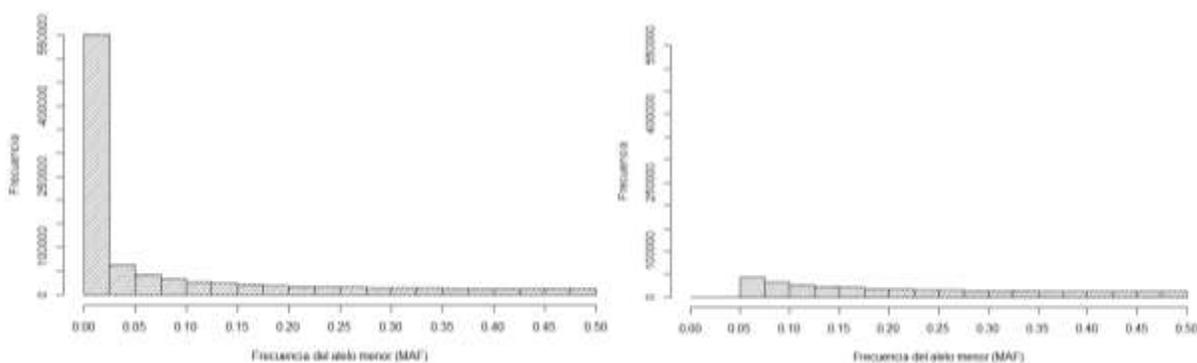


Figura 3. Histograma de frecuencias del alelo menor antes y después del control de calidad. Los marcadores con MAF menor de 0.05 fueron descartados.

Cuadro 3. Número total de SNPs contenidos en los cromosomas después del control de calidad.

Cromosoma	Número de SNPs	Cromosoma	Número de SNPs
1	51,322	6	26,210
2	41,232	7	27,664
3	38,023	8	28,047
4	30,072	9	25,168
5	38,809	10	23,145

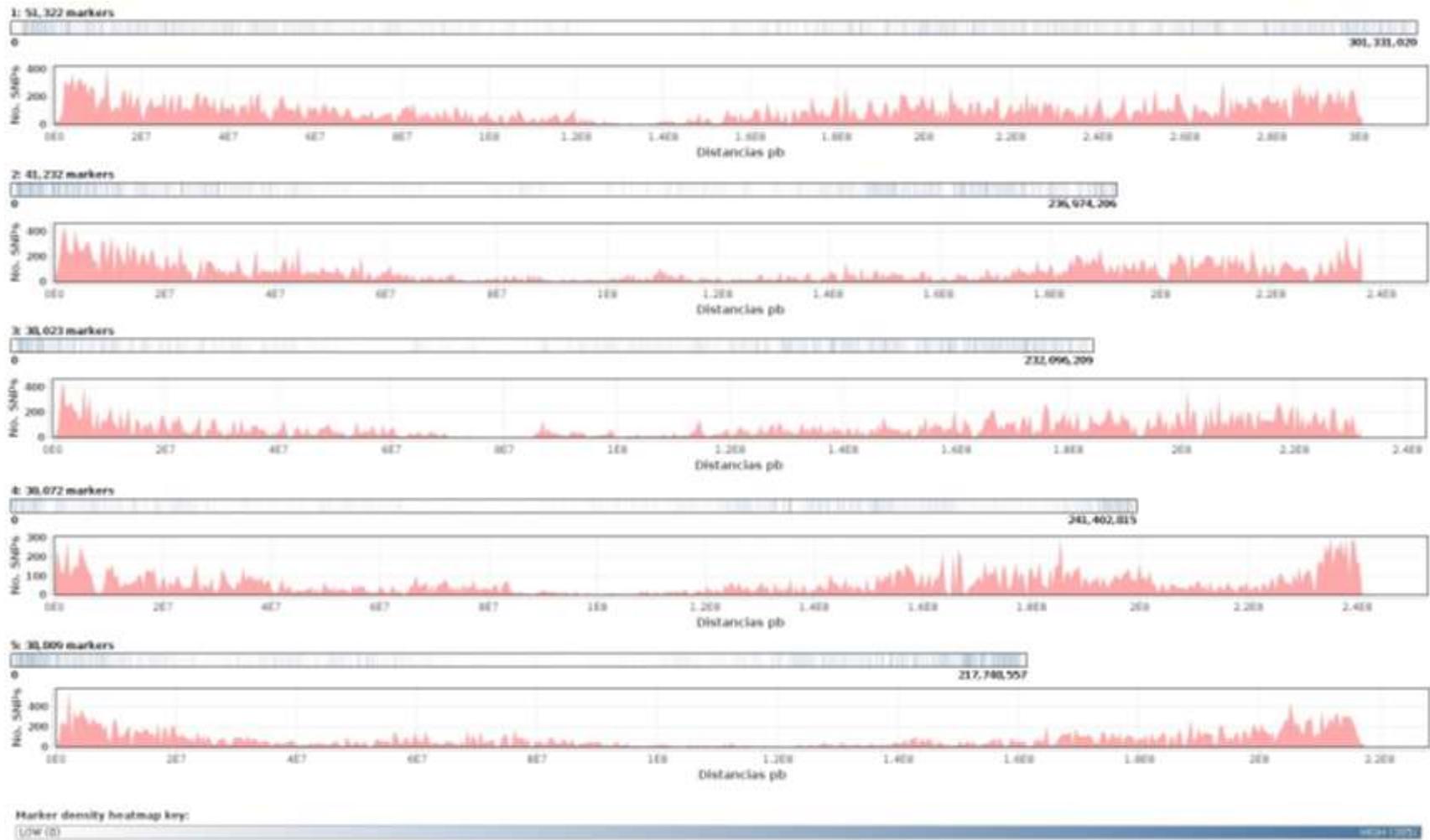


Figura 4. Distribución de los SNPs en los cromosomas del 1 al 5. Incluye dos tipos de gráficas por cromosoma, una de barras y la otra de líneas. Al final se incluye la clave del nivel de densidad de marcadores (Marker density heatmap key) para la interpretación de la gráfica de barras. En la gráfica de barras se indica el número total de marcadores en cada cromosoma.

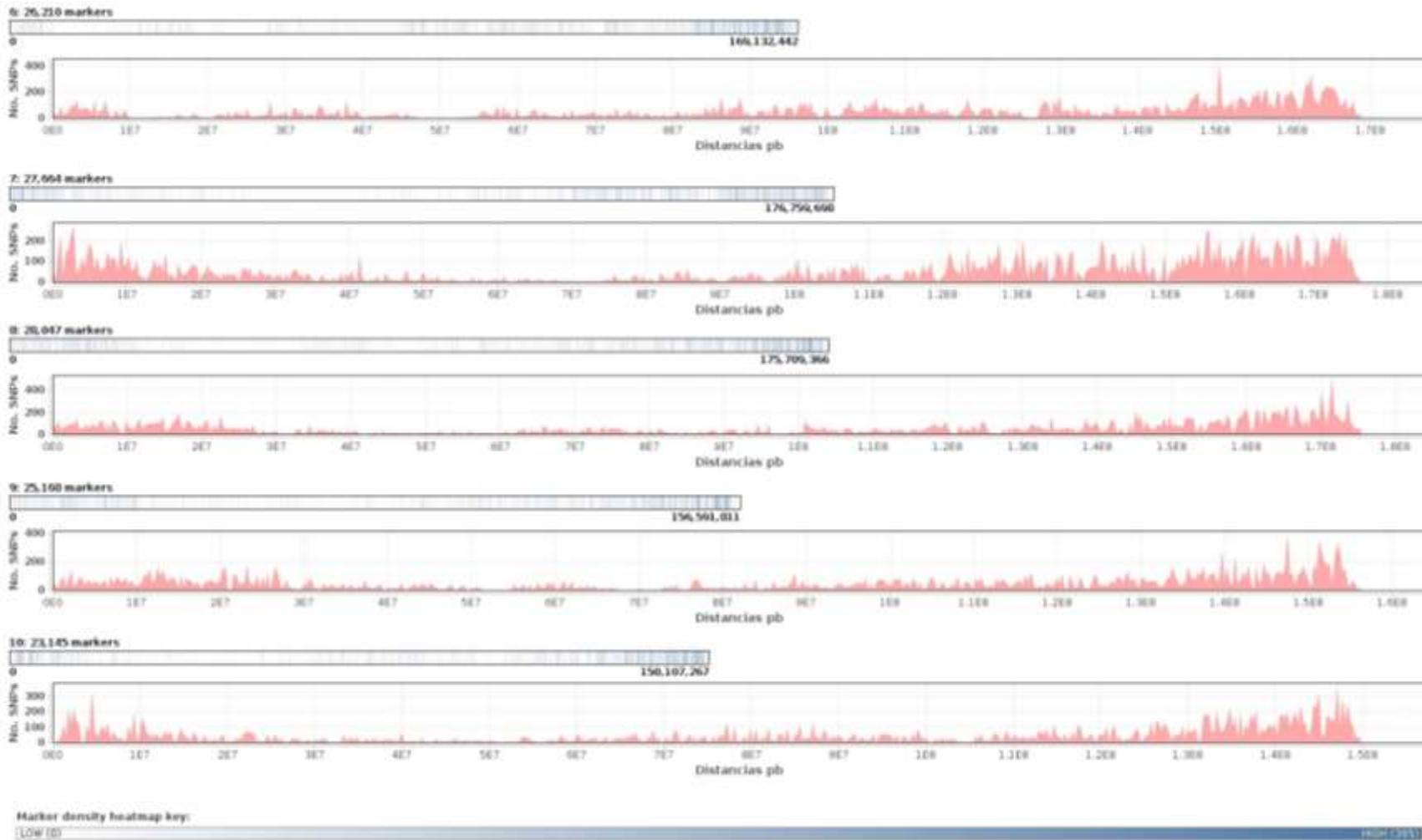


Figura 5. Distribución de los SNPs en los cromosomas del 6 al 10. Incluye dos tipos de gráficas por cromosoma, una de barras y la otra de líneas. Al final se incluye la clave del nivel de densidad de marcadores (Marker density heatmap key) para la interpretación de la gráfica de barras. En la gráfica de barras se indica el número total de marcadores en cada cromosoma.

4.11 Estimación de la heredabilidad de la resistencia al CMA basada en SNPs

A partir del análisis REML se obtuvieron los componentes de varianza ($\hat{\sigma}_g^2 = 0.0798$ y $\hat{\sigma}_e^2 = 0.0197$) y una estimación de la heredabilidad en sentido estricto $\hat{h}^2 = 0.8020$. El alto grado de heredabilidad revela que la tolerancia a mancha de asfalto es predominantemente controlada por efectos genéticos aditivos. Además, indica una alta repetibilidad y que los datos fenotípicos son confiables para los análisis genéticos posteriores.

4.12 Identificación de marcadores SNP asociados a tolerancia a mancha de asfalto

Un total de 40 marcadores SNPs fueron asociados significativamente a la tolerancia genética a mancha de asfalto mediante el análisis de asociación de genoma completo con ambos modelos, distribuidos de la siguiente manera: con el modelo single marker se obtuvieron 25 SNPs y con el modelo BayesB se identificaron 26, sumando un total de 51 SNPs, de los cuales 11 fueron identificados en común. Los marcadores se encuentran distribuidos en 23 regiones denominadas Bins: 1.04, 1.05, 1.06, 1.1, 1.11, 2.03, 3.04, 4.03, 4.08, 4.09, 5.03, 5.04, 6.01, 6.04, 6.05, 6.07, 7.02, 7.03, 7.04, 8.02, 10.01, 10.04, 10.05. La proporción de variación fenotípica total explicada (PVE) por los 40 SNPs fue de 56%, atribuida a efectos genéticos aditivos. Los 11 SNPs (S1_293221122, S1_79329381, S2_19259181, S4_234954673, S6_111966489, S6_111966499, S6_111966511, S6_111966523, S6_164681025, S8_15101936 y S10_135336611) identificados de manera independiente por ambos modelos sugiere que tales marcadores están fuertemente asociados a la resistencia/susceptibilidad al complejo mancha de asfalto. La proporción de variación fenotípica total explicada por los 11 marcadores en conjunto fue de 22.24%.

De manera independiente, con el método single marker con enfoque de modelo mixto implementado con TASSEL v.5.2 se identificaron 25 SNPs asociados a la tolerancia genética a la mancha de asfalto con un nivel de significancia de $\alpha=0.0001$, equivalente a un valor de 4 en la escala de $-\log_{10}(p)$. Los marcadores asociados significativamente se encuentran distribuidos en todos los cromosomas, excepto el nueve. Los cuatro marcadores más significativos (valor de p más pequeño) fueron S2_19259181, S1_293221122, S6_111966489, S6_111966511, los cuales explican de manera individual 5.38, 5.07, 4.94 y 4.94 % de la varianza fenotípica, respectivamente. La proporción de PVE en conjunto por los 25 SNPs fue de 34.25%. La Figura 6 muestra un gráfico manhattan para visualizar de manera global la

distribución de los marcadores en los cromosomas y sus respectivos puntajes en una escala $-\log_{10}(p)$; en el Cuadro 4 se muestra la lista de marcadores asociados significativamente, indicando el cromosoma, la posición física, valor de p y PVE.

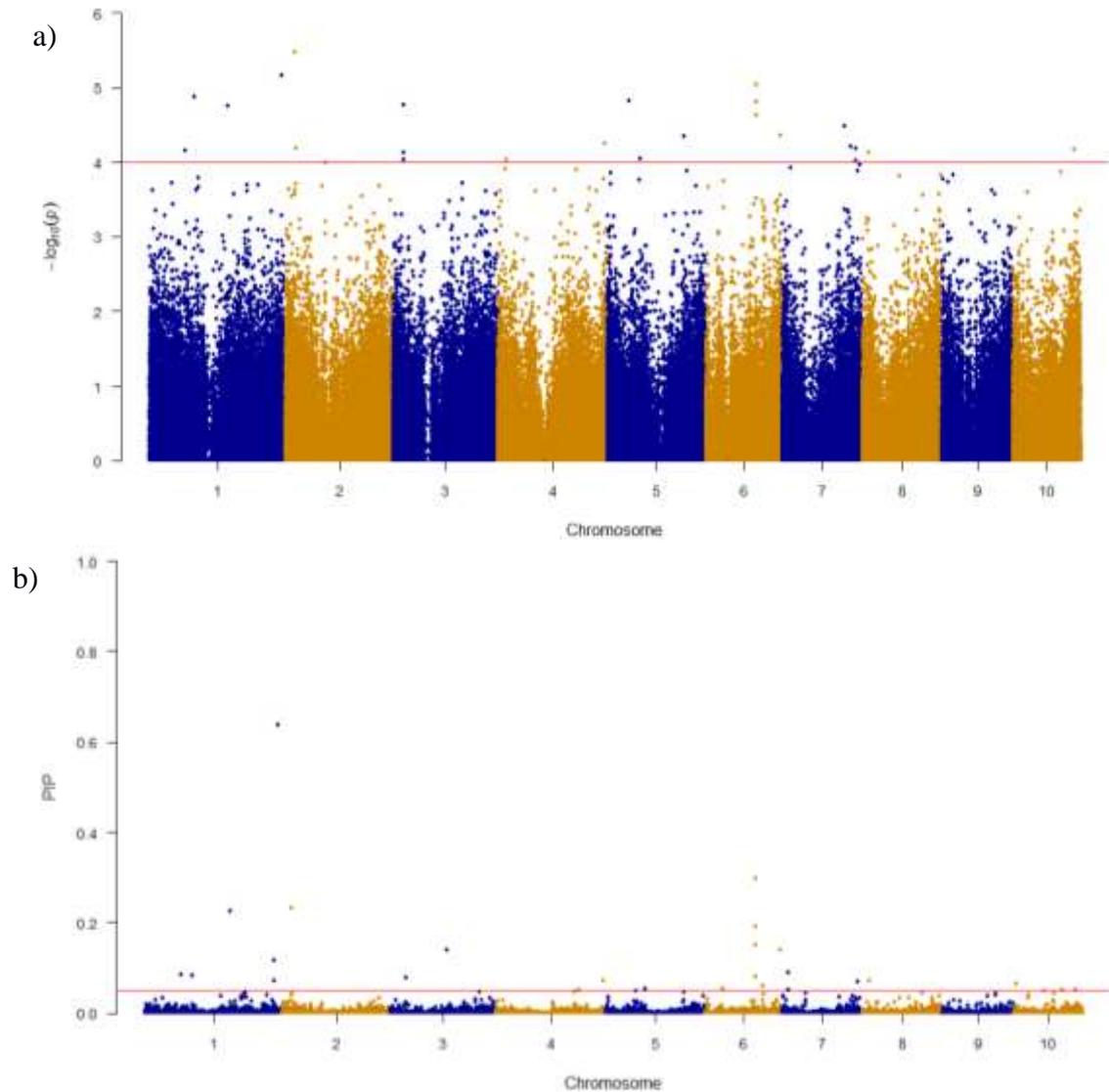


Figura 6. Gráficos tipo manhattan mostrando la distribución de los marcadores en los 10 cromosomas y su respectivo puntaje según el modelo y escala utilizada. Los colores diferencian a los cromosomas. La línea roja indica el umbral de significancia en ambos gráficos según modelo. a) Valores de probabilidad en escala $-\log_{10}(p)$ obtenidos con el modelo lineal mixto single marker y el umbral de significancia es $\alpha = 0.0001$ ($-\log_{10}(0.0001) = 4$). b) Valores PIP obtenidos con el modelo BayesB y el umbral de selección de marcadores es $PIP > 0.05$.

Por otra parte, con el modelo BayesB se logró identificar 26 marcadores asociados a la resistencia a la mancha de asfalto, los cuales superaron el umbral de $PIP > 0.05$. Al igual que en el método single marker, los marcadores asociados se encuentran distribuidos en todos los

cromosomas, excepto el nueve. El marcador con mayor probabilidad posterior de inclusión fue el S1_293221122, seguido por el S6_111966489, S2_19259181 y S1_187644290, con valores PIP de 0.638, 0.299, 0.233 y 0.225, respectivamente. Con el modelo BayesB, al ser una regresión simultánea sobre todos los marcadores, no es posible estimar la proporción de varianza fenotípica explicada por cada marcador puntual, por lo cual se presenta una estimación de la PVE por los 26 SNPs en conjunto igual a 52.26%, atribuida a efectos genéticos aditivos. La Figura 6 muestra el gráfico tipo manhattan para visualizar la distribución de los marcadores en los cromosomas y sus respectivos valores de probabilidad posterior de inclusión (PIP); el Cuadro 45 se muestra la lista de marcadores asociados significativamente, indicando el cromosoma, la posición física, valor PIP.

Respecto a la eficiencia del modelo lineal mixto sobre el control de falsos positivos, el gráfico tipo cuantil-cuantil (figura 7) nos muestra un buen ajuste de los valores de p observados sobre los valores teóricos esperados bajo una distribución normal, lo cual indica que se obtuvo un buen control de estructura de población y relaciones de parentesco y eficiente control de asociaciones espurias.

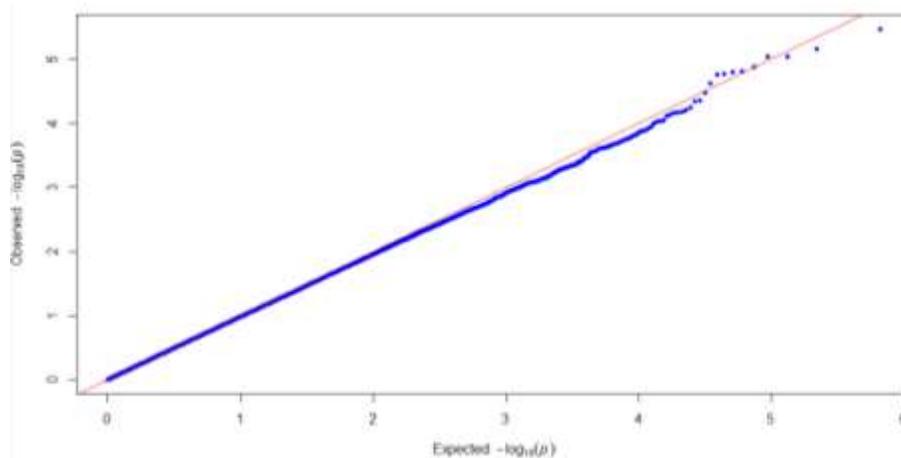


Figura 7. Gráfico cuantil-cuantil mostrando los valores observados de p del modelo lineal mixto single marker contra los valores teóricos esperados bajo una distribución normal en escala $-\log_{10}(p)$.

Cuadro 4. Marcadores SNP (25) asociados significativamente a la resistencia a mancha de asfalto identificados con el modelo lineal mixto single marker.

SNP ID	Cromosoma	Posición	Alelo menor	MAF	p	PVE
S2_19259181*	2	19259181	A	0.25378	3.35E-06	0.05377

S1_293221122*	1	293221122	C	0.17387	6.82E-06	0.05066
S6_111966489*	6	111966489	A	0.2635	9.04E-06	0.04942
S6_111966511*	6	111966511	C	0.2635	9.04E-06	0.04942
S1_98461847	1	98461847	G	0.31425	1.31E-05	0.04779
S5_48488908	5	48488908	C	0.41253	1.53E-05	0.04714
S6_111966523*	6	111966523	G	0.2635	1.57E-05	0.04702
S3_23495129	3	23495129	C	0.05616	1.69E-05	0.04669
S1_174002950	1	174002950	A	0.22246	1.76E-05	0.04651
S6_111966499*	6	111966499	A	0.2581	2.37E-05	0.04523
S7_137540103	7	137540103	A	0.13067	3.30E-05	0.04379
S6_164681025*	6	164681025	G	0.43305	4.44E-05	0.0425
S5_170062437	5	170062437	G	0.38661	4.53E-05	0.04241
S4_234954673*	4	234954673	G	0.12851	5.69E-05	0.04142
S7_151082785	7	151082785	G	0.28294	6.12E-05	0.04111
S7_160773966	7	160773966	C	0.05832	6.60E-05	0.04078
S2_21726221	2	21726221	C	0.14903	6.68E-05	0.04073
S10_135336611*	10	135336611	G	0.45032	6.79E-05	0.04066
S1_79329381*	1	79329381	C	0.28186	7.06E-05	0.04049
S8_15101936*	8	15101936	G	0.21058	7.37E-05	0.04031
S3_23495159	3	23495159	C	0.0594	7.56E-05	0.0402
S5_72696410	5	72696410	T	0.05724	8.90E-05	0.03949
S3_23495105	3	23495105	T	0.06479	9.22E-05	0.03934
S4_17333579	4	17333579	G	0.27862	9.34E-05	0.03928
S7_160773952	7	160773952	G	0.06911	9.45E-05	0.03923

PVE por los 25 SNPs en conjunto: 34.25%

PVE Proporción de la variación fenotípica total explicada por el marcador, obtenida a partir del r_{adj}^2 .

MAF Frecuencia del alelo menor

* SNPs identificados con ambos modelos.

Cuadro 5. Marcadores SNP (26) asociados significativamente a la resistencia a mancha de asfalto identificados con el modelo BayesB.

SNP ID	Cromosoma	Posición	Alelo menor	MAF	PIP
S1_293221122*	1	293221122	C	0.17387	0.638
S6_111966489*	6	111966489	A	0.2635	0.29875

S2_19259181*	2	19259181	A	0.25378	0.23275
S1_187644290	1	187644290	C	0.31857	0.22475
S6_111966511*	6	111966511	C	0.2635	0.191
S6_111966523*	6	111966523	G	0.2635	0.14975
S3_124181546	3	124181546	G	0.27862	0.14025
S6_164681025*	6	164681025	G	0.43305	0.13925
S1_282550335	1	282550335	T	0.42117	0.1165
S7_14658428	7	14658428	G	0.2095	0.0905
S1_79329381*	1	79329381	C	0.28186	0.08475
S1_103419465	1	103419465	A	0.4514	0.08275
S6_111966499*	6	111966499	A	0.2581	0.07975
S3_35126726	3	35126726	A	0.31317	0.07825
S1_282674577	1	282674577	C	0.35421	0.07175
S4_234954673*	4	234954673	G	0.12851	0.071
S8_15101936*	8	15101936	G	0.21058	0.0705
S7_166698140	7	166698140	T	0.09827	0.07025
S10_4820377	10	4820377	C	0.30346	0.06525
S6_127634563	6	127634563	A	0.45464	0.05975
S6_38117593	6	38117593	G	0.23218	0.05425
S5_86013493	5	86013493	C	0.43844	0.053
S4_182611991	4	182611991	G	0.35205	0.05175
S10_106065368	10	106065368	G	0.11447	0.05075
S7_14658316	7	14658316	G	0.1879	0.05075
S10_135336611*	10	135336611	G	0.45032	0.05075

PVE por los 25 SNPs en conjunto: 52.26%

PVE Proporción de la variación fenotípica total explicada por el marcador, obtenida a partir del r_{adj}^2 .

MAF Frecuencia del alelo menor

* SNPs identificados con ambos modelos.

PIP Probabilidad posterior de inclusión en el modelo

4.13 Identificación de genes candidatos

Un total de 68 genes candidatos fueron identificados entre los 40 SNPs asociados significativamente a la tolerancia a mancha de asfalto, de los cuales 61 genes codifican proteínas con función predicha conocida y 7 fueron proteínas hipotéticas no caracterizadas aun para maíz. La mayoría de los SNPs asociados significativamente fueron encontrados dentro de

la secuencia de los genes candidatos y otros de forma adyacente a los genes candidatos en una distancia menor de 10 Kbp. El número de genes candidatos asociados a cada SNP fue de uno hasta cuatro. Los 68 genes candidatos se encuentran distribuidos en las 23 regiones (bins) identificadas de la siguiente manera:

Cuadro 6. Distribución de genes candidatos en las regiones (bins) identificadas.

Región	No. Genes C.						
1.04	1	3.04	5	6.01	1	7.04	2
1.05	7	4.03	4	6.04	2	8.02	1
1.06	3	4.08	1	6.05	2	10.01	2
1.1	4	4.09	2	6.07	1	10.04	2
1.11	1	5.03	6	7.02	2	10.05	2
2.03	5	5.04	4	7.03	8		

Como puede observarse en el cuadro anterior, hay regiones que contienen solo un gen candidato y otras regiones que contienen hasta ocho genes asociados a mecanismos de resistencia, a las cuales debe prestársele especial atención.

La función predicha de los genes candidatos (Cuadro 7) indica que varios de ellos están asociados con resistencia a enfermedades e interactúan entre sí en diferentes mecanismos de defensa en plantas, entre ellos varios genes hipotéticos que codifican proteínas con función kinasa ampliamente conocidos por estar involucrados en resistencia a enfermedades, por ejemplo: GRMZM2G132607 (PFKB-like carbohydrate kinase family protein/ Ribokinase), GRMZM2G132763 (Leucine-rich repeat receptor-like protein kinase PEPR1), GRMZM2G072569 (Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1), GRMZM2G168917 (Putative protein kinase superfamily protein), GRMZM2G104384 (Leucine-rich repeat receptor-like protein kinase family protein), GRMZM5G802725 (Protein kinase superfamily protein, partial) y GRMZM2G163138 (Probable LRR receptor-like serine/threonine-protein kinase At1g51860).

También se identificaron genes que codifican proteínas con función predicha de factores de transcripción, entre ellos: GRMZM2G001930 (MYC transcription factor7), GRMZM2G080168 (BHLH141 - bHLH-transcription factor 141), GRMZM2G106798 (Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein), GRMZM2G092609 (Putative bZIP transcription factor superfamily protein),

AC207043.3_FG002 (VQ motif family protein; Transcriptional Regulators), GRMZM2G163054 (WRKY-transcription factor 50) y GRMZM2G144275 (BHLH transcription factor). Otros factores de transcripción importantes ampliamente conocidos por estar involucrados en mecanismos de defensa son los que codifican Leucine-zipper, entre los cuales se identificaron algunos genes candidatos: GRMZM2G351330 (Homeobox-leucine zipper protein HOX22) y GRMZM2G104204 (Homeobox-leucine zipper protein HOX14-like).

Cuadro 7. Genes candidatos y posible función asociados a los marcadores SNPs identificados en el análisis de asociación de genoma completo para tolerancia genética a mancha de asfalto.

SNP ID	Modelo	Gen candidato asociado	Bin*	Anotación: Posible función proteica
S1_79329381	Bayes B - Single Marker	GRMZM5G826389	1.04	SRG1 protein, c1945_1(263), leucoanthocyanidin dioxygenase
		GRMZM2G070271	1.05	Xyloglucan endotransglucosylase/hydrolase 27
S1_103419465	Bayes B	GRMZM5G855014	1.05	Uncharacterized protein.
		GRMZM2G022909	1.05	Uncharacterized protein. Análogo en arroz: heat shock protein putative expressed.
S1_174002950	Single Marker	GRMZM2G381822	1.05	Origin recognition complex subunit 3
		GRMZM2G083284	1.05	Hydrolase-like protein; Metallo-hydrolase/oxidoreductase superfamily protein
S1_98461847	Single Marker	GRMZM2G094390	1.05	ADP-ribosylation factor GTPase-activating protein AGD8
		GRMZM2G001930	1.05	MYC transcription factor7
S1_187644290	Bayes B	GRMZM2G132607	1.06	PFKB-like carbohydrate kinase family protein/ Ribokinase
		GRMZM2G132763	1.06	Leucine-rich repeat receptor-like protein kinase PEPR1
		GRMZM2G132748	1.06	Complex I subunit; pathways NADH dehydrogenase ubiquinon phosphorylation and dephosphorylation
S1_282550335	Bayes B	GRMZM2G072569	1.10	Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1
		GRMZM2G072806	1.10	NADH dehydrogenase [ubiquinone] iron-sulfur protein 8-B Mitochondrial; NADH-ubiquinone oxidoreductase 23 kDa subunit
		GRMZM2G072744	1.10	Malate dehydrogenase, glyoxysomal
S1_282674577	Bayes B	GRMZM2G122139	1.10	Cytosolic purine 5-nucleotidase
S1_293221122	Bayes B - Single Marker	GRMZM2G012280	1.11	Zinc knuckle (CCHC-type) family protein; Zinc finger
S2_19259181	Bayes B - Single Marker	GRMZM2G462261	2.03	Putative glucuronosyltransferase PGSIP7; Glycosyl transferase family 8
		GRMZM2G162333	2.03	Pectinesterase or Pectin methylesterase
		GRMZM2G462258	2.03	Inactive beta-amylase 9
S2_21726221	Single Marker	GRMZM2G351330	2.03	Homeobox-leucine zipper protein HOX22
		GRMZM2G168917	2.03	Putative protein kinase superfamily protein
S3_124181546	Bayes B	GRMZM2G063688	3.04	Galactosyltransferase family protein

S3_23495129,		GRMZM2G041352	3.04	Nuclease PA3
S3_23495159,	Single Marker	GRMZM2G121747	3.04	Protein ABA DEFICIENT 4 chloroplastic
S3_23495105		GRMZM2G345840	3.04	C2 domain-containing protein
S3_35126726	Bayes B	GRMZM2G428456	3.04	Uncharacterized protein
		GRMZM2G104384	4.03	Leucine-rich repeat receptor-like protein kinase family protein
		GRMZM2G080168	4.03	BHLH141 - bHLH-transcription factor 141
S4_17333579	Single Marker	GRMZM2G149108	4.03	Ubiquitin-like 1-activating enzyme E1A
		GRMZM2G106798	4.03	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein
S4_182611991	Bayes B	GRMZM2G439422	4.08	Peroxidase 65
S4_234954673	Bayes B - Single Marker	GRMZM2G143445	4.09	Putative autophagy domain family protein
		GRMZM2G143462	4.09	Ubiquitin receptor RAD23b
		AC225718.2_FG006	5.03	EF-hand Ca ²⁺ -binding protein CCD1
S5_48488908	Single Marker	GRMZM2G163544	5.03	Putative beta-glucosidase 41
		GRMZM2G463913	5.03	50S ribosomal protein L24 chloroplastic
		GRMZM2G042758	5.03	GDSL esterase/lipase
S5_72696410	Single Marker	GRMZM2G344212	5.03	Putative RING zinc finger domain superfamily protein
		GRMZM5G802725	5.03	Protein kinase superfamily protein, partial
S5_170062437	Single Marker	GRMZM2G092609	5.04	Putative bZIP transcription factor superfamily protein
		GRMZM2G071484	5.04	E3 ubiquitin-protein ligase PUB23
		AC207043.3_FG002	5.04	VQ motif family protein; Transcriptional Regulators
S5_86013493	Bayes B	GRMZM2G086243	5.04	Uncharacterized protein. Análogo en arroz: peptidyl-prolyl cis-trans isomerase putative expressed
S6_38117593	Bayes B	GRMZM2G055717	6.01	NADH dehydrogenase (ubiquinone)
S6_111966489,		GRMZM2G073700	6.04	RNA-binding region-containing protein 1
S6_111966499,	Bayes B - Single Marker			
S6_111966511,		GRMZM2G118657	6.04	UDP-glycosyltransferase 85A2
S6_111966523				
S6_127634563	Bayes B	GRMZM2G086761	6.05	Uncharacterized protein
		GRMZM2G163054	6.05	WRKY-transcription factor 50
S6_164681025	Bayes B - Single Marker	GRMZM2G022619	6.07	60S ribosomal protein L18a
S7_14658316,	Bayes B	GRMZM2G174625	7.02	Uncharacterized protein

S7_14658428	Bayes B	GRMZM5G830839	7.02	Subtilisin-like protease SBT4.3 y SBT4.4
		GRMZM2G330422	7.03	Análogo en arroz: calcium-transporting ATPase 9 plasma membrane-type putative expressed
S7_137540103	Single Marker	GRMZM2G065868	7.03	60S ribosomal protein L32
		GRMZM2G144275	7.03	BHLH transcription factor, DNA-binding domain superfamily protein
		GRMZM2G115757	7.03	Peptidyl-prolyl cis-trans isomerase; FK506-binding protein
		GRMZM2G356198	7.03	DSC E3 ubiquitin ligase complex subunit 1
S7_151082785	Single Marker	GRMZM2G163138	7.03	Probable LRR receptor-like serine/threonine-protein kinase At1g51860
		GRMZM2G163095	7.03	Probable ADP-ribosylation factor GTPase-activating protein AGD11
		GRMZM2G163045	7.03	Putative disease resistance RPP13-like protein 3; Disease resistance protein RPM1
S7_160773966, S7_160773952	Single Marker	GRMZM2G104204	7.04	Homeobox-leucine zipper protein HOX14-like
S7_166698140	Bayes B	GRMZM2G025281	7.04	Serine/arginine-rich splicing factor SC35
S8_15101936	Bayes B - Single Marker	GRMZM2G165692	8.02	Importin alpha-2 subunit
S10_4820377	Bayes B	GRMZM2G324231	10.01	Uncharacterized protein. Análogo en arroz: avr9/Cf-9 rapidly elicited protein putative expressed
		GRMZM2G314661	10.01	DDT domain-containing protein PTM
S10_106065368	Bayes B	GRMZM2G037335	10.04	Phospholipid-transporting ATPase 2
		GRMZM2G341741	10.04	Pentatricopeptide repeat-containing protein At2g13600
S10_135336611	Bayes B - Single Marker	GRMZM2G113800	10.05	Probable purine permease 11
		GRMZM2G113818	10.05	Probable purine permease 11

*Regiones del genoma: el mapa genético del maíz fue dividido en 100 segmentos llamados Bin en inglés.

También se identificaron algunos genes que codifican proteínas con función de factores de transcripción Zinc finger, tales como: GRMZM2G012280 (Zinc knuckle (CCHC-type) family protein; Zinc finger), el cual se encontró cercano al marcador S1_29322122 identificado en común por ambos modelos, y GRMZM2G344212 (Putative RING zinc finger domain superfamily protein), cercano al SNP S5_72696410 identificado con modelo single marker. Otro hallazgo importante es la identificación del gen candidato GRMZM2G163045 cuya función putativa es resistencia a enfermedades (RPP13-like protein 3; Disease resistance protein RPM1).

Otros genes menos comunes pero que recientemente han sido reportados que están involucrados e interactúan entre sí en mecanismos de defensa a enfermedades en plantas son GRMZM2G094390 (ADP-ribosylation factor GTPase-activating protein AGD8), GRMZM2G330422 (Análogo en arroz: calcium-transporting ATPase 9 plasma membrane-type putative expressed), GRMZM2G163095 (Probable ADP-ribosylation factor GTPase-activating protein AGD11) y GRMZM2G037335 (Phospholipid-transporting ATPase 2), GRMZM2G037335 (Phospholipid-transporting ATPase 2), GRMZM2G341741 (Pentatricopeptide repeat-containing protein At2g13600), GRMZM2G165692 (Importin alpha-2 subunit), GRMZM2G162333 (Pectinesterase or Pectin methylesterase), GRMZM2G115757 (Peptidyl-prolyl cis-trans isomerase; FK506-binding protein 2-1), GRMZM2G022619 (60S ribosomal protein L18a) y GRMZM2G042758 (GDSL esterase/lipase).

4.14 Evaluación preliminar en campo del nivel de tolerancia genética al CMA

Previo a realizar el ensayo de evaluación de materiales en campo, se hizo un incremento de semilla, dado que el total disponible en algunas líneas era limitado (50 granos o semillas). Para el incremento de las líneas endogámicas se utilizaron autofecundaciones y cuando se presentó asincronía floral se realizó la multiplicación por cruza fraternal. En la figura 8 puede observarse las fotografías del proceso de establecimiento del vivero en el municipio de Nueva Concepción, Escuintla.



Figura 8. Fotografías de las etapas del establecimiento de vivero, monitoreo del crecimiento, desarrollo vegetativo e identificación de parcela. Puede observarse en la fase de desarrollo vegetativo la variabilidad fenotípica entre los distintos materiales, particularmente altura de planta.

La figura 9 muestra las fases de cruzamientos y control de polinización, monitoreo en floración, cosecha del vivero y recolección e identificación de materiales. Es importante mencionar que en los cruzamientos se hicieron tanto autofecundaciones como polinizaciones cruzadas. Estas segundas cuando no se presentaba sincronía floral en la misma planta.



Figura 9. Fotografías fases finales del incremento de semilla en vivero.

La figura 10 muestra las fases iniciales del establecimiento del ensayo de evaluación de los genotipos a nivel de campo, en el municipio de Ixcán, Quiché. Estas fases fueron preparación del suelo, trazo del diseño experimental (alfa lattices) aplicación de lámina de riego y tratamiento de semillas



Figura 10. Fotografía de las fases iniciales del establecimiento del ensayo de evaluación de genotipos para tolerancia a mancha de asfalto, en el municipio de Ixcán, Quiché.

La figura 11 muestra las fases complementarias del establecimiento del ensayo de campo, así como las primeras etapas de desarrollo del cultivo en la evaluación de genotipos para medición de tolerancia genética a la enfermedad CMA.



Figura 11. Fotografías de las fases de desarrollo del cultivo y monitoreo de floración.

La figura 12 muestra las evaluaciones en post-floración del nivel de tolerancia al complejo mancha de asfalto en los distintos materiales establecidos en el ensayo. La medición se realizó de acuerdo a la escala de Ceballos y Deutsch (1992). No se realizó la evaluación del nivel de tolerancia al CMA en pre-floración porque en esa etapa del desarrollo del cultivo, aún no había presencia de la enfermedad.



Figura 12. Fotografía de las distintas expresiones del CMA en los genotipos de maíz evaluados en el ensayo de campo en post-floración.

Puede observarse en la figura 12 la presencia del complejo mancha de asfalto en las plantas; sin embargo, es necesario mencionar que aun y cuando la enfermedad se presentó, su expresión no estaba definida claramente, ya que todos los materiales se vieron influenciados por efectos de lluvias torrenciales e inundaciones en la zona del ensayo. Sin embargo, con estas consideraciones tomadas en cuenta se procedió con la evaluación, haciendo la salvedad al lector e investigador que los datos recolectados son datos preliminares y es necesario complementarlo con más evaluaciones a través de diferentes localidades y épocas de siembra.

La figura 13 muestra los resultados de la toma de datos de la repetición 1 del ensayo, se hizo la lectura de 6 plantas y se calculó el promedio. Se observa que la mayoría de materiales están

dentro de un nivel de tolerancia de acuerdo a la escala utilizada, destacando particularmente las líneas CML451, CML500, CML553, ICTAHB15 e ICTAB7.

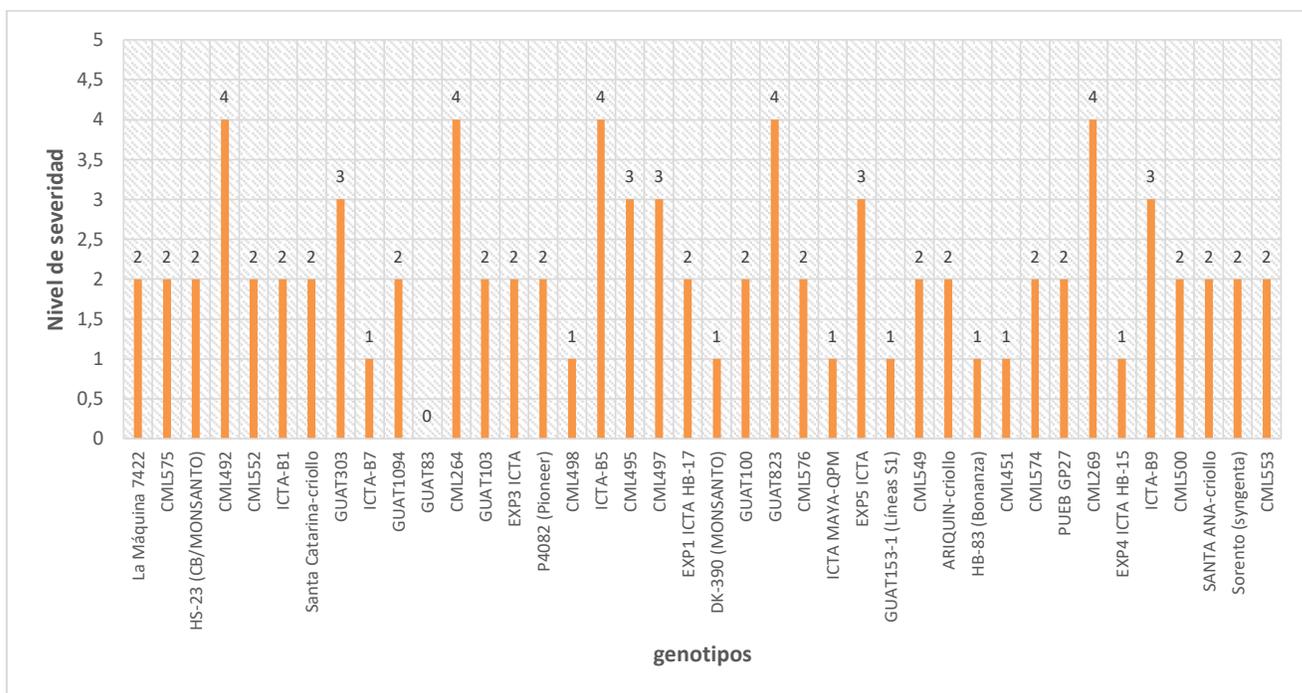


Figura 13. Expresión de tolerancia al CMA en la repetición 1 del ensayo a nivel de campo.

La figura 14 muestra los resultados de la toma de datos de la repetición 2 del ensayo. Se hizo la lectura de 6 plantas y se tomó el promedio, puede observarse que la mayoría de materiales están dentro de un nivel de tolerancia de acuerdo a la escala utilizada, destacando particularmente las líneas CML451, CML500, CML553, ICTAHB15 e ICTAB7.

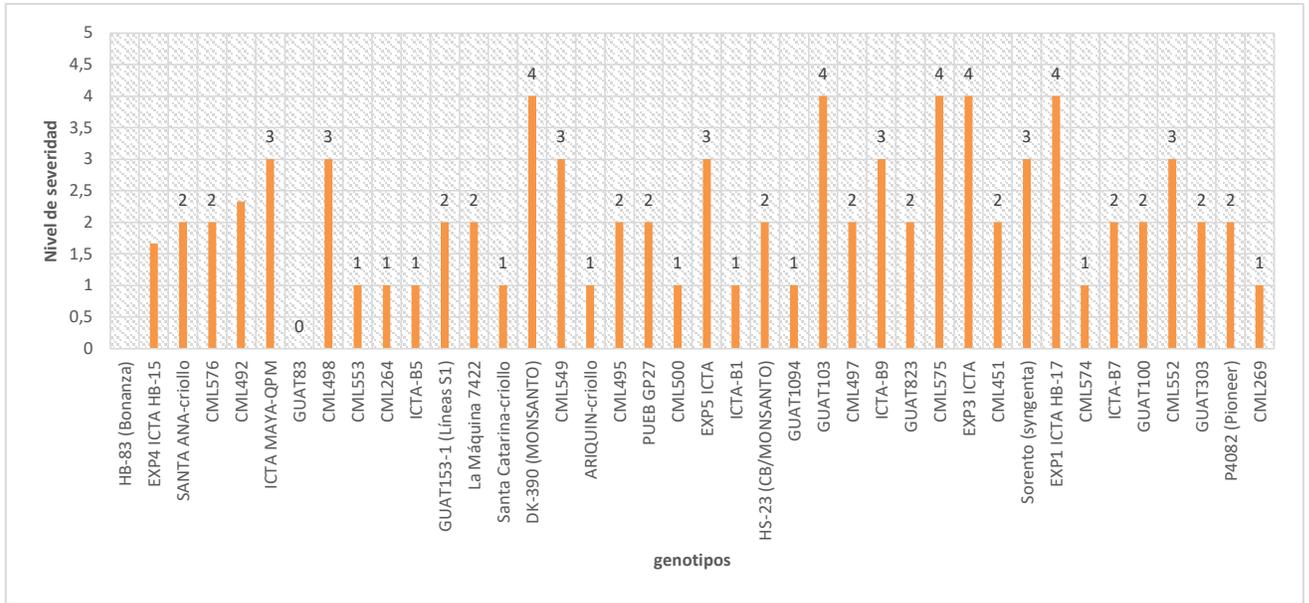


Figura 14. Expresión de tolerancia al CMA en la segunda repetición del ensayo a nivel de campo

La figura 15, muestra el promedio para cada material de su tolerancia al CMA, tomando en cuenta los datos de las dos repeticiones. Sobresalen las líneas CML451, CML500, CML553, ICTAHB15 e ICTAB7 como las más tolerantes al complejo mancha de asfalto.

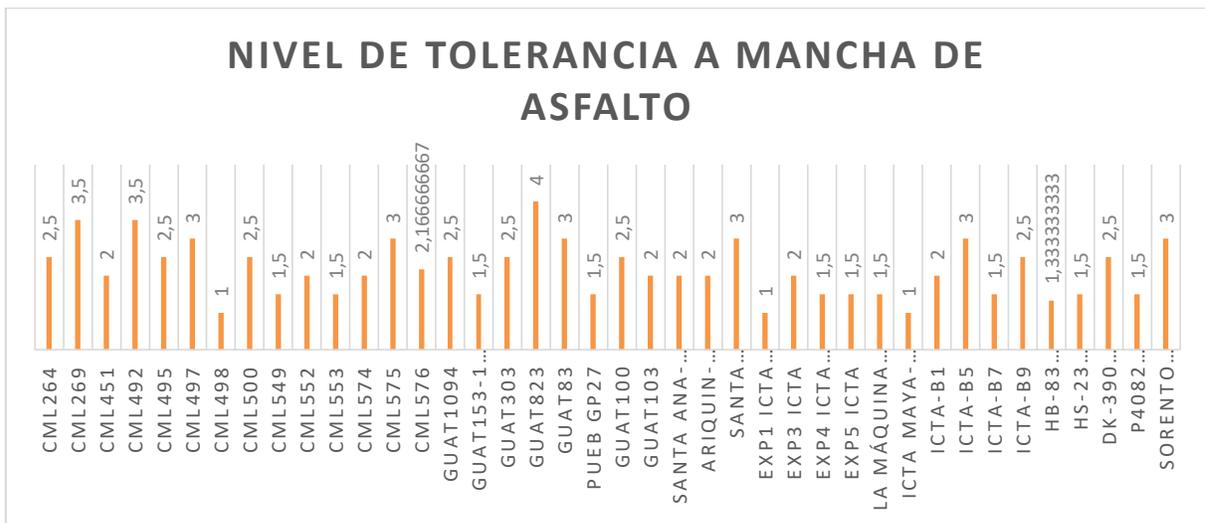


Figura 15. Resumen de la expresión de tolerancia al CMA tomando en cuenta el promedio de las dos repeticiones del ensayo a nivel de campo.

5 Discusión

5.9 Evaluación fenotípica, heredabilidad y donadores

La distribución continua del nivel de tolerancia a mancha de asfalto resalta el hecho que se trata de un carácter cuantitativo y de herencia poligénica (Shi et al., 2014), indicando que varios o muchos genes pueden estar involucrados. El valor de heredabilidad en sentido estricto estimado (0.8020) es cercano a los valores obtenidos por Ceballos & Deutsch (1992) (0.83 y 0.94) y está en el rango de los valores de heredabilidad en sentido amplio (incluyendo efectos no aditivos) obtenidos en otros estudios utilizando líneas endogámicas (0.54 a 0.96) (Cao et al., 2017; Mahuku et al., 2016), lo cual sugiere un predominante control genético de la tolerancia a la mancha de asfalto y una alta repetibilidad de los datos. Lo anterior confirma por una parte la presencia de un fuerte componente aditivo dado por la alta heredabilidad en sentido estricto congruente con los resultados de Hernández (2014); en segundo lugar, el valor positivo de la curtosis (0.157) sugiere que probablemente estén involucrados algunos genes de efecto mayor tal como lo refleja el estudio de Ceballos & Deutsch (1992). La alta heredabilidad observada también refleja la alta calidad de los datos fenotípicos de la base de datos, lo cual es un factor que favorece el poder de detección de QTL (Gowda et al., 2015) e indica que los datos son confiables para análisis genéticos. Los efectos aditivos además pueden ser acumulados mediante selección recurrente en poblaciones nativas.

Los mestizos derivados de las poblaciones nativas CHIS474, OAXA280, GUAT153, HAIT13 y TABA26 fueron los que mostraron mayores niveles de tolerancia, por lo que estas poblaciones pueden ser utilizadas como donadores de tolerancia genética al complejo mancha de asfalto mediante la derivación de líneas tolerantes e introgresión a materiales élite como parte de un programa de premejoramiento.

5.10 Identificación de marcadores SNP asociados a tolerancia a mancha de asfalto

Los marcadores y regiones identificadas en este estudio coinciden con los reportados en recientes estudios de asociación utilizando como germoplasma líneas endogámicas y poblaciones de mapeo de ligamiento. Estos estudios reportan asociación en el cromosoma 8 (bin 8.3), cromosomas 2 (bins 2.05 y 2.07), 3 (bins 3.04 y 3.09) y 7 (bin 7.02) (Cao et al., 2017; Li et al., 2016). Además, en este estudio se identificaron muchas regiones más asociadas a la tolerancia a mancha de asfalto. Esto se debe en parte a la gran diversidad genética existente en

poblaciones nativas, así como el análisis GWAS como poderosa herramienta para detectar alelos comunes de efecto menor, lo cual permitió la identificación de nuevas regiones con potencial para derivar resistencia a mancha de asfalto. Subsecuentes estudios deben realizarse para desarrollo de poblaciones biparentales a partir de las poblaciones nativas indicadas como potenciales donadores de resistencia.

En cuanto a los modelos empleados, cabe resaltar que los marcadores identificados con el modelo BayesB explican una mayor proporción de variación fenotípica, sin embargo, con ambos se identificaron genes muy importantes relacionados con recepción y reconocimiento de señal y mecanismos de defensa contra patógenos en plantas, por lo que la combinación de ambos modelos puede ser favorable en la identificación de QTL.

5.11 Identificación de genes candidatos

La mayoría de genes candidatos identificados producen proteínas ampliamente conocidas por estar asociadas a procesos enzimáticos de transporte, factores de transcripción que regula positivamente la expresión de genes de resistencia, así como receptores de señal involucrados en mecanismos de resistencia de enfermedades causadas por hongos en maíz, tales como, mancha gris de la hoja (*Cercospora zeaе maydis*), tizón norteño y sureño (*Setosphaeria turcica* y *Bipolaris maydis*) y carbón de la espiga (Kump et al., 2011; Poland, Bradbury, Buckler, & Nelson, 2011; Shi et al., 2014; M. Wang et al., 2012).

Dado que el nivel de expresión de los genes que codifican proteínas de resistencia es crítico para inhibir el desarrollo de la enfermedad, los factores de transcripción son de suma importancia, ya que al regular positivamente la expresión de genes de resistencia se limita e inhibe el desarrollo de la enfermedad (Alves et al., 2014). En este estudio se identificaron varios factores de transcripción, principalmente el factor WRKY, el cual ha sido encontrado regulando positivamente la expresión de genes de resistencia a enfermedades (Eulgem, Rushton, Robatzek, & Somssich, 2000; Michelmore et al., 2013; Mohr et al., 2010) y el factor MYC relacionado con la traducción de señal MAPK (del inglés, Mitogen-activated protein kinase) (Tena, Asai, Chiu, & Sheen, 2001). El factor de transcripción BHLH (del inglés, basic helix-loop-helix) es una proteína de la super familia de factores de transcripción que puede unirse a sitios específicos de ADN, actuando como un importante regulador de diferentes procesos

biológicos junto con el factor de transcripción MYC, incluyendo respuesta de defensa en plantas (Lorenzo, Chico, Sánchez-Serrano, & Solano, 2004; Pireyre & Burow, 2015; Toledo-Ortiz, Huq, & Quail, 2003; J. Wang et al., 2015). El factor de transcripción básico Leucine-zipper (bZIP) con dominios de unión a ADN ha sido identificado como responsable de desencadenar una respuesta de defensa sistémica en la planta para una gran variedad de patógenos (Alves et al., 2013; Després, DeLong, Glaze, Liu, & Fobert, 2000), así como en la producción de ROS (del inglés, reactive oxygen species) junto con genes con función enzimática peroxidasa en las paredes celulares, donde el ROS actúa como mecanismo de señalización en respuesta a patógenos (Bindschedler et al., 2006; Kaminaka et al., 2006, p. 10; Torres, Jones, & Dangl, 2006). Los factores de transcripción zinc finger identificados en este análisis, también se encuentran involucrados en resistencia de plantas a enfermedades (Alves et al., 2014; S. K. Gupta, Rai, Kanwar, & Sharma, 2012). El factor de transcripción VQ motif regula diferentes procesos de desarrollo, y cada vez hay más evidencia del rol que juega en la producción de ácido salicílico y ácido jasmónico en respuesta de defensa en plantas (Jing & Lin, 2015).

Varios genes identificados en este estudio codifican proteínas de la familia kinasa, especialmente los LRR-RPK (del inglés, Leucine-rich repeat receptor-like kinase) que actúan específicamente como receptores de señal para detección de patógenos. Los dominios LRR están implicados en resistencia a enfermedades en plantas y respuesta a estrés, por lo que son considerados genes de resistencia (Afzal, Wood, & Lightfoot, 2008; Ng & Xavier, 2011; Romeis, 2001; Torii, 2004; Zan et al., 2013). Estos genes también fueron identificados en dos estudios de mancha de asfalto en líneas endogámicas (Cao et al., 2017; Mahuku et al., 2016).

Otros genes identificados producen proteínas que recientemente han sido relacionadas a respuesta de defensa en plantas, tales como ubiquitins, que actúa como receptor de señal de defensa (Afzal et al., 2008; Michelmore et al., 2013); importin alpha subunit juega un papel importante en señalización de defensa en plantas (N. Kim et al., 2016; Wirthmueller, Roth, Banfield, & Wiermer, 2013); Serine/arginine-rich splicing factor se ha asociado en mecanismos de defensa en plantas mediante la aceleración de muerte celular e interacción con otros genes de defensa (Kumar & Kirti, 2012); GDSL esterase/lipase regula la resistencia sistémicamente asociada con la señalización de etileno (Kwon et al., 2009); Squamosa promoter-binding-like

proteína codifica factores de transcripción específicos en plantas involucrados en procesos de transducción de señal (H.-X. Zhang et al., 2016); pentatricopeptide repeats (PPRs, por sus siglas en inglés) y peroxidases han sido considerados recientemente como genes de resistencia análogos en plantas (Sekhwal et al., 2015); glucosyltransferase ha sido reportado en *Arabidopsis* en señalización de inducción de ácido salicílico y ácido jasmónico en respuesta a enfermedades (Blanco-Herrera et al., 2015); autophagy domain family protein, juega un papel importante en la defensa contra patógenos y productos tóxicos (S.-H. Kim, Kwon, Lee, & Chung, 2012); RNA binding proteínas involucradas en el control de infección por virus, bacterias y hongos (Huh & Paek, 2013; Musidlak, Buchwald, & Nawrot, 2014; Woloshen, Huang, & Li, 2011). Llama la atención que esta última proteína codificada por el gen GRMZM2G073700 está cercano a cuatro SNPs ubicados en el cromosoma 6, bin 6.04 (S6_111966489, S6_111966499, S6_111966511, S6_111966523), los cuales fueron identificados con ambos modelos. Sería interesante realizar un mapeo más fino y realizar análisis de haplotipos.

Los genes cuyas proteínas no estaban caracterizadas, posiblemente se deba a que no están contenidas en el genoma de referencia B73, ya que esta línea corresponde a germoplasma de zona templada, mientras que las poblaciones nativas evaluadas en este estudio están adaptadas a zonas tropicales.

En cuanto a los modelos empleados, con ambos se identificaron genes muy importantes relacionados con recepción y reconocimiento de señal y mecanismos de defensa contra patógenos en plantas, por lo que la combinación de ambos modelos puede ser favorable en la identificación de QTL.

La participación de múltiples genes con funciones distintas pero asociadas a respuesta de defensa en plantas confirma que se trata de un carácter de herencia poligénica. Hace falta investigar más sobre el tipo de interacción que puede ocurrir entre los genes involucrados y para un mejor entendimiento desde el punto de vista biológico y bioquímico. Los genes candidatos identificados en este estudio puede servir de base para subsecuentes investigaciones, en las cuales se realice un mapeo fino de los genes más importantes y desarrollar marcadores

funcionales para el mejoramiento asistido por marcadores moleculares para la selección de líneas e introgresión en materiales élite en el marco de un programa de mejoramiento.

5.12 Evaluación preliminar en campo del nivel de tolerancia genética al Complejo Mancha de Asfalto

El uso de distintos materiales en este estudio pretendía aprovechar la diversidad genética para ampliar el rango de tolerancia genética al complejo mancha de asfalto. Esto en acuerdo con Strange & Scott (2005) que proponen la variabilidad genética como una estrategia para restringir la incidencia de enfermedades en las plantas. En ese sentido las variedades nativas constituyen un importante recurso portador de genes con resistencia, para ampliar la base genética de los cultivos. Lo anterior en concordancia con los planteamientos de Pritch (2001) que considera la búsqueda de variabilidad genética como un objetivo de importancia en los programas de pre-mejoramiento. Los resultados en campo mostraron que la mayoría de los materiales tienen cierto nivel de tolerancia, destacando las líneas CML451, CML500, CML553, ICTAHB15 e ICTAB7.

Inicialmente previo al establecimiento del ensayo la información de los materiales indicaba que todos ellos habían manifestado cierto nivel de tolerancia al complejo CMA. (CIMMYT 2013 y comunicación personal con donadores del germoplasma utilizado) Este nivel de tolerancia general se presentó también en los genotipos evaluados en este estudio. Sin embargo, es necesario acotar que los datos tomados en esta investigación se hicieron bajo condiciones de expresión de la enfermedad, pero también en condiciones de los materiales afectados por lluvias torrenciales e inundaciones, por lo que los resultados deben tomarse con cierta reserva.

Por otra parte debe considerarse que los materiales evaluados en esta investigación se hicieron en condiciones climáticas y edafológicas diferentes a las condiciones en donde se reporta que inicialmente se hicieron evaluaciones de los mismos. Este componente también puede tener influencia en los resultados presentados en este estudio.

La tolerancia genética que los materiales presentan al CMA, pueden ser de carácter aditivo, tomando en cuenta los distintos rangos de tolerancia presentada. Esto estaría de acuerdo con los

datos indicados por Hernández (2014) cuyo estudio confirmó que los efectos génicos de dominancia y aditividad están presentes en la tolerancia genética a la enfermedad mancha de asfalto pero que los genes aditivos son los de mayor importancia.

6 Impacto esperado

Desde el punto de vista técnico-científico, es el primer proyecto de investigación biotecnológico en Guatemala donde se utilizarán marcadores moleculares de última generación que identifican polimorfismos de nucleótido simple (SNP), generados con la tecnología de última generación denominada genotipificación por secuenciación (GBS). El gran número de variedades nativas a analizar (463) constituye un proyecto de investigación robusto. Las técnicas de análisis, métodos y software a utilizar son herramientas bioinformáticas y bioestadísticas modernas que permitirá obtener mayor precisión en la detección de marcadores moleculares asociados a la tolerancia genética al CMA, los cuales servirán de base para subsecuentes investigaciones orientadas al desarrollo de marcadores funcionales que puedan ser utilizados en el mejoramiento genético asistido por marcadores para el desarrollo de nuevos materiales de maíz con resistencia durable al CMA. Con la evaluación de campo se espera identificar variedades nativas y CMLs tolerantes al CMA que puedan ser utilizados para el desarrollo de materiales tolerantes genéticamente adaptados a las condiciones de la región de Ixcán en el mediano y corto plazo. Considerando que la mayoría de variedades nativas son originarias de México, Brasil, Guatemala, y países de Centro y Sudamérica, se espera que los resultados de los análisis bioinformáticos y bioestadísticos trasciendan las fronteras, y pueden ser utilizados por otros países donde la enfermedad tiene importancia económica.

7 Conclusiones

La tolerancia al complejo mancha de asfalto es en carácter cuantitativo controlado principalmente por efectos genéticos por la alta heredabilidad observada ($\hat{h}^2 = 0.8$), cuya herencia poligénica esta mediada por muchos *loci* de efecto pequeño (efectos aditivos) y algunos pocos *loci* de efecto grande o medio (efectos dominante y epistáticos).

Las poblaciones nativas CHIS474, OAXA280, GUAT153, HAIT13 y TABA26 de la base de datos fenotípicas fueron los que mostraron mayores niveles de tolerancia en el análisis

descriptivo fenotípico, por lo que estas poblaciones pueden ser utilizadas como donadores de resistencia genética al complejo mancha de asfalto.

Un total de 40 marcadores SNPs fueron asociados significativamente a la tolerancia genética a mancha de asfalto mediante el estudio de asociación del genoma completo. La proporción de variación fenotípica total explicada (PVE) por los 40 SNPs fue de 56%, atribuida a efectos genéticos aditivos.

El análisis del gen candidato nos muestra que la tolerancia al complejo mancha de asfalto están involucrados múltiples genes con diferentes funciones relacionados a respuesta de defensa contra enfermedades que interactúan entre sí, principalmente factores de transcripción, receptores y transductores de señal, desencadenadores de respuesta de defensa sistémica, respuesta a estrés y genes cuya función proteica ha sido asociada recientemente a mecanismos de defensa en plantas.

Los materiales genéticos evaluados preliminarmente a nivel de campo presentaron en general cierto nivel de tolerancia al CMA, sin embargo, debido a los efectos de lluvias torrenciales e inundaciones que afectaron el desarrollo de la enfermedad, los resultados del ensayo de campo deben tomarse con reserva.

8 Vinculación

La investigación se realizó en colaboración con el Centro Internacional de Mejoramiento de Maíz y Trigo, quienes contribuyeron con el aporte de semilla de diverso germoplasma de maíz y líneas endogámicas para la evaluación de campo y asesoría para realización de análisis y discusión de los resultados de la investigación. También se realizaron alianzas estratégicas con la municipalidad de Ixcán y se involucraron productores de maíz de la región de Playa Grande Ixcán que colaboraron con el desarrollo de la fase de evaluación de campo.

9 Agradecimientos

Se agradece la cofinanciación de la Dirección General de Investigación (DIGI) dentro del Programa Universitario de Investigación en Alimentación y Nutrición (partida presupuestal 4.8.63.7.02).

10 Referencias

- 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.
- Barreto, H., Edmeades, G. & Crossa, J. (1997). The Alpha Lattice Design in Plant Breeding and Agronomy: Generation and analysis. México, DF. Centro Internacional de Mejoramiento de Maíz y Trigo -CIMMYT-.
- Berg, I. V., Fritz, S., & Boichard, D. (2013). QTL fine mapping with Bayes C(π): a simulation study. *Genetics Selection Evolution*, 45(1), 1-11.
- 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.
- Ceballos, H., & Deutsch, J. A. (1992). Inheritance of resistance to tar spot complex in maize. *Phytopathology*, 82(5), 505–512.
- Centro Internacional de Mejoramiento de Maíz y Trigo. (2013). *Complejo mancha de asfalto del maíz: Hechos y acciones*. Texcoco, México. Recuperado de http://conservacion.Cimmyt.org/index.php/es/component/docman/doc_view/922-tarspot
- Collard, B. C., Jahufer, M. Z., Brouwer, J. B., & Pang, E. C. (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.
- 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.
- 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.
- Hearne, Sarah; Chen, Charles; Buckler, Ed; Mitchell, Sharon; Romero, Alberto; Swarts, Kelly & Li, Huihui. (2014). *Seeds of discovery: Imputed GbS derived SNPs for maize landrace accessions represented in the SeeD-maize GWAS panel: Imputation using Beagle v.4*.

- México. International Maize and Wheat Improvement Center. Recuperado de <http://hdl.handle.net/11529/10035>
- Hernández, R. L. (2014). Genética de la resistencia al complejo mancha de asfalto en 18 genotipos tropicales de maíz. *Revista Fitotecnia Mexicana*, 38(1), 39-47.
- Hock, J., Dittrich, U., Renfro, B. L., & Kranz, J. (1992). Sequential development of pathogens in the maize tar spot disease complex. *Mycopathologia*, 117(3), 157–161.
- Hock, J., Kranz, J., & Renfro, B. L. (1989). El "complejo mancha de asfalto" de maíz, su distribución geográfica, requisitos ambientales e importancia económica en México. *Revista Mexicana de Fitopatología*, 7, 129-135.
- ICTA-Red SICTA-IICA-Cooperación Suiza. (2011). *Reconocimiento en campo de la mancha de asfalto*. Recuperado de <http://www.icta.gob.gt/maiz/Reconocimiento%20en%20campo%20de%20la%20mancha%20de%20asfalto%20en%20MAIZ.pdf>
- Instituto Nacional de Estadística. (2013). *Proyección de población por municipio*. Guatemala, Guatemala. Recuperado de <http://www.ine.gob.gt/np/poblacion/index.htm>.
- Larios, R. (30 de mayo de 2012). Se reportan pérdidas del 60% en cultivos de maíz. *Prensa Libre*, pp. 30-31.
- Lawrence, CJ, Dong, Q, Polacco, ML, Seigfried, TE, and Brendel, V. (2004). MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Research* 32:D393-D397.
- 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.
- Malaguti, G., & Subero, L. J. (1972). Tar spot of maize. *Agronomía Tropical*, 22(4), 443–445.
- McClintock, B., Yamakake, T. A., & Blumenschein, A. (1981). Chromosome constitution of races of maize: Its significance in the interpretation of relationships between races and varieties in the Americas. Chapingo, México, Colegio de Postgraduados.
- 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.

- Paliwal, R. L., Granados, G., Lafitte, H. R., Violic, A. D., & Marathée, J.-P. (2001). *El maíz en los trópicos: Mejoramiento y producción*. Roma, Italia: Food y Agriculture Organization of the United Nations.
- Pereyda-Hernández, J., Hernández-Morales, J., Sandoval-Islas, J. S., Aranda-Ocampo, S., León, C., & 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.
- 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: Precisur.
- Sharma, S., Upadhyaya, H. D., Varshney, R. K., & Gowda, C. L. (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. *Frontiers in Plant Science*, 4, 309.
- Strange, R. N., & Scott, P. R. (2005). Plant disease: A threat to global food security. *Annual Review of Phytopathology*, 43, 83-116.
- Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., ...Buckler, E. S. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, 43(2), 159-162.
- R-Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing. Recuperado de <https://www.r-project.org/>.
- Varón, F., & Sarria, G. (2007). *Enfermedades del maíz y su manejo*. Palmira, Colombia: Produmedios.
- 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.
- Weng, J., Liu, X., Wang, Z., Wang, J., Zhang, L., Hao, Z., ...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.
- Willcox, Martha; Burgueño, Juan; Mahuku, George; Guadarrama Espinoza, Armando; Chepetla Calderon, Daniel; Lopez, Manuel; Gonzalez, Jesus; Leyva, Mayolo; San Vicente, Felix; Maize Germplasm Bank; Dupont Pioneer. (2014). *Seeds of discovery: Evaluation of Tar Spot Resistance in 2011 in Guadalupe-Victoria, Chiapas. México*. International Maize and Wheat Improvement Center. Recuperado de <http://hdl.handle.net/11529/10031>
- White, D. G. (1999). *Compendium of corn diseases*. (3rd ed). Saint Paul, Minnesota: APS Press.
- Yi, N., & Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping. *Genetics*, 179(2), 1045-1055.
- 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.

11 Apéndice

Escala de severidad del complejo mancha de asfalto en maíz con base en la propuesta de (Ceballos & 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 %