

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

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

David O. González-Diéguez*, Carlos A. Orozco-Castillo

Facultad de Agronomía, Universidad de San Carlos de Guatemala

*Autor al que se dirige la correspondencia: dogrdc@gmail.com

Recibido: 20 de septiembre de 2019 / Fecha de 1era revisión: 24 de octubre de 2019 /
Aceptado: 03 de marzo de 2020 / Disponible en línea: 10 de noviembre de 2020

Resumen

El complejo mancha de asfalto (CMA) en maíz (*Zea mays* L.), causado por los hongos *Phyllachora maydis* Maubl. y *Monographella maydis* Müller & Samuels, es una enfermedad de importancia económica en Guatemala, que ha causado pérdida en el rendimiento entre 30 a 50%, inclusive del 100% si las condiciones son favorables. El objetivo de esta investigación fue identificar marcadores de un solo nucleótido o SNP (Single Nucleotide Polymorphism, por sus siglas en inglés) y genes candidatos asociados a la tolerancia genética al CMA. Para ello se analizaron 463 poblaciones nativas y 329,692 SNP, y se compararon dos modelos genómicos, single marker y BayesB, para la identificación de regiones asociadas a la tolerancia genética al CMA. Se identificaron 40 marcadores SNP asociados significativamente a la tolerancia genética al CMA 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 a enfermedades fueron identificados en las regiones señaladas por los marcadores SNP. Sus funciones principales son receptores y transductores de señal, factores de transcripción que regulan positivamente la expresión de genes de tolerancia y genes de la familia kinasa, por lo que potencialmente están involucrados en el mecanismo de defensa al CMA.

Palabras claves: modelos genómicos, estudio de asociación del genoma completo, genotipificación por secuenciación, resistencia genética, Loci de un carácter cuantitativo (QTL).

Abstract

The tar spot complex (TSC) disease in maize (*Zea mays* L.), caused by the fungi *Phyllachora maydis* Maubl. and *Monographella maydis* Müller & Samuels, is an economic important disease in Guatemala, producing yield losses between 30 to 50%, inclusive of 100% if the conditions are favorable. The objective of this research was to identify single nucleotide polymorphism markers (SNP) and candidate genes associated with genetic tolerance to TSC. A set of 463 native populations and 329,692 SNP were analyzed with two genomic models, single marker and BayesB, for the identification of regions associated with genetic tolerance to TSC. Forty 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 candidate genes for disease resistance were identified in the regions indicated by the SNP markers. Their main functions are signal transducers and receptors, transcription factors that positively regulate the expression of tolerance genes and family kinase genes, therefore, they are potentially involved in the defense mechanism to TSC.

Keywords: genomic models; genome-wide association study; genotyping by sequencing; genetic resistance; quantitative trait loci (QTL).



Introducción

La enfermedad complejo mancha de asfalto (CMA) en maíz (*Zea mays* L.), inducida por los hongos *Phyllachora maydis* Maubl., *Monographella maydis* Müller & Samuels y *Coniothyrium phyllachorae* Maubl., ha cobrado gran importancia en la última década por su amplia distribución y severidad debida a la interacción sinérgica de los patógenos involucrados. Ha sido reportada en la mayoría de países de Latino América (Hock et al., 1992; Mahuku et al., 2013), y en el 2015, *P. maydis* fue reportado en los Estados Unidos de América (Loladze, 2019). Los síntomas comienzan con la infección de *P. maydis*, seguido por la invasión de *M. maydis*, lo cual resulta en una interacción sinérgica que provoca la necrosis del tejido foliar (Hock et al., 1992; Pereyda-Hernández et al., 2009). La reducción en el rendimiento de grano y forraje varía entre 30 y 50%, dependiendo de la susceptibilidad del genotipo, si la enfermedad ataca antes o después de la floración y si las condiciones ambientales son favorables para producir una epidemia (Hock et al., 1989).

En Guatemala el primer reporte del CMA data del 1974, sin causar pérdida económica al cultivo (Monterroso et al., 1974). Fue hasta el 2007 cuando se reportó la presencia del CMA en el municipio de Ixcán, Quiché, con pérdidas de producción de grano de alrededor del 80%. En años subsiguientes, la enfermedad fue reportada en los departamentos del Petén, la región del Valle del Polochic en Alta Verapaz, en Jalapa e Izabal (Red SICTA/IICA/Cooperación Suiza, 2011). Como referencia, bajo el supuesto de que *M. maydis* llegue a los Estados Unidos, la pérdida potencial por cada 1% de reducción en producción se estiman en 1.5 millones de Tm de grano y US\$ 231.6 millones (Mottaleb et al., 2019).

El desarrollo de materiales genéticamente tolerantes a enfermedades es considerado la mejor estrategia para el control de enfermedades (Crute & Pink, 1996). Para alcanzar una tolerancia más durable es importante ampliar la base genética de los cultivares (Strange & Scott, 2005), lo cual es posible para el caso del CMA considerando la evidencia aportada por los estudios realizados por Ceballos y Deutsch (1992) y Hernández-Ramos y colaboradores (2015), 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 mediante selección.

Con el avance de la biotecnología y el desarrollo

de plataformas de genotipificación de alto rendimiento es posible rastrear regiones del genoma asociadas a determinados rasgos observables o fenotipos mediante estudios de asociación del genoma completo o genome-wide association studies (GWAS, por sus siglas en inglés). Elshire y colaboradores (2011) desarrollaron el método de Genotipificación por Secuenciación (GBS, por sus siglas en inglés), el cual es robusto y simple para especies de plantas con alta diversidad, capaz de analizar un gran número de muestras por ensayo y descubrir millones de marcadores de un solo nucleótido o *single nucleotide polymorphism* (SNP, por sus siglas en inglés). Diversos estudios GWAS han sido realizados en los principales cultivos de importancia mundial, en los cuales se han logrado identificar loci de caracteres cuantitativos (QTL, por sus siglas en inglés) asociados a tolerancia a estrés biótico y abiótico (Brachi et al., 2011). El maíz es una planta ideal para estudios a nivel molecular, ya que el desequilibrio de ligamiento genético decae entre 5 y 10 kilo pares de bases (kpb) en germoplasma tropical de maíz (Lu et al., 2011), lo cual lo hace especial para mapeo genético (Liu et al., 2014). Dos estudios GWAS para tolerancia al CMA han sido reportados en la literatura, en los cuales se analizaron líneas elite endogámicas utilizando marcadores SNP (Cao et al., 2017; Li et al., 2016). Ambos estudios coinciden en la identificación de un QTL de efecto mayor (qRtcs8-1) que confiere tolerancia al CMA y explica entre 13 y 43% de la variación fenotípica observada.

Las variedades nativas son una fuente importante de variación genética, ya que han sido seleccionadas por agricultores por su productividad y adaptabilidad a las condiciones ambientales locales (Pritsch, 2001). En este sentido, la evaluación de variedades nativas de maíz permitiría explorar una gran diversidad genética, convirtiéndose en potenciales portadoras de genes de resistencia a enfermedades. Los objetivos de este estudio fueron (a) identificar marcadores moleculares SNP asociados a la tolerancia genética al CMA en poblaciones nativas de maíz y (b) identificar genes candidatos potencialmente involucrados en mecanismos de defensa al CMA.

Materiales y Métodos

Base de datos fenotípica y genotípica

Se utilizaron las bases de datos fenotípica y genotípica generadas por el proyecto MasAgro-

Biodiversidad-Cimmyt, las cuales son accesibles al público. Para más detalles acerca de las bases de datos fenotípica y genotípica puede consultar en la web de CIMMYT. SeeD Maize Germinate (<http://germinate-stg.seedsofdiscovery.org/maize/>) (Centro Internacional de Mejoramiento de Maíz y Trigo [CIMMYT], 2017). La base de datos fenotípica contiene las estimas BLUP (Best Linear Unbiased Prediction, en inglés) del nivel de tolerancia al CMA de 535 poblaciones nativas, obtenidas a partir de la evaluación de sus respectivos mestizos derivados del cruce con probadores (híbridos de cruce simple). Los mestizos fueron evaluados en campo bajo condiciones naturales de infección en el 2011 en la localidad Guadalupe Victoria, Estado de Chiapas, México. La mayoría de las poblaciones nativas utilizadas fueron colectadas entre 1952 y 1960 en México, Brasil y Guatemala, y en menor proporción provenientes de otros países de Centro y Sudamérica. La base de datos genotípica contiene información de un panel de más de 4,000 poblaciones nativas del Banco de Germoplasma de Maíz del CIMMYT y contiene 946,072 marcadores SNP distribuidos en los 10 cromosomas del maíz (CIMMYT, 2017). El genotipado se realizó utilizando el método GBS, descrito por Elshire et al. (2011).

Procesamiento y control de calidad de la base de datos fenotípica y genotípica

El control de calidad de la base de datos fenotípica consistió en descartar todos los genotipos con fenotipo faltante y sin información de marcadores SNP. De las 535 poblaciones, solamente 463 pasaron el control de calidad. Se realizó un análisis descriptivo del nivel de tolerancia al CMA medido en post-floración y un gráfico de dispersión del nivel de severidad de la enfermedad y el rendimiento de grano por parcela (g) utilizando el programa R Statistics (R-Core Team, 2017).

De la base de datos genotípica se extrajo el genotipo de las 463 poblaciones nativas. Posteriormente se realizó el control de calidad que consistió en eliminar aquellos marcadores cuya frecuencia del alelo menor era menor de 5%. Se seleccionaron aquellos marcadores que no tenían genotipos faltantes. Después del filtrado se obtuvieron 329,692 SNPs polimórficos. El control de calidad se realizó con el software TASSEL (Bradbury et al., 2007).

Estimación de la heredabilidad de la tolerancia al CMA

Se utilizó el método LDAK (*linkage disequilibrium adjusted kinship*, en inglés) para estimar los componentes de varianza (varianza genética aditiva y varianza residual). Este método ajusta la matriz de relaciones genómicas aditivas por la correlación existente entre SNP-SNP debido al desequilibrio de ligamiento (LD) (Speed et al., 2017). Después de ajustar la matriz de relaciones genómicas, se estiman los componentes de varianza mediante un análisis de máxima verosimilitud restringida (REML, del inglés *restricted maximum likelihood*), incluyendo las ponderaciones del aporte de cada SNP. Para implementar este método se utilizó el software LDAK v.5 (Speed et al., 2017). A partir de las estimas de componentes de varianza se calculó la heredabilidad con la siguiente ecuación (Daetwyler et al., 2014):

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2} \quad (1)$$

Donde $\hat{\sigma}_g^2$ es la estimación de la varianza genética aditiva capturada por los marcadores moleculares y $\hat{\sigma}_e^2$ es la estimación de la varianza residual.

Estudio de asociación del genoma completo

Para la identificación de marcadores SNP asociados a la tolerancia genética al CMA se evaluaron dos modelos para estimar los efectos de los marcadores: *Single Marker* y BayesB. El primero consiste en evaluar un solo marcador a la vez mediante un modelo lineal mixto, de tal forma que se realizan tantas regresiones y pruebas de hipótesis como marcadores existan. Se utilizó el modelo lineal mixto descrito por Yu y colaboradores (2006), el cual incluye control de la estructura de población y de relaciones de parentesco para reducir la tasa de falsos positivos. El modelo puede ser escrito de la siguiente manera:

$$\mathbf{y} = \mathbf{W}\boldsymbol{\nu} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (2)$$

Donde \mathbf{y} es el vector de fenotipos (BLUPs) de las poblaciones nativas de dimensión n , $\boldsymbol{\nu}$ corresponde a los efectos fijos de los primeros tres componentes principales, $\boldsymbol{\beta}$ es el efecto fijo de los marcadores; \mathbf{u} son los efectos genéticos aditivos con distribución $\mathbf{u} \sim N_n(\mathbf{0}, \mathbf{G})$. \mathbf{G} es la matriz de covarianza $\mathbf{G} = 2 \mathbf{K}\sigma_a^2$, donde \mathbf{K} es la

matriz de relaciones genómicas de dimensiones $n \times n$, con elementos $\mathbf{K}_{ij}(i, j = 1, \dots, n)$ calculada a partir del conjunto de marcadores genéticos, codificados como 0, 1 y 2 para los genotipos aa, Aa y AA, respectivamente; y σ_a^2 es la varianza genética aditiva. \mathbf{W} , \mathbf{X} , y \mathbf{Z} son las matrices de incidencia para ν , β y \mathbf{u} respectivamente y $\boldsymbol{\varepsilon}$ es el vector de residuales aleatorios, el cual se distribuye $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \mathbf{R})$, donde $\mathbf{R} = \mathbf{I}\sigma_\varepsilon^2$ e \mathbf{I} es una matriz identidad y σ_ε^2 es la varianza residual desconocida. El umbral de significancia para seleccionar los marcadores fue $\alpha = .0001$. Para implementar este modelo se utilizó el software TASSEL. El modelo bayesiano BayesB, descrito originalmente por Meuwissen y colaboradores (2001), se basa en el hecho de que hay pocos loci que generan varianza genética para el carácter de interés. Para ello, el modelo 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 efecto genético nulo son llevados al punto de masa en cero, mientras que los pocos loci con efecto genético toman un valor de la distribución normal (Meuwissen et al., 2001). El modelo puede escribirse de la siguiente manera:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3)$$

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 SNP codificados como 0, 1 y 2 para los genotipos aa, Aa y AA, respectivamente. $\boldsymbol{\varepsilon}$ es el vector de residuos aleatorios, el cual se distribuye $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \mathbf{R})$, donde $\mathbf{R} = \mathbf{I}\sigma_\varepsilon^2$ e \mathbf{I} es una matriz identidad y σ_ε^2 es la varianza de residuos. $\boldsymbol{\beta}$ son los efectos de los marcadores, a los cuales se les asigna una distribución a priori mixta:

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

La distribución a priori de la varianza de los marcadores es $\sigma_j^2 \sim \chi^{-2}(q, S)$ para $j = 1, \dots, p$ y donde q 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 los marcadores tienen efecto diferente de cero. Como criterio para selección de marcadores se utilizó la probabilidad de inclusión posterior (PIP), 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 en R (Pérez-Rodríguez & de los Campos, 2014). Para calcular la proporción de variación fenotípica explicada (PVE) por los marcadores identificados, se ajustaron modelos de regresión lineal múltiple con el conjunto de marcadores identificados con cada modelo de manera independiente. El coeficiente de determinación ajustado r_{aj}^2 obtenido en la regresión es equivalente a la PVE por los marcadores en conjunto (Collard et al., 2005). También se ajustó un modelo de regresión combinando los marcadores identificados con ambos modelos (single marker y Bayes B). Adicionalmente, el software TASSEL calcula un coeficiente de determinación r^2 para cada marcador, lo cual da una idea de cuanto explica cada marcador en particular. Los modelos de regresión lineal múltiple se ajustaron utilizando la función `lm()` en R.

Identificación de genes candidatos

Para el análisis del gen candidato se consideró una ventana de 200 kpb (± 100 Kpb corriente abajo y corriente arriba) tomando como referencia la posición física de los marcadores identificados en el análisis GWAS. La búsqueda de genes candidatos se realizó en la base de datos MaizeGDB (Andorf et al., 2016; Portwood et al., 2019), basado en el genoma de referencia "B73" RefGen_v2(MGSC). Para identificar la posible función del gen candidato se utilizó la base de datos del National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>).

Resultados

Análisis fenotípico

El valor medio (desviación estándar) del nivel de tolerancia al CMA fue de 3.6(0.40), con un rango entre 2.4 y 4.5. El coeficiente de variación fue de 11.2 %. Las poblaciones nativas CHIS474, OAXA280, GUAT153, HAIT13 y TABA26 fueron las que mostraron mayores niveles de tolerancia mientras que las poblaciones OAXA130, BRAZI1 y QROO53 fueron las más susceptibles. La Figura 1 muestra la correlación entre el nivel de tolerancia y rendimiento observado. Se aprecia una correlación negativa (-0.4), por lo que las poblaciones más susceptibles tienden a presentar rendimientos más bajos. Llama la

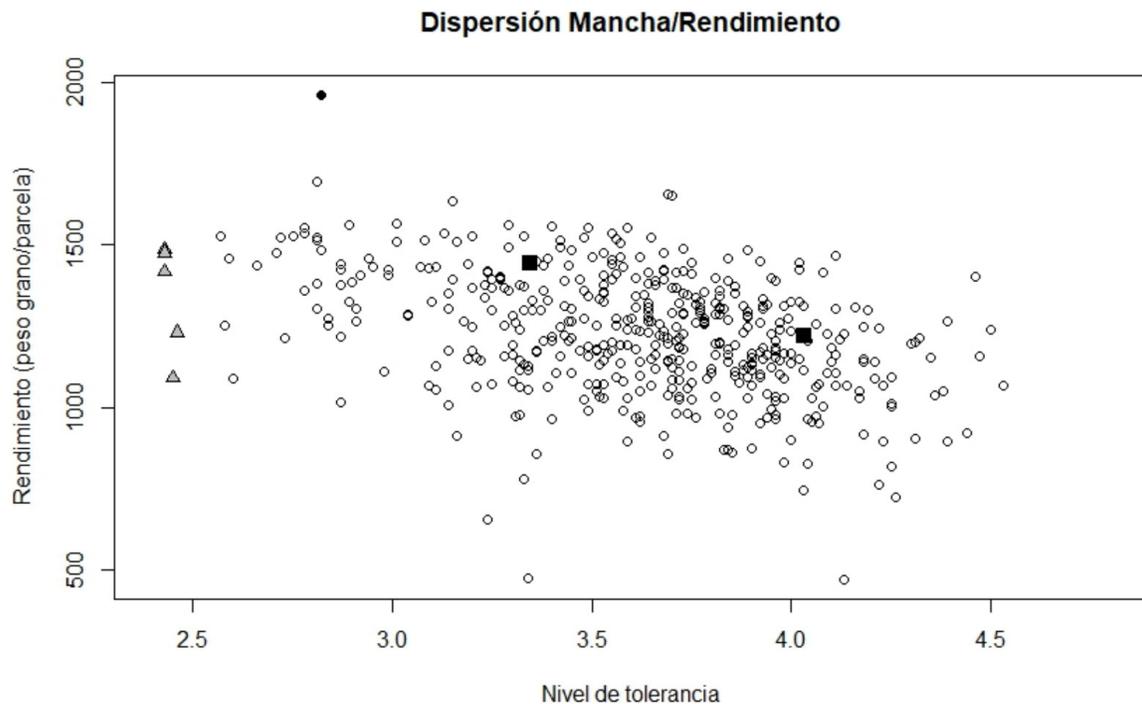


Figura 1. Gráfico de correlación entre nivel de tolerancia a mancha de asfalto y rendimiento de los mestizos derivados de las variedades nativas de maíz. Triángulos grises son los mestizos más tolerantes; los cuadrados negros son los testigos comerciales; el círculo negro es el mestizo derivado de la población VERAGP24_484.

atención la población VERAGP24_484, sobresaliente en rendimiento, pero con moderada susceptibilidad.

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

Se obtuvieron las siguientes estimas de los componentes de varianza, $\sigma_g^2 = 0.0798$ y $\sigma_e^2 = 0.0197$. La estima de la heredabilidad usando marcadores SNP fue $\hat{h}^2 = 0.80$, lo cual revela que la variación observada del nivel de tolerancia al CMA es predominantemente controlada por los efectos genéticos aditivos.

Identificación de marcadores SNP asociados a tolerancia a mancha de asfalto

Con el modelo single marker se identificaron 25 SNPs distribuidos en todos los cromosomas, excepto el nueve (Tabla 1). La proporción de PVE en conjunto por los 25 SNPs fue de 34.25%. En la Figura 2a se muestra la distribución de los marcadores en los cromosomas y sus respectivos puntajes en una escala $-\log_{10}(p)$. Con el modelo BayesB se identificaron 26 marcadores (Tabla 2). Al igual que con el método *single marker*, los marca-

dores asociados se encuentran distribuidos en todos los cromosomas, excepto el nueve. La proporción de PVE en conjunto por los 26 SNPs fue de 52.26%. Los marcadores con mayor probabilidad posterior de inclusión fueron S1_293221122, S6_111966489, S2_19259181 y S1_187644290. La Figura 2b muestra la distribución de los marcadores en los cromosomas y sus respectivos valores PIP.

Al comparar ambos modelos, 11 SNPs fueron identificados en común, lo cual resulta en 40 SNPs únicos al combinar ambos modelos. Los 11 SNPs en común fueron S1_293221122, S1_79329381, S2_19259181, S4_234954673, S6_111966489, S6_111966499, S6_111966511, S6_111966523, S6_164681025, S8_15101936 y S10_135336611.

Es posible que estos marcadores puedan estar fuertemente ligados a QTL que controlan la tolerancia/susceptibilidad al CMA. Los marcadores se encuentran distribuidos en 23 regiones o “Bins” (Tabla 3). La PVE fue de 56% y 22.24% para el grupo de 40 y 11 SNPs, respectivamente. De manera individual, la mayor PVE fue de 5.37% correspondiente al marcador S2_19259181, ubicado en el cromosoma 2 bin 2.03. No se identificaron marcadores asociados a QTL de efecto mayor en este conjunto de datos.

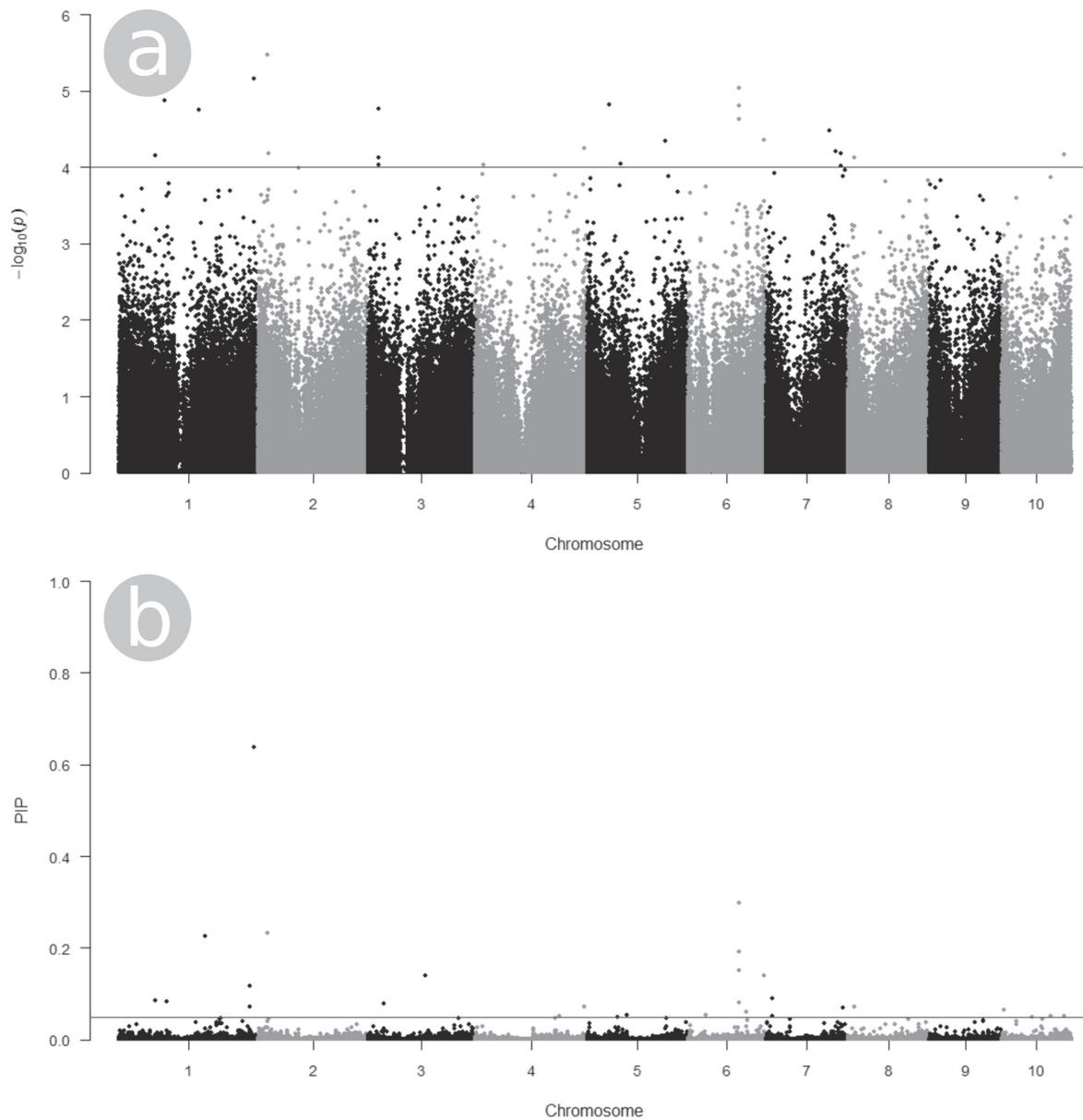


Figura 2. Gráficos tipo Manhattan mostrando la distribución de los marcadores en los 10 cromosomas y su respectivo puntaje según el modelo y escala utilizada. Los colores diferencian a los cromosomas. La línea gris indica el umbral de significancia. (a) Valores de probabilidad en escala $-\log_{10}(p)$ obtenidos con el modelo lineal mixto *single marker* y el umbral de significancia es $\alpha = .0001$ ($-\log_{10}(0.0001) = 4$). (b) Valores PIP obtenidos con el modelo BayesB y el umbral de selección de marcadores es $PIP > .05$.

Tabla 1.

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

SNP ID	Cromosoma	Posición	MAF	<i>p</i>	PVE
S2_19259181*	2	19259181	0.25378	3.35E-06	0.05377
S1_293221122*	1	293221122	0.17387	6.82E-06	0.05066
S6_111966489*	6	111966489	0.2635	9.04E-06	0.04942
S6_111966511*	6	111966511	0.2635	9.04E-06	0.04942
S1_98461847	1	98461847	0.31425	1.31E-05	0.04779
S5_48488908	5	48488908	0.41253	1.53E-05	0.04714
S6_111966523*	6	111966523	0.2635	1.57E-05	0.04702
S3_23495129	3	23495129	0.05616	1.69E-05	0.04669
S1_174002950	1	174002950	0.22246	1.76E-05	0.04651
S6_111966499*	6	111966499	0.2581	2.37E-05	0.04523
S7_137540103	7	137540103	0.13067	3.30E-05	0.04379
S6_164681025*	6	164681025	0.43305	4.44E-05	0.0425
S5_170062437	5	170062437	0.38661	4.53E-05	0.04241
S4_234954673*	4	234954673	0.12851	5.69E-05	0.04142
S7_151082785	7	151082785	0.28294	6.12E-05	0.04111
S7_160773966	7	160773966	0.05832	6.60E-05	0.04078
S2_21726221	2	21726221	0.14903	6.68E-05	0.04073
S10_135336611*	10	135336611	0.45032	6.79E-05	0.04066
S1_79329381*	1	79329381	0.28186	7.06E-05	0.04049
S8_15101936*	8	15101936	0.21058	7.37E-05	0.04031
S3_23495159	3	23495159	0.0594	7.56E-05	0.0402
S5_72696410	5	72696410	0.05724	8.90E-05	0.03949
S3_23495105	3	23495105	0.06479	9.22E-05	0.03934
S4_17333579	4	17333579	0.27862	9.34E-05	0.03928
S7_160773952	7	160773952	0.06911	9.45E-05	0.03923

Nota. PVE Proporción de la variación fenotípica total explicada por el marcador, obtenida a partir del r^2_{adj} . MAF Frecuencia del alelo menor. * SNPs identificados con ambos modelos.

Identificación de genes candidatos

Sesenta y ocho genes candidatos fueron identificados a partir de los 40 SNPs asociados significativamente a la tolerancia al CMA, de los cuales 61 genes codifican proteínas con función conocida y siete fueron proteínas 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 kpb. Los 68 genes candidatos se encuentran distribuidos en las 23 regiones (bins) identificadas en la Tabla 3. Existen regiones que contienen solo un gen candidato y otras regiones que contienen

Tabla 2.

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

SNP ID	Cromosoma	Posición	MAF	PIP
S1_293221122*	1	293221122	0.17387	0.638
S6_111966489*	6	111966489	0.2635	0.29875
S2_19259181*	2	19259181	0.25378	0.23275
S1_187644290	1	187644290	0.31857	0.22475
S6_111966511*	6	111966511	0.2635	0.191
S6_111966523*	6	111966523	0.2635	0.14975
S3_124181546	3	124181546	0.27862	0.14025
S6_164681025*	6	164681025	0.43305	0.13925
S1_282550335	1	282550335	0.42117	0.1165
S7_14658428	7	14658428	0.2095	0.0905
S1_79329381*	1	79329381	0.28186	0.08475
S1_103419465	1	103419465	0.4514	0.08275
S6_111966499*	6	111966499	0.2581	0.07975
S3_35126726	3	35126726	0.31317	0.07825
S1_282674577	1	282674577	0.35421	0.07175
S4_234954673*	4	234954673	0.12851	0.071
S8_15101936*	8	15101936	0.21058	0.0705
S7_166698140	7	166698140	0.09827	0.07025
S10_4820377	10	4820377	0.30346	0.06525
S6_127634563	6	127634563	0.45464	0.05975
S6_38117593	6	38117593	0.23218	0.05425
S5_86013493	5	86013493	0.43844	0.053
S4_182611991	4	182611991	0.35205	0.05175
S10_106065368	10	106065368	0.11447	0.05075
S7_14658316	7	14658316	0.1879	0.05075
S10_135336611*	10	135336611	0.45032	0.05075

Nota. PVE: Proporción de la variación fenotípica total explicada por los marcadores en conjunto, obtenida a partir del r^2_{adj} . MAF Frecuencia del alelo menor. * SNPs identificados con ambos modelos. PIP Probabilidad posterior de inclusión en el modelo

hasta ocho genes asociados a mecanismos de resistencia a enfermedades.

En la Tabla 4 se presenta la lista detallada de los 68 genes candidatos y su potencial función. La función predicha de los genes candidatos sugiere que varios de ellos están asociados con resistencia a enfermedades e

interactúan entre sí en diferentes mecanismos de defensa en plantas. Entre ellos están genes 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), GRMZ-

Tabla 3.

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		

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

M2G132763 (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 caracterizadas como factores de transcripción asociados a resistencia a enfermedades, 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 identificados en este estudio fueron: GRMZM2G351330 (Homeobox-leucine zipper protein HOX22) y GRMZM2G104204 (Homeobox-leucine zipper protein HOX14-like).

También se identificaron algunos genes que codifican proteínas con función de factores de transcripción tipo 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, el cual produce una proteína RPP13-like asociada a resistencia a enfermedades. 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).

Tabla 4.
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
S1_103419465	Bayes B	GRMZM2G070271	1.05	Xyloglucan endotransglucosylase/hydrolase 27
		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
		GRMZM2G462261	2.03	Putative glucuronosyltransferase PGSIP7; Glycosyl transferase family 8
S2_19259181	Bayes B - Single Marker	GRMZM2G162333	2.03	Pectinesterase or Pectin methyltransferase
		GRMZM2G462258	2.03	Inactive beta-amylase 9

Tabla 4 (Continuación)

SNP ID	Modelo	Gen candidato asociado	Bin*	Anotación: Posible función proteica
S2_21726221	Single Marker	GRMZM2G351330 GRMZM2G168917	2.03 2.03	Homeobox-leucine zipper protein HOX22 Putative protein kinase superfamily protein
S3_124181546	Bayes B	GRMZM2G063688	3.04	Galactosyltransferase family protein
S3_23495129, S3_23495159, S3_23495105	Single Marker	GRMZM2G041352 GRMZM2G121747 GRMZM2G345840	3.04 3.04 3.04	Nuclease PA3 Protein ABA DEFICIENT 4 chloroplastic C2 domain-containing protein
S3_35126726	Bayes B	GRMZM2G428456	3.04	Uncharacterized protein
S4_17333579	Single Marker	GRMZM2G104384 GRMZM2G080168 GRMZM2G149108 GRMZM2G106798	4.03 4.03 4.03 4.03	Leucine-rich repeat receptor-like protein kinase family protein BHLH141 - bHLH-transcription factor 141 Ubiquitin-like 1-activating enzyme E1A 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 GRMZM2G143462	4.09 4.09	Putative autophagy domain family protein Ubiquitin receptor RAD23b
S5_48488908	Single Marker	AC225718.2_FG006 GRMZM2G163544 GRMZM2G463913 GRMZM2G042758	5.03 5.03 5.03 5.03	EF-hand Ca2+-binding protein CCD1 Putative beta-glucosidase 41 50S ribosomal protein L24 chloroplastic GDSL esterase/lipase
S5_72696410	Single Marker	GRMZM2G344212 GRMZM5G802725	5.03 5.03	Putative RING zinc finger domain superfamily protein Protein kinase superfamily protein, partial
S5_170062437	Single Marker	GRMZM2G092609 GRMZM2G071484	5.04 5.04	Putative bZIP transcription factor superfamily protein E3 ubiquitin-protein ligase PUB23
S5_86013493	Bayes B	AC207043.3_FG002 GRMZM2G086243	5.04 5.04	VQ motif family protein; Transcriptional Regulators Uncharacterized protein. Análogo en arroz: peptidyl-prolyl cis-trans isomerase putative expressed

Tabla 4 (Continuación)

SNP ID	Modelo	Gen candidato asociado	Bin*	Anotación: Posible función proteica
S6_38117593	Bayes B	GRMZM2G055717	6.01	NADH dehidrogenase (ubiquinone)
S6_111966489, S6_111966499, S6_111966511, S6_111966523	Bayes B - Single Marker	GRMZM2G073700 GRMZM2G118657	6.04 6.04	RNA-binding region-containing protein 1 UDP-glycosyltransferase 85A2
S6_127634563	Bayes B	GRMZM2G086761	6.05	Uncharacterized protein
S6_164681025	Bayes B - Single Marker	GRMZM2G163054	6.05	WRKY-transcription factor 50
S7_14658316, S7_14658428	Bayes B Bayes B	GRMZM2G022619 GRMZM2G174625 GRMZM5G830839	6.07 7.02 7.02	60S ribosomal protein L18a Uncharacterized protein Subtilisin-like protease SBT4.3 y SBT4.4
S7_137540103	Single Marker	GRMZM2G330422	7.03	Análogo en arroz: calcium-transporting ATPase 9 plasma membrane-type putative expressed
		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
		GRMZM2G163138	7.03	Probable LRR receptor-like serine/threonine-protein kinase Atlg51860
S7_151082785	Single Marker	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

Tabla 4 (Continuación)

SNP ID	Modelo	Gen candidato asociado	Bin*	Anotación: Posible función proteica
S10_4820377	Bayes B	GRMZM2G324231	10.01	Uncharacterized protein. Análogo en arroz: avr9/Cf-9 rapidly elicited protein putative expressed
S10_106065368	Bayes B	GRMZM2G314661	10.01	DDT domain-containing protein PTM
S10_135336611	Bayes B - Single Marker	GRMZM2G037335	10.04	Phospholipid-transporting ATPase 2
		GRMZM2G341741	10.04	Pentatricopeptide repeat-containing protein At2g13600
		GRMZM2G113800	10.05	Probable purine permease 11
		GRMZM2G113818	10.05	Probable purine permease 11

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

Discusión

Análisis fenotípico

El rango de los valores fenotípicos fue suficiente para diferenciar claramente entre materiales tolerantes y susceptibles, confirmando la incidencia de la enfermedad por infección natural. El coeficiente de variación refleja niveles de variación aceptable (11.2%) considerando el elevado número de genotipos evaluados. 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 CMA mediante la derivación de líneas tolerantes e introgresión a materiales élite como parte de un programa de pre-mejoramiento.

Heredabilidad

El valor estimado de la heredabilidad concuerda con los reportados por Ceballos y Deutsch (1992) ($\hat{h}^2=0.83$) y está en el rango de los valores de heredabilidad en sentido amplio (incluyendo efectos genéticos no aditivos) obtenidos por Cao y colaboradores (2017) y Mahuku y colaboradores (2016) (entre 0.54 y 0.96). Lo anterior confirma la presencia de un fuerte componente aditivo, lo cual es congruente con los resultados de Hernández-Ramos y colaboradores (2015), quienes encontraron que los efectos del componente aditivo eran 45 veces mayor que los de dominancia. Considerando que no se detectaron QTL de efecto mayor en este conjunto de poblaciones, la tolerancia al CMA puede ser atribuida a los QTL de efecto pequeño identificados en el análisis GWAS, los cuales podrían ser acumulados mediante selección.

Identificación de marcadores SNP asociados a tolerancia a mancha de asfalto

Nuevas regiones asociadas a la tolerancia al CMA fueron identificadas en el presente estudio. Estas regiones se encuentran ubicadas en bins 1.04, 1.10, 1.11, 2.03, 4.03, 5.03, 6.01, 6.04, 6.05, 6.07, 7.04, 8.02, 10.01, 10.04 y 10.05. La mayoría de los genes candidatos identificados en estas regiones codifican proteínas asociadas a mecanismos de resistencia a enfermedades. Esto confirma que las poblaciones nativas de maíz cons-

tituyen una gran fuente de variación genética, permitiendo identificar nuevas regiones portadoras de genes de tolerancia a enfermedades. La utilización de las poblaciones nativas más sobresalientes en programas de pre-mejoramiento contribuiría a ampliar la base genética de los cultivares para tolerancia al CMA.

Algunas de las regiones identificadas en el presente estudio también fueron reportadas por Cao y colaboradores (2017) y Li y colaboradores (2016). Estas regiones se encuentran ubicadas en los cromosomas 1 (bin 1.05 y 1.06), 3 (bin 3.04), 4 (bins 4.08 y 4.09), 5 (bin 5.04) y 7 (bins 7.02 y 7.03) (Cao et al., 2017; Li et al., 2016). Esto sugiere que estas regiones son consistentes a través de diversas poblaciones, líneas endogámicas y poblaciones especiales de mapeo, así como a diversas condiciones ambientales. Sin embargo, el QTL de efecto mayor en el cromosoma 8 bin 8.03 detectado por Cao y colaboradores (2017) y Li y colaboradores (2016), no fue detectado en este estudio. Un solo marcador fue detectado en este cromosoma en el bin 8.02.

En cuanto a los modelos empleados, cabe resaltar que el conjunto de marcadores identificados con el modelo BayesB explican una mayor proporción de variación fenotípica comparado con el modelo *single marker* (52.26% y 34.25%, respectivamente), lo cual podría ser importante en términos de predicción genómica (Meuwissen et al., 2001). Lo anterior no significa que los marcadores identificados con el modelo *single marker* sean menos importantes, ya que la mayoría de estos marcadores también codifican proteínas relacionadas a la tolerancia a enfermedades. La utilización de ambos modelos puede contribuir a la identificación de diferentes regiones y mayor probabilidad de encontrar genes relacionados.

Identificación de genes candidatos

La mayoría de genes candidatos identificados codifican proteínas ampliamente conocidas por estar asociadas a procesos enzimáticos de transporte, factores de transcripción que regulan positivamente la expresión de genes de resistencia, así como receptores de señal involucrados en mecanismos de tolerancia a enfermedades causadas por hongos en maíz, tales como, mancha gris de la hoja (*Cercospora zeaе maydis* Tehon & Daniels), tizón norteño (*Exserohilum turcicum* (Pass.) K.J. Leonard & Suggs), tizón sureño (*Bipolaris maydis* Nisikado & Miyake) y carbón de la espiga (*Sphacelotheca reiliana* (J. G. Kühn) G. P. Clinton) (Kump et al., 2011; Poland et al., 2011; Shi et al., 2014; Wang et

al., 2012). En este estudio se identificaron los factores WRKY y MYC. El primero ha sido encontrado regulando positivamente la expresión de genes de tolerancia a enfermedades (Eulgem et al., 2000; Michelmore et al., 2013; Mohr et al., 2010). El factor MYC ha sido relacionado con la traducción de señal MAPK (del inglés, Mitogen-activated protein kinase) que interviene en mecanismos de resistencia (Tena et al., 2001). 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 tolerancia a enfermedades en plantas y respuesta a estrés, por lo que son considerados genes de tolerancia (Afzal et al., 2008; Ng & Xavier, 2011; Romeis, 2001; Torii, 2004; Zan et al., 2013). Varios genes que codifican estos tipos de proteínas también fueron identificados en líneas endogámicas (Cao et al., 2017; Mahuku et al., 2016). La identificación de múltiples genes con diversas funciones nos da una idea de la complejidad en el genoma, la expresión génica y la interacción de los metabolitos involucrados en el mecanismo de tolerancia a enfermedades en plantas.

Para concluir, las poblaciones nativas de maíz constituyen una gran fuente de variación genética importante, que permiten la identificación de nuevas regiones portadoras de genes de tolerancia a enfermedades. Las nuevas tecnologías de genotipado de alto rendimiento, así como el desarrollo de los métodos robustos para analizar la información, pueden ser empleadas con éxito para la identificación de regiones genómicas y genes asociados a caracteres de importancia económica. 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. Las poblaciones nativas aquí señaladas representan recursos genéticos valiosos para ser incluidas en programas de pre-mejoramiento con la finalidad de ampliar la base genética de líneas élite para tolerancia al CMA. Los genes candidatos identificados en este estudio pueden servir de base para subsecuentes investigaciones, en las cuales se realice un mapeo fino de tales regiones y desarrollar marcadores funcionales para el mejoramiento asistido por marcadores moleculares.

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 del año 2017. Especial agradecimiento a Liuba Cabrera, Carlos Salazar y a Astrid Mérida de la Roca por su apoyo en las diversas actividades del proyecto.

Referencias

- Afzal, A. J., Wood, A. J., & Lightfoot, D. A. (2008). Plant receptor-like serine threonine kinases: roles in signaling and plant defense. *Molecular Plant-Microbe Interactions: MPMI*, 21(5), 507-517. <https://doi.org/10.1094/MPMI-21-5-0507>
- Andorf, C. M., Cannon, E. K., Portwood II, J. L., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Braun, B. L., Campbell, D. A., Vinnakota, A. G., Sribalasu, V. V., Huerta, M., Cho, K. T., Wimalanathan, K., Richter, J. D., Mauch, E. D., Rao, B. S., Birkett, S. M., Sen, T. Z., & Lawrence-Dill, C. J. (2016). MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic Acids Research*, 44(D1), D1195-D1201. <https://doi.org/10.1093/nar/gkv1007>
- Berg, I. van den, Fritz, S., & Boichard, D. (2013). QTL fine mapping with Bayes C(π): a simulation study. *Genetics Selection Evolution*, 45(1), 19, 1-11. <https://doi.org/10.1186/1297-9686-45-19>
- Brachi, B., Morris, G. P., & Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology*, 12, 232, 1-8. <https://doi.org/10.1186/gb-2011-12-10-232>
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633-2635. <https://doi.org/10.1093/bioinformatics/btm308>

- Cao, S., Loladze, A., Yuan, Y., Wu, Y., Zhang, A., Chen, J., Huestis, G., Cao, J., Chaikam, V., Olsen, M., Prasanna, B. M., San Vicente, F., & Zhang, X. (2017). Genome-wide analysis of tar spot complex resistance in maize using genotyping-by-sequencing SNPs and whole-genome prediction. *Plant Genome*, *10*(2), 1-14. <https://doi.org/10.3835/plantgenome2016.10.0099>
- Ceballos, H., & Deutsch, J. A. (1992). Inheritance of resistance to tar spot complex in maize. *Phytopathology*, *82*(5), 505-512. <https://doi.org/10.1094/Phyto-82-505>
- Centro Internacional de Mejoramiento de Maíz y Trigo. (2017, March 12). SeeD Maize Germinate. Tar-Spot_2011B_Guadalupe-Victoria_Chiapas; SeeD GWAS GbS Analysis. [Data set]. <http://germinate-stg.seedsofdiscovery.org/maize/>
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, *142*(1-2), 169-196. <https://doi.org/10.1007/s10681-005-1681-5>
- Crute, I. R., & Pink, D. (1996). Genetics and utilization of pathogen resistance in plants. *The Plant Cell*, *8*(10), 1747-1755. <https://doi.org/10.1105/tpc.8.10.1747>
- Daetwyler, H. D., Bansal, U. K., Bariana, H. S., Hayden, M. J., & Hayes, B. J. (2014). Genomic prediction for rust resistance in diverse wheat landraces. *Theoretical and Applied Genetics*, *127*(8), 1795-1803. <https://doi.org/10.1007/s00122-014-2341-8>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Eulgem, T., Rushton, P. J., Robatzek, S., & Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends in Plant Science*, *5*(5), 199-206. [https://doi.org/10.1016/S1360-1385\(00\)01600-9](https://doi.org/10.1016/S1360-1385(00)01600-9)
- Hernández-Ramos, L., Sandoval-Islas, J. S., Mahuku, G., Benítez-Riquelme, I., & Cruz-Izquierdo, S. (2015). 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. <https://doi.org/10.1007/BF00442777>
- 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*(2), 129-135.
- Red SICTA/IICA/Cooperación Suiza. (2011). *Reconocimiento en campo de la mancha de asfalto*. <http://repiica.iica.int/docs/b3473e/b3473e.pdf>
- Kump, K. L., Bradbury, P. J., Wissler, R. J., Buckler, E. S., Belcher, A. R., Oropeza-Rosas, M. A., Zwonitzer, J. C., Kresovich, S., McMullen, M. D., Ware, D., Balint-Kurti, P. J., & Holland, J. B. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics*, *43*(2), 163-168. <https://doi.org/10.1038/ng.747>
- Li, X., Zhou, Z., Ding, J., Wu, Y., Zhou, B., Wang, R., Ma, J., Wang, S., Zhang, X., Xia, Z., Chen, J., & Wu, J. (2016). Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. *Frontiers in plant science*, *7*, 833. <https://doi.org/10.3389/fpls.2016.00833>
- Loladze, A., Rodrigues, F. A., Toledo, F., San Vicente, F., Gérard, B., & Boddupalli, M. P. (2019). Application of remote sensing for phenotyping tar spot complex resistance in maize. *Frontiers in Plant Science*, *10*, 552. <https://doi.org/10.3389/fpls.2019.00552>
- Lu, Y., Shah, T., Hao, Z., Taba, S., Zhang, S., Gao, S., Liu, J., Cao, M., Wang, J., Prakash, A. B., Rong, T., & Xu, Y. (2011). Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapid LD decay in tropical than temperate germplasm in maize. *PLoS ONE*, *6*(9), e24861. <https://doi.org/10.1371/journal.pone.0024861>

- Mahuku, G., Chen, J., Shrestha, R., Narro, L. A., Guerrero, K. V. O., Arcos, A. L., & Xu, Y. (2016). Combined linkage and association mapping identifies a major QTL (qRtsc8-1), conferring tar spot complex resistance in maize. *Theoretical and Applied Genetics*, *129*(6), 1217-1229. <https://doi.org/10.1007/s00122-016-2698-y>
- Mahuku, G., San Vicente, F., & Shrestha, R. (2013). *Complejo mancha de asfalto del maíz: Hechos y acciones. Folleto Técnico*. http://conservacion.cimmyt.org/index.php/es/component/docman/doc_download/922-tarspot
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829.
- Michelmore, R. W., Christopoulou, M., & Caldwell, K. S. (2013). Impacts of resistance gene genetics, function, and evolution on a durable future. *Annual Review of Phytopathology*, *51*(1), 291-319. <https://doi.org/10.1146/annurev-phyto-082712-102334>
- Mohr, T. J., Mammarella, N. D., Hoff, T., Woffenden, B. J., Jelesko, J. G., & McDowell, J. M. (2010). The Arabidopsis downy mildew resistance gene RPP8 is induced by pathogens and salicylic acid and is regulated by W box cis elements. *Molecular Plant-Microbe Interactions*, *23*(10), 1303-1315. <https://doi.org/10.1094/MPMI-01-10-0022>
- Monterroso, D., Gallardo, N., & Zuñiga, A. (1974). *Informe del programa de investigación de maíz y frijol. Proyecto de Colaboración Ministerio de Agricultura-Facultad de Agronomía*. Facultad de Agronomía, Universidad de San Carlos de Guatemala.
- Mottaleb, K. A., Loladze, A., Sonder, K., Kruseman, G., & San Vicente, F. (2019). Threats of tar spot complex disease of maize in the United States of America and its global consequences. *Mitigation and Adaptation Strategies for Global Change*, *24*(2), 281-300. <https://doi.org/10.1007/s11027-018-9812-1>
- Ng, A., & Xavier, R. J. (2011). Leucine-rich repeat (LRR) proteins. *Autophagy*, *7*(9), 1082-1084. <https://doi.org/10.4161/auto.7.9.16464>
- Pereyda-Hernández, J., Hernández-Morales, J., Sandoval-Islas, J. S., Aranda-Ocampo, S., León, C. de, & Gómez-Montiel, N. (2009). Etiology and management of tar spot (*Phyllachora maydis* Maubl.) of maize in Guerrero state, 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. <https://doi.org/10.1534/genetics.114.164442>
- Poland, J. A., Bradbury, P. J., Buckler, E. S., & Nelson, R. J. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences*, *108*(17), 6893-6898. <https://doi.org/10.1073/pnas.1010894108>
- Portwood II, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Walsh, J. R., Sen, T. Z., Cho, K. T., Schott, D. A., Braun, B. L., Dietze, M., Dunfee, B., Elsik, C. G., Manchanda, N., Coe, E., Sachs, M., Stinard, P., Tolbert, J., ... Andorf, C. M. (2019). MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Research*, *47*(D1), D1146-D1154. <https://doi.org/10.1093/nar/gky1046>
- Pritsch, C. (2001). *El pre-mejoramiento y la utilización de los recursos fitogenéticos. In IICA-PROCISUR. Estrategias en Recursos Fitogenéticos para los Países del Cono Sur. Montevideo, Uruguay*. Instituto Interamericano de Cooperación para la Agricultura.
- R-Core Team. (2017). *R: A language and environment for statistical computing*. <https://www.r-project.org/>
- Romeis, T. (2001). Protein kinases in the plant defence response. *Current Opinion in Plant Biology*, *4*(5), 407-414.
- Shi, L., Lv, X., Weng, J., Zhu, H., Liu, C., Hao, Z., Zhou, Y., Zhang, D., Li, M., Ci, X., Li, X., & Zhang, S. (2014). Genetic characterization and linkage disequilibrium mapping of resistance to gray leaf spot in maize (*Zea mays* L.). *The Crop Journal*, *2*(2), 132-143. <https://doi.org/10.1016/j.cj.2014.02.001>
- Speed, D., Cai, N., Consortium, the U., Johnson, M. R., Nejentsev, S., & Balding, D. J. (2017). Reevaluation of SNP heritability in complex

- human traits. *Nature Genetics*, *49*(7), 986-992. <https://doi.org/10.1038/ng.3865>
- Tena, G., Asai, T., Chiu, W.-L., & Sheen, J. (2001). Plant mitogen-activated protein kinase signaling cascades. *Current Opinion in Plant Biology*, *4*(5), 392-400. [https://doi.org/10.1016/S1369-5266\(00\)00191-6](https://doi.org/10.1016/S1369-5266(00)00191-6)
- Torii, K. U. (2004). Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. *International Review of Cytology*, *234*, 1-46. [https://doi.org/10.1016/S0074-7696\(04\)34001-5](https://doi.org/10.1016/S0074-7696(04)34001-5)
- 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. <https://doi.org/10.1016/j.plantsci.2012.08.004>
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203-208. <https://doi.org/10.1038/ng1702>
- Zan, Y., Ji, Y., Zhang, Y., Yang, S., Song, Y., & Wang, J. (2013). Genome-wide identification, characterization and expression analysis of populus leucine-rich repeat receptor-like protein kinase genes. *BMC Genomics*, *14*, 318. <https://doi.org/10.1186/1471-2164-14-318>