

EVALUACIÓN DEL POTENCIAL DE PRODUCIBILIDAD DE SEMILLA HÍBRIDA EN LÍNEAS ENDOCRIADAS DE MAÍZ (*Zea mays* L.): IDENTIFICACIÓN DE QTLs MEDIANTE MAPEO POR ASOCIACIÓN

Ramos ME¹; Torrent I¹; Lorea RD^{2,3}

¹Bayer Crop Science, Pergamino, Buenos Aires, Argentina

²Estación Experimental Agropecuaria Pergamino, INTA, Buenos Aires, Argentina

³Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Pergamino, Buenos Aires.

EVALUATION OF HYBRID SEED PRODUCTION POTENTIAL IN MAIZE (*Zea mays* L.) INBRED LINES: IDENTIFICATION OF QTLs VIA ASSOCIATION MAPPING

The lines were genotyped using 56,110 SNPs, allowing for the determination of population structure and the identification of nine subgroups. The genotypic matrix was filtered in successive stages prior to molecular characterization and association mapping (AM) analyses, following the parameters recommended by ASRgenomics v1.1.0 (Gezan, 2022). Through association mapping (AM), 169 SNPs associated with the evaluated traits were identified, corresponding to 103 QTLs distributed across the 10 chromosomes of the species. These included 13 QTLs for PMS, 20 for REND, 7 for AP, 11 for P50D, 11 for E50D, 10 for TAP, 15 for EXP, 7 for ALTE, and 9 for ALTP.

Palabras Clave: maíz, líneas endocriadas, producibilidad, mapeo por asociación, QTL

Keywords: maize, inbred lines, producibility, association mapping, QTL

Introducción

Los caracteres agronómicos de importancia económica, como el rendimiento, la floración o los atributos relacionados con la producibilidad de semilla híbrida, son generalmente cuantitativos y están controlados por múltiples genes de pequeño efecto, distribuidos en diferentes regiones del genoma. Estas regiones cromosómicas, conocidas como QTL (Quantitative Trait Loci), contribuyen a la variación fenotípica de un carácter y constituyen un recurso de gran interés para los programas de mejoramiento asistido por marcadores.

El mapeo por asociación (MA), o Genome-Wide Association Study (GWAS), es una herramienta de alta resolución ampliamente utilizada para identificar QTL asociados a caracteres cuantitativos. Esta metodología aprovecha el desequilibrio de ligamiento (DL) existente entre marcadores moleculares y variantes funcionales para detectar asociaciones entre polimorfismos de nucleótido simple (SNPs) y la variación fenotípica observada en una población (Mazaheri, 2019).

El maíz (*Zea mays* L.) constituye una especie particularmente adecuada para este tipo de estudios debido a su elevada diversidad genética, la disponibilidad de plataformas de alta densidad de SNPs, la rápida disipación del desequilibrio de ligamiento y la presencia de estructura poblacional bien definida, características que favorecen la identificación de QTL con alta resolución (Mazaheri, 2019; Cooper, 2019).

Uno de los principales desafíos del GWAS consiste en controlar los falsos positivos derivados del elevado número de pruebas estadísticas realizadas simultáneamente. Aunque la corrección de Bonferroni es ampliamente utilizada, su supuesto de independencia entre marcadores suele resultar excesivamente conservador en especies

donde existe desequilibrio de ligamiento. En este contexto, el método del número efectivo de pruebas independientes (Meff) proporciona un umbral de significancia más realista al considerar la correlación entre loci, mejorando el equilibrio entre la detección de asociaciones verdaderas y el control del error tipo I.

Con base en este enfoque, el objetivo del presente trabajo fue identificar QTLs asociados a caracteres relacionados con la producibilidad de semilla híbrida mediante un análisis de mapeo por asociación en un panel de líneas endogámicas de maíz templado pertenecientes al programa de mejoramiento de la EEA INTA Pergamino.

Materiales y Métodos:

Se evaluó un panel de 170 líneas endogámicas de maíz templado pertenecientes al programa de mejoramiento de la EEA INTA Pergamino, fenotipificadas durante la campaña 2021/22 en cuatro ambientes de la región núcleo maicera argentina para caracteres agronómicos relacionados con la producibilidad de semilla híbrida.

Las líneas contaban con información genotípica obtenida mediante la plataforma MaizeSNP50 BeadChip (Illumina), compuesta originalmente por 56.110 marcadores SNP distribuidos a lo largo del genoma de maíz (Ganal, 2011).

La matriz genotípica fue sometida a un control de calidad utilizando el paquete ASRgenomics v1.1.0 (Gezan, 2022), obteniéndose una matriz final integrada por 143 líneas y 31.803 SNPs de alta calidad. El filtrado de la matriz genotípica (MG) de los datos previo a los análisis de caracterización molecular y mapeo por asociación (MA) se realizaron en sucesivas etapas. En primer lugar, se eliminaron 19.329 SNPs con más de 20% de datos faltantes, considerándose también como dato faltante a los datos en estado heterocigota, dada la imposibilidad de su interpretación al analizar líneas endogámicas. En segundo lugar, se excluyeron 3.955 SNPs por coeficiente de endogamia dentro de poblaciones (FIS), eliminando de esta manera marcadores que presentaron un nivel de heterocigosidad excesiva para líneas endogámicas. Luego se eliminaron 2 marcadores porque su heterocigosidad era superior al 95%. Por último, se descartaron 1.021 marcadores cuya ubicación en el genoma no fuera específica, por falta de posicionamiento o por falta de vinculación a un cromosoma (figura 1).

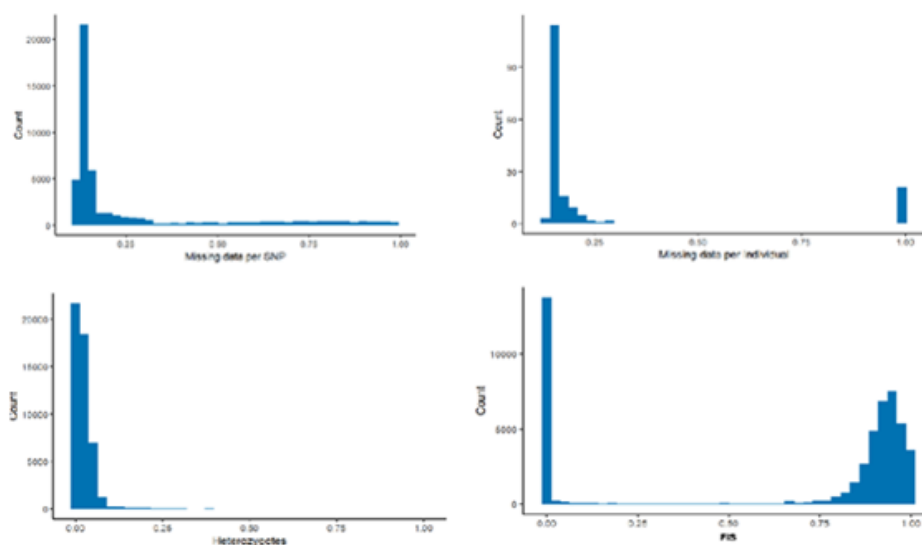


Figura 1 – Filtrado de la matriz genotípica a través de ASRgenomics.

Se evaluó el porcentaje de heterocigosis para cada una de las 170 líneas de la matriz genotípica (MG) (figura 2). Un total de 27 líneas fueron descartadas de los análisis de caracterización molecular y MA dado que presentaban un nivel de heterocigosis superior a 10% o no se contaba con información de estas. Las 143 líneas restantes conformaron el panel en base al cual se realizaron estos análisis.

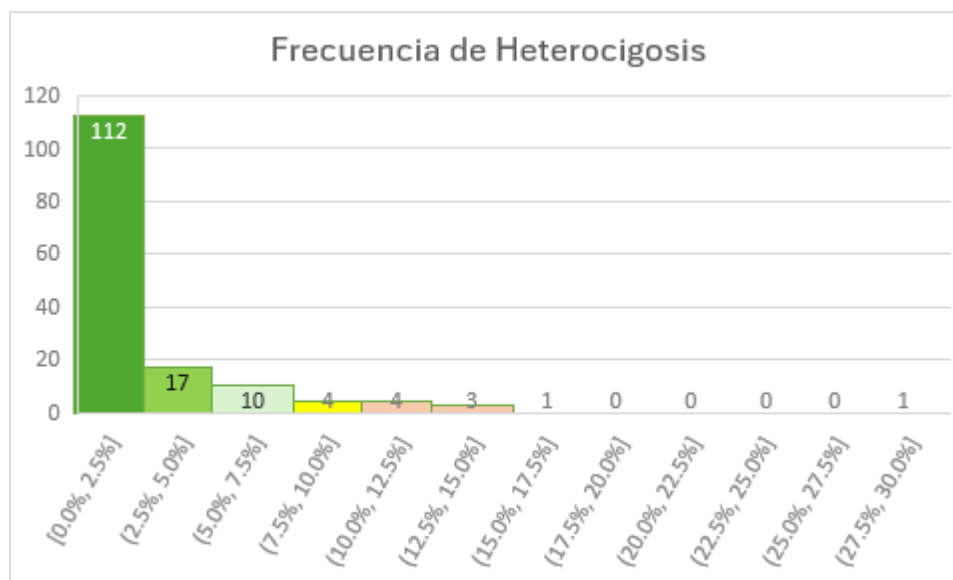


Figura 2: Distribución del nivel de heterocigosis calculado en base a 31.803 SNPs para el panel de 152 líneas de maíz. El número de líneas en cada categoría está indicado sobre las barras.

La matriz de marcadores M, una vez depurada, contiene 143 individuos y 31.803 marcadores (32k) y un 4,98 % de SNPs con datos faltantes.

El mapeo por asociación (GWAS) se realizó mediante un modelo lineal mixto (MLM), incorporando la matriz de estructura poblacional (Q) y la matriz de parentesco (K) para controlar asociaciones espurias a la matriz M. El umbral de significancia se estableció mediante el método del número efectivo de pruebas independientes (Meff), el cual considera la correlación entre marcadores y proporciona un control más adecuado del error tipo I que la corrección de Bonferroni.

Los SNPs significativamente asociados fueron agrupados en QTL considerando su proximidad física y el desequilibrio de ligamiento, delimitando regiones genómicas asociadas a los caracteres evaluados. Posteriormente, estas regiones fueron comparadas con el genoma de referencia de maíz para identificar posibles genes candidatos y contrastarlas con QTL previamente reportados en la literatura.

Resultados:

*Definición del umbral de significancia:

Debido a que los marcadores SNP no constituyen pruebas independientes, el umbral de significancia se estableció mediante el número efectivo de pruebas independientes (Meff), lo que permitió considerar la estructura de desequilibrio de ligamiento presente en el panel de líneas. El umbral obtenido fue de $p = 0,05/127 = 3,93 \times 10^{-4}$, equivalente a $-\log_{10}(p) = 3,40$. En consecuencia, para los análisis de mapeo por

asociación se consideraron asociaciones significativas aquellas con $\text{LOD} \geq 3,40$, proporcionando un equilibrio adecuado entre la detección de asociaciones reales y el control de falsos positivos.

*Evaluación de los modelos de GWAS:

Para seleccionar el modelo más apropiado de mapeo por asociación se comparó el comportamiento de diferentes modelos mediante gráficos QQ, evaluando su capacidad para controlar el error tipo I y minimizar asociaciones espurias.

Los modelos GLM, independientemente de que incorporaran estructura poblacional (Q), componentes principales (PCA) o componentes principales agrupados (PCA(k)), presentaron una marcada desviación respecto de la distribución esperada bajo la hipótesis nula, indicando una elevada inflación de valores p y, por consiguiente, un importante incremento de falsos positivos.

Entre los modelos mixtos, MLM + PCA + K y MLM + PCA(k) + K mostraron una corrección insuficiente de dichas desviaciones. En contraste, el modelo MLM + Q presentó el mejor ajuste a la distribución esperada para valores inferiores al umbral de significancia, manteniendo simultáneamente la capacidad para detectar asociaciones significativas (Figura 3). Por esta razón, este modelo fue seleccionado para realizar el análisis definitivo de mapeo por asociación.

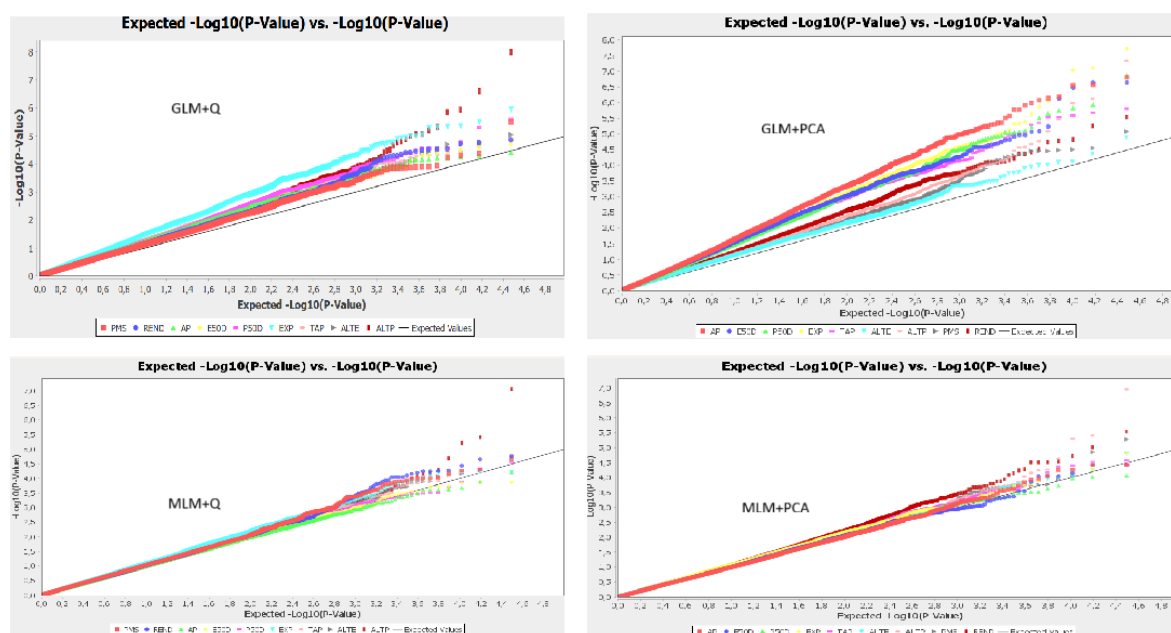


Figura 3: Valores de probabilidades resultantes de los análisis de asociación para la respuesta de los caracteres evaluados implementando diferentes modelos estadísticos. *GLM + Q*: Modelo lineal general con inclusión de la estructura poblacional calculada a través de STR; *GLM + PCA*: Modelo lineal general con inclusión de 4 componentes principales calculado a partir de Tassel; *MLM + Q*: modelo lineal mixto con inclusión de la estructura poblacional más la relación de parentesco calculada a través de STR; *MLM + PCA*: modelo lineal mixto con inclusión de 4 componentes principales más la relación de parentesco calculado a partir de Tassel.

*Identificación de regiones genómicas asociadas:

A partir del modelo MLM + Q se identificaron los SNPs significativamente asociados con los caracteres evaluados. Posteriormente, los marcadores significativos fueron agrupados en QTL considerando su proximidad física y el desequilibrio de ligamiento, evitando la sobreestimación del número de asociaciones mediante la integración de señales redundantes (Li, 2015).

Las regiones delimitadas fueron referenciadas sobre el genoma de maíz para identificar genes candidatos posicionales y compararlas con QTL previamente reportados en la literatura (Yang, 2010; Shamimuzzaman, 2020; Woodhouse, 2021). Este procedimiento permitió detectar regiones genómicas de interés para los caracteres asociados con la producibilidad de semilla híbrida y evaluar el grado de coincidencia con estudios previos.

En el presente trabajo se identificaron de 169 SNPs presentes en 103 QTL detectados con niveles de significancia superiores a $P = 1/127 = 3,93 \times 10^{-4}$, y $-\log_{10}(p) \approx 3,40$, respectivamente. A través de los resultados del análisis de mapeo de asociación, mediante Tassel se obtuvieron gráficos de dispersión (Manhattan Plot), donde podemos visualizar los p-valores de cada carácter evaluado ubicados dentro de los 10 cromosomas del maíz. A continuación, se muestran los gráficos de caracteres prioritarios para parentales femeninos: E50D (figura 4a), ALTE (figura 4b), PMS (Figura 4c) y REND (Figura 4d).

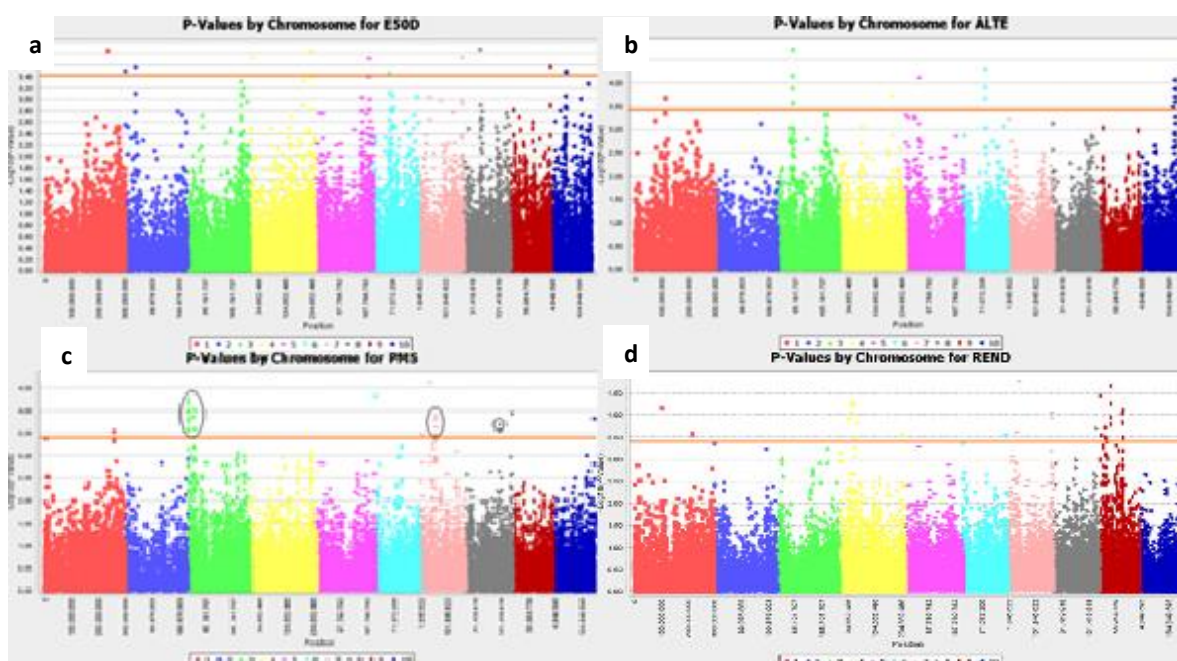


Figura 4: Manhattan Plot de los p-valores para el análisis de asociación (MLM + Q) genotipo-fenotipo. La asociación se llevó a cabo a partir de la evaluación para el carácter PMS. Los SNPs fueron asignados a los 10 cromosomas que se diferencian con distintos colores, sobre el eje X en el orden del 1 al 10. El eje Y indica la probabilidad en una escala logarítmica negativa.

La distribución de los QTLs encontrados para los caracteres evaluados se visualiza en la siguiente tabla 1:

Carácter	QTLs	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10
PMS	13	x		x			x	x	x		x
REND	20	x			X		x	x	x	x	
AP	7		x		X	x	x			x	
P50D	11				X	x	x	x	x	x	x
E50D	11	x	x		X	x	x	x	x	x	x
TAP	10		x	x	X	x	x		x	x	
EXP	15	x		x	X	x	x	x	x	x	x
ALTE	7	x		x	X	x	x				x
ALTP	9	x	x		X	x			x	x	x
Total	103										

Tabla 1: Distribución de los QTLs candidatos encontrados para los 9 caracteres a través de los 10 cromosomas de maíz.

Se detectaron de 16 QTL descritos en la tabla 2 que se solapan con regiones identificadas por diferentes autores:

Carácter	Cromosoma	Nº Genes candidatos	Estudios que respaldan la región
E50D	2	21	Peiffer et al. (2014)
ALTE	3	55	Wallace et al. (2014)
EXP	3	420	Wallace et al. (2014)
ALTP	5	20	Pan et al. (2017)
ALTE	6	70	Peiffer et al. (2014); Kusmec et al. (2017)
EXP	6	59	Wu et al.
PMS	8	63	Liu et al.
EXP	9	79	Wallace et al. (2014); Wu et al.
REND	9	92	Wallace et al. (2014); Liu et al. (2021)
ALTE	10	32	Wallace et al. (2014); Peiffer et al. (2014)

Tabla 2: QTL encontrados en este estudio que se solapan con regiones ya publicadas por otros autores.

Discusión:

Entre los resultados relevantes de este estudio, cabe destacar la coincidencia entre varios QTL detectados y los reportados en trabajos previos. Específicamente, se

observaron señales solapadas con regiones descritas por Peiffer (2014), Wallace (2014), Pan (2017) y Kusmec (2018), entre otros (tabla 2). Estos trabajos, basados en poblaciones complejas y estructura genética amplia, ya habían destacado la presencia de loci asociados a rasgos de desarrollo, ciclo madurativo y de productividad, lo que fortalece la validez de los hallazgos obtenidos en nuestro panel.

La detección de QTL compartidos entre estudios, especialmente en regiones como los cromosomas 3, 5 y 7, aporta evidencia sólida de que estos loci corresponden a hotspots genéticos (hace referencia a regiones del genoma donde se concentra una mayor cantidad de QTL asociados a uno o varios caracteres de interés) conservados en diferentes tipos de germoplasma. Esto refuerza su potencial función regulatoria en rasgos vinculados a la eficiencia reproductiva y, por ende, a la producibilidad de semilla híbrida.

Conclusión:

Estos resultados enfatizan la necesidad de combinar métodos clásicos de selección fenotípica con nuevas herramientas de selección, como por ejemplo el mapeo por asociación y la selección asistida por marcadores moleculares que elevan la eficiencia en la selección de caracteres de heredabilidad moderada.

La estrategia de mapeo por asociación permitió identificar regiones genómicas asociadas con caracteres determinantes de la producibilidad de semilla híbrida en maíz, evidenciando la naturaleza poligénica de este complejo de caracteres.

La coincidencia de varios QTL con regiones previamente reportadas en la literatura respalda la consistencia biológica de los resultados obtenidos, mientras que la identificación de regiones potencialmente novedosas amplía el conocimiento disponible sobre la arquitectura genética de estos caracteres.

En conjunto, los QTL identificados constituyen una base para el desarrollo futuro de herramientas de selección asistida por marcadores y selección genómica orientadas a optimizar la identificación de parentales femeninos y masculinos dentro de programas de mejoramiento de maíz.

Los 103 QTL encontrados pueden ser precedentes en la busca de caracteres de interés para la producción de semillas híbrida de maíz, diferenciando los atributos favorables en la búsqueda de parentales femeninos de aquellos preponderantes en el avance de parentales masculinos.

La combinación de loci validados, nuevos QTL y regiones con posible pleiotropía proporciona un marco sólido para el mejoramiento genético y genera oportunidades para avances aplicados y estudios funcionales posteriores.

Bibliografía:

- Adewale, S. A., Badu-Apraku, B., Akinwale, R. O., Paterne, A. A., Gedil, M., & Garcia-Oliveira, A. L. (2020). Genome-wide association study of Striga resistance in early maturing white tropical maize inbred lines. *BMC plant biology*, 20, 1-16.
- Alberti, M. M. (2015). Comportamiento de cruzamientos entre líneas endocriadas de Maíz (*Zea mays*) clasificadas en grupos basados en información de marcadores moleculares.
- Amadu, M. K., Beyene, Y., Chaikam, V., Tongoona, P. B., Danquah, E. Y., Ifie, B. E., ... & Gowda, M. (2025). Genome-wide association mapping and genomic prediction analyses reveal the genetic



50 AÑOS
PENSANDO
EL FUTURO

XIII
CONGRESO
NACIONAL
DE MAÍZ

architecture of grain yield and agronomic traits under drought and optimum conditions in maize. BMC Plant Biology, 25(1), 135.

Boakyewaa Adu, G., Badu-Apraku, B., Akromah, R., Garcia-Oliveira, A. L., Awuku, F. J., & Gedil, M. (2019). Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. PLoS one, 14(4), e0214810.

Burghardt, L. T., Young, N. D., & Tiffin, P. (2017). A guide to genome-wide association mapping in plants. Current Protocols in Plant Biology, 2(1), 22-38.

Gezan, S. A., de Oliveira, A. A., Galli, G., & Murray, D. (2022). User's manual for ASRgenomics v. 1.1.0: An R package with complementary genomic functions. VSN International.

Kumar, B., Rakshit, S., Kumar, S., Singh, B. K., Lahkar, C., Jha, A. K., ... & Hossain, F. (2022). Genetic diversity, population structure and linkage disequilibrium analyses in tropical maize using genotyping by sequencing. Plants, 11(6), 799.

Montenegro, L. (2022). Estudio de asociación de genoma completo para detectar resistencia a podredumbres de espiga en líneas endocriadas de maíz.

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci USA 108:6893–6898.

Soto-Cerda, B. J., & Cloutier, S. (2012). Association mapping in plant genomes. Genetic diversity in plants, 29-54.

Torrent, I. (2018). Caracterización fenotípica y genética de líneas de maíz templado (*Zea mays* L.) para el comportamiento frente al tizón común (*Exserohilum turcicum*) (Master's thesis).

Wang, J., Yu, J., Lipka, A. E., & Zhang, Z. (2022). Interpretation of Manhattan plots and other outputs of genome-wide association studies. In Genome-Wide Association Studies (pp. 63-80). New York, NY: Springer US.

Abreviatura	Significado	Abreviatura	Significado
ALTE	Altura de Espiga	MA	Mapeo por Asociación
ALTP	Altura de Planta	MAS	Selección Asistida por Marcadores (Marker-Assisted Selection)
AP	Aparición de Panoja	Meff	Número efectivo de pruebas independientes (Effective Number of Independent Tests)
DL	Desequilibrio de Ligamiento (Linkage Disequilibrium)	MG	Matriz Genotípica
EEA	Estación Experimental Agropecuaria	MLM	Modelo Lineal Mixto (Mixed Linear Model)
E50D	Días desde siembra hasta el 50 % de floración femenina (emisión de estigmas)	P50D	Días desde siembra hasta el 50 % de floración masculina (liberación de polen)
EXP	Excursión de Panoja	PCA	Análisis de Componentes Principales (Principal Component Analysis)
F1	Primera generación filial	PCA(k)	Componentes principales agrupados según clústeres
FIS	Coefficiente de endogamia dentro de poblaciones (Inbreeding coefficient)	PMS	Peso de Mil Semillas
GCT	Genes Candidatos	Q	Matriz de estructura poblacional
GLM	Modelo Lineal General (General Linear Model)	QQ Plot	Gráfico cuantil-cuantil utilizado para evaluar el ajuste de los modelos de asociación
GWAS	Genome-Wide Association Study (Estudio de Asociación del Genoma Completo)	QTL	Quantitative Trait Locus (Locus asociado a un carácter cuantitativo)
INTA	Instituto Nacional de Tecnología Agropecuaria	REND	Rendimiento
K	Matriz de parentesco (Kinship matrix)	SNP	Single Nucleotide Polymorphism (Polimorfismo de un solo nucleótido)
LOD	Logarithm of Odds; estadístico utilizado para evaluar la significancia de asociaciones genéticas	STR	Estructura poblacional estimada mediante STRUCTURE
M	Matriz de marcadores moleculares	TAP	Tamaño de Panoja