



**UNIVERSIDAD AUTÓNOMA CHAPINGO**  
**DEPARTAMENTO DE FITOTECNIA**  
**DOCTORADO EN CIENCIAS EN HORTICULTURA**

**MEJORAMIENTO GENÉTICO DE TOMATE ASISTIDO CON  
MARCADORES MOLECULARES DE ADN**

**TESIS**

Que como requisito parcial  
para obtener el grado de:

**Doctor en Ciencias en Horticultura**

**Presenta:**

**LUCAS HERNÁNDEZ IBÁÑEZ**

**Bajo la dirección de: JAIME SAHAGÚN CASTELLANOS, PhD.**

Chapingo, Estado de México, junio de 2017.



COSESION GENERAL ACADEMICA  
DEPTO. DE SERVICIOS ESCOLARES  
ORDINA DE EXAMENES PROFESIONALES



Instituto de Horticultura

**MEJORAMIENTO GENÉTICO DE TOMATE ASISTIDO CON MARCADORES  
MOLECULARES DE ADN**

Tesis realizada por **Lucas Hernández Ibáñez** bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

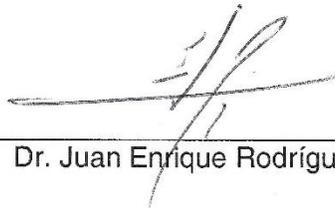
**DOCTOR EN CIENCIAS EN HORTICULTURA**

Director:



Dr. Jaime Sahagún Castellanos

Asesor:



Dr. Juan Enrique Rodríguez Pérez

Asesora:



Dra. Margarita Gisela Peña Ortega

Lector Externo



Dr. Carlos Cintora González

Chapingo, México, mayo de 2017.

## CONTENIDO

1	INTRODUCCIÓN GENERAL.....	3
2	REVISIÓN DE LITERATURA .....	7
2.1	Genética y marcadores moleculares de ADN en tomate .....	7
2.2	Estudios de asociación del genoma completo (GWAS) en tomate.....	9
2.3	Aplicación de la genómica al mejoramiento genético de tomate. ....	10
2.4	Necesidades de investigación .....	12
2.5	LITERATURA CITADA .....	13
3	SELECCIÓN GENÓMICA PARA RENDIMIENTO Y FIRMEZA DE FRUTO EN LÍNEAS ÉLITE F <sub>8</sub> DE TOMATE.....	16
3.1	RESUMEN Y ABSTRACT .....	16
3.1.1	RESUMEN.....	16
3.1.2	ABSTRACT .....	17
3.2	INTRODUCCIÓN.....	18
3.3	MATERIALES Y MÉTODOS.....	20
3.3.1	Evaluación de rendimiento y firmeza de fruto de líneas élite.....	20
3.3.2	Genotipificación de líneas élite .....	22
3.3.3	Modelo estadístico para predicción genómica de rendimiento y firmeza de fruto de líneas élite .....	22

3.4	RESULTADOS Y DISCUSIÓN .....	24
3.4.1	Caracterización fenotípica y genotípica de las líneas élite .....	24
3.4.2	Correlaciones entre rendimientos observados y predichos .....	25
3.5	CONCLUSIONES .....	31
3.6	AGRADECIMIENTOS.....	32
3.7	REFERENCIAS .....	32
4	PREDICCIÓN DE RENDIMIENTO Y FIRMEZA DE FRUTO DE HÍBRIDOS DE TOMATE MEDIANTE MARCADORES MOLECULARES.....	34
4.1	RESUMEN Y ABSTRACT .....	34
4.1.1	Resumen .....	34
4.1.2	Abstract .....	35
4.2	INTRODUCCIÓN.....	36
4.3	MATERIALES Y MÉTODOS.....	37
4.3.1	Selección de líneas progenitoras.....	38
4.3.2	Evaluación de rendimiento de fruto de cruzas intergrupales .....	40
4.3.3	Predicción de rendimiento de cruzas mediante BLUP y RR-BLUP ..	41
4.4	RESULTADOS Y DISCUSIÓN .....	44
4.4.1	Selección de líneas progenitoras mediante ISSR.....	44
4.4.2	Evaluación de rendimiento y firmeza de fruto de cruzas intergrupales .....	46

4.4.3	Predicción de rendimiento de cruzas mediante BLUP y RR-BLUP ..	48
4.4.4	Correlaciones entre rendimientos observados y predichos .....	49
4.5	CONCLUSIONES .....	51
4.6	AGRADECIMIENTOS.....	51
4.7	LITERATURA CITADA .....	52
5	DISCUSIÓN GENERAL .....	54
5.1	LITERATURA CITADA .....	59
6	CONCLUSIONES GENERALES .....	60
7	APÉNDICES.....	61

## LISTA DE CUADROS

Cuadro 4.1. Líneas élite F <sub>8</sub> de tomate cultivado, empleadas en la selección genómica. ....	21
Cuadro 4.2. Estadísticas básicas de tres variables estudiadas, registradas en tres ambientes, para 17 líneas élite de tomate. ....	24
Cuadro 4.3. Estadísticas básicas de tres variables estudiadas, registradas en tres ambientes, para 28 líneas élite de tomate. ....	25
Cuadro 4.4. Correlaciones promedio de la validación cruzada para predicción de líneas élite de tomate en tres variables y tres ambientes de producción. Los datos corresponden a 1,000 muestras aleatorias independientes de 17 líneas. ....	27
Cuadro 4.5. Correlaciones promedio de la validación cruzada para predicción de líneas élite de tomate en tres variables y tres ambientes de producción. Los datos corresponden a 1,000 muestras aleatorias independientes de 28 líneas. ....	28
Cuadro 5.1. Secuencia de iniciadores ISSR utilizados en la caracterización de la divergencia genética de 39 líneas élite de tomate, temperatura de alineamiento (Ta), número total de bandas amplificadas (TB) y polimórficas (BP) y porcentaje de polimorfismo (PP).....	45
Cuadro 5.2. Máximos, mínimos, medias y estimaciones de varianzas y heredabilidades de cuatro caracteres de 36 cruza simples de tomate en dos ciclos de producción (2014 y 2015), cultivados en condiciones de hidroponía e invernadero. ....	47
Cuadro 5.3. Correlaciones promedio de la validación cruzada para dos métodos (BLUP y RR-BLUP) de predicción de híbridos de tomate en cuatro variables. Los datos corresponden a 1,000 muestras aleatorias independientes de 36 híbridos predictores. ....	49

## LISTA DE FIGURAS

- Figura 4.1. Gráfica de valores genéticos predichos para los conjuntos de entrenamiento y prueba con  $n_t = 14$ . Modelo G-Blup que emplea marcadores moleculares en rendimiento de fruto total (izquierda); y G-Blup utilizando información de pedigrí con la variable firmeza de fruto (derecha).  
..... 29
- Figura 5.1. Agrupamiento de 39 líneas élite de tomate, genotipeadas con 29 iniciadores ISSR, mediante la distancia de Jaccard y el método de Ward. 46

## AGRADECIMIENTOS

A la **Universidad Autónoma Chapingo** y al **Instituto de Horticultura**, por mi formación como Doctor en Ciencias en Horticultura.

Al **Consejo Nacional de Ciencia y Tecnología (CONACyT)**, por la beca del programa de formación y desarrollo de científicos y tecnólogos, para la realización de mis estudios de doctorado.

A mi Comité Asesor integrado por el **Dr. Jaime Sahagún Castellanos**, **Dr. Juan Enrique Rodríguez Pérez** y la **Dra. Margarita Gisela Peña Ortega**, quienes me apoyaron en todas las etapas de la investigación, compartieron sus conocimientos, me orientaron para tomar las mejores decisiones y sobre todo me brindaron su amistad y confianza.

A mis padres **José Florencio Hernández Villalba** y **María Enriqueta Marcelina Ibáñez Padrón**, y hermanos **Enriqueta**, **Guadalupe**, **Margarita**, **Alejandra Leonardo**, **Lidia** y **Florencio**, por su apoyo total.

A mi esposa **Rosa Franco Amaro** y mis hijos **José Alejandro**, **Laisha Rosa Idalid** y **Marco Antonio** por su amor incondicional.

A los compañeros del posgrado en Horticultura, muy especialmente a Mariela. Y a los alumnos de Fitotecnia y Horticultura Protegida, de los cursos de estadística y diseños experimentales; así como a trabajadores académicos o administrativos, que colaboraron en alguna etapa de la investigación.

## DATOS BIOGRÁFICOS



### Datos personales

Nombre	Lucas Hernández Ibáñez
Fecha de nacimiento	18 de octubre de 1972
Lugar de nacimiento	Zacatlán, Puebla
CURP	HEIL721018HPLRBC01
Profesión	Ingeniero Agrónomo especialista en Fitotecnia
Cédula profesional	6361018

### Desarrollo académico

Bachillerato	Preparatoria Agrícola, Universidad Autónoma Chapingo
Licenciatura	Departamento de Fitotecnia, Universidad Autónoma Chapingo
Maestría	Maestría en Ciencias en Horticultura, Universidad Autónoma Chapingo

## RESUMEN GENERAL

### MEJORAMIENTO GENÉTICO DE TOMATE ASISTIDO CON MARCADORES MOLECULARES DE ADN

En el mejoramiento genético la predicción del rendimiento de los híbridos posibles, ante un elevado número de líneas progenitoras, se ha hecho indispensable por razones prácticas y económicas; el objetivo es crear sólo aquellos que cuenten con alta probabilidad de poseer desempeño adecuado. Existen varios métodos de predicción, y el uso de marcadores moleculares ha mejorado en mucho su precisión. La presente investigación tuvo como objetivo evaluar tres métodos de selección genómica de líneas élite y dos métodos de predicción de cruza simples de tomate. Inicialmente se evaluó la selección genómica en dos grupos de líneas élite F<sub>8</sub> de tomate, para rendimiento y firmeza de fruto en tres ambientes. Las líneas también fueron caracterizadas con marcadores de ADN tipo inter secuencias simples repetidas (ISSR). Posteriormente se predijo su rendimiento con el método del mejor predictor lineal e insesgado genómico (G-Blup). Las correlaciones entre rendimientos predichos y observados, fluctuaron entre 0.0 y 0.67; la magnitud de las correlaciones se incrementó al aumentar el número de líneas predictoras y el de ambientes de evaluación. Los tres métodos para seleccionar líneas fueron adecuados en la predicción del rendimiento y firmeza del tomate; aunque la combinación marcadores-pedigrí mostró mayor correlación. En una segunda etapa se compararon dos métodos de predicción de híbridos: mejor predictor lineal insesgado (BLUP) y regresión ridge BLUP. Para ello se realizaron 36 cruza simples, que fueron evaluadas en firmeza y rendimiento de fruto. Con 1,000 muestras aleatorias independientes de tamaños,  $n = 6, 12, 18, 24$  y 30 híbridos, se predijo el comportamiento de los 36- $n$  híbridos restantes. Las correlaciones entre valores observados y predichos fluctuaron entre 0.25 y 0.83. El BLUP, de manera consistente, registró los valores más altos. En ambos métodos cuando  $n$  aumentó, la magnitud de las correlaciones también se incrementó.

**Palabras clave:** *Solanum lycopersicum* L.; Selección genómica; Mejor predictor lineal e insesgado; G-Blup; Regresión ridge BLUP; componentes de rendimiento.

Tesis de Doctorado en Ciencias en Horticultura, Universidad Autónoma Chapingo.

Autor: Lucas Hernández-Ibáñez

Director de tesis: Jaime Sahagún-Castellanos

## ABSTRACT

### TOMATO GENETIC IMPROVEMENT ASSISTED WITH DNA MOLECULAR MARKERS

In genetic improvement, prediction of hybrid performance given a high number of parental lines has become indispensable for practical and economic reasons; the goal is to create only those with a high probability of having adequate performance. There are several prediction methods, and the use of molecular markers has greatly improved their accuracy. The objective of this research was to evaluate three methods for genomic selection of elite lines and two methods for prediction of hybrid performance in tomato. First, genomic selection for fruit yield and fruit firmness was evaluated in two groups of tomato F<sub>8</sub> elite lines, in three environments. Likewise, the lines were characterized with inter simple sequence repeats (ISSR) DNA markers. Their performance was predicted using the genomic best linear unbiased prediction method (G-Blup). Correlations between predicted and observed yields ranged from 0.0 to 0.67; correlations increased when the number of predicting lines and evaluation environments were larger. The three methods of line selection were adequate for prediction of tomato fruit yield and firmness; although the markers-pedigree combination showed greater correlation. In the second step we compared two methods for predicting hybrid performance: the best linear unbiased prediction (BLUP) and ridge regression BLUP. 36 single crosses were made and tested for fruit yield and fruit firmness. With 1,000 size-independent random samples,  $n = 6, 12, 18, 24$  and  $30$ , the behavior of the remaining  $36-n$  hybrids was predicted. The correlation coefficients between predicted and observed values fluctuated between 0.25 and 0.83. BLUP consistently recorded the highest values. In both methods of prediction when  $n$  increased the magnitude of correlations also increased.

**Keywords:** *Solanum lycopersicum* L.; Genomic selection; The best linear unbiased prediction; G-Blup; Ridge regression BLUP, fruit yield components.

Thesis: Doctor of Science in Horticulture, Universidad Autónoma Chapingo.

Author: Lucas Hernández-Ibáñez

Advisor: Jaime Sahagún-Castellanos

# 1 INTRODUCCIÓN GENERAL

El tomate (*Solanum lycopersicum* L.) es originario de la región andina en América del Sur. Se le encuentra en estado silvestre desde Ecuador hasta el norte de Chile. México y América central son considerados como su centro de domesticación, y desde aquí se dispersó a Europa y luego al resto del mundo (Rodríguez *et al.*, 2011).

En México se siembran al año 50 mil hectáreas de tomate, con una producción de poco más de 3 millones de toneladas, lo que genera fuentes de empleo y divisas por las exportaciones. Sin embargo, toda la semilla utilizada es producida fuera del país por compañías internacionales, esto genera dependencia científica, tecnológica y económica. Por lo anterior, es necesario aplicar los últimos avances en genómica, para generar nuevas variedades de tomate en los programas nacionales, con el objetivo de reducir el tiempo de los ciclos de mejoramiento y los recursos materiales y financieros necesarios.

El tomate es especie modelo en la investigación de plantas de fruto carnoso, ha sido utilizado para estudiar diferentes aspectos genéticos, desde la herencia mendeliana hasta la secuenciación de su genoma completo, pasando por el ligamiento, mapeo cromosómico y detección de caracteres de tipo cuantitativo. La secuenciación del genoma del tomate ha apoyado y acelerado los trabajos de investigación que buscan dilucidar el comportamiento e interacción de los genes en la especie (Menda *et al.*, 2013).

Desde el desarrollo de los primeros marcadores moleculares de ADN, se despertó el interés de los mejoradores por incorporar su uso en los programas de fitomejoramiento. Los marcadores iniciales presentaron dos serios inconvenientes: alto costo y baja densidad. En la actualidad los avances en genotipo han reducido los costos de los marcadores e incrementado la

cobertura del genoma (Heffner *et al.*, 2010), permitiendo su uso rutinario en programas de mejoramiento genético. Entre los primeros usos prácticos de los marcadores moleculares está la selección asistida (MAS). En esta selección se identifica marcadores asociados a caracteres de interés, lo que a su vez permite la selección de genotipos superiores. El método es eficiente para genes de efectos mayores, pero es poco efectivo para caracteres controlados por muchos genes con efectos muy pequeños. Ante esta situación, surgió la selección genómica (SG), que utiliza la información de todos los marcadores de que se disponga para estimar los valores reproductivos (genomic estimated breeding values GEBV) para caracteres complejos (Heffner *et al.*, 2009).

Si bien, la precisión en las predicciones basadas en SG en varios casos no ha sido muy alta, con ella se podría acelerar dramáticamente la ganancia genética mediante el acortamiento de los ciclos de mejoramiento (Heffner *et al.*, 2010), debido a una mayor respuesta a la selección por unidad de tiempo (Bernardo, 2010). En tomate hay pocos reportes de estudios que apliquen la SG. Hernández-Bautista *et al.* (2016) hicieron predicción para selección en familias  $F_{2:3}$ , y concluyeron que la selección fenotípica fue más eficiente que la selección genómica. Yamamoto *et al.* (2016), mediante marcadores de polimorfismo de un solo nucleótido (SNP), realizaron predicción y simulación para mejorar el rendimiento y sabor de tomate en variedades de fruto grande; con excelentes perspectivas.

Por otro lado, el manejo de la heterosis que permite la producción de variedades híbridas sobresalientes es probablemente el mayor éxito del mejoramiento genético de las plantas. El desarrollo de variedades híbridas sobresalientes, sin embargo, no es fácil: menos del 1 % de los híbridos de cruce simple de maíz (*Zea mays* L.) llegan a ser variedades comerciales (Bernardo, 1996), y algo similar debe suceder en otros cultivos, particularmente entre algunos que corresponden a especies autógamas, como el tomate (*Solanum lycopersicum* L.). Entre los obstáculos que encuentra el fitomejorador destaca la necesidad de evaluar en el

campo una cantidad elevada de híbridos experimentales, que usualmente rebasa las capacidades de los programas de mejoramiento. Para enfrentar esta dificultad se han desarrollado teoría y métodos de predicción del valor genotípico de variedades que no existen físicamente. Estos métodos se basan en la formación y evaluación experimental de una muestra pequeña de cruzas, y en las coancestrías de las líneas progenitoras, estimadas con base en sus registros de pedigrí o en sus huellas genómicas (Henderson, 1985; Bernardo, 1994; Gbur *et al.*, 2012).

En maíz se ha predicho el rendimiento de grano de cruzas simples mediante una ecuación de predicción que produce “los mejores predictores lineales insesgados” (BLUP, por sus siglas en inglés: best linear unbiased prediction) (Bernardo, 1993, 1995). Estas predicciones se basaron en marcadores moleculares RFLP’s (Bernardo, 1994, 1996), microsatélites (Balestre *et al.*, 2010) y en SNP (Massman *et al.*, 2013). Los coeficientes de correlación ( $r$ ) entre rendimientos observados y pronosticados fluctuaron entre 0.43 y 0.76. Aunque el tomate es la segunda especie hortícola en importancia a nivel mundial, y su mejoramiento genético en las últimas décadas se ha orientado a la producción y liberación de variedades híbridas  $F_1$ , los reportes sobre la predicción de rendimiento de híbridos son pocos.

Como se ha expuesto, son escasos los trabajos reportados de aplicación de la genómica en el mejoramiento de tomate. En virtud de que tanto para la producción de variedades de polinización libre como variedades híbridas  $F_1$  es necesario disponer de líneas puras, se requiere implementar metodologías que permitan aplicar la SG en los principales métodos de selección de líneas endogámicas, con el fin de acortar el tiempo necesario para llegar a liberar una nueva variedad al mercado, dado que el costo de producción del tomate es muy elevado, sobre todo si se hace en invernadero. La evaluación de campo de nuevos cultivares implica inversión financiera alta. Para hacerla más productiva se requiere implementar estrategias, como la predicción de rendimiento de fruto,

que permitan hacer y evaluar sólo las cruzas con mayor probabilidad de éxito. Por lo anterior, los objetivos generales de esta investigación fueron: a) Evaluar la selección genómica en líneas élite  $F_8$  de tomate desarrolladas mediante el método uniseminal y genotipificadas con marcadores de ADN del tipo intersecuencias simples repetidas (ISSR's); y b) predecir la firmeza, el número y el rendimiento total y comercial de frutos de cruzas simples de tomate mediante el empleo de los métodos BLUP y regresión ridge BLUP.

## 2 REVISIÓN DE LITERATURA

### 2.1 Genética y marcadores moleculares de ADN en tomate

El tomate tiene su centro de origen en la región andina, y su centro de domesticación en Mesoamérica. De México se llevó por los españoles a Europa, donde en un principio se le utilizó como especie ornamental y una vez que se agregó a la dieta, el gusto por esta especie se dispersó por el mundo entero. La domesticación de muchas plantas se ha correlacionado con un aumento espectacular en el tamaño del fruto. El ancestro del tomate domesticado, fue de fruto menor a 1 cm de diámetro y unos pocos gramos de peso, con cientos de semillas. Por el contrario, los tomates modernos pueden llegar a pesar hasta 1,000 gramos y superar los 15 cm de diámetro (Frery *et al.*, 2000). El fruto de tomate var. *cerasiforme*, también conocido como tomate cherry es más grande que el tipo silvestre pero de menor tamaño que el cultivado, lo que sugiere que es el progenitor directo del tomate cultivado (Rodríguez *et al.*, 2011).

La amplia diversidad fenotípica en el tomate cultivado es particularmente evidente en caracteres como la forma, color y tamaño del fruto, no es de extrañar que la guía de la Unión Internacional para la Protección de las Obtenciones Vegetales (UPOV) o del Instituto Internacional de los Recursos Genéticos Vegetales (IPGRI), consideren por lo menos nueve tipos diferentes de forma del fruto para caracterizar accesiones de tomate (Rodríguez *et al.*, 2011), además de variación también en el resto de caracteres morfológicos de la planta. Esta alta diversidad fenotípica no se corresponde con la estrecha diversidad genética del tomate cultivado, el proceso de domesticación y sobre todo el de dispersión generó un cuello de botella, de tal manera que la base genética de la cual se han derivado todos los cultivares modernos, es muy limitada y la variación observada son mutaciones puntuales espontáneas que el hombre ha seleccionado y mantenido para su beneficio. Lo anterior se constató al intentar utilizar los primeros marcadores moleculares de ADN como los RAPD's (ADN polimórfico

amplificado al azar), que no generaban polimorfismo o éste era muy bajo entre variedades de tomate cultivado. Un avance fue el uso de SSR's e ISSR's, que produjeron mayor polimorfismo y permitieron caracterizar genéticamente diversos cultivares (Bojinov y Danailov, 2008).

El avance de la biología molecular en las últimas décadas y en especial de los sistemas de secuenciación permitió implementar marcadores de ADN más precisos como los SNP's (polimorfismo de nucleótido único), y la secuenciación completa del tomate que se inició en 2005 y concluyó en 2012 con su publicación. El genoma secuenciado fue el de la línea endogámica de tomate "Heinz 1706", su tamaño predicho fue de aproximadamente 900 megabases (Mb). Se estimó una divergencia entre los genomas silvestre y domesticado del 0.6% (5.4 millones de SNPs distribuidos a lo largo de los cromosomas). Se identificó que los cromosomas de tomate consisten en heterocromatina pericéntrica y eucromatina distal, con repeticiones concentradas dentro y alrededor de los centrómeros, en cromómeros y en los telómeros. Se predijeron 34,727 genes que codifican para proteínas. De ellos, 30,855 son apoyados por datos de la secuenciación de ARN (Tomato Genome Consortium, 2012).

La publicación de la secuencia del tomate, generó optimismo en la comunidad científica, al suponer que el conocimiento del genoma entre otras cosas permitiría promover estudios genómicos a gran escala de especies emparentadas, además de ayudar a resolver preguntas específicas relacionadas al desarrollo básico y la regulación del metabolismo secundario de estas especies (Ranjan *et al.*, 2012). Por su parte, Menda *et al.* (2013) aseguran que la utilidad de este genoma ha quedado demostrada por el relativo gran número de estudios que han sido publicados desde que se liberó la secuencia, cubriendo un amplio rango de temas como expresión genética, diversidad genética, filogenética, genómica comparativa y epigenética. Esto último parece una sobrevaloración, pues los trabajos a que hacen referencia ya estaban en marcha y sólo se apoyaron en el

genoma publicado, más no partieron de él, como es el caso de algunas publicaciones recientes.

## **2.2 Estudios de asociación del genoma completo (GWAS) en tomate**

La relación causal entre polimorfismo genético en una especie y las diferencias fenotípicas observadas entre individuos, ha sido de mucho interés desde que Mendel postuló la existencia de “factores internos” que son heredados a la siguiente generación (Korte y Farlow, 2013). Así, desde el nacimiento de la Genética se ha trabajado en identificar caracteres morfológicos y su tipo de herencia, una vez identificados se les asignaba una posición en los cromosomas, ello originó los primeros mapas de ligamiento. Los marcadores moleculares de ADN permitieron construir mapas genéticos más precisos, que incluyeron caracteres de herencia simple y posicionaron caracteres cuantitativos (QTL's). Ejemplo de un QTL en tomate es el locus *fw2.2*, responsable de cambiar el peso del fruto hasta en un 30% (Frary *et al.*, 2000). El mapeo de QTL's ha provisto un método poderoso para identificar regiones del genoma que co-segregan con un carácter dado, identificado en poblaciones F<sub>2</sub> o familias de líneas homocigóticas recombinantes, y en los últimos años se ha trabajado en conjunto con estudios de asociación del genoma completo o GWAS (Genome-wide Association Studies). El enfoque básico de un GWAS es evaluar la asociación entre cada marcador genotipeado y un fenotipo de interés que ha sido registrado a través de un gran número de individuos (Korte y Farlow, 2013).

En la última década los polimorfismos de secuencia única o SNP, se han constituido como el sistema de marcadores de ADN preferido para identificar caracteres de interés y elaborar mapas de ligamiento. Sim *et al.* (2012) desarrollaron un arreglo para tomate con 8,784 SNPs que usaron para generar mapas de ligamiento de alta densidad para tres poblaciones F<sub>2</sub> específicas, e identificaron la posición física de 7,666 SNP's en relación al genoma del tomate.

Los estudios de asociación del genoma completo mediante SNPs han tenido éxito en la identificación de genes relacionados con caracteres poligénicos y son valiosos para el mejoramiento de los cultivos. No necesariamente abarcan todo el genoma, como lo demuestra el trabajo de Ranc *et al.* (2012), que se enfocaron al cromosoma 2 y con 352 SNP's encontraron asociación con caracteres como el peso de fruto, el número de lóculos y el contenido de sólidos solubles. Además determinaron que el polimorfismo en *S. lycopersicum* es menor que el de *S. lycopersicum* var. *cerasiforme* y mucho menor al de *Solanum pimpinellifolium*, mientras el desequilibrio de ligamiento se comportó en forma inversa. Shirasawa *et al.* (2013) optaron por el enfoque contrario, secuenciaron completamente el genoma de seis líneas de tomate cultivado, detectaron alrededor de 1.5 millones de SNP's candidatos de los cuales 1,293 fueron exitosamente genotipados. A través del GWAS identificaron SNP's que se asociaron significativamente con caracteres agronómicos. Sauvage *et al.* (2014), utilizaron una colección núcleo de 163 accesiones de tomate compuesta de *S. lycopersicum*, *S. lycopersicum* var *cerasiforme* y *Solanum pimpinellifolium* para mapear loci que controlan la variación en metabolitos del fruto. Los frutos fueron fenotipados para una amplia gama de metabolitos, incluyendo aminoácidos, azúcares y ascorbato. En paralelo, las accesiones fueron genotipadas con 5,995 marcadores SNP, repartidos en todo el genoma. Los autores detectaron un total de 44 loci que fueron asociados significativamente con un total de 19 caracteres, incluyendo niveles de citrato, ascorbato, malato y sacarosa.

### **2.3 Aplicación de la genómica al mejoramiento genético de tomate.**

Desde que se desarrollaron los primeros marcadores moleculares de ADN, se despertó el interés de los mejoradores por incorporar esta herramienta biotecnológica a los programas de fitomejoramiento. La posibilidad de asociar un conjunto de marcadores de ADN a sus correspondientes caracteres morfológicos, con mayor precisión, hizo pensar en una nueva revolución verde.

Los primeros marcadores presentaron dos serios inconvenientes: alto costo y baja densidad. Esto último se refiere a un número reducido de marcadores por iniciador empleado, lo que implicó segmentos grandes del genoma sin muestrear, con una baja precisión al momento de identificar QTL's. En años recientes, los avances en genotipo están rápidamente reduciendo los costos de los marcadores e incrementando la cobertura del genoma (Heffner *et al.*, 2010).

La aplicación de los marcadores se ha estudiado en las dos vertientes del mejoramiento: selección e hibridación. La primera estrategia empleada fue la selección asistida con marcadores (MAS), que identificaba marcadores asociados a QTL's y los utilizaba para selección de genotipos superiores. El método fue eficiente para genes de efectos mayores, pero poco efectivo para caracteres controlados por muchos genes con efectos muy pequeños. Ante esta situación, surgió la selección genómica (SG), que utiliza la información de todos los marcadores para calcular los valores genéticos de cría estimados (genomic estimated breeding values GEBV) para caracteres complejos (Heffner *et al.*, 2009). Si bien, la precisión en las predicciones en varios casos no es muy alta, la SG podría acelerar dramáticamente la ganancia genética mediante el acortamiento de los ciclos de mejoramiento (Heffner *et al.*, 2010), mejorando la respuesta a la selección por unidad de tiempo (Bernardo, 2010). Con estas perspectivas, la SG ha sido estudiada en diversos caracteres y cultivos, por ejemplo: Zhang *et al.* (2016) predijeron caracteres de calidad maltera en cebada, y Schmidt *et al.* (2016) hicieron lo propio para peso de semilla en soya, en ambos casos con excelentes resultados.

En tomate hay pocos reportes de estudios que apliquen la SG. Hernández-Bautista *et al.* (2016), evaluaron los caracteres rendimiento de fruto, peso de fruto promedio, diámetros polar y ecuatorial, número de lóculos y sólidos solubles totales; genotipearon con marcadores SSR's, y mediante los métodos Regresión Lineal Múltiple, Regresión Ridge y Bayes Lasso, hicieron predicción para selección en familias F<sub>2:3</sub>. Concluyeron que la selección fenotípica fue más

eficiente que la selección genómica. Yamamoto *et al.* (2016), mediante SNP's, realizaron predicción y simulación para mejorar el rendimiento y sabor de tomate en variedades de fruto grande; fenotipearon 20 caracteres y emplearon siete métodos de predicción. La regresión Ridge y Bayes Lasso predijeron valores de rendimiento, peso de fruto promedio y sólidos solubles totales más altos que Hernández-Bautista *et al.* (2016). La discrepancia anterior, posiblemente sea atribuible a las diferencias en el material vegetal y el sistema de marcadores utilizados.

Respecto al empleo de la genómica en la producción de híbridos, los avances han sido muy significativos en maíz. Los primeros trabajos de predicción de rendimiento en cruza simples se realizaron con el empleo del método del mejor predictor lineal e insesgado (BLUP) en combinación con marcadores moleculares RFLP's (Bernardo, 1994), posteriormente con microsatélites (Balestre *et al.*, 2010), y recientemente con SNP's (Massman *et al.*, 2013), mediante regresión ridge BLUP (RR-BLUP). La característica del método RR-BLUP radica en el uso de todos los marcadores moleculares y asume que cada uno de éstos representa la misma proporción de la varianza genética (Meuwissen *et al.*, 2001).

Hernández-Ibáñez *et al.* (2014), son los únicos que han reportado la predicción de rendimiento en híbridos de cruza simple de tomate mediante el método BLUP y marcadores ISSR (Inter secuencias simples repetidas), con correlaciones significativamente altas entre valores predichos y observados.

#### **2.4 Necesidades de investigación**

Como se ha expuesto, son pocos los trabajos reportados de aplicación de la genómica en el mejoramiento de tomate, por lo que existe aún un campo amplio por investigar. En virtud de que tanto para la producción de variedades de polinización libre como de variedades híbridas F<sub>1</sub>, es necesario disponer de líneas puras, se requiere generar metodologías que permitan aplicar la SG en los principales métodos de selección de líneas endogámicas (Pedigrí y uniseminal);

con el fin de acortar el periodo de tiempo necesario para llegar a liberar una nueva variedad al mercado.

Otro aspecto aún pendiente de desarrollar es concretar la construcción de un chip que incluya todos aquellos SNP's plenamente identificados por marcar caracteres de interés agronómico (resistencia a enfermedades, producción de metabolitos de interés, rendimiento y sus componentes, resistencia a condiciones ambientales adversas, etc.), para de una manera económica y práctica genotipar grandes volúmenes de genotipos y seleccionar los materiales superiores en las etapas iniciales del mejoramiento, con el consecuente ahorro de tiempo y dinero.

## 2.5 LITERATURA CITADA

- Balestre, M., Von Pinho, R. G., & Souza, J. C. (2010). Prediction of maize single-cross performance by mixed linear models with microsatellite marker information. *Genetics and Molecular Research* 9(2), 1054–1068.
- Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Science* 34(1), 20–25.
- Bernardo, R. (2010). *Breeding for Quantitative Traits in Plants*. Second Edition. Stemma Press. Minnesota, USA. 380 p.
- Bojinov, B. M., & Danailov, Z. P. (2008, September). Applicability of ISSRs for genotype identification in a tomato breeding collection. In *IV Balkan Symposium on Vegetables and Potatoes 830* (pp. 63-70).
- Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Van Der Knaap, E., Cong, B., ... & Tanksley, S. D. (2000). fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289(5476), 85-88.
- Heffner, E. L., Lorenz, A. J., Jannink, J. L., & Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop science*, 50(5), 1681-1690.
- Heffner, E. L., Sorrells, M. E., & Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Science*, 49(1), 1-12.

- Hernández-Bautista, A., Lobato-Ortiz, R., García-Zavala, J. J., Parra-Gómez, M. A., Cadeza-Espinosa, M., Canela-Doñan, D., ... & Chávez-Servia, J. L. (2016). Implications of genomic selection for obtaining F<sub>2:3</sub> families of tomato. *Scientia Horticulturae*, 207, 7-13.
- Hernández-Ibáñez, L., Sahagún-Castellanos, J., Rodríguez-Pérez, J. E., Peña-Ortega, M. G., & Rodríguez-Martínez, L. M. (2014). Predicción de rendimiento de fruto de híbridos de tomate con huellas genómicas. *Phyton (Buenos Aires)*, 83(2), 311-318.
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, 9(1), 1.
- Massman, J. M., Gordillo, A., Lorenzana, R. E., & Bernardo, R. (2013). Genomewide predictions from maize single-cross data. *Theoretical and applied genetics*, 126(1), 13-22.
- Menda, N., Strickler, S. R., & Mueller, L. A. (2013). Advances in tomato research in the post-genome era. *Plant Biotechnology*, 30(3), 243-256.
- Meuwissen T.H.E., Hayes B. J., & Goddard M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4), 1819–1829.
- Ranc, N., Muños, S., Xu, J., Le Paslier, M. C., Chauveau, A., Bounon, R., ... & Causse, M. (2012). Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*. *G3: Genes| Genomes| Genetics*, 2(8), 853-864.
- Ranjan, A., Ichihashi, Y., & Sinha, N. R. (2012). The tomato genome: implications for plant breeding, genomics and evolution. *Genome biology*, 13(8), 1.
- Rodríguez, G. R., Muños, S., Anderson, C., Sim, S. C., Michel, A., Causse, M., ... & van der Knaap, E. (2011). Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. *Plant physiology*, 156(1), 275-285.
- Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P. T., Nikoloski, Z., ... & Causse, M. (2014). Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant physiology*, 165(3), 1120-1132.
- Schmidt, M., Kollers, S., Maasberg-Prelle, A., Großer, J., Schinkel, B., Tomerius, A., ... & Korzun, V. (2016). Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. *Theoretical and Applied Genetics*, 129(2), 203-213.

- Shirasawa, K., Fukuoka, H., Matsunaga, H., Kobayashi, Y., Kobayashi, I., Hirakawa, H., ... & Tabata, S. (2013). Genome-wide association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato. *DNA research*, 20(6), 593-603.
- Sim, S. C., Durstewitz, G., Plieske, J., Wieseke, R., Ganai, M. W., Van Deynze, A., ... & Francis, D. M. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One*, 7(7), e40563.
- Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485(7400), 635-641.
- Yamamoto, E., Matsunaga, H., Onogi, A., Kajiya-Kanegae, H., Minamikawa, M., Suzuki, A., ... & Miyatake, K. (2016). A simulation-based breeding design that uses whole-genome prediction in tomato. *Scientific reports*, 6.
- Zhang, J., Song, Q., Cregan, P. B., & Jiang, G. L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theoretical and Applied Genetics*, 129(1), 117-130.

### **3 SELECCIÓN GENÓMICA PARA RENDIMIENTO Y FIRMEZA DE FRUTO EN LÍNEAS ÉLITE F<sub>8</sub> DE TOMATE**

Lucas Hernández-Ibáñez, Jaime Sahagún-Castellanos, Juan Enrique Rodríguez-Pérez, Margarita Gisela Peña-Ortega.

Instituto de Horticultura. Departamento de Fitotecnia. Universidad Autónoma Chapingo. Km. 38.5 Carretera México-Texcoco. Chapingo, Estado de México. C.P. 56230. MÉXICO.

#### **3.1 RESUMEN Y ABSTRACT**

##### **3.1.1 RESUMEN**

El tomate es una especie utilizada en estudios de genética clásica y molecular, por ser autógama diploide con genoma compacto. Su mejoramiento genético se hace mediante selección o hibridación. En ambos casos es necesario desarrollar líneas homocigóticas, proceso costoso y tardado. En esta investigación se evaluó la selección genómica de líneas élite F<sub>8</sub> de tomate (LE), desarrolladas mediante el método uniseminal y genotipificadas con marcadores de ADN del tipo Inter Secuencias Simples Repetidas (ISSR); Además, la eficiencia en la predicción del rendimiento y firmeza de fruto de tomate basada en información de pedigrí, de marcadores moleculares o de ambas fue comparada. El rendimiento de fruto comercial y total, así como la firmeza de fruto de 83 LE fueron evaluados entre 2010 y 2015. Dos grupos de LE con registros en tres ambientes fueron seleccionados, y con el método del mejor predictor lineal e insesgado genómico (G-Blup) se predijo el rendimiento. Se calcularon correlaciones entre rendimientos predichos y observados. Las correlaciones fluctuaron entre 0.0 y 0.67, valores que sugieren que este método tiene utilidad en la selección de líneas élite. El G-Blup registró consistentemente correlaciones altas al utilizar la

matriz de marcadores (que incluye el total de marcadores polimórficos). En todos los casos, la magnitud de las correlaciones siempre aumentó cuando el número de LE predictoras y los ambientes de evaluación fueron más numerosos. Por lo tanto, para alcanzar mayor precisión y confiabilidad de las predicciones con G-Blup se sugiere evaluar en campo la mayor cantidad de LE predictoras, en tantos ambientes como el presupuesto lo permita.

#### **3.1.1.1 Palabras clave**

*Solanum lycopersicum* L.; Selección genómica; G-Blup; componentes de rendimiento.

#### **3.1.2 ABSTRACT**

Tomato is a species used in classical and molecular genetic studies, because it is a diploid autogamous with a compact genome. Its genetic improvement is done by selection or hybridization. In both cases it is necessary to develop inbred lines, an expensive and time consuming process. In this research the genomic selection of tomato elite lines F<sub>8</sub> (LE), developed by the single seed descent method and genotyped with Inter Simple Sequence Repeat (ISSR) DNA markers was tested; In addition, the efficiency in the prediction of yield and firmness of tomato's fruit based on pedigree information, molecular markers or both was compared. The commercial and total fruit yield as well as the fruit firmness of 83 LE were tested between 2010 and 2015. Two groups of LE with records in three environments were selected, and with the Genomic best linear unbiased prediction method (G-Blup) the performance was predicted. Correlations between predicted and observed yields were calculated. The correlations fluctuated between 0.0 and 0.67, values that suggest that this method has utility in the selection of elite lines. The G-Blup consistently recorded high correlations when the marker array (which includes all polymorphic markers) was used. In all cases, the magnitude of the correlations always increased when the number of LE predictors and the

evaluation environments were larger. Therefore, in order to achieve greater accuracy and reliability of G-Blup predictions, it is suggested to evaluate in the field the greater number of LE predictors, in as many environments as the budget allows.

### **3.1.2.1 Keywords**

*Solanum lycopersicum* L.; Genomic selection; G-Blup; Performance components.

## **3.2 INTRODUCCIÓN**

Desde que se desarrollaron los primeros marcadores moleculares de ADN, se despertó el interés de los mejoradores por incorporar su uso en los programas de fitomejoramiento. Los marcadores iniciales presentaron dos serios inconvenientes: alto costo y baja densidad. Ésta se refiere a un número reducido de marcadores, que implica que segmentos grandes del genoma no son muestreados, lo que se asocia a una baja precisión para identificar caracteres de tipo cuantitativo (QTL's). En años recientes, los avances en genotipo han reducido los costos de los marcadores e incrementado la cobertura del genoma (Heffner *et al.*, 2010).

Entre los primeros usos prácticos de los marcadores moleculares está la selección asistida (MAS). En esta selección se identifica marcadores asociados a caracteres de interés, lo que a su vez permite la selección de genotipos superiores. El método es eficiente para genes de efectos mayores, pero es poco efectivo para caracteres controlados por muchos genes con efectos muy pequeños. Ante esta situación, surgió la selección genómica (SG), nuevo paradigma que utiliza la información de todos los marcadores de que se disponga para estimar los valores reproductivos (genomic estimated breeding values GEBV) para caracteres complejos (Heffner *et al.*, 2009). Si bien, la precisión en

las predicciones basadas en SG en varios casos no ha sido muy alta, con ella se podría acelerar dramáticamente la ganancia genética mediante el acortamiento de los ciclos de mejoramiento (Heffner *et al.*, 2010), debido a una mayor respuesta a la selección por unidad de tiempo (Bernardo, 2010). Con estas perspectivas, la SG ha sido estudiada en diversos caracteres y cultivos. Zhang *et al.* (2016) predijeron caracteres de calidad maltera en cebada, y Schmidt *et al.* (2016), lo hicieron para peso de semilla de soya; en ambos casos con excelentes resultados.

En tomate hay pocos reportes de estudios que apliquen la SG. Hernández-Bautista *et al.* (2016) evaluaron rendimiento de fruto, peso de fruto promedio, diámetros polar y ecuatorial, número de lóculos y sólidos solubles totales. Con base en genotipo con marcadores de secuencias simples repetidas (SSR's), mediante los métodos Regresión Lineal Múltiple, Regresión Ridge y Bayes Lasso, hicieron predicciones para selección en familias F<sub>2:3</sub>. Concluyeron que la selección fenotípica fue más eficiente que la selección genómica. Yamamoto *et al.* (2016) mediante marcadores de polimorfismo de un solo nucleótido (SNP's), realizaron predicción y simulación para mejorar el rendimiento y sabor de tomate en variedades de fruto grande; fenotipearon 20 caracteres y emplearon siete métodos de predicción. La regresión Ridge y Bayes Lasso predijeron rendimientos, pesos de fruto promedio y sólidos solubles totales más altos que los reportados por Hernández-Bautista *et al.* (2016). Posiblemente ello sea atribuible a las diferencias en el material vegetal, el sistema de marcadores utilizados y su densidad.

Como se ha expuesto, son pocos los trabajos reportados de aplicación de la genómica en el mejoramiento de tomate. En virtud de que tanto para la producción de variedades de polinización libre como variedades híbridas F<sub>1</sub> es necesario disponer de líneas puras, se requiere implementar metodologías que permitan aplicar la SG en los principales métodos de selección de líneas endogámicas, con el fin de acortar el periodo de tiempo necesario para llegar a

liberar una nueva variedad al mercado. Por lo anterior, esta investigación tuvo por objetivos: a) Evaluar la selección genómica en líneas élite F<sub>8</sub> de tomate desarrolladas mediante el método uniseminal y genotipificadas con marcadores de ADN del tipo inter secuencias simples repetidas (ISSR's); y b) comparar la eficiencia en la predicción del rendimiento y firmeza de fruto de tomate asociada al uso de información de pedigrí, de marcadores o de ambas.

### **3.3 MATERIALES Y MÉTODOS**

La presente investigación se diseñó en el Programa de Mejoramiento Genético de Tomate de la Universidad Autónoma Chapingo; y se realizó de febrero de 2010 a noviembre de 2015, en las instalaciones del campo agrícola experimental de esta Universidad.

#### **3.3.1 Evaluación de rendimiento y firmeza de fruto de líneas élite**

Para determinar el rendimiento y firmeza de fruto de 83 líneas élite (LE) de tomate cultivado, éstas se sembraron en febrero de 2010, 2011, 2014 y 2015, en charolas de poliestireno de 200 cavidades con turba como sustrato. El trasplante a bolsas se hizo 35 días después de la siembra (dds) en invernadero bajo un sistema hidropónico. Se usaron bolsas de polietileno negro con capacidad de 18 L y sustrato de espuma volcánica. La solución nutritiva utilizada fue la propuesta por Cadahia (2000) para tomate; la cantidad aplicada varió de acuerdo con la etapa fenológica y las condiciones climáticas. La densidad de plantación fue de 3.3 plantas·m<sup>-2</sup>.

La unidad experimental consistió en dos macetas, cada una con dos plantas. Se empleó un diseño experimental de bloques completos al azar con tres repeticiones. El rendimiento de fruto se evaluó en los primeros cuatro racimos. Las variables respuesta registradas fueron: rendimiento de fruto total (RFT) en kg·m<sup>-2</sup>, rendimiento de fruto comercial (RFC) en kg·m<sup>-2</sup> y firmeza del fruto (FF)

registrada en kgf·cm<sup>-2</sup>. Un fruto alcanzó la categoría de fruto comercial si su peso fue mayor a 70 g en tipo saladette y 100 g en tipo bola y no presentaba daño físico o mecánico, desorden fisiológico o afectación por plagas y enfermedades que demeritaran su apariencia. La FF se cuantificó en cuatro frutos por unidad experimental, doce días después de haber sido cosechados en la etapa de maduración rompiente y haber permanecido a temperatura ambiente. Se utilizó un penetrómetro de fruta manual modelo GY-1 (Sundoo Instruments).

Se determinaron dos grupos de LE para realizar la selección genómica. El primero constó de 17 líneas, con registros de rendimiento y firmeza de fruto en los ambientes 2010, 2011 y 2015. El segundo se integró con 28 LE evaluadas en 2010, 2014 y 2015 (Cuadro 3.1).

**Cuadro 3.1. Líneas élite F<sub>8</sub> de tomate cultivado, empleadas en la selección genómica.**

GRUPO 1				GRUPO 2							
Núm.	Línea	TF	HC	Núm.	Línea	TF	HC	Núm.	Línea	TF	HC
1	L10DB	B	D	1	L03IS	S	I	18	L65IB	B	I
2	L11IS	S	I	2	L04IS	S	I	19	L68IB	B	I
3	L14DB	B	D	3	L05IB	B	I	20	L69DB	B	D
4	L25DS	S	D	4	L06IB	B	I	21	L71IS	S	I
5	L28MIB	B	I	5	L11IS	S	I	22	L74IB	B	I
6	L42IB	B	I	6	L26SDS	S	D	23	L75IB	B	I
7	L43IB	B	I	7	L31IS	S	I	24	L76IB	B	I
8	L44IS	S	I	8	L35DM	S	D	25	L79IS	S	I
9	L51DB	B	D	9	L37IS	S	I	26	L80IB	B	I
10	L52DB	B	D	10	L45IB	B	I	27	L88IB	B	I
11	L58DB	B	D	11	L46IB	B	I	28	L91IS	S	I
12	L60IB	B	I	12	L47IB	B	I				
13	L66MIS	S	I	13	L54IB	B	I				
14	L72IS	S	I	14	L55DB	B	D				
15	L73IS	S	I	15	L59IB	B	I				
16	L76IB	B	I	16	L61IB	B	I				
17	L77IS	S	I	17	L64IB	B	I				

TF=Tipo de fruto, S=saladette, B=bola; HC= Hábito de crecimiento, D=determinado, I=indeterminado.

### 3.3.2 Genotipificación de líneas élite

En 2013 se sembraron las 83 líneas élite F<sub>8</sub> de tomate cultivado y 8 líneas del tipo silvestre en charolas de poliestireno de 200 cavidades con turba como sustrato. Entre 30 y 40 dds, de 10 plántulas se obtuvieron muestras de tejido de hojas tiernas, verdes y sin daño aparente, para obtener su ADN y genotipificarlas mediante 29 iniciadores ISSR; las condiciones de extracción de ADN y PCR se describen con detalle en Hernández-Ibáñez *et al.* (2017). Se construyó una matriz de datos binarios de las huellas genómicas obtenidas, definida como la matriz de marcadores (X) de dimensiones  $n \times m$ , donde  $n$  es el número de LE y  $m$  el número de marcadores polimórficos empleados.

Se calcularon los coeficientes de coancestría ( $f_{ij}$ ) con base en ISSR (Bernardo, 1993), de acuerdo con la expresión:

$$f_{ij} = [S_{ij} - 0.5(S_{i-} + S_{j-})] / [1 - 0.5(S_{i-} + S_{j-})]$$

donde  $S_{ij}$  es la similaridad de Jaccard estimada con marcadores moleculares entre las LE  $i$  y  $j$ ,  $S_{i-}$  y  $S_{j-}$  son la proporción promedio de variantes ISSR que comparte la línea  $i$  o la línea  $j$  con todas las líneas silvestres, respectivamente. Como por definición los coeficientes de coancestría no son negativos, aquéllos que así resultaron se igualaron a cero. La matriz generada se denominó matriz de pedigrí (A) con dimensiones  $n \times n$ .

### 3.3.3 Modelo estadístico para predicción genómica de rendimiento y firmeza de fruto de líneas élite

La selección genómica se practicó en los dos conjuntos de líneas, mediante el método del mejor predictor lineal e insesgado genómico (G-Blup) (VanRaden, 2008) con el modelo:

$$y = \mu 1 + u + e$$

donde  $y$  es un vector  $n \times 1$  de valores fenotípicos observados de las líneas,  $\mu$  es la media general,  $\mathbf{1}$  es un vector  $n \times 1$  con todos los elementos igual a 1,  $u$  es el vector  $n \times 1$  de valores reproductivos. En teoría  $u \sim NM(0, \sigma^2_{\beta} XX')$ ;  $e$  es un vector de errores aleatorios de dimensión  $n \times 1$  y  $e \sim NM(0, \sigma^2_e I)$ . En el caso de utilizar los coeficientes de pedigrí,  $u \sim NM(0, \sigma^2_{\beta} A)$ .

El modelo que integra la información de marcadores y pedigrí se expresa como:

$$y = \mu \mathbf{1} + X\beta + \theta + e$$

donde  $y$ ,  $\mu$  y  $e$  ya fueron definidos;  $\beta$  es el vector  $m \times 1$  de efecto de los marcadores moleculares y  $\theta$  es el vector  $n \times 1$  efecto de considerar las relaciones de parentesco entre individuos,  $\beta \sim NM(0, \sigma^2_{\beta} I)$ ,  $\theta \sim NM(0, \sigma^2_{\theta} A)$  y  $e \sim NM(0, \sigma^2_e I)$ . Además  $\sigma^2_{\beta}$ ,  $\sigma^2_{\theta}$  y  $\sigma^2_e$  representan componentes de varianza supuestos positivos.

### 3.3.3.1 Validación cruzada

La habilidad predictiva del método G-BLUP se determinó mediante validación cruzada. Las  $n$  LE se dividieron al azar en dos grupos, el primero se denominó conjunto de prueba conformado por  $n_t$  líneas ( $n_t = 12, 8$  y  $4$  para el grupo de 17 LE y  $n_t = 21, 14$  y  $7$  para el grupo de 28 LE), el segundo fue el conjunto de entrenamiento con las LE restantes ( $n - n_t$ ). El conjunto de entrenamiento sirvió para ajustar el modelo y predecir los valores de la variable respuesta de las líneas del conjunto de prueba. Para verificar la estabilidad de las predicciones, se realizaron 1000 muestreos aleatorios y se predijo el rendimiento de fruto de las  $n_t$  líneas en cada uno de los casos. Posteriormente, para cada valor de  $n_t$  se calcularon las correlaciones entre los rendimientos observados y los predichos correspondientes. El procedimiento descrito se aplicó en las tres variables.

La predicción con G-BLUP se hizo con base en la librería de funciones BGLR (Pérez & de los Campos, 2014) del paquete estadístico R.

### 3.4 RESULTADOS Y DISCUSIÓN

#### 3.4.1 Caracterización fenotípica y genotípica de las líneas élite

En el conjunto conformado por 17 LE, no se registró diferencias significativas entre genotipos para las variables RFT, RFC y FF. Líneas como la L76IB, L51DB y L10DB produjeron más kilogramos de fruto total y comercial, sin embargo presentaron valores de firmeza bajos. Las líneas L58DB, L66MIS y L14DB registraron la mejor FF. En 2015 se presentó el mayor RFT, seguido de 2010 y con los rendimientos más bajos en 2011 (Cuadro 3.2). El RFC más alto se consiguió en 2015, y fue igual estadísticamente en 2010 y 2011. La FF se evaluó sólo en 2010 y 2015, con los valores más altos el último año.

**Cuadro 3.2. Estadísticas básicas de tres variables estudiadas, registradas en tres ambientes, para 17 líneas élite de tomate.**

<i>Estadístico</i>	RFC (Kg·m <sup>2</sup> )				RFT (Kg·m <sup>2</sup> )				FF (Kgf·cm <sup>2</sup> )		
	Ambiente				Ambiente				Ambiente		
	2010	2011	2015	AP	2010	2011	2015	AP	2010	2015	AP
Media	5.70b	5.90b	8.68a	6.76	8.83b	7.09c	11.19a	9.04	6.46b	13.51a	9.98
D.E.	1.40	1.44	1.81	0.82	1.96	1.60	2.13	1.31	2.35	6.24	3.63
CV	24.5	24.4	20.9	12.1	22.2	22.6	19.1	14.5	36.4	46.2	36.3
Mín	3.58	3.99	5.72	5.89	5.24	4.87	7.99	6.93	2.28	7.79	6.22
Máx	8.30	8.73	12.52	8.78	12.34	10.85	14.89	11.81	9.50	28.13	18.81

RFT: rendimiento de fruto total, RFC: rendimiento de fruto comercial, FF: firmeza de fruto, AP: promedio de ambientes, D.E.: desviación estándar, CV: coeficiente de variación, Mín: mínimo y Máx: máximo. Letras diferentes entre columnas de la misma variable indican diferencias significativas ( $P \leq 0.05$ ) según prueba de Tukey para medias.

El análisis de varianza combinado del segundo conjunto de líneas, conformado por 28 LE, presentó diferencias estadísticas altamente significativas para genotipos y ambientes en las tres variables analizadas. Las líneas L76IB y L11IS registraron el mayor RFT y RFC, pero al mismo tiempo los valores más bajos de FF. La línea L26SDS fue la que menos fruto comercial y total produjo. Las líneas L69DB, L64IB y L65IB fueron las de mayor FF. En 2015 el RFT y RFC fue más alto y en 2010 el más bajo. En 2014 la FF registrada fue mayor que en 2010

(Cuadro 3.3). En los dos conjuntos de LE y las tres variables, los coeficientes de variación fueron altos; sin embargo, se pueden considerar aceptables debido a que los caracteres evaluados son muy influenciados por el ambiente.

**Cuadro 3.3. Estadísticas básicas de tres variables estudiadas, registradas en tres ambientes, para 28 líneas élite de tomate.**

<i>Estadístico</i>	RFC (Kg·m <sup>2</sup> )				RFT (Kg·m <sup>2</sup> )				FF (Kgf·cm <sup>2</sup> )		
	Ambiente				Ambiente				Ambiente		
	2010	2014	2015	AP	2010	2014	2015	AP	2010	2014	AP
Media	4.57b	6.05a	6.64a	5.75	7.13c	8.44b	9.96a	8.51	6.40b	9.27a	7.95
D.E.	1.49	2.30	2.28	1.49	1.96	2.29	3.02	1.90	3.39	3.01	3.50
CV	32.6	38.0	34.3	25.9	27.5	27.2	30.3	22.4	52.9	32.5	44.0
Mín	1.42	1.36	1.72	1.86	2.29	4.11	4.67	3.74	2.28	5.21	2.28
Máx	6.97	10.64	10.00	8.12	12.34	14.40	15.59	12.85	15.60	18.33	18.33

RFT: rendimiento de fruto total, RFC: rendimiento de fruto comercial, FF: firmeza de fruto, AP: promedio de ambientes, D.E.: desviación estándar, CV: coeficiente de variación, Mín: mínimo y Máx: máximo. Letras diferentes entre columnas de la misma variable indican diferencias significativas ( $P \leq 0.05$ ) según prueba de Tukey para medias.

Con los 29 iniciadores ISSR se amplificaron 276 bandas, de las cuales 215 fueron polimórficas, para las 91 LE cultivadas y silvestres. El grupo de 17 LE contó sólo con 148 marcadores polimórficos, por tal razón la matriz X fue de dimensiones 17x148. Los coeficientes de coancestría de la matriz A, de dimensiones 17x17, fluctuaron entre 0 (para varias combinaciones de líneas) hasta 0.66 (entre las líneas L42IB y L43IB), con un promedio general de 0.118. Por su parte el grupo de 28 LE, contó con 156 marcadores polimórficos (matriz  $X_{28 \times 156}$ ) y sus coeficientes de coancestría (matriz  $A_{28 \times 28}$ ) fluctuaron entre 0 y 0.46 (entre las líneas L64IB y L65IB), con una media general de 0.06.

### 3.4.2 Correlaciones entre rendimientos observados y predichos

En las 17 líneas de un grupo se observó un fuerte efecto ambiental, dado que las predicciones de rendimiento con G-Blup variaron considerablemente de un ambiente a otro, independientemente de si se empleó la información de marcadores moleculares, de pedigrí o su combinación. Para la variable RFT las correlaciones entre valores predichos y observados osciló entre 0.42 y 0.66 con datos de 2010; para 2015 de 0.02 a 0.2, y para datos de 2011 y para promedios de ambientes las correlaciones fueron negativas (Cuadro 3.4). Al aumentar el número de líneas en el conjunto de entrenamiento, el número de líneas en el conjunto de prueba se redujo ( $n_t$ ) y las correlaciones entre predichos y observados aumentó. Lo mismo ocurrió al utilizar la información de marcadores y pedigrí en forma conjunta.

La variable RFC mostró una tendencia similar al RFT. Sin embargo, la magnitud de las correlaciones entre rendimientos observados y predichos fue menor, en 2010 fluctuó entre 0.31 y 0.40, para 2015 fue entre 0.0 y 0.25, y en 2011 y el promedio de ambientes las correlaciones fueron negativas. En esta variable los valores más altos entre predichos y observados se lograron con 9 líneas en el conjunto de entrenamiento y 8 en el de prueba.

Las correlaciones promedio en FF fueron positivas en los ambientes 2010 y 2011, en el primero fluctuaron entre 0.38 y 0.66 y en 2011 se redujeron y variaron entre 0.05 y 0.26 (Cuadro 3.4). En esta variable el método G-Blup que combinó información de marcadores y pedigrí fue más consistente.

En el grupo de 28 LE, como en el grupo de 17, también se observó un marcado efecto ambiental en la predicción del rendimiento de fruto mediante el G-Blup. Las correlaciones entre rendimientos observados y predichos en RFT y RFC fueron más altas en 2014, se redujeron en 2015 y registraron los valores más bajos en 2010, incluso negativos para RFC (Cuadro 3.5). Las correlaciones en el RFT fluctuaron entre 0.27 y 0.48 para el promedio de los tres ambientes, con

ligeras variaciones entre las obtenidas con base en información de marcadores, pedigrí o combinada en el modelo.

**Cuadro 3.4. Correlaciones promedio de la validación cruzada para predicción de líneas élite de tomate en tres variables y tres ambientes de producción. Los datos corresponden a 1,000 muestras aleatorias independientes de 17 líneas.**

$n_t$	RFC (Kg·m <sup>2</sup> )				RFT (Kg·m <sup>2</sup> )				FF (Kgf·cm <sup>2</sup> )		
	Ambiente				Ambiente				Ambiente		
	2010	2011	2015	AP	2010	2011	2015	AP	2010	2011	AP
<b>Marcadores moleculares:</b>											
4	0.329	-0.129	0.226	-0.427	0.657	-0.404	0.018	-0.266	0.658	0.092	0.255
8	0.396	-0.109	0.183	-0.272	0.638	-0.286	0.022	-0.224	0.555	0.108	0.147
12	0.364	-0.047	0.080	-0.157	0.474	-0.187	0.061	-0.089	0.399	0.059	0.114
<b>Pedigrí:</b>											
4	0.366	-0.137	0.046	-0.299	0.663	-0.155	0.195	-0.129	0.595	0.182	0.224
8	0.373	-0.046	0.079	-0.266	0.556	-0.073	0.191	-0.049	0.503	0.210	0.175
12	0.296	-0.036	-0.008	-0.171	0.420	-0.066	0.135	-0.008	0.379	0.062	0.079
<b>Marcadores moleculares y pedigrí:</b>											
4	0.363	-0.103	0.247	-0.381	0.674	-0.363	0.113	-0.097	0.562	0.259	0.188
8	0.380	-0.017	0.224	-0.201	0.651	-0.230	0.100	-0.138	0.528	0.125	0.238
12	0.308	-0.036	0.125	-0.177	0.512	-0.158	0.048	-0.046	0.385	0.047	0.126

$n_t$ : número de líneas élite predichas, RFT: rendimiento de fruto total, RFC: rendimiento de fruto comercial y FF: firmeza de fruto, AP: promedio de ambientes.

Las correlaciones entre predichos y observados en RFC fueron de menor magnitud que las de RFT. En 2014 y 2015 los valores fueron ligeramente mayores a los calculados para el ambiente promedio. En cuanto al empleo de marcadores, pedigrí o ambos, ninguno mostró marcada superioridad.

Las correlaciones estimadas para FF, fluctuaron entre 0.20 y 0.51. Los mejores registros los produjo la información de pedigrí.

El método G-Blup consistentemente produjo mejores estimaciones al aumentar el número de líneas del conjunto de entrenamiento, por ello se observa que con  $n_t = 7$  LE (25 % del total) las correlaciones promedio entre predichos y observados para RFT, RFC y FF fueron mayores, con algunas excepciones donde el valor más alto fue al emplear  $n_t = 14$  (Figura 3.1).

**Cuadro 3.5. Correlaciones promedio de la validación cruzada para predicción de líneas élite de tomate en tres variables y tres ambientes de producción. Los datos corresponden a 1,000 muestras aleatorias independientes de 28 líneas.**

$n_t$	RFC (Kg·m <sup>2</sup> )				RTF (Kg·m <sup>2</sup> )				FF (Kgf·cm <sup>2</sup> )
	Ambiente				Ambiente				Ambiente
	2010	2014	2015	AP	2010	2014	2015	AP	AP
<b>Marcadores moleculares:</b>									
7	-0.124	0.348	0.224	0.196	0.279	0.422	0.232	0.458	0.346
14	-0.067	0.254	0.230	0.195	0.217	0.371	0.257	0.468	0.310
21	-0.035	0.157	0.155	0.114	0.120	0.271	0.221	0.350	0.221
<b>Pedigrí:</b>									
7	-0.296	0.191	0.250	0.205	0.085	0.401	0.211	0.483	0.514
14	-0.207	0.197	0.173	0.154	0.094	0.346	0.223	0.416	0.416
21	-0.151	0.098	0.125	0.091	0.038	0.258	0.152	0.265	0.302
<b>Marcadores moleculares + pedigrí:</b>									
7	-0.153	0.294	0.298	0.182	0.224	0.480	0.235	0.476	0.425
14	-0.122	0.191	0.230	0.196	0.205	0.403	0.253	0.466	0.371
21	-0.053	0.131	0.172	0.114	0.140	0.291	0.195	0.322	0.195

$n_t$ : número de líneas élite predichas, RFT: rendimiento de fruto total, RFC: rendimiento de fruto comercial y FF: firmeza de fruto, AP: promedio de ambientes.

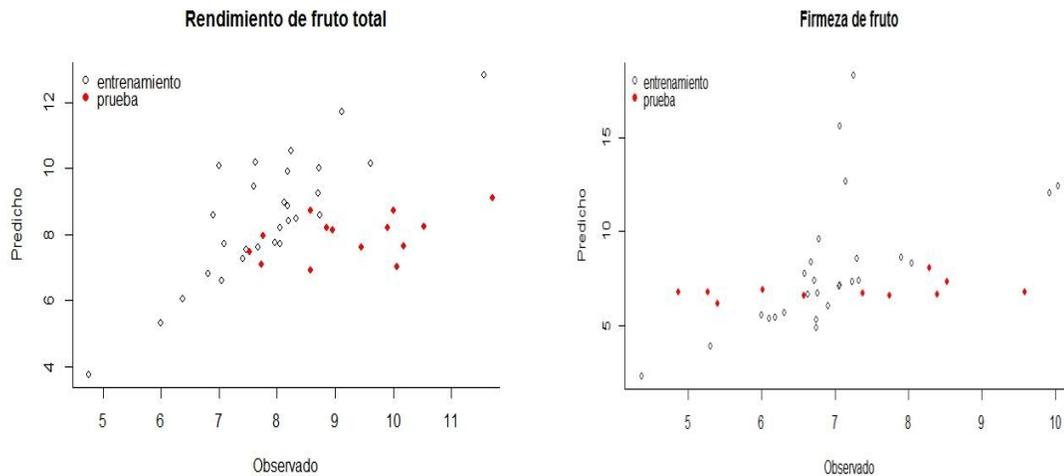


Figura 3.1. Gráfica de valores genéticos predichos para los conjuntos de entrenamiento y prueba con  $n_t=14$ . Modelo G-Blup que emplea marcadores moleculares en rendimiento de fruto total (izquierda); y G-Blup utilizando información de pedigrí con la variable firmeza de fruto (derecha).

El tomate ha sido utilizado ampliamente como especie modelo en estudios de genética clásica y molecular, entre otras razones por ser una especie autógama diploide ( $2n = 24$ ) y con un genoma relativamente compacto (~950 Mb) (The tomato genome consortium, 2012). El mejoramiento genético de esta especie ha sido por selección e hibridación; en ambos casos es necesario desarrollar líneas homocigóticas. Este proceso es largo y requiere la inversión de fuertes cantidades de recursos financieros, materiales y de mano de obra, razón por la que desde el surgimiento de los marcadores moleculares de ADN (MM) ha existido el interés por utilizarlos para mejorar el tomate de manera más eficiente. En este sentido, esta investigación ha explorado el uso de marcadores ISSR en combinación con el método G-Blup para hacer selección genómica en líneas élite.

Las variables analizadas (RFT, RFC y FF) son cuantitativas por lo que su expresión se afecta significativamente por el ambiente. Las LE fueron cultivadas para su evaluación durante cuatro ciclos agrícolas y la respuesta observada, de

los dos grupos de líneas empleados, fue diferente en cada ambiente. La estrategia de sembrar un genotipo en varios ambientes para obtener una predicción más precisa (Bernardo, 2010) funcionó en el conjunto 2, donde las correlaciones para RFT, RFC y FF para el promedio de ambientes fueron superiores a las de ambientes individuales. Esto no ocurrió con el primer conjunto de líneas, donde fue común obtener correlaciones promedio negativas.

Las diferencias de rendimiento en tomate entre ambientes coinciden con lo reportado por Hernández-Bautista *et al.* (2016), quienes obtuvieron predicciones superiores en su primer ambiente, de los dos analizados.

Es posible que además del efecto ambiental implicado, el bajo nivel predictivo en el grupo de 17 LE se deba también a los genotipos mismos y al tamaño reducido de MM, ya que el número de marcadores empleados en este grupo fue menor (148 MM) que en el dos (156 MM), lo anterior representa un marcador cada 6.42 Mb y 6.09 Mb, respectivamente. Esto implica regiones muy grandes del genoma sin cubrir, lo que limita la posibilidad de capturar los efectos genéticos buscados. El uso de una plataforma de marcadores de mayor densidad puede ser la opción ideal, Duangjit *et al.* (2016) con regresión ridge Blup (rr-Blup) y 2,313 marcadores SNP distribuidos a 0.1 cM obtuvieron correlaciones medias de 0.61 y 0.81 para firmeza y rendimiento de fruto, respectivamente, superiores ambas a las reportadas en esta investigación.

En cuanto al método G-Blup, la inclusión de sólo marcadores moleculares dio resultados ligeramente superiores al empleo de la información de pedigrí y fue superado por poco margen cuando se emplearon los marcadores y pedigrí. Lo anterior se puede deber a que la matriz A fue construida a partir de los marcadores moleculares. Esta ventaja, útil cuando se desconocen los antecedentes genéticos de las líneas, implica una redundancia en la información, es decir, que el modelo no fue capaz de extraer mayor información que la que aportan los marcadores mismos. En este sentido sería recomendable usar el

modelo que utiliza únicamente la matriz de marcadores ya que es eficiente y menos demandante de recursos informáticos para realizar los cálculos. El modelo G-Blup empleado representa una buena alternativa para selección genómica, ya que las correlaciones entre predichos y observados para rendimiento total de fruto en general fluctuaron entre 0.04 y 0.48 en el conjunto de 28 LE, en tanto que Hernández-Bautista *et al.* (2016) obtuvieron correlaciones entre 0.25 y 0.46 con los modelos rr-Blup y Bayes Lasso.

Finalmente, conviene precisar que para obtener las mejores predicciones es recomendable evaluar en campo como mínimo el 50 % de las líneas y en tantos ambientes como el presupuesto lo permita. En los dos grupos de LE analizados, el aumento en la correlación media ocurrió cuando en el conjunto de entrenamiento se incluyeron más LE.

### **3.5 CONCLUSIONES**

Los coeficientes de correlación entre valores observados y predichos para RFT, RFC y FF, estimados mediante G-BLUP en LE de tomate, fluctuaron entre 0.0 y 0.67, valores que sugieren que este método tiene utilidad en la selección de líneas élite.

El G-BLUP produjo consistentemente resultados altos de correlación entre valores observados y predichos, al utilizar la matriz de marcadores y la combinación marcadores-pedigrí.

En todos los casos, la magnitud de las correlaciones siempre aumentó cuando se incrementó el número de LE del conjunto de entrenamiento. Por lo tanto, para alcanzar mayor precisión y confiabilidad de las predicciones con G-BLUP se deben evaluar en campo la mayor cantidad de LE predictoras en, preferiblemente, varios ambientes.

### 3.6 AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) de México, por la beca doctoral otorgada al primer autor. Al Q.F.B. Ricardo Gaspar Hernández por el soporte técnico en marcadores moleculares de ADN.

### 3.7 REFERENCIAS

- Bernardo, R. (1993). Estimation of coefficient of coancestry using molecular markers in maize. *Theoretical and Applied Genetics* 85, 1055–1062. doi: 10.1007/BF00215047.
- Bernardo, R. (2010). *Breeding for Quantitative Traits in Plants*. Second Edition. Stemma Press. Minnesota, USA. 380 p.
- Cadahia, C. (2000). *Fertirrigación, cultivos hortícolas y ornamentales* (2da. Ed.). Madrid, España: Mundi Prensa.
- Heffner, E. L., Lorenz, A. J., Jannink, J. L., & Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop science*, 50(5), 1681-1690.
- Heffner, E. L., Sorrells, M. E., & Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Science*, 49(1), 1-12.
- Hernández-Bautista, A., Lobato-Ortiz, R., García-Zavala, J. J., Parra-Gómez, M. A., Cadeza-Espinosa, M., Canela-Doñan, D., Cruz-Izquierdo, S. & Chávez-Servia, J. L. (2016). Implications of genomic selection for obtaining F 2: 3 families of tomato. *Scientia Horticulturae*, 207, 7-13.
- Hernández-Ibáñez, L., Sahagún-Castellanos, J., Rodríguez-Pérez, J. E., & Peña-Ortega, M. G. (2017). Prediction of fruit yield and firmness of tomato hybrids with BLUP and RR-BLUP using ISSR molecular markers. *Revista Chapingo Serie Horticultura*, 23(1), 21-33.
- Duangjit, J., Causse, M., & Sauvage, C. (2016). Efficiency of genomic selection for tomato fruit quality. *Mol. Breeding*, 36:29.doi:10.1007/s11032-016-0453-3.

- Pérez, P., & De los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198(2), 483–495. doi:10.1534/genetics.114.164442.
- Schmidt, M., Kollers, S., Maasberg-Prelle, A., Großer, J., Schinkel, B., Tomerius, A., Graner, A., & Korzun, V. (2016). Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. *Theoretical and Applied Genetics*, 129(2), 203-213.
- The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485, 635-641.
- VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91(11), 4414-4423. doi:10.3168/jds.2007-0980.
- Yamamoto, E., Matsunaga, H., Onogi, A., Kajiya-Kanegae, H., Minamikawa, M., Suzuki, A., ... & Miyatake, K. (2016). A simulation-based breeding design that uses whole-genome prediction in tomato. *Scientific reports*, 6.
- Zhang, J., Song, Q., Cregan, P. B., & Jiang, G. L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theoretical and Applied Genetics*, 129(1), 117-130.

## **4 PREDICCIÓN DE RENDIMIENTO Y FIRMEZA DE FRUTO DE HÍBRIDOS DE TOMATE MEDIANTE MARCADORES MOLECULARES**

**Lucas Hernández-Ibáñez, Jaime Sahagún-Castellanos\*, Juan Enrique Rodríguez-Pérez, Margarita Gisela Peña-Ortega.**

Instituto de Horticultura. Departamento de Fitotecnia. Universidad Autónoma Chapingo. Km. 38.5 Carretera México-Texcoco. Chapingo, Estado de México. C.P. 56230. MÉXICO. Email: jsahagunc@yahoo.com.mx (\*autor para correspondencia). Teléfono de contacto: (595) 952 1500 Ext. 6185

### **4.1 RESUMEN Y ABSTRACT**

#### **4.1.1 Resumen**

En el desarrollo de híbridos de tomate, un número grande de líneas progenitoras implica un número de híbridos posibles demasiado alto, y su formación y evaluación se vuelve un problema. La magnitud de éste puede reducirse a niveles manejables con el uso de métodos de predicción del desempeño de los híbridos. En el presente estudio se compararon dos métodos basados en huellas genómicas, teoría de los modelos mixtos y en la evaluación experimental de una muestra de híbridos: 1) el mejor predictor lineal insesgado o BLUP y 2) la regresión ridge BLUP. Se realizaron 36 cruza simples con el objetivo de determinar el número, firmeza y rendimiento total y comercial de frutos. Con 1,000 muestras aleatorias independientes de tamaños:  $n = 6, 12, 18, 24$  y  $30$ , se predijo el comportamiento de los  $36-n$  híbridos restantes. Los coeficientes de correlación entre valores predichos y observados fluctuaron entre 0.25 y 0.83. El BLUP de

manera consistente registró los valores más altos. En ambos métodos de predicción cuando  $n$  aumentó la magnitud de las correlaciones también se incrementó.

#### **4.1.1.1 Palabras clave**

*Solanum lycopersicum* L.; Mejor predictor lineal e insesgado; Regresión ridge BLUP; componentes de rendimiento.

#### **4.1.2 Abstract**

In the development of tomato hybrids, a large number of parental lines involves an excessively high number of possible hybrids, making their formation and evaluation problematic. The magnitude of this can be reduced to manageable levels with the use of hybrid performance prediction methods. In this study, two methods based on genomic fingerprints were compared, mixed model theory and the experimental evaluation of a sample of hybrids: 1) best linear unbiased prediction or BLUP and 2) ridge regression BLUP. Thirty-six single crosses were performed with the objective of determining the number, firmness and total and commercial yield of fruits. With 1,000 size-independent random samples,  $n = 6, 12, 18, 24$  and  $30$ , the behavior of the remaining  $36-n$  hybrids was predicted. The correlation coefficients between predicted and observed values ranged between  $0.25$  and  $0.83$ . BLUP consistently recorded the highest values. In both prediction methods when  $n$  increased the magnitude of the correlations also increased.

#### **4.1.2.1 Keywords**

*Solanum lycopersicum* L., best linear unbiased prediction, ridge regression BLUP, fruit yield components.

## 4.2 INTRODUCCIÓN

El manejo de la heterosis que permite la producción de variedades híbridas sobresalientes es probablemente el mayor éxito del mejoramiento genético de las plantas. El desarrollo de variedades híbridas sobresalientes, sin embargo, no es fácil: menos del 1 % de los híbridos de cruce simple de maíz (*Zea mays* L.) llegan a ser variedades comerciales (Bernardo, 1996), y algo similar debe suceder en otros cultivos, particularmente entre algunos que corresponden a especies autóгамas, como el tomate (*Solanum lycopersicum* L.). Entre los obstáculos que encuentra el fitomejorador destaca la necesidad de evaluar en el campo una cantidad elevada de híbridos experimentales, que usualmente rebasa las capacidades de los programas de mejoramiento. Para enfrentar esta dificultad se han desarrollado teoría y métodos de predicción del valor genotípico de variedades que no existen físicamente. Estos métodos se basan en la formación y evaluación experimental de una muestra pequeña de cruces, y en las coancestrías de las líneas progenitoras, estimadas con base en sus registros de pedigrí o en sus huellas genómicas (Henderson, 1985; Bernardo, 1994; Gbur *et al.*, 2012).

En maíz se ha predicho el rendimiento de grano de cruces simples mediante una ecuación de predicción que produce “los mejores predictores lineales insesgados” (BLUP, por sus siglas en inglés: best linear unbiased prediction) (Bernardo, 1993, 1995). Estas predicciones se basaron en marcadores moleculares RFLP’s (Bernardo, 1994, 1996) y en marcadores microsatélites (Balestre, Von Pinho, & Souza, 2010). Los coeficientes de correlación ( $r$ ) entre rendimientos observados y pronosticados fluctuaron entre 0.43 y 0.76, y entre 0.55 y 0.70, respectivamente. También en maíz, Massman, Gordillo, Lorenzana, y Bernardo (2013) realizaron predicción genómica mediante regresión ridge BLUP (RR-BLUP) basada en marcadores SNP (polimorfismo de nucleótido único), con niveles de predicción similares a los de BLUP. Una característica atractiva del método RR-BLUP es que utiliza todos los marcadores moleculares

y asume que cada uno de éstos representa la misma proporción de la varianza genética (Meuwissen, Hayes, & Goddard, 2001).

Aunque el tomate es la segunda especie hortícola en importancia a nivel mundial, y su mejoramiento genético en las últimas décadas se ha orientado a la producción y liberación de variedades híbridas  $F_1$ , los reportes sobre la predicción de rendimiento de híbridos son pocos. Lo anterior posiblemente se deba a que el fitomejoramiento es realizado por el sector privado y en forma limitada por programas públicos. El costo de producción del tomate es muy elevado, sobre todo si se hace en invernadero. La evaluación de campo de nuevos cultivares implica inversión financiera alta. Para hacerla más productiva se requiere implementar estrategias, como la predicción de rendimiento de fruto, que permitan hacer y evaluar sólo las cruzas con mayor probabilidad de éxito.

Mirshamsi, Farsi, Shahriari, y Nemati (2008) utilizaron marcadores RAPD para estimar distancias genéticas entre progenitores y determinar la correlación entre éstas y los rendimientos de fruto de los híbridos producidos por sus cruzas. Sin embargo, los coeficientes de correlación obtenidos, por su reducido valor ( $r < 0.40$ ), sugieren que este método es de poca utilidad práctica.

En mejoramiento genético de hortalizas, y en especial en tomate (pese a su gran valor metodológico), son escasos los antecedentes de predicción de caracteres de rendimiento. Por esta razón, se planeó una investigación con el propósito de predecir la firmeza, el número y el rendimiento total y comercial de frutos de cruzas simples de tomate mediante el empleo de los métodos BLUP y regresión ridge BLUP.

### **4.3 MATERIALES Y MÉTODOS**

La presente investigación se realizó de febrero de 2013 a noviembre de 2015 y consistió en tres etapas: 1) análisis de líneas élite de tomate con marcadores

moleculares (ISSR), selección de líneas progenitoras y realización de cruzamientos dirigidos, 2) evaluación de firmeza, número y rendimiento de fruto de las cruzas simples y 3) predicción de rendimiento de cruzas mediante BLUP y RR-BLUP y validación de los métodos.

#### **4.3.1 Selección de líneas progenitoras**

En 2013 se sembraron 39 líneas élite F<sub>8</sub> de tomate en charolas de poliestireno de 200 cavidades con turba como sustrato. Entre 30 y 40 días después de la siembra (dds), de 10 plántulas se obtuvieron muestras de tejido de hojas tiernas, verdes y sin daño aparente. Las hojas se lavaron con alcohol al 70 % y se secaron con toallas de papel. Se colocó 0.3 g de hoja en un mortero de porcelana, se agregó nitrógeno líquido y se molió hasta obtener un polvo fino; el cual se transfirió a un microtubo de 1.5 mL con 600 µL de amortiguador de extracción (Tris HCl 100 mM, EDTA 20 mM, NaCl 1.4 M, CTAB 2.0 % y 2-Mercaptoetanol al 0.2 %), se precalentó a 65 °C en termoblock y se agitó hasta homogenizar. Nuevamente se calentó a 65 °C durante 20 min. En seguida, se retiró del calor, dejó enfriar y centrifugó a 21,000 x g durante 20 min. La solución se transfirió a otro microtubo, se agregaron 600 µL de cloroformo:alcohol isoamílico (24:1) y se agitó durante 20 min. Posteriormente, se centrifugó a 12,000 x g por 5 min y la fase acuosa (superior) se colocó en otro tubo con 500 µL de isopropanol frío, se agitó suavemente por inversión, se mantuvo a -20 °C por 60 min para precipitar el ADN, se centrifugó a 12,000 x g durante 5 min y decantó el sobrenadante. A la pastilla se le agregaron 700 µL de solución para disolver (Tris HCl 10 mM, EDTA 1 mM, NaCl 10 mM, pH 8.0) y se dejó en refrigeración (4 °C) durante 12 h.

Una vez resuspendido el ADN, se agregaron 4 µL de ARNasa e incubó a 37 °C por 1 h. El ADN se reprecipitó con 70 µL de acetato de sodio 3 M y 700 µL de isopropanol frío, se agitó por inversión e introdujo al congelador por 2 h. Se centrifugó a 8,000 x g durante 10 min, se eliminó el sobrenadante y se lavó la pastilla con 500 µL de etanol al 70 % (grado reactivo). Nuevamente se centrifugó,

se eliminó el sobrenadante y secó la pastilla a temperatura ambiente. Finalmente, ésta se disolvió con 50  $\mu\text{L}$  de TE y se almacenó a 4 °C.

Para cuantificar la pureza del ADN, se empleó un espectrofotómetro modelo NanoDrop Lite (Thermo Scientific). Se prepararon soluciones de trabajo con 10  $\mu\text{g}\cdot\mu\text{L}^{-1}$  de ADN. La calidad del ADN se verificó en un gel de agarosa al 0.8 % de 4 mm de grosor y TAE al 0.5 %. En cada pozo se depositaron 10  $\mu\text{L}$  de la solución de trabajo con ADN. La electroforesis se realizó por 2 h en una cámara modelo Owl A3-1 (Thermo Scientific). El gel se tiñó con solución de bromuro de etidio en concentración de 0.5  $\mu\text{g}\cdot\text{mL}^{-1}$  durante 20 min y se fotografió con fotodocumentador modelo TFM-26 transilluminator (UVP).

En la amplificación y separación de fragmentos mediante la reacción en cadena de la polimerasa (PCR, por sus siglas en inglés) se usaron 29 iniciadores ISSR (Inter Secuencias Simples Repetidas), las temperaturas de alineamiento aparecen en el Cuadro 4.1. La reacción de amplificación se realizó con: 5.2  $\mu\text{L}$  de  $\text{H}_2\text{O}$ , 10.0  $\mu\text{L}$  de dNTP's [500  $\mu\text{M}$ ], 2.5  $\mu\text{L}$  de amortiguador [10 x], 1.5  $\mu\text{L}$  de  $\text{MgCl}_2$  [50 mM], 3.0  $\mu\text{L}$  de iniciador ISSR [10  $\text{ng}\cdot\mu\text{L}^{-1}$ ], 0.3  $\mu\text{L}$  de enzima Taq ADN polimerasa [5  $\text{u}\cdot\mu\text{L}^{-1}$ ] y 2.5  $\mu\text{L}$  de ADN [10  $\text{ng}\cdot\mu\text{L}^{-1}$ ], dando un volumen total de 25  $\mu\text{L}$ .

La PCR se llevó a cabo en un termociclador modelo FTC41H2D marca TECHNE, con el siguiente programa: desnaturalización inicial a 93 °C por 1 min; seguido de 40 ciclos, cada uno con 20 s a 93 °C, 60 s a la temperatura de alineación del iniciador probado y 20 s a 72 °C, y una extensión final a 72 °C por 6 min. La muestra se enfrió a 10 °C. Los productos finales se llevaron a electroforesis en gel de agarosa al 2 % con TAE 1x; posteriormente se realizó la tinción con bromuro de etidio y fotodocumentación.

A partir de la matriz de datos binarios obtenida de las huellas genómicas, se construyó el coeficiente de similitud de Jaccard ( $S_{rs}$ ) para cada par de genotipos (Khattree & Naik, 2000). Con la matriz de disimilitud ( $d = 1 - S_{rs}$ ), se realizó el

agrupamiento mediante el método de varianza mínima de Ward con el programa Infostat. El resultado fue la formación de tres grupos de los que se eligieron dos, considerados genéticamente diferentes. A estos dos se les denominó X y Y.

#### **4.3.2 Evaluación de rendimiento de fruto de cruzas intergrupales**

Se seleccionaron 12 líneas élite F<sub>8</sub> de alto rendimiento de fruto; seis pertenecientes al grupo X (líneas L5BI, L37SI, L59BI, L61BI, L65BI y L76BI) y seis al Y (líneas L6BI, L45BI, L68BI, L69BD, L80BI y L88BI). Éstas se sembraron en febrero de 2013 y se realizaron las 36 (6x6) cruzas simples (CS) directas intergrupales; debido a que no se han reportado efectos maternos en tomate, las líneas de los grupos X y Y se usaron como hembras y machos, respectivamente. Al alcanzar la madurez de consumo, los frutos se cosecharon y la semilla de cada craza se separó y benefició.

Con el propósito de determinar el rendimiento de fruto de las CS, éstas se sembraron en febrero de 2014 y 2015, en charolas de poliestireno de 200 cavidades con turba como sustrato. El trasplante se hizo 35 dds en invernadero bajo un sistema hidropónico. Se usaron bolsas de polietileno negro con capacidad de 18 L y sustrato de espuma volcánica. La solución nutritiva utilizada fue la propuesta por Cadahia (2000) para tomate; la cantidad aplicada varió de acuerdo con la etapa fenológica y las condiciones climáticas. La densidad de plantación fue de 3.3 plantas·m<sup>-2</sup>.

La unidad experimental consistió en dos macetas, cada una con dos plantas. Se utilizó un diseño experimental de bloques completos al azar con tres repeticiones. El rendimiento de fruto se evaluó en los primeros cuatro racimos. Las variables respuesta registradas fueron: número de frutos (NF) por metro cuadrado, rendimiento de fruto total (RFT) en kg·m<sup>-2</sup>, rendimiento de fruto comercial (RFC) en kg·m<sup>-2</sup> y firmeza del fruto (FF) registrada en kgf·cm<sup>-2</sup>. Un fruto alcanzó la categoría de fruto comercial si su peso fue mayor a 100 g y no presentaba daño físico o mecánico, desorden fisiológico o afectación por plagas y enfermedades

que demeritaran su apariencia. La FF se cuantificó en cuatro frutos por unidad experimental, doce días después de haber sido cosechados en la etapa de maduración rompiente y haber permanecido a temperatura ambiente. Se usó un penetrómetro de fruta manual modelo GY-1 (Sundoo Instruments).

### 4.3.3 Predicción de rendimiento de cruzas mediante BLUP y RR-BLUP

#### 4.3.3.1 BLUP

Se calcularon los coeficientes de coancestría ( $f_{ij}$ ) con base en ISSR (Bernardo, 1993), de acuerdo con la expresión:

$$f_{ij} = [S_{ij} - 0.5(S_{i-} + S_{j-})] / [1 - 0.5(S_{i-} + S_{j-})]$$

donde  $S_{ij}$  es la similaridad de Jaccard estimada con marcadores moleculares entre las líneas  $i$  y  $j$  del grupo X,  $S_{i-}$  proporción promedio de variantes ISSR que comparte  $i$  con todas las líneas del grupo heterótico opuesto (Y) y  $S_{j-}$  proporción promedio de variantes ISSR que comparte la línea  $j$  con todas las líneas del grupo heterótico opuesto (X). Como por definición los coeficientes de coancestría no son negativos, aquéllos que así resultaron se igualaron a cero.

El modelo para el rendimiento de fruto ( $y$ ) fue (Bernardo, 1994, Massman *et al.*, 2013):

$$y = X\beta + Z_1g_1 + Z_2g_2 + Z_3s + e$$

en donde  $y$  es un vector  $p \times 1$  de  $p$  rendimientos observados de las  $N$  cruzas simples,  $\beta$  es un vector  $c \times 1$  de efectos fijos de los  $c$  ambientes de evaluación;  $g_1$  es un vector  $n_1 \times 1$  de efectos aleatorios de aptitud combinatoria general (ACG) de las líneas del grupo X;  $g_2$  es un vector  $n_2 \times 1$  de efectos aleatorios de aptitud combinatoria general (ACG) de las líneas del grupo Y;  $s$  es un vector  $N \times 1$  de los efectos aleatorios de aptitud combinatoria específica (ACE) de las  $N$  cruzas simples;  $e$  es el vector  $p \times 1$  de residuos y  $e \sim NM(0, \sigma_e^2 I_p)$  con  $\sigma_e^2 > 0$ .  $X$ ,  $Z_1$ ,  $Z_2$ , y  $Z_3$

son matrices cuyos elementos son 0 o 1 para denotar ausencia o presencia de efectos que relacionan a  $y$  con  $\beta$ ,  $g_1$ ,  $g_2$ , y  $s$ , respectivamente.

La covarianza (Cov) entre las CS intergrupales  $ij$  e  $i'j'$  se expresó con base en un modelo de  $k$  loci y en la simplificación que resulta de la que derivó Melchinger (1988):

$$\text{Cov}(ij, i' j') = f_{ii'} V_{A(X/Y)} + f_{jj'} V_{A(Y/X)} + f_{ii'} f_{jj'} V_{D(XY)}$$

En donde  $V_{A(X/Y)}$  es la varianza aditiva de los alelos  $i$  y  $j$  provenientes del grupo X o Y a través de los  $k$  loci de  $i'$  y  $j'$  involucrados y  $V_{D(XY)}$  es la varianza de dominancia a través de los  $k$  loci de los pares alélicos formados cada uno por un alelo de X y uno de Y. Las estimaciones de máxima verosimilitud restringida fueron obtenidas mediante iteración a partir de sendas ecuaciones para la  $V_E$  (varianza no genética),  $V_{A(X/Y)}$ ,  $V_{A(Y/X)}$  y  $V_{D(XY)}$  (Henderson, 1985; Bernardo, 1994; Gbur *et al.*, 2012). Además de la predicción, las estimaciones de estas varianzas se utilizaron para obtener heredabilidades.

La predicción de los rendimientos de fruto de los híbridos se basó en la teoría de los modelos mixtos que produce predictores que son lineales, insesgados y de varianza mínima (BLUP). La predicción del rendimiento de híbridos se hizo con base en los rendimientos observados de los híbridos restantes, llamados híbridos predictores. De acuerdo con Bernardo (1994) el rendimiento ajustado por efectos fijos de estos híbridos ( $y_T$ ) se expresó como:

$$y_T = (Z_3' R^{-1} Z_3)^{-1} Z_3' R^{-1} (y - X\beta)$$

donde  $R$  es una matriz diagonal, y los elementos son el recíproco del número de ambientes del experimento. Por otra parte, el vector formado con los rendimientos predichos de los híbridos desconocidos ( $y_U$ ) fue calculado según la expresión:

$$y_U = C_{UT} C_{TT}^{-1} y_T$$

en donde  $C_{UT}$  es una matriz  $N_U \times N_T$  cuyo  $ij$ -ésimo elemento es la covarianza genética entre el  $i$ -ésimo híbrido sujeto a predicción y el  $j$ -ésimo híbrido predictor, y  $C_{TT}$  es la matriz  $N_T \times N_T$  de las covarianzas de los  $n$  híbridos predictores. Los elementos de  $C_{UT}$  y los que van fuera de la diagonal de  $C_{TT}$  se calcularon según la expresión de la covarianza entre un híbrido desconocido y uno predictor, en términos de coancestría y varianzas genéticas.

#### 4.3.3.2 RR-BLUP

La predicción del rendimiento de CS mediante la utilización del genoma completo (genomewide), se hizo por medio de RR-BLUP con el modelo lineal propuesto por Massman *et al.* (2013) para  $N_M$  marcadores:

$$y = X\beta + W_1m_1 + W_2m_2 + W_3m_3 + e$$

donde  $y$ ,  $X$ ,  $\beta$  y  $e$  fueron ya definidos,  $m_1$  es un vector  $N_M \times 1$  de efectos aleatorios de ACG de los marcadores de las líneas maternas,  $m_2$  es un vector  $N_M \times 1$  de efectos aleatorios de ACG de los marcadores de las líneas paternas y  $m_3$  es un vector  $N_M \times 1$  de efectos aleatorios de ACE asociados a los marcadores alélicos de ambos grupos heteróticos.  $W_1$  y  $W_2$  son matrices cuyos elementos fueron igual a 1 ó -1. Los elementos de  $W_3$  en un locus dado fueron iguales al producto de los elementos de  $W_1$  y  $W_2$  en el mismo locus, según lo propuesto por Massman *et al.* (2013). En teoría  $m_1 \sim NM(0, \sigma^2_{m1}I)$ ,  $m_2 \sim NM(0, \sigma^2_{m2}I)$ ,  $m_3 \sim NM(0, \sigma^2_{m3}I)$  con  $\sigma^2_{m1} > 0$ ,  $\sigma^2_{m2} > 0$  y  $\sigma^2_{m3} > 0$ , además  $m_1$ ,  $m_2$  y  $m_3$  son independientes;  $e$  es un vector de errores aleatorios de dimensión  $p \times 1$  y  $e \sim NM(0, \sigma^2_e I)$  con  $\sigma^2_e > 0$ .

Para evaluar el método de predicción BLUP en términos de tamaño de muestra, se trabajó con conjuntos de  $n$  híbridos predictores ( $n = 6, 12, 18, 24$  y  $30$ ), seleccionados al azar, de los 36 posibles. Para verificar la estabilidad de las predicciones, en cada número de híbridos predictores ( $n$ ) se realizaron 1000 muestreos aleatorios y se predijo el rendimiento de fruto de los  $36-n$  híbridos restantes en cada uno de los casos. Posteriormente, para cada valor de  $n$  se

calcularon las correlaciones entre los rendimientos observados y los predichos correspondientes. En el caso de RR-BLUP,  $n$  correspondió al conjunto de entrenamiento y  $36-n$  al conjunto de prueba. El procedimiento descrito se aplicó en las cuatro variables.

Para predecir mediante el BLUP, se construyó un programa en el módulo IML del paquete estadístico *Statistical Analysis System* (SAS). La predicción con RR-BLUP se hizo con base en la librería de funciones BGLR (Pérez & De los Campos, 2014) del paquete estadístico R.

## **4.4 RESULTADOS Y DISCUSIÓN**

### **4.4.1 Selección de líneas progenitoras mediante ISSR**

Se amplificaron 276 bandas, de las cuales 215 fueron polimórficas. El número de bandas por iniciador ISSR varió desde 2 hasta 18, con promedio de 9.52 (Cuadro 4.1). Estos valores son bajos comparados con los de Aguilera *et al.* (2011), que muestran la amplificación de 9 a 22 bandas por iniciador, con promedio de 14.4, aunque reportaron un nivel reducido de polimorfismo (34 %), con promedio de 5.3 bandas polimórficas por iniciador. En el presente trabajo se amplificaron 7.4 bandas polimórficas por ISSR con 77.9 % de polimorfismo general, con mínimo de cero y máximo de 100 %.

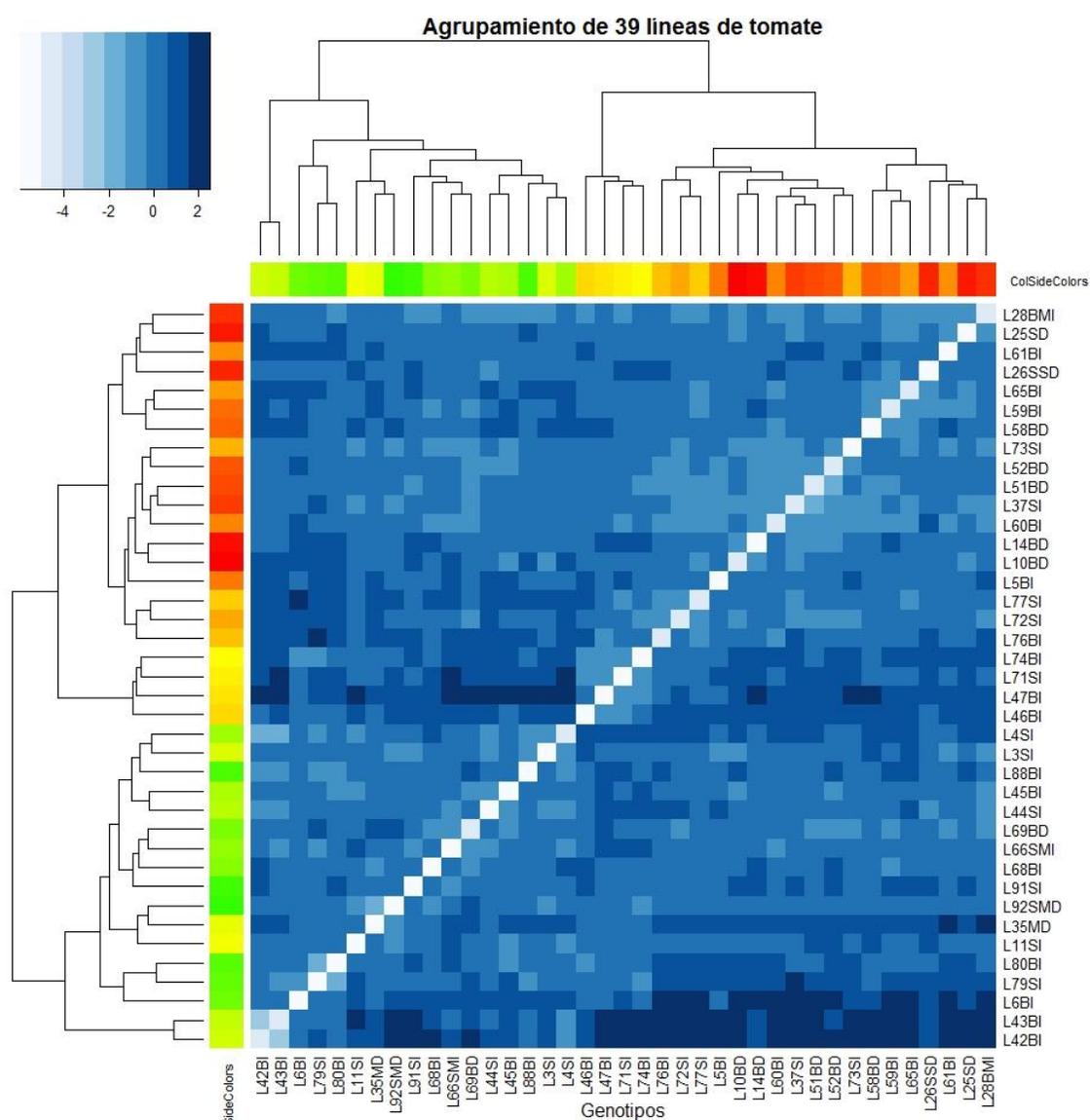
El método de agrupamiento de Ward definió tres conjuntos de líneas (Figura 4.1). El primer grupo estuvo constituido por 17 líneas (color verde a amarillo), el segundo con cuatro (tonos amarillos) y el tercero con 18 (naranja a rojo intenso). En dicha figura, los tonos de azul más claros indican distancia genética menor; por lo anterior, dentro de cada grupo predominan tonos azul claro y entre grupos tonalidades más oscuras. No se detectó asociación de los grupos generados con marcadores moleculares con alguna característica fenotípica. En cada grupo

hay diversidad en el tipo de crecimiento (determinado e indeterminado), tipo de fruto (saladette y bola) y color de fruto (amarillo, naranja y rojo).

**Cuadro 4.1.** Secuencia de iniciadores ISSR utilizados en la caracterización de la divergencia genética de 39 líneas élite de tomate, temperatura de alineamiento (Ta), número total de bandas amplificadas (TB) y polimórficas (BP) y porcentaje de polimorfismo (PP).

Iniciador	Secuencia 5'-3'	Ta (°C)	TB	BP	PP (%)	Iniciador	Secuencia 5'-3'	Ta (°C)	TB	BP	PP (%)
17898 B	(CA) <sub>6</sub> GT	42	7	4	57.1	PHV7	(GTG) <sub>2</sub> (TG) <sub>6</sub> T	58	2	1	50.0
17899 A	(CA) <sub>6</sub> AG	42	9	4	44.4	UBC807	(AG) <sub>8</sub> T	50	9	7	77.8
Echt 5	(AGAC) <sub>2</sub> GC	42	6	3	50.0	UBC811	(GA) <sub>8</sub> C	52	5	2	40.0
ISSR01	(CA) <sub>8</sub> AAGG	62	15	15	100.0	UBC815	(CT) <sub>8</sub> G	52	9	2	22.2
ISSR02	(CA) <sub>8</sub> AAGCT	62	15	15	100.0	UBC822	(TC) <sub>8</sub> A	50	3	0	0.0
ISSR03	(GA) <sub>8</sub> CTC	58	17	17	100.0	UBC827	(AC) <sub>8</sub> G	52	8	5	62.5
ISSR04	(AG) <sub>8</sub> CTC	58	18	17	94.4	UBC836	(AG) <sub>8</sub> CTA	52	11	9	81.8
ISSR05	(AC) <sub>8</sub> CTA	56	10	8	80.0	UBC844	(CT) <sub>8</sub> AC	54	9	8	88.9
ISSR06	(AC) <sub>8</sub> CTG	58	12	11	91.7	UBC845	(CT) <sub>8</sub> AGG	54	13	12	92.3
ISSR07	(AG) <sub>8</sub> CTG	53	10	10	100.0	UBC847	(CA) <sub>8</sub> AGC	52	7	3	42.9
ISSR08	(AC) <sub>8</sub> CTT	56	9	9	100.0	UBC848	(CA) <sub>8</sub> AGG	56	8	8	100.0
ISSR10	(GA) <sub>8</sub> T	50	9	8	88.9	UBC862	(AGC) <sub>6</sub>	68	9	6	66.7
LOL2	(CT) <sub>8</sub> GC	56	8	6	75.0	UBC873	(GACA) <sub>4</sub>	48	11	6	54.5
LOL9	(CAC) <sub>3</sub> GC	38	9	8	88.9	UBC886	YXY(CT) <sub>7</sub>	52	10	8	80.0
PHV6	CCA(CT) <sub>3</sub>	57	8	3	37.5	Total			276	215	77.9

Y=GAC, X=GAT.



**Figura 4.1.** Agrupamiento de 39 líneas elite de tomate, genotipeadas con 29 iniciadores ISSR, mediante la distancia de Jaccard y el método de Ward.

#### 4.4.2 Evaluación de rendimiento y firmeza de fruto de cruzas intergrupales

De las 36 cruzas intergrupales evaluadas (Cuadro 4.2), en 2014 los valores de las cuatro variables fueron superiores, mientras que en 2015 se produjeron en promedio menos frutos (7) y kilogramos de fruto total (0.94) y comercial (1.3). Sin

embargo, las cruzas se comportaron de manera similar dentro de cada ambiente. La misma craza registró el valor más alto (o más bajo) para el NF, RFC y RFT en ambas evaluaciones. En el caso de la FF, la craza con el mayor registro no fue la misma en los dos años evaluados; no obstante, la mayor FF estuvo asociada a cruzas que involucraron a la línea L68BI y la menor FF correspondió a cruzas con la línea L59BI (Cuadro 4.2).

**Cuadro 4.2.** Máximos, mínimos, medias y estimaciones de varianzas y heredabilidades de cuatro caracteres de 36 cruzas simples de tomate en dos ciclos de producción (2014 y 2015), cultivados en condiciones de hidroponía e invernadero.

Max, Min estimaciones de varianzas y $h^2$	NF	RFT ( $\text{kg}\cdot\text{m}^{-2}$ )	RFC ( $\text{kg}\cdot\text{m}^{-2}$ )	FF ( $\text{kgf}\cdot\text{cm}^{-2}$ )
Max	117.0 (L37SIxL80BI)	13.723 (L76BIxL69BD)	11.958 (L76BIxL68BI)	13.83 (L5BIxL68BI)
Min	51.7 (L76BIxL45BI)	8.153 (L59BIxL88BI)	4.373 (L37SIxL80BI)	5.98 (L59BIxL88BI)
Media	77.3	10.488	8.630	9.44
	Evaluación 2015			
Max	102.0 (L37SIxL80BI)	12.048 (L76BIxL69BD)	10.736 (L76BIxL68BI)	11.04 (L61BIxL68BI)
Min	49.2 (L76BIxL45BI)	7.358 (L65BIxL6BI)	2.658 (L37SIxL80BI)	6.81 (L59BIxL45BI)
Media	70.7	9.553	7.332	8.94
Varianzas				
$V_{ACG1}$	184.85	2.837	4.753	6.714
$V_{ACG2}$	147.53	2.083	4.120	5.881
$V_{ACE}$	39.67	0.586	1.390	1.662
$V_E$	10.77	0.192	0.371	0.491
$h^2$	0.986	0.983	0.982	0.983

Max: máximos, Min: mínimos, NF: número de frutos, RFT: rendimiento de fruto total, RFC: rendimiento de fruto comercial, FF: firmeza de fruto,  $V_{ACG1}$ : varianza de los efectos de aptitud combinatoria general de las líneas del grupo X,  $V_{ACG2}$ : varianza de los efectos de aptitud combinatoria general de las líneas del grupo Y,  $V_{ACE}$ : varianza de los efectos de aptitud combinatoria específica de las cruzas,  $V_E$ : varianza del error y  $h^2$ : heredabilidad. La estimación se hizo mediante marcadores moleculares y BLUP.

#### 4.4.3 Predicción de rendimiento de cruzas mediante BLUP y RR-BLUP

Los coeficientes de coancestría estimados mediante marcadores moleculares, en general, fueron bajos. En las líneas del grupo X, fluctuaron desde 0.01 entre el par de líneas L5BI y L65BI, hasta 0.327 entre L59BI y L65BI, con coancestría media de 0.108. En el grupo de líneas Y se presentaron valores inferiores, desde 0 (en seis combinaciones) hasta 0.175 entre el par L45BI y L80BI, con promedio de 0.06.

Las estimaciones de las varianzas genética y no genética de las cuatro variables estudiadas, con marcadores ISSR mediante BLUP (Cuadro 4.2), permitieron calcular la heredabilidad ( $h^2$ ) de cada carácter (Bernardo, 1996). De esta manera, para RFT, RFC, y FF, la heredabilidad fue muy alta (0.98); el NF registró  $h^2 = 0.99$ . En el caso del RFT, el valor estimado de  $h^2$  en este estudio dista del publicado por Dordevic, Zecevic, Zdravkovic, Zivanovic, y Todorovic (2010) de 0.451; además de las variaciones aleatorias del error, esta discrepancia se puede atribuir a diferencias entre materiales genéticos y ambientes. Wessel-Beaver y Scott (1992) obtuvieron estimaciones de  $h^2$  de 0.65 y 0.81 para el rendimiento de la misma población de tomate cultivada en Puerto Rico y Florida, respectivamente.

Los valores de heredabilidad en este estudio permiten predecir éxito en la selección de los mejores híbridos, así como en un programa de mejoramiento basado en selección. La proporción de la varianza de la aptitud combinatoria específica ( $V_{ACE}$ ) respecto de la varianza genética total entre las CS fue 0.107, 0.106, 0.135 y 0.117, para NF, RFT, RFC y FF, respectivamente. Valores similares fueron reportados por Massman *et al.* (2013) en maíz al emplear el BLUP (desde 0.10 para humedad hasta 0.18 para rendimiento de grano). Por lo anterior, se puede inferir que en esta investigación el BLUP explotó efectos aditivos en mayor medida que de dominancia.

#### 4.4.4 Correlaciones entre rendimientos observados y predichos

El promedio de 1,000 correlaciones entre el valor predicho y el registrado en campo para cada conjunto de  $n$  híbridos predictores se resume en el Cuadro 4.3. Tanto el BLUP como el RR-BLUP pronosticaron adecuadamente en las cuatro variables analizadas, y en todos los casos el BLUP estimó valores más cercanos a los reales que el RR-BLUP. Como era de esperarse, cuando se aumentó  $n$  se incrementaron también los valores de las correlaciones, aunque a partir de  $n = 24$  estos incrementos fueron de poca magnitud e incluso se registró reducción. En este estudio, con 18 híbridos predictores ( $n = 18$ ), se obtuvo la mayor estabilidad en las predicciones y sería el número de predictores ideal en un escenario de costos donde un incremento en  $n$  implica mayor inversión.

**Cuadro 4.3.** Correlaciones promedio de la validación cruzada para dos métodos (BLUP y RR-BLUP) de predicción de híbridos de tomate en cuatro variables. Los datos corresponden a 1,000 muestras aleatorias independientes de 36 híbridos predictores.

N	NF		RFT (kg·m <sup>-2</sup> )		RFC (kg·m <sup>-2</sup> )		FF (kgf·cm <sup>-2</sup> )	
	BLUP	RR-BLUP	BLUP	RR-BLUP	BLUP	RR-BLUP	BLUP	RR-BLUP
6	0.53	0.45	0.52	0.36	0.36	0.25	0.41	0.32
12	0.72	0.66	0.69	0.58	0.48	0.40	0.57	0.55
18	0.80	0.75	0.76	0.68	0.55	0.49	0.65	0.63
24	0.83	0.80	0.79	0.73	0.58	0.54	0.69	0.68
30	0.83	0.83	0.75	0.69	0.57	0.53	0.71	0.67
Media	0.74	0.70	0.70	0.60	0.51	0.44	0.61	0.57

N: número de híbridos predictores, BLUP: mejor predictor lineal e insesgado, RR-BLUP: regresión ridge BLUP, NF: número de frutos, RFT: rendimiento de fruto total, RFC: rendimiento de fruto comercial y FF: firmeza de fruto.

Al considerar cada variable por separado, el NF registró los coeficientes de correlación más altos (de 0.45 a 0.83), seguido del RFT (0.36 a 0.73), la FF (0.32

a 0.68) y los resultados más bajas fueron del RFC (0.25 a 0.54), con RR-BLUP. Los valores de correlación entre observados y predichos con BLUP mantuvieron la misma tendencia, el NF (0.53 a 0.83), el RFT (0.52 a 0.79), la FF (0.41 a 0.71) y por último el RFC (0.36 a 0.58).

Las correlaciones obtenidas para el RFT son, en general, similares a las reportadas por Hernández-Ibáñez, Sahagún-Castellanos, Rodríguez-Pérez, y Peña-Ortega (2014) en tomate (entre 0.45 y 0.79), y por Bernardo (1994) en rendimiento de grano de maíz. Este último utilizó RFLP's y las correlaciones más altas fueron de 0.80. Las correlaciones altas obtenidas sugieren que el método utilizado para predecir el rendimiento de los híbridos en tomate es prometedor. Sin embargo, la extrapolación debe hacerse con reservas, ya que la correlación entre híbridos observados y predichos aumenta conforme incrementa el número de híbridos predictores, pero hasta cierto nivel, después de éste se corre el riesgo de tener correlaciones más pequeñas e incluso negativas. Además, los resultados expuestos corresponden a un conjunto relativamente pequeño de CS. Bernardo (1996) aplicó esta metodología a una escala mayor (16 combinaciones de nueve grupos heteróticos en maíz), y las correlaciones entre rendimientos predichos y observados que obtuvo fueron ligeramente más bajas (de 0.426 a 0.762), muy parecidas a las obtenidas en el presente trabajo.

Massman *et al.* (2013) realizaron una comparación entre BLUP y RR-BLUP en cuatro variables en maíz y produjo correlaciones altas (0.87, 0.90, 0.69 y 0.84, para rendimiento de grano, humedad del grano, materia seca del tallo y raíz, respectivamente), sin diferencias significativas entre metodologías; por lo que afirmaron que RR-BLUP no superó los resultados obtenidos con BLUP. Estos resultados anteriores sugieren que el factor genético puede modificar la habilidad predictiva de los modelos.

Los resultados obtenidos en el presente trabajo en tomate se pueden considerar acordes con los reportados en maíz, a pesar de que se trata de especies

contrastantes. Mientras que el tomate es una especie autógama con nivel reducido de heterosis, el maíz es alógama con expresiones de heterosis altas. En tomate, además del RFT, es muy importante el RFC y la FF después de la cosecha, por lo que se ven involucrados otros genes además de los que controlan estrictamente el rendimiento; los cuales afectan negativamente las correlaciones entre los diversos componentes del rendimiento y la producción de fruto de tomate. Los efectos epistáticos y de ligamiento, entre otros factores genéticos, además del ambiente, también pueden afectar la correlación. Lo anterior explica la dificultad en predecir estos caracteres, y por ello los niveles bajos de predicción obtenidos.

#### **4.5 CONCLUSIONES**

Los coeficientes de correlación entre valores observados y predichos para NF, RFT, RFC y FF, estimados mediante BLUP y RR-BLUP en CS de tomate, fluctuaron entre 0.25 y 0.83, valores que sugieren que estos métodos pueden tener utilidad en el mejoramiento genético.

El BLUP registró consistentemente resultados más altos de correlación entre valores observados y predichos, comparado con RR-BLUP. En ambos casos, la magnitud de las correlaciones siempre aumentó cuando incrementó  $n$ . Por lo tanto, para alcanzar mayor precisión y confiabilidad de las predicciones con BLUP o RR-BLUP se deben evaluar en campo la mayor cantidad de cruzas predictoras, en tantos ambientes como el presupuesto lo permita.

#### **4.6 AGRADECIMIENTOS**

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) de México, por la beca doctoral otorgada al primer autor. Al Q.F.B. Ricardo Gaspar Hernández por el soporte técnico en marcadores moleculares de ADN.

#### 4.7 LITERATURA CITADA

- Aguilera, J. G., Pessoni, L. A., Rodrigues, G. B., Elsayed, A. Y., Da Silva, D. J. H., & De Barros, E. G. (2011). Genetic variability by ISSR markers in tomato (*Solanum lycopersicon* Mill.). *Revista Brasileira de Ciências Agrárias* 6(2), 243–252. doi:10.5039/agraria.v6i2a998
- Balestre, M., Von Pinho, R. G., & Souza, J. C. (2010). Prediction of maize single-cross performance by mixed linear models with microsatellite marker information. *Genetics and Molecular Research* 9(2), 1054–1068. doi:10.4238/vol9-2gmr791
- Bernardo, R. (1993). Estimation of coefficient of coancestry using molecular markers in maize. *Theoretical and Applied Genetics* 85, 1055–1062. doi:10.1007/BF00215047
- Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Science* 34(1), 20–25. doi:10.2135/cropsci1994.0011183X003400010003x
- Bernardo, R. (1995). Genetic models for predicting maize single-cross performance in unbalanced yield trial data. *Crop Science* 35(1), 141–147. doi:10.2135/cropsci1995.0011183X003500010026x
- Bernardo, R. (1996). Best linear unbiased prediction of maize single-cross performance. *Crop Science* 36(1), 50–56. doi:10.2135/cropsci1996.0011183X003600010009x
- Cadahia, C. (2000). *Fertirrigación, cultivos hortícolas y ornamentales* (2da. Ed.). Madrid, España: Mundi Prensa.
- Dordevic, R., Zecevic, B., Zdravkovic, J., Zivanovic, T., & Todorovic, G. (2010). Inheritance of yield components in tomato. *Genetika* 42(3), 575–583. doi:10.2298/GENSR1003575D
- Gbur, E. E., Stroup, W. W., Mc Carter, K. S., Durham, S., Young, L. J., Christman, M., ... Kramer, M. (2012). *Analysis of generalized linear mixed model in the agricultural and natural resources sciences*. Madison, Wisconsin, USA.: ASA, SSSA, CSSA, Inc.
- Henderson, C. R. (1985). Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *Journal of Animal Science* 60(1), 111–117. doi:10.2134/jas1985.601111x

- Hernández-Ibáñez, L., Sahagún-Castellanos, J., Rodríguez-Pérez, J. E., & Peña-Ortega, M. G. (2014). Predicción de rendimiento de fruto de híbridos de tomate con huellas genómicas. *ΦYTON Revista Internacional de Botánica Experimental* 83, 311–318. Obtenido de <http://www.revistaphyton.fundromuloraggio.org.ar/vol83/HernandezIbanez.pdf>
- Khattree, R., & Naik, D. N. (2000). *Multivariate data reduction and discrimination with SAS software*. Cary, N.C., U.S.A: SAS Institute Inc.
- Massman, J. M., Gordillo, A., Lorenzana, R. E., & Bernardo, R. (2013). Genomewide predictions from maize single-cross data. *Theoretical and Applied Genetics* 126, 13–22. doi:10.1007/s00122-012.1955-y
- Melchinger, A. E. (1988). Means, variances, and covariances between relatives in hybrids populations with disequilibrium in the parent populations. In B.S. Wei, & E. J. Eisen (Eds.), *Proc. 2<sup>nd</sup>. int. conf. quantit. genet.* (pp 400–405). Raleigh, NC.: Sinauer Assoc.
- Meuwissen T.H.E., Hayes B. J., & Goddard M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4), 1819–1829. Obtenido de <http://www.genetics.org/content/genetics/157/4/1819.full.pdf>
- Mirshamsi, A., Farsi, M., Shahriari, F., & Nemati, H. (2008). Use of random amplified polymorphic DNA markers to estimate heterosis and combining ability in tomato hybrids. *Pakistan Journal of Biological Sciences* 11(4), 499–507. doi:10.3923/pjbs.2008.499.507
- Pérez, P., & De los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198(2), 483–495. doi:10.1534/genetics.114.164442.
- Wessel-Beaver, L. & Scott, J. W. (1992). Genetic variability of fruit set, fruit weight, and yield in a tomato population grown in two high-temperature environments. *Journal of the American Society of Horticultural Science* 117(5), 867–870. Obtenido de <http://journal.ashspublications.org/content/117/5/867.full.pdf+html>

## 5 DISCUSIÓN GENERAL

El tomate, originario de la región andina y domesticado en Mesoamérica, como muchas otras especies cultivadas, ha registrado un aumento espectacular en el tamaño del fruto, resultado de un proceso largo de mejoramiento. El ancestro del tomate, de fruto menor a 1 cm de diámetro y unos pocos gramos de peso, con cientos de semillas, dista mucho de los tomates modernos que pueden llegar a pesar hasta 1,000 gramos y superar los 15 cm de diámetro (Frary *et al.*, 2000). En el último par de décadas, el proceso se ha acelerado, las variedades liberadas al mercado pasaron desde tener frutos de 100 y 200 g hasta más de 150 y 350 g en promedio para frutos tipo saladette y bola, respectivamente.

La demanda del mercado por variedades nuevas, con características superiores, presiona a los programas de mejoramiento. Éstos deben de responder no sólo con mejores genotipos, sino también, con ciclos de mejoramiento más cortos, entre la cruza inicial y la variedad liberada. En maíz y soya, la utilización de dobles haploides y selección genómica de líneas han acortado la generación de nuevos cultivares a periodos de 7 años (Bernardo, 2010). En tomate, no hay reportes de qué tanto se ha reducido el tiempo para generar una nueva variedad. Programas de mejoramiento recientes, como el de la Universidad Autónoma Chapingo, se ven en la necesidad de aplicar los últimos conocimientos en genómica al mejoramiento genético de la especie y conseguir genotipos más rendidores, menos demandantes de recursos y más adaptables a las condiciones cambiantes del ambiente en que tendrán que ser cultivados, para ser competitivos.

El tomate es una especie autógama, su mejoramiento en forma natural había sido por selección. Sin embargo, en la actualidad los materiales liberados son fundamentalmente híbridos F<sub>1</sub>. Debido a que en ambas metodologías se requiere producir líneas homocigóticas, en la primera fase de esta investigación se evaluó la selección genómica de líneas élite como una alternativa a la selección

fenotípica, cuyo principal inconveniente es el tiempo necesario para un ciclo de selección y su costo alto, sobre todo al hacer la producción en condiciones de invernadero e hidroponía. Por otro lado, en la explotación de la heterosis en un programa de hibridación, se puede generar un número muy grande de cruzas, cuando se cuenta con un número alto de LE (Bernardo, 1994). Por tal motivo, en la segunda etapa de la investigación se realizó la predicción de cruzas simples de tomate para los caracteres NF, RFT, RFC y FF, con la finalidad de explorar la reducción del número de híbridos a evaluar en campo, al considerar sólo los más prometedores, con los consecuentes ahorros.

La aplicación de la SG en el mejoramiento requiere utilizar una plataforma de marcadores moleculares de ADN para genotipar los materiales, segregantes o avanzados; y la evaluación fenotípica de los genotipos respectivos en campo para verificar la exactitud de las predicciones. En esta investigación se genotiparon todas las líneas élite del programa de tomate con 29 iniciadores ISSR; y los datos fenotípicos corresponden a la evaluación en campo de distintos conjuntos de líneas a lo largo de cinco años. Con los 29 ISSR se amplificaron 276 bandas, de las cuales 215 fueron polimórficas para el total de LE. Sin embargo, al considerar los dos grupos de líneas estudiadas, el número de marcadores polimórficos se redujo a 148 y 156. Esto incide en las predicciones, ya que un bajo número de marcadores limita la posibilidad de encontrar asociación que refleje la variación fenotípica observada.

En este estudio las correlaciones entre rendimientos observados y predichos, fueron consistentemente bajas para todas las variables y el método empleado (entre 0 y 66). Valores más cercanos a los de Hernández-Bautista *et al.* (2016) (entre 0.25 y 0.46), que a los de Duangjit *et al.* (2016) (desde 0.42 hasta 0.81), ambos en tomate. Hernández-Bautista *et al.* (2016) emplearon 31 marcadores SSR, que al parecer no fueron capaces de reflejar la variación de sus líneas F<sub>2:3</sub>, lo que los llevó a concluir que es más efectiva la selección fenotípica que la genómica, mientras que Duangjit *et al.* (2016) con 2,313 marcadores SNP

obtuvieron valores de predicción altos. En este sentido parece ser que la elección de la plataforma de marcadores y su número es pieza fundamental, incluso tan importante como el método de predicción (Crossa *et al.*, 2010). En diversos estudios las diferencias entre métodos de selección genómica no han mostrado ser muy grandes, e incluso en la mayoría sin diferencias significativas (Hernández-Bautista *et al.* 2016; Yamamoto *et al.*, 2016). Afortunadamente en años recientes, los avances en genotipo han reducido los costos de los marcadores e incrementado la cobertura del genoma (Heffner *et al.*, 2010). De seguir esta tendencia, los programas pequeños podrán migrar a plataformas de próxima generación y emplear marcadores SNP en mayor cantidad y con mejor precisión. Y si bien, la precisión en las predicciones obtenidas en el presente estudio no es muy alta, la SG podría acelerar dramáticamente la ganancia genética mediante el acortamiento de los ciclos de mejoramiento (Heffner *et al.*, 2010) y aumentar la respuesta a la selección por unidad de tiempo (Bernardo, 2010).

La precisión en la predicción de un carácter, depende del tipo de carácter estudiado. Caracteres de tipo cualitativo, con alta heredabilidad, presentan valores de correlación, entre predichos y observados, generalmente altos. Los caracteres cuantitativos en contraste, alcanzan valores de predicción de menor magnitud. En los resultados de esta investigación, se observó que de los caracteres cuantitativos analizados, el RFT registró las correlaciones entre predichos y observados más altas, seguido de FF y en último lugar RFC. Duangjit *et al.*, (2016) también registraron valores bajos en FF. Lo anterior puede sugerir que los caracteres que involucran calidad del fruto están determinados por la expresión de genes particulares que no se han asociado correctamente a los conjuntos de marcadores empleados. La razón posiblemente no sólo recae en el número de marcadores, sino en que el ADN utilizado para genotipar se toma en etapas tempranas del desarrollo de la planta, cuando los genes que tienen que ver con calidad no se están expresando, es decir que en esa zona el material

genético no se encuentra distendido y en consecuencia no se accede a él completamente.

Lo anterior está muy relacionado con el factor ambiental. Se observó que la predicción varió en función del ciclo de producción. Para algunos ambientes la predicción fue alta y en otros se redujo, incluso hubo correlaciones negativas. Utilizar el promedio de ambientes en las variables RFT y RFC mejoró la precisión en la predicción del grupo de 28 LE, no así para el de 17 LE, que registró valores negativos. Por esta razón, es fundamental evaluar los materiales en el mayor número de ambientes posibles para reducir el sesgo de un ambiente particularmente desfavorable.

La idea de que con el uso de marcadores se puede prescindir de la evaluación en campo no es del todo correcta. Se puede evitar fenotipar todos los materiales, al reducirlos a una muestra, que debe ser valorada con mayor precisión. El tamaño de la muestra según los resultados observados, es del 50 % si se quiere lograr valores altos de predicción. Esto permitiría evaluar dicho grupo de LE en el doble de ambientes con los mismos recursos que se invertirían en el paquete completo.

En la segunda etapa de la investigación se evaluó la predicción del NF, RFT, RFC y FF de híbridos de CS en tomate, mediante los métodos Blup y RR-Blup. Los dos grupos de seis LE utilizados, produjeron 36 cruzas que se evaluaron en los ciclos agrícolas 2014 y 2015. El ambiente 2014 fue más favorable y los rendimientos mayores. La predicción de las variables registradas en las cruzas, mediante la información genética de las LE progenitoras basada en los marcadores obtenidos con 29 ISSR fue exitosa. El método Blup fue ligeramente más preciso que el RR-Blup, como lo habían reportado Massman *et al.* (2013) en maíz. En ambos métodos se usaron los marcadores, en Blup para estimar los coeficientes de coancestría entre líneas y en RR-Blup directamente para la

predicción. En este sentido, RR-Blup presentó la ventaja de no requerir cálculos adicionales, con el inconveniente ya mencionado.

Como en la SG, la predicción en cruzas simples de tomate varió según el carácter estudiado. En NF y RFT se alcanzaron los valores de predicción más altos, en FF intermedios y en RFC se registró los más bajos. En FF y RFC se ven involucrados otros genes además de los que controlan estrictamente el rendimiento, genes que posiblemente no fueron genotipados debido entre otras razones a la baja densidad de la plataforma de marcadores utilizada. De ahí, los niveles bajos de predicción obtenidos.

Conforme aumentó el número de CS evaluadas en campo utilizadas en los modelos de predicción, los valores de correlación entre predichos y observados se incrementaron. En este estudio, emplear el 50 % de las CS como predictoras fue lo ideal, con esta cantidad se lograron los valores más altos de correlación. En la implementación práctica de esta metodología, evaluar en campo sólo el 25 % de CS implica una pérdida de precisión, compensada plenamente con la posibilidad de evaluar el mismo grupo de cruzas en el doble de ambientes y así conseguir una información fenotípica más robusta, que pronostique mejor las cruzas no evaluadas, identificar las más prometedoras, hacerlas y probarlas en campo en futuros ensayos. Por otro lado, la ventaja de esta metodología radica en no tener que hacer evaluaciones expofeso para la predicción, ya que se pueden utilizar datos de las diferentes pruebas de híbridos del programa de mejoramiento (Bernardo, 2010). De esta manera, conforme el programa avanza, se cuenta con más información y se obtiene mayor precisión.

Finalmente, el empleo de marcadores moleculares en un programa de mejoramiento genético no es optativo, ya debe ser considerado como parte fundamental. En este trabajo se ha dejado manifiesto el uso práctico de los marcadores para hacer SG en líneas élite de tomate desarrolladas por el método uniseminal. Así también, se comprobó su efectividad en la predicción de variables

de rendimiento y firmeza de fruto en híbridos F<sub>1</sub> de cruza simple. Además, a la par se utilizaron los marcadores para calcular relaciones de parentesco entre las líneas.

## 5.1 LITERATURA CITADA

Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Science* 34(1), 20–25.

Bernardo, R. (2010). *Breeding for Quantitative Traits in Plants*. Second Edition. Stemma Press. Minnesota, USA. 380 p.

Crossa, J., de Los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., ... & Arief, V. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186(2), 713-724.

Duangjit, J., Causse, M., & Sauvage, C. (2016). Efficiency of genomic selection for tomato fruit quality. *Mol. Breeding*, 36:29.doi:10.1007/s11032-016-0453-3.

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

Heffner, E. L., Lorenz, A. J., Jannink, J. L., & Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop science*, 50(5), 1681-1690.

Hernández-Bautista, A., Lobato-Ortiz, R., García-Zavala, J. J., Parra-Gómez, M. A., Cadeza-Espinosa, M., Canela-Doñan, D., ... & Chávez-Servia, J. L. (2016). Implications of genomic selection for obtaining F<sub>2</sub>: 3 families of tomato. *Scientia Horticulturae*, 207, 7-13.

Massman, J. M., Gordillo, A., Lorenzana, R. E., & Bernardo, R. (2013). Genomewide predictions from maize single-cross data. *Theoretical and applied genetics*, 126(1), 13-22.

Yamamoto, E., Matsunaga, H., Onogi, A., Kajiya-Kanegae, H., Minamikawa, M., Suzuki, A., ... & Miyatake, K. (2016). A simulation-based breeding design that uses whole-genome prediction in tomato. *Scientific reports*, 6.

## 6 CONCLUSIONES GENERALES

La selección genómica de líneas élite es factible en tomate. El G-Blup produjo consistentemente resultados altos de correlación entre valores observados y predichos, al utilizar la matriz de marcadores y la combinación marcadores-pedigrí.

La predicción de rendimiento y firmeza de fruto de tomate en cruza simples fue validada exitosamente. Los coeficientes de correlación entre valores observados y predichos para NF, RFT, RFC y FF, estimados mediante BLUP y RR-BLUP en CS de tomate, fluctuaron entre 0.25 y 0.83. El BLUP registró consistentemente resultados más altos de correlación entre valores observados y predichos, comparado con RR-BLUP.

Para obtener las mejores predicciones es recomendable evaluar en campo como mínimo el 50 % de las líneas élite o las cruza simples, y en tantos ambientes como se pueda. En la SG y predicción de CS analizadas, el aumento en la correlación media ocurrió cuando en el conjunto de entrenamiento se incluyeron más LE o CS.

Se confirmó que el empleo de marcadores moleculares de ADN tiene la capacidad de potenciar el mejoramiento genético de tomate y conseguir en menor tiempo los genotipos más rendidores, menos demandantes de recursos y más adaptables a las condiciones cambiantes del ambiente en que tendrán que ser cultivados, objetivos del Programa de Mejoramiento de Tomate de la Universidad Autónoma Chapingo.

## 7 APÉNDICES

**Apéndice 1.** Programa en R para selección genómica mediante el método G-Blup y cálculo de la correlación media entre valores predichos y observados

```
#Análisis de una variable (Y).
#Especificar el número de ciclos (cycles) que ejecutará el programa.
traits=1
cycles=1000
accuracy = matrix(nrow=cycles,ncol=traits)
for (r in 1:cycles) {
#1# Cargar y preparar el conjunto de datos
#Para importar la matriz de marcadores (MatX), especifique la ubicación del archivo de Excel, el
formato debe ser "separado por comas".
setwd("C:/Users/Lucas/Desktop/SET2")
datos<-read.csv(file="MatM_156MM.csv",header=TRUE)
str(datos)
M=datos[,-c(1)];M1=as.matrix(M)
MatX=M1;MatX ;rownames(MatX)<-datos[,1]
str(MatX)
MatX

#Importación de la matriz de pedigrí (MatA), archivo de Excel "separado por comas"
setwd("C:/Users/Lucas/Desktop/SET2")
datos1<-read.csv(file="MatA_28GEN.csv",header=TRUE)
str(datos1)
A=datos1[,-c(1)];A1=as.matrix(A)
MatA=A1;MatA ;rownames(MatA)<-datos1[,1]
str(MatA)
MatA

#Importación de la matriz de datos fenotípicos (MatF). Si el archivo incluye más de una variable,
especifique la columna que contiene los datos fenotípicos de la variable a analizar.
setwd("C:/Users/Lucas/Desktop/SET2")
datos2<-read.csv(file="MatF_28GEN.csv",header=TRUE)
str(datos2)
Y=datos2[,2];
Y=as.matrix(Y);
Y
MatF=Y; rownames(MatF)<-datos2[,1]
str(MatF)
MatF

#Cargar la librería de funciones BGLR.
library(BGLR);
MatX=as.matrix(MatX)
n<-nrow(MatF); p<-ncol(MatX)
y<-as.matrix(MatF)

#Computo de la Matriz de relaciones genómicas G3
X3=scale(MatX,center=TRUE,scale=TRUE)
G3=tcrossprod(X3)/ncol(X3)
```

```

#G3
#hist(G3)
#heatmap(G3)

#2# Crear un conjunto de prueba (testing set), size es el número de genotipos seleccionados al
azar, cuyo valor se da por perdido y predice con el resto de información.
yNA<-y
whichNa<-sample(1:n,size=14,replace=FALSE)
yNA[whichNa]<-NA
yNA
#3# Ajuste del modelo G-BLUP con datos de Marcadores. Eliminar el # para activar el ETA a
utilizar, el primero utiliza marcadores y pedigrí, el segundo sólo marcadores y el tercero
únicamente pedigrí.
#ETA<-list(list(X=MatX,model="BRR"),list(K=MatA,model="RKHS"))
#ETA<-list(list(X=MatX,model="BRR"))
#ETA<-list(list(K=MatA,model="RKHS"))
fm<-BGLR(y=yNA,ETA=ETA,nIter=10000, burnIn=5000)
#Elaboración de gráfica
plot(fm$yHat,y,main="Rendimiento de fruto total",xlab="Observado",
      ylab="Predicho",cex=.8,bty="L")
points(x=y[whichNa],y=fm$yHat[whichNa],col=2,cex=.8,pch=19)
legend("topleft", legend=c("entrenamiento","prueba"),bty="n",
      pch=c(1,19), col=c("black","red"))
#4# Evaluación de correlación entre conjuntos de entrenamiento y prueba (TRN y TST data sets)
COR.tst<-cor(fm$yHat[whichNa], y[whichNa])
COR.tst
COR.trn<-cor(fm$yHat[-whichNa], y[-whichNa])
COR.trn
MSE.tst<-mean((fm$yHat[whichNa] -y[whichNa])^2)
MSE.tst
MSE.trn<-mean((fm$yHat[-whichNa] -y[-whichNa])^2)
MSE.trn
#str(fm)
accuracy[r,1]<-COR.tst
}
#Cálculo de la correlación media
accuracy
mean(accuracy)
#Fin del programa

```

**Apéndice 2.** Programa en SAS para hacer estimaciones de máxima verosimilitud restringida mediante iteración a partir de sendas ecuaciones de Henderson (1985), para la  $V_E$  (varianza no genética),  $V_{A(X/Y)}$ ,  $V_{A(Y/X)}$  y  $V_{D(XY)}$ .

```

/*Inicia Programa SAS */
OPTIONS LS=256 PS=10000 NOCENTER;
PROC IML;
/* Las matrices X y Y son las matrices binarias de los marcadores moleculares de las líneas de
los grupos heteróticos X y Y respectivamente*/
/* Ejemplo del formato de las matrices X, Y y Rxy. Copiar las matrices del archivo de Excel y
pegarlas. Con fines didácticos, el ejemplo considera sólo 20 marcadores moleculares para X y
Y*//
X={1111011001110110011,
01011110101001100110,
01110000001000000110,
11111101111000100110,
00111101011101100010};

Y={00111000001101101110,
01101000001101101110,
01111110001001100110,
01101000001000000010,
01101000001100000010,
11101000001001100010,
00001000001001100010,
11111110001100000100};

Rxy={111110.292,
121210.896,
13138.946,
141410.475,
15159.844,
161611.017,
171710.089,
181811.718,
192112.070,
1102210.718,
1112311.008,
1122410.233,
1132511.317,
1142611.293,
1152711.532,
1162812.596,
1173111.118,
1183211.013,
1193310.116,
1203410.238,
1213511.482,
1223611.063,
1233710.881,
1243811.161,
1254110.427,
126429.974,

```

```

1274312.824,
1284410.649,
1294511.581,
1304611.703,
1314711.335,
1324814.077,
1335110.113,
134529.279,
135539.588,
136549.035,
137558.763,
1385610.334,
139579.608,
1405812.207};
/*LAS MATRICES X y Y TIENEN EL MISMO NÚMERO DE COLUMNAS (bandas), EL
NÚMERO DE HILERAS DE X DEBE SER MENOR AL NÚMERO DE HILERAS DE Y. LA
MATRIZ Rxy DEBE CONTENER COL1=AMBIENTE COL2=N° CRUZA COL3=PROG X
COL4=PROG Y COL=5 RENDIMIENTO */
/*El siguiente código calcula la matriz de relaciones genéticas fxxfy mediante marcadores
moleculares*/
PRINT X Y Rxy;
B=NCOL(X);nx=NROW(X);ny=NROW (Y); NOBS=NROW(Rxy); NAMB=1;
/*PROPORCIONES DE X*/
DO H=1 TO NX; Xh1=X[H,];DO I=1 TO NX; Xh2=X[I,]; Xh3=Xh1//Xh2; N=0; M=0;*PRINT H I;
*PRINT XH1 XH2 XH3;
DO J=1 TO B; K=Xh3[,J]; L=K[+]; IF L>0 THEN N=N+1; IF L=2 THEN M=M+1; DISx=M/N;
END; JAC1x=JAC1x//DISx; *PRINT M N DISx JAC1x; END;END;*PRINT JAC1x;
DO P=1 TO NX; Q1=P*NX-NX+1; Q2=P*NX; JAC2x=JAC1x[Q1:Q2,]; Sxx=Sxx||JAC2x;*PRINT
q1 q2 JAC2x Sxx; END;PRINT Sxx;

/*PROPORCIONES DE Y*/
DO H=1 TO NY; Yh1=Y[H,];DO I=1 TO NY; Yh2=Y[I,]; Yh3=Yh1//Yh2; N=0; M=0; *PRINT H I;
*PRINT YH1 YH2 YH3;
DO J=1 TO B; K=Yh3[,J]; L=K[+]; IF L>0 THEN N=N+1; IF L=2 THEN M=M+1; DISy=M/N;
END; JAC1y=JAC1y//DISy; *PRINT M N DISy JAC1y; END;END;*PRINT JAC1y;
DO P=1 TO NY; Q1=P*NY-NY+1; Q2=P*NY; JAC2y=JAC1y[Q1:Q2,]; Syy=Syy||JAC2y;*PRINT
q1 q2 JAC2y Syy; END;PRINT Syy;

/*PROPORCIONES DE XY*/
DO H=1 TO NX; Xh=x[H,];DO I=1 TO NY; Yh=Y[I,]; XYh=Xh//Yh; N=0; M=0;*PRINT H I; *PRINT
XH YH XYH;
DO J=1 TO B; K=XYh[,J]; L=K[+]; IF L>0 THEN N=N+1; IF L=2 THEN M=M+1; DISxy=M/N;
END;
JAC1xy=JAC1xy//DISxy; *PRINT M N DISxy H I JAC1xy; END;END;*PRINT JAC1xy;
DO P=1 TO NX; Q1=P*NY-NY+1; Q2=P*NY; JAC2xy=t(JAC1xy[Q1:Q2,]);
Sxy=Sxy//JAC2xy;*PRINT q1 q2 JAC2xy Sxy; END;PRINT Sxy;

/*PROMEDIOS XY y YX*/
DO R1=1 TO NX; AA=Sxy[R1,]; AA1=AA[:]; PRSxy=PRSxy//AA1; *PRINT R1 AA AA1
PRSxy;END; PRINT PRSxy;
DO S1=1 TO NY; BB=Sxy[,S1]; BB1=BB[:]; PRSyx=PRSyx||BB1; *PRINT S1 BB BB1
PRSyx;END; PRINT PRSyx;

/* fxx */

```

```

DO T=1 TO NX; DO U=1 TO NX;XX=Sxx[T,U]; XP1=PRSxy[T,]; XP2=PRSxy[U,]; *PRINT XX
XP1 XP2;
fxx1=(XX-(XP1+XP2)/2)/(1-(XP1+XP2)/2); IF fxx1<0 then fxx1=0; fxx2=fxx2/fxx1;*print fxx1
fxx2;END; END;
DO V=1 TO NX; w1=V*NX-NX+1; w2=V*NX; fxx3=fxx2[w1:w2,]; fxx=fxx||fxx3;*print fxx; END;
PRINT fxx;

```

```

/* fyy */
DO T=1 TO NY; DO U=1 TO NY;YY=Syy[T,U]; YP1=PRSyx[T,]; YP2=PRSyx[U,]; *PRINT YY
YP1 YP2;
fyy1=(YY-(YP1+YP2)/2)/(1-(YP1+YP2)/2); IF fyy1<0 then fyy1=0; fyy2=fyy2/fyy1;*print fyy1
fyy2;END; END;
DO V=1 TO NY; w1=V*NY-NY+1; w2=V*NY; fyy3=fyy2[w1:w2,]; fyy=fyy||fyy3;*print fyy; END;
PRINT fyy;

```

```

/* fxx*fyy */
DO RR=1 TO NX; DO SS=1 TO NX; DO TT=1 TO NY; DO UU=1 TO NY;
fxxij=fxx[ss,rr]; fyykl=fyy[uu,tt]; EE=fxxij*fyykl; EEE=EEE/EE;
*print rr ss TT UU fxxij tt uu fyykl EE; end;end;end;end;
*PRINT EEE;

```

```

NC=(NX*NY)**2/NY;
DO P=1 TO NC;
K1=P*NY-NY+1; K2=P*NY;
FF2=EEE[K1:K2,]; FF3=FF3||FF2;*PRINT K1 K2 ;*CC2 Cijkl; END;
*PRINT FF3;

```

```

NC=NX*NY;PRINT NC NX NY;
DO RRR=1 TO NX;
K3=RRR*NC-NC+1; K4=RRR*NC; *PRINT K3 K4;
FF4=FF3[K3:K4]; fxxfyy=fxxfyy/FF4; *PRINT FF4 FF5;
END;
PRINT fxxfyy;

```

```

Z1=DESIGN(Rxy[,3]); Z2=DESIGN(Rxy[,4]); Z=DESIGN(Rxy[,2]);Y=Rxy[,5];X=DESIGN(Rxy[,1]);
PRINT RXY Z1 Z2 Z Y X;

```

*/\* VALORES SEMILLA DE VARIANZAS. Introducir los valores semilla de las varianzas VE VX VY y VD.*

*Los valores estimados en la iteración (VEE VXE VYE y VDE) son los valores semilla de la siguiente iteración. Así sucesivamente hasta alcanzar una diferencia menor de 0.0001 \*/*

```

VE=0.0943908; VX=1.1333878; VY=0.933071; VD=0.4542054; FI1=VE/VX; FI2=VE/VY;
FI3=VE/VD;

```

```

C00=T(X)*X; C01=T(X)*Z1; C02=T(X)*Z2; C03=T(X)*Z;
C10=T(Z1)*X; C11=T(Z1)*Z1+INV(fxx)*FI1; C12=T(Z1)*Z2; C13=T(Z1)*Z;
C20=T(Z2)*X; C21=T(Z2)*Z1; C22=T(Z2)*Z2+INV(fyy)*FI2; C23=T(Z2)*Z;
C30=T(Z)*X; C31=T(Z)*Z1; C32=T(Z)*Z2; C33=T(Z)*Z+INV(fxxfyy)*FI3;

```

```

PRINT C00 C01 C02 C03 C10 C11 C12 C13 C20 C21 C22 C23 C30 C31 C32 C33;

```

```

C=(C00||C01||C02||C03)/(C10||C11||C12||C13)/(C20||C21||C22||C23)/(C30||C31||C32||C33);
PRINT C;

```

```

EE1=(t(X)*Y)/(T(Z1)*Y)/(T(Z2)*Y)/(T(Z)*Y);
PRINT EE1;
PARAM=INV(C)*EE1;
PRINT PARAM;

/*VALORES ESTIMADOS de PARAMETROS*/
Beta=PARAM[1,];
ax=PARAM[2:6,]; /*SELECCIONAR HILERAS CORRESPONDIENTES*/
ay=PARAM[7:14,];
d=PARAM[15:54,];
PRINT Beta AX AY D;

VEE=(T(Y)*Y-T(PARAM)*EE1)/(NOBS-NAMB);
VXE=((T(ax)*INV(fxx)*ax)+VEE*TRACE((INV(fxx))*C11))/nX;
VYE=(T(ay)*INV(fyy)*ay+VEE*TRACE((INV(fyy))*C22))/nY;
VDE=(T(D)*INV(fxxfyy)*D+VEE*TRACE((INV(fxxfyy))*C33))/nobs;
PRINT VE VEE VX VXE VY VYE VD VDE;

/* YP=Rendimientos predichos */
YP=INV(T(Z)*Z)*T(Z)*(Y-X*(Beta)); PRINT Y YP;

RESET ALL;
quit;
/* Fin del programa SAS*/

```

**Apéndice 3.** Programa en SAS para predecir el rendimiento de cruza simples de tomate con el empleo de marcadores moleculares de ADN y calcular la correlación entre valores predichos y observados.

```

/* Inicia programa SAS*/
/* VALIDACIÓN DE PREDICCIÓN DE HÍBRIDOS A PARTIR DE LA SELECCIÓN DE n
HÍBRIDOS OBSERVADOS (PREDICCIÓN CON UNA FRACCIÓN DE LOS HÍBRIDOS CON
RENDIMIENTO EVALUADO)*/
/* CÁLCULO DE MATRICES C Y V */
/* proc printto log='null'; quit; EVITA IMPRESIÓN
proc printto log=log; quit; PERMITE IMPRESIÓN
O BIEN: dm log 'clear'; LIMPIA LOG*/

OPTIONS LS=254 PS=1000 NOCENTER;
%LET NC=40; /* NÚMERO DE CRUZAS*/
%LET NRAN=35; /* NÚMERO DE LÍNEAS PREDICTORAS (SE ELIGEN NRAN DE NC
POSIBLES CRUZAS Y SE ESTIMAN LAS NC-NRAN NO ELEGIDAS) */
%let q=1000; /* N° DE MUESTREOS */

%macro enrique;
%do kkkZ=1 %to &q;

DATA UNO;DO R=1 TO &NC;Y=UNIFORM(0);OUTPUT;END;
PROC SORT;BY Y;
DATA DOS; SET UNO; IF _N_<=&NRAN; DROP Y;PROC SORT; BY R; TITLE'HÍBRIDOS
PREDICTORES'; *PROC PRINT; RUN;
DATA TRES; SET UNO; IF _N_>&NRAN; DROP Y;PROC SORT; BY R; TITLE'HÍBRIDOS A
PREDICHOS'; *PROC PRINT; RUN;
TITLE;RUN;
N O T A: Introducir el valor de las varianzas estimadas en el programa del apéndice 2*/
PROC IML;
VXE=1.1335234;
VYE=0.9331754;
VDE=0.4542061;
VEE=0.094406;
Beta=10.880124;

/* X MARCADORES MOLECULARES DE LÍNEAS DEL GRUPO HETERÓTICO X
Y MARCADORES MOLECULARES DE LÍNEAS DEL GRUPO HETERÓTICO Y
RXY RENDIMIENTO OBSERVADO DE HIBRIDOS PREDICTORES*/

/* LA MATRIZ Rxy DEBE CONTENER: COL1=AMBIENTE COL2=N° CRUZA COL3=PROG X
COL4=PROG Y COL=5 RENDIMIENTO de las cruza predictoras. Con fines didácticos, el
ejemplo considera sólo 20 marcadores moleculares para X y Y*/
X={1111011001110110011,
01011110101001100110,
01110000001000000110,
11111101111000100110,
00111101011101100010};

Y={001111000001101101110,
01101000001101101110,
01111110001001100110,

```

```
01101000001000000010,  
01101000001100000010,  
11101000001001100010,  
00001000001001100010,  
11111110001100000100};
```

```
Rxy={111110.292,  
121210.896,  
13138.946,  
141410.475,  
15159.844,  
161611.017,  
171710.089,  
181811.718,  
192112.070,  
1102210.718,  
1112311.008,  
1122410.233,  
1132511.317,  
1142611.293,  
1152711.532,  
1162812.596,  
1173111.118,  
1183211.013,  
1193310.116,  
1203410.238,  
1213511.482,  
1223611.063,  
1233710.881,  
1243811.161,  
1254110.427,  
126429.974,  
1274312.824,  
1284410.649,  
1294511.581,  
1304611.703,  
1314711.335,  
1324814.077,  
1335110.113,  
134529.279,  
135539.588,  
136549.035,  
137558.763,  
1385610.334,  
139579.608,  
1405812.207};
```

```
* PRINT X Y RXY;  
B=NCOL(X);nx=NROW(X);ny=NROW(Y); NOBS=NROW(Rxy); NAMB=1;  
/*PROPORCIONES DE X*/  
DO H=1 TO NX; Xh1=X[H,];DO I=1 TO NX; Xh2=X[I,]; Xh3=Xh1/Xh2; N=0; M=0;*PRINT H I;  
*PRINT XH1 XH2 XH3;  
DO J=1 TO B; K=Xh3[,J]; L=K[+]; IF L>0 THEN N=N+1; IF L=2 THEN M=M+1; DISx=M/N;  
END; JAC1x=JAC1x//DISx; *PRINT M N DISx JAC1x; END;END;*PRINT JAC1x;
```

```

DO P=1 TO NX; Q1=P*NX-NX+1; Q2=P*NX; JAC2x=JAC1x[Q1:Q2,]; Sxx=Sxx||JAC2x;*PRINT
q1 q2 JAC2x Sxx; END;*PRINT Sxx;

/*PROPORCIONES DE Y*/
DO H=1 TO NY; Yh1=Y[H,];DO I=1 TO NY; Yh2=Y[I,]; Yh3=Yh1//Yh2; N=0; M=0; *PRINT H I;
*PRINT YH1 YH2 YH3;
DO J=1 TO B; K=Yh3[,J]; L=K[+]; IF L>0 THEN N=N+1; IF L=2 THEN M=M+1; DISy=M/N;
END; JAC1y=JAC1y//DISy; *PRINT M N DISy JAC1y; END;END;*PRINT JAC1y;
DO P=1 TO NY; Q1=P*NY-NY+1; Q2=P*NY; JAC2y=JAC1y[Q1:Q2,]; Syy=Syy||JAC2y;*PRINT
q1 q2 JAC2y Syy; END;*PRINT Syy;

/*PROPORCIONES DE XY*/
DO H=1 TO NX; Xh=x[H,];DO I=1 TO NY; Yh=Y[I,]; XYh=Xh//Yh; N=0; M=0;*PRINT H I; *PRINT
XH YH XYH;
DO J=1 TO B; K=XYh[,J]; L=K[+]; IF L>0 THEN N=N+1; IF L=2 THEN M=M+1; DISxy=M/N;
END;
JAC1xy=JAC1xy//DISxy; *PRINT M N DISxy H I JAC1xy; END;END;*PRINT JAC1xy;
DO P=1 TO NX; Q1=P*NY-NY+1; Q2=P*NY; JAC2xy=t(JAC1xy[Q1:Q2,]);
Sxy=Sxy//JAC2xy;*PRINT q1 q2 JAC2xy Sxy; END;*PRINT Sxy;

/*PROMEDIOS XY y YX*/
DO R1=1 TO NX; AA=Sxy[R1,]; AA1=AA[:]; PRSxy=PRSxy//AA1; *PRINT R1 AA AA1
PRSxy;END; *PRINT PRSxy;
DO S1=1 TO NY; BB=Sxy[,S1]; BB1=BB[:]; PRSyx=PRSyx||BB1; *PRINT S1 BB BB1
PRSyx;END; *PRINT PRSyx;

/* fxx */
DO T=1 TO NX; DO U=1 TO NY;XX=Sxx[T,U]; XP1=PRSxy[T,]; XP2=PRSxy[U,]; *PRINT XX
XP1 XP2;
fxx1=(XX-(XP1+XP2)/2)/(1-(XP1+XP2)/2); IF fxx1<0 then fxx1=0; fxx2=fxx1/fxx1;*print fxx1
fxx2;END; END;
DO V=1 TO NY; w1=V*NX-NX+1; w2=V*NX; fxx3=fxx2[w1:w2,]; fxx=fxx||fxx3;*print fxx; END;
Pfx=fxx; *PRINT Pfx;

/* fyy */
DO T=1 TO NY; DO U=1 TO NX;YY=Syy[T,U]; YP1=PRSyx[T,]; YP2=PRSyx[U,]; *PRINT YY
YP1 YP2;
fyy1=(YY-(YP1+YP2)/2)/(1-(YP1+YP2)/2); IF fyy1<0 then fyy1=0; fyy2=fyy1/fyy1;*print fyy1
fyy2;END; END;
DO V=1 TO NX; w1=V*NY-NY+1; w2=V*NY; fyy3=fyy2[w1:w2,]; fyy=fyy||fyy3;*print fyy; END;
Pfy=fyy; *PRINT Pfy;

/*fxx*fyy*/
DO RR=1 TO NX; DO SS=1 TO NY; DO TT=1 TO NX; DO UU=1 TO NY;
fxxij=fxx[ss,rr]; fyykl=fyy[uu,tt]; EE=fxxij*fyykl; EEE=EEE//EE;
*print rr ss TT UU fxxij tt uu fyykl EE; end;end;end;end; *PRINT EEE;

NCT=(NX*NY)**2/NY; DO P=1 TO NCT; K1=P*NY-NY+1; K2=P*NY;
FF2=EEE[K1:K2,]; FF3=FF3||FF2;*PRINT K1 K2 ;*CC2 Cijkl; END; *PRINT FF3;
NC=NX*NY;*PRINT NC NX NY; DO RRR=1 TO NCT; K3=RRR*NC-NC+1; K4=RRR*NC; *PRINT
K3 K4;
FF4=FF3[K3:K4]; fxxfyy=fxxfyy//FF4; *PRINT FF4 FF5;END; Pfxfyy=fxxfyy;
*PRINT Pfxfyy;

```

```

/* Calculo de la matriz C (Matriz de varianzas genéticas entre el i-ésimo híbrido perdido y el j-
ésimo híbrido predictor) nombrada MatC */
/* NX=TOTAL DE LINEAS GRUPO X, NY=TOTAL DE LINEAS GRUPO Y,
NX=NROW(X);NY=NROW(Y);PRINT NX NY;

```

```

DO ii=1 TO NX;DO jj=1 TO NX;DO kk=1 TO NY;DO ll=1 TO NY;
Pfxxi=Pfxx[ii,jj]; Pfyyl=Pfy[kk,ll]; PMDXY=Pfxxi*Pfyyl;
PFVx=Pfxx[ii,jj]*VXE; PFVy=Pfy[kk,ll]*VYE; PFVd=PMDXY*VDE;
PFVxy=PFVx+PFVy+PFVd;MC2=PFVxy;MC1=MC1//MC2;
*PRINT ii kk jj ll MC2 PFVx PFVy PFVd PFVxy;end; end; end;
DDDDDD=NROW(MC1);*PRINT DDDDDD MC1;

```

```

/*NNX=NUMERO DE CRUZAS PREDICTORAS Y NNY=NUMERO DE CRUZAS A
PREDECIR*/

```

```

NNX=NX*NY;NNY=NX*NY;*PRINT NNX NNY;
NNC=(NNX*NNY)/NY;DO PP=1 TO NNC; KK1=PP*NY-NY+1; KK2=PP*NY;
FFFF2=MC1[KK1:KK2,]; FFFF3=FFFF3||FFFF2;*PRINT KK1 KK2 ; END;*PRINT FFFF3;
NNC1=NX*NX;DO PP1=1 TO NNC1; KK3=PP1*NY-NY+1; KK4=PP1*NY;
FFFF4=FFFF3[,KK3:KK4]; FFFF5=FFFF5//FFFF4;*PRINT KK3 KK4; END;*PRINT FFFF5;
DO PP2=1 TO NX; KK5=PP2*NNY-NY+1; KK6=PP2*NNY;
FFFF6=FFFF5[KK5:KK6,]; FFFF7=FFFF7||FFFF6;*PRINT KK5 KK6; END;*PRINT FFFF7;
MatC=FFFF7; *PRINT MatC;

```

```

/*Calculo de la Matriz V (Matriz de Varianzas-covarianzas fenotípicas entre los híbridos
predictores) Nombrada MatV. El número total de híbridos predictores es el producto de las
líneas de X y líneas de Y utilizadas para predecir.

```

```

N O T A: Capturar las varianzas calculadas */

```

```

NX=NROW(X);NY=NROW(Y);*PRINT NX NY;

```

```

/*

```

```

VXE=0.1812365; VYE=0.1957271; VDE=0.1060909;VEE=0.0128134;

```

```

*/

```

```

DO ii=1 TO NX;DO jj=1 TO NX;DO kk=1 TO NY;DO ll=1 TO NY;
Pfxxi=Pfxx[ii,jj]; Pfyyl=Pfy[kk,ll]; PMDXY=Pfxxi*Pfyyl;
PFVx=Pfxx[ii,jj]*VXE; PFVy=Pfy[kk,ll]*VYE; PFVd=PMDXY*VDE;
PFVxy=PFVx+PFVy+PFVd;MV2=PFVxy;MV1=MV1//MV2;
*PRINT ii kk jj ll MV2 PFVx PFVy PFVd PFVxy;end; end; end;
DDDDDD=NROW(MV1);*PRINT DDDDDD MV1;

```

```

/*NNX=NUMERO DE CRUZAS PREDICTORAS Y NNY=NUMERO DE CRUZAS A
PREDECIR*/

```

```

NNX=NX*NY;NNY=NX*NY;*PRINT NNX NNY;
NNC=(NNX*NNY)/NY;DO PP=1 TO NNC; KK1=PP*NY-NY+1; KK2=PP*NY;
FFFFFF2=MV1[KK1:KK2,]; FFFFF3=FFFFFF3||FFFFFF2;*PRINT KK1 KK2; END;*PRINT FFFFF3;
NNC1=NX*NX;DO PP1=1 TO NNC1; KK3=PP1*NY-NY+1; KK4=PP1*NY;
FFFFFF4=FFFFFF3[,KK3:KK4]; FFFFF5=FFFFFF5//FFFFFF4;*PRINT KK3 KK4; END;*PRINT
FFFFFF5;
DO PP2=1 TO NX; KK5=PP2*NNY-NY+1; KK6=PP2*NNY;
FFFFFF6=FFFFFF5[KK5:KK6,]; FFFFF7=FFFFFF7||FFFFFF6;*PRINT KK5 KK6 ;VP=FFFFFF7;
END;*PRINT VP;
er3=diag(j(1,nnx,vee)); *print er3; MatV=VP+er3; *print MatV;

```

```

/* HASTA AQUÍ SE HAN CONSTRUIDO LAS MATRICES COMPLETAS Fxx,Fyy,FxxFyy, C y V
1) SELECCIONAR AL AZAR COLUMNAS EQUIVALENTES DE C y V;*/

```

```

USE DOS; READ ALL INTO LS; USE TRES; READ ALL INTO LNS;

```

```

LS=LS` ;LNS=LNS` ;*PRINT LS LNS;
MATV=MATV[LS,LS]; MATC=MATC[LNS,LS];RXY1=RXY;
RXY=RXY[LS,]; RXY1=RXY1[LNS,];
*PRINT MATV MATC RXY;

/*El valor de Beta es generado en las iteraciones para calcular las varianzas*/
*Beta=2.332706;Z=DESIGN(Rxy[,2]);Y=Rxy[,5];X=DESIGN(Rxy[,1]);Cruza=RXY1[,2];Rend=RX
Y1[,5];*PRINT Rxy Z Y X;
/*Yp=Rendimientos de cruza predictoras y Ym=Rendimientos predichos*/
Yp=INV(T(Z)*Z)*T(Z)*(Y-(X*BETA)); *PRINT Y Yp;
Ym=MatC*inv(MatV)*Yp; *PRINT CRUZA REND YM;
R=REND; YPRED=YM; NMAT=R||YPRED; *PRINT NMAT;
CORR=CORR(NMAT); *PRINT CORR; VARNAMES={"OBSERVADO" "PREDICHO"};
MATTRIB CORR ROWNAME=VARNAMES COLNAME=VARNAMES; *PRINT CORR;
WA1=NMAT[,1]; PROMOBS=WA1[:];
WA2=NMAT[,2]; PROMPRE=WA2[:];
CORROP=CORR[1,2]; NUMPREDICTORAS=&NRAN; NUMPREDICHOS=NC-&NRAN;
IMPRIME=PROMOBS||PROMPRE||CORROP||NC||NUMPREDICTORAS||NUMPREDICHOS||&
KKKZ;
MATTRIB IMPRIME FORMAT=15.8;
PRINT IMPRIME;
RESET ALL;
dm log 'clear';
quit;
%end;
%mend Enrique;

%enrique;
run;
/*Fin del programa*/

```