

Prediction of fruit yield and firmness of tomato hybrids with BLUP and RR-BLUP using ISSR molecular markers

Predicción de rendimiento y firmeza de fruto de híbridos de tomate con BLUP y RR-BLUP mediante marcadores moleculares ISSR

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

Universidad Autónoma Chapingo, Instituto de Horticultura, Departamento de Fitotecnia.
Carretera México-Texcoco km 38.5, Chapingo, Estado de México, C. P. 56230, MÉXICO.
jsahagunc@yahoo.com.mx, tel.: (595) 952 1500, ext. 6185 (*Corresponding author)

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 we compared two methods based on genomic fingerprints, 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.

Keywords:

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

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.

Palabras clave: *Solanum lycopersicum* L., mejor predictor lineal e insesgado, regresión ridge BLUP, componentes de rendimiento.



Introduction

The management of heterosis that enables the production of outstanding hybrid varieties is probably the greatest success of plant breeding. The development of outstanding hybrid varieties, however, is no easy task: less than 1 % of single-cross maize (*Zea mays* L.) hybrids become commercial varieties (Bernardo, 1996), and something similar must occur in others crops, particularly among some that are autogamous species, such as tomato (*Solanum lycopersicum* L.). The greatest obstacle faced by plant breeders is the need to evaluate in the field a large number of experimental hybrids, which usually exceeds the capacity of breeding programs. To address this difficulty, theory and methods for predicting the genotypic value of varieties that do not physically exist have been developed. These methods are based on the formation and experimental evaluation of a small sample of crosses and on the coancestry of the parental lines, estimated on the basis of their pedigree records or genomic fingerprints (Henderson, 1985; Bernardo, 1994; Gbur et al., 2012).

In maize, the grain yield of single crosses has been predicted by a prediction equation that produces the best linear unbiased predictions (BLUP) (Bernardo, 1993, 1995). These predictions were based on RFLP molecular markers (Bernardo, 1994, 1996) and microsatellite markers (Balestre, Von Pinho, & Souza, 2010). Correlation coefficients (r) between observed and predicted yields ranged between 0.43 and 0.76, and between 0.55 and 0.70, respectively. Also in maize, Massman, Gordillo, Lorenzana, and Bernardo (2013) made a genomic prediction by ridge regression BLUP (RR-BLUP), based on SNP (single nucleotide polymorphism) markers, with prediction levels similar to those of BLUP. An attractive feature of the RR-BLUP method is that it uses all molecular markers and assumes that each of them represents the same proportion of genetic variance (Meuwissen, Hayes, & Goddard, 2001).

Although tomato is the second most important horticultural species in the world, and its breeding in recent decades has been oriented to the production and release of F_1 hybrid varieties, there are few reports on hybrid yield prediction. This may be due to the fact that plant breeding is mainly carried out by the private sector and only in a limited way by public programs. The cost of tomato production is very high, especially if it is done in a greenhouse. The phenotyping of new cultivars involves a high financial investment. To make it more productive, it is necessary to implement strategies, such as fruit yield prediction, that enable making and evaluating only those crosses with the greatest probability of success.

Mirshamsi, Farsi, Shahriari, and Nemati (2008) used RAPD markers to estimate genetic distances between

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 cruza 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 cruza, 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 cruza 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

parents and to determine the correlation between them and the fruit performances of hybrids produced by their crosses. However, the correlation coefficients obtained, due to their reduced value ($r < 0.40$), suggest that this method is of little practical use.

In breeding vegetables, especially tomato (despite its great methodological value), antecedents for predicting performance characters are scarce. Therefore, this study was undertaken in order to predict the firmness, number and total and commercial yield of fruits of tomato single crosses using the BLUP and ridge regression BLUP methods

Materials and methods

This research was carried out from February 2013 to November 2015 and consisted of three stages: 1) analysis of elite tomato lines with ISSR molecular markers, selection of parental lines and carrying out of directed crosses, 2) evaluation of fruit firmness, number and yield of single crosses and 3) prediction of performance of crosses using BLUP and RR-BLUP and validation of methods.

Selection of parental lines

In 2013, 39 elite F_8 tomato lines were sown in 200-well polystyrene trays with peat moss as substrate. Between 30 and 40 days after sowing (das), tissue samples of tender, green and apparently undamaged leaves were taken from 10 seedlings. The leaves were washed with 70 % alcohol and dried with paper towels. Next, 0.3 g of leaf were placed in a porcelain mortar, liquid nitrogen was added and it was ground to a fine powder, which was transferred to a 1.5-mL microtube with 600 μ L of extraction buffer (100 mM Tris HCl, 20 mM EDTA, 1.4 M NaCl, 2.0 % CTAB and 0.2 % 2-Mercaptoethanol), preheated to 65 °C in a thermoblock and stirred until homogenized. It was once again heated to 65 °C for 20 min. Next, it was removed from the heat, left to cool and centrifuged at 21,000 x g for 20 min. The solution was transferred to another microtube, 600 μ L of chloroform:isoamyl alcohol (24:1) were added and it was stirred for 20 min. Subsequently, it was centrifuged at 12,000 x g for 5 min and the aqueous (upper) phase was placed in another tube with 500 μ L of cold isopropanol, gently stirred by inversion, kept at -20 °C for 60 min to precipitate the DNA, centrifuged at 12,000 x g for 5 min and the supernatant was decanted. Then 700 μ L of solution (10 mM Tris HCl, 1 mM EDTA, 10 mM NaCl, pH 8.0) were added to dissolve the pellet and it was refrigerated (4 °C) for 12 h.

Once the DNA was resuspended, 4 μ L of RNase were added and incubated at 37 °C for 1 h. The DNA was reprecipitated with 70 μ L of 3 M sodium acetate and 700 μ L of cold isopropanol, stirred by inversion and stored

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.

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.

Selección de líneas progenitoras

En 2013 se sembraron 39 líneas élite F_8 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 μ L 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 μ L 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 μ L 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 μ L de solución para disolver (Tris HCl

in the freezer for 2 h. It was centrifuged at 8,000 x g for 10 min, the supernatant was removed and the pellet was washed with 500 µL of 70 % ethanol (reagent grade). It was centrifuged again, the supernatant was removed and the pellet dried at room temperature. Finally, it was dissolved with 50 µL of TE and stored at 4 °C.

To quantify DNA purity, a Thermo Scientific NanoDrop Lite spectrophotometer was used. Working solutions with 10 µg·µL⁻¹ DNA were prepared. DNA quality was verified on a 0.8 % agarose gel of 4 mm thickness and 0.5 % TAE, after which 10 µL of the working solution with DNA were deposited in each well. Electrophoresis was performed for 2 h in a Thermo Scientific Owl A3-1 chamber. The gel was stained with ethidium bromide solution at a concentration of 0.5 µg·mL⁻¹ for 20 min and photographed with a UVP TFM-26 transilluminator photodocumentation system.

In the amplification and separation of fragments by polymerase chain reaction (PCR), 29 ISSR (Inter-Simple Sequence Repeat) primers were used; the annealing temperatures are given in Table 1. The amplification reaction was performed with: 5.2 µL of H₂O, 10.0 µL of dNTP's [500 µM], 2.5 µL of buffer [10 x], 1.5 µL of MgCl₂ [50 mM], 3.0 µL of ISSR primer [10 ng·µL⁻¹], 0.3 µL of Taq DNA polymerase enzyme [5 u·µL⁻¹] and 2.5 µL of DNA [10 ng·µL⁻¹], giving a total volume of 25 µL.

PCR was carried out in a TECHNE-brand FTC41H2D thermocycler, with the following program: initial denaturation at 93 °C for 1 min, followed by 40 cycles, each with 20 s at 93 °C, 60 s at the annealing temperature of the tested primer and 20 s at 72 °C, and a final extension at 72 °C for 6 min. The sample was cooled to 10 °C. The final products were brought to electrophoresis in 2 % agarose gel with 1x TAE; ethidium bromide staining and photodocumentation were then performed.

From the matrix of binary data obtained from the genomic fingerprints, the Jaccard similarity coefficient (S_{ij}) was constructed for each pair of genotypes (Khattree & Naik, 2000). With the dissimilarity matrix ($d = 1 - S_{ij}$), clustering was performed using Ward's minimum variance method with Infostat statistical software. The result was the formation of three groups, of which two considered genetically different were chosen. These two were called X and Y.

Evaluation of fruit yield of intergroup crosses

Twelve elite F_8 high-fruit-yield lines were selected, six belonging to group X (lines L5BI, L37SI, L59BI, L61BI, L65BI and L76BI) and six to Y (lines L6BI, L45BI, L68BI, L69BD, L80BI and L88BI). They were sown in February 2013 and the 36 (6x6) intergroup direct single crosses (SCs) were performed; because no maternal effects in tomato

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 µL 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 µg·µL⁻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 µL 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 µg·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 emplearon 29 iniciadores ISSR (Inter Secuencias Simples Repetidas), las temperaturas de alineamiento aparecen en el Cuadro 1. La reacción de amplificación se realizó con: 5.2 µL de H₂O, 10.0 µL de dNTP's [500 µM], 2.5 µL de amortiguador [10 x], 1.5 µL de MgCl₂ [50 mM], 3.0 µL de iniciador ISSR [10 ng·µL⁻¹], 0.3 µL de enzima Taq ADN polimerasa [5 u·µL⁻¹] y 2.5 µL de ADN [10 ng·µL⁻¹], dado un volumen total de 25 µL.

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_{ij}) para cada par de genotipos (Khattree & Naik, 2000). Con la matriz de disimilitud ($d = 1 - S_{ij}$), 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.

have been reported, the lines of groups X and Y were used as females and males, respectively. Upon reaching consumption maturity, the fruits were harvested and the seed of each cross was separated and processed.

In order to determine the fruit yield of the SCs, they were sown in February 2014 and 2015, in 200-well polystyrene trays with peat moss as substrate. The transplant was made 35 das in a greenhouse under a hydroponic system. Black 18-L polyethylene bags with volcanic foam substrate were used. The nutrient solution used was the one proposed by Cadahia (2000) for tomato; the amount applied varied according to the phenological stage and climatic conditions. Plant density was 3.3 per m².

The experimental unit consisted of two pots, each with two plants. A randomized complete block experimental design with three replicates was used. Fruit yield was evaluated in the first four clusters. The response variables recorded were: fruit number (FN) per square meter, total fruit yield (TFY) in kg·m⁻², commercial fruit yield (CFY) in kg·m⁻² and fruit firmness (FF) recorded in kgf·cm⁻². A fruit reached the category of commercial fruit if its weight was greater than 100 g and it had no physical or mechanical damage, physiological disorder or damage by pests and diseases that demerit its appearance. FF was quantified in four fruits per experimental unit, twelve days after having been harvested in the breaker ripening stage and having remained at room temperature. A manual GY-1 fruit penetrometer (Sundoo Instruments) was used.

Prediction of single-cross performance using BLUP and RR-BLUP

BLUP

The coefficients of coancestry (f_{ij}) based on ISSR (Bernardo, 1993) were calculated according to the expression:

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

where S_{ij} is the Jaccard similarity estimated with molecular markers between the lines i and j of the group X, $S_{i.}$ is the average proportion of ISSR variants that i shares with all the lines of the opposite heterotic group (Y) and $S_{.j}$ is the average proportion of ISSR variants that the line j shares with all the lines of the opposite heterotic group (X). As by definition the coefficients of coancestry are not negative, those that resulted were equal to zero.

The model for fruit yield (y) was (Bernardo, 1994; Massman et al., 2013):

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

Evaluación de rendimiento de fruto de cruza intergrupales

Se seleccionaron 12 líneas elite F₈ 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) cruza 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 cruz 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⁻².

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: número de frutos (NF) por metro cuadrado, rendimiento de fruto total (RFT) en kg·m⁻², rendimiento de fruto comercial (RFC) en kg·m⁻² y firmeza del fruto (FF) registrada en kgf·cm⁻². 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 empleó un penetrómetro de fruta manual modelo GY-1 (Sundoo Instruments).

Predicción de rendimiento de cruza mediante BLUP y RR-BLUP

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

where y is a $p \times 1$ vector of p observed yields of the N single crosses, β is a $c \times 1$ vector of fixed effects of the c evaluation environments, g_1 is a $n_1 \times 1$ vector of random effects of general combining ability (GCA) of the lines of group X, g_2 is a $n_2 \times 1$ vector of random effects of GCA of the lines of group Y, s is a $N \times 1$ vector of the random effects of specific combining ability (SCA) of the N single crosses, e is the $p \times 1$ vector of residuals and X , Z_1 , Z_2 , and Z_3 are matrices whose elements are 0 or 1 to denote absence or presence of effects relating to y with β , g_1 , g_2 and s , respectively.

The covariance (Cov) between the intergroup SCs ij and $i'j'$ was expressed on the basis of a k loci model and the simplification resulting from the derivation of Melchinger (1988):

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

where $V_{A(X/Y)}$ is the additive variance of the alleles i and j from the group X or Y through the k loci of i' and j' involved and $V_{D(XY)}$ is the dominance variance through the k loci of the allelic pairs, each formed by an allele of X and one of Y. Restricted maximum likelihood estimates were obtained by iteration from equation routes for V_E (non-genetic variance), $V_{A(X/Y)}$, $V_{A(Y/X)}$ and $V_{D(XY)}$ (Henderson, 1985; Bernardo, 1994; Gbur et al., 2012). In addition to the prediction, estimates of these variances were used to obtain heritabilities.

Hybrid fruit yields were based on mixed model theory that produces linear, unbiased and minimum variance predictors (BLUP). Hybrid performance prediction was based on the observed yields of the remaining hybrids, called predictor hybrids. According to Bernardo (1994), the yield adjusted by fixed effects of these hybrids (y_T) was expressed as:

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

where R is a diagonal matrix and the elements are the reciprocal of the number of environments of the experiment. On the other hand, the vector formed with the predicted yields of the unknown hybrids (y_U) was calculated according to the expression:

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

where C_{UT} is a $N_U \times N_T$ matrix whose ij -th element is the genetic covariance between the i -th hybrid subject to prediction and the j -th predictor hybrid, and C_{TT} is the $N_T \times N_T$ matrix of the covariances of the n predictor hybrids. The C_{UT} elements and those outside the C_{TT} diagonal were calculated according to the expression of the covariance between an unknown hybrid and a predictor, in terms of coancestry and genetic variances.

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):

$$y = X\beta + Z_1 g_1 + Z_2 g_2 + Z_3 s + e$$

en donde y es un vector $p \times 1$ de p rendimientos observados de las N cruza simples, β 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 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 cruza simples, e es el vector $p \times 1$ de residuos y 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 β , 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 derivación de Melchinger (1988):

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

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.

Los rendimientos de fruto de los híbridos se basó en la teoría de los modelos mixtos que produce predictores 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.

RR-BLUP

The prediction of SC yield using genomewide prediction was made by means of RR-BLUP with the linear model proposed by Massman et al. (2013) for N_M markers:

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

where y , X , β and e were as already defined, m_1 is a $N_M \times 1$ vector of GCA random effects of markers of the maternal lines, m_2 is a $N_M \times 1$ vector of random effects of GCA of the markers of the paternal lines and m_3 is a $N_M \times 1$ vector of GCA random effects associated with the allelic markers of both heterotic groups. W_1 and W_2 are incidence matrices whose elements were equal to 1 or -1. The elements of W_3 at a given locus were the same as the product of the elements of W_1 and W_2 at the same locus, as proposed by Massman et al. (2013). In theory $m_1 \sim NM(0, \sigma^2_{m1}I)$, $m_2 \sim NM(0, \sigma^2_{m2}I)$, $m_3 \sim NM(0, \sigma^2_{m3}I)$, as well as m_1 , m_2 and m_3 are independent; e is a random error vector of $p \times 1$ dimension and $e \sim NM(0, \sigma^2_e I)$.

To evaluate the BLUP prediction method in terms of sample size, we worked with sets of n predictor hybrids ($n = 6, 12, 18, 24$ and 30), selected at random from the 36 possible options. To verify the stability of the predictions, 1,000 random samplings were performed on each number of predictor hybrids (n) and the fruit yield of the remaining $36-n$ hybrids was predicted in each case. Subsequently, for each value of n , the correlations between the observed and corresponding predicted yields were calculated. In the case of RR-BLUP, n corresponded to the training set and $36-n$ to the test set. The procedure described was applied to the four variables.

To predict using BLUP, a program was built in the IML module of the SAS statistical package. The prediction with RR-BLUP was made based on the BGLR function library (Pérez & de los Campos, 2014) of the statistical package R.

Results and discussion

Selection of parental lines using ISSR

A total of 276 bands, of which 215 were polymorphic, were amplified. The number of bands per ISSR primer ranged from 2 to 18, with an average of 9.52 (Table 1). These values are low compared to those of Aguilera et al. (2011), who reported the amplification of 9 to 22 bands per primer, with an average of 14.4, although they found a lower level of polymorphism (34 %), with an average of 5.3 polymorphic bands per primer. In the present work, 7.4 polymorphic bands were amplified per ISSR with 77.9 % general polymorphism, with a minimum of zero and a maximum of 100 %.

Ward's clustering method defined three sets of lines (Figure 1). The first group consisted of 17 lines (green to

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.

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 , β 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 o -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 $m_1 \sim NM(0, \sigma^2_{m1}I)$, $m_2 \sim NM(0, \sigma^2_{m2}I)$, $m_3 \sim NM(0, \sigma^2_{m3}I)$, además m_1 , m_2 y m_3 son independientes; e es un vector de errores aleatorios de dimensión $p \times 1$ y $e \sim NM(0, \sigma^2_e I)$.

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 1,000 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 *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.

Table 1. Sequence of ISSR primers used in the characterization of the genetic divergence of 39 elite tomato lines, annealing temperature (At), total number of amplified (TB) and polymorphic bands (PB) and percentage of polymorphism (PP).

Cuadro 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).

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

Y = GAC, X = GAT.

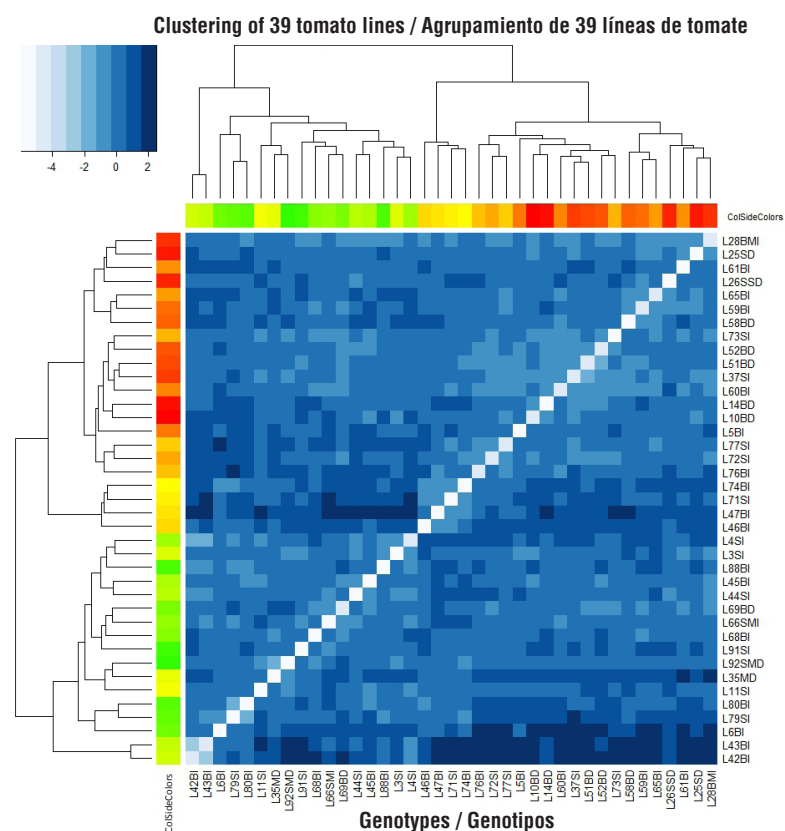


Figure 1. Clustering of 39 elite tomato lines, genotyped with 29 ISSR primers using the Jaccard distance and Ward's method.

Figura 1. Agrupamiento de 39 líneas élite de tomate, genotipeadas con 29 iniciadores ISSR mediante la distancia de Jaccard y el método de Ward.

yellow), the second with four (yellow tones) and the third with 18 (orange to intense red). In this figure, the lighter blue tones indicate shorter genetic distance; therefore, light blue tones predominate within each group and darker tonalities among groups. No association between the groups generated with molecular markers and a phenotypic characteristic was detected. In each group there is diversity in the type of growth (determinate and indeterminate), type of fruit (saladette and ball) and fruit color (yellow, orange and red).

Evaluation of fruit yield and firmness of intergroup crosses

Of the 36 intergroup crosses evaluated (Table 2), in 2014 the values of the four variables were higher, whereas in 2015 there were on average fewer fruits (7) and kilograms of total (0.94) and commercial fruit (1.3) produced. However, the crosses behaved similarly within each environment. The same cross recorded the highest (or lowest) value for FN, CFY and TFY in both evaluations. In the case of FF, the cross with the highest value was not the same in the two years evaluated; however, the

Resultados y discusión

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 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 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

Table 2. Maximums, minimums, means and estimates of variances and heritabilities of four characters of 36 tomato single crosses in two production cycles (2014 and 2015), grown hydroponically under greenhouse conditions.

Cuadro 2. Máximos, mínimos, medias y estimaciones de varianzas y heredabilidades de cuatro caracteres de 36 cruas simples de tomate en dos ciclos de producción (2014 y 2015), cultivados en condiciones de hidroponía e invernadero.

| Max, Min estimates of variances and h^2 / Max, Min estimaciones de varianzas y h^2 | FN / NF | TFY (kg·m ⁻²) / RFT (kg·m ⁻²) | CFY (kg·m ⁻²) / RFC (kg·m ⁻²) | FF (kgf·cm ⁻²) |
|---|------------------------|--|--|----------------------------|
| Evaluation 2014 / Evaluación 2014 | | | | |
| Max (cross / cruza) | 117.0 (L37SIxL80BI) | 13.723 (L76BIxL69BD) | 11.958 (L76BIxL68BI) | 13.83 (L5BIxL68BI) |
| Min (cross / cruza) | 51.7 (L76BIxL45BI) | 8.153 (L59BIxL88BI) | 4.373 (L37SIxL80BI) | 5.98 (L59BIxL88BI) |
| Mean / Media | 77.3 | 10.488 | 8.630 | 9.44 |
| Evaluation 2015 / Evaluación 2015 | | | | |
| Max (cross / cruza) | 102.0 (L37SIxL80BI) | 12.048 (L76BIxL69BD) | 10.736 (L76BIxL68BI) | 11.04 (L61BIxL68BI) |
| Min (cross / cruza) | 49.2 (L76BIxL45BI) | 7.358 (L65BIxL6BI) | 2.658 (L37SIxL80BI) | 6.81 (L59BIxL45BI) |
| Mean / Media | 70.7 | 9.553 | 7.332 | 8.94 |
| Variances / 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: maximums, Min: minimums, NF: fruit number, TFY: total fruit yield, CFY: commercial fruit yield, FF: fruit firmness, V_{ACG1} : variance of the effects of general combining ability of the lines of group X, V_{ACG2} : variance of the effects of general combining ability of the lines of group Y, V_{ACE} : variance of the effects of specific combining ability of the crosses, V_E : error variance and h^2 : heritability. Estimates were made using molecular markers and BLUP.

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 cruas, V_E : varianza del error y h^2 : heredabilidad. La estimación se hizo mediante marcadores moleculares y BLUP.

highest FF was associated with crosses involving the L68BI line and the lowest FF corresponded to crosses with the L59BI line (Table 2).

Prediction of single-cross performance using BLUP and RR-BLUP

The coefficients of coancestry estimated by molecular markers were, in general, low. In the group X lines, they ranged from 0.01 between the pair of L5BI and L65BI lines, to 0.327 between L59BI and L65BI, with average coancestry of 0.108. In the group Y lines, inferior values, from 0 (in six combinations) to 0.175 between the pair L45BI and L80BI, with an average of 0.06, were presented.

The estimates of genetic and non-genetic variances of the four variables studied, with ISSR markers using BLUP (Table 2), allowed calculating the heritability (h^2) of each character (Bernardo, 1996). Thus, for TFY, CFY, and FF, the heritability was very high (0.98); FN recorded $h^2 = 0.99$. In the case of TFY, the estimated value of h^2 in this study is far from that published by Dordevic, Zecevic, Zdravkovic, Zivanovic, and Todorovic (2010) of 0.451. In addition to random error variations, this discrepancy can be attributed to differences between genetic materials and environments. Wessel-Beaver and Scott (1992) obtained h^2 estimates of 0.65 and 0.81 for yield of the same tomato population grown in Puerto Rico and Florida, respectively.

The heritability values in this study allow predicting success in the selection of the best hybrids, as well as in a selection-based breeding program. The ratio of specific combining ability variance (V_{SCA}) to total genetic variance among SCs was 0.107, 0.106, 0.135 and 0.117, for FN, TFY, CFY and FF, respectively. Similar values were reported by Massman et al. (2013) in maize by using BLUP (from 0.10 for moisture to 0.18 for grain yield). Based on the above, it can be inferred that in this research BLUP exploited additive effects to a greater extent than dominance.

Correlations between observed and predicted yields

The average of 1,000 correlations between the predicted value and the recorded field value for each set of n predictor hybrids is summarized in Table 3. Both BLUP and RR-BLUP predicted adequately in the four variables analyzed, and in all cases BLUP estimated values closer to the real ones than RR-BLUP. As expected, when n was increased the values of the correlations were also increased, although from $n = 24$ these increases were of low magnitude and even a reduction was recorded. In this study, with 18 predictor hybrids ($n = 18$), the greatest stability in the predictions was obtained and it would be the number of ideal predictors in a cost scenario where an increase in n implies greater investment.

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).

Evaluación de rendimiento y firmeza de fruto de cruas intergrupales

De las 36 cruas intergrupales evaluadas (Cuadro 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 cruas se comportaron de manera similar dentro de cada ambiente. La misma crua registró el valor más alto (o más bajo) para el NF, RFC y RFT en ambas evaluaciones. En el caso de la FF, la crua con el mayor registro no fue la misma en los dos años evaluados; no obstante, la mayor FF estuvo asociada a cruas que involucraron a la línea L68BI y la menor FF correspondió a cruas con la línea L59BI (Cuadro 2).

Predicción de rendimiento de cruas 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 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

Table 3. Average cross-validation correlations for two methods (BLUP and RR-BLUP) for predicting tomato hybrids in terms of four variables. Data correspond to 1,000 independent random samples of 36 predictor hybrids.

Cuadro 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 | FN | | TFY (kg·m ⁻²)/ RFT (kg·m ⁻²) | | CFY (kg·m ⁻²)/ RFC (kg·m ⁻²) | | FF (kgf·cm ²) | |
|------------|------|---------|---|---------|---|---------|---------------------------|---------|
| | 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 |
| Mean/Media | 0.74 | 0.70 | 0.70 | 0.60 | 0.51 | 0.44 | 0.61 | 0.57 |

N: number of predictor hybrids, BLUP: best linear unbiased prediction, RR-BLUP: ridge regression BLUP, FN: fruit number, TFY: total fruit yield, CFY: commercial fruit yield and FF: fruit firmness.

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.

When considering each variable separately, FN recorded the highest correlation coefficients (from 0.45 to 0.83), followed by TFY (0.36 to 0.73), FF (0.32 to 0.68) and lastly CFY (0.25 to 0.54), with RR-BLUP. The correlation values between observed and predicted values with BLUP maintained the same trend: FN (0.53 to 0.83), TFY (0.52 to 0.79), FF (0.41 to 0.71) and lastly CFY (0.36 to 0.58).

The correlations obtained for TFY are generally similar to those reported by Hernández-Ibáñez, Sahagún-Castellanos, Rodríguez-Pérez, and Peña-Ortega (2014) in tomato (between 0.45 and 0.79), and by Bernardo (1994) in maize grain yield. The latter group used RFLPs and the highest correlations were 0.80. The high correlations obtained suggest that the method used to predict the yield of tomato hybrids is promising. However, extrapolation must be made with reservations, since the correlation between observed and predicted hybrids increases as the number of predictor hybrids increases, but only up to a certain level, after which one runs the risk of having smaller and even negative correlations. In addition, the results presented correspond to a relatively small set of SCs. Bernardo (1996) applied this methodology at a larger scale (16 combinations of nine heterotic groups in maize), and the correlations between predicted and observed yields that he obtained were slightly lower (from 0.426 to 0.762), very similar to those obtained in the present work.

Massman et al. (2013) made a comparison between BLUP and RR-BLUP on four variables in maize and produced high correlations (0.87, 0.90, 0.69 and 0.84, for grain yield, grain moisture, stalk and root lodging, respectively), without significant differences between methodologies; therefore, they affirmed that RR-BLUP

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.

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 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.

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,

did not surpass the results obtained with BLUP. These results suggest that the genetic factor may modify the predictive ability of the models.

The results obtained in the present work in tomato can be considered to be in agreement with those reported in maize, despite the fact that they are contrasting species. While tomato is an autogamous species with a low level of heterosis, maize is a cross-pollinated species with high heterosis expressions. In tomato, in addition to TFY, CFY and FF are very important after the harvest, so other genes in addition to those that strictly control yield are involved and thus adversely affect the correlations between the various yield components and the production of tomato fruit. Epistatic and linkage effects, among other genetic factors, in addition to the environment, can also affect the correlation. This explains the difficulty in predicting these characters, and therefore the low prediction levels obtained.

Conclusions

The correlation coefficients between observed and predicted values for FN, TFY, CFY and FF, estimated by BLUP and RR-BLUP in tomato SCs, ranged between 0.25 and 0.83, values that suggest that these methods have value for breeders.

BLUP consistently recorded higher correlation results between observed and predicted values, compared to RR-BLUP. In both cases, the magnitude of the correlations always increased when n increased. Therefore, in order to achieve greater accuracy and reliability in predictions made with BLUP or RR-BLUP, a greater number of predictor crosses must be evaluated in the field, in as many environments as funding allows.

Acknowledgments

The authors thank Mexico's National Science and Technology Council (CONACYT) for the Ph.D. scholarship granted to the first author, Q.F.B. Ricardo Gaspar Hernández for the technical support in DNA molecular markers, and the anonymous reviewers for their valuable contributions that substantially improved the manuscript.

End of English version

References / Referencias

- 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
- 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.

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

- 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., West, 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. *Phyton Revista Internacional de Botánica Experimental*, 83, 311-318. Retrieved from <http://www.revistaphyton.fund-romuloraggio.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
- 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 cruza predictoras, en tantos ambientes como el presupuesto lo permita.

Agradecimientos

Al Concejo 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. A los revisores anónimos por sus valiosas contribuciones que permitieron mejorar sustancialmente el manuscrito.

Fin de la versión en español

- Melchinger, A. E. (1988). Means, variances, and covariances between relatives in hybrid populations with disequilibrium in the parent populations. In: Wei, B. S., & Eisen, E. J. (Eds.), *Proc. 2nd. 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. Retrieved from <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. Retrieved from <http://journal.ashspublications.org/content/117/5/867.full.pdf+html>