

Predicción de valores genéticos aditivos en genotipos de guayabo (*Psidium guajava*) (Myrtaceae)

Prediction of breeding values in guava (*Psidium guajava*) (Myrtaceae) genotypes

Leneidy Pérez Pelea^{1,*}, Evelyn Bandera Fernández¹, Juliette Valdés-Infante Herrero², Bárbara Velázquez Palenzuela² y María Teresa Cornide Hernández³

RESUMEN

Durante el proceso de selección en plantas, es importante obtener estimados más precisos de la superioridad genética de los individuos. En las plantas perennes, se realizan mediciones durante varias cosechas en las mismas plantas, por lo que se debe utilizar un método de análisis que tenga en cuenta las medidas repetidas. El procedimiento óptimo para la estimación-predicción de parámetros genéticos en las especies perennes, es el de Máxima Verosimilitud Restringida-Mejor Predictor Lineal Insesgado. El presente trabajo se realizó con el objetivo de seleccionar la estructura de covarianza que mejor ajusta los datos y predecir el valor genético individual en progenies de guayabo (*Psidium guajava*), como criterio de selección de genotipos. Se evaluaron 10 caracteres cuantitativos vegetativos y del fruto durante tres años, en tres familias de hermanos completos de guayabo en Cuba. Se realizaron procedimientos *MIXED* para seleccionar la estructura de covarianza que mejor ajustaba los datos, para cada uno de los caracteres, en las tres familias. Posteriormente, se emplearon procedimientos *GLIMMIX* en los cuales se consideraron los factores genotipo e interacción genotipo-año como aleatorios, se especificó el factor año como medida repetida y la estructura de covarianza. Los análisis se realizaron en el programa SAS versión 9.3. Se determinó que la estructura de covarianza Toeplitz era la más adecuada. Se seleccionaron ocho genotipos con los mayores valores genéticos para varios caracteres en las tres familias, que pueden considerarse promisorios para disímiles propósitos.

Palabras clave: modelos lineales generalizados, mejor predictor lineal insesgado, máxima verosimilitud restringida, medidas repetidas

ABSTRACT

During the plant selection process, it is important to make more precise estimates of the genetic superiority of individuals. In perennial plants, repeated measurements are made during several crops in the same plants; therefore, an analysis method involving repeated measures is appropriate. The optimum procedure for the estimation-prediction of genetic parameters in perennial species is Restricted Maximum Likelihood -Best Linear Unbiased Predictor. The objective of the present work is to select the best covariance structure to the data and predict the individual breeding value in guava (*Psidium guajava*) progenies, as genotype selection criteria. Ten quantitative plant and fruit traits were evaluated during three years, in three guava full-sib families in Cuba. *MIXED* procedures were made to select the best covariance structure, for each trait in the three families. After that, *GLIMMIX* procedures were used, in which genotype and genotype-year interaction as random factors, year as repeated measure and covariance structure were specified. The analysis were made with the employ of SAS program version 9.3. It was determined that the Toeplitz covariance structure was the best. A group of eight genotypes promissories for dissimilar purposes were selected with the highest breeding values for various traits in the three families.

Keywords: generalized linear models, best linear unbiased predictor, restricted maximum likelihood, repeated measures

Recibido: abril 2019 **Aceptado:** junio 2019

Publicado online 2 de agosto de 2019. ISSN 2410-5546 RNPS 2372 (DIGITAL) - ISSN 0253-5696 RNPS 0060 (IMPRESA)

INTRODUCCIÓN

El guayabo (*Psidium guajava* L.) es considerado, dentro del género *Psidium*, como uno de los representantes de mayor valor nutricional, dado fundamentalmente por su aporte en vitaminas y sales minerales, entre las que se destaca el ácido ascórbico (vitamina C), por su importancia para la salud y la dieta humana (Ali & al. 2014). Sus frutos son utilizados en la industria de jugos, néctares, jaleas, helados, dulces y para el consumo en fresco. Las altas concentraciones de pectina en sus frutos, tienen

un papel importante en la reducción del colesterol y la disminución del riesgo de enfermedad cardiovascular. Por estas razones, el guayabo es considerado como “la manzana de los pobres”, porque es un frutal que tiene gran disponibilidad en el mercado y está accesible para las personas de menos recursos (Patel & al. 2015).

El guayabo es uno de los frutales tropicales que ha mostrado un gran incremento en su área de plantación. La demanda y expansión del cultivo se ha desarrollado por la necesidad de obtener cultivares, con frutos de buena calidad y larga duración, con alto contenido de nutrientes y resistentes a enfermedades (Rawls & al. 2015). El uso de cultivares adaptados a diferentes condiciones edafoclimáticas, con potencial para producir frutos de buena calidad, ha tenido gran influencia en el incremento del área cultivada y la consolidación del cultivo en varios países (Quintal & al. 2017).

¹Departamento de Biología Vegetal, Facultad de Biología, Universidad de la Habana, Calle 25 # 455, e/ I y J Vedado, La Habana, Cuba. C.P. 10400. ²Instituto de Investigaciones en Fruticultura Tropical, Ave 7ma # 3005 / 30 y 32, Miramar, La Habana, Cuba. ³Instituto de Investigaciones de la Caña de Azúcar (INICA), Grupo Empresarial AZCUBA, Carretera al CAI Martínez Prieto Km 2½, Boyeros, La Habana, Cuba. *Autor para correspondencia (e-mail: lene@fbio.uh.cu).

Editor encargado: José Angel García-Beltrán.

Durante el proceso de selección de plantas, es importante obtener estimados más precisos de la superioridad genética de los individuos lo que permitirá seleccionar aquellos genotipos que presenten los caracteres de interés para el mejorador, y que a su vez, puedan ser transmitidos a la descendencia. En las plantas perennes como el guayabo, se realizan mediciones repetidas en los individuos, pues se cosechan las mismas plantas durante varios años. Por tal razón, resulta apropiado emplear métodos de análisis que tengan en cuenta las mediciones repetidas (Cargnin 2016).

Los experimentos con medidas repetidas no cumplen con premisas básicas del análisis de varianza como la normalidad, la homogeneidad de varianzas y la independencia de errores, porque las mediciones se realizan sobre la misma unidad experimental en el tiempo, lo que resulta en una correlación entre ellas. La correlación entre las mediciones cambia con el tiempo, de manera que las mediciones tomadas en puntos de tiempo adyacentes están más correlacionadas que las tomadas en puntos de tiempo más distantes; lo que resulta en una compleja estructura de covarianza que debe ser modelada para brindar los apropiados errores estándar de la prueba estadística (Davis 2017).

La metodología de los modelos mixtos permite analizar correcta y eficientemente los datos de experimentos con medidas repetidas, a partir de la modelación de la estructura de covarianza, que considera las correlaciones entre las medidas y la presencia de varianzas heterogéneas (Bandera & Pérez 2018). Se han desarrollado varios métodos para facilitar la selección de la estructura de covarianza que explique mejor la variabilidad y correlación entre las medidas repetidas. Los principales criterios utilizados en los programas estadísticos, para seleccionar los modelos, son el criterio de Akaike y el Bayesiano, los cuales están basados en valores de verosimilitud y dependen del número de observaciones y parámetros del modelo. La metodología Máxima Verosimilitud Restringida – Mejor Predictor Lineal Inssegado (*REML – BLUP*), es el procedimiento analítico adecuado para la estimación de componentes de varianza y la predicción de valores genéticos, en experimentos de este tipo (Quintal & al. 2017).

Basados en los elementos antes expuestos, los objetivos del presente trabajo fueron: seleccionar la estructura de covarianza más apropiada para los datos con el empleo de los criterios de Akaike y Bayesiano, y predecir el valor genético aditivo individual en progenies de guayabo, como criterio de selección de genotipos, con el empleo del procedimiento *REML – BLUP*.

MATERIALES Y MÉTODOS

Área de estudio

El estudio fue desarrollado en la Unidad Científico-Tecnológica de Base (UCTB) de Alquízar, provincia Artemisa, perteneciente al Instituto de Investigaciones en Fruticultura Tropical (IIFT) de Cuba. La UCTB se encuentra ubicada en los 22° 47' de latitud Norte y los 82° 31' de longitud Oeste, a 11 msnm, sobre un suelo Ferralsol éutrico, con una topografía llana de pendiente cero (Hernández & al. 2015).

Material vegetal empleado en los experimentos

Se evaluaron tres familias de hermanos completos de guayabo, obtenidas a partir de cruzamientos intraespecíficos realizados mediante polinización controlada en el año 2001. Se utilizaron como progenitores femeninos, tres plantas del cultivar 'Enana Roja Cubana' ('EEA 18-40'), que se emplea a escala comercial en el país. Como progenitores masculinos se emplearon los cultivares: 'N6', 'Suprema Roja' y 'Belic L-207'. Estos cultivares fueron seleccionados como progenitores, por presentar gran variabilidad fenotípica y genotípica, detectada al realizar la evaluación del banco de germoplasma del cultivo por medio de marcadores morfológicos y moleculares (Rodríguez & al. 2004).

Las semillas obtenidas de cada cruzamiento, se sembraron en semilleros y posteriormente se trasplantaron a bolsas individuales de 26 x 46 cm que contenían suelo Ferralsol éutrico y materia orgánica (cachaza) a la relación 3:1. Cuando las plantas tenían entre 50 y 60 cm de altura se plantaron en la UCTB, con un marco de plantación de 6 x 5 m. Cada una de las plantas obtenidas se consideró un genotipo diferente, debido a que se obtuvieron de semillas, por lo que había solo una réplica por genotipo. Las tres familias se plantaron de forma adyacente una a la otra, formando un bloque compacto en el mismo lote, junto a los progenitores.

En la familia 1 ('EEA 18-40' x 'N6') se obtuvieron 96 plantas (descendientes), en la familia 2 ('EEA 18-40' x 'Suprema Roja') 73 plantas y en la familia 3 ('EEA 18-40' x 'Belic L-207') 47 plantas. Las progenies difirieron en el número de descendientes debido a que los frutos obtenidos en cada cruzamiento no tenían el mismo número de semillas, y algunas plantas se perdieron por afectaciones producidas por factores bióticos y abióticos en los diferentes estadios de su crecimiento y desarrollo.

Las plantas se mantuvieron con riego localizado, mediante la técnica de riego por goteo con emisores marca RAM de 2,3 L.h⁻¹, espaciados a 0,65 m dentro de un lateral de 20 mm de diámetro. El riego se aplicó con dosis fijas e intervalos fijos (días alternos) y fue suspendido durante los eventos de fuertes lluvias. Las

labores culturales, la fertilización y el control fitosanitario fueron realizados según el Instructivo Técnico del Cultivo (MINAG 2011).

Evaluación de caracteres cuantitativos

En las progenies resultantes de los tres cruzamientos y sus progenitores, se evaluaron diez caracteres cuantitativos, teniendo en cuenta las recomendaciones del descriptor del cultivo publicado por la Unión Internacional para la Protección de Nuevas Variedades de Plantas (UPOV, por sus siglas en inglés) (UPOV 1987) y por (Rodríguez & *al.* 2010). Los caracteres evaluados fueron: largo de la hoja (mm), ancho de la hoja (mm), altura de la planta (m), masa del fruto (g), largo del fruto (mm), ancho del fruto (mm), espesor externo de la pulpa (mm), espesor interno de la pulpa (mm), número de semillas por fruto y masa total de las semillas por fruto (g).

Las plantas se comenzaron a evaluar a los cinco años de edad, a partir del 2006, y se realizaron mediciones durante tres años consecutivos (2006 - 2008) en cada uno de los genotipos de las tres familias. Los caracteres vegetativos fueron evaluados en el período marzo - abril y los caracteres del fruto en el período agosto - septiembre, que es el pico de cosecha de verano. Los frutos se cosecharon en su madurez fisiológica y fueron evaluados en completa maduración, dos o tres días después de cosechados. Para la medición de estos caracteres se utilizaron balanza técnica, reglas milimetradas y pie de rey.

Estudio de los caracteres cuantitativos

Primeramente, se efectuó un estudio de los caracteres evaluados para comprobar si los datos experimentales se ajustaban a las premisas del Análisis de Varianza (ANOVA), y detectar la presencia de valores atípicos (*outliers*). El estudio de los caracteres se realizó a partir del desarrollo de pruebas de bondad de ajuste a la distribución normal, con el empleo del procedimiento *UNIVARIATE*. Posteriormente, al verificar el comportamiento no normal de los caracteres, se empleó el procedimiento *SEVERITY* para determinar a cuál de las distribuciones de la familia exponencial se ajustaba cada uno de los caracteres. Los procedimientos se desarrollaron con el empleo del programa Sistema de Análisis Estadístico (SAS, por sus siglas en inglés), versión 9.3 (SAS Institute 2012).

Selección de la estructura de covarianza y predicción de valores genéticos aditivos, para los caracteres evaluados en las tres familias de hermanos completos de guayabo
Como las mediciones se realizaron en las mismas plantas durante tres años consecutivos, se tuvo en cuenta la presencia de medidas repetidas en el experimento, para obtener estimados más precisos de los parámetros genéticos. En esta situación, se incumple con las pre-

misas de normalidad e independencia de errores, por lo que se debe especificar en el modelo la mejor estructura de covarianza.

Una vez conocida la distribución de ajuste de cada uno de los caracteres, se empleó el procedimiento *MIXED* para seleccionar la estructura de covarianza que mejor ajustaba los datos, en cada uno de los caracteres, en las tres familias. Se modelaron cinco estructuras de covarianza con diferente grado de complejidad: Simetría Compuesta, Componente de Varianza, Toeplitz, Autorregresiva de orden 1 y No estructurada. Se usaron varios criterios de bondad de ajuste, basados en los valores de los estadísticos de Akaike (*AIC*, *Akaike information criterion*), Akaike corregido para poblaciones finitas (*AICC*, *Akaike information criterion corrected*) y Bayesianos (*BIC*, *Bayesian information criterion*). Estos criterios son funciones del logaritmo de la verosimilitud y pueden ser comparados entre modelos. Para seleccionar el modelo más apropiado, se consideró aquel en el cual se obtuvieron los menores valores de los criterios de información, según lo planteado por Littell & *al.* (2006).

Los mejores predictores lineales insesgados (*BLUP*, *Best Linear Unbiased Predictors*) fueron obtenidos con el empleo del procedimiento *GLIMMIX*, que considera un modelo lineal generalizado mixto, que se utiliza en datos no normales y puede modelar efectos aleatorios y/o repetidos. Este procedimiento realiza la estimación de los efectos por el método de Máxima Verosimilitud Restringida (*REML*, *Restricted Maximum Likelihood*). En el modelo mixto se consideró la media general y el factor año como efectos fijos y el error, el factor genotipo y la interacción genotipo - año como efectos aleatorios. Además, se especificó el factor año como medida repetida, la estructura de covarianza que mejor ajustó el modelo y la función de enlace correspondiente a la distribución de los datos en cada carácter, previamente determinada con el empleo del procedimiento *SEVERITY*.

La ecuación del modelo mixto empleado fue la siguiente: $y = Xr + Zg + Wa + e$, donde: y es el vector de observaciones, r es el vector de mediciones que incluye la media general y se considera de efectos fijos, g es el vector de efectos genotípicos individuales considerados aleatorios, a es el vector de efectos ambientales (años de evaluación) considerado de efectos fijos, e es el vector de errores o residuales (aleatorios), X , Z y W son las matrices de incidencia conocidas, asociadas con los vectores r , g y a , respectivamente.

Los valores *BLUP* obtenidos se emplearon para predecir los valores genéticos aditivos (*Breeding value*) de cada uno de los genotipos en las tres familias de hermanos completos. Los valores genéticos aditivos obtenidos

se pueden utilizar para predecir el comportamiento de algunos genotipos con valores superiores en caracteres cuantitativos de interés para el mejoramiento.

Los procedimientos se desarrollaron con el empleo del programa Sistema de Análisis Estadístico (SAS, por sus siglas en inglés), versión 9.3 (SAS Institute 2012).

RESULTADOS

Estudio de los caracteres cuantitativos evaluados

Con el empleo de procedimiento *UNIVARIATE* se realizaron pruebas de bondad de ajuste a la distribución normal, como las pruebas de *Shapiro-Wilk*, *Kolmogorov-Smirnov*, *Cramer-von Mises* y *Anderson-Darling*, las cuales indicaron que los caracteres analizados, no se ajustaban a la distribución normal, en cada una de las familias objeto de estudio. No se detectaron valores atípicos en los caracteres evaluados.

Al comprobar el ajuste no normal de los caracteres en estudio, se procedió a realizar el procedimiento *SEVERITY*, que permitió determinar a cual distribución de la familia exponencial se ajustaba cada carácter, en cada una de las familias. Este procedimiento utiliza los criterios de Akaike (*AIC*), Akaike corregido (*AICC*) y Bayesiano (*BIC*), para seleccionar la distribución de mejor ajuste de los datos, a partir del menor valor generado para estos criterios de ajuste.

Se determinó que, en las tres familias, los caracteres se ajustaban a las distribuciones Gamma o Log-normal (Log-N) (Tabla I), las cuales pertenecen a la familia exponencial. Por esta razón, se requiere el uso de un modelo

TABLA I

Distribuciones de la familia exponencial observadas en los caracteres evaluados en las tres familias de guayabo

TABLE I

Distributions of the exponential family observed in the traits evaluated in three guava families

Carácter	Familia 1	Familia 2	Familia 3
Largo de la hoja	Gamma	Gamma	Gamma
Ancho de la hoja	Gamma	Gamma	Gamma
Altura de la planta	Gamma	Gamma	Gamma
Masa del fruto	Log-Normal	Gamma	Log-Normal
Largo del fruto	Log-Normal	Log-Normal	Gamma
Ancho del fruto	Log-Normal	Log-Normal	Log-Normal
Espesor externo de la pulpa	Log-Normal	Log-Normal	Log-Normal
Espesor interno de la pulpa	Gamma	Gamma	Gamma
Número de semillas por fruto	Gamma	Gamma	Gamma
Masa total de semillas por fruto	Gamma	Gamma	Gamma

lineal generalizado mixto para analizar los valores de estos caracteres que no se ajustan a la normalidad y no son independientes, porque provienen de mediciones repetidas en el tiempo sobre las mismas plantas. Los modelos lineales generalizados utilizan una función de enlace de tipo no lineal, que transforma los datos para lograr un mejor ajuste de los mismos y permiten retransformar las medias de mínimos cuadrados resultantes a la unidad de medida original de la variable. En las variables con distribución Gamma se utilizó la función logarítmica y en aquellas con distribución log-Normal, la función Identidad, según lo planteado por Davis (2018).

Selección de la estructura de covarianza y predicción de valores genéticos aditivos, para los caracteres evaluados en las tres familias de hermanos completos de guayabo

Con los resultados obtenidos de los procedimientos *MIXED* se pudo seleccionar la estructura de covarianza que mejor ajustaba a los datos con medidas repetidas, a partir de los valores de los estadísticos de ajuste que brindan los criterios de información de Akaike (*AIC*), Akaike corregido para poblaciones finitas (*AICC*) y Bayesiano (*BIC*). En la Tabla II se muestran los valores de los estadísticos de ajuste obtenidos en las estructuras de covarianza Autorregresiva de orden 1, Toeplitz, Simetría Compuesta, No estructurada y Componente de Varianza, para el carácter masa del fruto evaluado en la familia 1 ('EEA 18-40' x 'N6'). Resultados similares se observaron en los restantes caracteres en las tres familias.

Se observaron valores muy similares de los estadísticos de ajuste (*AIC*, *AICC* y *BIC*) en las cinco estructuras de covarianza empleadas, aunque los menores valores se obtuvieron con el criterio de Akaike (*AIC*). Estos resultados se obtuvieron en los 10 caracteres, en las tres familias. En las estructuras Toeplitz, Componente de Varianza y No estructurada se obtuvieron los menores valores de los estadísticos de ajuste en cada carácter, los cuales fueron iguales (Tabla II). La estructura de covarianza Toeplitz fue la seleccionada para la realización de los análisis posteriores (procedimientos *GLIMMIX*), debido a que es la estructura que tiene sentido cuando las observaciones están igualmente espaciadas y la estructura de correlación no cambia apreciablemente en el tiempo, según lo planteado por Littell & al. (2006) y Bello (2016); condiciones que se cumplen en las mediciones obtenidas en el presente trabajo.

Una acción de relevancia fundamental en el proceso de mejoramiento genético de especies perennes es la predicción de los valores genéticos aditivos de los candidatos a selección. Cuando se consideraron el factor genotipo y la interacción genotipo-año como efectos aleatorios en el procedimiento *GLIMMIX*, con estructura de covarianza Toeplitz y el factor año como medida repetida, se obtuvieron los Mejores Predictores Lineales

TABLA II

Valores de los estadísticos de ajuste de los criterios de información de Akaike (AIC), Akaike corregido (AICC) y Bayesiano (BIC) obtenidos para el carácter masa del fruto en la familia 1 ('EEA 18-40' x 'N6')

TABLE II

Values of the fit statistics from Akaike's information criterion (AIC), Akaike corrected (AICC) and Bayesian (BIC), obtained by mass of the fruit in family 1 ('EEA 18-40' x 'N6')

Estadísticos de ajuste	Estructuras de covarianza				
	Autorregresiva de orden 1	Toeplitz	Simetría Compuesta	No estructurada	Componente de varianza
AIC	16066,8	16064,8	16066,8	16064,8	16064,8
AICC	16081,6	16079,3	16081,6	16079,3	16079,3
BIC	16596,1	16588,8	16596,1	16588,8	16588,8

Insesgados (*BLUP*). Estos valores constituyen los predictores de los valores genéticos aditivos (*Breeding value*), y se obtuvieron por el método de Máxima Verosimilitud Restringida en el modelo mixto. Con los valores predichos, se pueden ordenar los genotipos en cada una de las familias de acuerdo a sus valores genéticos aditivos. Dentro de cada familia, los genotipos con mayores valores de *BLUP* serán los de mayor valor genético aditivo para el carácter analizado.

A partir de los valores genéticos individuales obtenidos en cada una de las familias, se pudo determinar que un grupo de 109 genotipos tendrán los mayores valores genéticos aditivos en más de dos de los caracteres evaluados. En las tres familias, las plantas del cultivar 'Enana Roja Cubana', empleado como progenitor femenino en los cruzamientos, mostraron altos valores genéticos en la mayoría de los caracteres del fruto evaluados. Este resultado es de gran importancia porque cuando se desea elegir individuos como progenitores con fines de mejora, es importante estimar su valor genético pues a mayor valor genético aditivo de un progenitor, mayor será su contribución al valor genotípico de la progenie (Cornide 2018).

Entre los 109 genotipos seleccionados con altos valores genéticos aditivos, se encontraban 54 considerados por Pérez (2013) de comportamiento estable en los caracteres vegetativos y del fruto durante los tres años de evaluación. La selección de los genotipos estables se realizó a partir de un análisis de la interacción genotipo-ambiente con el empleo del Modelo de Efectos Principales Aditivos e Interacción Multiplicativa (*AMMI*, por sus siglas en inglés). Por esta razón, los 54 genotipos pueden ser considerados estables y promisorios por presentar altos valores genéticos aditivos.

Con el objetivo de seleccionar genotipos con altos valores genéticos predichos, de comportamiento estable en los caracteres y que a su vez muestren los caracteres de interés para el mejoramiento del cultivo, los resultados

del presente trabajo fueron comparados también, con una preselección realizada en campo por Rodríguez & *al.* (2009), de genotipos de las progenies evaluadas. Estos investigadores seleccionaron 40 genotipos que presentaban varios atributos de interés económico como porte bajo, altos valores de rendimiento, contenido de vitamina C, sólidos solubles totales, masa del fruto y espesor externo de la pulpa, bajos valores de número, masa total y masa promedio de semillas por fruto, así como, caracteres cualitativos relacionados con la forma del fruto, el color de la pulpa, entre otros. Al combinar los resultados obtenidos en los tres análisis, se puede proponer un grupo de ocho genotipos entre las tres familias, que fueron estables en los caracteres evaluados durante los tres años, mostraron valores deseables en los caracteres de interés para el mejoramiento del cultivo, y en los cuales, se pudo predecir altos valores genéticos individuales para tales caracteres con la estimación de los predictores lineales insesgados. Los genotipos seleccionados se muestran en la Tabla III, donde se señalan los caracteres en los que mostraron los mayores valores genéticos aditivos individuales.

Los ocho genotipos, que presentaron valores genéticos aditivos superiores en varios de los caracteres evaluados, y fueron considerados de comportamiento estable durante los años de evaluación en esos caracteres, podrán ser incorporados al banco de germoplasma para incrementar su variabilidad. Además, podrán ser utilizados como progenitores en programas de cruzamientos del cultivo y replicarse en otras localidades.

DISCUSIÓN

Existen varias opciones para trabajar con datos que no cumplen con las premisas de los métodos paramétricos, pues pocas variables biológicas muestran un ajuste exacto a la normalidad (Whitlock & Schluter 2009). Primero se debe estar seguro que la no normalidad es debida a la naturaleza de los datos, y no a datos perdidos, no declarados o a errores en la entrada de los mismos que se pueden corregir. Por esta razón, una de

TABLA III

Genotipos seleccionados en las tres familias de guayabo con altos valores genéticos en caracteres vegetativos y del fruto

LH: Largo de la hoja. AH: Ancho de la hoja. AP: Altura de la planta. MF: Masa del fruto. LF: Largo del fruto. AF: Ancho del fruto. EEP: Espesor externo de la pulpa. EIP: Espesor interno de la pulpa. NS: Número de semillas. MTS: Masa total de las semillas.

TABLE III

Genotypes selected in three guava families with high breeding values in plant and fruit traits

LH: Leaf length. AH: Leaf width. AP: Plant height. MF: Mass of the fruit. LF: Fruit length. AF: Fruit width. EEP: External flesh thickness. EIP: Internal flesh thickness. NS: Number of seeds per fruit. MTS: Total mass of seeds per fruit.

Familia	Genotipos	LH	AH	AP	MF	LF	AF	EEP	EIP	NS	MTS
'EEA 18-40' x 'N6'	H8	X	X		X	X		X			
'EEA 18-40' x 'N6'	H22	X	X		X	X		X			X
'EEA 18-40' x 'N6'	H44				X	X					X
'EEA 18-40' x 'N6'	H100			X	X		X	X			
'EEA 18-40' x 'Suprema Roja'	H115				X					X	X
'EEA 18-40' x 'Suprema Roja'	H139			X	X					X	
'EEA 18-40' x 'Belic L-207'	H233			X				X		X	X
'EEA 18-40' x 'Belic L-207'	H309					X			X	X	X

las formas de analizar correctamente los datos obtenidos en una investigación, es identificar la distribución apropiada de estos, al realizar un estudio de la variable antes de hacer los análisis (Bandera 2018). En el presente trabajo, se emplearon los procedimientos *UNIVARIATE* y *SEVERITY* para verificar el ajuste de los datos a la distribución normal, detectar la presencia de valores atípicos y determinar la distribución de ajuste de los datos en cada carácter, en las tres familias de guayabo.

Tabachnick & Fidell (2007) plantearon que, aunque las transformaciones de los datos son recomendadas como una alternativa para lograr el cumplimiento de las premisas de normalidad, linealidad y homocedasticidad, no son recomendadas universalmente, debido a que un análisis es interpretado por las variables que contiene, y cuando se transforman son a veces difíciles de interpretar.

Los datos no normales pueden ser producto de una mezcla de múltiples distribuciones o procesos, o que se ajusten a otro tipo de distribución no normal como: log-Normal, Weibull, Exponencial, Gamma, entre otras; según la naturaleza de la variable discreta o continua (McDonald 2014). La falta de normalidad en los datos conlleva muchas veces a la presencia de heterogeneidad de varianzas, otra de las premisas que se debe cumplir para poder desarrollar el Análisis de Varianza.

Para seleccionar la distribución a la cual se ajusta cada carácter, el procedimiento *SEVERITY* tiene en cuenta los valores de los estadísticos de ajuste de Akaike (*AIC*, *AICC*) y Bayesiano (*BIC*) y se selecciona la distribución con el menor valor de tales estadísticos (SAS/STAT 2017). Los valores de estos criterios de información son

relativos e indican cuáles opciones proveen el mejor ajuste; no cuán bien el modelo ajusta los datos en términos absolutos (Davis 2017). El criterio Bayesiano tiende a seleccionar modelos menos complejos que el Akaike, y al elegir los modelos más simples se incrementa la probabilidad de cometer error tipo I. Es por ello, que se prefiere el criterio de Akaike porque el control del error tipo I tiene mayor prioridad (Guerin & Stroup 2000). El criterio de Akaike corregido reduce el sesgo producido por el *AIC* en muestras pequeñas, y en muestras grandes ambos convergen.

Varios estudios se han realizado para determinar los mejores estadísticos de ajuste en la selección de las estructuras de covarianza (Al-Marshadi 2014). Ferron & al. (2002) encontraron que se selecciona la estructura de covarianza correcta con el criterio de Akaike, en el 79% de las veces como promedio; mientras que con el criterio Bayesiano solo se logra en el 66% de las veces. En contraste, Keselman & al. (1998) encontraron que los criterios *AIC* y *BIC* fueron exitosos en la selección de la estructura de covarianza en el 47 y 35% de las veces, respectivamente.

En la presente investigación, el criterio que se utilizó para seleccionar la mejor estructura de covarianza en todos los caracteres en las tres familias, fue el Akaike (*AIC*) y los menores valores de los estadísticos de ajuste se obtuvieron con las estructuras Toeplitz, No estructurada y Componente de Varianza. Se observaron valores muy similares de los estadísticos de ajuste *AIC*, *AICC* y *BIC*. Las pequeñas diferencias entre los valores de los criterios de información, indican que las diferentes estructuras empleadas modelan los datos de manera similar (Davis 2017). Al obtenerse los mismos valores de los estadísticos

cos en las estructuras Toeplitz, Componente de Varianza y No estructurada, se puede seleccionar cualquiera de las tres para los análisis de medidas repetidas.

La estructura Toeplitz asume que dos observaciones cualesquiera de un individuo dado, con la misma diferencia de tiempo, están igualmente correlacionadas, y que la correlación no disminuye exponencialmente con el tiempo (Littell & *al.* 2006). La estructura de Componente de Varianza es más simple, asume que la correlación de los errores dentro de un sujeto es cero y no se recomienda en experimentos con medidas repetidas. Mientras que el modelo de covarianza No estructurado es el más complejo, en el cual los errores dentro de sujetos para cada par de unidades de tiempo tienen su propia correlación única. Es una estructura muy flexible que puede teóricamente ajustar cualquier grupo de datos, porque estima un parámetro separado para cada elemento de la matriz de covarianza. Una estructura que no asume patrones matemáticos en la matriz de covarianza puede ser atractiva, pero si estima muchos parámetros, como la No estructurada, puede conducir a problemas de convergencia (Davis 2017).

Se debe seleccionar la estructura de covarianza con el menor número de parámetros, que mejor ajuste los datos. La estructura más parsimoniosa es la de mejor ajuste. Se debe lograr un equilibrio entre el mejor ajuste y el número de parámetros del modelo que debe ser pequeño, pero adecuado. Si se hace una mala selección de la covarianza del modelo, se puede comprometer severamente la inferencia de los efectos fijos de interés porque se obtienen estimados sesgados (Bello 2016). Si se ignora la correlación al usar un modelo que es demasiado simple, como Componente de Varianza o Simetría Compuesta, se puede incrementar la probabilidad de cometer error tipo I y subestimar los errores estándar. Si el modelo es muy complejo, como el No estructurado, se puede sacrificar la potencia y la eficiencia de la prueba estadística. Los análisis con medidas repetidas son tan robustos como el modelo de covarianza empleado sea aproximadamente correcto (Littell & *al.* 2006). Por todas estas razones, fue seleccionada la estructura de covarianza Toeplitz como la de mejor ajuste a los datos en cada uno de los caracteres evaluados en las tres familias de hermanos completos de guayabo.

Después de seleccionar la mejor estructura de covarianza para cada uno de los caracteres vegetativos y del fruto, en las tres familias, se realizaron los procedimientos *GLIMMIX* que consideran un modelo lineal generalizado mixto. Estos modelos extienden los modelos lineales clásicos, para incluir efectos aleatorios, datos correlacionados, desbalanceados o incompletos y con distribuciones de la familia exponencial. Además, permiten que la media de una población dependa de un

predictor lineal a través de una función de enlace de tipo no lineal (Stroup 2015). Con el empleo de los modelos generalizados se pudieron obtener los mejores predictores lineales insesgados que ayudan al mejorador a predecir los valores genéticos aditivos de los genotipos en estudio.

El éxito en la selección de progenies superiores está intrínsecamente ligado a una experimentación adecuada, pues el análisis estadístico resultará en buenas predicciones del valor genético aditivo por medio de estos predictores insesgados. Si dentro de las familias se observan genotipos con los mayores valores genéticos aditivos predichos para varios caracteres, se pueden seleccionar como individuos que concentran alelos favorables para esos caracteres (Pimentel & *al.* 2014). Los valores genéticos aditivos superiores reflejan los progresos para la selección. Además, los valores de los predictores lineales insesgados, son útiles para seleccionar los mejores genotipos que pueden ser utilizados como progenitores en programas de cruzamientos (Noerwijati & *al.* 2013).

La predicción de los valores genéticos aditivos, con el empleo de la metodología *REML-BLUP*, ha mostrado resultados satisfactorios en los programas de mejoramiento en frutales como: papayo (*Carica papaya* L.) (Oliveira & *al.* 2012), guayabo (*Psidium guajava* L.) (Quintal & *al.* 2016), guanábano (*Annona muricata* L.) (Sánchez & *al.* 2017) y camu camu (*Myrciaria dubia* Kunth) McVaugh) (Pinedo & *al.* 2017).

Es importante destacar las principales ventajas que tiene el uso de la metodología de los modelos mixtos (*REML-BLUP*) en la estimación de parámetros genéticos y la predicción de valores genéticos para el mejoramiento de especies perennes. Los *BLUP* maximizan la exactitud de la selección y permiten utilizar de manera simultánea un número grande de datos provenientes de varias fuentes, por ejemplo, varios experimentos evaluados en una o más localidades y durante una o más cosechas. Los *BLUP* individuales utilizan todos los efectos del modelo estadístico, manejan el desbalance, consideran la cercanía genética entre las plantas evaluadas y la coincidencia entre las unidades de selección y recombinación. También, permiten comparar individuos o variedades en el tiempo (generaciones, años) y en el espacio (localidades, bloques); de manera simultánea corrigen los datos de los efectos ambientales, estiman componentes de varianza y predicen valores genéticos individuales de manera precisa e insesgada. Además, se pueden utilizar en datos desbalanceados, no normales y con estructuras complejas (medidas repetidas, diferentes años, localidades y diseños) (de Resende 2016).

Se recomienda que los ocho genotipos seleccionados por presentar valores genéticos aditivos superiores en varios de los caracteres evaluados y haber sido considerados de comportamiento estable durante los años de evaluación en esos caracteres, se genotipifiquen con el empleo de los marcadores moleculares, que ya fueron secuenciados para el cultivo y los resultados fueron previamente publicados por Rodríguez & al. (2007) y Valdés-Infante & al. (2011). Los resultados que se obtengan de la genotipificación de los individuos seleccionados, constituyen una información de gran importancia para complementar los avances que se han alcanzado en el programa de mejoramiento genético del cultivo en el país.

CONCLUSIONES

Con el empleo de diferentes criterios de ajuste, se puede seleccionar la estructura de covarianza Topelitz como la que mejor ajusta los datos experimentales en los caracteres evaluados. Los genotipos seleccionados, por presentar los mayores valores genéticos aditivos en varios de los caracteres evaluados, pueden ser considerados genotipos promisorios, y deben incluirse en el banco de germoplasma del cultivo. Además, deben ser utilizados como progenitores en programas de cruzamientos.

REFERENCIAS BIBLIOGRÁFICAS

- Ali, D.O., Ahmed, R.A. & Babikir, E.B. 2014. Physicochemical and nutritional value of red and white guava cultivars grown in Sudan. *J. Agric. Fd. Applied Sci.* 2(2): 27-30.
- Al-Marshadi, A.H. 2014. Selecting the covariance structure in mixed model using statistical methods calibration. *J. Math. Stat.* 10(3): 309-315.
- Bandera, E. & Pérez, L. 2018. Los modelos lineales generalizados mixtos. Su aplicación en el mejoramiento de plantas. *Cult. Trop.* 39(1): 127-133.
- Bandera, E. 2018. Estudio de la variabilidad genética y de la asociación entre caracteres cuantitativos en familias de hermanos completos de guayabo (*Psidium guajava* L.). Tesis de Maestría, Facultad de Biología, Universidad de La Habana, Cuba.
- Bello, N.M. 2016. Mixed Models for Analysis of Repeated Measures Data. Kansas State University. International Biometric Society. <http://exa.unne.edu.ar/GAB2016/GAB2016/docs/Repeated%20measures%20GAB2016%20-%20NBello.pdf>. 4 de Marzo de 2019.
- Cargnin, A. 2016. Repetibilidade e número de colheita de características para seleção de clones de variedades viníferas. *Cienc. Rural* 46(2): 221-226.
- Cornide, M.T. 2018. Bases genéticas y principales métodos del fitomejoramiento. En: Cornide, M.T. (ed.). Genética, genómica y fitomejoramiento. Editorial UH, La Habana, Cuba.
- Davis, J.W. 2017. Linear Mixed Models with Repeated Effects. Introduction and Examples Using SAS/STAT® Software. University of Georgia, Griffin Campus, USA. <https://site.caes.uga.edu/expstat-grif/files/2018/07/RepeatedMixedFinal1.pdf>. 5 de Marzo de 2019.
- Davis, J.W. 2018. Introduction to Generalized Linear Mixed Models. A Count Example. University of Georgia, Griffin Campus, USA. <https://site.caes.uga.edu/expstat-grif/files/2018/07/Counts-GLMMfin1.pdf>. 5 de Marzo de 2019.
- de Resende, M.D.V. 2016. Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breed. Appl. Biot.* 16: 330-339.
- Ferron, J., Dailey, R. & Yi, Q. 2002. Effects of misspecifying the first-level error structure in two level models of change. *Multivariate Behav. Res.* 37: 379-403.
- Guerin, L. & Stroup, W.W. 2000. A simulation study to evaluate PROC MIXED analysis of repeated measures data. Proceedings of the 12th Annual Conference on Applied Statistics in Agriculture. Manhattan, KS, Kansas State University, USA.
- Hernández, A., Pérez, J.M., Boch, D. & Castro, N. 2015. Clasificación de los suelos de Cuba 2015. Ediciones INCA, San José de las Lajas, Mayabeque, Cuba.
- Keselman, H.J., Algina, J., Kowalchuk, R.K. & Wolfinger, R.D. 1998. A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Commun. Stat. Simulat. Comput.* 27: 591-604.
- Littell, R.C., Milliken, G.A., Stroup, W.W. & Wolfinger, R.D. 2006. SAS for Mixed Models. 2nd Edition, SAS Institute Inc., Cary, NC, USA.
- McDonald, J.H. 2014. Handbook of Biological Statistics. Sparky House Publishing, Baltimore, Maryland, USA.
- MINAG. 2011. Instructivo Técnico para el cultivo de la Guayaba. Instituto de Investigaciones en Fruticultura Tropical, La Habana, Cuba.
- Noerwijati, K., Taryono, N. & Prajitno D. 2013. Breeding value estimation of fifteen related cassava genotypes using BLUP. *Journal of Agricultural and Biological Science* 8(4): 317-321.
- Oliveira, E.J., Andrade, F.F.G., de Freitas, X.J., de Loyola, D.J. & de Resende, M.D.V. 2012. Plant selection in F2 segregating populations of papaya from commercial Hybrids. *Crop Breed. Appl. Biot.* 12: 191-198.
- Patel, R.K., Maiti, C.S., Deka, B.C., Vermav, V.K., Deshmukh, N.A. & Verma, M.R. 2015. Genetic variability, character association and path coefficient study in guava (*Psidium guajava* L.) for plant growth, floral and yield attributes. *International Journal of Bio-resource and Stress Management* 6(4): 457-466.
- Pérez L. 2013. Análisis de la variabilidad genética y de la interacción genotipo – ambiente en poblaciones de guayabo (*Psidium guajava* L.). Tesis de Doctorado, Facultad de Biología, Universidad de La Habana, Cuba.
- Pimentel, A.J.B., Guimarães, J.F.R., de Souza, M.A., de Resende, M.D.V., Moura, L.M., Rocha, J.R.S. & Ribeiro, G. 2014. Estimación de parámetros genéticos e predição de valor genético aditivo de trigo utilizando modelos mistos. *Pesq. Agropec. Bras.* 49(11): 882-890.
- Pinedo, M.P., Chagas, E.A., Davila, E.P., Rodriguez, C.A., Lozano, R.B., Chagas, P.C. & Melo, V.F. 2017. Selection of superior genotypes in 37 clones of Camu-Camu by repetitivity analysis. *J. Agr. Sci.* 9(6): 175-187.
- Quintal, S.S.R., Viana, A.P., Campos, B.M., Vivas, M. & do Amaral, A.T. 2016. Selection via mixed models in segregating guava families based on yield and quality traits. *Rev. Bras. Frutic.* 29(2): p.e-866.

Quintal, S.S.R., Viana, A.P., Campos, B.M., Vivas, M. & do Amaral, A.T. 2017. Analysis of structures of covariance and repeatability in guava segregating population. *Rev. Caatinga* 30(4): 885-891.

Rawls, B., Harris-Shultz, K., Dhekney, S., Forrester, I. & Sitther, V. 2015. Clonal fidelity of micropropagated *Psidium guajava* L. plants using microsatellite markers. *Am. J. Plant Sci.* 6: 2385-2392.

Rodríguez, N.N., Valdés-Infante, J., Becker, D., Velázquez, B., Coto, O., Ritter E. & Rohde, W. 2004. Morphological, agronomic and molecular characterization of Cuban accessions of guava (*Psidium guajava* L.). *J. Genet. Breed.* 58(1): 79-89.

Rodríguez, N.N., Valdés-Infante, J., Becker, D., Velázquez, B., González, G., Sourd, D., Rodríguez, J., Billotte, N., Risterucci, A.M., Ritter, E. & Rohde, W. 2007. Characterization of guava accessions by SSR markers, extension of the molecular linkage map, and mapping of QTLs for vegetative and reproductive characters. *Acta Hortic.* 735: 210-215.

Rodríguez, N.N., Valdés-Infante, J., Rodríguez, J.A., Velázquez, J.B., Rivero, D. & Martínez, F. 2009. Preselección de híbridos de guayabo (*Psidium guajava* L.) con potencial productivo y calidad de la fruta. *CitriFrut* 26(2): 13-20.

Rodríguez, N.N., Fermín, E., Valdés-Infante, J., Velázquez, B., Rivero, D., Martínez, F., Rodríguez, J. & Rohde, W. 2010. Illustrated descriptor for guava (*Psidium guajava*). *Acta Hortic.* 849: 103-109.

Sánchez, C.F., Alves, R.S., García, A.D., Teodoro, P.E., Peixoto, L.A., Silva, L.A., Bhering, L.L. & de Resende, M.D.V. 2017. Estimates of repeatability coefficients and the number of the optimum measure to select superior genotypes in *Annona muricata* L. *Genet. Mol. Res.* 16(3): gmr16039753.

SAS Institute Inc. 2012. SAS version 9.3. SAS Institute Inc., Cary, NC, USA. Available from: <http://www.sas.com>

SAS/STAT. 2017. SAS/STAT Software version 14.1. SAS Institute Inc., Cary, NC, USA. Available from: <http://www.sas.com>

Stroup, W.W. 2015. Rethinking the Analysis of Non-Normal Data in Plant and Soil Science. *Agron. J.* 107: 811-827.

Tabachnick, B.G & Fidell, L.S. 2007. Using multivariate statistics. Allyn and Bacon, Boston, USA.

Valdés-Infante, J., Rodríguez, N.N., Ritter, E. & Rohde, W. 2011. Mapa de ligamiento genético y análisis de QTLs en guayabo (*Psidium guajava* L.). Población de mapeo de 'Enana Roja Cubana' x 'Belic L-207'. *CitriFrut* 28(2): 56-66.

Whitlock, M.C. & Schluter, D. 2009. The Analysis of Biological Data. Roberts and Company Publishers, Greenwood Village, Colorado, USA.