

Original Article

## **Methodology of validation and management of crosses in the genetic improvement in sugarcane**

Reynaldo Rodríguez-Gross<sup>1\*</sup>

Yaquelin Puchades-Isaguirre<sup>1</sup>

Wilfre Aiche-Maceo<sup>1</sup>

<sup>1</sup>Grupo de Genética y Protección de Plantas. Estación Territorial de Investigaciones de la Caña de Azúcar Oriente Sur. Santiago de Cuba, Cuba

\*Author for correspondence. [reynaldo.rodriguez@inicas.azcuba.cu](mailto:reynaldo.rodriguez@inicas.azcuba.cu)

### **ABSTRACT**

The Sugarcane Research Institute in Cuba develops a selection program to respond to obtaining new cultivars. However, there is no procedure available to determine and classify the family response from the progeny selection data. The objective of the work was to establish a methodology of classification and management of biparental crosses in the program of genetic selection of sugarcane in Cuba, based on the selection information and the estimation of its genetic value. Data from the stage of Clonal Lot 1 of the selection program of the southeast region were used. The period included the selection of the years from 2000 to 2015. A linear model was used to quantify the interaction between the crossing and the environment and quantify the estimated genetic value. Seven variables related to the sugar content of the progenies were used, as well as agricultural yield, aroya resistance and coal from sugarcane. Seven categories were established for the classification of the crossings: crossing much discarded, discarded, moderately discarded, exploratory, moderately verified, verified and very tested. The proposed method allows to make more efficient the programs of crossing and genetic selection of sugarcane.

**Key words:** selection, simulation, progeny

## INTRODUCTION

Selection as a concept is the process of discrimination of the best clones of a population and is in turn the highest cost of breeding programs. This is a multi-stage process that aims to identify and extract superior individuals as efficiently as possible, maximizing genetic gains and at a lower cost <sup>(1)</sup>.

The choice of parents and the prediction of their value is one of the important actions in any crossing program <sup>(2)</sup>. Therefore, the increase in selection and genetic gain is a measure of the success of breeding programs <sup>(3)</sup>.

Some of the genetic improvement programs have prioritized family selection followed by individual selection <sup>(4)</sup>. A simultaneous selection methodology is necessary for the first stages, which constitutes the basis for the determination of an index from multiple characters rather than intensive selection on a particular character <sup>(5)</sup>. This objective requires new strategies to optimize profits in the programs of genetic improvement of sugarcane <sup>(6)</sup>. In this sense, one of the important research directions is the optimization of family behavior prediction methods, using generalized mixed linear models and their Bayesian counterparts <sup>(7)</sup>.

The Sugarcane Research Institute of Cuba (INICA) develops a genetic improvement program to obtain new, more productive cultivars, resistant to the main pests and adapted to the different commercial exploitation conditions <sup>(8)</sup>. Every year it works with large populations of progenies at different stages.

For the management of the information that is generated in the selection process, a computer program is available to capture, store and process this information <sup>(9)</sup>. This computer application has a mathematical model that allows estimating the genetic value (VGE) of parents and crosses. However, there is no methodology available to predict the family response and its management in the sugarcane selection program in Cuba.

The objective of this work was to establish a methodology for classification and management of biparental crosses in the program of genetic selection of sugarcane in Cuba, based on the determination of the VGE through the evaluation of the progeny in the clonal lot 1.

## MATERIALS AND METHODS

Selection data from clonal lot 1 of the selection program were used to obtain new sugarcane cultivars that are developed in the southeast region of Cuba <sup>(10)</sup>. The study period included 16 years from 2000 to 2015.

The trials were established in areas of the experimental block of the América Libre locality in the Contramaestre municipality, Santiago de Cuba province (-76.2° longitude and 20.3° latitude) on a Sialitic Brown soil <sup>(11)</sup>. The clones were planted at a distance of one meter between clones and 1.6 m between rows without repetition. The progeny were evaluated in the first offspring strain in January at 12 months of age. C87-51 commercial cultivar with high sucrose content was used as a control or control.

The database comprised 16 years of clonal selection and all families with more than 40 individuals evaluated were considered in the study. The information of 24092 clones of 230 biparental combinations was collected. All data were captured and validated in the SASSEL software <sup>(12)</sup>.

On the other hand, information on the stages of replicated and extensive studies of the sugarcane selection process in Cuba was used to determine biparental crossings that individuals contributed to the final stages of this process.

### **Mathematical model used to estimate the genetic value of crosses**

A linear model was used to quantify the interaction between the crossing and the environment <sup>(9)</sup>. Seven evaluated variables related to the sucrose content of the progenies, components of agricultural yield (diameter and length of the stems) and resistance to two of the main pests that affect the crop were used (Table 1).

The weighting coefficients were adjusted according to the relative importance of the variables and their heritability. The maximum weighting value was given to the selection percentage, as this is a variable that encompasses the desirable characteristics in the progeny of the crosses.

**Table 1.** Evaluated variables considered in the relation functions of the mathematical model to determine the estimated genetic value (VGE)

No	Variable	Coefficient of weighing
1	Selection Percentage	1
2	Refractometric °Brix	0.8
3	Diameter of stem (cm)	0.8
4	Length of stem (cm)	0.8
5	Percentage of individual affected by rust	0.7
6	Percentage of individuals affected by coal	0.7
7	Percentage of individual affected by low °Brix	0.8

### Statistical processing

To counteract the effect of the genotype-environment interaction (IGxA) over the years, in this case crosses x series, the VGE was standardized with its independent determination for each year or series studied.

The population of families comprised 230 biparental crosses. From these, only 204 crosses were evaluated, which were used for the elaboration of the methodology of classification and management of the crosses. The rest of the crosses (26) evaluated two or more times during the study period were used to validate the proposed methodology.

The original data of the VGE were evaluated with respect to their normality by means of the Shapiro-Wilk's W test and they fulfilled the requirements regarding these parameters, so that in no case was their transformation necessary. A frequency histogram was performed with the 204 crosses that were evaluated once. The number of classes were formed according to the normal VGE distribution and the homogeneity in the standard deviation of the groups formed according to the class intervals.

With the VGE values of the 26 crosses, evaluated two or more times, classification was made according to the groups determined by the histogram of the crosses evaluated once. With this categorization the difference between the groups or repeatability between them was determined and in this way the validity of the procedure was verified.

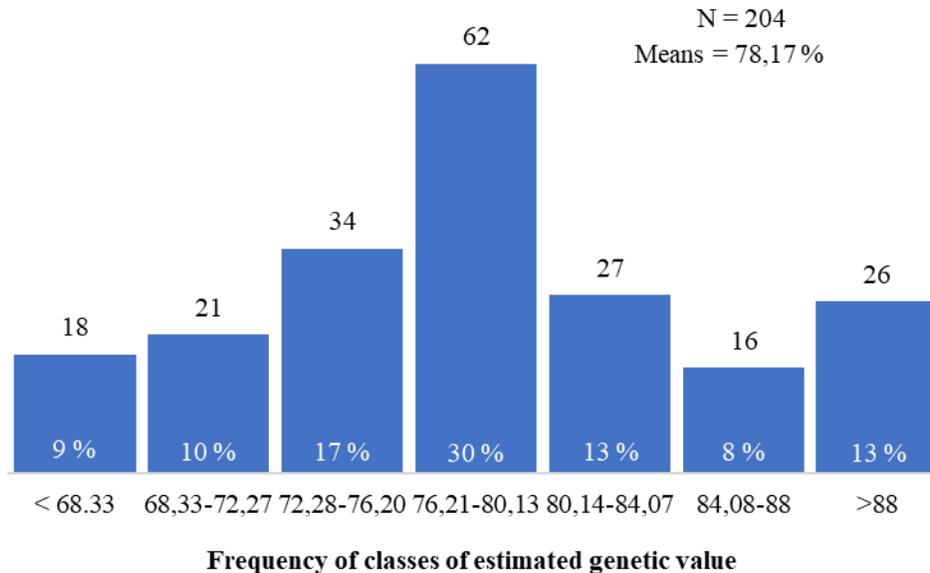
A new verification of the classification of the crosses with the VGE obtained from the families that contributed individuals to the final stages of the selection process (replicated and extensive studies) was carried out. The absolute and relative frequency of each category

was determined as a criterion for the validation of the groups, based on the fact that they were more represented as tested crossings.

## RESULTADS AND DISCUSSION

The frequency histogram of the VGE of 204 crosses allowed to obtain seven class intervals (Figure 1). The frequency histogram is adjusted to a normal curve, where the average values are grouped to the center of the curve. Class intervals are due to the fact that there is a homogeneity of variance between the classes and the average in the middle class. The average of the VGE was 78.17 % and the middle class ranges from 76.20 to 80.13 %.

El histograma de frecuencia del VGE de 204 cruces permitió obtener siete intervalos de clases (Figura 1). El histograma de frecuencia se ajusta a una curva normal, donde los valores medios se agrupan al centro de la curva. Los intervalos de clases obedecen a que exista entre clases homogeneidad de varianza y la media en la clase del medio. La media del VGE fue de 78.17 % y la clase media oscila desde 76.20 a 80.13 %.



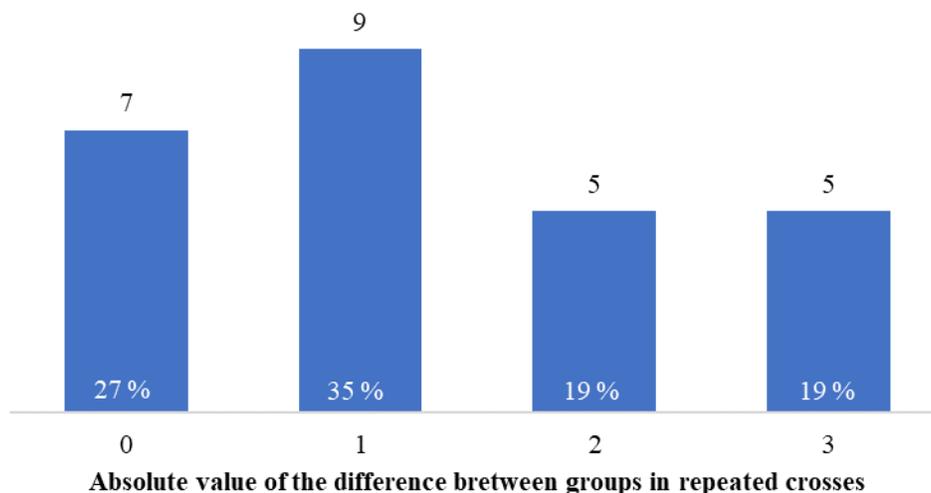
**Figure 1.** Absolute and relative frequency histogram of estimated genetic value (%) for seven groups of classes intervals

The class intervals in Figure 1 allow a classification of the family value to be established in correspondence with the results of the simultaneous selection made on several characters (Table 2). Crossings of the middle class were classified as exploratory or with no tendency to define a prominent value or not as a family. From this group the rest of the classification was defined when considering the lower than average classes with a tendency to be discarded crosses and higher classes to it as verified. That is, categories were established from moderately tested to highly tested (VGE greater than 80.13 %) and from moderately discarded to very discard with values below 76.20 %.

**Table 2.** Classification of the crossover according to its estimated genetic value (VGE) and frequency histogram classes

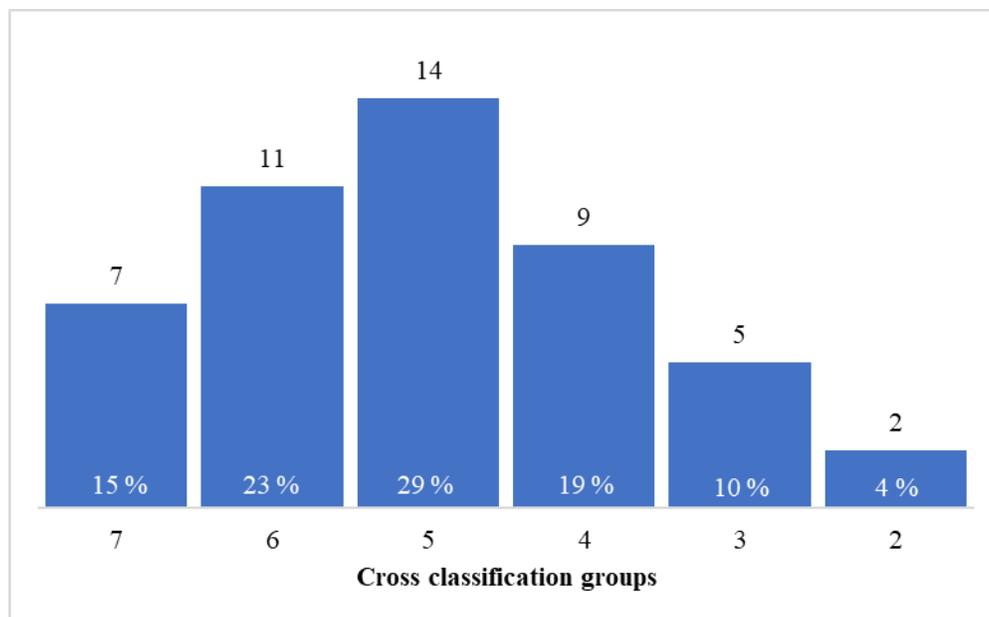
VGE class intervals (%)	Group	Classification crossing
< 68.33	1	Very discarded
68.33-72.27	2	Discarded
72.28-76.20	3	Moderately discarded
76.21-80.13	4	Exploratory
80.14-84.07	5	Moderately tested
84.08-88	6	Tested
>88	7	Very tested

The frequency histogram, of the 26 crosses evaluated two or more times in the clonal lot stage 1, demonstrated the effectiveness of the classification from the VGE (Figure 2). In this case, 27 % of the crossings showed similar VGE and therefore classified in the same group (zero difference). 35 % showed differences of a group, which together with the previous one amounts to 62 %. This value means the probability of classifying a family and matching its category or varying in a group. Only 38 % changed the classification in two and three groups which shows the effectiveness of the proposed method.



**Figure 2.** Difference of the classification between groups for the crosses evaluated two or more times in the clonal lot 1

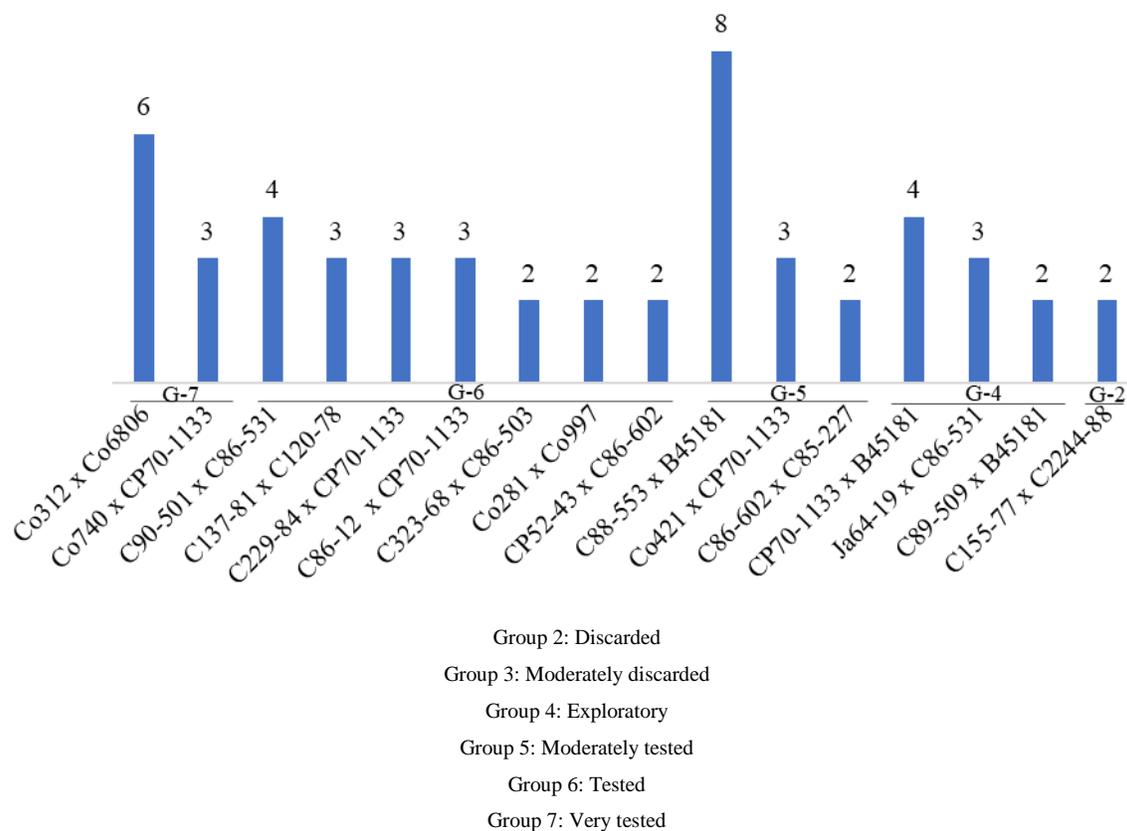
Regarding the crossings evaluated that provides individuals to replicated studies, 67 % classified in groups 5, 6 and 7 ranging from modernly tested to highly tested (Figure 3). Of group 1 (very discarded) no crossings were quantified and of groups 2 and 3 (moderately discarded and discarded) only 7 crossings were found (14 %). Group 4 (exploratory crosses) accounted for 19 %, which together with groups 5, 6 and 7 accounted for 86 %, which evidences a good classification to establish a tested family value system or to discard for the selection program genetics in sugarcane in Cuba.



Group 2: Discarded  
 Group 3: Moderately discarded  
 Group 4: Exploratory  
 Group 5: Moderately tested  
 Group 6: Tested  
 Group 7: Very tested

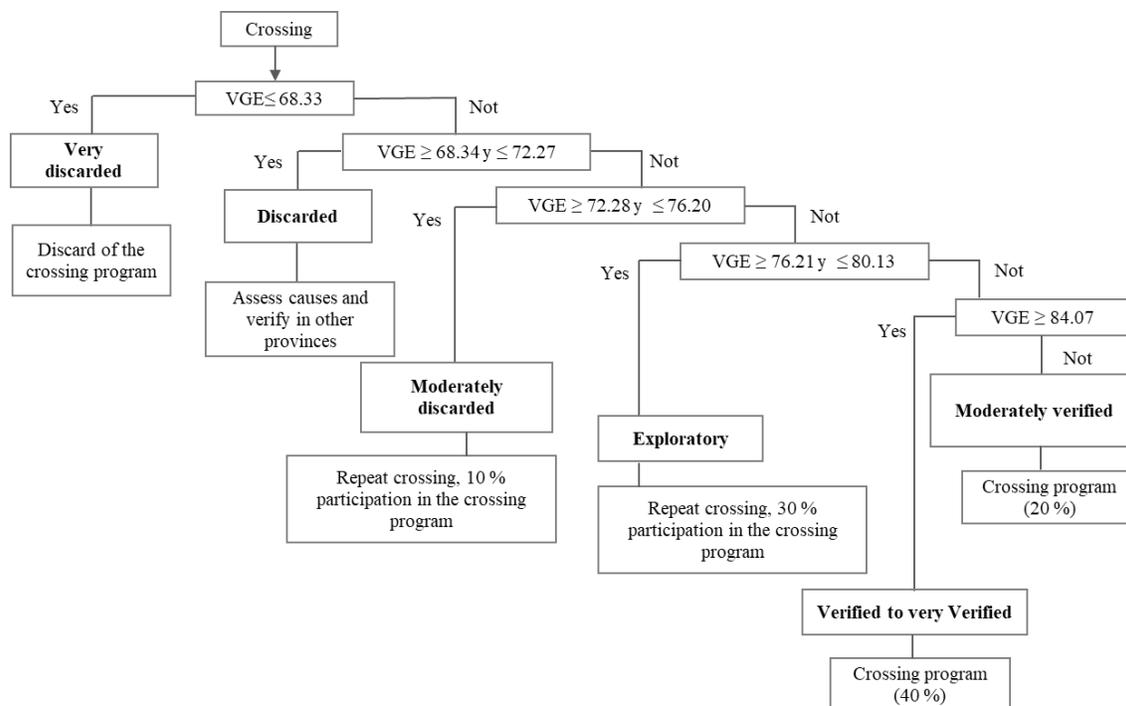
**Figure 3.** Frequency histogram of the classification of crosses by groups that contributed individuals to replicated studies according to the estimated genetic value (VGE)

Among the crossings with individuals in replicated studies, Co312 x Co6806 and Co740 x CP70-1133 are classified as highly verified (group 7) with contribution of six and three cultivars respectively (Figure 4). Other families, highlighted and classified in group 6 (tested) were: C90-501 x C86-531, C137-81 x C120-78, C229-84 x CP70-1133 and C86-12 x CP70-1133. The latter contributed the cultivation C00-575 recommended for commercial exploitation by the improvement program of the southeast region, which confirms the reliability of the procedure used. The ClonT96-40 x CSG87-508 crossover, classified in group 5 (moderately tested), produced the recommended C04-553 cultivar to extension.



**Figure 4.** Frequency of individuals by crossings in replicated studies and classification according to their estimated genetic value

These results show the effectiveness of the classification made at the crossings from the VGE. In this way, a methodology or algorithm is proposed to determine the strategy to follow in the hybridization and selection campaigns in the program of genetic improvement of sugarcane in Cuba (Figure 5).



**Figure 5.** Algorithm of classification and management of biparental crosses according to their estimated genetic value (VGE) in the program of genetic improvement of sugarcane in Cuba

When determining the VGE of a crossing if it is less than 68.33 %, it is classified as very discarded and removed from the crossings program. From this value, the rest of the categories, their management and proportion in the crossing program are determined. This methodology allows giving greater weight to the prediction of family value and its conduction in the process of improvement, as well as reorienting the individual selection in those families with lower VGE.

In the literature, it is common to find work related to different methods of selection, prediction of genetic value or advances in genetic improvement programs. The use of BLUP (Best Non-Biased Linear Predictor) models is the most common procedure. In this sense, the work carried out in Brazil stands out <sup>(13)</sup>, where they suggest the use of BLUP methods in family selection followed by the individual for the identification of the best genotypes. However, BLUP (s) methods require a great deal of information from individuals and family, as well as their field evaluation under an experimental design <sup>(6)</sup>.

In a study of parent selection in Brazil, through the use of mixed models, it was reported that the parent CP70-1133 presented a high additive value and selection index with respect to the

°Brix variable, which suggests its use in the programs of crossing <sup>(14)</sup>. The VGE data obtained for this cultivar confirm this result.

Another method used in Florida was the use of an index on stalks vigor and the °Brix for the selection of sugarcane progeny <sup>(15)</sup>. These researchers suggest the use of individual selection over family selection to find clones with desired characteristics.

On the other hand, the use of a decision tree as a tool in the selection of families in sugarcane has been reported <sup>(6)</sup>. These researchers propose this method to overcome the difficulties of the BLUP methods in relation to the volume of information that needs to be evaluated in the field. Thus, they recommend the use of the regression and classification tree for the selection of the best families with an accuracy of 74 %. The use of this procedure, too, has been reported previously <sup>(16,17)</sup>.

Other researchers recommend the use of selection indices to determine the best parents and thus obtain progenies with higher yields <sup>(14)</sup>. However, the use of one method or another or its combination are valuable tools to evaluate parents and progenies and increase efficiency in genetic improvement programs.

## CONCLUSIONS

- The determination of the estimated genetic value of the crosses allowed them to be classified into seven categories that include the highly discarded, discarded, moderately discarded, exploratory, moderately proven, proven and highly verified crosses.
- The classification methodology, based on the determination of the estimated genetic value of the biparental crosses, allowed the design of the strategy to be followed in the hybridization and selection campaigns in the program of genetic improvement of sugarcane in Cuba.

## BIBLIOGRAPHY

1. Jackson P. What is the optimal selection intensity of seedlings (stage 1) in sugarcane breeding programs? *INTERNATIONAL SUGAR JOURNAL*. 2018;120(1435):542-6.
2. Atkin FC, Dieters MJ, Stringer JK. Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. *Theoretical and Applied Genetics*. 2009;119(3):555-65.
3. Zhou MM, Gwata ET. Location effects and their implications in breeding for sugarcane yield and quality in the midlands region in South Africa. *Crop Science*. 2015;55(6):2628-38.
4. Stringer JK, Cox MC, Atkin FC, Wei X, Hogarth DM. Family selection improves the efficiency and effectiveness of selecting original seedlings and parents. *Sugar Tech*. 2011;13(1):36-41.
5. Mohammad T, Khalil IH, McCord PH, Glaz B, Todd J. Phenotypic selection of sugarcane *Saccharum* spp. genotypes using best linear unbiased predictors. *Journal-American Society of Sugar Cane Technologists*. 2014;34:44-56.
6. Peternelli LA, Bernardes DP, Brasileiro BP, Barbosa MH, Silva RH. Decision Trees as a Tool to Select Sugarcane Families. *American Journal of Plant Sciences*. 2018;9(2):216-30.
7. Cornide M. *Genética, genómica y fitomejoramiento*. La Habana: Universidad la Habana; 2018. 262 p.
8. González R. *Variedades de caña de azúcar cultivadas en Cuba. Cronología, legislación, metodologías y conceptos relacionados*. La Habana, Cuba: ICIDCA; 2019. 216 p.
9. Rodríguez-Gross R, Puchades-Isaguirre Y, Aiche-Maceo W, Cornide-Hernández MT. Modelo matemático para estimar el valor genético de progenitores y cruces en caña de azúcar. *Cultivos Tropicales*. 2018;39(2):81-8.
10. Jorge H, González R, Casas M, Jorge I. *Normas y Procedimientos del Programa de Mejoramiento Genético de la Caña de Azúcar en Cuba*. INICA: PUBLINICA La Habana; 2011.
11. Hernández A, Pérez J, Bosch D, Rivero L. *Nueva versión de clasificación genética de los suelos de Cuba*. AGRINFOR. 1999. 64 p.

12. Rodríguez R, Puchades Y, Abiche W, Rill S, García H. SASEL: software for data management generated in the Cuban sugarcane-breeding program. Proceedings of the International Society of Sugar Cane Technologists. 2016;29:63-6.
13. Brasileiro BP, de Paula Mendes TO, Peternelli LA, da Silveira LCI, de Resende MDV, Barbosa MHP. Simulated individual best linear unbiased prediction versus mass selection in sugarcane families. Crop Science. 2016;56(2):570-5.
14. Neto H, Oliveira R, Bessalho J, Daros E, Camargo J, Weber H, De Medeiros J, Souza G. Parent selection for long-term crosses in sugarcane (*Saccharum* spp.) through the use of mixed models. Proc. Int. Soc. Sugar Cane Technol. 2013. 28 p.
15. Zhao D, Comstock JC, Glaz B, Edme SJ, Glynn NC, Del Blanco IA, et al. Vigor rating and Brix for first clonal selection stage of the Canal Point sugarcane cultivar development program. Journal of Crop Improvement. 2012;26(1):60-75.
16. Zhou MM, Kimbeng CA, Tew TL, Gravois KA, Pontif M, Bischoff KP. Logistic regression models to aid selection in early stages of sugarcane breeding. Sugar Tech. 2014;16(2):150-6.
17. Zhao X, Liu Z, Dan F, Wang K. Plant breeding evaluation with rank entropy-based decision tree. IFAC-PapersOnLine. 2016;49(16):336-40.