Date:	13 Nov 2020
To:	"Santiago Ostengo" santiagostengo@eeaoc.org.ar
cc:	"Angélica Rueda Calderón" angelicarueda@agro.unc.edu.ar, "Cecilia Bruno" cebruno@agro.unc.edu.ar, "María Inés Cuenya" micuenya@eeaoc.org.ar, "Mónica Balzarini" mbalzari@agro.unc.edu.ar
From:	"G.P. Rao" sugartech@rediffmail.com
Subject:	SUTE: Your manuscript entitled A protocol for identifying characteristic sucrose accumulation curves of sugarcane genotypes (Saccharum spp.)

Ref.: Ms. No. SUTE-D-19-00320R6
A protocol for identifying characteristic sucrose accumulation curves of sugarcane genotypes (Saccharum spp.)
Sugar Tech

Dear Mr. Ostengo,

I am pleased to tell you that your work has now been accepted for publication in Sugar Tech.

You will soon receive an e-mail from Springer production team. Please respond to this e-mail in the timely manner to help us process your manuscript to the next stage. Please note that the article will be processed and the proofs be sent to you only after we receive your responses to this e-mail. If you have any queries on the same please contact the production editor.

Thank you for submitting your work to this journal.

With kind regards

G.P. Rao
Editor-in-Chief
Sugar Tech

COMMENTS TO THE AUTHOR:

—

****

Our flexible approach during the COVID-19 pandemic

If you need more time at any stage of the peer-review process, please do let us know. While our systems will continue to remind you of the original timelines, we aim to be as flexible as possible during the current pandemic.

This letter contains confidential information, is for your own use, and should not be forwarded to third parties.

Recipients of this email are registered users within the Editorial Manager database for this journal. We will keep your information on file to use in the process of submitting, evaluating and publishing a manuscript. For more information on how we use your personal details please see our privacy policy at https://www.springernature.com/production-privacy-policy. If you no longer wish to receive messages from this journal or you have questions regarding database management, please contact the Publication Office at the link below.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: https://www.editorialmanager.com/sute/login.asp?a=r). Please contact the publication office if you have any questions.

Sugar Tech

A protocol for identifying characteristic sucrose accumulation curves of sugarcane genotypes (Saccharum spp.)

--Manuscript Draft--

Manuscript Number:	SUTE-D-19-00320R6
Full Title:	A protocol for identifying characteristic sucrose accumulation curves of sugarcane genotypes (Saccharum spp.)
Article Type:	Original paper
Corresponding Author:	Santiago Ostengo Estacion Experimental Agroindustrial Obispo Colombres Las Talitas, Tucumán ARGENTINA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Estacion Experimental Agroindustrial Obispo Colombres
Corresponding Author's Secondary Institution:	
First Author:	Santiago Ostengo
First Author Secondary Information:	
Order of Authors:	Santiago Ostengo
	Angélica Rueda Calderón
	Cecilia Bruno
	María Inés Cuenya
	Mónica Balzarini
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Sucrose accumulation curves represent the maturity profile of sugarcane cultivars, which is considered as a character of interest for the selection of genotypes in breeding programs. However, variations due to the environment (E) and interaction between genotype and environment (G×E) may be confused with the effect of genotype (G) and hinder the selection process of promising clones. The objective of this study was to identify a group of accumulation curves with high intra-group genotypic variability in the sucrose accumulation process throughout several E. This group is then used to select genotypes according to their maturity profile. A protocol is presented whereby the following statistical tools are integrated: (i) classification of non-linear accumulation curves according to parameters associated with the beginning of the maturity process, sucrose accumulation rate and the time elapsed until the accumulation rate decreases, (ii) estimation of the genotypic contribution to intragroup variability of each accumulation curve parameter within each group, and (iii) identification of the group of accumulation curves with the higher contribution of genotypic variability to total variance of sucrose accumulation parameters. The novelty of the work lies in the sequence of analytical steps to identify information useful to select genotypes according to their maturity profile. The protocol involves estimating parameters of nonlinear models for fitting maturity curves in multi-environment trials, clustering of curves according to the sucrose accumulation parameters and estimation of variability due to G, E, and G×E within each cluster to identify the group with characteristic genotypic curves. Its implementation is illustrated using 175 sucrose accumulation curves of nine sugarcane clones evaluated in different crop cycles (first and second ratoons) and several environments (7 to 50 for each clone) in Tucumán, Argentina. The proposed protocol allows identifying sucrose accumulation curves that exhibit a high genotypic variance, thus facilitating the selection of the best clones. Key words: Accumulation curves, nonlinear models, cluster analysis, variance components

Response to Reviewers:	Tucumán, Argentina, November 10th, 2020 Professor Sushil Solomon, Govind Pratap Rao Editor in chief, Sugar Tech: An International Journal of Sugar Crops and Related Industries Reference: SUTE-D-19-00320R1A We would like to thank the reviewers for their interest, complimentary comments and suggestions that certainly improved the paper. The reference style in reference section was adapted to journal style format and a brief conclusion was added at the end of discussion.
Suggested Reviewers:	Goolam BADALOO goolam.badaloo@msiri.mu
	Edison SILVA esilva@cincae.org
	James TODD james.todd@ars.usda.gov

A protocol for identifying characteristic sucrose accumulation curves of sugarcane genotypes (*Saccharum* spp.)

Santiago Ostengo[1], Angélica Rueda Calderón[2], Cecilia Bruno[2], María Inés Cuenya[1] and Mónica Balzarini[2]

[1]Estación Experimental Agroindustrial Obispo Colombres, Sección Caña de Azúcar, Tucumán, Argentina.
[2] Consejo Nacional de Ciencia y Tecnología de Argentina (CONICET), Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias-Universidad Nacional de Córdoba, Córdoba, Argentina.

Corresponding author e-mail: santiagostengo@eeaoc.org.ar

Author's Contribution MB defined the research theme and led this study with input on analytical approaches. SO and MIC performed the experiments and contributed to phenotyping. ARC, CB and SO performed statistical analyses. All authors contributed to the results interpretation, as well as manuscript preparation.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interest.

A protocol for identifying characteristic sucrose accumulation curves of sugarcane genotypes (*Saccharum* spp.)

Abstract

Sucrose accumulation curves represent the maturity profile of sugarcane cultivars, which is considered as a character of interest for the selection of genotypes in breeding programs. However, variations due to the environment (E) and interaction between genotype and environment (G×E) may be confused with the effect of genotype (G) and hinder the selection process of promising clones. The objective of this study was to identify a group of accumulation curves with high intra-group genotypic variability in the sucrose accumulation process throughout several E. This group is then used to select genotypes according to their maturity profile. A protocol is presented whereby the following statistical tools are integrated: (i) classification of non-linear accumulation curves according to parameters associated with the beginning of the maturity process, sucrose accumulation rate and the time elapsed until the accumulation rate decreases, (ii) estimation of the genotypic contribution to intra group variability of each accumulation curve parameter within each group, and (iii) identification of the group of accumulation curves with the higher contribution of genotypic variability to total variance of sucrose accumulation parameters. The novelty of the work lies in the sequence of analytical steps to identify information useful to select genotypes according to their maturity profile. The protocol involves estimating parameters of nonlinear models for fitting maturity curves in multi-environment trials, clustering of curves according to the sucrose accumulation parameters and estimation of variability due to G, E, and G×E within each cluster to identify the group with characteristic genotypic curves. Its implementation is illustrated using 175 sucrose accumulation curves of nine sugarcane clones evaluated in different crop cycles (first and second ratoons) and several environments (7 to 50 for each clone) in Tucumán, Argentina. The proposed protocol allows identifying sucrose accumulation curves that exhibit a high genotypic variance, thus facilitating the selection of the best clones.

Key words: Accumulation curves, nonlinear models, cluster analysis, variance components

INTRODUCTION

Chaining analytical processes in a logical sequence facilitates information retrieval when large amount of data is analyzed. In sugarcane, successive samplings are performed on each genotype in order to monitor sucrose accumulation (Wagih et al. 2004). The accumulation curve may show fluctuations due to the environment (E) and to the genotype by environment interaction (G×E), rather than to the effect of genotype (G). In this paper we develop a statistical protocol to identify a group of accumulation curves with high genetic variability in the curve parameters. These sucrose accumulation curves differ more in the G effect than E related effects and would, therefore, be useful for selection. The statistical workflow to identify the groups of curves used for genotype selection from a large set of sucrose accumulation curves with total variability composed of G, E and G×E effects, involves clustering of curves according to the sucrose accumulation parameters and variance components estimation within each one of the formed clusters. Two different unsupervised classification techniques (Brock et al. 2008), UPGMA and *k*-means, are used to identify accumulation curves with similar patterns in the four parameters of a non-linear two-polynomial with breakpoint model. The clustering obtained from these procedures is the one that best represents the underlying sucrose accumulation structure. The evaluation of clustering efficiency is included to obtain reliable clusters (Charrad et al. 2014). Our aim was to integrate these statistical tools in a protocol that facilitates the characterization and selection of sugarcane genotypes according to their maturity profile.

MATERIALS AND METHODS

We evaluated 175 maturity curves for nine sugarcane genotypes grown in several environments (combinations of location and year) in Tucumán, Argentina. The cultivars were: CP65-357, LCP85-384, RA87-3, TUC00-19, TUCCP77-42, TUC89-28, TUC95-10, TUC95-37, and TUC97-8. This set of genotypes represent commercial varieties widely cultivated across the sugarcane area of Tucumán (Ostengo et al. 2018) and varieties released by the Sugar Cane Breeding Program of the Estación Experimental Agroindustrial “Obispo Colombres” (EEAOC)

in recent years (TUC89-28, TUC95-37, TUC97-8, TUC95-10, and TUC 00–19) (Cuenya et al. 2009a, 2009b, 2010, 2011, 2013). Some of them are genetically related; cultivar TUCCP 77-42 is parent of TUC00-19, TUC97-8 and TUC89-28 (self-pollination of TUCCP77-42), meanwhile CP65-357 is parent of TUC95-37.

The variety trials were conducted in ten sites and evaluated during two crop cycles (first ratoon and second ratoon). The trial locations represent the environmental variability of the sugar cane area of Tucuman. The locations were: Cevil Pozo (26° 50'S, 65° 07′O), Fronterita (26° 00'S, 65° 27′O), El Colmenar (26° 47'S, 65° 11′O), Ingas (27° 26'S, 65° 21′O), La Banda (26° 59'S, 65° 23′O), Los Córdoba (27° 29'S, 65° 36′O), Los Quemados (27° 13'S, 65°14′O), Mercedes (26° 56'S, 65°19′O), Palá Palá (27°02'S, 65°13′O) and Santa Ana (27° 28'S, 65° 40′O). Not all genotypes were evaluated in all environments; with evaluations in 7 to50 environments for each clone. Each trial was planted according to a randomized complete-block design (RCBD) with two replicates per environment. Sucrose accumulation values (Pol% cane) were recorded every two weeks during the harvest season (from May to September) between 2006 and 2015. Pol% cane was determined from a sample composed of 10 stalks that were cleaned and detopped in each sampling. The samples were processed in laboratory using a cane hammer shredder (about 95% open cell) and the juice was extracted using a hydraulic press (subjected to pressure of 240 kg cm-2 per minute).

Protocol proposed for identifying characteristic sucrose accumulation curves of sugarcane genotypes

Step 1. Estimation of accumulation curve parameters

For each genotype in each environment we estimated a non-linear regression model composed of two-polynomial with a breakpoint also called "segmented polynomial" model. The model parameters are associated with the Pol% cane initial level (α), the sucrose accumulation rates in the initial ($\beta 1$) and final phases ($\beta 2$) of the harvest period and a break-point or threshold time where the curve slope changes (δ). Models were adjusted using the library STATS in R (R Core Team 2020).

Step 2. Clustering of the estimated curves

In order to clustering the estimated curves, UPGMA and *k*-means were used using the accumulation curves parameters as input variables. UPGMA was performed from the Euclidean distance matrix between accumulation curves. The software used to cluster was R (R Core Team 2020). To compare the clusters, the following validation indices were used: Connectivity, Silhouette width and Dunn (Brock et al. 2008). Those indices are available in clValid package in R. The optimum number of clusters was identified using the package NbClust in R, which calculates several indices and then the majority rule was applied to determine the number of clusters that better explain the underlying structure in the accumulation sucrose curves.

The statistical significance of the identified cluster structure was assessed by MANOVA, using the four sucrose curve parameters as response variables, and the Wilk's lambda statistics ($p < 0.05$).

Step 3. Identification of the group of curves to support genotype selection

To identify the group of curves with high genetic variance, the total variability for each parameter among curves of the same cluster, was decomposed into G, E, and G×E variance components using a linear mixed model. The library *nlme* in the software R was used in this step. The linear mixed model adjusted in order to explain the total variability of each parameter within each cluster was the following:

$$y_{ij} = \mu + E_i + G_j + \varepsilon_{ij}$$

where y_{ij} is the response variable (a parameter of the accumulation curve), μ is the overall mean, E_i is the randon effect of the i^{th} environment, $G_j$is the random effect of the j^{th} genotype, and ε_{ij} is a random term accounting for interaction (G×E) and net error, with variability estimated as residual variance.

RESULTS

Step 1

The segmented polynomial model was fitted for each sucrose curve (n=175 curves), and the four parameters collected from each fitting in a *n×4* matrix which was later subject to cluster analysis. The model equation for the population of sucrose accumulation curves is:

$$y_{ij} = 14.06 + (0.69 \times f) \times (f < 3.29) + \big(0.69 \times 3.29 + 0.24 \times (f - 3.29)\big) \times (f \geq 3.29)$$

where 14.06 Pol% cane is the estimates of the intercept α, f represent the sampled fortnight, an increase of 0.69 Pol% cane per fortnight is the estimates of β_1, 3.29 fortnight is the estimates of γ, and an increase of 0.24 Pol% cane per fortnight is the estimates of β_2, respectively.

Step 2

For both clustering methods, the optimum number of clusters was two (data not shown). We excluded nine atypical curves from the cluster analysis. UPGMA was more sensitive than *k*-means in identifying differences among accumulation curves. UPGMA exhibited a better connectivity than *k*-means and highest values of Silhouette and Dunn indices (Table 1). Low Connectivity and high values of Silhouette and Dunn indexes indicate better clustering. Moreover, according to the multivariate analysis of variance including all model parameters as responses, the two formed clusters were statistically different ($p<0.001$). The two clusters structure of sucrose curves formed by UPGMA was selected to further analysis.

Table 1. Validation of sucrose curves clusters formed by two clustering methods.

Method	Validation indices†			MANOVA††
	Connectivity	Silhouette	Dunn	p-value Hotelling Test
UPGMA	13.94	0.35	0.11	<0.001
k-means	26.32	0.34	0.10	<0.001

† Low Connectivity and high values of Silhouette and Dunn indices indicate better clustering.
†† significant of two cluster structure (MANOVA; Wilks´s λ statistics)

Step 3

Cluster C1 (n= 127 curves), obtained by UPGMA, was characterized by a high genetic variability in all the sucrose accumulation parameters (higher than E and G×E effects), and all genotypes had sucrose curves in this cluster (Figure 1). Cluster 2 (C2) showed 39 curves, but no curves from genotype CP65-357 or RA87-3 were present in this cluster with prevailing environmental variability (Table 2). Therefore, C1 was identified as the group of curves to base selection. Most locations were common to each cluster. The G×E effects were relatively low within each group (Table 2). Figure 1 shows the characteristic genotype profile in each cluster. Cluster C1 shows the highest genotypic discriminative ability.

Frequency information on the curves assigned to each group can be useful to evaluate the stability in those genotypes that present a large number of sucrose curves. In most varieties, about 70% of their curves made up the cluster C1. However, more than 50% of the curves of the TUC 95-37 variety were assigned to the cluster C2, this implies that this cultivar presents a low stability with respect to sucrose accumulation parameters during the harvest period and highly different sucrose accumulation curves can be found for TUC 95-37. In the same way, environments that were extensively evaluated, can be characterized according to their aptitude for genotypic selection. In El Colmenar and Santa Ana location, 100% of the curves were grouped as C1, while in Cevil Pozo only 37%. Therefore, El Colmenar and Santa Ana are more favourable than Cevil Pozo to select genotypes according to their maturity profile.

Table 2. Relative contribution (%) of genotype (G), environment (E) and their interaction (G×E) effects to the total variability of four sucrose curve parameter in two clusters of 166 sucrose accumulation curves.

Cluster	Effects	α	β_1	γ	β_2
C1 n=127	G	47.2	57.6	73.9	63.2
	E	44.2	33.2	17.2	26.7
	G×E	8.6	9.2	8.9	10.1
C2 n=39	G	29.1	32.9	37.4	42.6
	E	64.6	60.4	54.9	50.0
	G×E	6.3	6.7	7.7	7.4

Sucrose accumulation parameters: Pol% cane initial level (α), sucrose accumulation rate in the initial (β_1) and final phases (β_2) of the harvest period and threshold time where the accumulation rate changes (γ). All the variance components are expressed as percentage of total variance for each parameter.

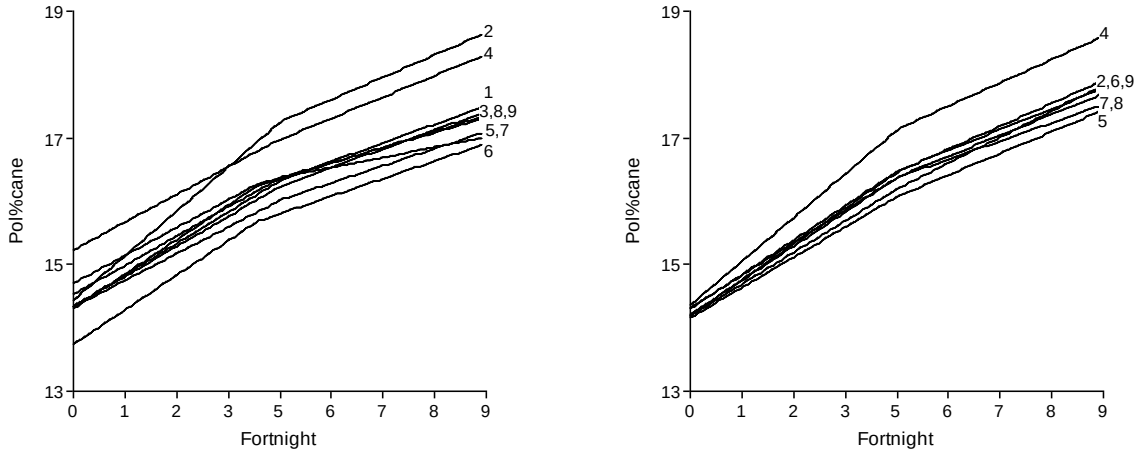

Figure1. Genotypic average of Pol% cane at each fortnight from beginning of harvest. Two groups of sucrose accumulation curves are formed, one is a group with higher genetic variance (left) and the other (right) is a group with more environmental than genetic variance. The numbers at the end of the lines are genotype codes: 1: CP65-357; 2: LCP85-384; 3: RA87-3; 4: TUC00-19; 5: TUC77-42; 6: TUC89-28; 7: TUC95-10; 8: TUC95-37; 9: TUC97-8.

DISCUSSION

Sugarcane genotype characterization according to sucrose content is usually performed at a specific time during the maturity process. In our study, the trait was measured as a parameter set (α, β_1, γ and β_2) that describes the sucrose accumulation process. These parameters represent a novel way to assess the maturity progress of sugarcane from the beginning of harvest. Saez et al. (2011) reported genotypic differences in initial content and in sucrose accumulation rate when evaluating different varieties in one single location of Tucumán.The location-by-year combinations analysed in the present work cover a wide range of rainfall (annual average rainfall between 700 and 1400 mm approximately) and different soils regarding organic matter content (low to high), drainage capacity (poor to excessively drained) and texture classes. G × E interaction was found relevant for sucrose content at the beginning of the harvest season in several analyses conducted in sugarcane area of Tucumán (Aybar Guchea et al. 2019). However, monitoring nine fortnight from the beginning of harvest, the G × E was relatively low and the sucrose accumulation curves from a given genotype in different environments

were more impacted by main environmental effects. Such environmental variations can hinder the selection process. Therefore, the identification of a subset of accumulation curves with more genetic than environmental effect is a useful tool for sugarcane breeding. To group sucrose curves characterized for many parameters a multivariate clustering approach is needed. Several works have implemented multivariate techniques for genotype clustering regarding different traits with breeding purposes (Wanessa et al. 2015; Koij and Saba 2015). In our protocol, multivariate classification of accumulation curves followed by variance component estimation allowed identify a subset of curves with high genotypic discriminative capacity. This methodology represents a breeding tool that can be used for an efficient selection of sugarcane genotypes according to their maturity profile.

REFERENCES

Aybar Guchea, Matías, Santiago Ostengo, Ernesto Raúl Chavanne, and María Inés Cuenya. 2019. Pattern analysis of genotype-by-environment interaction for yield and sucrose content of sugarcane in Tucumán, Argentina. *Proceedings of the International Society of Sugar Cane Technologists* 30: 453-458.

Brock, Guy, Vasyl Pihur, Susmita Datta, and Somnath Datta. 2008. ClValid: An R package for cluster validation. *Journal of Statistical Software* 25: 1–22. https://doi.org/10.18637/jss.v025.i04.

Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. 2014. **NbClust** : An *R* Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* 61. https://doi.org/10.18637/jss.v061.i06.

Cuenya, María Inés, Ernesto Raúl Chavanne, María Beatriz García, Santiago Ostengo, Miguel Ángel Ahmed, Diego Daniel Costilla, Carolina Díaz Romero, and Modesto Antonio Espinosa. 2009a. Comportamiento productivo y fitosanitario de TUC 95-37, una nueva variedad de caña de azúcar para la provincia de Tucumán. *Gacetilla Agroind. EEAOC* 73: 13 pp.

Cuenya, María Inés, Ernesto Raúl Chavanne, María Beatriz García, Santiago Ostengo, Miguel Ángel Ahmed, Diego Daniel Costilla, Carolina Díaz Romero, and Modesto Antonio Espinosa. 2009b. Comportamiento productivo y fitosanitario de TUC 97-8, una nueva variedad de caña de azúcar para la provincia de Tucumán. *Gacetilla Agroind. EEAOC* 74: 11 pp.

Cuenya, María Inés, Ernesto Raúl Chavanne, María Beatriz García, Santiago Ostengo, Miguel Ángel Ahmed, Diego Daniel Costilla, Carolina Díaz Romero, and Modesto Antonio Espinosa. 2010. Comportamiento productivo y fitosanitario de TUC 89-28, una nueva variedad de caña de azúcar "energética". *Avance Agroind.* 31 (3) 13-18.

Cuenya, María Inés, Ernesto Raúl Chavanne, Santiago Ostengo, María Beatriz García, Miguel Ángel Ahmed, Diego Daniel Costilla, Carolina Díaz Romero, Modesto Antonio Espinosa, Nicolás Delgado, and Jorge Díaz. 2011. TUC 95-10: Una nueva variedad de caña de azúcar altamente productiva. *Gacetilla Agroind. EEAOC* 75: 16 pp.

Cuenya, María Inés, Ernesto Raúl Chavanne, Santiago Ostengo, María Beatriz García, Miguel Ángel Ahmed, Diego Daniel Costilla, Carolina Díaz Romero, Modesto Antonio Espinosa, Jorge Díaz and Nicolás Delgado. 2013. TUC 00-19: una nueva variedad de caña de azúcar altamente productiva y de maduración temprana. *GacetillaAgroind. EEAOC* 77: 16 pp.

Koij, Fariba Shafiei, and Jalal Saba. 2015. Using Cluster Analysis and Principal Component Analysis to Group Lines and Determine Important Traits in White Bean. *Procedia Environmental Sciences* 29: 38–40. https://doi.org/10.1016/j.proenv.2015.07.145.

R Core Team. 2020. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna. Vienna, Austria.

Ostengo, Santiago, Modesto Antonio Espinosa, Jorge Díaz, Ernesto Raúl Chavanne, Diego Daniel Costilla, and María Inés Cuenya. 2018. Relevamiento de la distribución de variedades y de otras tecnologías aplicadas en el cultivo de caña de azúcar en la provincia de Tucumán: campaña 2016/2017. *Gacetilla Agroind. EEAOC* 81.

Saez, Julio Victor, Arturo Felipe, Roberto Alfredo Sopena, Claudia Rosa Celia Vega, and Alberto Mariotti Martínez Jorge. 2011. Dinámica de la acumulación de sacarosa en tallos de caña de azúcar (*Saccharum*spp.). *Ciencia y Tecnología de los Cultivos Industriales* 1: 20–23.

Wagih, Mohamed, Ambo Ala, and Yumu Musa. 2004. Evaluation of sugarcane varieties for maturity earliness and selection for efficient sugar accumulation. *Sugar Tech* 6: 297–304. https://doi.org/10.1007/BF02942512.

Wanessa, Jacqueline, De Lima Pereira, Elaine Cristina, Lucas Nunes, Jurema Mansur, Custódio Nogueira, Péricles De Albuquerque, Melo Filho, Liziane Maria De Lima, and Roseane Cavalcanti. 2015. Cluster analysis to select peanut drought tolerance lines 9: 1095–1105.