

## VARIABILIDAD GENÉTICA DE LA CAÑA DE AZÚCAR MEDIANTE EL USO DE MARCADORES DArT EN MÉXICO

### GENETIC VARIABILITY OF SUGARCANE USING DArT MARKER IN MEXICO

Hilda Victoria Silva-Rojas<sup>1</sup>, Paulino Pérez-Rodríguez<sup>1</sup>, Carlos Flores-Revilla<sup>2</sup>, Carlos Aarón Rangel-Ortega<sup>2</sup>, José Manuel Aguirre-Rayó<sup>1</sup>, Jazmín Lavín-Castañeda<sup>1</sup>, Gustavo Alonso Pinto-Hernández<sup>2</sup> y Roberto Loyo-Joachín<sup>2</sup>

<sup>1</sup>Colegio de Postgraduados, Campus Montecillo, Texcoco, Estado de México CP 56230 México

<sup>2</sup>Centro de Investigación y Desarrollo de la Caña de Azúcar A.C., Ciudad de México CP 06500 México

#### Resumen

La caña de azúcar (*Saccharum* spp. híbridos) es uno de los cultivos agroindustriales más importante para la producción de azúcar a nivel mundial. Es una monocotiledónea poliploide, con un complejo genoma que ha restringido el uso de estudios genéticos clásicos en los programas de mejoramiento de este cultivo. Una de las técnicas que ha dado los mejores resultados para la genotipificación de clones de caña de azúcar es el uso de marcadores Diversity Array Technology (DArT) que permite estimar la diversidad genética de la caña de azúcar basado en la reducción eficiente del genoma. Por tal razón, el objetivo del presente trabajo fue genotipificar y determinar las distancias genéticas de 282 clones procedentes de la Estación de Hibridación del CIDCA en Chiapas. La extracción del DNA se realizó a partir de hojas jóvenes de cada clon, los que se enviaron a la compañía DArT en Australia para el análisis del genoma con una alta profundidad. Los resultados mostraron 52,828 marcadores de tipo “presencia” (1) y “ausencia” (0) de un fragmento de restricción que contiene el marcador de interés. Se eliminaron marcadores de baja calidad, obteniéndose 43,250 marcadores en total. Una pequeña fracción de los datos presentó datos faltantes, razón por la cual fueron imputados generando muestras aleatorias de la distribución marginal de los genotipos observados. A partir de los marcadores se obtuvo la matriz de distancias Euclidianas (*E*) y la matriz de relaciones genómicas (*G*). La matriz *E* fue utilizada para construir un agrupamiento jerárquico de los individuos. Los clones también se agruparon utilizando partición basada en medioides, la gráfica de siluetas resultó en la formación de seis grupos. Una de las principales contribuciones de la presente investigación fue el conocimiento de las distancias genéticas de 282 clones que podrán ser utilizadas en el programa anual de mejoramiento genético del CIDCA en México.

**Palabras clave:** cruzamiento, DArT, distancias Euclidianas, allopoliploide, matriz de relaciones genómicas

#### Abstract

Sugarcane (*Saccharum* spp. hybrids) is one of the most important agro-industrial crops for sugar production worldwide. It is a polyploid monocotyledonius, with a genome complex that has restricted the use of classical genetic studies in the breeding programs of this crop. One of the techniques that has given the best results for the genotyping of sugarcane clones is the use Diversity Array Technology (DArT) markers that allows estimating the genetic diversity of sugar cane based on the efficient reduction of the genome. For this reason, the objective of this work was to genotype and determine the genetic distances of 282 clones from the CIDCA Hybridization Station in Chiapas. The extraction of the DNA was done from young leaves of each clone, which were sent to the company DArT in Australia for the analysis of the genome with a high depth. The results showed 52,828 markers of type "presence" (1) and

"absence" (0) of a restriction fragment containing the marker of interest. Markers of low quality were eliminated, obtaining 43,250 markers. A small fraction of the data presented missing data, which is why they were imputed generating random samples of the marginal distribution of the observed genotypes. The Euclidean distances matrix (E) and the genomic relations matrix (G) were obtained from the markers. Matrix E was used to construct a hierarchical grouping of individuals. The clones were also grouped using partition based on medioides, the graph of silhouettes resulted in the formation of six groups. One of the main contributions of this research was the knowledge of the genetic distances of 282 clones that could be used in the annual genetic improvement program of CIDCA in Mexico.

**Keywords:** Crosses, DArT, Euclidean distances, allopolyploid, matrix of genomic relations

### Introducción

La caña de azúcar (*Saccharum* spp. híbridos) es una de las gramíneas de uso industrial más importante para la producción de azúcar y etanol a nivel mundial (Dookun-Saumtally et al., 2012). En una monocotiledónea allopoliploide originaria de Oceanía (Nueva Guinea), obtenida por recombinación genética *Saccharum officinarum* y *S. spontaneum* con un número no definido de cromosomas que varía de  $2n = 100$  a 130. La alta poliploidía y heterocigosidad que se genera al momento de la hibridación en condiciones naturales es restringida por lo que se dificulta la aplicación de métodos de genética clásica en este cultivo, además del tiempo que se necesita para validar los resultados (Premachandran et al., 2011). Sin embargo, esta condición de poliploide le permite tener un gran rango de adaptabilidad a diferentes condiciones ambientales en regiones tropicales y subtropicales de mundo, sembrándose en más de 103 países en los cinco continentes (Mendes de Paula et al., 2014).

En México, desde su introducción en 1522, se ha sembrado en diferentes climas y condiciones topográficas, aunque tiene un gran rango de adaptación, se requiere de la utilización de variedades que tengan un alto índice de respuesta en cada zona y que permita expresar su máximo potencial productivo de tal manera que sea económicamente rentable para el productor (Aguilar-Rivera et al., 2010).

En el ámbito internacional, México es reconocido por ser el sexto país productor de caña de azúcar, sin embargo el 77% de las aproximadamente 824,000 ha se siembran con cuatro variedades entre las que se encuentran la CP 72-2086, MEX 69-290, MEX 79-431 e ITV 92-1424 (CONADESUCA, 2016; Senties-Herrera et al., 2014). Considerando que estas variedades se obtuvieron hace más de 30 ó 40 años, se están volviendo vulnerables a las condiciones de cambio climático, por lo que se tiene la necesidad de renovar y diversificar las variedades utilizadas en las zonas productoras de este cultivo, que traerá beneficios directos tanto al sector primario como a la industria mejorando los rendimientos de campo y fábrica (Aguilar-Rivera et al., 2010).

El programa de mejoramiento genético para la obtención de nuevas variedades de caña de azúcar estuvo a cargo del Instituto para el Mejoramiento de la Producción de Azúcar (IMPA) quien coordinaba trabajos de mejoramiento genético de la caña hasta 1990. A la fecha estas investigaciones han sido retomadas por el Centro de Investigación y Desarrollo de la Caña de Azúcar A.C., quien cuenta con un banco de germoplasma con alrededor de 1,200 accesiones que florecen de manera natural. Este programa de mejoramiento está enfocado en la producción de nuevas variedades con características agronómicas (buen porte de las plantas, resistentes al acame, resistentes a plagas y enfermedades etc.), genéticas (presencia de genes de rendimiento, producción de azúcar etc.) e industriales (porcentaje de sacarosa, fibras, mieles etc.).

Debido al periodo que se necesita para poder liberar una variedad entre 14 y 16 años se ha buscado alternativas que permitan reducir estos tiempos, por lo que se ha considerado el uso de técnicas modernas basadas en marcadores moleculares. Esta metodología ha sido empleada de forma exitosa en otras especies con diferentes nivel de ploidía (e.g. maíz, trigo, cebada, etc.).

Estudios en caña de azúcar realizados en Australia (Aitken et al., 2014; Al-Khayri et al., 2015; Berkman et al., 2013; Huang et al., 2010) y Brasil (García et al., 2006; García et al., 2013) han permitido identificar numerosos marcadores asociados con caracteres de importancia agronómica, además del desarrollo de mapas genéticos, entre otros avances (Jaccoud et al., 2001; Manners, 2011). Una de las técnicas desarrolladas para la genotipificación de clones de caña de azúcar es la técnica de Diversity Array Technology (DArT) desarrollada por Heller-Uszynska et al. (2010) se ha convertido en una novedosa plataforma para el análisis genético, genotipificación de todo el genoma y estudios de la diversidad genética de caña de azúcar basado en la reducción eficiente del genoma.

Ante este panorama, resulta evidente la posibilidad de complementar las características agronómicas e industriales de la caña de azúcar con la información genética que pueda apoyar los programas de mejoramiento para la obtención de variedades adaptadas a cada región en los que se pueda determinar la resistencia o susceptibilidad a plagas, enfermedades y con un alto contenido de sacarosa.

Por lo cual se llevó a cabo el presente trabajo con el objetivo de determinar las distancias genéticas entre 282 clones de caña de azúcar obtenidas en la Estación de Hibridación del CIDCA A.C. para ser utilizados en el programa anual de cruzamientos.

## Materiales y Métodos

### Lugar de estudio

Se colectaron hojas jóvenes de 282 clones de caña de azúcar mantenidos en el Banco de Germoplasma de la Estación de Hibridación del CIDCA A.C., ubicado en Tuxtla Chico, Chiapas (Figura 1).



**Figura 1.** Regiones productoras de caña de azúcar en México, indicando la ubicación de la Estación de Hibridación del Centro de Investigación y Desarrollo de la Caña de Azúcar A.C. en Tuxtla Chico, Chiapas, México.

### **Material de estudio**

Se tomó una porción de tejido de hojas jóvenes de 282 clones seleccionados previamente con base a su genealogía e inclusión en el programa anual de mejoramiento genético. El tejido se colocó dentro de una bolsa plástica, la que se etiquetó y colocó en un recipiente de unicel con hielo a 4 °C. Los recipientes se trasladaron al laboratorio donde se procedió a lavar la muestra con agua destilada, posteriormente se desinfectó con una solución de hipoclorito de sodio al 1% durante 3 min y se enjuagó tres veces con agua destilada estéril. Las muestras se colocaron en papel toalla estéril para eliminar el exceso de agua, se re-etiquetó y se colocaron en una nueva bolsa de plástico las que se mantuvieron a -20°C hasta su procesamiento.

### **Extracción de DNA**

Se realizó con el método de Bromuro de Cetiltrimetilamonio al 2% (CTAB) /cloroformo/ alcohol isoamílico (Doyle y Doyle, 1987) con la adición de acetato de sodio.

La concentración de DNA de cada una de las muestras se determinó por espectrofotometría en un NanoDrop 2000 UV-Vis (Thermo Scientific, USA). Se consideró que el DNA es de buena calidad cuando las lecturas de las absorbancias  $A_{260}/A_{280}$  y  $A_{260}/A_{230}$ , estuvieron en un rango entre 1.8 y 2.0.

### **Análisis de los marcadores DArT**

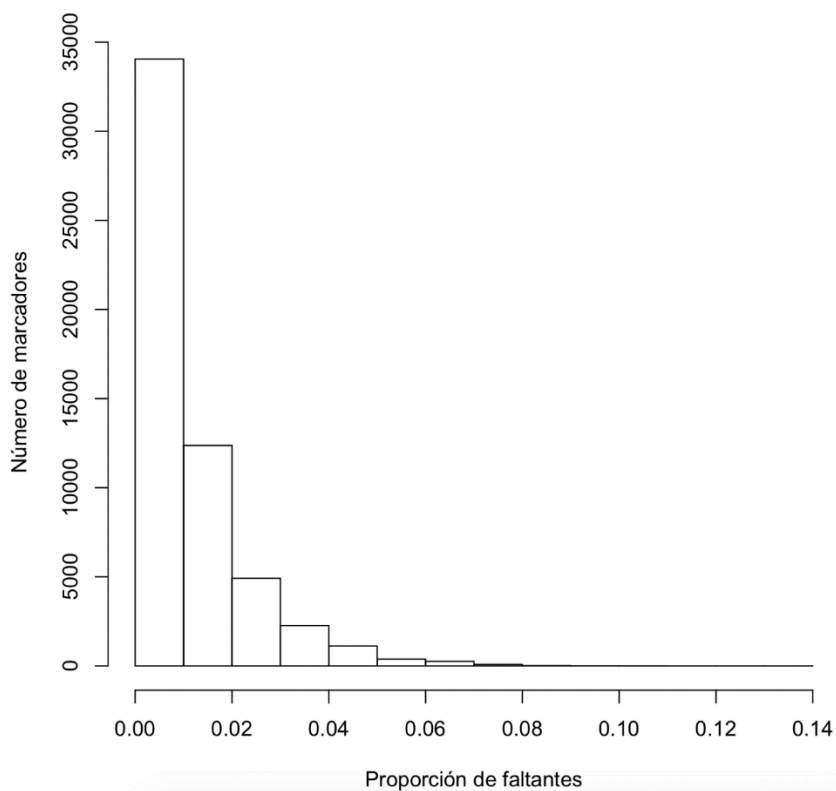
La preparación de las representaciones genómicas individuales de los genotipos de caña de azúcar “Targets” (=objetivos) para la hibridación con las matrices DArT, se realizó con los metadatos y los polimorfismos SNP obtenidos en el HiSeq 2500 de la plataforma de Illumina. Tanto el análisis de la imagen, la identificación del polimorfismo y el registro de datos, se realizaron con el programa DArTSoft especialmente desarrollado para este propósito por Diversity Arrays Technology Pty. Ltd. ([www.diversityarrays.com/software.html](http://www.diversityarrays.com/software.html)) como lo mencionó Wenzl et al. (2004) y Akbari et al. (2006). DArT analizó 52,828 marcadores para cada muestra e identificó, anotó los polimorfismos y calculó los parámetros de calidad para cada marcador. Los resultados finales se analizaron utilizando el programa R (R Core Team, 2016, <https://www.R-project.org/>) para determinar los datos faltantes, el mapa de calor (Heat map) y las distancias euclidianas, así como los análisis de datos fenotípicos y genotípicos en forma conjunta con la finalidad de realizar predicciones genómicas con los datos obtenidos en el presente análisis.

## **Resultados y Discusión**

### **Caracterización genética**

Se obtuvo información genética de 282 clones de caña de azúcar genotipificados con la tecnología DArT. Los marcadores obtenidos son de tipo ‘presencia’ (1) y ‘ausencia’ (0) de un fragmento de restricción que contiene el marcador de interés. En total se obtuvieron 55,487 marcadores con un porcentaje de marcadores con valores faltantes.

La Figura 2 representa la distribución de valores faltantes para los 55,487 marcadores, también se observa que más de 30,000 marcadores tienen una proporción de entre 0 y 0.01 de valores faltantes, lo cual coincide con la moda de la distribución, lo que indica que la información genotípica es de muy buena calidad.



Min. 1st Qu. Median Mean 3rd Qu. Max.  
 0.000000 0.000000 0.007092 0.009706 0.014184 0.131206

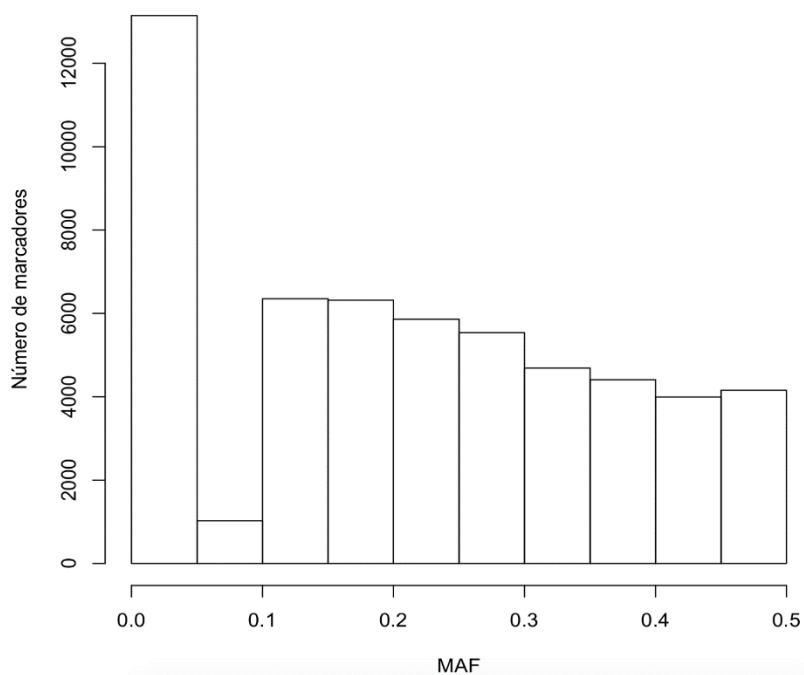
**Figura 2.** Distribución de valores faltantes.

Debido a que no se puede trabajar con las tasas de valores faltantes por marcador que son bastante bajas, se realizó una imputación basada en las frecuencias de los marcadores observados. Después de imputar y eliminar marcadores cuya MAF es menor del 0.05 se obtuvieron 42,338 marcadores.

Los genotipos faltantes se imputaron generando muestras aleatorias de la distribución marginal de los genotipos observados, es decir:

$$x_{ij} \sim \text{Bernoulli}(p^*_j)$$

donde  $p^*_j$  representa la frecuencia alélica calculada usando los genotipos no faltantes (Cossa-Iriarte et al., 2010).

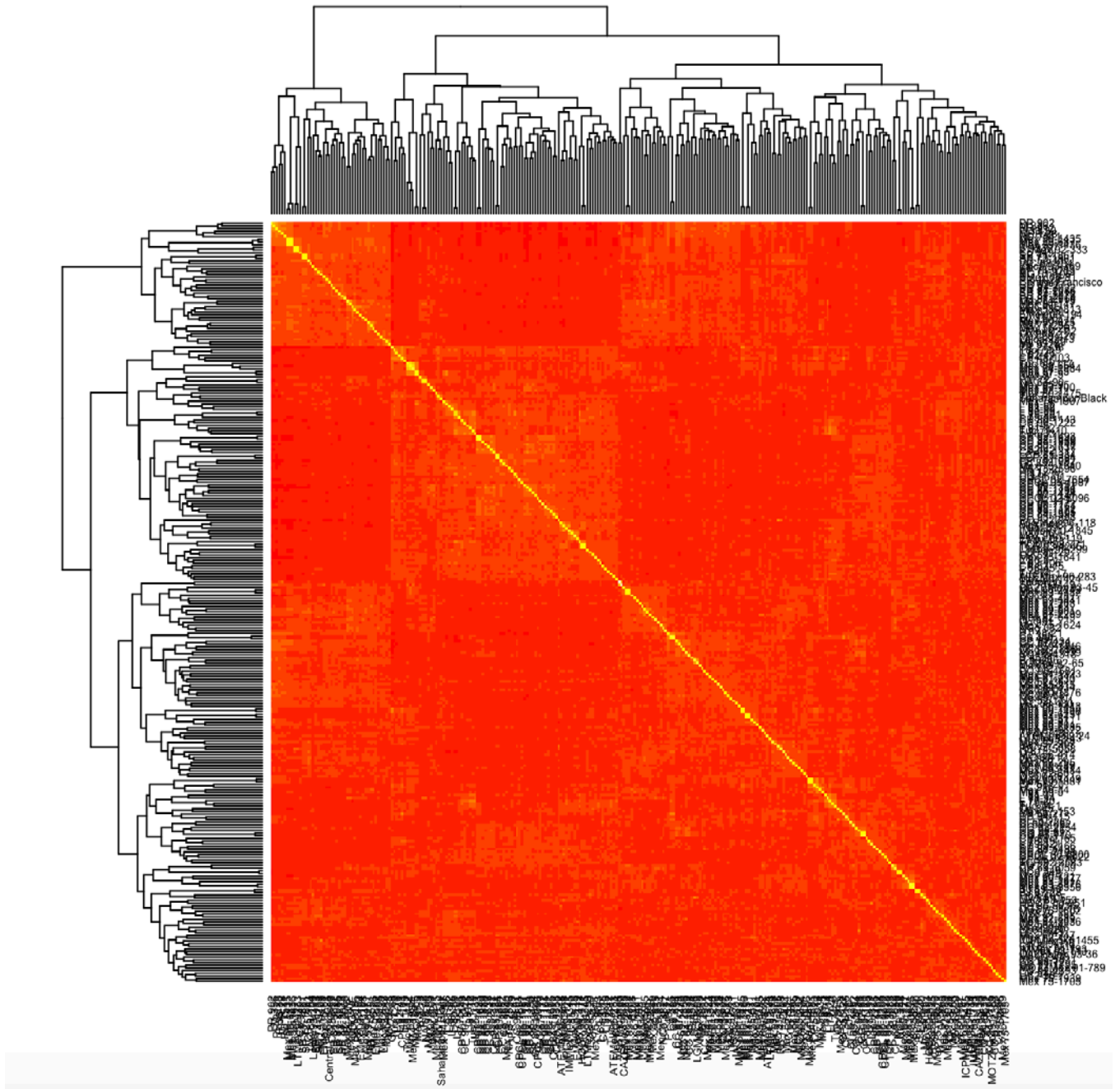


**Figura 3.** Distribución de frecuencias del alelo menor (MAF).

Después de imputar y eliminar marcadores en base a MAF se obtuvo la matriz de relaciones genómicas  $G$ , misma que puede ser utilizada para estudios de la estructura de la población de interés o en predicción genómica. La matriz puede calcularse fácilmente usando la expresión siguiente:

$$G = ZZ^T/p, \dots(1)$$

donde  $Z$  es la matriz de marcadores de dimensiones  $n=282$  individuos y  $p=42,338$ , que se obtiene al centrar y estandarizar las columnas de la matriz de marcadores con 0's y 1's (López-Cruz et al., 2015). La Figura 4 presenta un mapa de calor (heatmap) de las 282 líneas que se obtiene usando la matriz  $G$ .

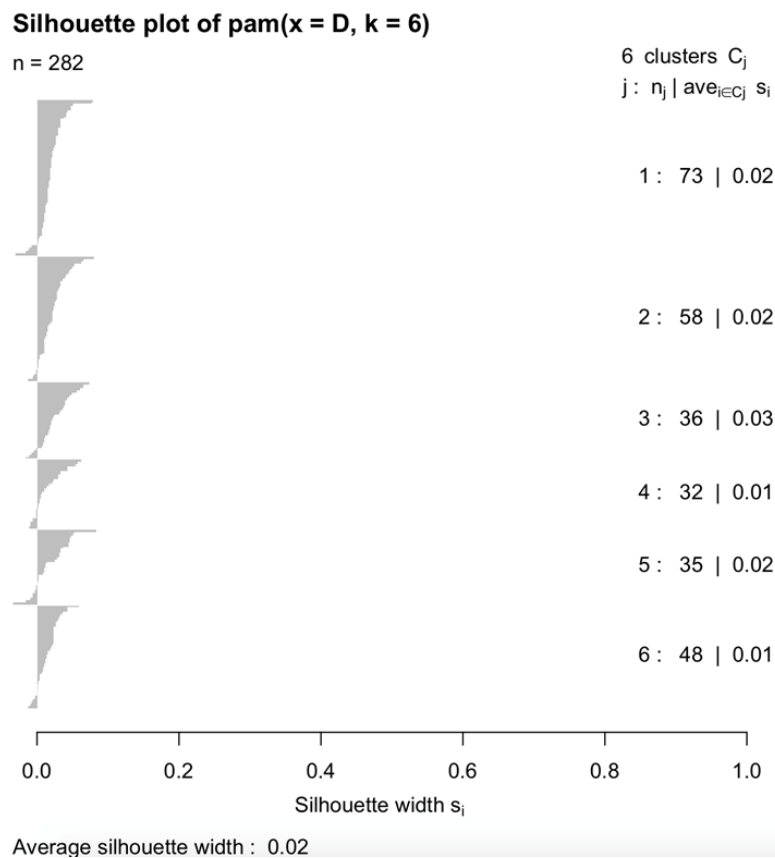


**Figura 4.** Mapa de calor para los 282 clones de caña de azúcar procedente de la Estación de Hibridación del CIDCA A.C. analizadas mediante la técnica de Diversity Array Technology (DARt).

Basada en la matriz de marcadores codificada con 0/1 se realizó también un agrupamiento de los 282 clones. Primero se calculó una matriz de distancias Euclidianas y luego se hizo un agrupamiento

jerárquico utilizando encadenamiento completo, de tal manera que el dendrograma muestra las distancias entre los individuos con base al marcador.

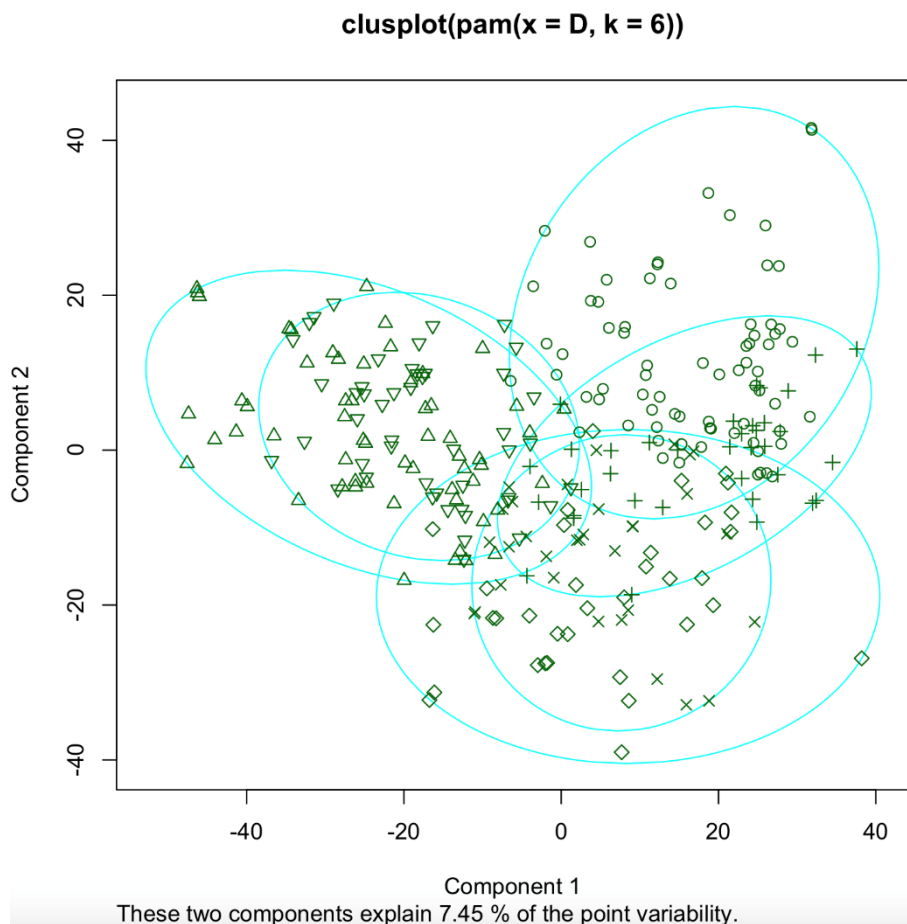
Las variedades también se agruparon utilizando partición basada en medioides (Kaufman y Rousseuw, 2005). La Figura 5 muestra la gráfica de siluetas resultante para 6 grupos. El ancho promedio de silueta mide la calidad del agrupamiento, valores cercanos a 1 indican que una variedad está bien clasificada y valores negativos que la variedad está mal clasificada.



**Figura 5.** Diagrama de siluetas para agrupamientos resultante usando partición basada en medioides.

La Figura 6 muestra una gráfica de los 2 primeros componentes principales basada en la matriz de distancias Euclidianas, muestra también óvalos que incluyen las variedades pertenecientes a cada uno de los grupos, mismos que se identifican con diferentes tipos de símbolos en la gráfica. Los dos primeros componentes principales explican solo el 7.45% de la variabilidad total. Los análisis de datos fueron realizados en el paquete estadístico R (R Core Team, 2016).





**Figura 6.** Componente principal 1 vs Componente principal 2 explican el 11.55% de la variabilidad, asimismo el agrupamientos para los 282 clones de caña de azúcar.

El uso de marcadores moleculares DArT desarrollado para especies poliploides como caña de azúcar, permitió realizar el análisis genético de los clones analizados. Con estas herramientas se podrán determinar las distancias genéticas de otros individuos que se encuentran en el banco de germoplasma del CIDCA AC., el conocimiento de la información genética más distante permitirán en algunos casos obtener híbridos con las mejores características agronómicas, genéticas, sanitarias y agroindustriales que puedan ser evaluados en las diferentes regiones productoras del país.

### Conclusiones

El uso marcadores moleculares DArT desarrollado para especies poliploides como caña de azúcar, permitió realizar el análisis genético de los 282 clones analizados. Con estas herramientas se determinaron las distancias genéticas entre los clones en estudio, los que serán utilizados en el programa anual de cruzamientos en la Estación de Hibridación del CIDCA AC.

## Referencias

- Aguilar-Rivera, N.; Galindo-Mendoza, G.; Contreras-Servin, C.; Fortanelli-Martínez, J. (2010). Competitiveness and productivity of Mexico's sugar mills. *Theoria* 19:7-30.
- Aitken, K. S.; McNeil, M. D.; Hermann, S.; Bundock, P. C.; Kilian, A.; Heller-Uszynska, K.; Henry, R. J.; Li, J. (2014). A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput Diversity Array Technology (DArT) markers. *BMC Genomics* 15:152. DOI: 10.1186/1471-2164-15-152.
- Akbari, M.; Wenzl, P.; Caig, V.; Carling, J.; Xia, L.; Yang, S.; Uszynski, G.; Mohler, V.; Lehmensiek, A.; Kuchel, H.; Hayden, M. J.; Howes, N.; Sharp, P.; Vaughan, P.; Rathmell, B.; Huttner, E.; Kilian, A. (2006). Diversity Arrays Technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor Appl Genet* 113:1409–1420.
- Al-Khayri, J. M.; Mohan, J. S.; Johnson, D. V. (2015). *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*. Springer International Publishing. 645pp.
- Berkman, P. J.; Casu, R. E.; Stiller, J.; Rae, A. L.; Aitken, K. S. (2013). Towards the sugarcane genome sequence and an understanding of polyploid. *Proc. Int. Soc. Sugar Cane Technology* 28:1-10.
- Dookun-Saumtally, A.; Saumtally, S.; Joomun, N. (2012). Biotechnological applications to sugarcane pathogens. *Functional Plant Science and Biotechnology* 6:1-11.
- CONADESUCA. (2016). Sistema de Información de la Investigación en la Agroindustria de la Caña de Azúcar. Consultada el 02 de octubre de 2016.  
<http://www.gob.mx/conadesuca/acciones-y-programas/sistema-infocana?idiom=es>
- Crossa, J.; de los Campos, G.; Pérez-Rodríguez, P.; Gianola, D.; Burgueño, J.; Araus, J. L.; Makumbi, D.; Singh, R. P.; Dreisigacker, S.; Yan, J.; Arief, V.; Banziger, M.; Braun, H.-J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724.
- Doyle, J. J.; Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:1–15.
- Garcia, A. A.; Kido, E. A.; Meza, A. N.; Souza, H. M.; Pinto, L. R.; Pastina, M. M.; Leite, C. S.; Silva, J. A.; Ulian, E. C.; Figueira, A.; Souza, A. P. (2006). Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theor Appl Genet* 112:298-314.
- Garcia, A. A.; Mollinari, M.; Marconi, T. G.; Serang, O. R.; Silva, R. R.; Vieira, M. L.; Vicentini, R.; Costa, E. A.; Mancini, M. C.; Garcia, M. O.; Pastina, M. M.; Gazaffi, R.; Martins, E. R.; Dahmer, N.; Sforça, D. A.; Silva, C. B.; Bundock, P.; Henry, R. J.; Souza, G. M.; van Sluys, M. A.; Landell, M. G.; Carneiro, M. S.; Vincentz, M. A.; Pinto, L. R.; Vencovsky, R.; Souza, A. P. (2013). SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Nature: Scientific Reports* 3, Article number: 3399. DOI: 10.1038/srep03399.
- Heller-Uszynska, K.; Uszynski, G.; Huttner, E.; Evers, M.; Carlig, J.; Caig, V.; Aitken, K.; Jackson, P.; Piperidis, G.; Cox, M.; Gilmour, R.; D'Hont, A.; Buterfield, M.; Glaszmann, J. C.; Kilian, A. (2010). Diversity arrays technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Mol Breed* 28:37-55.
- Huang, E.; Aitken, K.; George, A. (2010). Association Studies. pp. 43-68. In: Henry, R. J., and Kole, C. (eds.). *Genetics, Genomics and Breeding of Sugarcane*. CRC Press. Science Publishers, USA. 300pp.

- Jaccoud, D.; Peng, K.; Feinstein, D.; Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 29 (e25):1-7.
- Jackson, P. A.; Morgan, T. E. (2003). Early stage selection for commercial cane sugar (CCS) in sugarcane clones: effects of time of sampling and irrigation. *Crop and Pasture Science* 54:389-396.
- Kaufman, L.; Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York. 342pp.
- López-Cruz, M.; Crossa-Iriarte, J.; Bonnet, D.; Dreisigacker, S.; Poland, J.; Jannink, L-L.; Singh, R. P.; Autrique, E.; de los Campos, G. (2015). Increased prediction accuracy in wheat breeding trials using a marker  $\times$  environment interaction genomic selection model. *G3:Genes|Genome|Genetics* doi:10.1534/g3.114.016097.
- Manners, J. M. (2011). Functional Genomics of Sugarcane. pp. 91-146. In: Jean-Claude Kader and M. Delseny (eds.). *Advances in Botanical Research*. Volume 60. Academic Press, USA. 506pp.
- Mendes de Paula, T. O.; Marinho, C. D.; Souza, V.; Barbosa, M. H.; Peternelli, L. A.; Kimbeng, C. A.; Zhou, M. M. (2014). Relationships between methods of variety adaptability and stability in sugarcane. *Genetic and Molecular Research* 13:4216-25.
- Premachandran, M. N.; Prathima, P. T.; Lekshmi, M. (2011). Sugarcane and polyploidy, a review. *J Sugarcane Res* 1:1-15.
- R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Senties-Herrera, H.; Gómez- Merino, F. C. (2014). Nuevas directrices en mejoramiento genético de caña de azúcar. *Colegio de Postgraduados Campus Córdoba, México. Agroproductividad* 7:9-15.
- Wenzl, P.; Carling, J.; Kudrna, D.; Jaccoud, D.; Huttner, E.; Kleinhofs, A.; Kilian, A. (2004). Diversity Arrays Technology (DART) for whole-genome profiling of barley. *Proc Natl Acad Sci USA* 101:9915–9920.