



# Seamaíz

XI Congreso Nacional de Maíz

GENÉTICA Y MEJORAMIENTO  
GENÉTICO VEGETAL



## **ANÁLISIS DE LA ESTRUCTURA DE UN PANEL DE LÍNEAS DE MAÍZ DEL PROGRAMA DE MEJORAMIENTO DE INTA BASADO EN MARCADORES SSR Y SNP**

---

Solmi, L.<sup>1</sup>; Delucchi, C.<sup>2</sup>; Lorea, R.<sup>2</sup>; Eyhérbide, G. y Díaz Paleo, A.<sup>2\*</sup>

<sup>1</sup> Universidad Nacional del Noroeste de la provincia de Buenos Aires, Monteagudo 2332, Pergamino, Buenos Aires, Argentina.

<sup>2</sup> Instituto Nacional de Tecnología Agropecuaria, Estación Experimental Agropecuaria Pergamino, Ruta 32 Km4,5, Buenos Aires, Argentina.

\*diazpaleo.antonio@inta.gob.ar

---

### **ANALYSIS OF THE STRUCTURE OF A PANEL OF MAIZE LINES OF THE INTA BREEDING PROGRAM BASED ON SSR AND SNP MARKERS**

#### **ABSTRACT**

The diversity and population structure of a collection of 157 maize lines belonging to the INTA Breeding Program were estimated with STRUCTURE 2.3.4 program. In a first approach a set of 1000 SNP, uniformly distributed through the genome, were run in STRUCTURE. The resulting population substructure was compared with those achieved by two subsets of 250 and 500 SNP. The best results were obtained with 1000 SNP and these were later compared with a set of 50 SSR. Each of the two types of markers clustered the lines in K=8 subpopulations. The SSR grouped the lines with higher Q membership values than the SNP. Five of these groups were mostly similar in the members included and three differed in the member lines they included. When the substructure was analyzed with the 50 SSR and the 1000 SNP together, 9 subpopulations were recognized, improving the overall Q values. When feasible, it is recommended to use both types of markers.

#### **Palabras Clave**

SSR, SNP, STRUCTURE 2.3.4, CLUMPP, Estructura poblacional.

#### **Key Words**

SSR, SNP, STRUCTURE 2.3.4, CLUMPP, Population structure.

## INTRODUCCIÓN

La estimación de diversidad genética y estructura poblacional dentro de una colección de germoplasma son informaciones útiles para el mejoramiento vegetal que se pueden obtener a partir de distintos marcadores moleculares. Población estructurada hace referencia a la existencia de diversas clases de individuos, que corresponden a subpoblaciones producidas a partir de variables que modifican a la población. Dentro de tales variables, el origen de los individuos a partir de los fundadores de la población produce relaciones genealógicas que introducen diferencias entre subpoblaciones, por medio de cambios en las frecuencias génicas y genotípicas. Los individuos pueden entonces agruparse a partir de la información alélica.

En la actualidad, los marcadores moleculares son masivamente utilizados para la caracterización de poblaciones, en especial los de polimorfismo de nucleótido simple (SNP, Single Nucleotide Polimorphysm) y los microsatélites (SSR, Simple Sequence Repeats). Las distintas propiedades de los marcadores SSR y SNP están asociadas con los procesos de mutación propios de cada uno. Debido a

que las tasas de mutación son bajas en los SNP (en el orden de  $10^{-8}$ pb/generación), la gran mayoría son bialélicos, y por lo tanto, tienen una heterocigosidad máxima de 0,5. La tasa de mutación de los SSR es mucho más alta (para las repeticiones de dinucleótidos en maíz,  $5,2 \times 10^{-4}$  a  $1,1 \times 10^{-3}$ pb/generación) (Hamblin *et al.*, 2007), por lo que la heterocigosidad máxima puede acercarse a 1.

El objetivo del trabajo fue realizar un análisis comparativo de los resultados de estructura poblacional obtenidos a partir de marcadores SSR y SNP en una población de líneas endocriadas de maíz pertenecientes al Programa de Mejoramiento de la Estación Experimental Pergamino de INTA.

## MATERIALES Y MÉTODOS

**Material vegetal:** Se seleccionó una población de 157 líneas de maíz, del Programa de Mejoramiento de la EEA Pergamino, las cuales concentran una amplia variabilidad fenotípica y representan orígenes diversos.

**Genotipado con SNP y SSR:** A través de un convenio de colaboración con la empresa Syngenta, se disponía de la información de genotipado con SNP, con el chip MaizeSNP50 Beadchip, (Illumina Inc.), de un panel de 270 líneas de maíz de INTA Pergamino. En total se obtuvieron datos de aproximadamente más de 50.000 loci SNP distribuidos a lo largo del genoma de cada genotipo. La información de cada SNP se obtuvo de los trabajos de Galal *et al.* (2011) y Cook *et al.* (2012). Los marcadores SNP fueron filtrados por posición,

utilizando el programa Tassel 5 (Bradbury *et al.*, 2007), y se eligieron 1000 SNP uniformemente distribuidos en el genoma. Asimismo, se efectuó el genotipado con 50 marcadores SSR, distribuidos uniformemente, de las 157 líneas que componen el panel. La información de los SSR como así también el protocolo de genotipado se encuentran publicados (Olmos *et al.*, 2014).

**Análisis de la estructura poblacional:** El análisis de estructura poblacional se efectuó sobre las 157 líneas con los 50 marcadores SSR y con 250, 500 o 1000 SNP de los elegidos, manteniendo la distribución uniforme. Se aplicó el programa STRUCTURE 2.3.4. (Pritchard *et al.* 2000). Se realizaron 5 corridas independientes para cada grupo de marca-

dores, fijando el número de subpoblaciones (K) de 2 a 12, el número de iteraciones del período burn-in y de MCMC (Monte Carlo MarkovChain) de 10000, 50000 y 100000, y el modelo de ancestría mezcla (admixture). Se calcularon los valores de probabilidad posterior de los datos a través de múltiples va-

lores de K mediante el método de Evanno *et al.* (2005) para detectar el número de subpoblaciones originarias que mejor se adaptaban a los datos. Mediante el programa CLUMPP (Jakobsson *et al.*, 2007) se obtuvo la matriz Q ajustada entre repeticiones sucesivas, para el K más probable.

## RESULTADOS Y DISCUSIÓN

Los agrupamientos llevados adelante con SSR o SNP mostraron valores de K similares utilizando 50000 o 100000 iteraciones. Los valores K estimados con 10000 iteraciones fueron más variables, por lo tanto, se descartó como alternativa para agrupar correctamente las líneas. Para el caso de los SSR el valor de K determinado fue de 8 subpoblaciones. El número de subpoblaciones K estimado fue 8 con 500 y 1000 SNP. El valor de K fue igual a 9 con 250 SNP.

La utilización de 50 SSR arrojó resultados similares a 500 o 1000 SNP. Este resultado es consistente con lo publicado por otros autores que en maíz encontraron que 7 a 11 veces más cantidad de SNP que de SSR se deberían utilizar para analizar la estructura de las poblaciones. En la tabla 1 se observa que tan solo 50 SSR agrupan las líneas en subpoblaciones con altos valores de membresía, Q, superiores o comparables a 1000SNP.

Para todos los grupos de marcadores, se obtuvieron 5 subpoblaciones que están formadas por líneas genéticamente similares, de origen genético común y alta proporción de líneas con valores de membresía mayor o igual a 0,7. Estas subpoblaciones incluyen líneas derivadas de P465, derivadas del híbrido Cargill 280, derivadas de la sintética BS13P, derivadas de la sintética 34 de INTA Leales, y una subpoblación mixta de líneas derivadas de híbridos Pioneer más un grupo de líneas derivadas de una sintética semi dentada semi precoz.

En cambio, en las 3 subpoblaciones (derivadas de sintética colorada, LP122 + LP199, Familia Argentino-Canadá+híbrido M11 + los híbridos AX88#) restantes se encontraron diferencias más importantes en los agrupa-

mientos según los conjuntos de marcadores utilizados. Los agrupamientos con 1000 SNP tuvieron mayores valores de asignación a los grupos y por lo tanto más homogéneos, que los de 500 o 250 SNP, por lo que se considera mejor la utilización de 1000SNP. Se observan a su vez diferencias entre los 1000SNP y los 50SSR en cuanto a orígenes de las líneas y porcentajes de asignación. La cantidad de líneas con valores Q mayor o igual a 0,7 variaron según el tipo de marcador utilizado. En estos agrupamientos se destaca que si bien el número total de subpoblaciones es el mismo, los marcadores agrupan de forma distinta.

Al agrupar con 50 SSR y 1000 SNP en conjunto, se obtuvieron 9 subpoblaciones, una más que con ambos tipos de marcadores separados (Figura 1). El grupo extra está formado por la línea LP562 y derivadas. La información adicional de los 50 SSR permitió diferenciar este subconjunto a partir de las subpoblaciones más heterogéneas para ambos tipos de marcadores, mejorando los valores Q de asignación de las líneas del nuevo grupo.

## CONCLUSIÓN

La asignación por membresía de las líneas a las subpoblaciones fue alta tanto a partir de los SSR como de los SNP, siendo superior con los SSR.

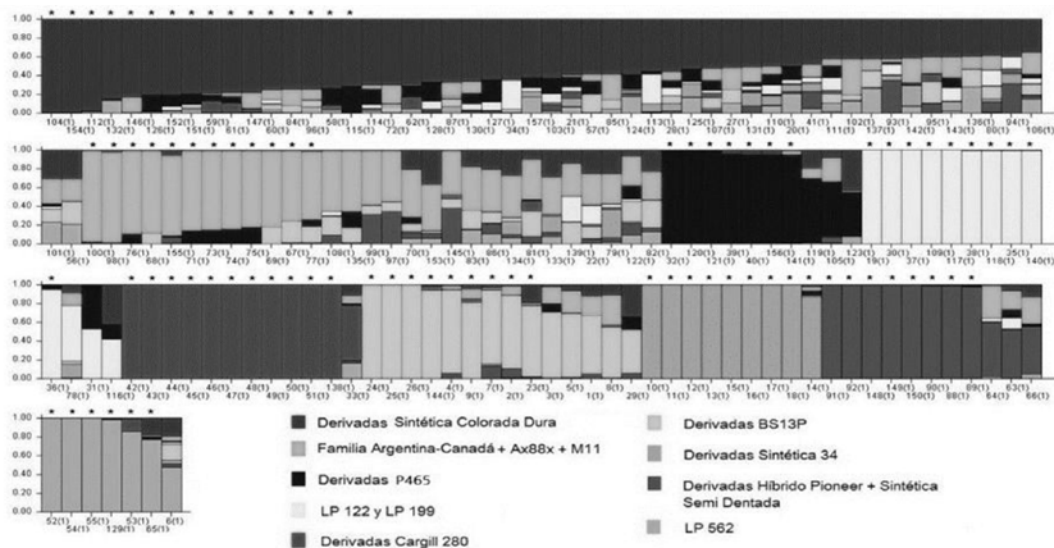
Es llamativo el alto poder de discriminación de los SSR con respecto a los SNP. En nuestro caso la relación SNP/SSR fue igual a 20. Mayores valores de PIC de los SSR podrían explicar esta ventaja.

Es recomendable la utilización conjunta de SSR+SNP porque aumenta la discriminación de las líneas a subpoblaciones con valores de asignación superiores.

Este proyecto ha sido parcialmente financiado por PNCyO 1127043 y el PNBIO 1131042 de INTA.

Umbral de Q	50 SSR	1000 SNP
Q > 0,8	47,13%	45,22%
Q > 0,7	64,97%	55,41%
Q > 0,6	73,89%	68,15%

**Tabla 1.** Porcentaje de líneas asignadas a subpoblaciones (K=8) en relación a los umbrales de Q y los marcadores utilizados.



**Figura 1.** Representación de la matriz Q, para K=9, obtenida a partir de 50 SSR y 1000 SNP en conjunto. \* indica que la línea posee un valor Q mayor o igual a 0,7.

## Referencias

Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. (2007) *TASSEL: Software for association mapping of complex traits in diverse samples*. *Bioinformatics* 23:2633-2635.

Cook, J.P.; McMullen, M.D.; Holland, J.B.; Tian, F.; Bradbury, P.; Ross-Ibarra, J.; Buckler, E.; Flint-Garcia, S.A. (2012) *Genetic Architecture of Maize Kernel Composition in the Nested*

*Association Mapping and Inbred Association Panels*. Plant Physiology 158 (2): 824-834;  
DOI: 10.1104/pp.111.185033

Evanno, G.; Regnaut, S.; Goudet, J. (2005) *Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study*. Molecular Ecology 14(8), 2611-2620.

Ganal, M.W.;Durstewitz, G.;Polley, A.;Bérard, A.;Buckler, E.S.;et al.(2011) *A Large Maize (Zea mays L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome*. PLoS ONE 6(12): e28334. <https://doi.org/10.1371/journal.pone.0028334>

Hamblin, MT; Warburton, ML; Buckler, ES. (2007) *Empirical Comparison of Simple Sequence Repeats and Single Nucleotide Polymorphisms in Assessment of Maize Diversity and Relatedness*. PLoS ONE, 2(12), e1367. <https://doi.org/10.1371/journal.pone.0001367>

Jakobsson, M.; Rosenberg, N.A. (2007) *CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure*. Bioinformatics 23(14), 1801-1806.

Olmos,S.; Delucchi, C.; Ravera,M.; Negri, M.E.; Mandolino, C.; Eyhérbide, G.H.(2014) *Genetic relatedness and population structure within the public Argentinean collection of maize inbred lines*. Maydica 59(1), 16-31.

Pritchard, JK; Stephens, M; Donnelly, P. (2000). *Inference of Population Structure Using MultilocusGenotype Data*. Genetics, 155, 945–959.