



MAÍZIFICANDO CONCIENCIA

XII CONGRESO NACIONAL DE MAÍZ

Eje

Genética y mejoramiento

8, 9 y 10 de Noviembre
Pergamino, BA
UNNOBA



Secretaría de Agricultura,
Ganadería y Pesca
Ministerio de Economía
Argentina

2022





DIVERSIDAD GENÉTICA Y ESTRUCTURA POBLACIONAL EN MAÍZ TEMPLADO ARGENTINO A TRAVÉS DE MÉTODOS DE FILOGENÉTICA COMPUTACIONAL

Perdomo S. I.¹; Baricalla A.^{1,2}; Iglesias J.^{1,3}

¹ Universidad Nacional del Noroeste de la Provincia de Buenos Aires.

² Centro de Investigación y Transferencia del Noroeste de la Provincia de Buenos Aires (CITNOBA) UNNOBA-CONICET.

³ Instituto Nacional de Tecnología Agropecuaria- EEA Pergamino.

iglesias.juliana@inta.gob.ar

GENETIC DIVERSITY AND POPULATION STRUCTURE STUDY OF A PUBLIC PANEL OF TEMPERATE MAIZE INBREDS LINES FROM ARGENTINA

Abstract

In maize, genetic diversity and population structure assessment by using new sequencing technologies is crucial since inbreeding reduces not only the fitness, but also the production-associated traits. The present work aims to evaluate the genetic diversity and genetic structure of a panel of 191 inbred lines from the Maize Breeding Program at EEA INTA Pergamino with the widely-used Illumina® MaizeSNP50 BeadChip. To that end, two different Bayesian methodologies were investigated: STRUCTURE and fastSTRUCTURE altogether with NJ and UPGMA clustering methods. t-SNE was used for data visualization. Clustering results were validated by comparison with pedigree and heterotic pattern information. Population structure analysis provided an estimated $K=10$ groups (subpopulations) constituted by genotypes derived from families of syn. hybrid Pioneer, syn. BS13P, syn. 34 from INTA Leales, from AX882(F2) and M11 (F2), lines derived from P465 among some other mixed groups. F_{st} values from the 10-subpopulations ranged from 0.43-0.75, with an overall $F_{st}=0.6$ indicating a medium to a high amount of genetic differentiation. Therefore, inbreds from each subpopulation can be crossed and tested for heterosis. Overall, these results showed substantial genetic variability that could be exploited to obtain new maize varieties to meet future challenges. Important information is also provided for future genome wide association studies and marker-assisted selection.

Palabras claves

Mejoramiento, variabilidad, subpoblaciones, F_{st}

Keywords

Breeding, variability, subpopulations, F_{st}



Introducción

Durante un gran periodo de la historia humana, entre 5.000 y 10.000 años atrás, varios cultivos fueron domesticados, dentro de los cuales se encuentra el maíz (*Zea mays L.*). Considerando que es un cultivo que a lo largo de los años se ha ido adaptando a un gran rango ambiental, y que también es tolerante a muchos de los estreses bióticos y abióticos, su producción a nivel nacional y mundial ha ido en auge.

Un punto de partida para obtener un buen progreso en cualquier programa de mejora de cultivos es la información sobre la diversidad genética y la estructura poblacional del germoplasma base. Debido a esto, muchos investigadores han resaltado la importancia y la necesidad de una evaluación precisa de la diversidad genética en los programas de mejora.

Los conocimientos e información sobre la caracterización molecular de la diversidad de la población, su estructura y las relaciones filogenéticas entre los materiales de mejoramiento, permiten comprender cómo utilizar el germoplasma escogido, seleccionar las líneas parentales y asignar grupos heteróticos permitiendo distinguir a los individuos que presenten un mayor vigor híbrido con respecto a la media de sus padres.

Con el advenimiento de las técnicas modernas de biología molecular y los métodos que permiten medir el polimorfismo genético a nivel de ADN, se logran realizar los análisis de caracterización molecular y evaluación de diversidad que proporcionan a los mejoradores información detallada, ayudando a la selección de parentales para iniciar un programa de mejora.

El análisis de la estructura poblacional conlleva a considerar que la estructura consiste en dos partes distintas pero interrelacionadas: la estructura demográfica y la estructura genética. Aquello que en realidad caracteriza a la estructura genética de una población es su conjunto de genes, es decir, las frecuencias génicas y genotípicas. En maíz, la determinación de este tipo de estructura es importante ya que, la endogamia reduce el fitness y los rasgos asociados a la producción en las diferentes poblaciones. Por este motivo, es conveniente calcular diferentes parámetros como por ejemplo el coeficiente de endogamia entre poblaciones (F_{st}).

En este trabajo se espera conocer la estructura poblacional dentro de un panel de líneas frecuentemente utilizadas para el mejoramiento en la Estación Experimental Agropecuaria INTA Pergamino ensayando diversas metodologías derivadas de la filogenética computacional. En este sentido, se plantea evaluar los marcadores SNPs como fuente de información genotípica, para realizar un análisis de la estructura-poblacional mediante diferentes métodos y poder luego considerarlo en análisis de mapeo de asociación posteriores. Adicionalmente se podrá contar con metodología de análisis disponible, testada y optimizada para su uso posterior en este y otros paneles de genotipos para mejoramiento.



Materiales y métodos

Material vegetal y filtrado de datos genotípicos.

El material vegetal estuvo constituido por un panel de 191 líneas endocriadas provenientes de más de 45 orígenes distintos, incluyendo materiales de endosperma flint y dentado del programa de mejoramiento genético de maíz de la EEA INTA Pergamino (Olmos *et al.*, 2014 y otras fuentes). La información genotípica fue obtenida a partir del chip MaizeSNP50 de Illumina que consta de 56.658 loci de SNPs utilizando B73v5 como línea de referencia. Se eliminaron los marcadores y las líneas con más del 20% de datos faltantes y de heterocigosidad. Se determinaron marcadores monomórficos y polimórficos, luego se eliminaron marcadores con frecuencia alélica $< 5\%$.

Inferencia de estructuración.

La inferencia de la estructura poblacional se realizó mediante metodología bayesiana. Se utilizaron los softwares STRUCTURE 2.3.4 (Raj *et al.*, 2014) en primer lugar a partir de un servidor de alta efectividad, y luego fastSTRUCTURE mediante el servidor CyVerse (Merchant *et al.*, 2016). Principalmente, la manera de ejecutarse STRUCTURE es ajustando un modelo para cada valor posible de K que determina la cantidad de subpoblaciones presentes (para este trabajo $K=1$ a 20). Como existe cierta variación en el análisis se realizaron tres réplicas con el objetivo de disminuir el error durante el proceso. A su vez se realizaron 100.000 iteraciones de las cuales 10.000 se utilizaron para *burn-in* y para el procedimiento MCMC. Finalmente, el software brinda como resultado qué número efectivo de grupos existe en los datos, representado por el mayor valor de delta K. Todo este cálculo es realizado por STRUCTURE Harvester, a partir de las matrices, que va generando STRUCTURE. Un resultado importante de STRUCTURE es que también brinda los valores de Fst de las diferentes subpoblaciones.

En el caso de fastSTRUCTURE, es un software que tiene un algoritmo que no permite calcular el número de grupos de manera apropiada. Por este motivo, se debió considerar la ejecución previa de STRUCTURE Harvester para obtener un número óptimo de subpoblaciones. De esta forma, STRUCTURE Harvester, mediante el método de Evanno (Evanno *et al.*, 2005), procesa los resultados estructurales, genera archivos para su uso posterior con CLUMPP y brinda el número óptimo de grupos.

Análisis de distancia.

La obtención de las matrices de distancia y *kinship* se llevó a cabo mediante TASSEL (Bradbury *et al.*, 2007), a partir de los datos filtrados como se especificó anteriormente, así como también la realización de los árboles de distancia a partir de los métodos *Neighbor Joining* (NJ) y UPGMA. Posteriormente, se utilizó ITOL (Letunic & Bork, 2016) para observar los árboles de manera gráfica y comparativa.

Reducción dimensional.

Este paso se realizó mediante el método t-SNE, a partir de las matrices de Distancia, PCA y *Kinship* obtenidas mediante TASSEL. Luego, sus respectivos gráficos de reducción dimensional, fueron realizados en R mediante el paquete ggplot2 (Wickham *et al.*, 2022). Se utilizaron cuatro valores de perplejidad diferentes (1, 3, 15 y 30), parámetro que permite realizar una distribución de probabilidad representando similitudes entre los vecinos más cercanos. Para este proceso se realizaron 200.000 iteraciones.



Resultados y discusión

Se realizó un filtrado de los datos genotípicos, proceso que llevó a la eliminación de 11.290 SNPs, dando como resultado 44.820 SNPs que fueron utilizados en el análisis posterior.

Como se observa en la Figura 1, la cifra de delta K de mayor valor obtenido (21.609,641) permitió inferir la presencia de 10 grupos definidos y estructurados con valores de F_{st} superiores a 0,5 (Figura 1). En este sentido y corroborando la información obtenida por STRUCTURE ($K=10$) el método de Evanno también brindó un valor probable de subpoblaciones igual a 10 para los 191 genotipos evaluados en este trabajo. Esto demuestra una elevada estructuración poblacional debido a la existencia de una diferenciación genética significativa entre las subpoblaciones. En este caso los individuos presentan fondos genéticos similares dentro de una misma subpoblación y una marcada diferenciación con individuos de otras subpoblaciones.

Más en detalle con respecto a la estructuración, se observaron agrupamientos equivalentes con las dos metodologías bayesianas testeadas ya que el patrón de estructura de las subpoblaciones en STRUCTURE se mantuvo también en fastSTRUCTURE (no mostrado). Esto permitió realizar una comparación sobre la utilización de ambos algoritmos en términos de reducción de tiempo de cálculo y de recursos computacionales. Recordemos que para establecer el número de subpoblaciones posibles existentes con fastSTRUCTURE se debió implementar el método de Evanno de STRUCTURE Harvester. De esta manera se deduce que, si bien fastSTRUCTURE permite agilizar el tiempo de cálculo siendo un software más rápido y sencillo de implementar, no permite definir el número óptimo de subpoblaciones presentes en el set de datos para lo cual hay que utilizar alguna otra herramienta de soporte como es el caso de STRUCTURE Harvester y Evanno.

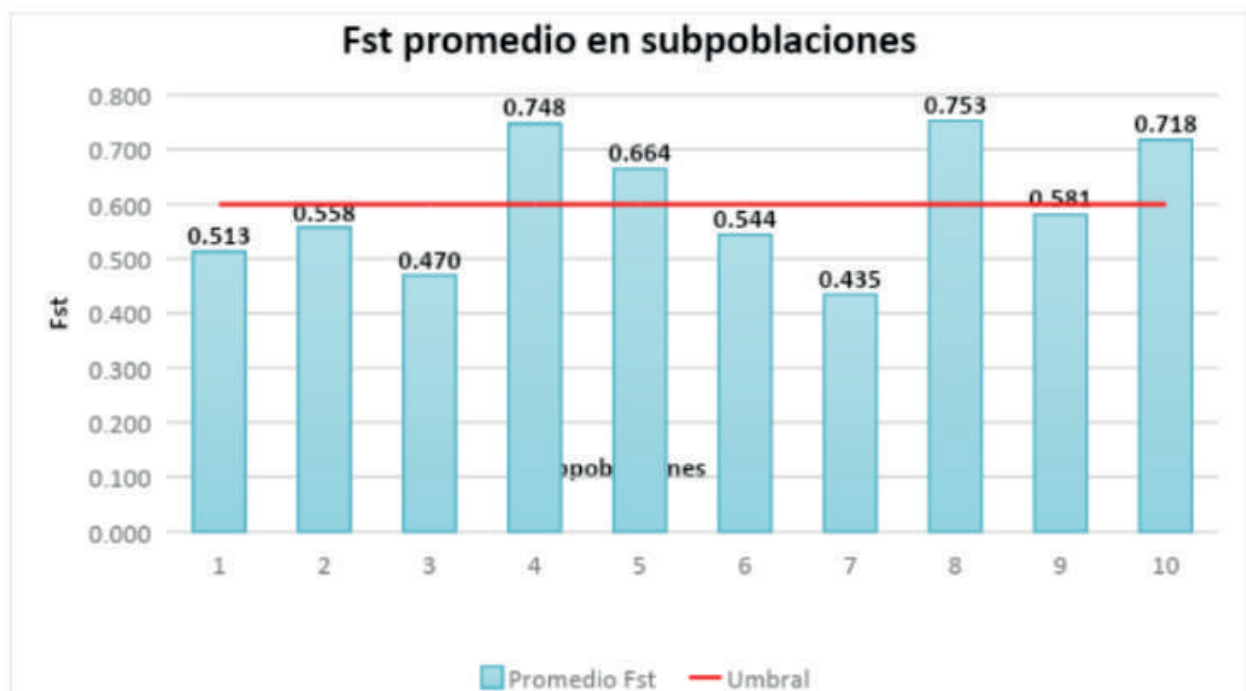


Figura 1. Gráfico de los promedios de F_{st} para cada subpoblación obtenida, a partir del estudio de un panel de 191 líneas endocriadas de maíz del Programa de Mejoramiento Genético de Maíz de la EEA INTA Pergamino.

Posteriormente, surgió la posibilidad de establecer un análisis comparativo de los grupos obtenidos por inferencia bayesiana y los obtenidos mediante árboles de distancia, observando un patrón de agrupamiento diferencial. Esto se debió a que, en el caso de los árboles se obtuvieron líneas “mal clasificadas” en subpoblaciones con cuyos individuos no comparten fondos genéticos (Figura 2). De todas formas, estas líneas corresponden a genotipos que según STRUCTURE o fastSTRUCTURE/Evanno presentan sus proporciones de pertenencia particionada entre varias subpoblaciones lo cual, puede dar un indicio de la singularidad de la ubicación de agrupamiento mediante los arboles de distancia. De esta forma, un 10% de las líneas se agruparon de manera inexacta a partir del método NJ (Figura 2) y un 12% a partir del método UPGMA (no mostrado).

En la figura 2, se pueden observar genotipos o líneas que no se agruparon por distancia en las subpoblaciones a las que corresponden considerando la estructuración por inferencia bayesiana. Comenzando en sentido antihorario y desde la parte inferior izquierda, tenemos el caso de L006 y L008. En cuanto a distancias estas líneas se agruparon de manera invertida entre las subpoblaciones 2 y 8. L006 según inferencia bayesiana presenta mayor pertenencia en la subpoblación 2 y L008 en la subpoblación 8. Ambas pertenecen al grupo de líneas derivadas de Sintética BS13P.

Las líneas L202, L229 y L247 se agruparon por distancia genética junto a la subpoblación 3 (formada mayormente por líneas derivadas sintéticas de híbrido Pioneer) siendo que poseen diversos orígenes: L202 (BS13P, Grupo 1), L229 (SR en (P465×D)F2, Grupo 2) y L247 ((N28×B73)×(N28×199)]F2, Grupo 2).

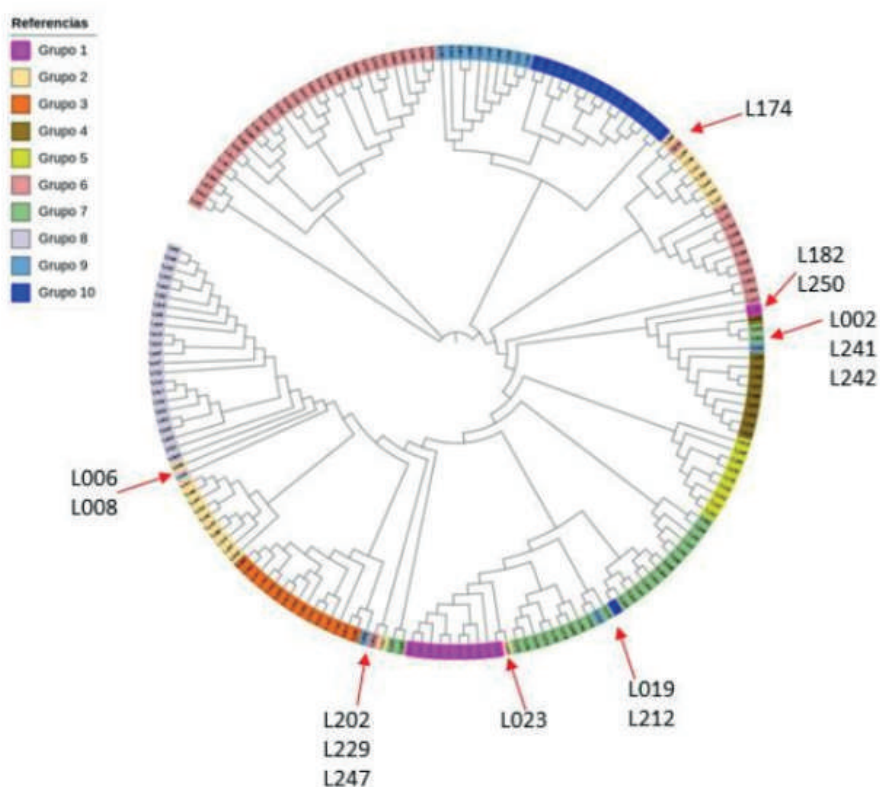


Figura 2. Representación gráfica del árbol de distancia NJ a partir de un panel de 191 líneas endocriadas del Programa de Mejoramiento Genético de Maíz de la EEA INTA Pergamino. Vista en colores de las 10 subpoblaciones (descriptas en Tabla 1). Las flechas señalan casos que se discuten en el texto.

La línea L023 derivada de sintética BS13P, pertenece por inferencia bayesiana al grupo 2 y se agrupó por distancia con líneas derivadas de híbrido AX888F2 y M11(F2) (Grupo 1). Lo mismo ocurre con la línea L212 que se agrupó con derivadas de este híbrido, pero en este caso, esta línea pertenece según inferencia bayesiana a la familia de líneas derivadas de P465xD.

La línea L019, con origen a partir de la sintética BS13P, en el árbol se ubica formando parte del grupo 7 correspondientes a la familia de híbrido AX888. En realidad, esta línea tiene mayor pertenencia, según inferencia bayesiana en la subpoblación 10. En este grupo también aparece L212 que por inferencia bayesiana pertenece al grupo 9.

Las líneas L002 (derivada de sintética BS13P), L241 (derivada de población base amplia de EEUU) y L242 (derivada de Suwan) en el árbol se agruparon en el grupo 4 donde se encuentran derivadas de sintética 34 de INTA Leales. Este también fue el caso para L182. Por último L174 (Grupo 6), aparece entremezclada con miembros del grupo 2.

Finalmente se definió la estructura poblacional de manera aún más simple, en un mapa de menos dimensiones a través del algoritmo t-SNE. De esta forma, se obtuvo una mayor visualización y comprensión sobre las 10 subpoblaciones obtenidas (Figura 3).

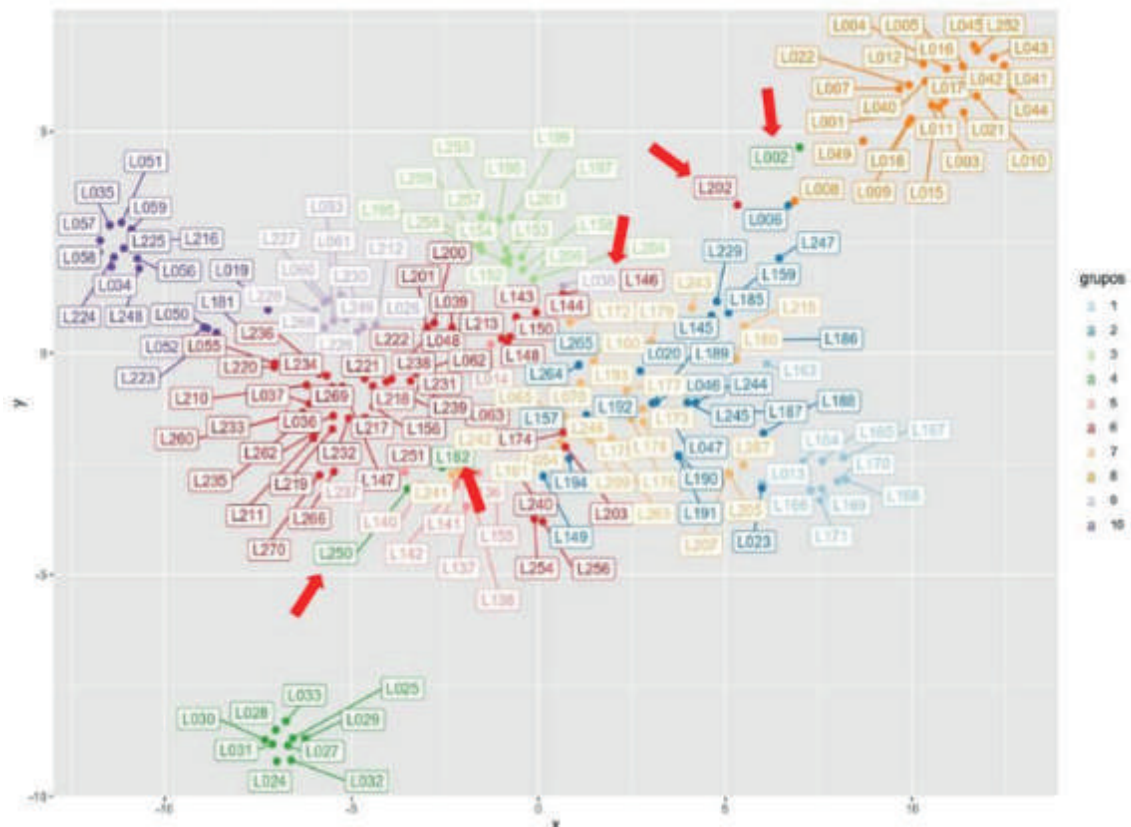


Figura 3. Representación gráfica de los agrupamientos en dos dimensiones realizado con t-SNE a partir de la matriz de PCA de un panel de 191 líneas endocriadas de maíz del Programa de Mejoramiento Genético de Maíz de la EEA INTA Pergamino. Perplejidad=30, 200.000 iteraciones. Las flechas rojas señalan casos que se explican en el texto. Vista de las 10 subpoblaciones (descriptas en Tabla 1).

Cotejando los resultados obtenidos con los datos previos de pedigrí y grupos heteróticos a los que pertenecían los genotipos, se validó el grado de pertenencia a los diferentes grupos resultantes y su distribución espacial diferenciada. Con respecto a los agrupamientos, al observar el gráfico obtenido mediante t-SNE (Figura 3) se puede destacar que las subpoblaciones que demostraron una mayor solidez en los agrupamientos (según la numeración de estructuración basada en la inferencia bayesiana) son las subpoblaciones 1, 3, 4, 8, 9 y 10. Estas subpoblaciones mostraron diferenciarse del resto y destacarse en el espacio al momento de reducir las dimensiones del análisis. En el caso de las subpoblaciones 4, 8, 9 y 10 se sostiene que son grupos de líneas muy definidas ya que sus valores de F_{st} se encuentran por encima de 0,5. Asimismo, las subpoblaciones restantes (2, 5, 6, 7) se presentaron más dispersas en el t-SNE e incluso más cercanas entre sí con valores de $F_{st} > 0,5$ para las subpoblaciones 2, 5 y 6. Por el contrario, las subpoblaciones 3 y 7 fueron las de los menores valores de $F_{st} < 0,5$ y por ende las menos estructuradas de todas. En la tabla 1 se pueden observar las familias que conforman cada subpoblación y también la cantidad de individuos que se encuentran dentro de cada una.

Grupo	Orígenes*	N° de genotipos
1 ($F_{st}=0,513$)	(M11) F2 Híbrido AX888 F2	10
2 ($F_{st}=0,558$)	Sintética líneas coloradas duras BSGA (DK638) (DK752×B73) F2 – Fam. DK [(N28×B73)×(N28×199)] Compuesto II (QUICHUA)	24
3 ($F_{st}=0,470$)	Sintética Semidentada Semiprecoz Sintética de híbridos Pioneer Sintética de híbrido Pioneer- Fam. LP299-2 Flint×Dentada F2 - Fam .Arg-Canadá	15
4 ($F_{st}=0,748$)	Sintética 34 de INTA Leales (SR en Perg) BS29 Suwan	12
5 ($F_{st}=0,664$)	Sintética BS13P (LP581×LP562) F2 Sintética Semidentada Semiprecoz R49022×M370	9
6 ($F_{st}=0,544$)	SR en (L2565 × D)F2 [(5842×LP125)×(2×P1338)]F2 Compuesto Dentado 3:3:B Sintética de híbridos (ACA 2000)F2 LP 923 Sintética Semidentada (Sin Cuarentín) × (L1571 (Sin A)) Población flint colorada	45
7 ($F_{st}=0,435$)	(FESTIVAL)F2 Flint×Dentada F2 – Fam. Arg-Canadá Población base amplia-EE. UU. Suwan (AX888) F2 (Cargill 280) F2 (Sintética Líneas Arg.) × (Líneas Canadá) Compuesto I	25
8 ($F_{st}=0,753$)	Sintética BS13P (LP561×BS13)BS13 Compuesto BS13	24
9 ($F_{st}=0,581$)	SR en (P465×D)F2 [(P465×D)Fn* × ZN6]F2 – Fam. P465 (LP662×LP611)F2 Landrace Argentina Sintética 34 de INTA Leales (SR en Pergamino) LP611×LP613 Población 582W	11

10 (Fst=0,718)	LP611×LP199 LP521×L28 L196×LP122 BS29 Compuesto II Compuesto Argentino Caribe (L3178×L196)F2 – Fam. 122 (LP199×L3178)F2 – Fam. 122 (L28×LP521)F2-Sint. Colorada Dura	16
---------------------------------	--	-----------

Tabla 1: Descripción de los orígenes de las líneas que conforman cada subpoblación según K=10 obtenido por inferencia bayesiana. *Basado en Olmos et al., 2014.

Los agrupamientos obtenidos en este trabajo se basaron en los marcadores SNPs los cuales agruparon las líneas endocriadas en función de su ascendencia y filogenia. Sin embargo, la agrupación inesperada de algunas líneas estaría indicando que éstas no son líneas extraídas de la misma población de origen o no presentan necesariamente el mismo historial de selección como se pensaba. En este sentido las líneas L002 (Sintética BS13P), L038, L182 (BS29), L202 (derivada L10612×B14) y L250 (Suwan) mostraron patrones de agrupamiento llamativos en t-SNE y por inferencia bayesiana. Algunas de ellas como es el caso de la L002 y L202 también demostraron agrupamientos llamativos según los cálculos de distancia. Esto permite analizar más detalladamente estos genotipos y su patrón de agrupamiento.

La línea L002 (Sintética BS13P) (Figura 3), por ejemplo, se agrupa con una proporción de pertenencia de 0,46 en la subpoblación 4 (derivadas de sintética 34 de INTA Leales) y una proporción de pertenencia de 0,42 en la subpoblación 8 (derivadas de sintética BS13P). Así mismo el agrupamiento de t-SNE la ubica espacialmente cercana con líneas provenientes de su mismo bulk próxima a la subpoblación 8 (derivadas de Sintética BS13P). Líneas con información de pedigrí similares se repartieron entre las subpoblaciones 2 y 8, siendo esta línea L002 la única de su tipo que se ubica en la subpoblación 4. Una subpoblación con un notable valor de Fst.

La línea L038 figura particionada entre varias subpoblaciones con valores de estructuración menores a 0,1; por inferencia bayesiana se ubica en la subpoblación 9 (derivadas de P465). No obstante, según su información de pedigrí (derivada de Compuesto I) podría ubicarse en las subpoblaciones 6 y/o 7. Por t-SNE se ubica espacialmente cercana a estas subpoblaciones. En el árbol de distancia NJ se ubicó dentro del grupo 7.

En el caso de la línea L202 (derivada L10612×B14) que es agrupada en la subpoblación 6 según inferencia bayesiana, t-SNE la ubicó cercana a líneas provenientes de las familias sintéticas coloradas duras BSGA y los híbridos AX888, subpoblaciones 2, 3 y 8 (ver también figura 2).

Por último, L182 (BS29) y L250 (Suwan) poseen patrones de agrupamiento similares; subpoblación 4 según inferencia bayesiana, pero con bajos valores de estructuración. Indagando en su información de pedigrí (BS29 y Suwan-1 respectivamente) se observa que son cercanas a las subpoblaciones 5 y 7 y así queda evidenciado por la ubicación espacial que le asigna el t-SNE.

Es decir que en estos casos donde los valores de estructuración no fueron suficientemente determinantes para la asignación de pertenencia a un grupo, la herramienta t-SNE (que permite sumar más información) fue de gran ayuda y permitió visualizar en dos dimensiones la cercanía entre genotipos y grupos.

En el caso de las familias Sintética de híbrido Pioneer- Fam. LP299-2 o derivadas, junto con líneas de las Sintética Semidentada se agruparon de manera conjunta y clara, así como líneas provenientes de la Sintética 34 de INTA Leales. También las líneas derivadas a partir de la familia de los híbridos AX888F2 y M11 compartieron un grupo. En el caso de las líneas derivadas de la sintética BS13P, demostraron agruparse definitivamente (salvo la excepción mencionada con L002).

Dos subpoblaciones bien destacadas del resto y que son bastante heterogéneas, en cuanto a sus orígenes, son la 9 y 10. En la subpoblación 9 se agruparon líneas derivadas de P465, derivadas del cruzamiento P465×D, derivadas de la Población 582W, y de los cruzamientos LP611×LP613 y LP662×LP611. Mientras tanto, en la subpoblación 10 se encuentran líneas principalmente derivadas de la línea LP122, LP611, LP521 y derivadas de la familia Compuesto II.

En suma, este estudio demuestra la posibilidad de la utilización de diferentes métodos, algoritmos y software que permiten una aproximación a la estructuración verdadera presente en los datos.

El t-SNE, un algoritmo de machine learning que permite la reducción de datos altamente dimensionales, no solo permitió una óptima visualización de los datos, sino que también demostró ser una poderosa herramienta en aquellos casos en los que la estructuración no era clara. Es decir, cuando STRUCTURE/fastSTRUCTURE-Evanno arrojaba valores de estructuración o pertenencia de un genotipo, repartidos entre varias subpoblaciones, t-SNE colaboró con el resultado final de pertenencia de cada línea.

Con respecto a la evaluación del tiempo de cálculo, el desempeño de fastSTRUCTURE fue notablemente más óptimo en el uso del tiempo y recursos computacionales que STRUCTURE (30 minutos vs 3 días en un servidor informático). No obstante, es importante mencionar que el algoritmo fastSTRUCTURE no asigna un valor de delta K óptimo para los agrupamientos. Por lo que fue necesario usar el método de Evanno a través del software STRUCTURE harvester o sino tomar el valor delta K de STRUCTURE para establecer el número óptimo de subpoblaciones.

Nuevamente se destaca que STRUCTURE es computacionalmente demandante, mientras que el método de Evanno en combinación con STRUCTURE harvester y fastSTRUCTURE no lo es.

Conclusión

En este trabajo se utilizó la información de un chip de 44820 SNPs para explorar la diversidad genética y la estructura poblacional de un panel de 191 líneas endocriadas del Programa de Mejoramiento Genético de Maíz de la EEA INTA Pergamino.

En base a los resultados aquí presentados se puede concluir que el panel bajo estudio, es un panel diverso y con altos niveles de estructuración y diferenciación genética. Este nivel de diversidad genética podría ser la base para desarrollar nuevos genotipos de maíz con características deseables para responder a las demandas actuales y los cambiantes escenarios agroclimáticos.

Por otra parte, el conocimiento de la diversidad genética y la estructura poblacional es de suma importancia para llevar adelante estudios utilizando selección genómica, selección asistida por marcadores y estudios de mapeo por asociación. Como es sabido, la presencia de estructuración en

una población de mapeo puede dar lugar a falsos positivos o falsas asociaciones marcador-carácter de interés. En este sentido, por medio del análisis de inferencia aquí presentado, se obtuvieron datos que ya fueron incorporados en modelos estadísticos para estudios de asociación de genoma completo para resistencia a enfermedades (ver este congreso) considerando el ajuste de la variabilidad debida a la estructuración genética presente en el panel de líneas.

Finalmente, se puede destacar como conclusión final de este trabajo que, gracias a las diferentes herramientas disponibles, fue posible trabajar con una gran cantidad de datos, considerando el número total de SNPs utilizados. Se logró disminuir errores de cálculo, incorporando mayor relevancia y precisión en los análisis. El algoritmo variacional que incorpora fastSTRUCTURE es casi dos órdenes de magnitud más rápido que STRUCTURE, no requiere de un servidor informático robusto y no sacrifica precisión o potencia, sin embargo, presenta como dificultad el hecho de que hasta el momento no se ha logrado propiamente establecer el número óptimo de subpoblaciones presentes en los datos sin involucrar algún otro método de análisis.

De todas formas, siendo fastSTRUCTURE un software rápido de ejecutar y dinámico, permite consolidar los agrupamientos obtenidos por STRUCTURE. A su vez, la posterior utilización de t-SNE brinda una visualización de los datos a nivel espacial, lo cual, en varios casos ayuda al entendimiento de la distribución de los datos que se encuentran en condición dudosas al momento del análisis. Es decir que t-SNE es una herramienta de gran utilidad ya que, ayuda a dilucidar los resultados obtenidos con facilidad, de manera práctica y con más detalle.

Apoyo financiero

Programa de Promoción de la Investigación Científica de la UNNOBA- EXP 2691/2019 Resol. N° 1002/2019; Proyecto INTA 2019-PE-E6-I114-072.

Agradecimientos

Los autores agradecen al grupo de Mejoramiento de Maíz de la EEA INTA Pergamino por la información provista y por su colaboración en la realización del presente trabajo. Este trabajo forma parte de la tesina de grado de S.P.

Referencias bibliográficas

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. *Bioinformatics* 23(19):2633–5.

Earl DA. 2012. *Conservation. Genetic. Resources.* 4(2):359–361.

Evanno G, Regnaut S, Goudet J. 2005. *Molecular. Ecology.* 14(8):2611–2620.

Letunic I., Bork P. 2016. *Nucleic Acids Research* 44(W):1 W242–W245.

Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, et al. 2016. *PLoS Biology* 14(1): e1002342.

Olmos SE, Delucchi C, Ravera M, Negri ME, Mandolino C, Eyhérabide GH. 2014. *Maydica* 59(1): 16–31.

Raj A, Stephens M, Pritchard JK. 2014. *Genetics* 197 (2): 573–589.

Xu C, Ren Y, Jian Y, et al. 2017. *Molecular Breeding* 37 (3).

Wickham H, Chang W, Henry L, Lin Pedersen T. 2022. <https://ggplot2.tidyverse.org>