

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

SUBDIRECCIÓN DE POSTGRADO



ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE MAÍCES NATIVOS DEL  
SURESTE DE COAHUILA, MÉXICO POR EL MÉTODO DArTseq

TESIS


Que presenta Elías Méndez Hernández  
Como requisito parcial para obtener el Grado de:

MAESTRÍA EN CIENCIAS EN FITOMEJORAMIENTO

ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE MAICES NATIVOS DEL SURESTE  
DE OAHUQUILA, MÉXICO POR EL MÉTODO DE RAPD

TESIS

Elaborada por ELIAS MENDEZ HERNANDEZ como requisito parcial para obtener el  
grado de MAESTRIA EN CIENCIAS EN TIPOLOGIA AMBIENTAL con la supervisión y  
aprobación del Comité Particular de Asesoría

  
Dr. Miguel Ángel Pérez Quiroga  
Asesor Principal

  
Dr. Enríque Ríos Sánchez  
Asesor

  
Dra. Norma Angélica Ruiz Torres  
Asesor

  
Dr. César Daniel Petrucci Espinoza  
Asesor

  
Dr. Marcelino Gutiérrez de la Fuente  
Subdirector de Posgrado  
UNAHU

Saltillo, Coahuila

Julio 2020

## AGRADECIMIENTOS

A **Dios** principalmente por prestarme la vida y darme la oportunidad de lograr terminar una de mis metas en la vida.

A la **Universidad Autónoma Agraria Antonio Narro** (UAAAN) por brindarme todas las facilidades para continuar con mi formación profesional.

Al **Consejo Nacional de Ciencia y Tecnología** (CONACYT) por brindarme la oportunidad y el apoyo financiero para realizar mis estudios de postgrado y la presente investigación.

A **MasAgro**, financiado por SADER, en especial a la unidad de Desarrollo de Capacidades dentro del componente MasAgro Biodiversidad, por el apoyo financiero para realizar parte de la presente investigación con la producción de datos genotípicos.

Al **Centro de Internacional de Mejoramiento de Maíz y Trigo** (CIMMYT), por abrirme sus puertas para la realización de una estancia académica y contribución en la capacitación y colaboración en aplicar las técnicas de Laboratorio básicas para el desarrollo de esta investigación.

Al **Dr. Miguel Ángel Pérez Rodríguez**, por ser una excelente persona, por su valiosa orientación, paciencia y tiempo dedicado en la realización de esta investigación, y por su amistad brindada.

Al **Dr. Froylán Rincón Sánchez**, por su amabilidad, apoyo y sus valiosas contribuciones aportadas.

A la **Dra. Norma Angélica Ruíz Torres**, por su apoyo en cada una de las revisiones y valiosas sugerencias aportadas para la realización del presente trabajo.

Al **Dr. César Daniel Petroli**, como asesor externo, por su apoyo incondicional en esta investigación y su amistad brindada. De igual manera al equipo de trabajo en SAGA para la realización y obtención de los datos genotípicos.

A todas las personas que de una u otra manera contribuyeron en la realización de este trabajo.

## **DEDICATORIA**

### **A MIS PADRES:**

**Nicolás Méndez Pérez y María Hernández Pérez**, por creer en mí, apoyarme en cada etapa de mi vida sabiendo que jamás existirá una forma de agradecerles una vida de lucha, sacrificio y esfuerzo constante, y que el logro mío es logro suyo, Gracias por sus consejos y ánimos.

### **A MIS HERMANOS:**

**Mary, Gloria, Abi y Fer**, por todo su amor y su apoyo incondicional.

### **A MI HIJA:**

**Dulce Arisbeth Itzayana Méndez Andrade**, quien es mi fortaleza y fuente de amor e inspiración.

## Índice de Contenido

	<b>Página</b>
Lista de Cuadros.....	viii
Lista de Figuras .....	ix
RESUMEN.....	x
ABSTRACT .....	xiii
INTRODUCCIÓN .....	15
OBJETIVOS E HIPÓTESIS .....	17
Objetivo general .....	17
Objetivos específicos:.....	17
Hipótesis .....	17
REVISIÓN DE LITERATURA.....	18
Importancia del maíz en México .....	18
Importancia del maíz en el sureste de Coahuila .....	18
Genética de poblaciones .....	19
Diversidad genética .....	19
Descripción de la diversidad genética .....	20
Polimorfismo .....	21
Contenido de información polimórfica (PIC).....	21
Heterocigosidad observada .....	22
Heterocigosidad esperada .....	22
Número de alelos efectivos.....	22
Alelos raros .....	23
Marcadores moleculares .....	23
Isoenzimas .....	24
Polimorfismo de Longitud de Fragmentos de Restricción (RFLP) .....	24
Polimorfismo de ADN Amplificados al Azar (RAPD) .....	25
Polimorfismo de Longitud de Fragmentos Amplificados (AFLP) .....	26
Microsatélites o secuencias simples repetidas (SSR) .....	26
Polimorfismos de un solo nucleótido (SNP).....	27
Microarreglos - Diversity Arrays Technology (DArT) .....	27
Marcadores moleculares en el estudio de diversidad genética de maíz .....	28

Genotipificación por secuenciación (GBS).....	30
Plataforma DArTseq.....	31
<b>MATERIALES Y MÉTODOS</b> .....	<b>35</b>
Material genético.....	35
Proceso de genotipificación.....	36
Obtención de muestras de tejido.....	36
Liofilización.....	37
Molienda .....	37
Extracción del ADN genómico.....	37
Cuantificación y calidad del ADN genómico .....	38
Reducción de la complejidad genómica .....	38
Amplificación de fragmentos.....	40
Control de calidad de los productos de PCR .....	41
Purificación y cuantificación .....	41
Genotipado por secuenciación .....	42
Validación de datos genéticos .....	43
Identificación de marcadores .....	43
Tamizaje de datos .....	44
Análisis de diversidad genética .....	45
Heterocigosidad .....	45
Número de alelos efectivos.....	45
Índice de Shannon.....	46
Número de alelos raros .....	46
Coeficientes de divergencia genética .....	46
Ordenamiento de poblaciones nativas de maíz .....	47
Relación entre poblaciones de maíz .....	47
<b>RESULTADOS Y DISCUSIÓN</b> .....	<b>48</b>
Análisis exploratorio de marcadores SNP.....	48
Índices de diversidad genética.....	51
Coeficientes de divergencia genética .....	55
Ordenamiento de poblaciones nativas.....	56
Relación entre poblaciones nativas .....	62

CONCLUSIONES .....68  
REFERENCIAS.....69  
ANEXO.....77

## Lista de Cuadros

	<b>Página</b>
<b>Cuadro 1.</b> Clasificación racial y frecuencia de las 76 poblaciones de maíz nativo en base a la altitud del sitio de colecta, Coahuila, México.....	35
<b>Cuadro 2.</b> Soluciones y sus concentraciones para la digestión del ADN genómico.	40
<b>Cuadro 3.</b> Soluciones y sus concentraciones para la ligación del ADN genómico. ..	40
<b>Cuadro 4.</b> Componentes y concentraciones para la PCR. ....	41
<b>Cuadro 5.</b> Índices de diversidad genética de las 76 poblaciones de maíz secuenciados. ....	51
<b>Cuadro 6.</b> Solución amortiguadora con CTAB para extracción. <sup>1</sup> .....	77
<b>Cuadro 7.</b> Lavado 1: 76% de EtOH, 0.2 M NaOAc. ....	77
<b>Cuadro 8.</b> Lavado 2: 76% EtOH, 10 mM NH <sub>4</sub> OAc.....	77
<b>Cuadro 9.</b> Cloroformo: Octanol, 24:1.....	77
<b>Cuadro 10.</b> Preparación de gel de agarosa (0.8 %) para muestras de ADN. ....	78
<b>Cuadro 11.</b> Preparación de gel de agarosa (01.2 %) para productos de PCR.....	78



## Lista de Figuras

## Página

- Figura 1.** Preparación de bibliotecas mediante la metodología de Genotipado por Secuenciación (GBS) (Figura tomada de Elshire *et al.*, 2011). .....30
- Figura 2.** Plataforma de genotipificación de alto rendimiento DArTseq (Figura tomada de Kilian *et al.*, 2012). .....32
- Figura 3.** Representación esquemática del método de reducción de la complejidad del genoma para generar representaciones genómicas (Figura tomada de Singh, 2015). .....39
- Figura 4.** Representación esquemática de la secuenciación por Illumina HiSeq 2500. (Figura tomada de Illumina Inc., 2016). .....42
- Figura 5.** Distribuciones de marcadores SNPs en las poblaciones de maíz. (a) Reproducibilidad de los SNPs; (b) Tasa de llamada de los SNPs. ....48
- Figura 6.** Distribución de frecuencia de los contenidos de información de polimorfismo (PIC) de los marcadores SNP. ....50
- Figura 7.** Dispersión gráfica de los valores de distancia genética y fenotípica de los pares posibles entre 76 poblaciones de maíz. ....55
- Figura 8.** Dispersión gráfica de 76 poblaciones de maíz nativo con base en los dos primeros componentes principales (CP1 y CP2), obtenidos de datos morfológicos. En las poblaciones se indica el número de entrada; las letras, la identificación de cada raza: C: Celaya; CN: Cónico Norteño; EC: Elotes Cónicos; EO: Elotes Occidentales; O: Olotillo; R: Ratón; T: Tuxpeño; TN: Tuxpeño Norteño. ....57
- Figura 9.** Dispersión gráfica de 76 poblaciones de maíz nativo con base en los dos primeros componentes principales (CP1 y CP2), obtenidos de datos genéticos. En las poblaciones se indica el número de entrada; las letras, la identificación de cada raza: C: Celaya; CN: Cónico Norteño; EC: Elotes Cónicos; EO: Elotes Occidentales; O: Olotillo; R: Ratón; T: Tuxpeño; TN: Tuxpeño Norteño. ....58
- Figura 10.** Dendrograma jerárquico de las relaciones entre 76 poblaciones de maíz del sureste de Coahuila, México mediante caracteres morfológicos. Las letras indican la identificación de cada raza: C= Celaya, O= Olotillo; CN= Cónico Norteño, R= Ratón, EC= Elotes Cónicos, EO= Elotes Occidentales, T= Tuxpeño y TN= Tuxpeño Norteño y los dígitos el número de colecta. ....62
- Figura 11.** Dendrograma jerárquico de las relaciones entre 76 poblaciones de maíz del sureste de Coahuila, México mediante marcadores SNP. Las letras indican la identificación de cada raza: C= Celaya, O= Olotillo; CN= Cónico Norteño, R= Ratón, EC= Elotes Cónicos, EO= Elotes Occidentales, T= Tuxpeño y TN= Tuxpeño Norteño y los dígitos el número de colecta. ....65

## **RESUMEN**

ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE MAICES NATIVOS DEL  
SURESTE DE COAHUILA, MÉXICO POR EL MÉTODO DArTseq

**POR**

ELIAS MÉNDEZ HERNÁNDEZ

MAESTRÍA EN CIENCIAS EN FITOMEJORAMIENTO  
UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

DR. MIGUEL ÁNGEL PÉREZ RODRÍGUEZ – ASESOR –

SALTILLO, COAHUILA

JULIO 2020

## RESUMEN

La evaluación y el entendimiento de la diversidad de maíces nativos son importantes para el planteamiento de estrategias de conservación, caracterización y uso del germoplasma en el mejoramiento genético. En este estudio se caracterizó un grupo de accesiones de germoplasma de maíz utilizando marcadores de polimorfismo de un solo nucleótido (SNP), mediante la tecnología DArTseq, con el objetivo de analizar la diversidad genética de 76 poblaciones de maíz nativas del sureste de Coahuila, México. La extracción del ADN genómico se realizó a partir de hojas jóvenes, a través del método de Bromuro de Cetil trimetil amonio (CTAB), y sometido a un análisis de genotipificación mediante secuenciación de última generación, creando perfiles genómicos completos de los individuos caracterizados. Se determinaron los índices de diversidad genética; Heterocigosidad esperada ( $H_e$ ), Heterocigosidad observada ( $H_o$ ), Número de alelos efectivos ( $A_e$ ), Índice de Shannon ( $SH$ ), Alelos raros ( $R_j$ ) y los análisis de agrupamiento basados en la distancia genética de Rogers modificada entre las 76. En este estudio se identificaron 73,834 marcadores (SNPs) polimórficos, de los cuales 53,471 fueron de alta calidad con valores de contenido de información polimórfica de 0.10 a 0.50, una reproducibilidad de 97.50% entre las muestras replicadas y un promedio de tasa de llamadas del 82.84%. La  $H_o$  varió de 0.406 a 0.524, con media de 0.474, mientras que la  $H_e$  fue de 0.408 a 0.445, con una media de 0.426. El análisis de correlación entre la matriz de distancias fenotípicas y genéticas reveló una asociación positiva, con un valor de  $r=0.42^{**}$ . A través de técnicas de análisis multivariantes se estudió la diversidad genética, lo que permitió identificar dos grupos de poblaciones de acuerdo con su área de adaptación y el tipo de mazorca. El primero incluyó a las poblaciones adaptadas a las áreas de Transición-Altura (Cónico Norteño, Elotes Cónicos), clasificada como Complejo mazorca cónica. El segundo representó a las poblaciones adaptadas a áreas Bajas-Intermedias (Tuxpeño, Tuxpeño Norteño, Ratón, Celaya y Olotillo), identificadas como Complejo mazorca cilíndrica. Las poblaciones del Complejo mazorca cónica albergan una amplia diversidad genética, superior a la variación de las poblaciones adaptadas a áreas intermedias. Se encontró una variación amplia entre las poblaciones de los tipos Cónico Norteño, y dentro de las poblaciones de las razas Ratón y Tuxpeño Norteño; dentro de estos grupos raciales existe una variación

determinada por las combinaciones y formas intermedias entre ellos. El uso de los marcadores moleculares SNPs, desarrollados por la tecnología de DArTseq, fue crucial para entender las relaciones genéticas entre poblaciones de las razas nativas del sureste de Coahuila, México.

**Palabras clave:** Maíces nativos, diversidad genética, marcadores SNP, DArTseq.

**ABSTRACT**

ANALYSIS OF THE GENETIC DIVERSITY OF COAHUILA NATIVE MAIZE BY  
THE METHOD DArTseq

BY

ELIAS MÉNDEZ HERNÁNDEZ

MASTER OF SCIENCE IN PLANT BREEDING

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

DR. MIGUEL ÁNGEL PÉREZ RODRÍGUEZ – ADVISOR –

SALTILLO, COAHUILA

JULIO 2020

The assessment and understanding of local maize diversity are very important for conservation strategies, characterization and germplasm utilization in genetic improvement. In our study, the genetic diversity of 76 native maize populations from southeastern Coahuila were characterized by single nucleotide polymorphism (SNP) analysis based on DArTseq. Genomic DNA was extracted from young leaves using Cetyl trimethyl ammonium bromide (CTAB) protocol. Genotyping analysis by next-generation sequencing method was performed to create complete genomic profiles of the characterized individuals. Genetic diversity indices: Expected heterozygosity ( $H_e$ ), Observed heterozygosity ( $H_o$ ), Number of effective alleles ( $A_e$ ), Shannon index ( $SH$ ), Rare alleles ( $R_j$ ) and cluster analyzes, were determined based on the modified Rogers genetic distance between 76 individuals. In this study, a total of 73,834 polymorphic markers (SNPs) were identified, of which 53,471 have high quality of polymorphic information content values of 0.10 to 0.50, a reproducibility of 97.50% among the replicated samples and an average call rate of 82.84%. The  $H_o$  ranged from 0.406 to 0.524, with a mean of 0.474, while the  $H_e$  ranged from 0.408 to 0.445, with a mean of 0.426. Correlation analysis between the phenotypic and genetic distance matrix revealed a positive association, with a value of  $r = 0.42^{**}$ . Genetic diversity was assessed through multivariable analysis, which can identify two groups of populations according to their adaptation area and the corncob type. The first group included the populations adapted to the Transition-Height areas (Conico Norteño, Elotes Conicos), classified as a conical cob complex. The second group represented the populations adapted to Low-Intermediate areas (Tuxpeño, Tuxpeño Norteño, Ratón, Celaya and Olotillo), identified as Cylindrical cob complex. The populations of the conical cob complex harbor a wide genetic diversity, superior to the variation of the populations adapted to intermediate areas. A wide variation was found between the Conico Norteño types populations, and inside the Raton and Tuxpeño Norteño breeds populations. In these racial groups there is a variation determined by the combination, and an intermediate form is formed between them. SNPs markers, was crucial to investigate the genetic relationships between populations of native breeds in southeastern Coahuila, Mexico.

**Keywords:** Native maize, genetic diversity, SNP markers, DArTseq.

## INTRODUCCIÓN

El maíz (*Zea mays* L.) es el cultivo agrícola más importante desde el punto de vista alimentario, industrial y social; se produce en dos ciclos productivos: primavera-verano (PV) y otoño-invierno (OI), bajo las más diversas condiciones agroclimáticas de humedad (temporal y riego) y diferentes tecnologías (SAGARPA-SIAP, 2018). Al ser originario y domesticado en México, es aquí donde se encuentra la mayor riqueza en diversidad genética documentada por diferentes grupos raciales (Wellhausen *et al.*, 1951).

Gran parte de la diversidad genética del maíz nativo de México se encuentra en los campos agrícolas en forma de variedades criollas, sembrándose a una altura de 0 a 2900 metros sobre el nivel del mar (msnm), en regiones con una temperatura media anual de 11.3 a 26.6 °C y una precipitación media anual de 426 a 4245 mm (Ruíz *et al.*, 2008).

La recolección y estudio de los maíces nativos mexicanos han sido objeto de investigación por dos causas importantes: 1) contar con fuentes de germoplasma para su uso en los programas de mejoramiento genético, y 2) para obtener un planteamiento de estrategias de conservación y caracterización del cultivo (Hernández y Esquivel, 2004).

La diversidad de maíz en el sureste del estado de Coahuila está constituida de manera compleja por tipos raciales, variantes dentro de raza e intermedias entre razas desarrollados por los agricultores, que han sido clasificada por medio de caracteres morfológicos, usando principalmente caracteres de la mazorca (Rincón *et al.*, 2010). En este mismo sentido, en un análisis de la expresión de los caracteres en 148 colectas de 50 razas mexicanas de maíz evaluadas a través de 10 ambientes, Sánchez *et al.* (1993) determinaron que los caracteres de la mazorca, tienen una interacción con el ambiente de baja a moderada, lo cual indica la importancia en la descripción y clasificación de las poblaciones de maíz.

Los caracteres morfológicos han sido de gran utilidad en la discriminación y organización racial del maíz, frecuentemente se emplean guías y descriptores básicos que permiten explicar y entender la diversidad genética e interrelaciones del maíz en forma rápida y sencilla. Sin embargo, muchos de estos caracteres morfológicos se ven

afectados por su baja heredabilidad y alta interacción con el ambiente, los que influyen en la expresión del fenotipo, como es el caso de los caracteres vegetativos y caracteres de mazorca, por lo que en muchos casos se requieren de nuevos métodos para realizar los estudios de diversidad del maíz, para ampliar el conocimiento logrado con estudios morfológicos (Khush, 2001; Beyene *et al.*, 2005).

Las nuevas técnicas biotecnológicas para la evaluación genética rápida, tal es el caso de los marcadores moleculares, están siendo utilizadas con mayor precisión debido a que no consideran los efectos ambientales, como sucede con las técnicas clásicas para el análisis y clasificación de la diversidad genética de las especies (Govindaraj *et al.*, 2015).

El uso de marcadores moleculares ha permitido caracterizar la diversidad de materiales nativos. Dentro de dichos marcadores, los polimorfismos de un solo nucleótido (SNPs), se han utilizado extensamente para la caracterización de la diversidad y en la descripción de poblaciones en cuanto a su estructura genética, debido a que se encuentran muy extendidos en el genoma, son muy precisos y reproducibles, adecuados para la detección de alto rendimiento y no se ven afectados por las condiciones ambientales o el estado fisiológico de la planta (Warburton *et al.*, 2010; Jonah *et al.*, 2011).

Debido a las condiciones ambientales del Sureste del estado de Coahuila, y a las modificaciones del ambiente, existe la posibilidad creciente de la existencia de un riesgo a la permanencia de la diversidad genética del maíz tal como la conocemos en esta región. Esta problemática ha dado lugar a investigaciones destinadas a describir y comprender los factores que influyen en la dinámica de la diversidad del maíz, a través de la obtención de información detallada sobre las diferencias a nivel genómico de los maíces. Esto podría estimular su uso para el diseño de mejores estrategias de conservación de esta diversidad, además de ser una alternativa más en la búsqueda de alelos o genes nuevos y útiles que puedan ser incorporados en los programas del mejoramiento genético.

La presente investigación tiene por objetivo analizar la diversidad genética de maíces nativos del sureste del estado de Coahuila, a través de su caracterización genómica, utilizando marcadores moleculares SNP de tipo DArTseq desarrollados en México.



## OBJETIVOS E HIPÓTESIS

### Objetivo general

Estudiar la diversidad genética de 76 poblaciones de maíces nativos del sureste del estado de Coahuila, México.

### Objetivos específicos:

- Realizar la genotipificación de 76 poblaciones nativas de maíz mediante la tecnología de DArTseq.
- Realizar el análisis de diversidad genética utilizando polimorfismos de un solo nucleótido (SNP).
- Analizar la relación entre caracteres morfológicos y genéticos en el estudio de diversidad del maíz.

### Hipótesis

La genotipificación por medio del método DArTseq permitirá estudiar con mayor precisión las relaciones entre y dentro de grupos raciales que determinan la diversidad genética de maíces nativos en el sureste de Coahuila.

## REVISIÓN DE LITERATURA

### Importancia del maíz en México

En México, el maíz (*Zea mays* L.) es el cereal básico e importante en la alimentación humana, y también donde se encuentra la mayor parte de la riqueza en cuanto a diversidad genética (Ortega *et al.*, 1991).

México es considerado como el centro de origen y domesticación de maíz (Matsuoka *et al.*, 2002), posee una amplia variabilidad genética expresada en más de 59 grupos raciales (Sánchez *et al.*, 2000). Las poblaciones de amplia base genética representan el material base para los programas de mejoramiento genético, por su potencial como fuente de características nuevas, exóticas y favorables (Vigouroux *et al.*, 2008).

México ocupa el séptimo lugar como productor de maíz grano a nivel mundial, con un promedio de 27, 169, 997 t (FAOSTAT, 2018). A nivel nacional, la superficie sembrada, promedio anual fue de 7,366, 947.47 ha, de las cuales el 78% se cultiva bajo condiciones de riego y el 22% se realiza en condiciones de temporal (SAGARPA-SIAP, 2018).

### Importancia del maíz en el sureste de Coahuila

En el estado de Coahuila, en 2018 se sembraron 30,521.93 ha de maíz para grano (85 % de temporal y 15 % de riego), con rendimiento promedio de grano de 1.23 t ha<sup>-1</sup>, principalmente con variedades locales nativas (variedades criollas). En el sureste del estado de Coahuila, la superficie sembrada fue de alrededor de 26,782 ha, de las cuales el 94 % se sembró bajo condiciones de temporal o secano, y el restante para condiciones de riego, con un rendimiento promedio de 0.73 t ha<sup>-1</sup> (SAGARPA-SIAP, 2018).

La siembra de maíz para grano, bajo la modalidad de temporal, en el Sureste de Coahuila, involucra el uso de materiales genéticos adaptados a las condiciones particulares de producción, lo que infiere una amplia diversidad genética a lo largo de esta región. En el 80% de la superficie donde se cultiva maíz, se siembran maíces

nativos, que a lo largo de los años los productores han conservado por selección de semilla a partir de sus cosechas.

En el estado de Coahuila, se ha identificado la presencia de ocho grupos raciales: Celaya, Cónico Norteño, Elotes Cónicos, Elotes Occidentales, Olotillo, Ratón, Tuxpeño y Tuxpeño Norteño, poblaciones que representan la variación del maíz (Rincón *et al.*, 2010; Rincón y Ruiz, 2015).

### **Genética de poblaciones**

De acuerdo con Templeton (2006), la genética de poblaciones ha desempeñado un papel central en la biología evolutiva, ya que se ocupa de los mecanismos por los cuales la evolución ocurre dentro de las poblaciones y las especies, estos cambios evolutivos son la interacción de los factores principales como: mutación, migración, recombinación, selección y deriva genética. La genética de poblaciones ofrece herramientas básicas para la biología de la conservación, no solo porque involucra el estudio del origen, cantidad y distribución de la variación genética en las poblaciones, sino que también se interesa en el destino de esta variación a través del tiempo y el espacio.

El término “Población”, en genética de poblaciones, se refiere a un grupo de organismos de la misma especie que viven en un área geográfica suficientemente restringida para que cada individuo tenga la oportunidad potencial de aparearse con algún otro individuo. Uno de los aspectos principales en genética de poblaciones es, precisamente, descubrir la arquitectura genética de las poblaciones naturales e identificar las fuerzas y factores evolutivos que expliquen dicha variación (Hartl y Clark, 1997).

Las tecnologías modernas utilizadas en el mejoramiento genético prometen promover el uso de la diversidad genética existente en las poblaciones, siendo inherente en cualquier programa de selección el conocimiento básico de la diversidad genética y la estructura de la población del material a mejorar.

### **Diversidad genética**

La diversidad genética de una especie representa la variación heredable entre y dentro de sus poblaciones, y en especies cultivadas tienen gran trascendencia porque en ello operan los procesos de selección que aplican los agricultores y fitomejoradores (López *et al.*, 2009).

La diversidad del maíz se puede atribuir a la selección que es practicada por el hombre a través del proceso de domesticación, así como a los numerosos entornos microecológicos y los efectos ambientales que cada condición climática ejerce sobre las poblaciones que determinan la capacidad de adaptación (Caraballosa *et al.*, 2000).

El estudio de la diversidad genética dentro y entre poblaciones se realiza mediante el uso de marcadores morfológicos, bioquímicos y moleculares. Los marcadores de ADN (moleculares) han ganado importancia en la caracterización del germoplasma, debido a la cantidad de polimorfismos distribuidos en el genoma, la rapidez y calidad de datos generados. (Govindaraj *et al.*, 2015). Actualmente los Single Nucleotide Polymorphisms (SNPs), o polimorfismos de un solo nucleótido, se han desarrollado y utilizado con mayor frecuencia como los marcadores preferidos en varios cultivos, para la evaluación de la variabilidad genética entre diferentes accesiones de una especie. Esto, debido a su amplia cobertura del genoma, herencia codominante, bajo costo y rápido seguimiento en comparación con otros marcadores moleculares (Gupta y Rustgi, 2004; Kujur *et al.*, 2015).

En estudios de diversidad genética y de estructura poblacional es de suma importancia contar con marcadores moleculares que permitan identificar los niveles de diversidad, así como la relación entre los individuos. Un mayor número de marcadores moleculares permite encontrar un mayor número de sitios que presentan variaciones dentro de los genomas de los individuos a evaluar y al igual obtener de manera más confiable y representativa la diversidad genética (Glaubitz *et al.*, 2014; Xu *et al.*, 2017).

### **Descripción de la diversidad genética**

Para analizar la variabilidad genética existen diversos procedimientos que resumen la información en términos de estimaciones manejables. Los parámetros más empleados y

que se describen a continuación son: Polimorfismo, Contenido de información polimórfica (PIC), Heterocigosidad esperada ( $H_e$ ), Heterocigosidad observada ( $H_o$ ), Número de alelos efectivos ( $A_e$ ) y Alelos raros ( $R_j$ ).

### **Polimorfismo**

La existencia de múltiples alelos presentes en cada gen, demuestra la existencia de la variación genética y es uno de los factores con mayor influencia en su generación, esta diversidad se presenta en la mayoría de las poblaciones naturales. A estos múltiples estados que puede poseer un gen se le conoce como polimorfismos (Hartl & Jones, 2009).

El polimorfismo, o la proporción de loci polimórficos ( $P$ ), es una medida del número de loci variable en una población y se calcula dividiendo el número de loci polimórficos entre el número total de loci analizados (Sosa *et al.*, 2002).

$$p = \frac{\text{N}^\circ \text{ total de loci polimórfico}}{\text{N}^\circ \text{ total de loci}}$$

Para que un gen se considere polimórfico, cada alelo debe aparecer en la población a una frecuencia de al menos el 1%, es decir, cuando el alelo más frecuente de un gen en una población alcance un 95 % en frecuencia o 0.95 se determina que existe polimorfismo. Para ello, es necesario el uso de un criterio arbitrario a fin de precisar el polimorfismo de un locus de acuerdo con su frecuencia alélica, debido a que cuando se aumenta el número de individuos muestreados en una población, la probabilidad de encontrar un nuevo alelo también aumenta (Hedrick, 1983).

### **Contenido de información polimórfica (PIC)**

El contenido de información polimórfica (PIC) es similar al valor de heterocigosidad, y su valor puede variar entre 0 a 1. Este índice estima la eficiencia de cada marcador para determinar polimorfismos en un locus dado en la población de acuerdo con las frecuencias de los alelos, valores más cercanos a 1 existe un mayor nivel de

polimorfismo. Para su cálculo se multiplica la probabilidad de cada posible cruzamiento (estimado a partir de las frecuencias alélicas), por la probabilidad que sean informativos, es decir, que se pueda identificar al progenitor del que procede el alelo (Botstein *et al.*, 1980).

### **Heterocigosidad observada**

La heterocigosidad observada ( $H_o$ ) se define como la proporción de organismos heterocigotos calculada a partir de los genotipos observados en una muestra poblacional. Estima la probabilidad, de que dos alelos tomados al azar de la población sean diferentes; sin embargo, no refleja bien la cantidad de variación genética en las poblaciones de organismos que se reproducen asexualmente o por autofecundación (Frankham *et al.*, 2009).

### **Heterocigosidad esperada**

Nei (1987) asegura que el promedio de heterocigosidad o diversidad genética es el método de medición más adecuado de la variación genética, pues no depende de la arbitrariedad de la definición de polimorfismo y puede ser definida sin ambigüedad en términos de frecuencias alélicas.

La heterocigosidad esperada de una población ( $H_e$ ) es la media de los valores obtenidos para cada locus en dicha población. Cuando las poblaciones están bajo equilibrio Hardy-Weinberg, la heterocigosidad puede ser calculada a partir de la frecuencia alélica. Por lo tanto, la heterocigosidad esperada cuantifica la igualdad o equitatividad de las frecuencias alélicas en los loci (Hoelzel, 1998; Sosa *et al.*, 2002).

### **Número de alelos efectivos**

El número de alelos efectivos ( $n$ ), hace referencia a los alelos con capacidad de pasar a la siguiente generación y es otro de los parámetros que se usa para medir la variabilidad genética. Es el recíproco de la probabilidad de que dos genes homólogos elegidos al azar

en la población sean idénticos por descendencia ( $n = \frac{1}{F}$ ). Por lo tanto, la única forma en que dos genes puedan ser el mismo alelo es si son idénticos por descendencia. (Kimura y Crow, 1964).

### **Alelos raros**

Los alelos raros se han definido en términos de sus frecuencias relativas. Raychaudhuri (2011) define a los alelos raros como una variante poco frecuente, y que la inclusión de polimorfismos con baja frecuencia alélica puede determinar la arquitectura de los caracteres en complejos genéticos, por lo tanto, los alelos raros se consideran alelos polimórficos que tienen <1% de frecuencia en una población específica (Reyes-Valdés *et al.*, 2018).

Los estudios de asociación de genoma completo (GWAS, por sus siglas en inglés Genome-Wide Association Study), es una estimación que permite detectar la influencia de variaciones alélicas a nivel genómico sobre la expresión de caracteres complejos en los cultivos. La mayoría de las especies tienen variaciones frecuentes de diversidad genética, incluidos alelos raros, que aparentan desaparecer en una gran población debido a una mutación perjudicial (Raychaudhuri, 2011).

### **Marcadores moleculares**

Los marcadores moleculares se basan en el análisis de las diferencias en secuencias del ADN (Ácido Desoxirribonucleico) entre individuos. Las técnicas empleadas para ello son muy diversas, y se puede clasificar como de carácter dominante o codominante (Karp & Edwards, 1998).

El mejoramiento genético es un proceso a largo plazo, por lo que es necesario recurrir al apoyo de técnicas que aceleren este proceso. Una alternativa podría ser la utilización de herramientas moleculares, basadas en el uso del ADN y la identificación de variaciones genéticas de interés para el mejorador.

En la actualidad existen varias técnicas moleculares que permiten conocer cómo se encuentran las proporciones alélicas en las poblaciones naturales de manera indirecta, como con los análisis de proteínas, o de manera directa con estudios de ADN.

### **Isoenzimas**

Las isoenzimas, o aloenzimas fueron originalmente definidas como marcadores moleculares múltiples de las enzimas que tienen funciones idénticas o similares, y que están presentes en el mismo individuo, y estas fueron las primeras en ser usadas en genética de plantas (Tanksley *et al.*, 1981). En su mayoría son selectivamente neutras y se utilizan como marcadores hereditarios para cuantificar las frecuencias alélicas y genotípicas de los individuos, que son los estimadores básicos de la composición genética de una población.

La aplicación de las isoenzimas está dirigida a la cuantificación de heterocigosis, diversidad genética, diferenciación genética y otras medidas de variación genética intra e interpoblacional. También han sido aplicadas exitosamente para evaluar y entender aspectos de biología evolutiva como los sistemas de reproducción y patrones de fecundación cruzada, relaciones entre fenotipo y ambiente, filogenias, diversidad en plantas (Pérez-Nasser y Piñero, 1997).

### **Polimorfismo de Longitud de Fragmentos de Restricción (RFLP)**

El análisis RFLP es una técnica usada desde los años 1970. Esta técnica se basa en la variación en la longitud de los fragmentos de ADN, los cuales son producidos por una endonucleasa de restricción específica a partir de ADN genómicos de dos o más individuos de una especie. Los RFLPs se generan por rearrreglos o mutaciones que dan lugar a la creación o delección de sitios de reconocimiento para las endonucleasas específicas (Valadez y Kahl, 2000).

Los RFLPs han sido ampliamente usados en plantas con diferentes objetivos: caracterización de germoplasma, estudios filogenéticos, pureza de semillas híbridas, selección y/o localización de genes específicos (mediante análisis de ligamiento) de



características agronómicas importantes, etc. (Phillips *et al.*, 1995; Valadez y Kahl 2000).

Según Paterson (1996), el costo, la complejidad técnica y la gran cantidad de ADN que se requiere pueden ser limitantes para el empleo de los RFLP como técnica de diagnóstico. En algunos casos, aunque la variación interespecífica es alta, los individuos de una población o los cultivares de una especie muestran una variación baja, lo cual reduce las posibilidades de encontrar polimorfismos útiles. (Bernatzky, 1988).

### **Polimorfismo de ADN Amplificados al Azar (RAPD)**

Son marcadores que amplifican aleatoriamente segmentos de ADN en una gran variedad de especies, son de carácter dominante, es decir, no pueden discernir los homocigotos dominantes de los heterocigotos para un segmento particular. Los RAPDs se basan en la probabilidad estadística de que se presenten sitios complementarios al oligonucleótido de 10 pares de bases (pb) a lo largo del genoma. El polimorfismo de las bandas entre los individuos se debe a cambios en la secuencia de los nucleótidos en los sitios de acoplamiento del oligonucleótido y por inserción o deleción de los fragmentos en estos sitios (Williams *et al.*, 1990).

Los RAPDs son útiles en la elaboración de mapas genéticos, en el estudio de parentesco y en el análisis de la estructura poblacional, ya que ayudan a estimar tamaño efectivo, aislamiento reproductivo y niveles de fecundación cruzada (Otero *et al.*, 1997).

Entre las principales ventajas de los RAPDs, están la amplificación de regiones tanto codificantes del ADN como las no codificantes y revelan niveles de variación más altos que los RFLPs e isoenzimas (Williams *et al.*, 1990). Sin embargo los problemas prácticos detectados son la presencia de bandas “erróneas” (artefactos), la reproducibilidad de los resultados y la comigración de bandas. Por otro lado, como los loci son dominantes, los RAPDs dan menos información genética por locus que los marcadores codominantes.

### **Polimorfismo de Longitud de Fragmentos Amplificados (AFLP)**

Los AFLPs son considerados marcadores de alta eficacia, permiten el análisis de un elevado número de loci por experimento sin requerir información previa sobre su secuencia, son en su mayoría dominantes y altamente reproducibles. Sin embargo, es más complicada de ejecutar a comparación con otras técnicas, y además requieren una mayor cantidad de ADN (Karp *et al.*, 1997). Se pueden emplear, en la construcción del armazón de los mapas genéticos en el que se localizan los marcadores codominantes, para discriminar entre individuos cercanamente relacionados y para localizar genes específicos en genomas complejos (Valadez y Kahl, 2000).

### **Microsatélites o secuencias simples repetidas (SSR)**

Los microsatélites o secuencias simples repetidas (SSR), son regiones de secuencias pequeñas (dos a 10 pares de bases) repetidas, las cuales se asume que están distribuidas azarosamente por todo el ADN. Son secuencias de ADN altamente variables dispersas a través de los genomas de plantas y animales, los cuales pueden o no estar asociadas con genes, son loci altamente mutables que pueden estar presentes en muchos sitios del genoma. Dado que, la repetición por sí misma no codifica para formar ninguna proteína, y debido a que las secuencias de ADN repetitivo pueden recombinarse y expandirse más frecuentemente que otros tipos de secuencias, estas regiones son a menudo altamente variables y consecuentemente útiles para medir el polimorfismo entre especies o variedades muy relacionadas (Phillips *et al.*, 1995; Valadez y Kahl, 2000).

Los microsatélites han tomado ventaja sobre otros marcadores genéticos, debido a que: a) tienen el más alto grado de polimorfismo; b) segregan de manera mendeliana y son codominantes; c) la presencia de un solo locus genético por microsatélite hace que la lectura de las bandas sea clara y fácil de interpretar y d) son selectivamente neutros (Golstein y Pollock, 1994).

Estos marcadores son ideales para el estudio de ligamiento genético en plantas y el mapeo físico, los estudios poblacionales y la identificación de variedades (SIDTA, 1999).

### **Polimorfismos de un solo nucleótido (SNP)**

Los marcadores basados en polimorfismos de base simple o SNPs son variaciones de secuencias de ADN, que ocurren cuando un solo nucleótido (A, T, C o G) cambia la secuencia de un par de bases del genoma, es decir, variaciones de un solo nucleótido en la secuencia del genoma de uno o más individuos pertenecientes a una población (Edwards *et al.*, 2007).

Los SNP están ampliamente distribuidos dentro del genoma y pueden encontrarse en regiones codificantes o no codificantes de genes o entre dos genes (región intergénica), su frecuencia oscila entre 1 cada 600 y 1 cada 1000 pares de bases, y se encuentran presentes en abundancia en plantas (Xu, 2010).

### **Microarreglos - Diversity Arrays Technology (DArT)**

La tecnología de matriz de diversidad (DArT), es una de las técnicas moleculares basada en hibridación de microarreglos, que permite la detección de polimorfismos de ADN en varios miles de loci en un único ensayo, sin depender de la información de la secuencia del material genómico (Jaccoud *et al.*, 2001; Wenzl *et al.*, 2004).

Para identificar los marcadores polimórficos, se aplica un método de reducción de complejidad en el genoma, la representación genómica obtenida de este grupo se clona y los insertos individuales se ordenan en una micromatriz, dando como resultado una "matriz de descubrimiento". Las representaciones genómicas etiquetadas preparadas a partir de los genomas individuales incluidos en el grupo, se hibridan a la matriz de descubrimiento y los clones polimórficos (marcadores DArT) muestran intensidades de señal de hibridación variables para diferentes individuos (Jaccoud *et al.*, 2001).

Semagn *et al.* (2006) mencionan a DArT como uno de los perfiles genéticos útiles para caracterización de colecciones de germoplasma, estudios de diversidad, mapeo de *loci* de caracteres cuantitativos (QTL, Quantitative Trait Loci), y en la construcción de mapas genéticos. Es una técnica de carácter dominante (presencia o ausencia), con diferencias de intensidad, lo que posiblemente tenga alguna limitación de su valor en alguna aplicación.

Una de las ventajas de DArT sobre otros sistemas de marcadores es que proporciona una genotipificación donde no se necesita información de secuencia previa para especies a estudiar, es de alto rendimiento, rápido, altamente reproducible y rentable con un costo estimado, diez veces menor que los marcadores SSR por *data point*. Hasta la fecha, los marcadores DArT se han aplicado con éxito en el análisis de diversidad genética, el mapeo de enlaces y en el descubrimiento de la estructura poblacional de las colecciones en varias especies de cultivos (Xia *et al.*, 2005).

### **Marcadores moleculares en el estudio de diversidad genética de maíz**

Los marcadores moleculares son de gran utilidad por su eficiencia, distribución a lo largo del genoma y sobre todo por no tener influencia del ambiente; proporcionan un enfoque alternativo importante para caracterizar la diversidad genética, la estructura de la población y las relaciones genéticas entre poblaciones o materiales de mejoramiento de una colección de germoplasma de maíz determinado (Wu *et al.*, 2016).

Estudios sobre la diversidad morfológica y genética del maíz, muestran que existe gran variabilidad en las poblaciones mexicanas (Sánchez *et al.*, 1998). Se han llevado a cabo estudios sobre la diversidad genética mediante el uso de herramientas moleculares como las isoenzimas (Sánchez *et al.* 2000), los Polimorfismos de ADN Amplificados al Azar (RAPD) (Lanza *et al.*, 1997; Arias *et al.*, 2018), los Polimorfismos de Longitud de Fragmentos Amplificados (AFLP) (Rodríguez-Pérez *et al.*, 2012), los Microsatélites o Secuencias Repetidas Cortas (SSR) (Matsuoka *et al.*, 2002; Reif *et al.*, 2006; Rocandio-Rodríguez *et al.*, 2014; Vega-Álvarez *et al.*, 2017), e inclusive con SNPs (Yan *et al.*, 2009; Romay *et al.*, 2013; Arteaga *et al.*, 2015). Muchos de estos estudios han analizado muestras de bancos de germoplasma representativas de todo el país y se han tomado como base para la selección y la clasificación racial de las muestras utilizadas.

Doebley *et al.* (1985) analizaron el patrón de 13 isoenzimas en 94 accesiones que representaban 34 razas de maíz, reportando que el 72% de la variación de isoenzimas residía dentro de las accesiones y el 27% entre las accesiones, con una frecuencia de 7.09 alelos por locus. Sánchez *et al.* (2000) estudiaron morfológicamente y por isoenzimas diferentes razas mexicanas de maíz, utilizando 209 accesiones de 59 razas

donde más del 65% de los alelos encontrados en las colectas estudiadas eran raros en frecuencias menores a 0.01. Además, reportan varias poblaciones con bajos niveles de diversidad y de diferenciación genética, que son aquellas variedades que se caracterizan por sus usos especiales como palomitas de maíz, maíz dulce, el pozole, atole, pinole o tamales.

López *et al.* (2009) mediante polimorfismos de 19 loci de isoenzimas, analizaron la diversidad genética de 40 poblaciones de maíces nativos representativas del Istmo de Tehuantepec, Oaxaca, México, y mediante estimadores de diversidad genética, reportaron valores altos para las poblaciones de maíz del grupo maíz grande, con mayor número de alelos por locus (2.474), polimorfismos (68%) y heterocigosidad esperada (0.224), así como una diferenciación genética entre sus poblaciones mayor que la del Zapalote Chico.

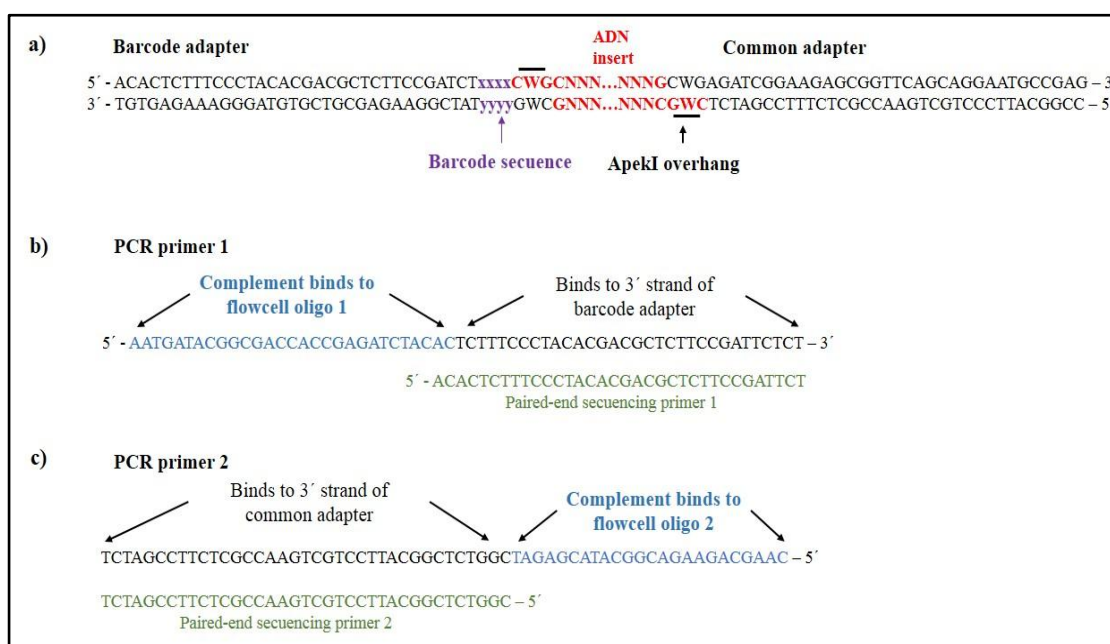
Reif *et al.* (2006) utilizaron 25 marcadores SSR para caracterizar 25 accesiones de 24 razas mexicanas, también reportaron gran variabilidad en el material utilizado, confirmando así mismo la amplia base genética de las razas de maíz en México.

En un estudio realizado por Arteaga *et al.* (2015), utilizando un total de 36,931 SNPs genotipados en 46 variedades de maíz nativas de México, así como la subespecie teosinte encontraron hasta 70,749 alelos totales en una sola raza local (Cónico, n= 16) y un mínimo de 44,134 (Zapalote Chico, n= 1). Reportaron niveles de variación genética para esas variedades locales de maíz ( $He = 0.234$  a  $0.318$ ), con una media de  $0.311$ , mientras que para *Zea m. Mexicana* y *Zea m. parviglumis* detectaron niveles bajos ( $He = 0.262$  y  $0.234$ , respectivamente). Estos autores concluyen que las variedades locales de maíz mexicano tienen una alta diversidad genética que se estructura principalmente de acuerdo con la interacción de latitud y altitud.

Por su parte Rossi *et al.* (2019), evaluaron la diversidad genética de 291 líneas de maíz del CIMMYT utilizando 18,082 SNPs, permitiéndoles estimar el contenido de información polimórfica (PIC) y la diversidad genética. Con el uso de esta cantidad de marcadores encontraron la formación de dos grupos de acuerdo a sus adaptaciones ambientales, y reportaron un total de 71.996 alelos con una media de 3.98 alelos por locus.

## Genotipificación por secuenciación (GBS)

En la actualidad, tecnologías con mayor rendimiento en la producción de marcadores moleculares, son empleadas para obtener una mayor representación del germoplasma evaluado en un solo ensayo. La técnica de GBS, por lo general se basa en la secuenciación de alto rendimiento de la próxima generación de subconjuntos genómicos dirigidos por enzimas de restricción tipo II, para reducir la complejidad del genoma, de esta forma producir la biblioteca de fragmentos de ADN que será sometida a secuenciación (Figura 1). Este método, permite la caracterización de colecciones de germoplasma, mejoramiento, mapeo de rasgos en diversos organismos y estudios de población y proporciona una gran cantidad de marcadores en todo el genoma a bajo costo por muestra. (Elshire *et al.*, 2011).



**Figura 1.** Preparación de bibliotecas mediante la metodología de Genotipado por Secuenciación (GBS) (Figura tomada de Elshire *et al.*, 2011).

Las secuencias de código de barras bicatenario y adaptadores comunes se muestra en la Figura (1a). Los adaptadores son ligados al ADN genómico y cortado por ApeKI. Las posiciones de la secuencia del código de barras y los voladizos de ApeKI se muestran en relación con el inserto de ADN. En la Figura (2b), se muestran las secuencias del cebador de PCR 1 y el cebador de secuenciación final emparejado 1(PE-1), donde se

indican los sitios de unión para el oligonucleótido 1 de celda de flujo y el adaptador de código de barras. Por último, se indican los sitios de unión para el oligonucleótido 2 de celda de flujo y el adaptador común Figura (3c).

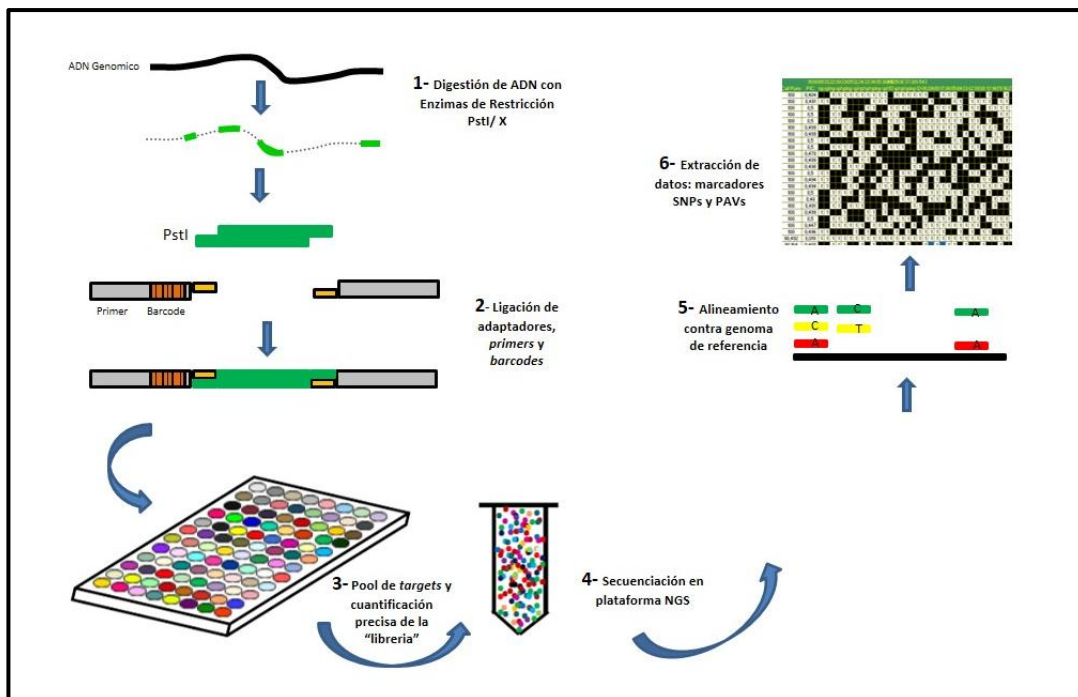
El enfoque genotipado por secuenciación (GBS), ha dado lugar a notables avances en la secuenciación de genomas completos, como es el caso a especies con alta diversidad y genomas grandes como el maíz, debido a esto, Elshire *et al.* (2011), mencionan que deben emplearse el enriquecimiento o la reducción de la complejidad del genoma para asegurar una superposición suficiente en la cobertura de secuencia para especies con genomas grandes, independientemente que la secuenciación (GBS), sea bastante sencillo para genomas pequeños.

Gore *et al.* (2007) mencionan que reducir la complejidad del genoma con enzimas de restricción es fácil, rápido, altamente reproducible y puede llegar a regiones importantes del genoma que son inaccesibles para los enfoques de captura de secuencia.

### **Plataforma DArTseq**

El método de genotipificación DArTseq fue desarrollado por la empresa Diversity Arrays Technology de Australia, y permite la identificación de decenas o inclusive centenas de miles de polimorfismos a partir de una muestra genómica. Está basado en la reducción de la complejidad del genoma por medio de la combinación de dos enzimas de restricción, y en la utilización de *barcodes* tipo código de barras que permiten el genotipado de múltiples muestras en paralelo obteniendo marcadores de alta densidad (Kilian *et al.*, 2012).

En la Figura 2 se muestra una descripción grafica de la genotipificación de alto rendimiento basada en la tecnología de DArTseq.



**Figura 2.** Plataforma de genotipificación de alto rendimiento DArTseq (Figura tomada de Kilian *et al.*, 2012).

Los métodos utilizados en la reducción de la complejidad genómica, generalmente dependen de la digestión con enzimas de restricción (RE), ligación de adaptadores (*barcodes*). Estos *barcodes*, actúan como códigos de barra, que permiten identificar a que muestra de ADN pertenece cada uno de los fragmentos genómicos obtenidos a partir de la digestión con las enzimas (1-2). El agrupamiento de los *targets* en una sola solución se muestra en la etapa 3 de la Figura 2. Estos conglomerados de muestras de ADN son llamados *pools*, el objetivo es simplificar el manejo de los fragmentos a la hora de incorporarlos al secuenciador (4).

El análisis de secuencia es llevada a cabo alineando las lecturas con la representación del genoma basada en las secuencias obtenidas y etiquetadas por este mismo método (5). La obtención de marcadores de polimorfismos de un solo nucleótido (SNPs) y variación de presencia/ausencia del fragmento (PAVs) mediante esta tecnología se ilustra en la etapa 6 de la Figura 2. Los cuales son determinados en el proceso de digestión por enzimas.

DArTseq ofrece un perfil genómico a través de la generación de polimorfismos de un solo nucleótido (SNP) de alta densidad, y también la identificación de variaciones de presencia y ausencia (PAV), pudiéndose obtener de 40,000 a 200,000 marcadores de



este tipo (Sansaloni *et al.*, 2011; Chen *et al.*, 2016). Ambas tecnologías, tanto como la genotipificación por secuenciación (GBS) y DArTseq, en maíz, se producen una enorme cantidad de datos perdidos (hasta un 50%), con la diferencia que DArTseq, es un método más confiable para la detección de heterocigotos, y se generan una menor cantidad de marcadores (50,000 a 350,000 SNPs), pero tiene una cobertura mucho mayor y presenta una menor cantidad de datos perdidos (20% o menor) (Chen *et al.*, 2016; Wu *et al.*, 2016).

Cyrine *et al.* (2019) demostraron la relevancia de la tecnología de DArTseq como un producto confiable y rentable herramienta para evaluar la diversidad dentro y entre las variedades locales, al compararlas con diferentes germoplasmas de trigo duro, produciendo una gran cantidad de loci polimórficos e informativos. También en trigo, un banco de germoplasma de Azerbaiyán ha sido caracterizado con esta tecnología, siendo de gran utilidad para representar la diversidad y las relaciones genéticas entre especies de *Aegilops* L. en distintas regiones geográficas de ese país (Abbasov *et al.*, 2019). El desarrollo de la tecnología DArTseq y el genotipado mediante secuenciación, se utiliza cada vez más para estudiar la diversidad genética de diferentes cultivos, ya que permiten estudiar la diversidad genética de grandes entradas de números y genomas complejos (Chen *et al.*, 2018; Nadeem *et al.*, 2018).

Debido a su carácter rápido, de alto rendimiento y rentable, DArTseq ha sido ampliamente utilizado con éxito en programas de fitomejoramiento, mapeo de ligamiento, identificación QTL en población de mapeo biparental, estudios de asociación de genoma completo (GWAS), estudios de diversidad genética, selección asistida por marcadores y genómica que está ganando una considerable importancia en el incremento de la eficiencia de la mejora de las plantas al permitir transferir de forma más precisa, que los métodos clásicos de selección, regiones genómicas de interés (Baloch *et al.*, 2017).

Los métodos tradicionales de genotipificación se basan en ensayos que detectan los polimorfismos de ADN, de manera indirecta, a través de las diferencias en la movilidad de los fragmentos de ADN en geles de electroforesis, la amplificación diferencial por PCR, o la hibridación diferencial en microarreglos (Elshire *et al.*, 2011). DArTseq se ha

desarrollado con un enfoque técnicamente simple, altamente multiplexado que hace posible omitir el paso de desarrollo del ensayo, no sólo para descubrir, sino también para la clasificación directa de los polimorfismos de ADN (Sansaloni *et al.*, 2011).

## MATERIALES Y MÉTODOS

### Material genético

En este estudio se analizaron 76 poblaciones de maíz nativo del estado de Coahuila, que han sido recolectadas en diversas etapas como parte de las actividades de varios proyectos de investigación. Las poblaciones nativas además de la clasificación racial (Wellhausen *et al.*, 1951), han sido caracterizadas morfológicamente, usando principalmente caracteres de la mazorca (Rincón *et al.*, 2010; Rincón-Sánchez y Ruiz-Torres, 2018), las cuales son representativas de ocho grupos raciales y que han sido agrupadas de acuerdo con la altitud de procedencia (Cuadro 1).

**Cuadro 1.** Clasificación racial y frecuencia de las 76 poblaciones de maíz nativo con base a la altitud del sitio de colecta, Coahuila, México.

Clasificación racial	Grupos (Altitud, en msnm) <sup>†</sup>				Total
	Bajo (<1000)	Intermedio (1001-1800)	Transición (1801-2000)	Altura (>2000)	
Celaya		3			3
Celaya x Ratón <sup>‡</sup>		1			1
Cónico Norteño		1	7	20	28
Cónico Norteño x Ratón		1	1	2	4
Cónico Norteño x Tuxpeño Norteño		1		1	2
Elotes Cónicos		1	1	1	3
Elotes Occidentales		2			2
Olotillo x Ratón		1			1
Ratón	4	8	1		13
Ratón x Celaya		1			1
Ratón x Cónico Norteño		1			1
Ratón x Olotillo		1			1
Ratón x Tuxpeño Norteño		4			4
Tuxpeño		1			1
Tuxpeño Norteño	1	7	1		9
Tuxpeño Norteño x Celaya		1			1
Tuxpeño Norteño x Cónico Norteño		1			1
<b>Total</b>	<b>5</b>	<b>36</b>	<b>11</b>	<b>24</b>	<b>76</b>

<sup>†</sup> Adaptado de Rincón-Sánchez y Ruiz-Torres (2018); <sup>‡</sup> El símbolo “x” entre grupos raciales indica algún grado de combinación o forma intermedia entre los dos grupos raciales.

Los materiales genéticos en este estudio fueron caracterizados con base en caracteres de la mazorca (Rincón-Sánchez y Ruíz-Torres, 2018). A partir de un tamaño de muestras de 10 mazorcas representativas de cada población, se obtuvieron 15 caracteres cuantitativos de la mazorca para analizar la diversidad del maíz. Siete caracteres de la mazorca: diámetro de mazorca y olote (cm), longitud de mazorca (cm), número de hileras de la mazorca, porcentaje de desgrane y dos caracteres de índice: diámetro de mazorca/olote (cm), y diámetro/longitud de la mazorca (cm). Se utilizaron ocho caracteres del grano: longitud, ancho y espesor (mm), número de granos por hilera, peso seco de 100 granos (g) y tres caracteres de índice del grano: Espesor/longitud del grano (mm), espesor/ancho del grano (mm) y ancho/longitud del grano (mm).

La altitud de los sitios de procedencia de las poblaciones en estudio está asociada a los diferentes nichos ecológicos en la región de estudio, y por consiguiente a la diversidad y tipos de maíces. Por tal motivo, esta información es de utilidad en la presente investigación para la valoración y análisis de la diversidad genética de las poblaciones.

### **Proceso de genotipificación**

La caracterización genotípica de las muestras evaluadas en este estudio se llevó a cabo en el Laboratorio del Servicio de Análisis Genético para la Agricultura (SAGA), del Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), ubicado en Carretera México-Veracruz Km. 45, El Batán, Texcoco, México.

### **Obtención de muestras de tejido**

Se realizó la siembra de 30 semillas representativas de cada población de maíz en los invernaderos con sede en CIMMYT (El Batán), bajo condiciones controladas. A los 15 días después de la siembra, se realizó la selección de las hojas de manera visual de aquellas entradas que germinaron. A partir de 15 plántulas de cada población, se tomó una fracción de 10 a 15 cm de hojas frescas y jóvenes, sin zonas necróticas. Esta estrategia de toma de 15 entradas por población, muestra compuesta de ADN (“bulk”),

tuvo como objetivo representar de manera amplia la mayor diversidad posible de cada una de las poblaciones.

### **Liofilización**

Las muestras de cada población fueron colocadas en glases a  $-80\text{ }^{\circ}\text{C}$  durante 24 horas. Posteriormente, el material foliar fue colocado en una liofilizadora (Labconco<sup>TM</sup>, modelo 230 v), en condiciones de vacío de  $\leq 100$  micrones durante 5 días y una temperatura de condensador de  $\leq 50\text{ }^{\circ}\text{C}$ . Una vez liofilizadas, las muestras se secaron por 72 horas y se conservaron en bolsas de plástico herméticamente cerradas a temperatura ambiente.

### **Molienda**

El tejido vegetal liofilizado, se cortó en pedazos pequeños de 0.5 a 1.5 cm y los fragmentos se depositaron en tubos de extracción de 1.5 ml, junto con una esfera metálica pequeña de 4 mm de diámetro. La molienda se llevó a cabo en un molino mecánico de la marca SPEX<sup>TM</sup> GenoGrinder a 1750 revoluciones por minuto (rpm) por 3 min, obteniendo un polvo fino.

### **Extracción del ADN genómico**

Siguiendo el método de Murray y Thompson (1980), se colocaron 300 mg de tejido y molido en un tubo para centrifugación de 15 ml, posteriormente se agregaron 9.0 ml de solución amortiguadora CTAB (Bromuro de hexadeciltrimetilamonio) a  $65\text{ }^{\circ}\text{C}$ , se mezcló invirtiendo el tubo varias veces suavemente. Las muestras fueron incubadas a  $65\text{ }^{\circ}\text{C}$  de 60 a 90 min, agitando los tubos periódicamente con suavidad. Posteriormente, los tubos fueron colocados a temperatura ambiente durante 5 min y se agregaron 4.5 ml de cloroformo/octanol (24:1), la mezcla se colocó en un agitador rotativo vertical (Fisher Scientific) 15 min y se centrifugó a 3750 rpm durante 30 min (Beckman Coulter, Allegra<sup>®</sup> x-15R Centrifuge), tras lo cual se recuperaron 750  $\mu\text{l}$  de la capa acuosa superior y se colocó en tubos limpios de 1.5 ml, se agregaron nuevamente 4.5 ml de cloroformo/octanol (24:1), se agitó suavemente durante 10 min y se centrifugó igual que

en las condiciones previas. Con una pipeta se transfirió la capa acuosa superior a un tubo limpio de 1.5 ml, se agregaron 30  $\mu$ l de ARNasa (10 mg/ml) y se incubó a temperatura ambiente por 30 min, posteriormente se agregaron 6.0 ml de isopropanol (2-propanol) y se mezcló suavemente. Finalmente, el ADN precipitado se retiró mediante un gancho de vidrio y fue almacenado en el refrigerador a 4 °C.

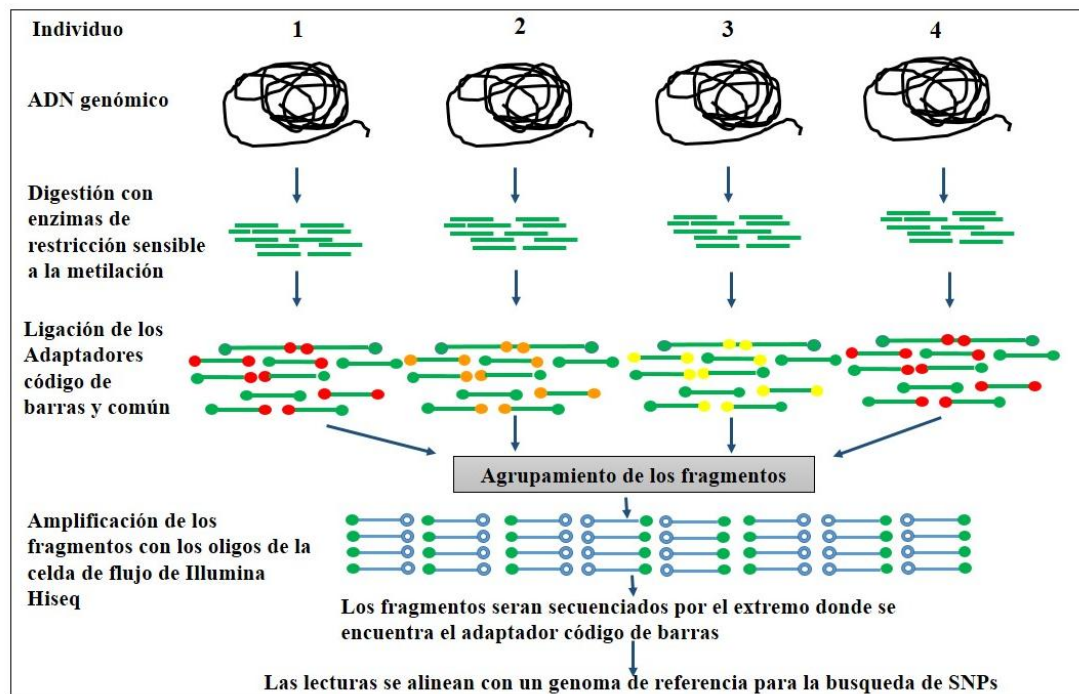
### **Cuantificación y calidad del ADN genómico**

El ADN fue cuantificado mediante espectrofotometría, se centrifugaron brevemente las muestras a 1500 rpm, después se añadió 3  $\mu$ l de ADN en el Nanodrop (Thermo Scientific NanoDrop ND8000 multi-muestra Micro-Volumen UV-Vis Espectrofotómetro) a través de lecturas de absorbancia de 260/280 nm a partir de 1  $\mu$ L de la muestra. El ADN se aceptó como “puro” cuando las muestras tuvieron una razón 260/280 > 1.8 (Madigan *et al.*, 1998). Después de la cuantificación las muestras fueron almacenadas a 4 °C.

Para analizar la calidad del ADN se prepararon diluciones de 50 ng/ $\mu$ l de cada muestra. Enseguida se cargaron 10  $\mu$ l de ADN más 30  $\mu$ l bromuro de etidio para ocho placas en un pocillo de un gel de agarosa al 0.8 %, adicionando en un carril por peine PhiX174/HaeIII como marcador de peso molecular y sometido a electroforesis a 120 volts/20 minutos, finalmente se fotografiaron utilizando un sistema de documentación en gel (BioDocAnalyze Live, Biometra).

### **Reducción de la complejidad genómica**

En la Figura 3 se muestra una descripción gráfica de la preparación de las bibliotecas de DArTseq, de acuerdo con el método de reducción de la complejidad descrita por Sansaloni *et al.* (2011). Se utilizaron dos enzimas de restricción sensibles a metilación para reducir la complejidad del genoma, una de corte específico o raro llamada *Pst*-I (CTGCAG), y una de corte frecuente conocida como *Nsp*-I (CATG). Así se generó una representación genómica de las muestras evaluadas, lo que permitiría la identificación de los marcadores.



**Figura 3.** Representación esquemática del método de reducción de la complejidad del genoma para generar representaciones genómicas (Figura tomada de Singh, 2015).

Para la digestión del ADN, se usaron las enzimas *Pst*-I y *Nsp*-I con las siguientes soluciones y concentraciones, tal como se muestra en el Cuadro 2. El Cuadro 3 presenta las soluciones y concentraciones para la ligación del ADN genómico. Con los fragmentos resultantes de la digestión se llevó a cabo la ligación de fragmentos específicos para *Pst*-I con un adaptador de sitio específico etiquetado con 96 códigos de barras diferentes. Finalmente, las mezclas de digestión y ligación de cada muestra fueron colocadas individualmente en cada pocillo de placas de 96 muestras (Eppendorf twin.tec<sup>®</sup>, PCR plate 96), en un volumen final de 20  $\mu$ l por pocillo. Las placas fueron posteriormente incubadas en termocicladores por un periodo de 3 horas a 37 °C y 20 minutos a 60 °C.

**Cuadro 2.** Soluciones y sus concentraciones para la digestión del ADN genómico.

Unidad	Cantidad	50 de RXN (µl)
H <sub>2</sub> Odd	50 µl	50.0
Amortiguador 10X para <i>Nsp</i> -I	1X	5.0
<i>Nsp</i> -I (5 U/ µl)	2.5 U/ µg de ADN	0.5
ADN genómico (0.3 µg/ µl)	1 µg	4.5
<i>Pst</i> -I (10 U/ µl)	2.5 U/ µg de ADN	0.2
NaCl (2.5 M)	50 µM	1.0

H<sub>2</sub>Odd= Agua bidestilada; NaCl= Cloruro de Sodio; µl= Microlitros; µg= Microgramo; µM=Micromolar; X= Cantidad ajustada; RXN= Mezcla; *Nsp*-I= Enzima de corte frecuente; *Pst*-I= Enzima de corte específico.

**Cuadro 3.** Soluciones y sus concentraciones para la ligación del ADN genómico.

Solución	Cantidad	Volúmenes 10 de RXN (µl)
H <sub>2</sub> Odd		5
Amortiguador de ligazón (5X )	1X	2
Adaptador <i>Nsp</i> -I (50 µM)	50 pmoles	1
Adaptador <i>Pst</i> -I (5 µM)	5 pmoles	1
Ligasa ADN T4 (1 U/ µl)	1 U	1

H<sub>2</sub>Odd= Agua bidestilada; µl= Microlitros; pmoles= Concentración molar; µM=Micromolar; X= Cantidad ajustada; U= Unidad; RXN= Mezcla; *Nsp*-I= Enzima de corte frecuente; *Pst*-I= Enzima de corte específico.

### Amplificación de fragmentos

Se tomaron 10 µl de producto de reacción de digestión / ligación y se le agregó un iniciador hacia adelante: (5'-CGCCAAATGACGACTTCTAC-3'), más un iniciador reverso: (5'-GCGCATAACGATACCACTGA-3') para la amplificación por PCR, en una mezcla como se indica en el Cuadro 4. Los iniciadores corresponden a la posición 1547 y 2050, respectivamente, del plásmido Phi y amplifican un fragmento con 523 pares base de largo. Esta mezcla fue puesta bajo el siguiente programa: 20 segundos a 94 °C, 30 segundos a 58 °C y 45 segundos a 72 °C, por 30 ciclos, adicionalmente se tuvo un ciclo inicial de 1 minuto de desnaturalización a 94 °C y al final uno de elongación de 7 minutos a 72 °C, para después dejarse en modo reposo a 10 °C durante 5 minutos en un Termociclador MJ Research (Ramse, Minnesota, USA).



**Cuadro 4.** Componentes y concentraciones para la PCR.

<b>Cantidad</b>	<b>Final</b>	<b>10 RXN (µl)</b>
ddH <sub>2</sub> O <sup>1</sup>		0.05
5X Green	1X	2.00
25 Mm mGCl <sub>2</sub> <sup>2</sup>	1.5 mM	0.60
Mezcla de dNTP (2.5 Mm)	200 µM	0.80
Iniciador F + R (1.0 µM) <sup>3</sup>	0.25 µM	2.50
Go taq ( <sup>®</sup> DNA Polymerase (5U/ µl)	0.25 U	0.05
DNA (10-50 ng/ µl)	50-100 ng	4.00

ddH<sub>2</sub>O<sup>1</sup>= Agua destilada; µl= Microlitros; ng= Nanogramo; µM=Micromolar; X= Cantidad ajustada; RXN= Mezcla; *Nsp-I*= Enzima de corte frecuente; *Pst-I*= Enzima de corte específico; dNTP= 5'-trifosfato de desoxineucleósidos; F+R; Iniciador hacia adelante + iniciador reverso.

### **Control de calidad de los productos de PCR**

Los productos de PCR fueron verificados por medio de electroforesis en gel de agarosa a una concentración de 1.2 %, mediante un campo electroforético a 120 volts/20 minutos. En este control de calidad se buscó obtener un pequeño barrido de ADN, lo que demostraba un rango de tamaño de diferentes fragmentos. Aquellos pocillos donde no hubo productos de PCR, o se detectaron fragmentos de tamaño muy pequeños (dímeros), fueron removidos del proceso y no fueron secuenciados.

### **Purificación y cuantificación**

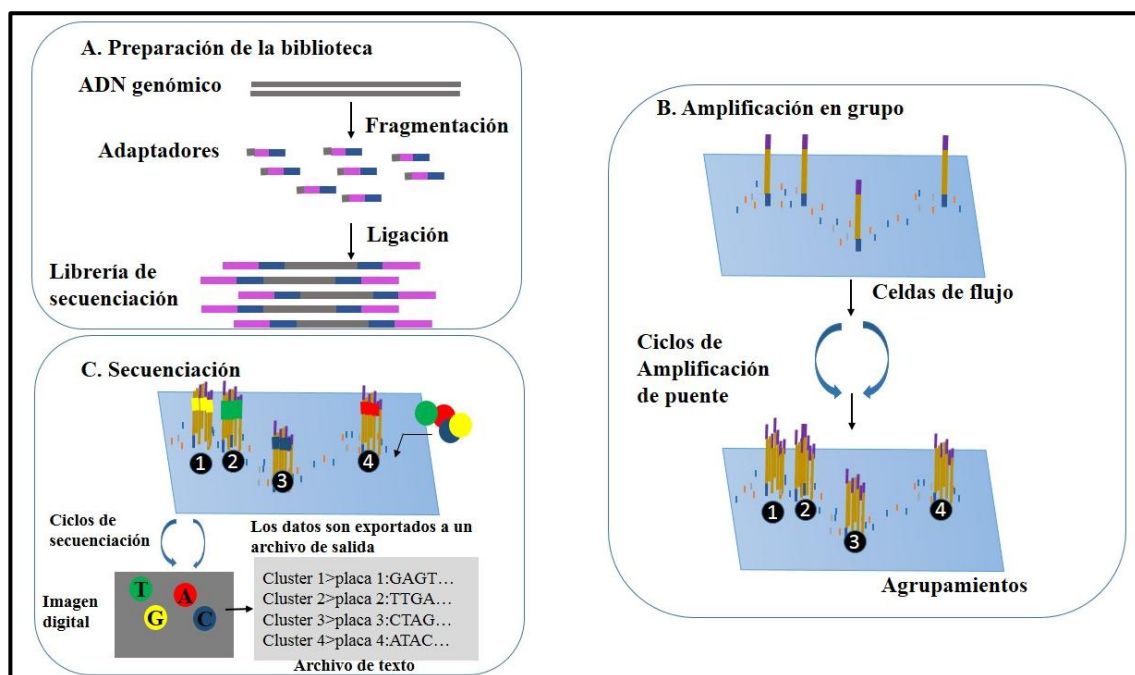
En este proceso se llevó a cabo el agrupamiento de los productos de PCR en una sola solución mezcla. Se tomaron 10 µl de producto de PCR de cada pocillo de una placa y se mezclaron en un tubo de 2 ml, este procedimiento fue realizado para cada placa y poder simplificar el manejo de los fragmentos a la hora de incorporarlos al secuenciador. Posteriormente, se llevó a cabo la purificación de los productos de PCR agrupados en cada una de las soluciones mezcla, utilizando el estuche comercial (*QIAquick PCR Purification*) de la marca QUIAGEN<sup>®</sup>. Se realizó un último control de calidad de los productos de PCR purificados, por medio de electroforesis en gel de agarosa a 1.2 % de concentración para confirmar la presencia de ADN amplificado en forma de barrido. Para este control de calidad se emplearon referencias de corridas anteriores como controles positivos y así comparar el resultado obtenido con muestras ya secuenciadas y

con resultados exitosos. Una vez que se obtuvieron los productos de PCR purificados se procedió a la cuantificación con UV, usando un espectrofotómetro NanoDrop 8000 (Thermo Scientific, EUA.), para estimar la cantidad necesaria de ADN que se introduciría al secuenciador.

### Genotipado por secuenciación

Después de generar las bibliotecas (purificadas y cuantificadas), obtenidas mediante la digestión de la enzima de restricción del ADN genómico total, la ligación del adaptador y la amplificación de fragmentos ligados al adaptador, se empleó el método de genotipado de alto rendimiento utilizando la tecnología DArTseq como lo describen Sansaloni *et al.* (2011). En cada placa de 96 pocillos se replicó el 16 % de las muestras para evaluar la reproducibilidad.

En la Figura 4, se observa una representación del método de genotipado por DArTseq, mediante la plataforma de Illumina HiSeq 2500.



**Figura 4.** Representación esquemática de la secuenciación por Illumina HiSeq 2500. (Figura tomada de Illumina Inc., 2016).

El agrupamiento de fragmentos por amplificación en puente se muestra en la Figura (4B). Los productos de amplificación de cada muestra se agruparon por placa para clonarlos por medio de PCR en un equipo cBot (SY-301-2002, Illumina, San Diego, CA, EUA), el cual utiliza un proceso conocido como amplificación por puente, esto generó agrupamientos de fragmentos (clusters) de acuerdo a la secuencia de los mismos.

En la Figura (4C) se muestra la técnica de secuenciación de fragmentos por medio de la plataforma HiSeq 2500 en SAGA-CIMMYT. Consistió en introducir los fragmentos creados a partir de la digestión del ADN en los canales de las celdas de flujo llamadas *flow-cell*. Los fragmentos templados fueron secuenciados por el extremo en que se encuentra el adaptador código de barras, el cual permitió identificar por medios bioinformáticas a que individuo pertenecen cada una de las lecturas producidas por el secuenciador.

### **Validación de datos genéticos**

#### **Identificación de marcadores**

Terminada la secuenciación, se analizó la calidad de las lecturas mediante un *primary report* que permitió revisar la calidad de las imágenes tomadas. Luego se realizó el filtrado o eliminación de los sitios de corte de las enzimas, *Pst*-I (CTGCAG) y *Nsp*-I (CATG), y del código de barras que le asigna a cada fragmento su pertenencia a cada individuo.

Posteriormente, se realizó la identificación de los SNPs presentes en las poblaciones evaluadas a través del sistema analítico DArTsoft 14 desarrollado por la empresa DArT. Se obtuvieron como resultado marcadores SNP y marcadores Silico-DArT como datos binarios (1/0), indicando la variación de presencia/ausencia (PAV) del fragmento. Para este estudio se utilizaron solamente los polimorfismos codominantes (SNPs). La particularidad de esta tecnología es que el llamado de SNPs es independiente del genoma de referencia, sino que se realiza con base en una biblioteca propia, formada por todos los fragmentos producidos por este método para diferentes germoplasmas de este cultivo. Por eso, la conveniencia de la utilización de DArTseq para estudios de

diversidad, la cual estaría mejor representada con base en varios germoplasmas diferentes, y no en el genoma de una muestra solamente.

### Tamizaje de datos

Para determinar la calidad de los marcadores SNPs a ser empleados en el análisis, se realizaron varios filtros para identificar aquellos con mejores propiedades, que fueron tomados en cuenta para caracterizar y calificar a estos polimorfismos.

- La frecuencia alélica fue estimada con el objetivo de obtener la proporción en la cual cada alelo marcador está presente con relación al total de marcadores, usando el paquete estadístico R de un script propio del CIMMYT.
- El porcentaje de datos perdidos fue otro factor importante a tener en cuenta para eliminar aquellos marcadores con una alta tasa de valores no reconocidos por el sistema de análisis. Con la librería *adegenet* en R se determinó el porcentaje de datos perdidos presente en los SNPs identificados.
- Usando Microsoft Excel 2010 se obtuvo: el porcentaje de reproducibilidad, realizando un reordenamiento de menor a mayor y eliminando aquellos que tenían menos del 95% de reproducibilidad. La tasa de llamada de los SNPs también se ordenó de menor a mayor, descartando aquellos valores menores del 75% y el Contenido de Información Polimórfica (PIC) de los 53,471 SNPs seleccionados se calcularon con el fin de conocer el grado de información que brinda un determinado marcador genético empleando la siguiente fórmula (Botstein *et al.*, 1980).

$$PIC_I = 1 - \sum_{j=1}^{n-1} P^2_{ij} - \sum_{j=1}^{n-1} \sum_{k=j+1}^n 2P^2_{ij} P^2_{ik}$$

Dónde:  $P^2_{ij}$  y  $P^2_{ik}$  son las frecuencias alélicas del  $j$ -ésimo y el  $k$ -ésimo alelo del marcador  $i$ , respectivamente, y la sumatoria se extiende sobre todos los alelos. Este índice hace referencia a la información de un marcador en la población de acuerdo con las frecuencias de los alelos.

## Análisis de diversidad genética

Los parámetros de diversidad genética se obtuvieron con el software BIO-R (Análisis de biodiversidad con R para Windows) Versión 2.0 (Pacheco *et al.*, 2016), en el cual se estimó:

### Heterocigosidad

Este parámetro se puede estudiar como heterocigosidad observada (Ho) y esperada (He).

La (Ho) se calculó para cada población, como la media inter-loci de la frecuencia genotípica de los individuos heterocigotos mediante la expresión:

$$H_o = \frac{1}{L} \sum_i^L \frac{(N_{heterocigotos})_i}{N_i}$$

Dónde:  $L$  es el número de loci analizado en población objeto de estudio,  $(N_{heterocigotos})_i$  es el número absoluto de individuos heterocigotos para el locus  $i$  y  $N_i$  es el número total de individuos analizados para el locus  $i$ .

La (He) de cada población es la media de los valores obtenidos para cada locus en dicha población y se calculó usando la siguiente fórmula:

$$H_e = \frac{1}{L} \sum_{j=1}^{nL} h_j$$

Dónde:  $h_j$  es la heterocigosidad por locus y  $L$  es el número total de loci. Los valores de  $H_e$  están entre 0 y 1 y un mínimo de 30 loci para ser analizados en 15 individuos por población, para reducir el riesgo de sesgo estadístico.

### Número de alelos efectivos

$$A_e = \sum_{l=1}^L \frac{1}{1 - h_l} = \sum_{l=1}^L \frac{1}{\sum P^2_i}$$

Dónde:  $p^2i$  es la frecuencia en el alelo  $i$ -ésimo en un locus y  $hl = \sum p^2i$  es la heterocigosidad en un locus.

### Índice de Shannon

$$SH = - \sum_{a=1}^A p^{\wedge}_a \log(p^{\wedge}_a)$$

Dónde:  $p^{\wedge}_a$  es la frecuencia estimada del alelo  $a$  en toda la muestra y  $A$  es el número de alelos en la muestra.

### Número de alelos raros

$$R_j = \sum_{i=1}^k p_{ij} S_i$$

Dónde:  $k$  es el número total de alelos,  $p_{ij}$  la frecuencia alélica del  $i$ -ésimo alelo dentro de la accesión  $j$ .

### Coefficientes de divergencia genética

En el estudio de la diversidad del maíz se usaron caracteres fenotípicos y genotípicos, los cuales se analizaron con técnicas multivariadas de componentes principales y del análisis de conglomerados (Rohlf, 2012; Team, 2015).

Los valores fenotípicos estandarizados de los quince caracteres se utilizaron para calcular un coeficiente de distancia fenotípica, obtenido por el complemento del coeficiente de correlación de Pearson entre pares de poblaciones ( $d_{ij} = 1 - r$ ), para obtener la matriz de distancias fenotípicas (Rohlf, 2012).

A partir de los polimorfismos de un solo nucleótido (SNP) se estimaron las frecuencias alélicas presentes en cada una de las poblaciones de maíz. Posteriormente, las frecuencias alélicas se utilizaron para computar la matriz de distancia genética de Rogers modificada (Wright, 1978), la cual se determinó mediante el software BIOR, a partir de la siguiente fórmula:

$$MR_{xy} = \frac{\sqrt{\sum_{l=1}^L \sum_{a=1}^{nl} (plax - play)^2}}{2L}$$

Dónde:  $plax$  es la frecuencia estimada del alelo  $a$ , dentro del locus  $l$ , en el genotipo  $x$ ;  $L$  es el número de loci;  $nl$  es el número de alelo dentro del locus;  $play$  es la frecuencia estimada del alelo  $a$ , dentro del locus  $l$ , en el genotipo  $y$  (Pacheco *et al.*, 2016).

Con las matrices de distancias fenotípicas y genéticas se realizó un análisis gráfico y de correlación a través del procedimiento de comparación de matrices MXCOMP de NTSYS-pc (Rohlf, 2012).

### **Ordenamiento de poblaciones nativas de maíz**

Se realizó la transformación de las matrices de distancias fenotípicas y genéticas en forma de producto escalar para obtener sus valores propios y vectores característicos dando como resultado un análisis de coordenadas principales, o componentes principales, con el software NTSYS-pc (Rohlf, 2012). En cada caso, se usaron los dos primeros componentes principales (CP1 y CP2) para representar las poblaciones en un análisis de dispersión gráfica.

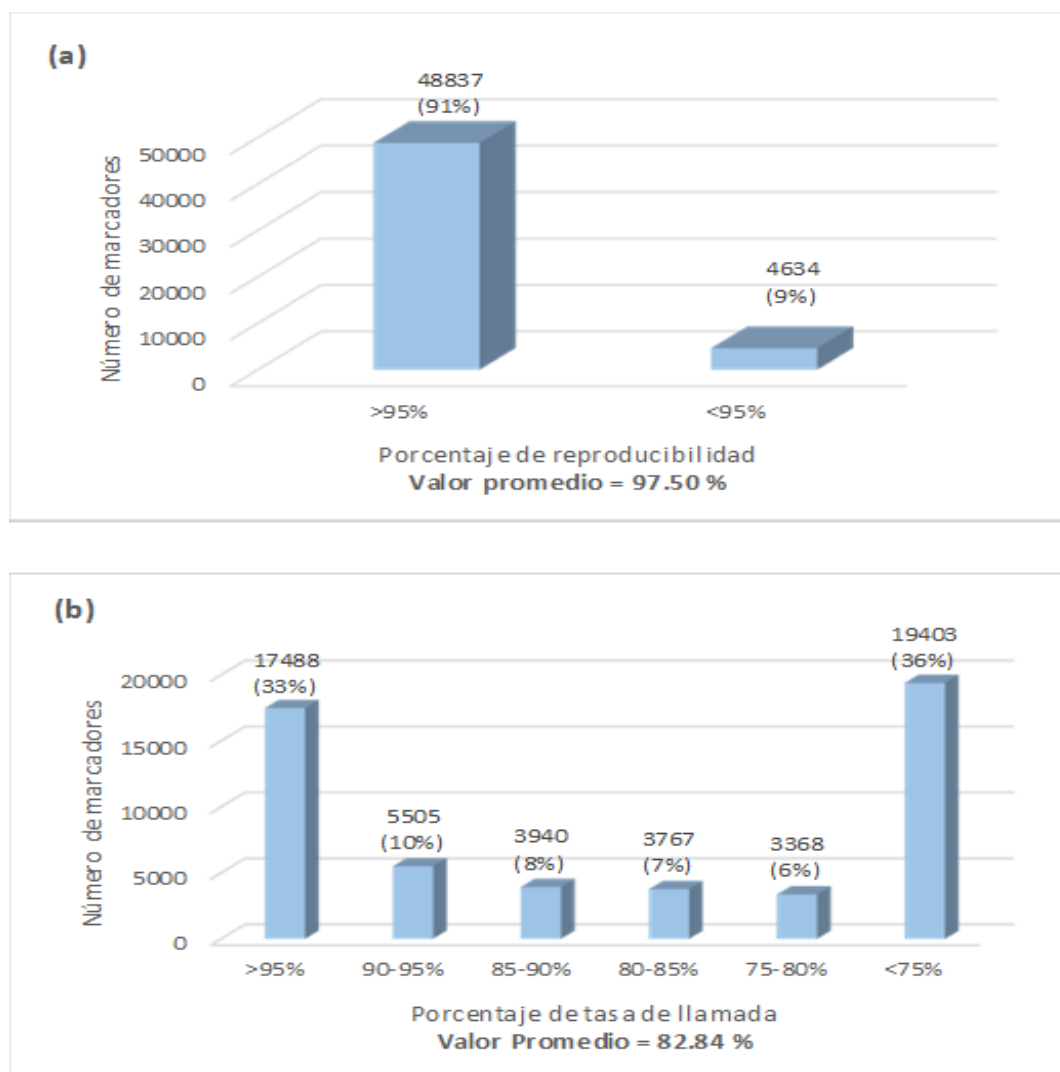
### **Relación entre poblaciones de maíz**

Para resumir las relaciones entre todas las poblaciones examinadas, a partir de la matriz de distancias fenotípicas se realizó un análisis de conglomerados jerárquico, donde se obtuvo un dendrograma, implementando el algoritmo *hclust* del paquete estadístico R (<http://cran.r-project.org>) (Team, 2015), a través del método de grupo de pares no ponderados utilizando promedios aritméticos (UPGMA). Con la matriz de distancias genéticas se realizó un análisis de conglomerados usando el método de Ward (Ward, 1963). Este análisis fue realizado con el software BIO-R (Pacheco *et al.*, 2016).

## RESULTADOS Y DISCUSIÓN

### Análisis exploratorio de marcadores SNP

Se identificaron 73,834 marcadores codominantes, polimorfismos de un solo nucleótido (SNP), que fueron sometidos a un proceso de filtrado de calidad. Primeramente, fueron eliminados los marcadores monomórficos, y se utilizaron la tasa de llamada de los SNPs y la Reproducibilidad o Consenso entre los datos generados para una muestra y su réplica técnica. De los datos obtenidos, se seleccionaron 53,471 como marcadores polimórficos (Figura 5).



**Figura 5.** Distribución de marcadores SNPs en las poblaciones de maíz. (a) Reproducibilidad de los SNPs; (b) Tasa de llamada de los SNPs.

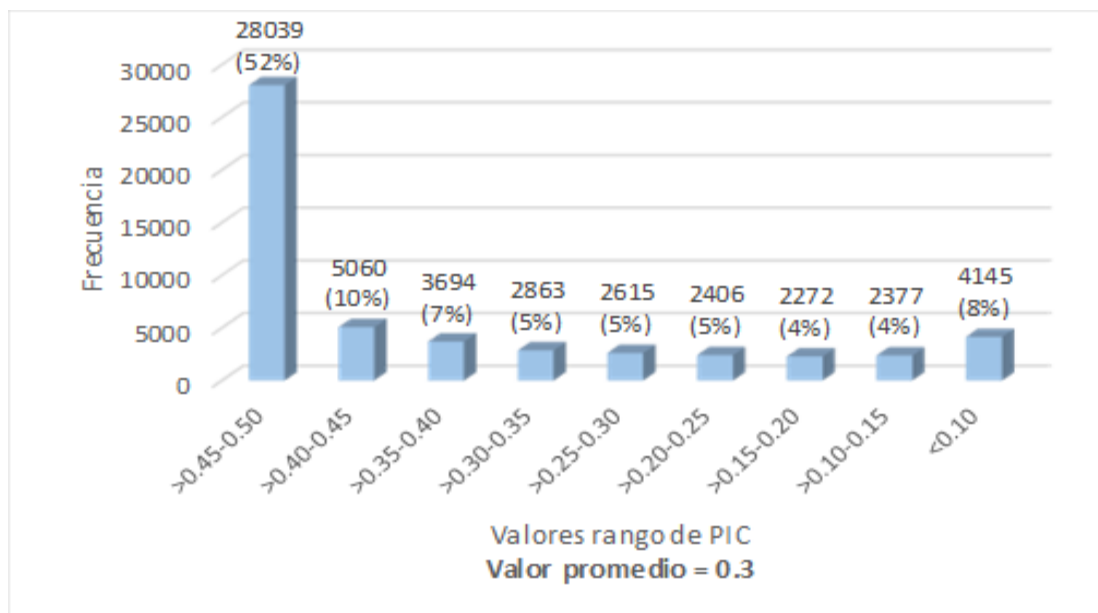


Se encontró un promedio de reproducibilidad del 97.50 %, y alrededor del 91% (48,837) de los marcadores SNP mostraron valores superiores o igual a 95% de reproducibilidad (Figura 5a). La tasa de llamada exhibió un promedio de 82.84% (Figura 5b).

La reproducibilidad y tasa de llamadas fueron también los criterios usados por Yan *et al.* (2009) al evaluar el valor de los marcadores en un estudio donde caracterizaron un panel de diversidad de 154 líneas endogámicas de maíz. Estos autores reportaron una tasa de llamadas del 92 % y una reproducibilidad del 98%. Así mismo al analizar un total de 2,815 accesiones de maíz con marcadores SNP distribuidos por todo el genoma, mediante genotipado por secuenciación (GBS), la cobertura promedio (Tasa de llamadas SNP) reportados por Romay *et al.* (2013) fue del 35% por muestra, con valores que van del 2 al 75%. El porcentaje significativamente mayor encontrado en el presente estudio, se debe a la tasa de nombres o de valores (1/0) de los individuos detectados y la cantidad de datos que coinciden entre una muestra y su réplica. Por lo tanto, los resultados indicaron que la matriz de genotipado DArT generó un número mayor de marcadores polimórficos para estudios de diversidad de población, estos datos sugieren la solidez de la puntuación suficiente para evaluar con precisión las relaciones genéticas (Wenzl *et al.*, 2004; Sansaloni *et al.*, 2010).

La reproducibilidad y tasa de llamadas son dos parámetros de calidad para evaluar a los marcadores moleculares para estudios genómicos posteriores. Sansaloni *et al.* (2010) utilizaron criterios de umbral de reproducibilidad >97% y una tasa de llamadas >80% para la selección de marcadores polimórficos. La tasa de llamadas obtenida en el presente estudio (Figura 5b) se encuentra dentro de los parámetros mencionado anteriormente. Lo anterior indica que los marcadores evaluados en este trabajo son considerados polimórficos.

La distribución de los marcadores SNP obtenidos mediante DArTseq de acuerdo con los valores para el contenido de información polimórfica (PIC) se muestra en la Figura 6.



**Figura 6.** Distribución de frecuencia de los contenidos de información de polimorfismo (PIC) de los marcadores SNP.

Los marcadores SNPs mostraron una amplitud en los valores de PIC entre 0.10 y 0.50 (Figura 6). Los grupos con rango entre 0.40-0.45 y 0.45-0.50 agruparon el 62 % de los marcadores identificados, el resto de las clases presentaron valores entre 4 y 8 %, mientras que el grupo de 0.10-0.15 y 0.15-0.20 únicamente exhibieron el 8 % del total de los marcadores, lo que indica el valor relativo de cada marcador con respecto a la cantidad de polimorfismo exhibido (Wu *et al.*, 2016).

El valor promedio de PIC encontrado fue de 0.3 (Figura 6), esto fue ligeramente superior a un estudio previo por Wu *et al.* (2014) quienes reportaron valores de promedio de PIC de 0.29 en 41,819 marcadores polimórficos, al caracterizar 367 líneas endogámicas de maíz, utilizando polimorfismos de un solo nucleótido (SNP). Mientras tanto, Zhang *et al.* (2016), reportaron un valor de PIC mayor a 0.35. Esto se deduce que los marcadores identificados en el estudio son polimórficos, confiables y al igual informativos que los SNPs reportados en las publicaciones mencionadas.

El valor del PIC, indica que tan informativo puede ser un marcador. En este trabajo el valor de PIC se mantuvo por encima de 0.5. Botstein *et al.* (1980) establecieron rangos para los valores de PIC con base en las frecuencias alélicas que se presentan en cada locus como altamente informativos (PIC > 0.5), razonablemente informativos (PIC < 0.5

pero  $> 0.25$ ) y ligeramente informativos ( $PIC < 0.25$ ). Por lo tanto, los marcadores obtenidos en este trabajo son considerados como razonablemente informativos.

### Índices de diversidad genética

Los resultados del análisis de diversidad genética de las 76 poblaciones, representada por los índices de diversidad, heterocigosidad esperada ( $H_e$ ), heterocigosidad observada ( $H_o$ ), número de alelos efectivos ( $A_e$ ), índice de Shannon ( $SH$ ) y número de alelos raros ( $R_j$ ) se presentan en el Cuadro 5.

**Cuadro 5.** Índices de diversidad genética de las 76 poblaciones de maíz secuenciados.

<b>Genotipo</b>	<b><math>H_e</math></b>	<b><math>H_o</math></b>	<b><math>A_e</math></b>	<b><math>SH</math></b>	<b><math>R_j</math></b>
48CN <sup>†</sup>	0.445 **	0.507 **	1.803 **	0.920 **	185.003 **
138CN	0.445 **	0.505 **	1.801 **	0.919 **	182.659 **
143CN	0.445 **	0.502 **	1.801 **	0.919 **	181.557 **
52CN	0.445 **	0.502 **	1.801 **	0.919 **	183.328 **
50CN	0.445 **	0.497 **	1.801 **	0.919 **	181.786 **
141CN	0.444 **	0.516 **	1.800 **	0.918 **	182.625 **
156CN	0.443 **	0.521 **	1.796 **	0.917 **	182.696 **
47CN	0.443 **	0.485 **	1.795 **	0.916 **	181.871 **
45CN	0.443 **	0.496 **	1.795 **	0.916 **	181.972 **
46CN	0.442 **	0.495 **	1.794 **	0.915 **	181.593 **
49CN <sub>x</sub> TN	0.442 **	0.506 **	1.794 **	0.915 **	179.945 **
44EC	0.442 **	0.489 **	1.794 **	0.915 **	180.849 **
139CN	0.442 **	0.499 **	1.793 **	0.915 **	180.281 **
54CN	0.442 **	0.508 **	1.792 **	0.914 **	178.996 **
136CN	0.441 **	0.502 **	1.790 **	0.914 **	180.197 **
53CN	0.441 **	0.504 **	1.790 **	0.913 **	178.127 **
137CN	0.441 **	0.492 **	1.789 **	0.913 **	177.941 **
140CN	0.441 **	0.497 **	1.789 **	0.913 **	178.346 **
51CN	0.441 **	0.492 **	1.788 **	0.913 **	179.299 **
88CN	0.439 **	0.488 **	1.784 **	0.911 **	175.784 **
57CN	0.438 **	0.524 **	1.778 **	0.908 **	173.168 **
158CN	0.437 **	0.518 **	1.775 **	0.907 **	172.053 **
67CN	0.436 **	0.484 **	1.773 **	0.906 **	172.082 **
142CN	0.436 **	0.496 **	1.772 **	0.905 **	171.516 **
58CN	0.436 **	0.524 **	1.772 **	0.905 **	172.855 **
61CN	0.435 **	0.503 **	1.770 **	0.904 **	169.265 **
56CN	0.435 **	0.519 **	1.769 **	0.903 **	171.775 **
55CN	0.435 **	0.504 **	1.769 **	0.903 **	171.067 **
151EO	0.435 **	0.498 **	1.768 **	0.903 **	170.054 **

60CN	0.434 **	0.496 **	1.766 **	0.902 **	170.059 **
71CN	0.433 **	0.476 **	1.764 **	0.901 **	167.047 **
64CNxR	0.430 **	0.524 **	1.756 **	0.897 **	167.028 **
63CN	0.430 **	0.519 **	1.756 **	0.897 **	167.865 **
72CNxTN	0.430 **	0.499 **	1.754 **	0.897 **	166.222 **
153EO	0.430 **	0.485 **	1.754 **	0.896 **	167.047 **
135EC	0.428 *	0.482 **	1.749 *	0.894 *	163.593
159CN	0.427 *	0.481 *	1.745	0.892	163.850
91EC	0.427 *	0.496 **	1.745	0.892	163.969
62R	0.422	0.499 **	1.729	0.884	158.296
155RxTN	0.420	0.469	1.725	0.882	154.872
66R	0.420	0.467	1.725	0.882	157.287
154TN	0.420	0.473	1.723	0.881	155.021
65TN	0.419	0.476	1.722	0.880	155.267
134C	0.418	0.472	1.717	0.878	154.015
157RxC	0.417	0.469	1.716	0.877	153.253
152C	0.417	0.468	1.715	0.876	151.473
73R	0.417	0.452	1.714	0.876	152.242
145TN	0.415	0.449	1.711	0.874	156.351
90R	0.415	0.475	1.710	0.874	153.207
146R	0.415	0.457	1.710	0.874	151.277
68RxTN	0.415	0.451	1.710	0.874	152.101
105R	0.415	0.451	1.709	0.874	150.636
77TN	0.414	0.472	1.708	0.873	155.333
104TN	0.414	0.474	1.707	0.873	150.354
109TN	0.414	0.460	1.705	0.872	150.016
147C	0.413	0.457	1.705	0.871	151.039
103R	0.413	0.480	1.705	0.871	152.139
76TN	0.413	0.409	1.704	0.871	149.619
70RxCN	0.413	0.464	1.704	0.871	154.392
108TN	0.413	0.455	1.703	0.870	149.656
69R	0.412	0.423	1.702	0.870	148.624
74RxTN	0.412	0.410	1.701	0.869	149.343
106OxR	0.412	0.437	1.700	0.869	152.561
148R	0.412	0.406	1.700	0.869	148.970
75RxO	0.412	0.415	1.700	0.869	149.136
86R	0.411	0.427	1.699	0.868	149.699
144R	0.411	0.421	1.699	0.868	149.413
78TN	0.411	0.420	1.698	0.868	148.112
87T	0.411	0.419	1.698	0.867	149.360
150TNxC	0.411	0.449	1.697	0.867	149.995
107RxTN	0.410	0.434	1.696	0.866	150.935
149CxR	0.410	0.418	1.696	0.866	148.054

79R	0.409	0.415	1.693	0.865	148.334
59TN	0.409	0.447	1.692	0.865	155.500
85R	0.409	0.422	1.692	0.865	147.207
133R	0.408	0.420	1.689	0.863	146.509
<b>Media</b>	0.426	0.474	1.743	0.890	163.355
<b>Máximo</b>	0.445	0.524	1.803	0.920	185.003
<b>Mínimo</b>	0.408	0.406	1.689	0.863	146.509
<b>EE</b>	0.001	0.004	0.005	0.002	1.479

\* , \*\* = Poblaciones con mayor diversidad genética con base  $\mu+EE$  y  $\mu+2EE$ , respectivamente; † Las letras indican la identificación de cada raza y algunas combinaciones entre ellas: C= Celaya, O= Olotillo, CN= Cónico Norteño, R= Ratón, EC= Elotes Cónicos, EO= Elotes Occidentales, T= Tuxpeño y TN= Tuxpeño Norteño y los dígitos, el número asignado a la colecta; He= Heterocigosidad esperada; Ho= Heterocigosidad observada; Ae= Número de alelos efectivos; SH= índice de Shannon; R<sub>j</sub>= Número de alelos raros; EE= Error estándar de la media.

De acuerdo con Nei (1987) la heterocigosidad esperada (He), es la probabilidad de que un individuo sea heterocigoto para un loci. La amplitud de He encontrada entre el total de poblaciones analizadas osciló entre 0.408 y 0.445, con un valor medio de 0.426 (Cuadro 5). Se puede dilucidar del Cuadro 5, la relación entre los índices de diversidad, y que, en conjunto, indican que, de la muestra analizada, los maíces procedentes del área de transición y altura (Cónico Norteño y Elotes Cónicos) muestran mayor diversidad genética que los maíces adaptados a áreas intermedias a bajas (Ratón, Tuxpeño Norteño) (Cuadro 1).

Lo anterior difiere ligeramente con un estudio previo por Vega-Álvarez *et al.* (2017), quienes reportaron un promedio de He=0.49, usando 31 loci de microsatélites, al estimar la diversidad genética de 107 poblaciones perteneciente a nueve razas nativas de maíz del noroeste de México. Estos autores confirmaron que razas de maíz del noroeste de México albergan una amplia diversidad genética tanto entre como dentro de sus poblaciones.

En el estudio de poblaciones de Valles Altos de México con base en microsatélites, Rocandio-Rodríguez *et al.* (2014) reportan un valor medio de He=0.710, valor superior a los encontrados en el presente estudio (0.408 – 0.445), lo cual puede deberse a la diferencia en la técnica utilizada y en la cantidad de polimorfismos encontrados.

Al caracterizar genéticamente 367 accesiones de maíz, en su mayoría pertenecientes a la región templada, y utilizando genotipado SNP de alto rendimiento, Wu *et al.* (2014) reportaron un valor de  $He=0.364$ , el cual fue ligeramente inferior con lo que se reporta en el presente estudio, indicando que los valores de  $He$  variaron a través de diferentes regiones genéticas incluso en el mismo cromosoma. Los mismos autores mencionan que los resultados obtenidos implicaron que diferentes regiones genéticas habían sufrido diferentes presiones de selección durante los procesos de reproducción.

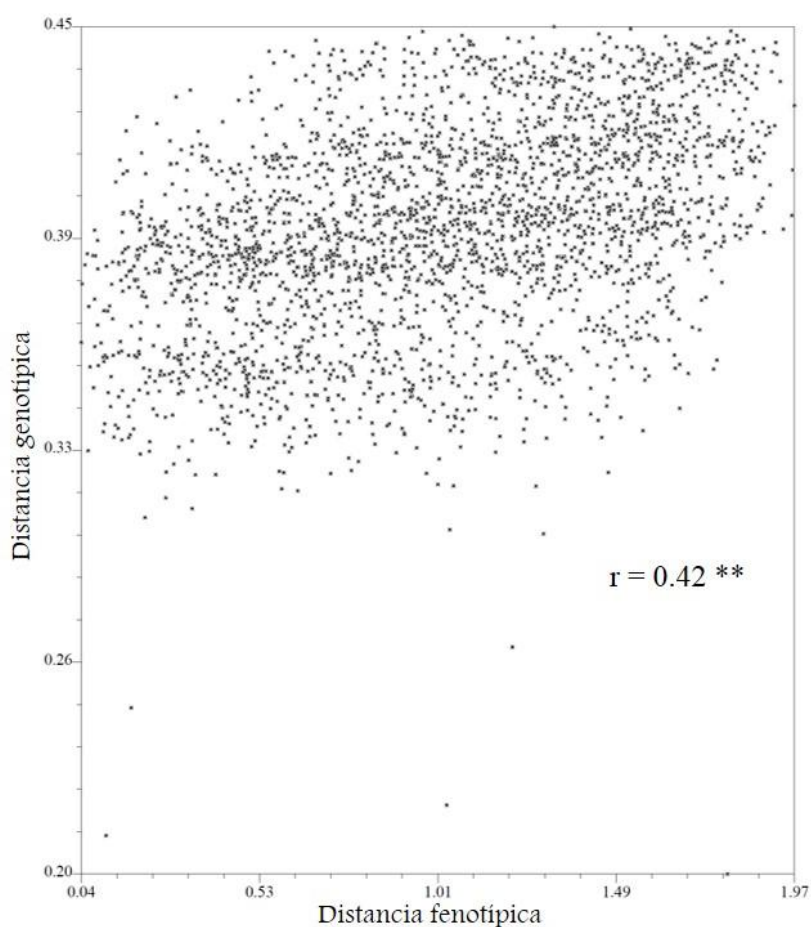
Las poblaciones de Olotillo, Celaya, Ratón, Tuxpeño y Tuxpeño Norteño, procedentes de áreas de adaptación intermedias, presentaron una diversidad genética inferior, comparada con los materiales de Transición-Altura (Cuadro 5). En un estudio donde se incluyen estos complejos raciales, al evaluar la variación genómica de 46 poblaciones locales de maíces nativos de México usando marcadores SNP, Arteaga *et al.* (2015), reportan una amplitud de  $He=0.234$  a  $0.318$  (media  $0.311$ ). Sánchez *et al.* (2000) mencionan que valores bajos de  $He$  en accesiones de maíces que son de uso especial, y se utilizan para palomitas de maíz, maíz dulce, pozole, pinole o tamales. Estos mismos autores reportan un valor promedio de  $He=0.212$  e indican que valores bajos de diversidad están presentes en las poblaciones de maíz que estudiaron y que dichos valores son muy similares a los que se encuentran en cultivos de especies con altos niveles de endogamia.

Los bajos valores de heterocigosidad esperada ( $He$ ) encontrados en este presente estudio, en comparación con lo reportado por Matsuoka *et al.* (2002) y Rocandio-Rodríguez *et al.* (2014) al analizar la diversidad genética de maíz, valorados por SSR, encontraron en promedio una  $He= 0.73$  y  $0.71$ . Aranguren-Méndez *et al.* (2005), mencionan que la máxima heterocigosidad esperada para cada polimorfismo de un solo nucleótido (SNP) es de solamente el 50%, siendo por lo tanto menos informativos que los microsatélites (SSR), las cuales generalmente presentan múltiples alelos y sus valores de heterocigosis superan el 70%, lo que significa que esta diferencia de resultados encontrados, posiblemente se deba a la técnica y el tipo de marcador utilizado.

### Coefficientes de divergencia genética

En este estudio se analiza la utilidad de los polimorfismos de un solo nucleótido (SNP) en la clasificación de poblaciones de maíz como estrategia para estudiar las relaciones genéticas a partir de una muestra de individuos.

La estimación de divergencia genética obtenida con 15 caracteres fenotípicos (coeficiente de distancia fenotípica) y la estimación de divergencia genética a partir de las frecuencias alélicas de 53,471 SNPs (distancia genética) se compararon con base en el coeficiente de correlación entre ambas matrices (Figura 7).



**Figura 7.** Dispersión grafica de los valores de distancia genética y fenotípica de los pares posibles entre 76 poblaciones de maíz.

Con excepción de cuatro puntos fuera del patrón general, la Figura 7 muestra una asociación positiva, que, de acuerdo con el análisis de correlación se obtuvo un valor de

$r=0.42^{**}$ . El análisis de los datos morfológicos a través de 15 caracteres cuantitativos de la mazorca y de grano se usaron para describir la diversidad genética del maíz en la región del sureste de Coahuila, México (Rincón-Sánchez y Ruiz-Torres, 2018).

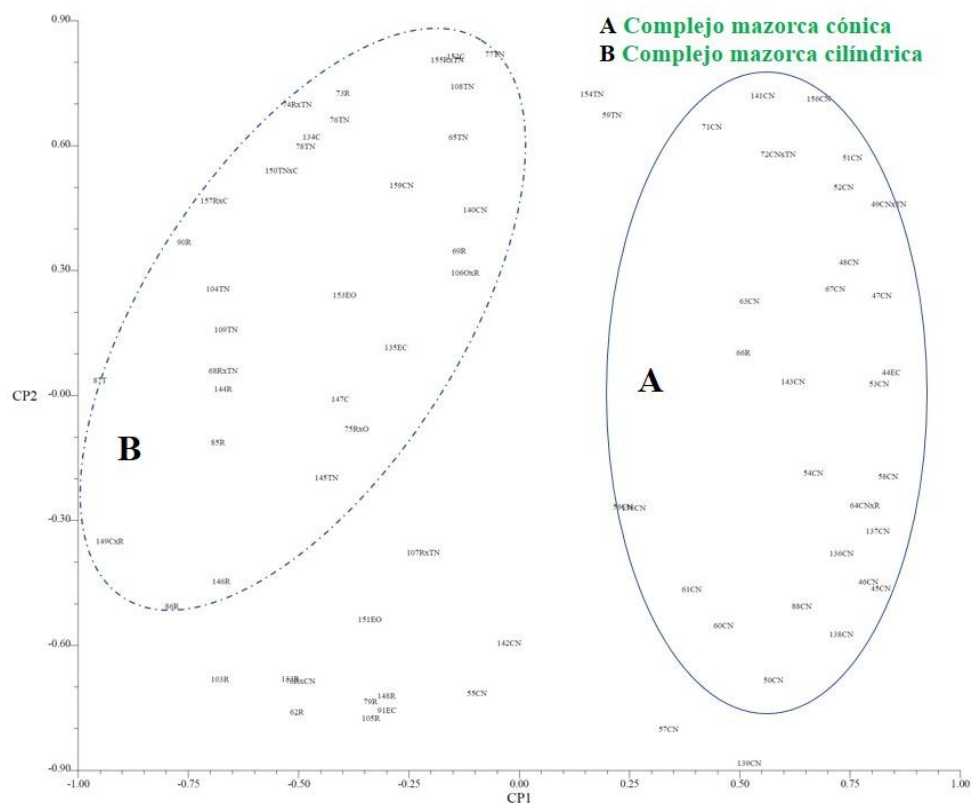
Las correlaciones significativas indican que estos dos conjuntos independientes de datos reflejan en gran medida el mismo patrón de diversidad genética y validan el uso de estos datos para calcular las diferentes estadísticas de diversidad para accesiones de maíz en la región de estudio. Beyene *et al.* (2005) encontraron correlación significativa entre la matriz de distancias morfológicas con las matrices de distancias utilizando marcadores moleculares SSR y AFLP con valores de  $r=0.43^{**}$  y  $r=0.39^{**}$ , respectivamente; en tanto que la correlación entre las matrices de distancias SSR y AFLPs fue de  $r=0.67^{**}$ . Sin embargo, Ristic *et al.* (2013) encontraron correlaciones bajas y no significativas entre marcadores moleculares y caracteres morfológicos.

Los bajos coeficientes de correlación entre los datos fenotípicos y genéticos, aun cuando son los mismos genotipos, son explicados, ya que son valores diferentes obtenidos a diferente nivel. Por un lado, los datos fenotípicos explican la expresión de cada genotipo expuesto a las fluctuaciones ambientales y su interacción, representado en la mayoría de los casos por pocos caracteres. En lado opuesto, los datos genéticos provienen de muestras de ADN, y por lo tanto, explican solo la contribución genética, independientemente de los factores ambientales. En el presente estudio, los datos fenotípicos representados por 15 caracteres cuantitativos de la mazorca y grano se comparan con 53, 471 polimorfismos de un solo nucleótido (SNP).

### **Ordenamiento de poblaciones nativas**

Las relaciones entre las poblaciones en estudio se analizan con base en el Análisis de Componentes Principales (ACP), obtenido a partir de las matrices de distancias fenotípicas y genéticas. El diagrama de dispersión de las 76 poblaciones de maíz con base en los datos morfológicos se presenta en la Figura 8.



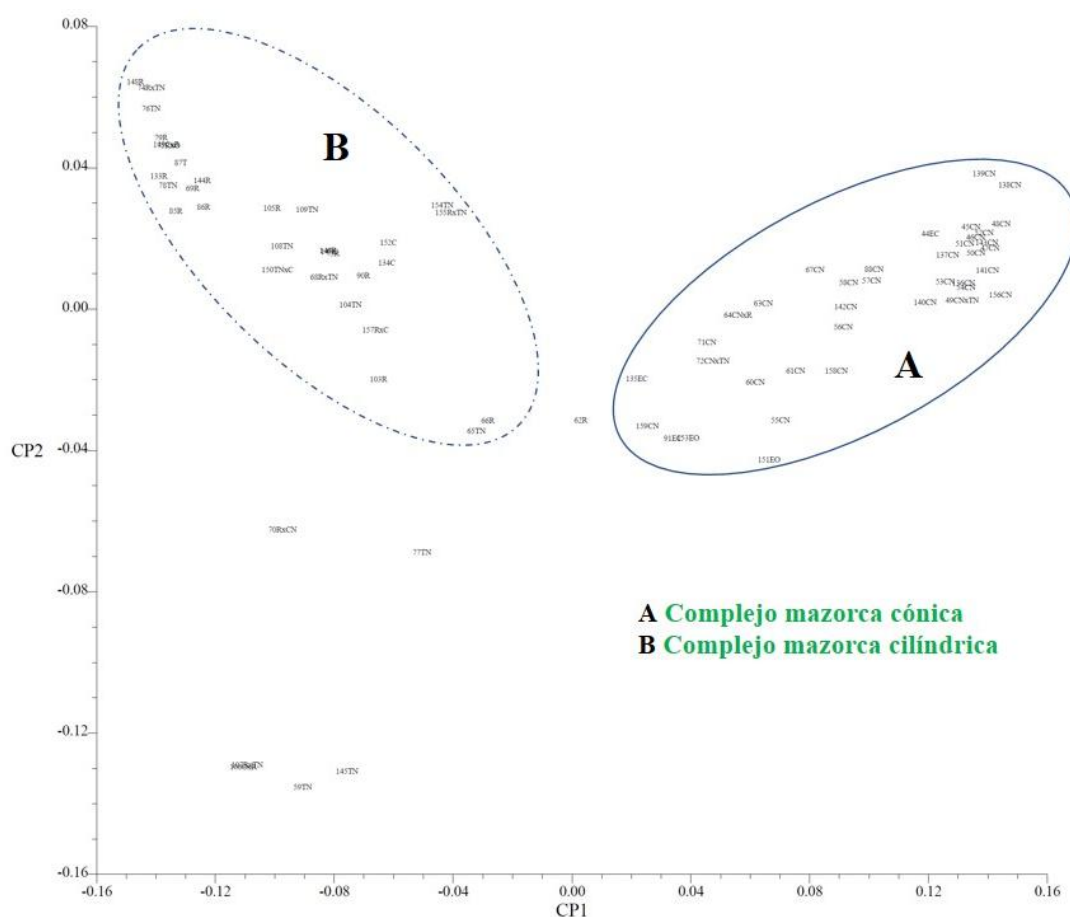


**Figura 8.** Dispersión gráfica de 76 poblaciones de maíz nativo con base en los dos primeros componentes principales (CP1 y CP2), obtenidos de datos morfológicos. En las poblaciones se indica el número de entrada; las letras, la identificación de cada raza: C: Celaya; CN: Cónico Norteño; EC: Elotes Cónicos; EO: Elotes Occidentales; O: Olotillo; R: Ratón; T: Tuxpeño; TN: Tuxpeño Norteño.

Se puede identificar en la Figura 8 dos grandes grupos: el primero formado por poblaciones de maíz adaptadas a las áreas de transición y a los valles altos (A), representado por las razas Cónico Norteño y Elotes Cónicos (Cuadro 1), identificado como Complejo mazorca cónica; el segundo, localizado en el lado opuesto en el primer componente (CP1), constituido por poblaciones procedentes de áreas intermedias de las razas Ratón y Tuxpeño Norteño (Cuadro 1), identificado como Complejo mazorca cilíndrica (B). Adicionalmente se identifican poblaciones que de acuerdo con los caracteres morfológicos pueden representar factores de interacción entre los dos grupos. Estos resultados indican que, los caracteres de la mazorca y grano incluidos en el estudio diferencian a las poblaciones por su adaptación a áreas intermedias y de transición y altura como lo señalan Rincón-Sánchez y Ruiz Torres (2018).

Las distinciones entre poblaciones de maíz dentro de estos dos complejos en la región de estudio, demuestra la utilidad de los caracteres de mazorca para describir la diversidad del maíz como lo señala Goodman y Paterniani (1969).

La dispersión gráfica de las poblaciones de maíz con base en los datos genéticos se presenta en la Figura 9.



**Figura 9.** Dispersión gráfica de 76 poblaciones de maíz nativo con base en los dos primeros componentes principales (CP1 y CP2), obtenidos de datos genéticos. En las poblaciones se indica el número de entrada; las letras, la identificación de cada raza: C: Celaya; CN: Cónico Norteño; EC: Elotes Cónicos; EO: Elotes Occidentales; O: Olotillo; R: Ratón; T: Tuxpeño; TN: Tuxpeño Norteño.

Los resultados basados en SNP distinguieron dos grupos bien definidos: el primero identificado como Complejo mazorca cónica, compuesta por poblaciones procedentes de áreas de Transición-Altura de la raza Cónico Norteño y Elotes Cónicos; el segundo distribuido en el extremo superior izquierdo de la Figura 9, constituido por poblaciones de las razas Ratón y Tuxpeño Norteño, identificados como Complejo mazorca

Cilíndrica, con base a la altitud de procedencia, este complejo se encuentra distribuido en zonas Bajas-Intermedias del sureste de Coahuila, México (Cuadro 1).

Por la dispersión de cada uno de los grupos, se puede detectar que en función del área de altitud existe la agrupación de cada población evaluada. Los maíces, Cónico Norteño y Elotes Cónicos son poblaciones que se adaptan muy bien a tierras altas, altitudes superiores a 1700 msnm, mientras que las razas Ratón y Tuxpeño Norteño se encuentran a altitudes de 84 a 1300 msnm (Ruíz *et al.*, 2008). Los resultados indican que, para ambos grupos, tanto para el Complejo mazorca cónica como para el de mazorca cilíndrica existe una alta variación y heterogeneidad debido a la diversidad genética y consecuente diversidad racial, y a la variación en las condiciones ambientales de su procedencia (Cuadro 1), tal como ha sido reportado en el estudio de los maíces nativos distribuidos en el estado de Coahuila (Rincón *et al.*, 2010).

Las poblaciones clasificadas como Ratón y Tuxpeño Norteño, están distribuidas a lo largo del grupo B en ausencia de un patrón que las diferencie (Figura 9), lo cual contrasta con las relaciones encontradas en estudios de la diversidad usando caracteres morfológicos (Rincón *et al.*, 2010; Rincón-Sánchez y Ruiz-Torres, 2018). En un estudio realizado por Wen *et al.* (2012), al caracterizar genéticamente mediante 1,433 marcadores SNP, 321 accesiones de la raza Tuxpeño junto con otras variedades locales, concluyeron que la agrupación se debe a una diversidad de genes y heterocigosidad relativamente alta, además de las ubicaciones geográficas donde se recolectaron las accesiones de Tuxpeño que muestran un amplio rango climático.

Es importante señalar que, dentro de la agrupación Complejo mazorca cónica (Figura 9), se incluye a las poblaciones 151EO y 153EO clasificadas como Elotes Occidentales, agrupadas entre los 1001 y 1800 m (Cuadro 1). Elotes Occidentales es una subraza de Harinoso de ocho, tipo de mazorca cilíndrica, de ocho hileras, de tipo de grano harinoso y de acuerdo con Wellhausen *et al.* (1951), Harinoso de Ocho ha dado lugar directamente a la Tabloncillo e indirectamente a Jala, Bolita, Celaya y Cónico Norteño.

Los Elotes Occidentales se distribuyen en elevaciones de 1200 a 1600 msnm, se usan para maíces eloteros y por consiguiente representan el complejo del maíz harinoso. Aunque las variedades eloteras que se encuentran al norte y al este de la Altiplanicie de Jalisco a elevaciones mayores, muestran una fuerte influencia genética del Cónico (Wellhausen *et al.*, 1951). Lo anterior puede ser una posible explicación, debido a que los Elotes Occidentales comparten en su genoma alelos presentes en la raza Cónico Norteño, los cuales son identificados por polimorfismos de un solo nucleótido en el presente estudio.

Las poblaciones de las razas Elotes Cónicos y Cónico se ubicaron en torno al origen; tal convergencia es congruente con lo propuesto por Wellhausen *et al.* (1951), en el sentido a que Elotes Cónicos corresponde a una subraza de Cónico.

Aunque la mayoría de las poblaciones se ubicaron dentro de los dos grupos principales (Cónica y Cilíndrica), pocos de ellos se encuentran afuera (Figura 9), esto puede deberse a una combinación genética que posiblemente resultó de un alto grado de interacción entre los dos complejos raciales y causado por el intercambio de semillas entre agricultores dentro y entre comunidades en la región, que también esto genera una gran variación (Rincón *et al.*, 2010; Aguirre *et al.*, 2011). Rice *et al.* (1998), mencionan que debido a que los agricultores realizan una selección direccional hay una divergencia de flujo genético extenso (intercambio de semillas y dispersión de polen natural), por lo que la adaptación local es contrastante a diferentes condiciones ambientales.

Los resultados de la Figura 9 concuerdan en parte con el estudio realizado por Wu *et al.* (2016), que al utilizar 362,008 SNP para caracterizar la diversidad genética de 539 líneas endogámicas de maíz, se agruparon por separado basado a su área de adaptación (tierras altas, altitud media, tierras bajas). En el presente estudio se utilizaron 53,471 SNP para diferenciar dos grandes grupos, constituidos por poblaciones con adaptación a áreas específicas (Áreas de Transición-Altura y áreas Bajas-Intermedias), indicando la divergencia genética entre ambos. Esto resalta la importancia de la selección de marcadores SNP óptimos para el análisis de diversidad genética, independientemente de la plataforma de genotipado que se use. La evaluación y la selección de marcadores informativos mediante parámetros de calidad, ayudarían a una mejor separación del

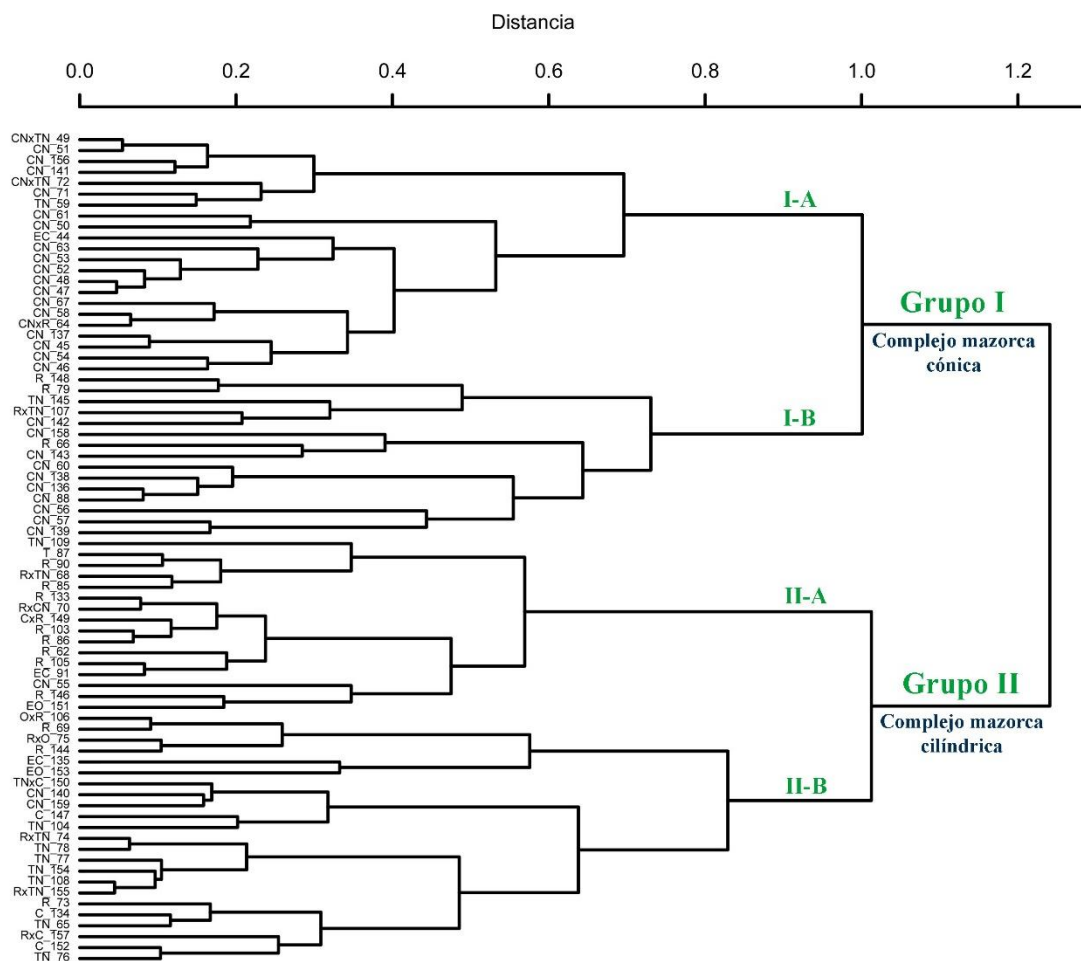
germoplasma del maíz, por lo tanto, la generación de datos informativos, es un requisito previo fundamental (Chen *et al.*, 2016).

Por el contrario, en la mayoría de los sistemas basados en secuenciación donde el número de marcador es en gran medida de hasta decenas a cientos de miles de marcadores, los niveles más altos de datos faltantes pueden dar lugar a una interpretación errónea, por lo que en la sub-selección de los miles de marcadores generados, es necesario contar solamente con marcadores con una representación alta y repetible a través de muestras.

Los grupos raciales y sus combinaciones genéticas incluidos en el presente estudio tienden a ocupar un área ecológica específica de acuerdo con su zona de adaptación y condiciones ambientales y que mediante ello, pueden explicar los patrones de la diversidad identificados, tal como lo mencionan Brush y Perales (2007). De ahí la importancia de la información ecogeográfica donde se hacen las colectas para sugerir otros sitios, en los cuales dicho germoplasma o germoplasma derivado de razas nativas se pueda adaptar cuando se piensa usar en programas de mejoramiento genético (Sánchez y Ruiz, 1995).

## Relación entre poblaciones nativas

Los resultados de la relación entre las poblaciones nativas, con base en los coeficientes de distancia obtenidos con los caracteres morfológicos se presenta en la Figura 10.



**Figura 10.** Dendrograma jerárquico de las relaciones entre 76 poblaciones de maíz del sureste de Coahuila, México mediante caracteres morfológicos. Las letras indican la identificación de cada raza: C= Celaya, O= Olotillo; CN= Cónico Norteño, R= Ratón, EC= Elotes Cónicos, EO= Elotes Occidentales, T= Tuxpeño y TN= Tuxpeño Norteño y los dígitos el número de colecta.

A una distancia de 1.05, en el dendrograma se distinguen dos grandes grupos. En la parte superior de la Figura (Grupo I) se ubicaron las poblaciones consideradas del Complejo mazorca cónica (Transición-Altura), y en la parte inferior (Grupo II) las poblaciones del Complejo mazorca cilíndrica (Bajas-Intermedias), en congruencia con el ordenamiento obtenido por el Análisis de Componentes Principales (Figura 8).

La diversidad que se encontró en este análisis se manifiesta en cuatro Subgrupos, a partir de una distancia de 0.81, aproximadamente. La subdivisión del Grupo I en dos subgrupos: el Subgrupo I-A que aglomeró en su totalidad a las poblaciones clasificadas como Cónicos Norteños a excepción del 59\_TN. El Subgrupo I-B congregó a diez poblaciones de Cónico Norteño, cuatro de Ratón y una de Tuxpeño Norteño. En el Grupo II se distinguen dos Subgrupos: el Subgrupo II-A, formado la mayor parte por poblaciones de Ratón, y el Subgrupo II-B compuesto mayormente por poblaciones de la raza Tuxpeño Norteño y Ratón. De manera general, el ordenamiento y relaciones entre las poblaciones se explica por los 15 caracteres de la mazorca (Rincón-Sánchez y Ruiz-Torres, 2018), sin embargo, la presencia de combinaciones raciales adicionales a la frecuencia que caracteriza al subgrupo II-A y II-B, indica que estos caracteres pueden estar limitando para explicar con mayor precisión la diversidad del maíz en las poblaciones en estudio. Perales *et al.* (2005) mencionan que usando características morfológicas que están bajo la selección de agricultores, se ha demostrado que las poblaciones de maíz se mantienen a distintas escalas regionales mucho más pequeñas, lo que sugiere que hay una tendencia a que las poblaciones de maíz en una región están estructurado fenológicamente y morfológicamente de acuerdo con su adaptación local.

En el análisis previo de la diversidad de los maíces incluidos en este estudio mediante caracteres morfológicos, mismas que han sido analizadas en este estudio por (Rincón-Sánchez y Ruiz-Torres, 2018), mencionan que a una distancia de 1.05 el análisis de conglomerados reveló dos grupos principales, el Grupo 1 representado por 37 poblaciones (Complejo de tipo mazorca cónica), y el segundo, constituido por 40 poblaciones, (Complejo de mazorca cilíndrica). Sin embargo, los mismos autores también señalan que algunas poblaciones no concordaban con las clasificaciones de razas asignadas tales como: 149EC, 155CN, 48EC, 12CN, 163R y 36R, las cuales no pudieron separarse claramente, lo que demuestra las limitaciones de estos resultados en la evaluación de la variabilidad genética.

Esta relación entre poblaciones puede ser explicada por el número de caracteres utilizados, lo que repercute en la constitución de los subgrupos de la Figura 10.

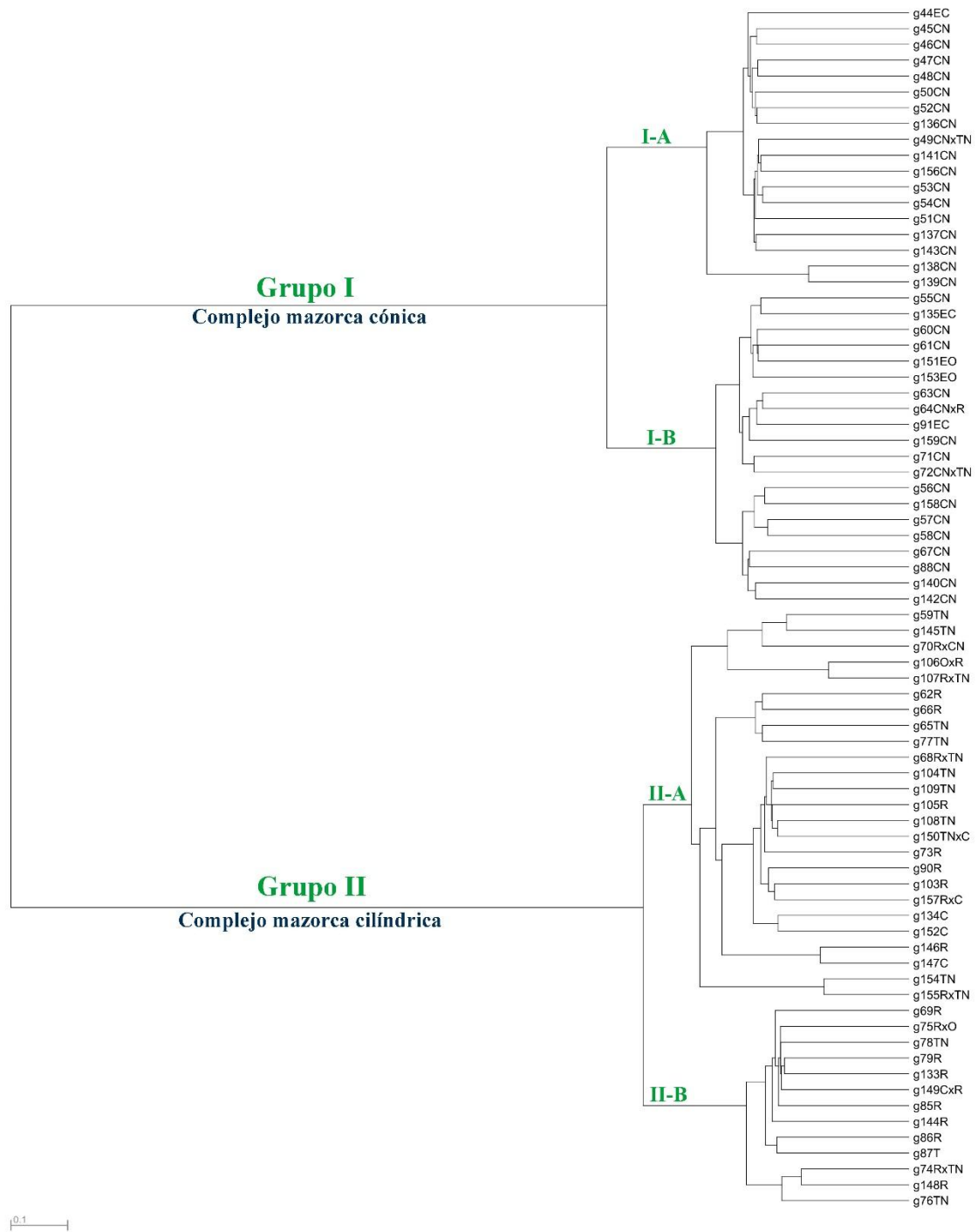
El análisis de conglomerados basado en el método de Ward (Ward, 1963), utilizando la distancia genética de Rogers modificada (Wright, 1978) se presenta en la Figura 11.

El agrupamiento de las poblaciones en la Figura 11 es una extensión y correspondencia del análisis de componentes principales (Figura 9), en donde, las poblaciones se definen en grupos de una manera más precisa, lo cual incluye la forma de mazorca, que caracteriza a los maíces por su adaptación específica y delimitación geográfica altitudinal (Wellhausen *et al.*, 1951; Ruíz *et al.*, 2008).

En el dendograma se logra identificar dos grupos bien definidos: el Grupo I, identificado como el Complejo mazorca cónica (razas adaptadas a áreas Transición-Altura), y el Grupo II con poblaciones consideradas del Complejo mazorca cilíndrica (razas adaptadas a áreas Bajas-Intermedias), tal y como se presenta en la Figura 11.

Dentro del Grupo I, se integró principalmente por poblaciones del Complejo mazorca cónica de amplia variación, Cónico Norteño y dos poblaciones de Elotes Cónicos (135EC y 91EC). La presencia de las colectas clasificadas como Elotes Cónicos en este grupo, demuestra la estrecha relación de estos maíces con la raza Cónico; que, de acuerdo con Wellhausen *et al.* (1951), Elotes Cónicos es una subraza de la raza Cónico, y esta es el antecesor de la raza Cónico Norteño.





**Figura 11.** Dendrograma jerárquico de las relaciones entre 76 poblaciones de maíz del sureste de Coahuila, México mediante marcadores SNP. Las letras indican la identificación de cada raza: C= Celaya, O= Olotillo; CN= Cónico Norteño, R= Ratón, EC= Elotes Cónicos, EO= Elotes Occidentales, T= Tuxpeño y TN= Tuxpeño Norteño y los dígitos el número de colecta.

El análisis de conglomerados también permitió hacer una distinción dentro del Complejo mazorca cilíndrica (Grupo II), el cual se constituyó principalmente por poblaciones de las razas Ratón, Olotillo, Celaya, Tuxpeño y Tuxpeño Norteño. En general, esta agrupación de maíces nativos está de acuerdo con las clasificaciones de las poblaciones de maíz recolectadas en esta región de Coahuila (Cuadro 1).

De igual manera, en la Figura 11 se puede identificar cuatro subgrupos. La división del Complejo mazorca cónica (Grupo I), se pueden distinguir dos subgrupos: el Subgrupo I-A, incluyó a 17 poblaciones de Cónico Norteño y una población de Elotes Cónicos. El Subgrupo I-B formado por 16 poblaciones de Cónico Norteño, dos de Elotes Cónicos y Elotes Occidentales. En el Grupo II del Complejo mazorca cilíndrica, se distinguen dos subgrupos: el Subgrupo II-A, integrado por 25 poblaciones, la mayor parte por poblaciones de Ratón y Tuxpeño Norteño, tres de Celaya y uno de Olotillo, el Subgrupo II-B compuesto por 13 poblaciones principalmente de la raza Ratón y Tuxpeño Norteño.

En este estudio, la similitud o divergencia genética entre las poblaciones evaluadas, obtenida por polimorfismos de un solo nucleótido (SNP), logró separar claramente los maíces adaptados a áreas de Transición-Altura de aquellos con adaptación a áreas Bajas-Intermedias (Cuadro 1), lo cual coincide con la investigación realizada por Reif *et al.* (2006), quienes caracterizaron molecularmente 25 razas pertenecientes a 11 estados de México. Estos autores señalan que, una distribución de la diversidad genética presenta un patrón de aislamiento por distancia; esto es, poblaciones más cercanas genéticamente y geográficamente se parecen más entre sí.

En el estudio de Rincón-Sánchez y Ruíz-Torres (2018), donde se incluyen las poblaciones analizadas en el presente estudio, con base en caracteres morfológicos de mazorca, se reporta de manera general la separación en dos grandes grupos, tipificados como Complejo mazorca cónica y cilíndrica. El uso de marcadores moleculares como los polimorfismos de un solo nucleótido (SNP), utilizados en el presente estudio, permitieron agrupar las poblaciones de una manera más precisa, por lo que concluye que los datos moleculares serían más precisos en la discriminación de accesiones estrechamente relacionadas y más eficientes el análisis de sus relaciones genéticas. Por lo tanto, los marcadores SNPs constituyen una técnica con alto poder de discriminación

de poblaciones a nivel de ADN y de utilidad en estudios de diversidad genética (Melchinger *et al.*, 1991). Cabe recalcar que tanto los marcadores morfológicos como los genéticos son necesarios y complementarios para la conservación y el desarrollo efectivo de los recursos genéticos, con el fin de obtener resultados integrales en la evaluación de la diversidad genética (Ristic *et al.*, 2013).

Las diferencias en la clasificación de poblaciones mediante los datos con marcadores moleculares y estudios previos con datos morfológicos, se debe a que las muestras a nivel de secuencia de ADN, definen un agrupamiento a nivel genético, en tanto que los marcadores morfológicos, definen una relación a nivel del fenotipo, en donde el genotipo es afectado por factores ambientales (Govindaraj *et al.*, 2015).

## CONCLUSIONES

La tecnología DArTseq basada en polimorfismos de nucleótido único (SNP) fue eficiente para entender las relaciones genéticas entre poblaciones de las razas nativas del sureste de Coahuila, en comparación con el uso de caracteres cuantitativos de la mazorca y grano en maíz.

Los valores de heterocigosidad esperada ( $H_e$ ) como indicador de la diversidad genética, fue de utilidad para determinar que las poblaciones adaptadas al área de transición y altura son más diversas en relación con las poblaciones adaptadas a áreas intermedias.

Se encontró una relación positiva ( $r=0.42^{**}$ ) entre las distancias fenotípicas y las distancias genéticas, indicando congruencia en la clasificación de poblaciones de maíz. Validando trabajos previos que utilizaron únicamente datos morfológicos para realizar la clasificación en grandes grupos de las poblaciones evaluadas aquí, y que están ligados a las áreas de adaptación.

El análisis con métodos multivariantes a partir de SNP, permitió corroborar la diversidad del maíz en el sureste de Coahuila, México, representada por dos grupos: Complejo mazorca cónica constituido por las razas Cónico Norteño y Elotes Cónicos (Transición-Altura) y el Complejo mazorca cilíndrica por las razas Ratón y Tuxpeño Norteño (Bajas-Intermedias).

## REFERENCIAS

- Abbasov, M., C. P. Sansaloni, J. Burgueño, C. D. Petroli, Z. Akparov, M. Aminov, S. Babayeva, V. Izzatullayeva, E. Hajiyev, K. Rustamov, S. A. Mammadova, A. Amri, and T. Payne. 2019. Genetic diversity analysis using DArTseq and SNP markers in populations of *Aegilops* species from Azerbaijan. *Genetic Resources and Crop Evolution* 22:7-19.
- Aguirre, M. V. J., S. F. Rincón, S. R. Ramírez, A. O. G. Colón and M. M. G. Razo. 2011. Modelo Para la Conservación de Maíces Criollos en el Sureste de Coahuila Saltillo Coahuila, México.
- Aranguren-Méndez, J. A., R. Román-Bravo, W. Isea, Y. Villasmil, y J. Jordana. 2005. Los microsatélites (STR's), marcadores moleculares de ADN por excelencia para programas de conservación: una revisión. *Arch. Latinoam. Prod. Anim.* 13(1):1-6.
- Arias, V. Y., I. G. Marquetti, I. M. Cabrera, L. F. Granda, B. P. Delgado-Oramas. 2018. Diversidad genética en maíz (*Zea mays* L.) procedente del Pinar del Río y Guantánamo, Cuba, mediante el empleo de RAPD. *Rev. Protección Veg.* 33:1-9.
- Arteaga, C. M., A. Moreno, A. Mastretta, A. Vázquez, A. Breña, A. Moreno, L. Eguiarte, D. Pinero. 2015. Variación genómica en variedades de maíz recolectadas recientemente de México. *Genomics Data* 7:38-45.
- Baloch, F. S., A. Alsaleh., M. Q. Shahid, V. Çiftçi, L. E. Sáenz de Miera, M. Aasim, M. Azhar Nadeem, H. Aktaş, H. Ozkan, R. Hatipoğlu. 2017. A Whole Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from Central Fertile Crescent. *PLoS ONE* 12(1):1-16.
- Bernatzky, R. 1988. Restriction fragment length polymorphism. In: Gelvin, S.B. Schilperoort, R.A. eds. *Plant Molecular Biology Manual*. Dordrecht, The Netherlands. Kluwer Academic Publishers. 1-18p.
- Beyene, Y., A. M. Botha and A. A. Myburg. 2005. A comparative study of molecular and morphological methods of describing genetic relationships in traditional Ethiopian highland maize. *African Journal of Biotechnology* 4(7):586–595.
- Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32:314–331.
- Brush, S. B y H. R. Perales. 2007. A maize landscape: Ethnicity and agro-biodiversity in Chiapas Mexico. *Agriculture, Ecosystems and Environment*, 121: 211-221.
- Carballoso, V., A. Mejía, S. Valderrama, A. Carballo, F. V. González. 2000. Divergencia en poblaciones de maíz nativas de Valles Altos de México. *Agrociencia* 34(2):167-174.

- Chen, J., C. Zavala, N. Ortega, C. Petroli, J. Franco, J. Burgueño, and S. J. Hearne. 2016. The Development of Quality Control Genotyping Approaches: A Case Study Using Elite Maize Lines. *PLoS one* 116, e0157236.
- Chen, T., P. Alisha, W. Wang, X. Gao and L. Zhang. 2018. Population structure of Chinese southwest wheat germplasms resistant to stripe rust and powdery mildew using the DArT-seq technique. *Ciencia Rural* 48(4):1-9.
- Cyrine, R., Z. Kehel, M. Berek, C. Sansaloni, F. Bassi, and A. Amri. 2019. Genome-Wide Genetic Diversity and Population Structure of Tunisian Durum Wheat Landraces Based on DArTseq Technology. *International Journal of Molecular Sciences* 20:1-21.
- Doebley, J. F., M. M. Goodman y C.W. Stuber. 1985. Isozyme variation in the races of maize from Mexico. *American Journal of Botany* 72(5):629-639.
- Edwards, D., J. Forster, D. Chagné and J. Batley. 2007. What is SNPs? In: Oraguzie, N.C., E. Rikkerink, S. Gardiner and H. De Silva (eds) *Association Mapping in Plants*. Springer, Berlin, pp. 41–52.
- Elshire, R. J., J. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6(5):1-9.
- FOASTAT. 2018. Datos sobre alimentación y agricultura. FAOSTAT. Food and Agriculture Organization of the United Nations (FAO). Roma, Italia. Disponible en línea: <http://www.fao.org/faostat/es/#data/QC/visualize>. Fecha de consulta 12 de febrero de 2020.
- Frankham, R., J. D. Ballou, & D. A. Briscoe. 2009. *Introduction to Conservation Genetics*. Cambridge University Press. Cambridge, UK.
- Glaubitz, J. C., T. M Casstevens, F. Lu, J. Harriman, R. J. Elshire, Q. Sun, E. S. Buckler. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9(2): 1-9.
- Goodman, M. M. and E. Paterniani. 1969. The Races of Maize: III. Choices of appropriate characters for racial classification. *Econ. Bot.* 23: 265-273.
- Golstein, D. B. y D. D. Pollok. 1994. Least-squares estimation of molecular distance – noise abatement in phylogenetic reconstruction. *Theoretical Population Biology* 45:219-226.
- Gore, M., P. Bradbury, R. Hogers, M. Kirst, E. Verstege, J. V. Oeveren, J. Peleman, E. Buckler, and M. V. Eijk. 2007. Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. *Crop Sci* 47(2): 135–148.
- Govindaraj, M., M. Vetriventhan, and M. Srinivasan. 2015. Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives. *Genetics Research International* 2015:1-10.

- Gupta, P. K. and S. Rustgi. 2004. Molecular Markers from the transcribed/expressed region of the genome in higher plants. *Funct. Integr. Genomics* 4:139-162.
- Hartl, D. L. and A. G. Clark. 1997. *Principles of Population Genetics*. Sinauer Associates, Inc. Publishers. Massachusetts, USA.
- Hartl, D. and E. Jones. 2009. *Genetics: Analysis of Genes and Genomes*. Jones and 7th Ed. Barttler Publishers, Canada.
- Hedrick, P. W. 1983. *Genetic of Populations*. Jones and Bartlett Publishers, Inc. Boston, USA.
- Hernández, C., J. M. y G. Esquivel E. 2004. Rendimiento de grano y características agronómicas en germoplasma de maíz de valles altos de México. *Revista Fitotecnia Mexicana* 27: 27-31.
- Hoelzel, A. R. 1998. *Molecular Genetic Analysis of Populations: A practical approach*. IRL Press at Oxford University Press. New York, USA.
- Illumina, Inc. 2016. An introduction to Next-Generation Sequencing technology. Disponible en: [http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina\\_sequencing\\_introduction.pdf](http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf).
- Jaccoud, D., K. Peng, D. Feinstein, A. Kilian. 2001. Diversity arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Research* 29(4): 1-7.
- Jonah, P. M., L. L. Bello, O. Lucky, A. Midau, S. M. Moruppa. 2011. The importance of Molecular Markers in Plant Breeding Programs. *Global Journal of Science Frontier Research* 11(5):1-9.
- Karp, A., S. Kresovich, S. Bhat, K. Ayad, W. Hodgking. 1997. *Molecular tools in plant genetic resources conservation: a guide to the technologies*. International Plant Genetic Resources Institute (IPGRI). Roma Italia. 47p.
- Karp, A. y K. Edwards. 1998. DNA makers: a global overview. In: G. Caetano Anollés, P.M. (eds.). *DNA markers: protocols, applications and overviews*. Gresshoff New York. Pp. 1-13.
- Kilian, A., P. Wenzl, E. Huttner, J. Carling, L. Xia, H. Blois, V. Caig, K. Heller-Uszynska, D. Jaccoud, C. Hopper, M. Aschenbrenner-Kilian, M. Evers, K. Peng, C. Cayla, P. Hok, and G. Uszynski. 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. *Data Production and Analysis in Population Genomics: Methods and Protocols* 888:67-89.
- Kimura, M. and J. F. Crow. 1964. The number of alleles that can be maintained in a finite population, *Genetics* 49:725-738.
- Kujur, A., D. Bajaj, H. Upadhyaya, S. Das, R. Ranjan, T. Shree, M. Saena, S. Badon, V. Kumar, S. Tripathi, C. Gowda, S. Sharma, S. Singh, A. Tyagi and S. Parida. 2015. Employing genome-wide SNP discovery and genotyping strategy to

- extrapolate the natural allelic diversity and demonstration patterns in chickpea. *Front. Plants Sci.* 6(162):1-20.
- Khush, G. 2001. Green revolution the way forward. *Nature Reviews Genetics* 2: 815-822.
- Lanza, L. L. B., C. L. de Souza, L. M. M. Ottoboni, M. L. C. Viera, and A. P. de Souza. 1997. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. *Theoretical and Applied Genetics* 94:1023-1030.
- López, R. G., A. Santacruz, A. Muñoz, F. Castillo, L. Córdova, H. Vaquera. 2009. Perfil Isoenzimático de maíces nativos del Istmo de Tehuantepec, Oaxaca, México. II. Variación dentro de grupos. *Revista Fitotecnia Mexicana* 32(3):177-188.
- Lorenz, A. J., S. Chao, F. G. Asoro, E. L. Heffner, T. Hayashi, H. Iwata, and J. L. Jannink. 2011. Genomic Selection in Plant Breeding: Knowledge and Prospects. *Advances in agronomy* 110:77.
- Madigan, M., J. Martinico, J. Parker, T. 1998. Brock *Biología de los microorganismos*. Madrid, España. Prentice Hall. 1064 P.
- Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. Sánchez, E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences* 99(9):1-5.
- Melchinger. A. E., M. M. Messmer, M. Lee, W. L. Woodman, K. R. Lamkey. 1991. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. *Crop Sci.* 31:669–678.
- Murray, M. G. and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8(19): 4321–4325.
- Nadeem, M., E. Habyarimana, V. Ciftci, M. Amjad, T. Karakoy, G. Comertpay, M. Qasim, R. Hatipoglu, M. Zahit, F. Ali, S. Ercisli, G. Chung, F. Baloch. 2018. Characterization of genetic diversity in Turkish common bean gene pool using phenotypic and whole-genome DArTseq-generated silicoDArT marker information. *PLOS ONE* 13(10)1-28.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press. New York, USA. 512p.
- Ortega, P., R. A., J. J. Sánchez G., F. Castillo G. y J. M. Hernández C. 1991. Estado actual de los estudios sobre maíces nativos de México. In: *Avance del Estudio de los Recursos Fitogenéticos de México*. Ortega P. R., G. Palomino H., F. Castillo G., V. A. González H. y M. Livera M. (eds). SOMEFI. Chapingo, México. pp: 161-185.
- Otero, A. A., M. De la Cruz y K. Oyama. 1997. El uso de los RAPDs como marcadores moleculares en plantas. *Boletín de la Sociedad Botánica de México* 60:85-117.



- Pacheco, A., G. Alvarado, F. Rodríguez, J. Crossa and J. Burgueño. 2016. BIO-R (Biodiversity Analysis with R) Versión 2.0.
- Paterson, A. H. 1996. Genome mapping in plants. Academic Press, Inc. and R.G. Landes Co., New York and Austin. 330 pp.
- Perales, R. H, B. F. Benz, S. B. Brush. 2005. Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. Proc. Natl. Acad. Sci. USA 102(3):949-954.
- Pérez-Nasser, N. y D. Piñero. 1997. Isoenzimas. Boletín de la Sociedad Botánica de México 60:77-84.
- Phillips, W., H. Rodríguez y P. Fritz. 1995. Marcadores de ADN: Teoría, aplicaciones y protocolos de trabajo con ejemplos de investigaciones en cacao (*Theobroma cacao*). Serie técnica. Informe técnico #252. Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). Turrialba, Costa Rica. 183 p.
- Raychaudhuri, S. 2011. Mapping rare and common causal alleles for complex human diseases. Cell. 147(1):57-69.
- Reif, J. C., M. L. Warburton, X. C. Xia, D. A. Hoisington, J. Crossa, S. Taba, J. Muminovic, M. Bohn, M. Frisch, A. E. Melchinger. 2006. Grouping of accessions of Mexican races of maize revisited with SSR markers. Theoretical and Applied Genetics 113:177-185.
- Reyes-Valdés, M. H., J. Burgueño, S. Singh, O. Martínez and C. P. Sansaloni. 2018. An informational view of accession rarity and allele specificity in germplasm banks for management and conservation. PLoS One 13:1-15.
- Rice, E and M. Smale. 1998. Farmers' use of improved seed selection practices in Mexican maize: evidence and issues from the Sierra de Santa Marta. World Development 26(9):1625-1640.
- Rincón, S. F., F. Castillo G. y N. A. Ruiz T. 2010. Diversidad y Distribución de los Maíces Nativos en Coahuila, México. Sociedad Mexicana de Fitogenética. Chapingo, México. 123 p.
- Rincón, S. F. y N. A. Ruiz T. 2015. Maíces nativos en el estado de Coahuila, México. Revista Claridades Agropecuarias. Edición especial: 40-44. SAGARPA-ASERCA. Disponible en línea <http://www.infoaserca.gob.mx/claridades/especial-201509.asp>.
- Rincón-Sánchez, F., N. A. Ruiz-Torres. 2018. Assessing maize diversity and defining representative subset by means of selected morphological ear traits. Plant Genetic Resources; 1-10.
- Ristic, D., V. Babic, V. Andelkovic, J. Vancetovic, S. Mladenovic, and D. Ignjatovic. 2013. Genetic diversity in maize dent landraces assessed by morphological and molecular markers. Genetika 45(3):811-824.
- Rocandio-Rodríguez, M., A. Santacruz-Varela, L. Córdova-Téllez, H. López-Sánchez, F. Castillo-González, R. Lobato-Ortiz, J. J. García-Zavala. 2014. Detection of

genetic diversity of seven maize races from the high central valleys of Mexico using microsatellites. *Maydica* 59:144-151.

- Rodríguez-Pérez, G., F. Zavala-García, C. Ojeda-Zacarías, A. Gutiérrez-Diez, J. E. Trevino-Ramírez, F. Rincón-Sánchez. 2012. Diversidad de maíces criollos de Nuevo León, México, mediante AFLP y caracteres morfológicos. *Agronomía Mesoamericana* 23(2):29-39.
- Rohlf, F. J. 2012. NTSYSpc: Numerical Taxonomy System. Ver. 2.21q for Windows. Setauket, New York: Exeter Software.
- Romay, M. C., M. J. Millard, J. C. Glaubitz, J. A. Peiffer, K. L. Swarts, et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology* 14:1–18.
- Rossi, E. A., M. Ruíz, M. Di Renzo, N. C. Bonamico. 2019. Diversidad genotípica de 291 líneas de maíz de CIMMYT y caracterización fenotípica en el sur de Córdoba, Argentina. *Journal of Basic and Applied Genetics* 3:25-33.
- Ruiz, C. J. A., P. N. Durán, G. J. J. Sánchez, P. J. Ron, E. D. R. González, J. B. Holland and G. G. Medina. 2008. Climatic adaptation and ecological descriptors of 42 Mexican maize races. *Crop Science* 48(4):1502–1512.
- SAGARPA-SIAP. 2018. Anuario Estadístico de la Producción Agrícola. Servicio de información Agroalimentaria y Pesquera (SIAP), Secretaria de Agricultura, Ganadería, Pesca y Alimentación (SAGARPA). Disponible en línea <https://nube.siap.gob.mx/cierreagricola/>. Fecha de consulta 12 de febrero del 2020.
- Sanchez, G., J. J., M. M. Goodman, and J. O. Rawlings 1993. Appropriate characters for racial classification in maize. *Econ. Bot.* 47:44-59.
- Sánchez, G. J. J., J. A. Ruiz. 1995. Teosinte distribution in Mexico. In: *Gene Flow among Maize Landraces, Improved Maize Varieties, and Teosinte; Implication for Transgenic Maize*. J. A. Serratos, M. C. Willcox, F. Castillo (eds.). México, D. F. CIMMYT. Pp. 18-39.
- Sánchez, G. J. J., T. A. Kato, M. Aguilar, J. M. Hernández, A. López, J. A. Ruiz. 1998. Distribución y caracterización del teocintle. Centro de Investigación Regional del Pacífico Centro, Instituto Nacional de 100 Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Libro Técnico Núm. 2. Guadalajara, Jalisco, México. Pp. 1-29.
- Sánchez, G., J. J. M. M. Goodman, and C. W. Stuber. 2000. Isozymatic and morphological diversity in the races of maize of México. *Economic Botany* 54:43-59.
- Sansaloni, C. P., C. D. Petrolí, J. Carling, C. J. Hudson, D. A. Steane, A. A. Myburg, et al. 2010. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. *Plant Methods* 6(1):1-11.

- Sansaloni, C. Petroli, C. Jaccoud, D. Carling, J. Detering, F. Grattapaglia, D. Kilian A. 2011. Diversity Arrays Technology (DART) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proceedings. BioMed Central Ltd, Pp54.
- Semagn, K., A. Bjørnstad and M. N. Ndjiondjop. 2006. An overview of molecular marker methods for plants. African Journal of Biotechnology 5(25):2540–2568.
- SIDTA. 1999. Los marcadores moleculares en ingeniería genética y en mejora vegetal. España. Disponible en línea <http://www.jcyl.es/jcyl/cag/dgiadr/svidta/boletin/dic99/bold.html#los%20marcadores%20moleculares%20en%20ingenieria%20genetica>. Fecha de consulta 15 de junio del 2020.
- Singh, B. D., y A. K. Singh. 2015. High-Throughput SNP Genotyping. In Marker-Assisted Plant Breeding: Principles and Practices pp. 367-400. Springer India.
- Sosa, P., F. Batista, M. A. González y N. Bouza. 2002. La Conservación Genética de las Especies Amenazadas. En: Bañares, Á. (ed.). Biología de la Conservación de Plantas Amenazadas: Técnicas de diagnóstico del estado de conservación. Organismo Autónomo de Parques Nacionales. España, pp. 133-160.
- Tanksley, S., H. Medina and C. M. Rick. 1981. The effect of isozyme selection on metric characters in an interspecific backcross of tomato basis of an early screening procedure. Theoretical and Applied Genetics 60:291-296.
- Team, R. C. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- Templeton, A. R. 2006. Population Genetics and Microevolutionary Theory. John Wiley & Sons. Hoboken, New Jersey, USA.
- Valadez, E. y G. Kahl. 2000. Huellas de ADN en genomas de plantas. (Teoría y Protocolos de Laboratorio) Mundiprensa. D. F. México, 182 p.
- Vega-Alvarez, I., A. Santacruz-Varela, M. Rocandio-Rodríguez, L. Córdova-Téllez, H. López-Sánchez, A. Muñoz-Orozco and A. Hernández-Bautista. 2017. Genetic diversity and structure of native maize races from Northwestern Mexico. Pesquisa Agropecuaria Brasileira 55(11):1023-1032.
- Vigouroux, Y., J. C. Glaubitz, Y. Matsuoka, M. M. Goodman, J. Sánchez, J. Doebley. 2008. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. American Journal of Botany 95(10):1240-1253.
- Warburton, M. L., P. Setimela, J. Franco, H. Cordova, K. Pixley, M. Banziger, S. Dreisigacker, C. Bedoya, and J. MacRobert. 2010. Toward a Cost-Effective Fingerprinting Methodology to Distinguish Maize Open-Pollinated Varieties. Crop Science 50:467-477.

- Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(301): 236-244.
- Wellhausen, E. J., L. M. Roberts y E. Hernández X., y P. C. Mangelsdorf. 1951. Razas de Maíz en México, su Origen, Características y Distribución. Oficina de Estudios Especiales. Secretaría de Agricultura y Ganadería. Folleto Técnico 5. México, D. F. 237 p.
- Wen, W., J. Franco, V. H. Chavez-Tovar, J. Yan, S. Taba. 2012. Genetic Characterization of a Core Set of a Tropical Maize Race Tuxpeño for Further Use in Maize Improvement. *PLoS ONE* 7(3):e32626.
- Wenzl, P., J. Carling, D. Kudrna, D. Jaccoud, E. Huttner, A. Kleinhofs, A. Kilian. 2004. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *The National Academy of Sciences of the USA* 101(26): 9915-9920.
- Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18(22):6531-6535.
- Wright, S. 1978. *Evolution and the genetics of populations: a treatise in four volumes: variability within and among natural populations*. Chicago: University of Chicago, v.4, 590p.
- Wu, X., L. Yongxiang, S. Yunsu, S. Yanchun, W. Tianyu, H. Yubi, L. Yu. 2014. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. *Theoretical and Applied Genetics* 127:621-631.
- Wu, Y., F. San Vicente, K. Huang, T. Dhliwayo, D. Costich, E. Semagn, and R. Babu. 2016. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theoretical and Applied Genetics* 129:753-765.
- Xia, L., K. Peng, S. Yang, P. Wenzl, M. de Vicente, M. Fregene, A. Kilian. 2005. DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. *Theor Appl Genet* 110(6):1092-1098.
- Xu, Y. 2010. *Molecular plant breeding*. Experimental Agriculture Wallingford: CABI. 47(1). 734 p.
- Xu, C., Y. Ren, Y. Jian, Z. Guo, Y. Zhang, C. Xie, J. Fu, H. Wang, G. Wang, and Y. Xu. 2017. Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. *Molecular Breeding* 37:1-20.
- Yan, J., X. Yang, T. Shah, H. Sánchez-Villeda, J. Li, M. Warburton, Y. Zhou, J. H. Crouch, and Y. Xu. 2009. High-throughput SNP genotyping with the GoldenGate assay in maize. *Mol Breeding* 25:441-451.
- Zhang, X., H. Zhang, L. Li, Z. Ren, D. Liu, L. Wu, H. Liu, J. Jaqueth, B. Liu, G. Pan, and S. Gao. 2016. Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. *BMC Genomics* 17:1-16.

## ANEXO

**Cuadro 6.** Solución amortiguadora con CTAB para extracción.<sup>1</sup>

<b>Sol. Conc.</b>	<b>(Final)</b>	<b>50 ml</b>	<b>200 ml</b>	<b>500 ml</b>
dH <sub>2</sub> O		32.5 ml	130.0 ml	325.0 ml
1M Tris Ph 7.5	100 mM	5.0 ml	20.0 ml	50.0 ml
5M NaCl	700 mM	7.0 ml	28.0 ml	70.0 ml
0.5M EDTA Ph 8.0	50 mM	5.0 ml	20.0 ml	50.0 ml
CTAB <sup>2</sup>	1 %	0.5 gr	2.0 gr	5.0 gr
BME <sup>3</sup>	140 mM	0.5 ml	2.0 ml	5.0 ml

<sup>1</sup> Solución recién hecha; calentar a 60-65°C antes de agregar CTAB y BME.

<sup>2</sup> CTAB = bromuro mixto de alquiltrimetil-amonio (Sigma M-7635).

<sup>3</sup> Agregar BME (β-mercaptoetanol) bajo una campana extractora, justo antes de usarse. ml=Mililitro; dh<sub>2</sub>O= Agua destilada; Ph= Potencial de hidrogeno; Tris= Tris (hidroximetil) aminometano; EDTA= Tetracetato de etilendiamina.

**Cuadro 7.** Lavado 1: 76% de EtOH, 0.2 M NaOAc.

<b>Unidad</b>	<b>100 ml</b>	<b>200 ml</b>	<b>300 ml</b>	<b>400 ml</b>	<b>500 ml</b>
EtOH absoluto	76 ml	152 ml	228 ml	304 ml	380 ml
2.5 M NaOAc	8 ml	16 ml	24 ml	32 ml	40 ml
dH <sub>2</sub> O	16 ml	32 ml	48 ml	64 ml	80 ml

ml= Mililitro; EtOH= Alcohol absoluto; NaOAc= Acetato de sodio; dh<sub>2</sub>O= Agua destilada.

**Cuadro 8.** Lavado 2: 76% EtOH, 10 mM NH<sub>4</sub>OAc.

<b>Unidad</b>	<b>100 ml</b>	<b>200 ml</b>	<b>300 ml</b>	<b>400 ml</b>	<b>500 ml</b>
EtOH absoluto	76 ml	152 ml	228 ml	304 ml	380 ml
1 M NH <sub>4</sub> OAc	1 ml	2 ml	3 ml	4 ml	5 ml
dH <sub>2</sub> O	23 ml	46 ml	69 ml	92 ml	115 ml

ml= Mililitro; EtOH= Alcohol absoluto; NH<sub>4</sub>OAc= Acetato de amonio; dh<sub>2</sub>O= Agua destilada.

**Cuadro 9.** Cloroformo: Octanol, 24:1.

<b>Unidad</b>	<b>100 ml</b>	<b>200 ml</b>	<b>300 ml</b>	<b>400 ml</b>	<b>500 ml</b>
Cloroformo	96 ml	192 ml	288 ml	384 ml	480 ml
Octanol	4 ml	8 ml	12 ml	16 ml	20 ml

ml= Mililitro.

**Cuadro 10.** Preparación de gel de agarosa (0.8 %) para muestras de ADN.

Número de placas	-	1	2	3	4	5	6	7	8
Número de geles	-	1/2	1/2	1	1	11/2	11/2	2	2
TAE (ml)	100	150	150	300	300	450	450	600	600
Agarosa	0.8	1.2	1.2	2.4	2.4	3.6	3.6	4.8	4.8
Bromuro (μl)	0.8	7.5	7.5	15	15	22.5	22.5	30	30

TAE= tris-acetato EDTA (amortiguador); ml= Mililitro; μl= Microlitro.

**Cuadro 11.** Preparación de gel de agarosa (01.2 %) para productos de PCR.

Número de placas	-	1	2	3	4	5	6	7	8
Número de geles	-	1/2	1/2	1	1	11/2	11/2	2	2
TAE (ml)	100	150	150	300	300	450	450	600	600
Agarosa	1.2	1.8	1.8	3.6	3.6	5.4	5.4	7.2	47.2
Bromuro (μl)	0.8	7.5	7.5	15	15	22.5	22.5	30	30

TAE= tris-acetato EDTA (amortiguador); ml= Mililitro; μl= Microlitro.