



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**CARACTERIZACIÓN DE LAS RAZAS DE MAÍZ (*Zea  
mays L.*) TEHUA, ZAPALOTE CHICO, ZAPALOTE  
GRANDE, COMITECO Y CONEJO, POR MÉTODOS  
GENÓMICOS**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**BIÓLOGO**

**P R E S E N T A:**

**JOSÉ LUIS CALDÚ PRIMO**



**DIRECTOR DE TESIS:**

**DRA. ANA LAURA WEGIER BRIUOLO**

**2016**

1. Datos del alumno

Caldú

Primo

José Luis

55 49 25 15

Universidad Nacional Autónoma de México

Facultad de Ciencias

Biología

307746450

2. Datos del tutor

Dra

Ana Laura

Wegier

Briuolo

3. Datos del sinodal 1

Dr

Jorge

Nieto

Sotelo

4. Datos del sinodal 2

Dr

Lev Orlando

Jardón

Barbolla

5. Datos del sinodal 3

Dra

Alicia

Mastretta

Yanes

6. Datos del sinodal 4

Dra

Alma Amparo

Piñeyro

Nelson

7. Datos del trabajo escrito

Caracterización de las razas de maíz (*Zea mays L.*) Tehua, Zapalote Chico, Zapalote Grande, Comiteco y Conejo, por métodos genómicos

64 p.

2016

## **Agradecimientos**

Primero, agradezco a la UNAM y a la Facultad de Ciencias en donde, no sólo me convencieron del acierto de seguir una carrera científica, sino que por su carácter de universidad pública pude conocer mejor la situación mi país y situarme en ella como una persona consciente.

Gracias a la CONABIO, a la Dirección General del Sector Primario y Recursos Naturales Renovables (DGSPRNR) perteneciente a la SEMARNAT y al Dr. Daniel Piñero por el apoyo financiero y las facilidades otorgadas para hacer este trabajo dentro de la segunda etapa del proyecto “Monitoreo de las razas de maíz en México usando un enfoque genómico, una estrategia para su uso sustentable y su conservación”. Espero que mi trabajo signifique una buena inversión de sus recursos.

Agradezco al CIMMYT por las semillas otorgadas y por el gran trabajo que hacen, sin el que mucha investigación sobre maíz no sería posible.

Siguiendo con los indispensables, agradezco a Ana. El apoyo constante y los consejos que me has dado me han permitido conocer los caminos que puedo seguir como biólogo y ubicarme como tal en la sociedad.

Esta tesis no la habría podido acabar sin la ayuda de Alicia y de José, gracias por enseñarme a sacar información de las marañas de datos. Muchas gracias a mis otros sinodales Jorge, Lev y Alma por tomarse el tiempo de revisar mi tesis y darme comentarios y correcciones, que además de ayudarme a redondear la tesis, me dieron perspectivas que me ayudan a pensar en lo que hice.

Pasando a los agradecimientos más coloquiales, agradezco a todos mis amigos: los del moderno, los biólogos, los de la fac, los del lab, los del fut, mis primos, los de Suecia y los no clasificados.

Terminando con los más importantes, a Fersi por todo tu cariño y apoyo.

A Isa por escucharme aunque hable poco, a Tania y Anahí que siempre están conmigo a pesar de la distancia y a mis papás por ser siempre mi base y ejemplo.

## Índice

RESUMEN.....	1
INTRODUCCIÓN.....	2
HIPÓTESIS.....	9
OBJETIVOS.....	10
MATERIALES Y MÉTODOS.....	11
Selección de razas.....	11
Obtención de muestras.....	12
Extracción de ADN.....	13
Obtención de SNPs.....	15
Selección de marcadores candidatos.....	16
Análisis de componentes principales e identificación de marcadores diferenciados entre razas.....	18
Estructura altitudinal.....	21
Ubicación de los SNPs identificados en el genoma del maíz.....	21
RESULTADOS.....	23
Grupos de marcadores candidatos.....	23
Análisis de componentes principales.....	24
Marcadores útiles para distinguir entre tres razas.....	25
Clasificación de las muestras por su altitud.....	32
Análisis discriminante de las cinco razas.....	40
ANÁLISIS Y DISCUSIÓN.....	47
Agrupamiento por grupos de marcadores candidatos.....	47
Identificación de marcadores útiles para distinguir entre razas.....	48
CONCLUSIÓN.....	54
PERSPECTIVAS.....	55
REFERENCIAS.....	58

## RESUMEN

México es centro de domesticación y diversificación del maíz (*Zea mays* subsp. *Mays*). Durante este proceso que ha durado alrededor de 10,000 años, se ha acumulado una enorme diversidad fenotípica del cultivo. Actualmente se reconocen 59 razas nativas de maíz en México, estas son poblaciones que mantienen características en común que las distinguen del resto de las poblaciones. En trabajos previos se ha reportado que la variación genómica del maíz mexicano se explica mejor por factores ambientales (altitud) que por categorías raciales. En este trabajo se compararon datos genómicos (46,518 SNPs obtenidos con el MaizeSNP50 BeadChip de Illumina) de 50 muestras de cinco razas de maíz (Comiteco, Conejo, Tehua, Zapalote Grande y Zapalote Chico) buscando marcadores que sirvan para clasificar a las muestras por raza. Se realizaron análisis discriminantes de componentes principales (DAPC) agrupando a las muestras por razas y por categorías altitudinales, a partir de éstos se obtuvieron los marcadores más informativos para clasificar a las muestras por raza y por categoría altitudinal. Se identificaron 18 *loci* que sirven para clasificar a las muestras por raza, de los cuáles 5 *loci* coinciden con marcadores que distinguen entre categorías altitudinales. En otras palabras, se encontraron 13 *loci* asociados a la clasificación por razas independientemente del factor altitudinal. Estos resultados muestran que las razas de maíz mexicano contienen una gran diversidad genómica que no puede ser explicada únicamente por factores ambientales, sino que probablemente está también influida por la selección artificial de los campesinos que lo cultivan.

## INTRODUCCIÓN

La domesticación ha sido estudiada como un proceso evolutivo desde el planteamiento de la teoría de selección natural por Charles Darwin en 1859 (Darwin, 1859). Desde entonces la domesticación ha recibido especial atención desde la Biología Evolutiva al ser el ejemplo más cercano y evidente de cambio en las especies.

Dependiendo de la amplitud de la definición, la domesticación puede ser vista como un proceso guiado exclusivamente por el humano o extendido a otras especies. Paul Gepts la define como “el resultado de un proceso de selección que lleva a un aumento en la adaptación de plantas y animales por el cultivo o crianza y utilización por los humanos” (Gepts, 2004). Otros autores, como Meyer y Purugganan, dan una definición más amplia “La domesticación es una interacción coevolutiva que lleva al establecimiento de nuevas especies domesticadas, cuyo crecimiento y reproducción están en gran medida controladas para el beneficio de otra especie.” Tomando esta segunda definición, los autores identifican, además de la domesticación por el hombre, el surgimiento de la domesticación en especies de insectos como hormigas, escarabajos y termitas que cultivan colonias de hongos (Meyer *et al.*, 2013; Purugganan *et al.*, 2009). Es importante resaltar el sentido coevolutivo de la domesticación cuando se habla de plantas domesticadas; ya que a pesar de que a primera vista parece existir una relación vertical en la que los cultivos dependen del hombre para sobrevivir, igualmente los humanos dependen de los cultivos para mantener sus poblaciones.

La evolución de plantas en domesticación comenzó hace alrededor de 10,000 años asociado a cambios ambientales en el planeta y cambios conductuales en los grupos humanos existentes que los llevaron a pasar de la caza-recolecta al cultivo de plantas como la fuente principal de alimentos (Gepts, 2004; Meyer *et al.*, 2013). Esta transición fue la base

de los cambios civilizatorios de la Revolución Neolítica. Acompañando la transición al cultivo de plantas como fuente principal de alimentos, surgieron nuevas técnicas, herramientas y oficios diferentes a buscar comida que permitieron a las sociedades humanas tener el desarrollo subsecuente que llevado al mundo hasta las condiciones en que nos encontramos hoy en día (Stiner, 2001; Weisdorf, 2005). Meyer y Purugganan (2013) han propuesto un patrón general del proceso de domesticación que identifica diferentes etapas de domesticación y diversificación en los cultivos. De acuerdo a este modelo, la primera etapa en la domesticación sucedió cuando los humanos empezaron a plantar y cultivar deliberadamente especies silvestres con caracteres de interés, convirtiendo a esas especies silvestres en especies cultivadas en proceso de domesticación. Siguiendo la domesticación inicial, vino una fase que involucró la dispersión y adaptación de las especies domesticadas a ambientes ecológicos y culturales diferentes. Esta segunda fase fue guiada por migraciones humanas y transmisión cultural. Esta segunda etapa llevo a una divergencia genética y fenotípica entre las poblaciones de especies domesticadas que fueron objeto de diferentes presiones selectivas. La última etapa consiste en el cruzamiento deliberado de variedades de un cultivo para maximizar la producción y generar nuevas variedades con características particulares (Meyer y Purugganan, 2013).

Durante el proceso de domesticación y diversificación de los cultivos, las especies domesticadas tienen cambios drásticos a nivel genético. En términos generales, la domesticación implica un cuello de botella genético que reduce la variación de las especies domesticadas con respecto a sus parientes silvestres. Durante este proceso, las especies son objeto de presiones selectivas conscientes e inconscientes que hacen que la distribución de la variación genética cambie. La presión selectiva consciente es llevada a cabo por los agricultores que seleccionan positiva o negativamente caracteres fenotípicos y consecuentemente los genes asociados a estos caracteres. La presión

selectiva inconsciente sucede porque la transición de las plantas cultivadas a los espacios donde el hombre las cultiva implica cambios en las condiciones en que esas plantas viven y se reproducen, los cuales ejercen una presión selectiva sin que el hombre tenga un objetivo predeterminado (Wright *et al.*, 2005; Ross-Ibarra, 2007; Zohary, 2004). Sin embargo, para que exista selección inconsciente, el hombre tiene que deliberadamente sembrar semillas en un ambiente controlado por el hombre, lo que algunos autores llaman domesticación del paisaje. Estas acciones son conscientes poniendo en duda el grado de inconsciencia asociado a la selección de ciertos caracteres (Heiser, 1988; Baleé y Erickson, 2006). Sin importar el modo en que se dé la selección, las diferentes presiones selectivas y los métodos de propagación de los cultivos actúan sobre el genoma ocasionando barridos selectivos que reducen la variación genética en los sitios del genoma asociados a las características que están siendo seleccionadas. A partir de la década de 1990, se han realizado muchos estudios que buscan identificar regiones del genoma seleccionadas durante la domesticación de diversas especies (Meyer *et al.*, 2013; Purugganan *et al.*, 2009; Hufford *et al.*, 2012; Yamasaki *et al.*, 2007).

El maíz (*Zea mays* subsp. *mays*) fue domesticado en México a partir del teocintle (*Zea mays* subsp. *parviglumis*) en la cuenca del Balsas entre 6000 y 9000 años antes del presente (Yamasaki *et al.*, 2005; Piperno *et al.*, 2009; van Heerwardeen *et al.*, 2011; Vigoroux *et al.*, 2008). Posterior a su domesticación, el maíz fue dispersado a otros lugares de Mesoamérica, se cree que esta dispersión se dio en un período corto de tiempo por los restos arqueobotánicos que se han encontrado (Piperno *et al.*, 2009; van Heerwardeen *et al.*, 2011). La dispersión del maíz, así como el uso que le dan los diferentes grupos humanos que lo adoptaron, están asociados a la adaptación del cultivo a los nuevos ambientes agroecológicos en los que se cultiva. Este es un proceso complejo de diversificación en el que hay adaptación local a partir de la variación genética disponible, introducción de nueva



variación por flujo génico y sistemas de intercambio de semilla e introgresión de genes de parientes silvestres. El resultado de esta diversificación es la enorme variedad fenotípica y los cientos de razas que existen de maíz (van Heerwarden *et al.*, 2011; Hufford *et al.*, 2012; Dyer *et al.*, 2008).

La diversidad de maíz en México ha sido descrita y clasificada por diversos autores durante el siglo XX. Estos esfuerzos fueron guiados principalmente por el interés de utilizar los recursos genéticos del maíz en el mejoramiento agronómico (Perales y Golicher, 2014). El concepto de raza de maíz fue concebido durante esta búsqueda por describir y clasificar la heterogeneidad que existe en el maíz. Dos trabajos pioneros en este sentido fueron las monografías de Chávez en 1913 y de Anderson en 1946 (Ortega Paczka, 2003). Anderson y Cutler en 1942 definieron el concepto de raza de maíz como “Un grupo de individuos emparentados, con suficientes características en común para permitir su reconocimiento como grupo”. Posteriormente Efraín Hernández Xolocotzi y Glafiro Alanís en 1970 ampliaron la definición de raza a “una población con un conjunto sustancial de características en común que la distinguen como grupo y la diferencian de otras poblaciones, con capacidad de transmitir con fidelidad dichas características a las generaciones posteriores y que ocupa un área ecológica específica”.

A partir de estas definiciones de raza de maíz se han llevado a cabo trabajos de clasificación del maíz en sus razas. El trabajo de clasificación de Wellhausen *et al.* en 1951 fue el primer estudio organizado y sistemático de la diversidad del maíz en México. En ese trabajo se describieron 25 razas clasificándolas en cuatro grupos y tres subrazas y se dejaron pendientes siete tipos como “no bien definidas”, utilizando caracteres morfológicos, fisiológicos genéticos y citológicos (Mera Ovando, 2009; Ortega Paczka, 2003). Hernández Xolocotzi y Alanís en 1970 identificaron cinco razas más del norte y noroeste de México. Más adelante, Benz con estudios de sistemática numérica de caracteres

morfológicos agrupa a los maíces mexicanos en 30 razas, describiendo 5 razas nuevas (Benz, 1986). En 1991, Ortega Paczka *et al.* propone una clasificación en 41 razas de maíz. El número de razas identificadas para México aumentó a 59 siguiendo el trabajo de Sánchez *et al.* en el que suma las 35 razas de Wellhausen *et al.* (1951), cinco de Hernández y Alanís (1970), diez de Ortega *et al.* (1991) y nueve de Benz (1986) (Sánchez *et al.*, 2000; Ortega Paczka, 2003). Actualmente, la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) reconoce 64 razas de maíz colectadas en México, de estas 59 se consideran nativas y 5 razas descritas inicialmente en otros países. Dichas 64 razas están agrupadas en siete grupos o complejos raciales, con base en caracteres morfológicos y genéticos (CONABIO, 2011).

Es importante resaltar que las razas del maíz son categorías artificiales creadas por el hombre y reflejan un nivel de variación dentro de la heterogeneidad del maíz. En el maíz se encuentra variación a nivel de individuos, poblaciones locales, subrazas, razas y grupos raciales. Las poblaciones locales de maíz son un grupo de individuos que se desarrollan lo bastante cerca unos de otros para efectuar cruces e intercambiar genes y diferenciadas por la selección de semillas que ejerce el agricultor que las siembra. En este sentido las razas son un conjunto de poblaciones locales con un conjunto de características en común que las diferencian de otras razas. La variación entre las poblaciones que forman una raza puede ser muy grande y es difícil encontrar poblaciones locales que no presenten características combinadas de diferentes razas (Ortega Paczka, 2003). Sin embargo, las razas de maíz son reconocidas por expertos en el tema y son usadas por la utilidad y necesidad de organizar la variación del maíz para su estudio.

El estudio de la diversidad del maíz mexicano es importante para entender la evolución del maíz, al ser México su centro de origen, y porque su material genético puede servir como fuente para el

mejoramiento del maíz para nuevas condiciones ambientales (Goodman *et al.*, 1988; Arteaga *et al.*, 2015). Las razas mexicanas de maíz presentan una enorme diversidad resultado de la heterogeneidad cultural y ambiental en las que evolucionan. En un análisis global de las razas mexicanas se encontró que la distribución de la variación genética en muestras de maíz mexicano se explica mejor por variables ambientales resultado de la interacción entre latitud y altitud, que por asignaciones de raza (Arteaga *et al.*, 2015; Breña Ochoa, 2013). Sin embargo, el reconocimiento de razas por expertos en maíz y su caracterización e identificación a partir de muestras en tiempos y lugares diferentes, hace pensar que debe de haber diferencias genéticas que subyacen las características distintivas de las diferentes razas. En este sentido, se han realizado trabajos buscando encontrar la “huella genética” de las razas de maíz (SAGARPA, 2014), sin embargo no existe todavía una distinción de las razas de maíz mexicano por marcadores moleculares. Una explicación alternativa a estos resultados es que el concepto de razas de maíz es una categoría artificial impuesta sobre la heterogeneidad fenotípica del maíz como una forma de clasificarla pero que la variación genómica no responda realmente a las categorías raciales.

En el presente estudio se compararon datos genómicos de muestras de cinco razas de maíz (Comiteco, Conejo, Tehua, Zapalote Chico y Zapalote Grande) buscando marcadores que sirvieran para clasificar a las muestras por su raza asignada. Se obtuvieron marcadores que diferencian muestras de tres razas (Tehua, Zapalote Chico y Zapalote Grande), marcadores que diferencian a las muestras por su categoría altitudinal y marcadores que diferencian entre las cinco razas incluidas en el estudio. La intención de obtener estos tres grupos de marcadores que distinguen entre diferentes categorías era averiguar si los marcadores que sirven para distinguir entre razas se pueden explicar por diferencias ambientales causadas por la altitud y si los mismos marcadores que sirven para distinguir entre tres razas se mantienen

cuando se agregan más razas. Se encontraron marcadores diferentes para distinguir entre razas y por categorías altitudinales, la mayoría de estos marcadores no se han caracterizado funcionalmente. Estos resultados muestran que en las poblaciones locales de maíz hay una gran diversidad genética mediada por condiciones ambientales y culturales, por lo que es importante conservar las condiciones por las cuales los procesos evolutivos mantienen la diversidad del maíz.

## **HIPÓTESIS**

Los caracteres morfológicos que definen las diferentes razas de maíz están asociados a genes seleccionados por factores ambientales y culturales en las poblaciones de una misma raza, por lo que existe una combinación de marcadores moleculares que distinguen muestras de diferentes razas.

## **OBJETIVOS**

### **Objetivo general**

Encontrar marcadores moleculares que sirvan para la clasificación de maíz de acuerdo a su raza asignada.

### **Objetivos particulares**

- 1.** Caracterizar muestras de cinco razas mexicanas de maíz de acuerdo a una combinación de marcadores genómicos que presentan.
- 2.** Identificar entre los marcadores que sirven para clasificar muestras por raza aquellos que son resultado de diferencias ambientales dadas por la altitud.

## MATERIALES Y MÉTODOS

### Selección de razas

La selección de razas se hizo a partir de la lista reconocida por CONABIO (2011) buscando una pareja de razas que presentaran caracteres fenotípicos contrastantes (Tabla 1). De este modo se seleccionaron las razas Zapalote Chico, del grupo Tropicales precoces, y Tehua, del grupo Maduración tardía. Estas dos razas tienen características fenotípicas completamente opuestas. Zapalote Chico tiene plantas y mazorcas muy cortas, maduración precoz y poca sensibilidad al fotoperiodo; por el contrario, Tehua tiene plantas altas, mazorcas grandes, maduración tardía y alta sensibilidad al fotoperíodo (Tabla 1). Posteriormente, se seleccionó una raza más de cada grupo de razas a los cuales pertenecían el primer par elegido, así se eligieron las razas Conejo de Tropicales precoces y Comiteco de Maduración tardía. Finalmente, se añadió la raza Zapalote Grande del grupo Dentados Tropicales, cuyo origen propuesto es la hibridación entre Zapalote Chico y Tehua (Wellhausen *et al.*, 1951; CONABIO, 2011).

Tabla 1. Características de las razas, información obtenida de Wellhausen *et al.*, 1951, CONABIO, 2011 y Ruiz Corral, *et al.*, 2013.

Raza	Zapalote Chico	Zapalote Grande	Tehua	Comiteco	Conejo
Grupo de razas	Tropicales Precoces	Dentados Tropicales	Maduración Tardía	Maduración Tardía	Tropicales Precoces

<b>Distribución geográfica</b>	Istmo de Tehuantepec, estados de Oaxaca y Chiapas. (0 a 880 m.s.n.m.)	Estados de Chiapas y Oaxaca (0-1200 m.s.n.m.)	Zonas altas de Chiapas. (600 a 1000 m.s.n.m.)	Alrededor de Comitán, Chiapas y zonas altas de Oaxaca, (820 a 2100 m.s.n.m.).	Cuenca del Balsas, Oaxaca, Guerrero y Michoacán. (74 a 2000 m.s.n.m.)
<b>Planta</b>	Muy corta (1-2 m). Pocas hojas.	Corta a intermedia (1.5 a 2.5 m). Número medio de hojas	Muy alta, (hasta 6 m. de altura). Muchas hojas.	Muy alta (4 - 5 m), tallos gruesos y hojas numerosas.	Corta (1.6 - 1.9 m)
<b>Maduración</b>	Precoz	Intermedio	Muy tardía	Tardía	Precoz
<b>Mazorca y tipo de grano</b>	Muy pequeña, con 10 a 12 hileras, granos cortos y dentados	Corta de grano dentado.	Diámetro grande, grano grande, dentado y semicristalino.	Largas y gruesas con granos semicristalinos a dentados.	Corta y delgada (12 - 18 cm) con 8 a 10 hileras. Granos cristalinos a semidentados medianos,
<b>Origen propuesto</b>	Derivado de Nal-Tel y Tepecintle.	Derivado de Zapalote Chico con introgresión de Tehua.		Derivado de Tehua y Olotón.	Derivado de Nal-Tel y Tabloncillo.

### Obtención de muestras

Las semillas para este estudio fueron proporcionadas por el Banco de Germoplasma del Centro Internacional para el Mejoramiento de Maíz y Trigo (CIMMYT). Se solicitaron todas las accesiones disponibles en el banco de germoplasma del CIMMYT para estas 5 razas que no presentaran influencia de otras razas. En total se recibieron 37 muestras de 5 razas de maíz (4 a 10 muestras por raza); colectadas, entre 1946 y



2010, en localidades de 4 entidades federativas de México (Chiapas, Guerrero, Michoacán y Oaxaca) y Guatemala (Tabla 2).

Las semillas fueron germinadas en un invernadero del Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP). Tres semillas de cada muestra se sembraron en sustrato (50% agrolita, 50 % turba), dentro de un almácigo de 50 tubos. Todas las muestras se sembraron durante febrero de 2015. Todas las muestras sembradas germinaron y fueron mantenidas en almácigo dentro del invernadero con riego diario hasta que se realizaron las extracciones. La extracción se realizó entre tres y cuatro semanas después de la germinación a partir de tejido foliar de una planta germinada de cada muestra. Adicionalmente a estas 37 muestras, se agregaron los datos de 13 muestras de las cinco razas que fueron previamente genotipificadas siguiendo el mismo protocolo como seguimiento del Proyecto Global de Maíces Nativos de CONABIO (Arteaga *et al.*, 2015) (Tabla 2). Estas 13 muestras se agregaron durante el procesamiento bioinformático de los datos genómicos. El número total de muestras incluidas fue de 50 muestras.

### **Extracción de ADN**

La extracción de ADN genómico se hizo en el Laboratorio de Biotecnología Forestal del Instituto de Investigaciones Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) siguiendo el protocolo del kit de extracción de DNA genómico (K0512) de Thermo-Fisher Scientific (Waltham, Massachusetts, E.U.A.).

El extracto de ADN se purificó siguiendo el protocolo del kit de purificación de productos de PCR MinElute de Qiagen (Valencia, California, E.U.A.) y fue resuspendido en agua bidestilada. El ADN se visualizó mediante electroforesis en gel de agarosa al 1% teñido con bromuro de etidio en un transiluminador de luz ultravioleta con cámara digital. Antes de procesar las muestras para la obtención de SNPs, se

determinó su calidad mediante cuantificación en un espectrofotómetro Nanodrop 2000 de Thermo-Fisher Scientific (Waltham, Massachusetts, E.U.A.). Para ser enviadas a genotipificar, las muestras requirieron tener una relación de absorción de 260/280 nm mayor a 1.6. Las muestras que no cumplieron con esa característica fueron descartadas y se repitió la extracción. El ADN extraído se mantuvo en refrigeración a  $-20^{\circ}\text{C}$  hasta el traslado para su genotipificación.

Tabla 2: Datos de colecta de las muestras utilizadas

Muestra	Origen	Raza	Año	Estado	Altura (msnm)	Muestra	Origen	Raza	Año	Estado	Altura (msnm)
14_2166	CIMMYT	Comiteco	1946	Chiapas	100	15_2270	CIMMYT	Zapalote Chico	1946	Oaxaca	100
17_10037	CIMMYT	Comiteco	1952	Guatemala	1943	18_10463	CIMMYT	Zapalote Chico	1972	Chiapas	700
32_26872	CIMMYT	Comiteco	1972	Chiapas	900	19_10473	CIMMYT	Zapalote Chico	1973	Chiapas	100
33_29890	CIMMYT	Comiteco	2010	Chiapas	1510	22_23130	CIMMYT	Zapalote Chico	1998	Oaxaca	681
35_29971	CIMMYT	Comiteco	2010	Chiapas	1955	23_23142	CIMMYT	Zapalote Chico	1998	Oaxaca	255
36_30008	CIMMYT	Comiteco	2010	Chiapas	1631	26_23547	CIMMYT	Zapalote Chico	1999	Oaxaca	8
37_3052	CIMMYT	Comiteco	2010	Chiapas	2008	28_23597	CIMMYT	Zapalote Chico	1999	Oaxaca	5
8_805	CIMMYT	Comiteco	1961	Chiapas	2241	29_25029	CIMMYT	Zapalote Chico	1976	Oaxaca	50
9_1058	CIMMYT	Comiteco	1961	Chiapas	1400	34_29952	CIMMYT	Zapalote Chico	2009	Chiapas	573
maiz_150	Arteaga et al.	Comiteco	2009	Chiapas	1550	7_802	CIMMYT	Zapalote Chico	1961	Chiapas	100
maiz_183	Arteaga et al.	Comiteco	2009	Chiapas	1550	maiz_142	Arteaga et al.	Zapalote Chico	2009	Chiapas	654
maiz_186	Arteaga et al.	Comiteco	2009	Chiapas	1550	10_1222	CIMMYT	Zapalote Grande	1985	Chiapas	-
maiz_80	Arteaga et al.	Comiteco	2009	Chiapas	455	11_1773	CIMMYT	Zapalote Grande	1946	Chiapas	100
maiz_82	Arteaga et al.	Comiteco	2009	Chiapas	1553	16_2601	CIMMYT	Zapalote Grande	1962	Chiapas	31
1_33	CIMMYT	Conejo	1946	Guerrero	210	21_16210	CIMMYT	Zapalote Grande	1972	Chiapas	41
2_44	CIMMYT	Conejo	1947	Guerrero	480	24_25020	CIMMYT	Zapalote Grande	1976	Oaxaca	50
3_68	CIMMYT	Conejo	1960	Guerrero	180	25_23217	CIMMYT	Zapalote Grande	1962	Oaxaca	820
5_168	CIMMYT	Conejo	1952	Michoacán	250	27_23552	CIMMYT	Zapalote Grande	1999	Oaxaca	8
maiz_152	Arteaga et al.	Conejo	2008	Guerrero	1493	4_137	CIMMYT	Zapalote Grande	1958	Chiapas	1000
maiz_182	Arteaga et al.	Conejo	2008	Guerrero	1402	maiz_144	Arteaga et al.	Zapalote Grande	2009	Chiapas	654
maiz_21	Arteaga et al.	Conejo	2008	Guerrero	1312						
maiz_23	Arteaga et al.	Conejo	2008	Guerrero	1685						
12_1849	CIMMYT	Tehua	1953	Guatemala	122						
13_1861	CIMMYT	Tehua	1952	Guatemala	792						
20_13297	CIMMYT	Tehua	1972	Chiapas	820						
30_23861	CIMMYT	Tehua	1947	Chiapas	1604						
31_25784	CIMMYT	Tehua	1972	Chiapas	730						
6_777	CIMMYT	Tehua	1944	Chiapas	1400						
maiz_124	Arteaga et al.	Tehua	2009	Chiapas	1545						
maiz_125	Arteaga et al.	Tehua	2009	Chiapas	1525						

## **Obtención de SNPs**

La obtención de los SNPs se llevó a cabo en la Unidad de Microarreglos del Instituto Nacional de Medicina Genómica (INMEGEN). El protocolo que se siguió fue el del ensayo Infinium HD de Illumina que usa el MaizeSNP50 BeadChip (Illumina, San Diego, California, E.U.A.) para detectar los alelos de 56,110 SNPs para cada muestra.

Los SNPs detectados mediante este microarreglo se localizan en más del 50% de los 17,520 genes de alta confiabilidad, localizados en los 10 cromosomas, del genoma de referencia de la línea mejorada endogámica B73 (Schnable et al. 2009; Ganai et al. 2011). En teoría, una distribución homogénea de 50,000 SNPs en el genoma de este cultivo da como resultado en promedio 40 a 50 kilopares de bases entre marcadores, sin embargo, la mayoría de los SNPs vecinos detectados por el chip varían en la cantidad de nucleótidos que los separan de menos de 40 a más de 100 kilopares de bases (Ganal et al. 2011). En el diseño de este chip hay un sesgo hacia SNPs en regiones genómicas que se ha reportado tienen asociación con alguna característica de interés agronómico, esto implica que pueden haber regiones con variación entre razas nativas que no van a estar representadas.

La identificación y el análisis de calidad de los 56,110 SNPs obtenidos de 37 muestras se hizo con el módulo de genotipificación (Genotyping (GT) Module v.1.7.4) del programa GenomeStudio v.2010.1 de Illumina ([www.illumina.com/software/genomestudio\\_software.ilmn](http://www.illumina.com/software/genomestudio_software.ilmn)). El algoritmo GenCall compara la intensidad de la señal de los dos colores fluorescentes, que representan los dos alelos de un SNP, con agrupaciones de genotipos canónicos. Los genotipos obtenidos son representados gráficamente, de acuerdo a la frecuencia alélica (NormTheta; eje x) en relación a la intensidad de la señal (NormR; eje y), como puntos de color diferente dependiendo de si son homocigotos o heterocigotos. Se aplicaron dos filtros de calidad de SNPs, primero se eliminaron los SNPs para los que más del 30% de las muestras no

podieron ser asignadas a un grupo (missingcalls > 30%), el segundo filtro fue eliminar los SNPs para los que la separación entre los grupos fuera menor a 0.3 (separation score < 0.3). En total, 9,592 SNPs no cumplieron con estos requisitos y fueron eliminados, resultando 46,518 SNPs por muestra.

Posteriormente, las muestras con los 46,518 SNPs fueron procesadas en el programa PLINK v.1.07 (Purcell et al. 2007). A las 37 muestras se le sumaron los datos de 13 muestras de estas razas que fueron previamente genotipificadas como seguimiento del Proyecto Global de Maíces Nativos de CONABIO (Arteaga *et al.*, 2015). Las muestras adicionales fueron genotipificadas con el mismo chip y usando los mismos filtros de calidad, resultando en 36,931 SNPs. De esos 36,931 SNPs se eliminaron 1,022 SNPs que no estaban entre los 46,518 SNPs de las muestras genotipificadas para este trabajo. El número final de SNPs fue 37 muestras con 46,518 SNPs y 13 muestras con 35,909 SNPs.

### **Selección de marcadores candidatos**

Para la búsqueda de marcadores asociados a las características fenotípicas distintiva de cada raza, se definieron un grupo de marcadores reportados de domesticación y mejoramiento en maíz y un grupo de marcadores con  $F_{ST}$  alta entre las razas. El índice  $F_{ST}$  es una medida de la diferenciación genética entre subdivisiones de una población, esta medida es utilizada ampliamente en genética de poblaciones. El valor de  $F_{ST}$  puede ir de 0 a 1. Una  $F_{ST}$  de cero indica que las frecuencias alélicas son idénticas en todas las poblaciones estudiadas, o sea que no ha habido nada de diferenciación entre ellas, mientras más alto es el valor de  $F_{ST}$ , significa que hay mayor diferenciación entre las subpoblaciones (Hedrick, 2011). Calculando la  $F_{ST}$  de todos los SNPs agrupando a las muestras por raza, los que presenten valores más altos serán los SNPs que mayor diferenciación tienen entre las razas.

El grupo de marcadores de domesticación y mejoramiento se basa en los marcadores reportados por Meyer *et al.*, 2013 y Hufford *et al.*, 2012. El primer conjunto se desprende del artículo de revisión publicado por Meyer *et al.*, 2013 en el que reúnen 60 *loci* que han sido reportados como genes de domesticación o diversificación en diferentes especies. Estos *loci* fueron seleccionados por haber estado incluidos en estudios de genética de poblaciones y/o haber sido analizados y validados funcionalmente para sustentar su rol en la evolución de los cultivos. De los 60 *loci*, 23 fueron clasificados como genes de domesticación, 32 como genes de diversificación y 5 genes que fueron seleccionados tanto durante domesticación como durante diversificación. (Meyer *et al.*, 2013). A partir de esos 60 *loci* reportados para diferentes especies se buscaron los *loci* presentes en maíz haciendo un *Basic Local Alignment Search Tool* (BLAST) de nucleótidos con el algoritmo *megablast* en el portal (Altschul *et al.*, 1990; <http://blast.ncbi.nlm.nih.gov/>), el criterio de inclusión fue una cobertura mayor al 50% del gen en el maíz.

El segundo conjunto fue tomado del artículo de Hufford *et al.*, 2012 en el que realizaron un análisis de 75 genomas de maíz y sus parientes silvestres (17 parientes silvestres, 23 razas y 35 líneas mejoradas). Para identificar regiones del genoma seleccionadas durante la evolución del cultivo, hicieron un análisis estadístico que rastrea diferencias extremas en frecuencias alélicas en regiones con desequilibrio de ligamiento extenso. Esto se basa en el principio de que los eventos de domesticación y mejoramiento ocasionan cuellos de botella genómicos que tienen como efecto un decremento en la diversidad genética, cambios en frecuencias alélicas y un incremento en el desequilibrio de ligamiento. Para designar *loci* de domesticación se identificaron barridos selectivos entre los genomas de parientes silvestres y razas de maíz, mientras que para designar *loci* de diversificación o mejoramiento se identificaron barridos selectivos entre genomas de razas y líneas mejoradas. A partir de estos análisis se formaron dos conjuntos de regiones seleccionadas con 484 regiones de domesticación y 695

regiones de mejoramiento. Dentro de estos barridos de selección, se propone como gen candidato de selección al que se encontrará en el centro de la región con el índice de diferenciación más alto (Hufford *et al.*, 2012). Para este estudio se utilizaron únicamente los genes candidatos de cada región propuesta.

Una vez que se tuvieron los *loci* reportados para domesticación y mejoramiento, se ubicaron los SNPs del MaizeSNP50 BeadChip que se encontraran en el rango genómico de esos *loci*. Esto se realizó a través del buscador de marcadores de Panzea (Zhao *et al.*, 2005, <http://www.panzea.org/>).

El grupo de marcadores con  $F_{ST}$  alta se obtuvo a partir de los datos genómicos de SNPs obtenidos para las muestras. Se agruparon las muestras por raza y se calculó el valor de  $F_{ST}$  para cada uno de ellos. El cálculo del índice  $F_{ST}$  de cada SNP se llevó a cabo con el paquete *snpStats* (Clayton, 2015) en el programa R v.2.15.2 (R Development Core Team, 2008). El grupo de marcadores con  $F_{ST}$  alta se definió tomando el 1% de SNPs con mayor valor de  $F_{ST}$  entre las razas. Para comprobar que los grupos de SNPs de domesticación y de SNPs con  $F_{ST}$  alta no se traslapan, se ubicaron a los SNPs de domesticación de acuerdo a sus valores de  $F_{ST}$  y se contaron los que pertenecían al grupo de SNPs con  $F_{ST}$  alta.

### **Análisis de componentes principales e identificación de marcadores diferenciados entre razas**

Usando los tres grupos de SNPs previamente definidos, se realizó un análisis de componentes principales buscando si había un grupo de SNPs que rescatara mejor el agrupamiento por raza. El análisis de componentes principales se llevo a cabo con el paquete *SNPRelate* (Zheng *et al.*, 2012) de R v.2.15.2 (R Development Core Team, 2008)

Para buscar marcadores que sirvan para distinguir entre razas, se realizó un análisis discriminante de componentes principales (DAPC, por

sus siglas en inglés). El DAPC es un método multivariado diseñado para identificar y describir agrupaciones de individuos genéticamente relacionados. Este método requiere datos genéticos de los individuos e información sobre el agrupamiento de estos individuos que se busca describir. El análisis consiste en generar un modelo de clasificación con funciones lineales que, tomando los componentes principales como variables, disminuyan la variación intragrupal y aumenten la variación entre los grupos (Jombart *et al.*, 2010). Este análisis se realizó utilizando el paquete *adegenet* (Jombart, 2008) en R v.2.15.2 (R Development Core Team, 2008).

Las variables que utiliza el DAPC son componentes principales obtenidos a partir de los SNPs, esto con la intención de disminuir el número de variables manteniendo la mayor cantidad de información posible. Un paso previo al DAPC es un agrupamiento por modelo buscando encontrar agrupaciones de individuos basadas únicamente en los datos genéticos, sin información previa. El modelo que se utiliza es agrupamiento por *k-means* y el número óptimo de grupos se decide por el criterio de información Bayesiana. El agrupamiento por modelo da una idea de qué tan sustentados están los grupos propuestos de acuerdo a la información genética. Para continuar con el DAPC se puede elegir entre utilizar los grupos resultantes del agrupamiento por modelo o utilizar otro agrupamiento. Otro parámetro que se tiene que tomar en cuenta para correr el análisis es el número de componentes principales a utilizar. Al realizar un modelo usando DAPC hay un compromiso entre la precisión al separar los grupos y la capacidad de generalizar el modelo. El número de componentes principales incluidos es el modulador en este compromiso; si se incluyen muchos componentes principales se retiene mucha variación y el modelo es muy preciso para separar los grupos pero pierde generalidad, si se toman pocos el modelo se puede generalizar pero pierde precisión. El paquete *adegenet* incluye el cálculo de *a-score*, una medida para solucionar este compromiso. Esta medida consiste en medir la diferencia entre los grupos formados por el análisis

y grupos formados al azar, repitiendo esta medida para un rango de componentes principales retenidos, el número de componentes principales para el cuál se obtenga un valor mayor para la *a-score* es el número óptimo de componentes principales a retener (Jombart *et al.*, 2010).

Para evaluar la precisión del agrupamiento por modelo, se calcularon la pureza y la entropía de éste. Estas medidas miden la certeza de un agrupamiento por modelo cuando se tiene información previa de los grupos que se busca reconstruir. La pureza y la entropía buscan reflejar el número de casos en los que el agrupamiento por modelo difiere del grupo original (Amigó *et al.*, 2009). Ambas medidas van de 0 a 1, en la pureza un valor de 1 significa un agrupamiento perfecto y en la entropía un valor de 0 significa un agrupamiento perfecto. Para demostrar que la precisión del agrupamiento no puede ser resultado del azar, se creó una distribución aleatoria de estas dos medidas tomando 1000 grupos aleatorios de SNPs, agrupando de las muestras con el mismo modelo y midiendo la pureza y la entropía de cada agrupamiento.

Las funciones discriminantes del DAPC pueden ser usadas para asignar la pertenencia de cada individuo a los diferentes grupos. Esto se presenta calculando la probabilidad de pertenencia a los diferentes grupos, de acuerdo a los valores de las muestras para las funciones discriminantes. Estos resultados muestran qué tan diferenciados están los grupos entre sí, así como identificar muestras que genéticamente son una mezcla de diferentes grupos.

Se realizaron tres análisis discriminantes de componentes principales. El primero se hizo utilizando únicamente muestras de las razas Tehua, Zapalote Chico y Zapalote Grande. El primer análisis se hizo utilizando Tehua y Zapalote Chico por ser las primeras razas seleccionadas por sus características contrastantes y se incluyó Zapalote Grande por su origen propuesto de hibridización entre Zapalote



Grande y Tehua. Después se realizó un DAPC de todas las muestras agrupadas en las cinco razas para probar si los marcadores que diferenciaron entre tres razas (Zapalote Chico, Zapalote Grande y Tehua) son los mismos que diferencian entre las cinco razas. Por último se realizó un DAPC de todas las muestras agrupadas por categoría altitudinal para identificar marcadores que diferenciaran por altitud.

### **Estructura altitudinal**

Para poder descartar que la estructura se deba a la distribución altitudinal de las razas, se realizó un análisis discriminante agrupando a las muestras por categorías altitudinales en lugar de por razas. Las muestras se dividieron en las categorías “Alto” y “Bajo”. El rango de altitudes de las dos categorías fue definido a partir de la distribución altitudinal de las mismas muestras, tomando 750 m.s.n.m. como la división altitudinal entre las categorías (Bajo=0-750 m.s.n.m, Alto= 751-2241 m.s.n.m). Usando este criterio, se dividieron las muestras de las cinco razas en “Bajo” o “Alto” de acuerdo a la altitud a la que fueron colectadas, para comprobar que los grupos de muestras diferían significativamente en su distribución de altitudes se realizó una prueba  $U$  de Mann-Whitney. Se definió el grupo de SNPs con  $F_{ST}$  alta para la altitud tomando el 1% de SNPs con mayor  $F_{ST}$  entre las categorías altitudinales, a este grupo se le llamara en adelante SNPs de altitud. Utilizando este grupo de SNPs y los grupos altitudinales definidos se realizó un análisis discriminante igual al descrito en la sección anterior para categorías raciales.

### **Ubicación de los SNPs identificados en el genoma del maíz**

La posición de los SNPs que resultaron informativos en el DAPC se buscó en el portal de *SNPblast* (<http://www.ncbi.nlm.nih.gov/projects/SNP/SNPblast.html>). Ahí se obtuvo la posición genómica del marcador, su *loci* de referencia y el gen asociado, en caso de que tuviera. Teniendo el

*loci* de referencia y el gen asociado, la información se corroboró en el buscador de *locus* del portal *maizeGDB* ([http://www.maizegdb.org/locus\\_lookup](http://www.maizegdb.org/locus_lookup), Andorf *et al.* 2015). La información de la proteína asociada a los genes anotados fue obtenida a través del portal de información genómica de gramíneas *Gramene* (<http://www.gramene.org/>, Liang.C., *et al.* 2008).

## RESULTADOS

### Grupos de marcadores candidatos

Se definieron tres grupos de SNPs para buscar en ellos marcadores útiles para diferenciar entre razas. Los grupos son:

- Todos los SNPs: grupo de 46,518 SNPs que pasaron los filtros de calidad de genotipificación y llamado.
- SNPs de domesticación y mejoramiento: Grupo de 974 SNPs obtenidos a partir de la búsqueda de SNPs en genes reportados de domesticación y mejoramiento para maíz en Hufford *et al.*, 2013 y Meyer *et al.*, 2013. En los genes reportados por Meyer se ubicaron 84 SNPs, habiendo 5 genes que no estuvieron asociados a ningún SNP. En los *loci* reportados por Hufford se ubicaron 501 SNPs en *loci* de mejoramiento y 389 SNPs en *loci* de domesticación, con 295 y 251 *loci* sin SNPs asociados, respectivamente.
- SNPs con  $F_{ST}$  alta: grupo de 435 SNPs, que son el 1% de los SNPs con valores más altos de  $F_{ST}$  entre razas. La distribución de  $F_{ST}$  de todos los SNPs tiene un intervalo de 0 a 0.597, con una media de 0.056 y desviación estándar de 0.055. La mayoría de los SNPs tienen valores de  $F_{ST}$  cercanos a 0, indicando que no se diferencian entre las razas. El grupo de SNPs con  $F_{ST}$  alta está formado por 435 SNPs con  $F_{ST}$  mayor a 0.2283, éstos son el 1% de SNPs con valores de  $F_{ST}$  más altos, y por lo tanto los que más diferenciación entre las razas presentan (Figura 1, izquierda).

La mayoría de los SNPs de domesticación y mejoramiento presentan valores de  $F_{ST}$  relativamente bajos; solamente 6 están compartidos con el grupo de SNPs con  $F_{ST}$  alta, estos son: PZE-103025094,

SYN23628, SYN6642, PZE-107089944, PZE-102042184 y PZE-106035079 (Figura 1, derecha).

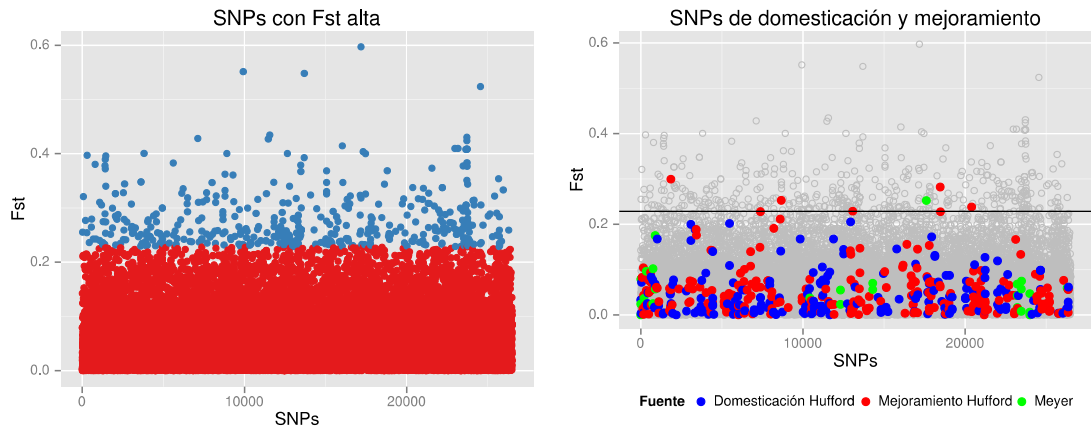


Figura 1:  $F_{ST}$  de los grupos de SNPs. Izquierda: SNPs con  $F_{ST}$  alta, grupo con 1% de SNPs con mayor  $F_{ST}$  en azul, resto de los SNPs en rojo. Derecha: SNPs de domesticación y mejoramiento, los SNPs de domesticación y mejoramiento se muestran de diferente color de acuerdo a su origen; la mayoría de estos SNPs tienen valores de  $F_{ST}$  bajos y sólo 6 superan el umbral de 0.2283.

## Análisis de componentes principales

### Análisis de componentes principales

Se realizó un análisis de componentes principales para cada grupo de SNPs. El porcentaje de varianza explicado por los primeros tres componentes principales para el análisis con cada uno de los tres grupos es de: Todos los SNPs 4.006, 2.943 y 2.800; SNPs de domesticación y mejoramiento 5.395, 3.807 y 3.554; SNPs con  $F_{ST}$  alta 18.016, 6.659 y 5.843, respectivamente.

En la gráfica de los primeros dos componentes principales (Figura 2), se puede observar que usando todos los SNPs y los SNPs con  $F_{ST}$  alta las muestras de una misma raza se localizan cercanas entre sí y hasta cierto punto separadas de las otras razas. En el análisis de componentes principales con SNPs de domesticación y mejoramiento no se pueden

identificar las razas por la posición de sus muestras en la gráfica. En el análisis de SNPs con  $F_{ST}$  alta, se alcanzan a distinguir bien cuatro grupos, Zapalote Chico, Zapalote Grande, Conejo y Tehua con Comiteco; por lo que se ve aquí, Tehua y Comiteco forman un sólo grupo genético. El análisis usando todos los SNPs muestra un patrón en el que algunas razas agrupan, pero en general no hay una separación entre los grupos de razas ni una localización de los grupos de la misma raza.

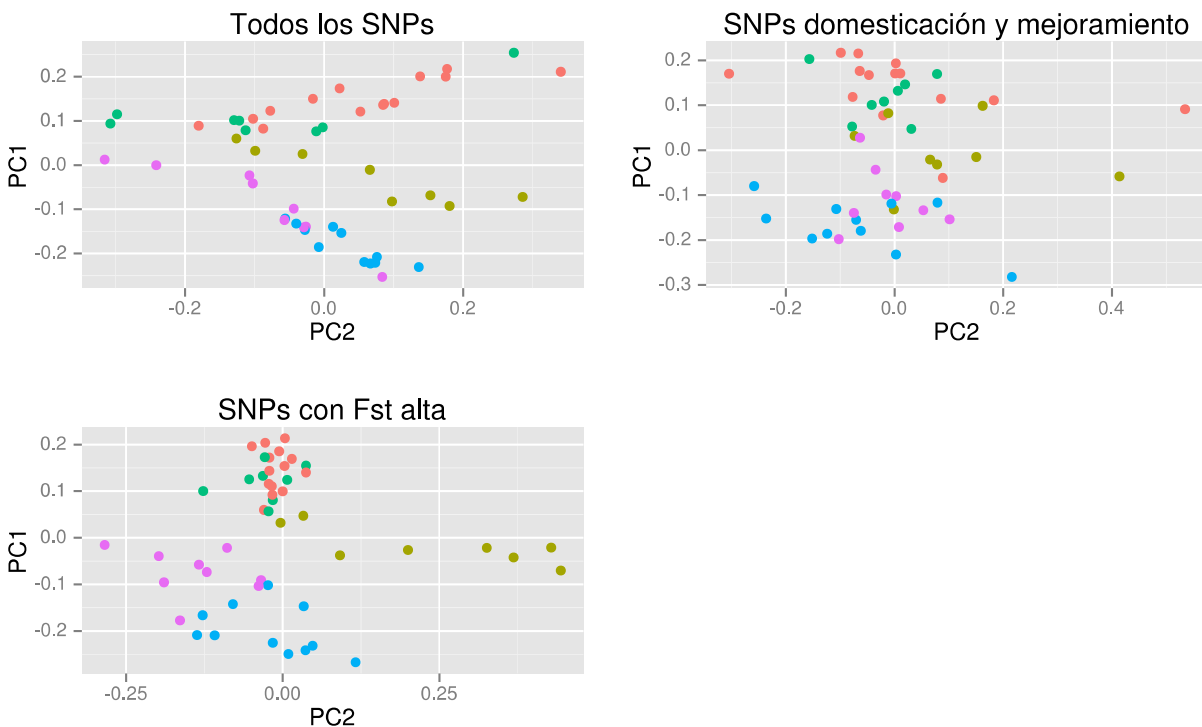


Figura 2: Gráficas con los dos componentes principales del análisis usando los tres grupos de SNPs. Las muestras están coloreadas de acuerdo a su raza.

### Marcadores útiles para distinguir entre tres razas

Se realizó el análisis discriminante de componentes principales (DAPC) utilizando los tres conjuntos de SNPs y únicamente las muestras de las

razas Tehua, Zapalote Chico y Zapalote Grande. Se utilizaron estas tres razas porque Tehua y Zapalote Chico fueron las primeras razas seleccionadas por sus características contrastantes y Zapalote Grande por su origen propuesto de hibridación entre Zapalote Grande y Tehua. El agrupamiento basado en modelo mostró resultados diferentes para los tres conjuntos de datos. Para el grupo de todos los SNPs y el grupo de SNPs de domesticación el mejor agrupamiento recuperado fue que todas las muestras formaran un sólo grupo ( $K=1$ ); por otro lado, para el grupo de SNPs con  $F_{ST}$  alta el mejor agrupamiento resultó utilizando tres grupos ( $K=3$ ) (Figura 3).

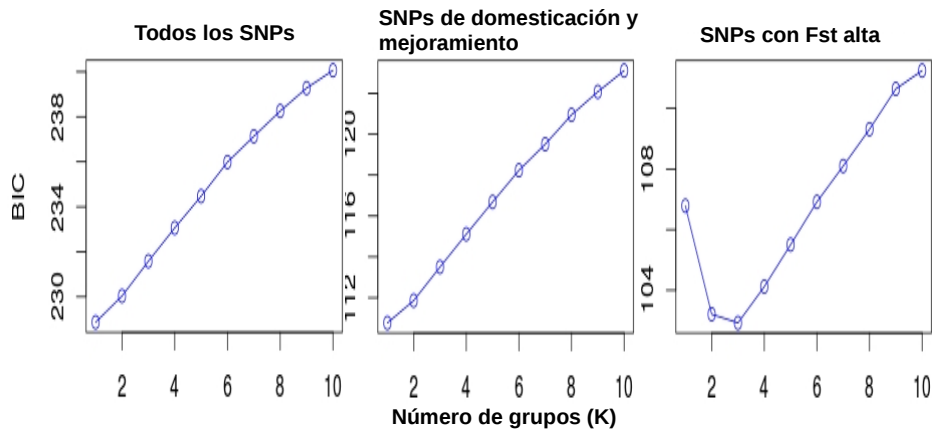


Figura 3: Agrupamiento por modelo usando los tres grupos de SNPs. El número de grupos para el que el criterio de información Bayesiana (BIC) es el menor corresponde al número de grupos óptimo. El grupo de SNPs con  $F_{ST}$  alta es el único para el que el mejor agrupamiento forma más de un grupo ( $K=3$ ).

Debido a que el grupo de los SNPs con  $F_{ST}$  alta es el único que sirve para agrupar a los datos, el resto del análisis se siguió únicamente utilizando a este grupo de datos. El agrupamiento por modelo de las muestras corresponde casi perfectamente al agrupamiento de estas por razas (Tabla 3). De acuerdo a esto, el grupo 1 corresponde a las muestras de raza Tehua, el grupo 2 a muestras de la raza Zapalote Chico y el grupo 3 a muestras de la raza Zapalote Grande. Sólo una muestra (maiz\_142) se encuentra en un grupo

diferente al de su raza asignada (Zapalote Chico), agrupándola con las muestras de Zapalote Grande. Esta muestra y la muestra maiz\_144 (Zapalote Grande) fueron colectadas en la misma localidad y del mismo productor (Mastretta-Yanes, comunicación personal), por lo que se podría argumentar que ambas son Zapalote Grande realmente.

La pureza del agrupamiento fue de 0.964 y la entropía de 0.072, estos valores muestran un agrupamiento casi perfecto. En la Figura 4 se muestra la posición de estos valores en una distribución aleatoria de la pureza y la entropía calculadas a partir de 1000 agrupaciones hechas con 435 SNPs tomados al azar. En ambos casos el valor se encuentra fuera de la distribución aleatoria de las medidas, mostrando que el agrupamiento obtenido con los SNPs de  $F_{ST}$  alta es mejor que un agrupamiento realizado con SNPs aleatorios.

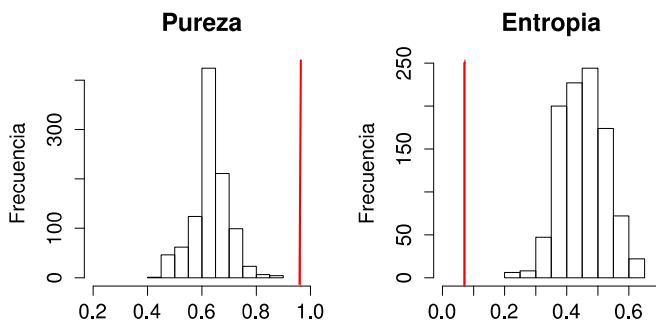
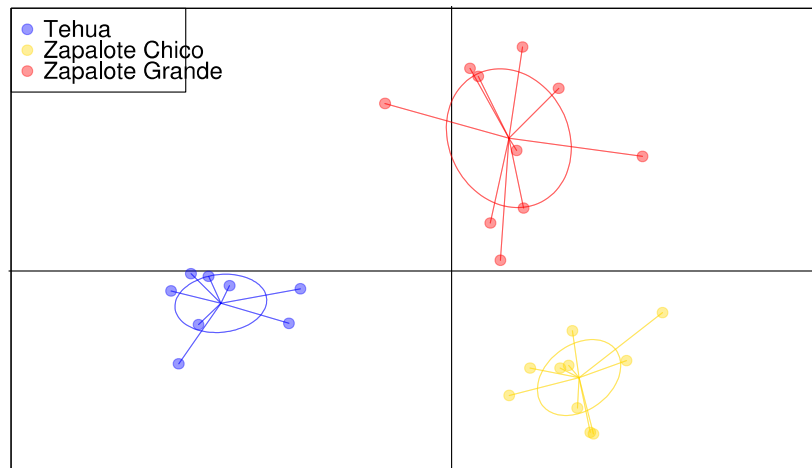


Figura 4: Histogramas de la pureza y la entropía. Las barras muestran la distribución de las medidas para agrupaciones hechas con grupos aleatorios de SNPs, las líneas rojas muestran el valor de las medidas para el agrupamiento con  $F_{ST}$  alta.

Tabla 3: Agrupamiento de las muestras, la muestra que no agrupó con el resto de las muestras de su raza (Zapalote Chico) se muestra sombreada.

Muestra	Grupo	Raza
13_1861	1	Tehua
20_13297	1	Tehua
6_777	1	Tehua
12_1849	1	Tehua
31_25784	1	Tehua
30_23861	1	Tehua
maiz_124	1	Tehua
maiz_125	1	Tehua
7_802	2	Zapalote Chico
19_10473	2	Zapalote Chico
15_2270	2	Zapalote Chico
22_23130	2	Zapalote Chico
23_23142	2	Zapalote Chico
18_10463	2	Zapalote Chico
26_23547	2	Zapalote Chico
28_23597	2	Zapalote Chico
34_29952	2	Zapalote Chico
29_25029	2	Zapalote Chico
21_16210	3	Zapalote Grande
4_137	3	Zapalote Grande
16_2601	3	Zapalote Grande
10_1222	3	Zapalote Grande
11_1773	3	Zapalote Grande
24_25020	3	Zapalote Grande
25_23217	3	Zapalote Grande
27_23552	3	Zapalote Grande
maiz_142	3	Zapalote Chico
maiz_144	3	Zapalote Grande

Se continuó el DAPC utilizando los grupos definidos por el agrupamiento. El número de componentes principales óptimo para generar el modelo fue dos componentes principales, los cuales generan un modelo formado por dos funciones discriminantes. Si se corren estas dos funciones con los datos de cada muestra, se obtienen dos valores (uno para cada función discriminante) para cada muestra y a partir de estos datos se genera una gráfica de DAPC (Figura 5). En la gráfica se observa como la primera función discriminante separa a las muestras de Tehua de las de Zapalote Chico y Zapalote Grande sobre el eje horizontal, y la segunda función separa a las muestras de Zapalote Chico de las de Zapalote Grande en el eje vertical.



*Figura 5: Gráfica de DAPC. La primera función discriminante se grafica sobre el eje horizontal y la segunda función sobre el vertical, cada punto corresponde a una muestra coloreada de acuerdo a su raza.*

A pesar de que el DAPC toma como variables componentes principales, se puede evaluar la contribución de cada SNP a las funciones discriminantes, esto quiere decir que se puede evaluar cuáles son los SNPs más útiles para separar entre los grupos. Debido a que este análisis discriminante está formado por dos funciones, se pueden obtener los marcadores más útiles para separar a las muestras en el eje horizontal y en el eje vertical. En la Figura 6 se muestra la distribución



de los SNPs de acuerdo a su contribución a las funciones tanto para el eje horizontal como para el vertical. En ambos casos, la distribución presenta un patrón exponencial con unos cuantos SNPs con contribuciones mucho más grandes que las del resto. Se tomaron los SNPs con mayor contribución haciendo un corte en distribución en donde a vista se observa que se separan los SNPs de mayor contribución. Esto dio como resultado la selección de 14 SNPs para la primera función discriminante (eje horizontal) y 7 SNPs para la segunda función (eje vertical) (Tabla 4).

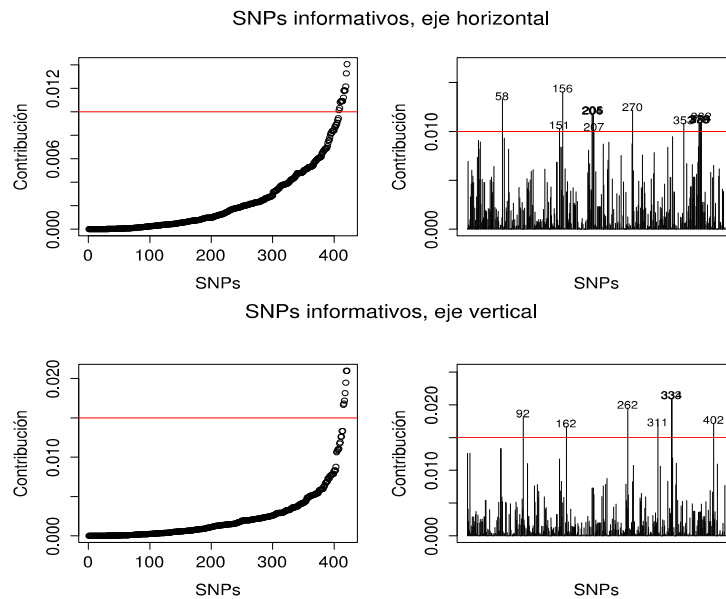


Figura 6: Contribución de los SNPs al DAPC. Las gráficas de la izquierda muestran los valores de contribución de los SNPs ordenados de manera creciente, la línea roja corresponde al corte a partir del cuál se tomaron los SNPs significativos. En las gráficas de la derecha cada barra corresponde a la contribución de un SNP, los números corresponden al índice del SNP significativo en la lista de SNPs con  $F_{ST}$  alta.

A partir de la información de los SNPs (Tabla 4), se puede ver que algunos corresponden al mismo gen y otros por su cercanía probablemente se encuentren ligados. Si se toman únicamente los SNPs que corresponden a genes diferentes, quedan 11 loci informativos para

la primera función discriminante y seis loci para la segunda función. Se encontraron 9 genes asociados a 13 de los 21 SNPs identificados, los otros 8 SNPs se encuentran en regiones del genoma en las que no se ha descrito un gen. De los 9 genes asociados a los SNPs, sólo 3 se encuentran anotados y corresponden a: una chaperonina de 10 kDa (cpn10), al factor de transcripción de choque de calor 30 (*Heat shock factor proteín 30*) y a un factor de transcripción (*Transcription factor IIIA*).

*Tabla 4: SNPs con contribuciones altas para las dos funciones discriminantes del DAPC. Los grupos de SNPs que corresponden al mismo gen se muestran sombreados.*

SNP	Eje	Cromo-Soma	Posición	Gen asociado	Anotación
SYN13273	X	1	63,211,705	GRMZM2G073401	Chaperonina
PZE-103162678	X	3	213,304,333		
SYN35555	X	4	193,404,738	GRMZM2G058120	No caracterizado
PZE-104117392	X	4	193,784,241		
PZE-104117394	X	4	193,787,328	GRMZM5G866082	No caracterizado
PZE-104117395	X	4	193,787,435	GRMZM5G866082	No caracterizado
PZE-104117476	X	4	193,936,668		
PZE-106027700	X	6	66,225,297		
PZE-109030021	X	9	34,676,000	AC216886.3_FG003	No caracterizado
PZE-109039020	X	9	59,462,222	GRMZM2G173090	Heat shock factor Protein HSF30
PZE-109039021	X	9	59,462,243	GRMZM2G173090	Heat shock factor Protein HSF30
PZE-109039022	X	9	59,462,302	GRMZM2G173090	Heat shock factor Protein HSF30
PZE-109039048	X	9	59,507,310		
PZE-109039430	X	9	60,412,695	GRMZM2G476230	No caracterizado
PZE-101239396	Y	1	286,726,652	GRMZM2G061084	No caracterizado
PZE-104021070	Y	4	22,553,514		
PZE-106012648	Y	6	32,203,571		
PZE-108004274	Y	8	4,285,891	GRMZM2G171122	No caracterizado
PZE-108072700	Y	8	125,607,200	GRMZM2G380515	Transcription Factor IIIA
PZE-108072703	Y	8	125,610,778	GRMZM2G380515	Transcription Factor IIIA
PZE-109094812	Y	9	141,330,945		

Teniendo los SNPs informativos para las dos funciones discriminantes, se extrajeron los estados de estos en cada una de las muestras para poder observar las diferencias alélicas en estos

marcadores para las tres razas (Figura 7). Se puede observar una distinción clara entre Zapalote Grande y Zapalote Chico con respecto a Tehua en el bloque de los marcadores discriminantes de la primera función. En el segundo bloque de marcadores, se ve una distinción entre Zapalote Chico y Tehua con respecto a Zapalote Grande, sin embargo esta distinción no es tan clara como la de los marcadores de la primera función.

Muestra	SNPs eje horizontal										SNPs eje vertical					Raza						
	cpn10	HSF30					TF IIIA															
	SNV13273	PZE-103162678	SVN35555	PZE-104117392	PZE-104117394	PZE-104117395	PZE-104117476	PZE-106027700	PZE-109030021	PZE-109039020	PZE-109039021	PZE-109039022	PZE-109039048	PZE-109039430	PZE-101239396	PZE-104021070	PZE-106012648	PZE-108004274	PZE-108072700	PZE-108072703	PZE-109094812	
13_1861																						Tehua
20_13297																						Tehua
6_777																						Tehua
12_1849																						Tehua
31_25784																						Tehua
30_23861																						Tehua
maiz_124																						Tehua
maiz_125																						Tehua
7_802																						Zapalote Chico
19_10473																						Zapalote Chico
15_2270																						Zapalote Chico
22_23130																						Zapalote Chico
23_23142																						Zapalote Chico
18_10463																						Zapalote Chico
26_23547																						Zapalote Chico
28_23597																						Zapalote Chico
34_29952																						Zapalote Chico
29_25029																						Zapalote Chico
maiz_142																						Zapalote Chico
21_16210																						Zapalote Grande
4_137																						Zapalote Grande
16_2601																						Zapalote Grande
10_1222																						Zapalote Grande
11_1773																						Zapalote Grande
24_25020																						Zapalote Grande
25_23217																						Zapalote Grande
27_23552																						Zapalote Grande
maiz_144																						Zapalote Grande
		Homocigo alelo mayor					Heterocigo					Homocigo alelo menor										

Figura 7: Estado alélico de los SNPs informativos para las muestras de las tres razas. El color de la celda indica el estado alélico del SNP para cada muestra, las celdas en blanco son SNPs para los que no hay información, los SNPs encerrados en un recuadro punteado están asociados al mismo gen.

## Clasificación de las muestras por su altitud

El análisis discriminante demuestra que las muestras de Tehua, Zapalote Grande y Zapalote Chico pueden ser asignadas a su raza por la combinación ciertos marcadores moleculares. Sin embargo, considerando que se ha encontrado que la altitud es el factor que mejor explica la variación genética en maíz y estas razas tienen una distribución en altitudes contrastantes (Zapalote Chico en zonas bajas, Zapalote Grande bajas a intermedias y Tehua en zonas altas), las diferencias alélicas encontradas podrían ser resultado de adaptación ambiental a las diferentes altitudes o de la selección artificial de caracteres morfológicos por parte de las personas que la cultivan.

Las muestras fueron clasificadas por su altitud. El límite entre las categorías altitudinales “Alto” y “Bajo” definido fue 750 metros sobre el nivel del mar debido a que esta altura separa la parte principal de la distribución de alturas de Zapalote Chico y Zapalote Grande de Tehua (Figura 8). Las medianas de la altitud de muestras de los grupos “Bajo” y “Alto” fueron 122 y 1535, respectivamente; la distribución de altitudes difiere significativamente (Mann-Whitney  $U= 2449$ ,  $p=2.2 \times 10^{-16}$ ).

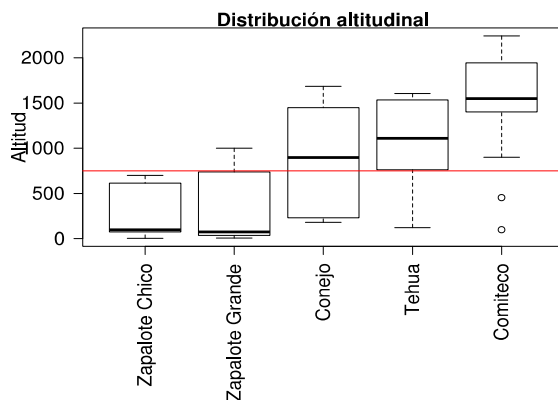


Tabla 5: Número de muestras de cada raza por categoría altitudinal

Raza	Bajo	Alto
Comiteco	2	12
Conejo	4	4
Tehua	2	6
Zapalote Chico	11	0
Zapalote Grande	6	2

Figura 8: Distribución altitudinal de las razas. Gráfica de caja con la distribución altitudinal de las muestras de cada raza. La línea horizontal roja corresponde a los 750 m.s.n.m, el límite entre las categorías Alta y Baja.

La clasificación altitudinal de las muestras de las cinco razas se muestra en la Tabla 5. Sólo una muestra de Zapalote Grande no pudo ser asignada a una categoría debido a que no se contaba con información de la localidad en la que fue colectada. Zapalote Chico es la única raza en la que todas las muestras pertenecen a la misma categoría (Bajo).

Debido a que en las razas Zapalote Grande y Tehua hay muestras de las dos categorías altitudinales, un primer acercamiento para ver si los marcadores informativos para las funciones discriminantes definidas anteriormente están asociados a adaptación altitudinal o a otras características morfológicas que diferencian a las razas, fue repetir la gráfica resultante del DAPC de las tres razas (Zapalote Chico, Zapalote Grande y Tehua), agregando el factor altitudinal de cada muestra (Figura 9). En la gráfica se puede ver que, aunque la primera función (eje horizontal) separa a las razas de diferentes alturas, las muestras de estas razas pertenecientes a diferentes categorías altitudinales se

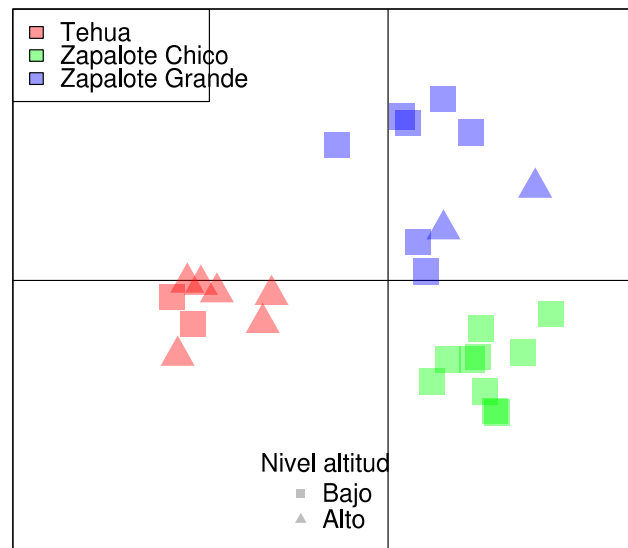


Figura 9: Gráfica de DAPC con categorías altitudinales. Las muestras tienen color de acuerdo a su raza y forma de acuerdo a su categoría altitudinal (cuadrado altas y triángulo bajas).

agrupan con las muestras de su misma raza y no con las de su misma categoría altitudinal.

El cálculo de  $F_{ST}$  de los SNPs por categorías altitudinales presentó una distribución con un intervalo de 0 a 0.5251, una media de 0.0345 y desviación estándar de 0.043. El grupo de SNPs de altitud, formado por el 1% de los SNPs con mayor  $F_{ST}$  para categorías altitudinales, estuvo integrado por 423 SNPs con valores de  $F_{ST}$  entre 0.1728 y 0.5251. Se comparó este grupo de SNPs de altitud con los SNPs con  $F_{ST}$  alta entre razas, estos dos grupos tienen 98 SNPs en común, esto es el 23.16% para el grupo de SNPs de altitud y 22.52% para el grupo de SNPs con  $F_{ST}$  alta entre razas.

Usando el grupo de SNPs de altitud, se realizó el análisis discriminante de componentes principales (DAPC) igual al que se hizo con los SNPs con  $F_{ST}$  alta entre las razas. El agrupamiento por modelo dio como resultado que el mejor agrupamiento era formar dos grupos ( $k=2$ ). Los grupos formados por el modelo corresponden casi perfectamente con los grupos definidos “Bajo” y “Alto”. Sólo hay una muestra que no coincide, la muestra 31\_25784 de raza Tehua, perteneciente al grupo Bajo y agrupada con todas las muestras de Alto.

A partir de este agrupamiento, se calculó el DAPC usando un sólo componente principal. El resultado fue un modelo de una sola función discriminante que calcula la probabilidad de cada muestra de pertenecer a una de las dos categorías altitudinales (Figura 10, izquierda). Usando esta función se asignó cada muestra a una categoría altitudinal, todas las muestras excepto una (31\_25784, raza Tehua) fueron asignadas a su categoría altitudinal original (Figura 10, derecha).

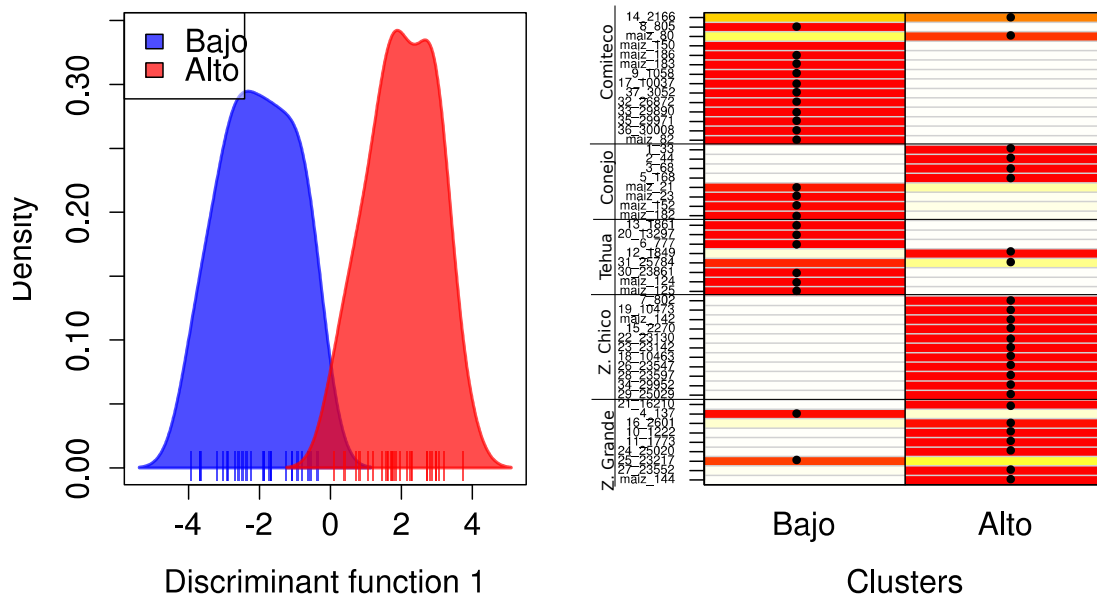


Figura 10: DAPC para grupos altitudinales. Izquierda. Función discriminante, se grafica la densidad de muestras de acuerdo a la probabilidad que tienen de pertenecer a cada categoría de acuerdo al modelo. Derecha. Asignación de cada muestra a un grupo altitudinal, el tono corresponde a la probabilidad de cada muestra a pertenecer a una categoría (rojo probabilidad alta; blanco baja probabilidad). El punto negro indica el grupo al que pertenece cada muestra, cuando coincide con un fondo rojo indica una buena asignación en la clasificación.

Siguiendo con el análisis, se obtuvo la contribución de cada SNP de altitud a la función discriminante. Observando la distribución de las contribuciones, se separó un grupo de los 14 SNPs que más contribuyen a la clasificación por altitud (Figura 11). La ubicación de estos 14 SNPs y el gen al que están asociados se encuentra en la Tabla 6. De estos 14 SNPs, ocho SNPs corresponden a seis genes y los seis SNPs restantes se encuentran en regiones en las que no se han reportado genes, algunos por su cercanía pueden estar ligados entre sí. Sólo tres de los seis genes asociados a los SNPs informativos tienen una proteína anotada, estas son: una proteína de respuesta a auxinas (SAUR40), una proteína transformante relacionada con la edición de RNA mensajero

(Transformer-2) y al factor de transcripción de choque de calor 30 (Heat shock factor protein 30).

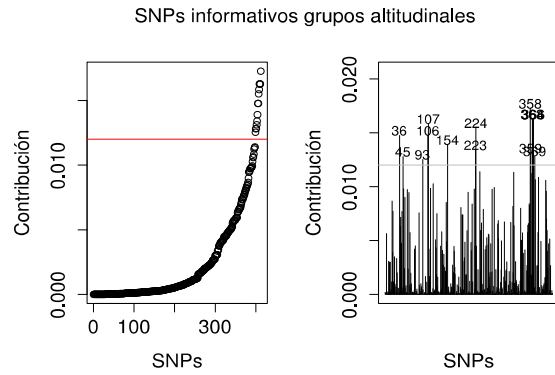


Figura 11: Contribución de los SNPs al DAPC de categorías de altitud. Izquierda. Valores de contribución de los SNPs ordenados de manera creciente, la línea roja corresponde al corte a partir del cuál se tomaron los SNPs significativos. Derecha. Los números corresponden al índice del SNP significativo en la lista de SNPs de altitud.

Tabla 6: SNPs con contribuciones altas para distinguir entre grupos altitudinales. Los grupos de SNPs que corresponden a la misma región del genoma se muestran sombreados.

SNP	Cromo-Soma	Coordenada	Gen	Anotación
PZE-102150323	2	197,663,601	GRMZM2G042741	SAUR40
SYN25763	2	197,663,065	GRMZM2G042741	Auxin-responsive SAUR40
PZE-102099063	2	116,694,897	GRMZM2G071744	No caracterizado
PZE-102099083	2	116,801,868		
PZE-102099223	2	117,348,457		
PZE-104009802	4	6,691,056	GRMZM2G334631	No caracterizado
PZE-105062945	5	62,229,361	GRMZM2G152111	Transformer-2 protein
PZE-105062976	5	62,231,214		
PZE-109036755	9	50,385,286		
PZE-109037166	9	51,669,851		
SYN15460	9	52,052,243	GRMZM2G100103	No caracterizado
SYN15462	9	52,052,297	GRMZM2G100103	No caracterizado
PZE-109037353	9	52,059,708		
PZE-109039002	9	59,461,144	GRMZM2G173090	Heat shock factor protein HSF30



Comparando este grupo de SNPs con los 21 SNPs informativos del análisis anterior entre tres razas, el resultado es que no hay SNPs compartidos entre los dos grupos. Sin embargo, hay un gen presente en ambos grupos de SNPs informativos, este es el gen GRMZM2G173090 correspondiente a al factor de transcripción de choque de calor 30 (HSF30). Este gen se encuentra asociado a tres SNPs en en análisis de las tres razas y a un SNP diferente en el análisis de altitud. Este análisis demuestra que, con la excepción de un gen, los marcadores que sirven para separar las muestras por categorías altitudinales son independientes de los SNPs que sirven para separar a las razas Tehua, Zapalote Chico y Zapalote Grande. El estado de los SNPs informativos para grupos altitudinales en cada muestra se puede ver en la Figura 12.

Muestra	SAUR40	TP	HSF	Altitud	Nivel	Raza
8_805	PZE-102099083	PZE-105062945	PZE-109037353_A	2241	Alto	Comiteco
37_3052	PZE-102150323	PZE-105062945	PZE-109037353_A	2008	Alto	Comiteco
35_29971	SYN125763	PZE-105062945	PZE-109037353_A	1955	Alto	Comiteco
17_10037	PZE-102099063	PZE-105062945	PZE-109037353_A	1943	Alto	Comiteco
maiz_23	PZE-102099223	PZE-105062945	PZE-109037353_A	1685	Alto	Conejo
36_30008	PZE-104009802	PZE-105062945	PZE-109037353_A	1631	Alto	Comiteco
30_23861	PZE-105062976	PZE-105062945	PZE-109037353_A	1604	Alto	Tehua
maiz_82	PZE-109036755	PZE-105062945	PZE-109037353_A	1553	Alto	Comiteco
maiz_150	PZE-109037166	PZE-105062945	PZE-109037353_A	1550	Alto	Comiteco
maiz_186	SYN15460	PZE-105062945	PZE-109037353_A	1550	Alto	Comiteco
maiz_183	SYN15462	PZE-105062945	PZE-109037353_A	1550	Alto	Comiteco
maiz_124	SYN15460	PZE-105062945	PZE-109037353_A	1545	Alto	Tehua
maiz_125	SYN15462	PZE-105062945	PZE-109037353_A	1525	Alto	Tehua
33_29890	SYN15460	PZE-105062945	PZE-109037353_A	1510	Alto	Comiteco
maiz_152	SYN15462	PZE-105062945	PZE-109037353_A	1493	Alto	Conejo
maiz_182	SYN15460	PZE-105062945	PZE-109037353_A	1402	Alto	Conejo
9_1058	SYN15462	PZE-105062945	PZE-109037353_A	1400	Alto	Comiteco
6_777	SYN15460	PZE-105062945	PZE-109037353_A	1400	Alto	Tehua
maiz_21	SYN15462	PZE-105062945	PZE-109037353_A	1312	Alto	Conejo
4_137	SYN15460	PZE-105062945	PZE-109037353_A	1000	Alto	Z. Grande
32_26872	SYN15462	PZE-105062945	PZE-109037353_A	900	Alto	Comiteco
20_13297	SYN15460	PZE-105062945	PZE-109037353_A	820	Alto	Tehua
25_23217	SYN15462	PZE-105062945	PZE-109037353_A	820	Alto	Z. Grande
13_1861	SYN15460	PZE-105062945	PZE-109037353_A	792	Alto	Tehua
31_25784	SYN15462	PZE-105062945	PZE-109037353_A	730	Alto	Tehua
18_10463	SYN15460	PZE-105062945	PZE-109037353_A	700	Bajo	Z. Chico
22_23130	SYN15462	PZE-105062945	PZE-109037353_A	681	Bajo	Z. Chico
maiz_142	SYN15460	PZE-105062945	PZE-109037353_A	654	Bajo	Z. Chico
maiz_144	SYN15462	PZE-105062945	PZE-109037353_A	654	Bajo	Z. Grande
34_29952	SYN15460	PZE-105062945	PZE-109037353_A	573	Bajo	Z. Chico
2_44	SYN15462	PZE-105062945	PZE-109037353_A	480	Bajo	Conejo
maiz_80	SYN15460	PZE-105062945	PZE-109037353_A	455	Bajo	Comiteco
23_23142	SYN15462	PZE-105062945	PZE-109037353_A	255	Bajo	Z. Chico
5_168	SYN15460	PZE-105062945	PZE-109037353_A	250	Bajo	Conejo
1_33	SYN15462	PZE-105062945	PZE-109037353_A	210	Bajo	Conejo
3_68	SYN15460	PZE-105062945	PZE-109037353_A	180	Bajo	Conejo
12_1849	SYN15462	PZE-105062945	PZE-109037353_A	122	Bajo	Tehua
14_2166	SYN15460	PZE-105062945	PZE-109037353_A	100	Bajo	Comiteco
7_802	SYN15462	PZE-105062945	PZE-109037353_A	100	Bajo	Z. Chico
19_10473	SYN15460	PZE-105062945	PZE-109037353_A	100	Bajo	Z. Chico
15_2270	SYN15462	PZE-105062945	PZE-109037353_A	100	Bajo	Z. Chico
11_1773	SYN15460	PZE-105062945	PZE-109037353_A	100	Bajo	Z. Grande
29_25029	SYN15462	PZE-105062945	PZE-109037353_A	50	Bajo	Z. Chico
24_25020	SYN15460	PZE-105062945	PZE-109037353_A	50	Bajo	Z. Grande
21_16210	SYN15462	PZE-105062945	PZE-109037353_A	41	Bajo	Z. Grande
16_2601	SYN15460	PZE-105062945	PZE-109037353_A	31	Bajo	Z. Grande
26_23547	SYN15462	PZE-105062945	PZE-109037353_A	8	Bajo	Z. Chico
27_23552	SYN15460	PZE-105062945	PZE-109037353_A	8	Bajo	Z. Grande
28_23597	SYN15462	PZE-105062945	PZE-109037353_A	5	Bajo	Z. Chico
10_1222	SYN15460	PZE-105062945	PZE-109037353_A			Z. Grande

Figura 12: Estado alélico para los SNPs informativos para categorías altitudinales en todas las muestras ordenadas por altitud. El color de la celda indica el estado alélico, las celdas en blanco son SNPs para los que no hay información, los SNPs encerrados en un recuadro punteado están asociados al mismo gen.

## Análisis discriminante de las cinco razas

Por último, se realizó el análisis con las muestras de las cinco razas. El resultado del agrupamiento por modelo usando los diferentes grupos de SNPs fue que tanto para el grupo de todos los SNPs, como para el grupo de SNPs de domesticación el modelo no era capaz de agrupar las muestras y el mejor agrupamiento era un sólo grupo ( $K=1$ ). En cambio, para el grupo de SNPs con  $F_{ST}$  alta el mejor agrupamiento es usando tres grupos. El análisis continuó usando únicamente los SNPs con  $F_{ST}$  alta, ya que fueron los únicos que sirvieron para agrupar las muestras.

Los grupos inferidos por el modelo son: un grupo con muestras de la raza Zapalote Chico, el segundo grupo contiene muestras de las razas Conejo y Zapalote Grande, y el tercer grupo muestras de las razas Tehua y Comiteco (Tabla 7). Sólo hay dos muestras que no se agrupan con el resto de las muestras de su raza, estas son la muestra 25\_23217 (Zapalote Grande en el grupo 1) y la muestra maiz\_142 (Zapalote Chico en el grupo 2). La segunda muestra

Tabla 7: Agrupamiento de las muestras usando los SNPs con  $F_{ST}$  alta.

Muestra	Grupo	Raza	Nivel Altitud
7_802		1 Zapalote Chico	bajo
19_10473		1 Zapalote Chico	bajo
15_2270		1 Zapalote Chico	bajo
22_23130		1 Zapalote Chico	bajo
23_23142		1 Zapalote Chico	bajo
18_10463		1 Zapalote Chico	bajo
25_23217		1 Zapalote Grande	alto
26_23547		1 Zapalote Chico	bajo
28_23597		1 Zapalote Chico	bajo
34_29952		1 Zapalote Chico	bajo
29_25029		1 Zapalote Chico	bajo
1_33		2 Conejo	bajo
2_44		2 Conejo	bajo
3_68		2 Conejo	bajo
21_16210		2 Zapalote Grande	bajo
4_137		2 Zapalote Grande	alto
16_2601		2 Zapalote Grande	bajo
10_1222		2 Zapalote Grande	NA
5_168		2 Conejo	bajo
11_1773		2 Zapalote Grande	bajo
24_25020		2 Zapalote Grande	bajo
27_23552		2 Zapalote Grande	bajo
maiz_21		2 Conejo	alto
maiz_23		2 Conejo	alto
maiz_142		2 Zapalote Chico	bajo
maiz_144		2 Zapalote Grande	bajo
maiz_152		2 Conejo	alto
maiz_182		2 Conejo	alto
13_1861		3 Tehua	alto
14_2166		3 Comiteco	bajo
8_805		3 Comiteco	alto
20_13297		3 Tehua	alto
9_1058		3 Comiteco	alto
17_10037		3 Comiteco	alto
6_777		3 Tehua	alto
12_1849		3 Tehua	bajo
37_3052		3 Comiteco	alto
31_25784		3 Tehua	bajo
32_26872		3 Comiteco	alto
33_29890		3 Comiteco	alto
35_29971		3 Comiteco	alto
30_23861		3 Tehua	alto
36_30008		3 Comiteco	alto
maiz_82		3 Comiteco	alto
maiz_80		3 Comiteco	bajo
maiz_124		3 Tehua	alto
maiz_125		3 Tehua	alto
maiz_150		3 Comiteco	alto
maiz_186		3 Comiteco	alto
maiz_183		3 Comiteco	alto

(maiz\_142) es la misma que en el análisis con tres razas se agrupaba con las muestras de Zapalote Grande.

El DAPC se realizó agrupando las muestras por raza y no por los grupos resultantes del agrupamiento por modelo. Para el análisis se utilizaron los 4 primeros componentes principales y dos funciones discriminantes. En la gráfica del DAPC (Figura 13) se ve que el análisis logra separar a las muestras de acuerdo a su raza. El eje horizontal de la gráfica, correspondiente a la primera función discriminante, separa a las razas siguiendo un patrón altitudinal, yendo de izquierda a derecha de la raza más alta (Comiteco) hasta la más baja (Zapalote Chico) con las otras tres razas ordenadas por su altura. La segunda función discriminante (eje vertical) sirve para separar las razas de niveles altitudinales similares (Comiteco y Tehua, Zapalote Chico y Zapalote Grande).

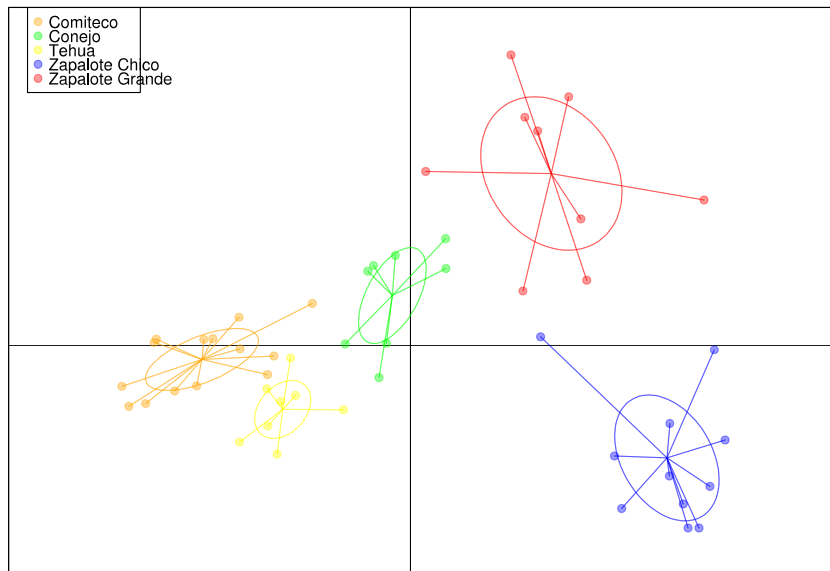


Figura 13: DAPC con muestras de las 5 razas, el color de cada muestra corresponde a su raza.

Se calculó la probabilidad de pertenencia de cada muestra a las diferentes razas. En la Figura 14 se presentan estas probabilidades en

una gráfica de barras en la que cada color es la probabilidad de pertenencia a una raza. Ahí se puede ver que, en términos generales, usando el grupo de SNPs con  $F_{ST}$  alta las razas están bien diferenciadas y que algunas de las muestras presentan bastante mezcla entre razas (por ejemplo, maiz\_150).

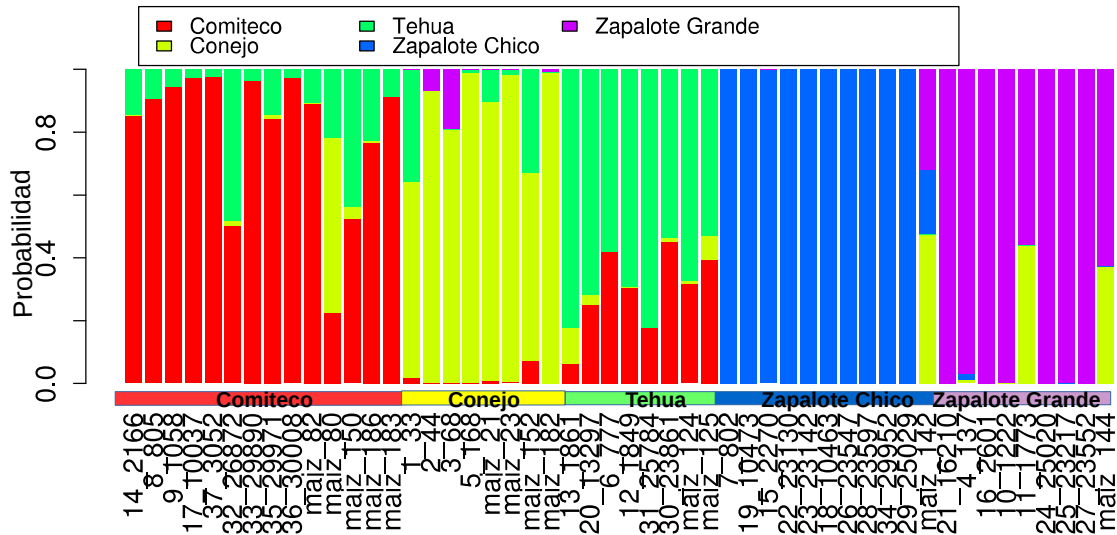


Figura 14: Probabilidad de pertenencia a razas. Cada barra corresponde a una muestra y el color de la barra es la proporción de probabilidad de pertenencia a las diferentes razas.

El nivel de mezcla es especialmente alto entre muestras de Comiteco y Tehua, razas que formaban un mismo grupo en el agrupamiento por modelo y para las cuales la distinción como dos razas diferentes no está bien sustentada de acuerdo a estos datos genómicos. Las muestras de Zapalote Chico están claramente diferenciadas del resto, la única excepción es la muestra maiz\_142, la misma que en el análisis de las tres razas se agrupó con Zapalote Grande y que probablemente corresponda a una asignación errónea en campo. Las muestras de Zapalote Grande muestran bastante homogeneidad, excepto dos que contienen mezcla con Conejo. Es interesante resaltar que las muestras de Zapalote Grande no muestran parentesco con Zapalote Chico ni Tehua, razas propuestas por Wellhausen *et al.* como

sus progenitoras. Finalmente, las muestras de Conejo forman un grupo bien definido, aunque todas las muestras presentan cierto grado de mezcla con las razas Comiteco, Tehua o Zapalote Grande. En la Figura 15 se muestra esta misma información como mapa de calor. Ahí se puede ver que para la mayoría de las muestras la asignación de raza por el DAPC corresponde a su raza original.

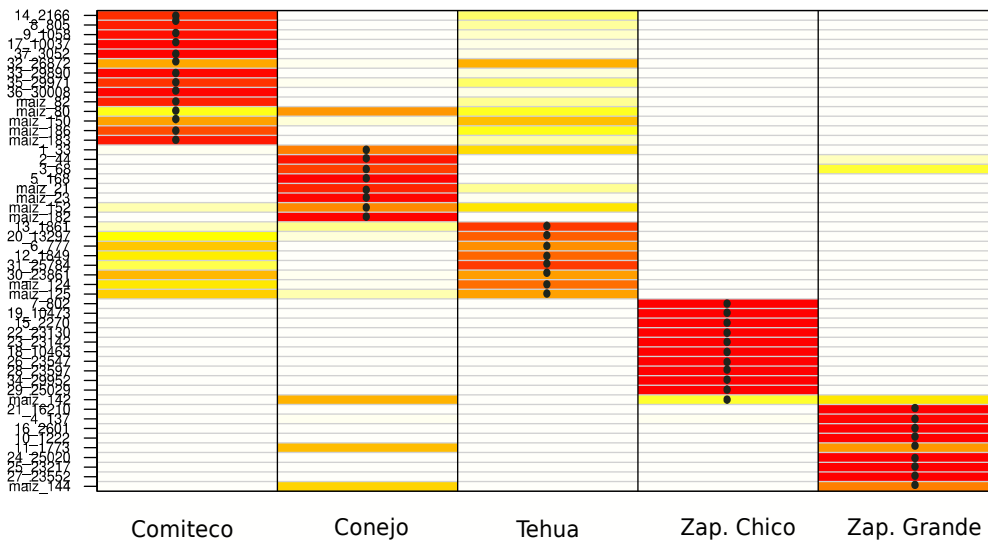


Figura 15: Mapa de calor de la probabilidad de pertenencia de cada muestra a razas. El punto en las celdas indica la raza original de la muestra, el color de la celda muestra la probabilidad de pertenencia a una raza (blanco=0, rojo=1). El punto negro indica el grupo al que pertenece cada muestra, cuando coincide con un fondo rojo indica una buena asignación en la clasificación.

Para este DAPC también se obtuvieron los SNPs que más contribuyen a las funciones discriminantes. Al igual que para el análisis con tres razas, se obtuvieron los SNPs informativos para la función del eje horizontal y para la del eje vertical (Figura 16); sin embargo, en este caso al ser más razas la interpretación de diferenciación de estos SNPs es menos directa. El resultado obtenido son 11 SNPs informativos para el eje horizontal y nueve SNPs informativos para el eje vertical (Tabla 8). Nueve de los 20 SNPs identificados están asociados a 7 genes, de los

cuales cuatro no han sido caracterizados y los tres restantes están anotados como una enzima poligalacturonasa, el factor de transcripción de choque de calor 30 (HSF30) y el factor de transcripción IIIA (Transcription factor IIIA). El factor de transcripción de choque de calor 30 es el mismo que había resultado informativo en los análisis de tres razas (Zapalote Chico, Zapalote Grande y Tehua) y de categorías altitudinales y el factor de transcripción IIIA es el mismo que se había obtenido en el análisis con tres razas.

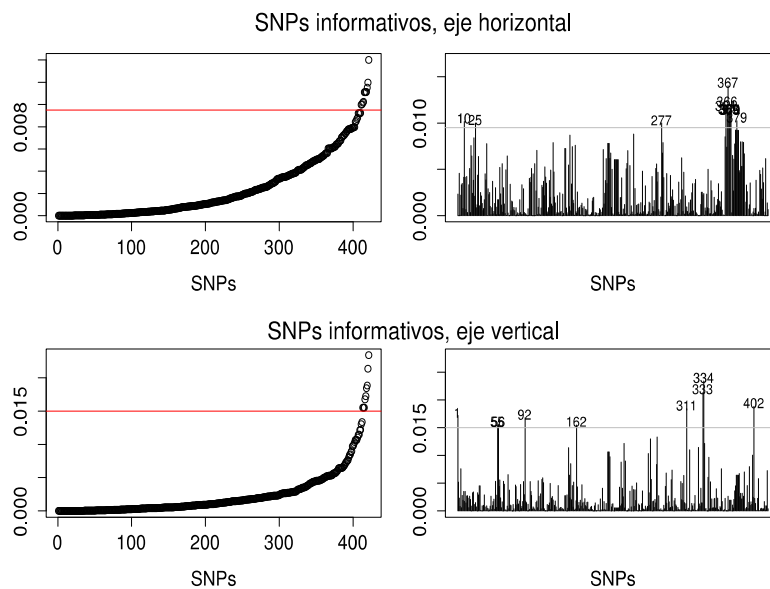


Figura 16: Contribución de los SNPs al DAPC con cinco razas.

Comparando este grupo de SNPs con los obtenidos en los dos análisis anteriores, en los SNPs informativos del eje horizontal hay cinco SNPs (PZE-109036755, PZE-109037166, SYN15460, SYN15462, PZE-109037353) que también se encuentran en el grupo de SNPs informativos para el análisis de categorías altitudinales y un SNP (PZE-109039021) que se encuentra entre los SNPs informativos para la función del eje horizontal del análisis de tres razas (Tabla 8). De los SNPs informativos del eje vertical, seis de los nueve SNPs (PZE-101239396, PZE-104021070, PZE-108004274, PZE-108072700, PZE-108072703,

PZE-109094812) se encuentran también entre los SNPs informativos de la función vertical del análisis con tres razas. Esto indica que para el eje horizontal 5 de 11 SNPs que diferencian entre estas cinco razas explican un factor altitudinal de la variación y un SNP de estos once está compartido con los SNPs informativos para el mismo eje en el análisis con tres razas. El SNP compartido con el análisis de las tres razas corresponde al factor de transcripción de choque térmico 30 que también es informativa para el análisis altitudinal, por lo tanto este SNP también explica un factor altitudinal de la variación. Para el eje horizontal seis de nueve SNPs se encuentran también entre los SNPs informativos para el mismo eje en el análisis usando sólo tres razas. En la Figura 17 se muestra el estado de estos SNPs en las muestras.

*Tabla 8: Información de los SNPs informativos para las funciones del DAPC usando las cinco razas. Los SNPs sombreados son SNPs que pertenecen al mismo gen. Asteriscos junto al nombre de los SNPs indican si el SNP se encuentra en otro de los grupos de SNPs informativos.*

SNP	Eje	Cromo- Soma	Coordenada	Gen	Anotación
PZB01114.1	x	9	64,695,530		
PZE-101080378	x	1	65,092,399	GRMZM2G042683	No caracterizado
PZE-106049782	x	6	99,291,757	GRMZM2G153666	Polygalacturonasa
PZE-109036638	x	9	50,198,296		
PZE-109036749	x	9	50,383,446		
PZE-109036755*	x	9	50,385,286		
PZE-109037166*	x	9	51,669,851		
SYN15460*	x	9	52,052,243	GRMZM2G100103	No caracterizado
SYN15462*	x	9	52,052,297	GRMZM2G100103	No caracterizado
PZE-109037353*	x	9	52,059,708		
PZE-109039021**	x	9	59,462,243	GRMZM2G173090	Heat shock factor Protein HSF30
PHM10525.11	y	8	124,752,481		
PZE0006576497	y	2	50,310,642		
PZE-100002058	y	2	49,681,375		
PZE-101239396**	y	1	286,726,652	GRMZM2G061084	No caracterizado
PZE-104021070**	y	4	22,553,514		
PZE-108004274**	y	8	4,285,891	GRMZM2G171122	No caracterizado
PZE-108072700**	y	8	125,607,200	GRMZM2G380515	Transcription Factor IIIA
PZE-108072703**	y	8	125,610,778	GRMZM2G380515	Transcription Factor IIIA
PZE-109094812**	y	9	141,330,945		

\* SNPs compartidos con análisis de categorías altitudinales  
\*\* SNPs compartidos con análisis de tres razas



Muestra	SNPs										Raza
	TF IIIA					TF IIIA					
14_2166											Comiteco
8_805											Comiteco
9_1058											Comiteco
17_10037											Comiteco
37_3052											Comiteco
32_26872											Comiteco
33_29890											Comiteco
35_29971											Comiteco
36_30008											Comiteco
maiz_82											Comiteco
maiz_80											Comiteco
maiz_150											Comiteco
maiz_186											Comiteco
maiz_183											Comiteco
13_1861											Tehua
20_13297											Tehua
6_777											Tehua
12_1849											Tehua
31_25784											Tehua
30_23861											Tehua
maiz_124											Tehua
maiz_125											Tehua
1_33											Conejo
2_44											Conejo
3_68											Conejo
5_168											Conejo
maiz_21											Conejo
maiz_23											Conejo
maiz_152											Conejo
maiz_182											Conejo
21_16210											Z. Grande
4_137											Z. Grande
16_2601											Z. Grande
10_1222											Z. Grande
11_1773											Z. Grande
24_25020											Z. Grande
25_23217											Z. Grande
27_23552											Z. Grande
maiz_144											Z. Grande
7_802											Z. Chico
19_10473											Z. Chico
15_2270											Z. Chico
22_23130											Z. Chico
23_23142											Z. Chico
18_10463											Z. Chico
26_23547											Z. Chico
28_23597											Z. Chico
34_29952											Z. Chico
29_25029											Z. Chico
maiz_142											Z. Chico

* SNPs compartidos con análisis altitudinal	Homócigo alelo mayor
** SNPs compartidos con análisis de 3 razas	Heterócigo
	Homócigo alelo menor

Figura 17: Estado alélico de los SNPs informativos para el DAPC de todas las razas. El color de la celda indica el estado alélico, las celdas en blanco son SNPs para los que no hay información, los SNPs encerrados en un recuadro punteado están asociados al mismo gen, los SNPs con uno o dos asteriscos junto a su nombre son SNPs compartidos en los otros dos análisis de DAPC.

## ANÁLISIS Y DISCUSIÓN

### Agrupamiento por grupos de marcadores candidatos

El análisis de componentes principales y el análisis de agrupamiento muestran que el agrupamiento de las muestras depende del grupo de SNPs que se utilice. En particular, el grupo de SNPs con  $F_{ST}$  alta sí muestra distinciones entre razas, mientras que el grupo de SNPs de domesticación y mejoramiento no.

Los SNPs de domesticación han sido encontrados por diferencias entre la especie domesticada y su pariente silvestre, es congruente que estos SNPs formen un solo grupo entre las razas de maíz. Los SNPs de mejoramiento fueron obtenidos por diferencias entre razas de maíz y variedades mejoradas (Hufford *et al.* 2012). El desarrollo de líneas mejoradas de maíz se lleva a cabo a partir de un proceso de selección de características de interés agronómicas que reduce la variación genética considerablemente. En ese proceso se dirige la selección buscando desarrollar líneas uniformes que puedan ser cultivadas en diferentes ambientes con la finalidad de tener un aumento en el rendimiento. En cambio, las razas nativas están adaptadas a ambientes locales muy particulares y a usos diversos por lo que los patrones de selección a los que están sujetos son muy diversos y dependen de la comunidad que los siembra (Jardón Barbolla y Benítez Keinrad, 2015). Por ende, es muy probable que existan muchos genes que diferencien a las razas que no han sido utilizados en los programas de mejoramiento. En específico, los marcadores que distinguen entre estas cinco razas no están entre los genes de domesticación y mejoramiento de maíz reportados por Meyer *et al.* y Hufford *et al.*

El hecho de que al utilizar todos los SNPs el agrupamiento por modelo no sea capaz de distinguir entre los grupos de razas y el mejor agrupamiento sea considerar todas las muestras como un sólo grupo demuestra que tomando el genoma entero, la distribución de la

variación no sigue una estructura en categorías raciales. Esto es congruente con análisis de más razas que muestran que la estructura genética corresponde a grupos altitudinales y distancias geográficas, más que a la asignación de raza (Arteaga *et al.*, 2016). Sin embargo, tomando el grupo de SNPs con  $F_{ST}$  alta el modelo es capaz de recuperar los grupos raciales estudiados aquí. Por lo tanto, es posible sugerir que los SNPs que sirven para distinguir entre las razas están sujetos a diferentes presiones selectivas en estas razas lo cuál hace que varíen de forma diferente al resto del genoma. Estas presiones selectivas son, posiblemente, aquellas guiadas por las y los campesinos al sembrar y seleccionar la semilla con las características que buscan año con año. Este manejo permite que se mantenga una diferenciación entre las razas de maíz a pesar del intenso flujo génico. La estabilidad que se observa en las razas de maíz es entonces resultado de la interacción entre procesos contradictorios, los procesos de flujo génico que tienden a homogeneizar la variación en el maíz y la selección continua de los campesinos que mantienen las razas diferenciadas, adaptadas a los sitios donde son sembradas y con características fenotípicas particulares que permiten distinguirlas.

### **Identificación de marcadores útiles para distinguir entre razas**

El análisis discriminante de componentes principales fue útil para generar un modelo que distingue entre razas y entre categorías altitudinales, así como para encontrar SNPs informativos en estos modelos. A pesar de que hubieron coincidencias en los SNPs informativos de los tres análisis, los marcadores que distinguen entre tres razas, cinco razas y categorías altitudinales son diferentes (Tablas 4, 6 y 8). Poco más de la mitad (52.2%) de los marcadores están asociados a una proteína, de estas 16 proteínas diferentes identificadas sólo 6 se encuentran caracterizadas. Es importante resaltar que, dado que el chip de SNPs del que se obtuvieron los datos genómicos ha sido diseñado para identificar regiones del genoma asociadas a características de

interés agronómico a partir del genoma de la línea mejorada B73, es probable que haya más regiones del genoma en las que haya variaciones entre las razas pero que no sean identificadas por el chip utilizado. Esto se ve reforzado por la diferencia en tamaños genómicos que existe entre las razas de maíz y la reducción del tamaño genómico de líneas mejoradas comparadas con el de razas nativas (Diez *et al.*, 2013). Estos datos justifican la idea de que existe mucha variación funcional en el genoma de las razas de maíz que no ha sido estudiada.

Entre los SNPs que sirven para diferenciar entre tres razas (Tehua, Zapalote Chico y Zapalote Grande), un SNP está asociado a la chaperonina 10 (cpn10). Las proteínas chaperoninas son indispensables para el funcionamiento de un organismo, su función principal es mediar en el plegamiento de proteínas. La chaperonina 10 se clasifica como una chaperonina del tipo I, su funcionamiento lo lleva a cabo en conjunto con la chaperonina 60 (cpn60) también del tipo I. Estas chaperoninas del tipo I están presentes en bacterias, mitocondrias y cloroplastos, dada su alta conservación, su funcionamiento se ha descrito en bacterias en donde la acción conjunta de la cpn60 y cpn10 sirve para el correcto plegamiento de proteínas en situaciones de estrés. En *A.thaliana* se han encontrado cuatro homólogos de la chaperonina 10 con expresión tanto mitocondrial como en cloroplasto mostrando que el funcionamiento de las chaperoninas de tipo I es más complejo en plantas que en bacterias, pero probablemente también contribuyen al correcto plegamiento de proteínas en situaciones de estrés (Levy-Rimler *et al.*, 2002).

Los SNPs informativos de categorías altitudinales y los SNPs informativos de razas comparten el factor de transcripción de choque de calor 30 (Hsf30). Esta proteína forma parte de la familia de factores de transcripción de choque de calor, moléculas responsables de la respuesta celular a estrés térmico (Westerheide *et al.*, 2012). El factor Hsf30 identificado en el presente trabajo no se encuentra funcionalmente caracterizado en maíz, sin embargo es homólogo del

factor de choque de calor HsfA2 presente en otras especies de plantas en las que se ha estudiado su funcionamiento (ejemplo: *A.thaliana*, *S.lycopersicum*). En *A.thaliana*, el factor HsfA2 es el factor de transcripción de choque térmico con mayor expresión bajo condiciones de estrés de calor y funciona como un regulador un conjunto de genes de respuesta al estrés (Schramm *et al.*, 2006). En el tomate, se ha mostrado que la síntesis del factor HsfA2 es estrictamente dependiente de estrés por calor, siendo un activador transcripcional fuerte para la expresión de proteínas de choque de calor. La proteína sintetizada por Hsf2A puede encontrarse en tres diferentes formas: 1) como factor de transcripción en el núcleo, 2) insoluble en el citoplasma formando parte de complejos chaperones en granulos de choque por calor y 3) forma soluble en el citoplasma. Las tres formas en que se puede encontrar la proteína Hsf2A están asociadas a la respuesta a choque térmico (Nover *et al.*, 2001). Dada la homología del factor de choque térmico identificado (Hsf30) con Hsf2A, se puede asociar su función con la respuesta al choque por calor. La correlación entre altitud y temperatura, sugiere que la diferencia alélica en este gen se ha sido fijado en las razas de diferentes altitudes porque tiene una influencia en su adecuación para el ambiente en el que se desarrollan.

En concreto, contando a la proteína de choque térmico, 5 de los 18 marcadores del análisis de razas coinciden con los SNPs del análisis de categorías altitudinales. En otras palabras, el análisis para identificar SNPs que diferencien las razas arrojó al menos 13 SNPs independientes de la altitud. Esto es importante debido a que la altitud ha sido descrita como la principal variable detrás de la estructura genética del maíz en México (Breña Ochoa, 2013, Arteaga *et al.*, 2016). Tomando en cuenta que el rango de fecha de muestreo de cada raza va de la década de 1940 a la década de 2000, los SNPs identificados muestran un mantenimiento temporal que refuerza la idea de que estas regiones del genoma están bajo una presión selectiva que hace que se mantengan uniformes dentro de las razas y diferenciadas entre éstas.

Un grupo de SNPs que podrían estar asociados a características fenotípicas interesantes son los SNPs que tanto en el análisis de tres como en el de cinco razas son informativos para el eje vertical. El eje vertical de estos análisis separa razas de categorías altitudinales similares. Una posible interpretación de esto es que estos SNPs están asociados con caracteres diferenciados en las razas independientes del ambiente y posiblemente resultado de la selección artificial de los campesinos. El factor de transcripción IIIA se encuentra en un gen asociado a estos SNPs. Un homólogo de este factor de transcripción ha sido caracterizado para *A.thaliana*, donde se sabe que tiene un papel regulatorio en la transcripción de genes de RNA ribosomal 5S y tiene un funcionamiento especial al unirse tanto al gen del rRNA 5S como al rRNA producido tras la transcripción (Mathieu *et al.*, 2003). La función de un factor de transcripción de este tipo también se estudió en arroz, donde se asoció al desarrollo de la planta y a la tolerancia a estrés abiótico por salinidad, bajas temperaturas y sequía (Huang *et al.*, 2012). La variación en un factor de transcripción puede tener grandes efectos en el desarrollo de las plantas debido a los cambios en regulación genética que implica una modificación a un factor de transcripción. Los factores de transcripción forman parte de redes de regulación transcripcional que determinan el patrón de expresión génica en las células de un organismo durante su desarrollo. Modificaciones en los elementos de la red transcripcional pueden acarrear cambios no lineales en la dinámica de la red que significan grandes cambios en el fenotipo del organismo que los porta (Dávila-Velderraín y Álvarez-Buylla, 2014).

En el análisis DAPC utilizando las cinco razas, llama la atención que las muestras de Zapalote Chico son el grupo más alejado del resto de las razas y también son las muestras con mayor probabilidad de pertenecer a su propia raza. La separación de las muestras de Zapalote Chico de las otras razas llama especialmente la atención porque pertenece al mismo grupo de razas que Conejo y se le atribuye ser una de las razas a partir de las cuales surgió Zapalote Grande, lo que hace

suponer que tendría que tener una posición cercana a estas razas. La particularidad y estabilidad de la raza Zapalote Chico ya ha sido estudiada por otras vías, Gutierrez-Nava y colaboradores reportaron en 1998 la presencia en muestras de Zapalote Chico de transposones del tipo *MuDR* característicos de las líneas de maíz *Mutator* características por su alta tasa de mutaciones somáticas. En su trabajo demuestran que en muestras de Zapalote Chico hay elementos *Mu* activos asociados a mutaciones somáticas en plantas surgidas por el cruzamiento de Zapalote Chico con muestras de razas diferentes, en un fenómeno de disgenesia híbrida. De alguna forma, la raza Zapalote Chico presenta elementos *MuDR* cuya actividad mutagénica es reprimida cuando hay cruzamientos dentro de la misma raza, pero que ocasionan mutaciones cuando es cruzado con razas diferentes. A partir de estos descubrimientos, los autores explican la estabilidad y homogeneidad de la raza Zapalote Chico al tener un aislamiento de las otras razas por el efecto de disgenesia híbrida. Esta característica también explica el mito zapoteco según el cuál el maíz Zapalote Chico “mata” a otras líneas de maíz cuando se cruzan, lo que ocasiona que los campesinos zapotecos que siembran Zapalote Chico evitan la introducción de germoplasma de otras razas (Gutierrez-Nava *et al.*, 1998).

El presente estudio se centró en diferencias a nivel genómico entre las razas de maíz, por lo tanto no considera otros factores que ocasionan variedad fenotípica y que pueden tener una fuerte influencia en la diferenciación de las razas de maíz. En particular, el manejo del sistema agroecológico en el que se desarrollan las diferentes razas de maíz puede ocasionar cambios fenotípicos entre las razas (Jardón Barbolla y Benítez Keinrad, 2015). Considerando que el maíz se desarrolla en sistemas agroecológicos complejos en los que interactúan factores biológicos (interacción con otras especies, flujo génico, presencia de polinizadores, microbioma del suelo, etc.), sociales y culturales (técnicas de cultivo, prácticas de intercambio de semilla, extensión y patrón de tierras de cultivo, usos religiosos y simbólicos, etc.), la comprensión de

la diversidad del maíz requiere de estudiar al maíz y el agroecosistema en el que vive como un todo en constante evolución. Es importante también siempre tener en cuenta que el maíz sólo puede vivir asociado a grupos humanos, por lo que para entender su evolución es necesario tomar en cuenta a los humanos que lo están sembrando. Asimismo, la conservación de la diversidad del maíz requiere de la conservación de las culturas y los agroecosistemas en los que es cultivado.



## CONCLUSIÓN

Este trabajo demuestra que las distinciones raciales del maíz tienen un sustento en la variación genómica de éstas, se logró el objetivo general de caracterizar muestras de razas mexicanas de maíz de acuerdo a la combinación de marcadores genómicos que presentan. Siguiendo esta lógica, en este trabajo se encontraron 18 marcadores moleculares que sirven para la diferenciación de las cinco razas, además de otros 12 marcadores independientes que sirven para diferenciar entre tres de las cinco razas incluidas. Entre estos 30 marcadores, 12 se encuentran asociados a proteínas de las cuales sólo cuatro están caracterizadas. De acuerdo a este resultado, se puede asignar raza a una muestra de maíz a partir de datos moleculares sin la necesidad de analizar caracteres morfológicos de la mazorca.

La variación en las razas de maíz mexicano es resultado de la interacción de múltiples factores ambientales y culturales. Estos diferentes factores ocasionan presiones selectivas en regiones distintas del genoma. Esto se pudo comprobar en el trabajo al distinguir marcadores asociados a diferencias altitudinales y otros asociados a categorías raciales. Se encontraron 12 marcadores asociados a altitud, cuatro de éstos estaban también entre los marcadores que sirven para distinguir razas. Entre estos 12 marcadores, seis se encuentran asociados a proteínas y tres de estas proteínas están caracterizadas.

Los resultados obtenidos permiten aceptar la hipótesis. Los caracteres morfológicos que definen las diferentes razas de maíz están asociados a genes seleccionados por factores ambientales y culturales en las poblaciones de una misma raza, por lo que existe una combinación de marcadores moleculares que distinguen muestras de diferentes razas.

El maíz tiene una enorme diversidad fenotípica como consecuencia de la heterogeneidad cultural de los grupos humanos que lo utilizan y de la variedad de ambientes en los que crece. Esta diversidad fenotípica ha sido descrita y organizada en razas formadas por grupos de poblaciones

con características en común. La conservación de la riqueza genómica y fenotípica de las razas de maíz requiere que se mantengan los factores y procesos bioculturales en los que éstas evolucionan.

## **PERSPECTIVAS**

Los resultados de este trabajo muestran que existe una enorme riqueza genómica en las razas de maíz mexicano que es la base de sus diferencias fenotípicas. Se encontraron marcadores moleculares que sirven para diferenciar cinco razas de maíz mexicano a partir de datos de SNPs genómicos. Hay muchas direcciones por las que se puede continuar con este trabajo. Algunas propuestas para continuar el trabajo son:

- Validación de los resultados. Los resultados obtenidos en este trabajo tienen que ser validados, un primer acercamiento para su validación sería probar la capacidad de clasificación del modelo utilizando más muestras de cada raza. Si bien los análisis mostraron marcadores que diferenciaban a las muestras de cada razas, para poder afirmar que esas diferencias son características de la raza se tendría que tener una muestra más amplia para cada muestra.
- Ampliar el número de razas. Una dirección lógica para continuar con el estudio es aumentar el número de razas incluidas. Sin embargo, es importante considerar que al ampliar el número de razas habrá mayor variación y encontrar marcadores que sirvan para distinguirlas será más complicado. Este trabajo se logró el objetivo de encontrar marcadores que distinguieran entre razas en parte por el método para elegir razas y muestras de éstas. Se incluyeron pocas razas (5) con características fenotípicas contrastantes y de cada raza se buscaron alrededor de diez muestras colectadas en diferentes tiempos y localidades, buscando abarcar toda la distribución geográfica de la raza. Para aumentar el número de razas se deben tomar en cuenta estos detalles, realizar análisis eligiendo razas representativas de caracteres fenotípicos contrastantes y así hacer una inclusión progresiva de razas.
- Utilizar datos de secuenciación genómica. Este trabajo se realizó utilizando un chip de SNPs que, aunque buscan cubrir la mayor

parte del genoma, han sido elegidos a partir del genoma de la línea comercial endogámica de maíz B73 y sólo recuperan información de *loci* asociados a características de interés para el mejoramiento agronómico. Utilizando información de secuenciación de genoma completo se podrían encontrar regiones con variación en las razas de maíz que no son rescatadas utilizando un chip de SNPs. Tomando en cuenta que lo que se busca son regiones genómicas con efectos fenotípicos que pueden ser objeto de selección natural o artificial, se pueden buscar opciones en tecnologías de secuenciación que reduzcan la complejidad del genoma tales como genotipo por secuenciación (GBS, por sus siglas en inglés). Otra opción sería hacer análisis de transcriptoma para encontrar diferencias en las regiones expresadas.

- Resequenciación de los *loci* identificados. Si bien se encontraron SNPs diferenciados entre las razas, para poder encontrar el efecto de estas diferencias alélicas y la verdadera causa de su diferenciación entre las razas es necesario secuenciar las regiones genómicas de estos *loci*, analizar si estas diferencias se mantienen y que posibles efectos tienen.
- Análisis de genómica de paisaje. En este análisis se identificaron *loci* asociados a diferencias ambientales tomando en cuenta únicamente factores altitudinales. Sin embargo, la variación ambiental tiene muchos factores que no están ligados a la altitud y que deben generar sus propias presiones selectivas en el genoma. Realizando un análisis de genómica del paisaje se tomarían en cuenta más factores ambientales y se podrían identificar regiones del genoma con variación asociada a estos factores. Es importante saber que aunque se encuentren diferencias genómicas correlacionadas con factores ambientales, dada la naturaleza del maíz como especie domesticada asociada siempre con el cultivo por humanos, hay factores sociales y culturales que no serían tomados en cuenta en estos análisis y que deben de tener influencia en el genoma.

- Analizar fuentes de variación no genóticas. La relación entre genotipo y fenotipo no es lineal. Por lo tanto, a pesar de la importancia de encontrar variación genómica asociada a factores fenotípicos, es importante estar consciente de existen otras fuentes de variación fenotípica (por ejemplo, influencia ambiental, manejo durante el desarrollo, interacciones con otras especies) que tienen que ser tomadas en cuenta si se quiere entender como surge y es mantenida la variación en las especies domesticadas.

## REFERENCIAS

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

Amigó, E., Gonzalo, J., Artiles, J., & Verdejo, F. (2009). A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information retrieval*, 12(4), 461-486.

Anderson, E. y Cutler., H.C. (1942). Races of *Zea mays*: I. Their recognition and classification. *Ann. Missouri Bot. Gard.* 29:69-88

Andorf, CM, Cannon, EK, Portwood, JL, Gardiner, JM, Harper, LC, Schaeffer, ML, Braun, BL, Campbell, DA, Vinnakota, AG, Sribalusu, VV, Huerta, M, Cho, KT, Wimalanathan, K, Richter, JD, Mauch, ED, Rao, BS, Birkett, SM, Richter, JD, Sen, TZ, Lawrence, CJ. (2015) MaizeGDB 2015: New tools, data, and interface for the maize model organism database. *Nucleic Acids Research*

Arteaga, MC., Moreno-Letelier A., Mastretta-Yanes, A., Vázquez-Lobo, A., Breña Ochoa, A., Moreno-Estrada, A., Eguiarte E. L. and Piñero, D. (2016) Genomic variation in recently collected maize landraces from Mexico. *Genomics Data*

Balée, W.L., Erickson, C. (Ed.). (2006).-*Time and Complexity in Historical Ecology. Studies in the Neotropical Lowlands.* Columbia University Press. E.U.A. 433 p.

Benz, B.F. (1986). Taxonomy and evolution of Mexican maize. Ph.D. dissertation. University of Wisconsin, USA. 433 p.

Breña Ochoa, A. (2013) *Estructura genética espacial asociada a la selección en razas de maíz de México.* Tesis (Maestría en Ciencias

Biológicas). México D.F. Universidad Nacional Autónoma de México. 60 pp.

Clayton D. (2015). *snpStats: SnpMatrix and XSnpmatrix classes and methods*. R package version 1.20

CONABIO (2011) *Proyecto Global de Maíces Nativos: recopilación, generación, actualización y análisis de información acerca de la diversidad genética de maíces y sus parientes silvestres en México*. México. <http://www.biodiversidad.gob.mx/usos/maices/razas2012.html>

Davila-Velderrain, J., & Alvarez-Buylla, E. R. (2014). Bridging the Genotype and the Phenotype: Towards An Epigenetic Landscape Approach to Evolutionary Systems Biology. *BioRxiv*.

Diez, C. M., Gaut, B. S., Meca, E., Scheinvar, E., Montes-Hernandez, S., Eguiarte, L. E., & Tenaillon, M. I. (2013). Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist*, 199(1), 264–276. <http://doi.org/10.1111/nph.12247>

Dyer, G. a, & Taylor, J. E. (2008). A crop population perspective on maize seed systems in Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, 105(2), 470–475. <http://doi.org/10.1073/pnas.0706321105>

Ganal M. W., G. Durstewitz, A. Polley, A. Bérard, E. S. Buckler, A. Charcosset, J. D. Clarke, E.- M. Graner, M. Hansen, J. Joets, M.-C. Le Paslier, M. D. McMullen, P. Montalent, M. Rose, C.-C. Shön, Q. Sun, H. Walter, O. C. Martin, M. Falque. (2011). A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE* 6: e28334.

Gepts, P. (2004). Crop Domestication as a Long term Selection Experiment. *Plant Breed Rev.* 2004;24:1–44.

Gutiérrez-Nava, M. L., Warren, C. a., León, P., & Walbot, V. (1998). Transcriptionally active MuDR, the regulatory element of the mutator transposable element family of *Zea mays*, is present in some accessions of the Mexican land race Zapalote chico. *Genetics*, 149(1), 329-346.

Hedrick, P. (2011). *Genetics of Populations*. Jones and Bartlett Publishers. 4°Ed. U.S.A. 675 pp.

Hernández-Xolocotzi, E. y G. Alanís Flores. (1970). Estudio morfológico de cinco nuevas razas de maíz de la Sierra Madre Occidental de México. *Agrociencia* 5(1): 3-30.

Heiser, C. B. (1988). Aspects of Unconscious Selection and the Evolution of Domesticated Plants. *Euphytica*, 37, 77-81.

Huang, J., Sun, S., Xu, D., Lan, H. Sun, H., Wang, Z., Bao, Y., Wang, J., Tang, H., Zhang, H. (2012) A TFIIIA-type zinc finger protein confers multiple abiotic tolerances in transgenic rice (*Oriza sativa* L.) *Plant Molecular Biology*. 80(3)

Hufford, M. B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R. a,... Ross-Ibarra, J. (2012). Comparative population genomics of maize domestication and improvement. *Nature Genetics*, 44(7), 808-11

Jardón Barbolla L. y Benítez Keinrad M. (2015). La comunidad agroecológica como unidad ecológica, de domesticación y de conservación. En: *Naturaleza, Ciencia y Sociedad: 40 años en la Facultad de Ciencias*. Editor: F. Guerrero McManus. Facultad de Ciencias-CEIICH, UNAM

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405



Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94.

Levy-Rimler, G., Bell, R. E., Ben-Tal, N., & Azem, A. (2002). Type I chaperonins: Not all are created equal. *FEBS Letters*, 529(1), 1-5.

Liang C, Jaiswal P, Hebbard C, Avraham S, Buckler ES, Casstevens T, Hurwitz B, McCouch S, Ni J, Pujar A, Ravenscroft D, Ren L, Spooner W, Teclé I, Thomason J, Tung CW, Wei X, Yap I, Youens-Clark K, Ware D, Stein L. (2008). Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.* 36:947-53

Mathieu, O., Yukawa, Y., Prieto, J. L., Vaillant, I., Sugiura, M., & Tourmente, S. (2003). Identification and characterization of transcription factor IIIA and ribosomal protein L5 from *Arabidopsis thaliana*. *Nucleic Acids Research*, 31(9), 2424-2433. <http://doi.org/10.1093/nar/gkg335>

Mera Ovando, L.M. (2009) Diversificación y distribución reciente del maíz en México. En: Bye Boettler, R. A., Serratos Hernández, J. A., Mera Ovando, L. M., Mapes Sánchez, C., & Kato Yamakake, T. *Origen y Diversificación del Maíz. Una revisión analítica*. México D.F.: Universidad Nacional Autónoma de México, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. México. pp. 69-81

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature Reviews. Genetics*, 14(12), 840-52.

Nover, L., Bharti, K., Döring, P., Mishra, S. K., Ganguli, a, & Scharf, K. D. (2001). *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress & Chaperones*, 6(3), 177-189.

Ortega Paczka, R. (2003). La diversidad del maíz en México. En: Esteva, G., Marielle, C. (coords.). *Si maíz no hay país*. pp. 123-154. México D.F.: Consejo Nacional para la Cultura y las Artes.

Perales, H., & Golicher, D. (2014). Mapping the diversity of maize races in Mexico. *PloS One*, 9(12),

Piperno, D. R, Ranere, A. J., Holst, I., Iriarte, J., Dickau, R. (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences of the United States of America*. 106:5019-5024.

R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ross-Ibarra, J., Morrell, P. L., & Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, 104 Suppl , 8641-8.

Ruiz Corral, J. A., Hernández Casillas, J. M., Sánchez Gonzáles, J. D. J., Corona Ortega, A., Ojeda Ramírez, G., Guerrero Herrera, M. D. J., ... De la Cruz Larios, L. (2013). *Ecología, adaptación y distribución actual y potencial de las razas mexicanas de maíz*. INIFAP. México D.F. 149 p.

Sánchez González, J.J., Goodman, M.M, and C.W. Stuber (2000) Isozymatic and morphological diversity in the races of maize of Mexico. *Economic Botany* 54(1):43-59.

Schnable et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112-1115.

Schramm, F, Ganguli, A., Kiehlmann, E., Englich, G., Walch, D., y von Koskull-Döring, P. (2006). The heat stress transcription factor HsfA2

serves as a regulatory amplifier of a subset of genes in the heat stress response in *Arabidopsis*. *Plant Molecular Biology*. 60(5): 759-772

Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA). (2014). Protección, utilización, desarrollo y aprovechamiento sustentable de las razas y variedades de maíz nativos en México. En <http://www.sinarefi.org.mx/redes/politicaspUBLICASmaiz.pdf>

Stiner, M. (2001). Thirty years on the “Broad Spectrum Revolution” and paleolithic demography. *Proc. Natl. Acad. Sci. (USA)* 98:6993–6996.

van Heerwaarden, J. Van, Doebley, J., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., & Sánchez González, J. J. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences of the United States of America*, 108(3).

Vigouroux, Y., Glaubitz, J. C., Matsuoka, Y., Goodman, M. M., Sánchez G., J. J., & Doebley, J. (2008). Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *American Journal of Botany*, 95(10), 1240–53.

Weisdorf, J. L. (2005). From Foraging To Farming: Explaining The Neolithic Revolution. *Journal of Economic Surveys*. 19(4), 561–586.

Wellhausen, E., Roberts, L.M., Hernández X., E., Mangelsdorf, P.C. (1951) *Razas de maíz en México, su origen, características y distribución*. Folleto Técnico no.5, México, D. F: Oficina de Estudios Especiales, Secretaría de Agricultura y Ganadería.

Westerheide, S. D., Raynes, R., Powell, C., Xue, B., & Uversky, V. N. (2012). HSF transcription factor family, heat shock response, and protein intrinsic disorder. *Current Protein & Peptide Science*, 13(1), 86–103.

Wright, S. I., Bi, I. V., Schroeder, S. G., Yamasaki, M., Doebley, J. F., McMullen, M. D., & Gaut, B. S. (2005). The effects of artificial selection on the maize genome. *Science*. New York, N.Y., 308(5726), 1310-1314. <http://doi.org/10.1126/science.1107891>

Yamasaki, M., Tenailon, M. I., Bi, V., Schroeder, S. G., Sanchez-villeda, H., Doebley, J. F., McMullen, M. D. (2005). A Large-Scale Screen for Artificial Selection in Maize Identifies Candidate Agronomic Loci for Domestication and Crop Improvement. *The Plant Cell*. 17, 2859-2872.

Zhao, W., Canaran, P., Jurkuta, R., *et al.* (2005) Panzea: a database and resource for molecular and functional diversity in the maize genome. *Nucleic Acid Research* ,34,D752-D757.

Zheng X, Levine D, Shen J, Gogarten S, Laurie C and Weir B (2012). A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics*, 28(24), pp. 3326-3328.

Zohary, D. (2004). Unconscious Selection and the Evolution of Domesticated Plants, *Economic Botany*, 58(1), 5-10.