



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS**

Instituto de Biología
Biología Evolutiva

**ESTRUCTURA DE LA DIVERSIDAD GENÓMICA DEL CHILTEPÍN (*CAPSICUM
ANNUUM* VAR. *GLABRIUSCULUM*) EN MÉXICO, EVALUADO A PARTIR DE ADN
DE DIFERENTES FUENTES DE TEJIDO Y GRADO DE DOMESTICACIÓN**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRA EN CIENCIAS BIOLÓGICAS

PRESENTA:

ELSA GABRIELA MARTÍNEZ ANDRADE

TUTORA PRINCIPAL DE TESIS:

Dra. Alicia Mastretta Yanes
CONABIO

COTUTORA DE TESIS:

Dra. Alejandra Citlalli Moreno Letelier
Instituto de Biología, UNAM

COMITÉ TUTOR:

Dr. Lev Orlando Jardón Barbolla
CEIICH, UNAM

CUIDAD UNIVERSITARIA, CD. MX., MAYO, 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS**

Instituto de Biología
Biología Evolutiva

**ESTRUCTURA DE LA DIVERSIDAD GENÓMICA DEL CHILTEPÍN (*CAPSICUM
ANNUUM* VAR. *GLABRIUSCULUM*) EN MÉXICO, EVALUADO A PARTIR DE ADN
DE DIFERENTES FUENTES DE TEJIDO Y GRADO DE DOMESTICACIÓN**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRA EN CIENCIAS BIOLÓGICAS

PRESENTA:

ELSA GABRIELA MARTÍNEZ ANDRADE

TUTORA PRINCIPAL DE TESIS:

Dra. Alicia Mastretta Yanes
CONABIO

COTUTORA DE TESIS:

Dra. Alejandra Citlalli Moreno Letelier
Instituto de Biología, UNAM

COMITÉ TUTOR:

Dr. Lev Orlando Jardón Barbolla
CEIICH, UNAM

CUIDAD UNIVERSITARIA, CD. MX., MAYO, 2023

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE BIOLOGÍA

OFICIO CPCB/277/2023

ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que el Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día 23 de agosto de 2021 se aprobó el siguiente jurado para el examen de grado de MAESTRA EN CIENCIAS BIOLÓGICAS en el campo de conocimiento de BIOLOGÍA EVOLUTIVA de la estudiante MARTÍNEZ ANDRADE ELSA GABRIELA con número de cuenta 307194325 con la tesis titulada "Estructura de la diversidad genómica del chiltepín (*Capsicum annuum* var. *glabriusculum*) en México, evaluado a partir de ADN de diferentes fuentes de tejido y grado de domesticación", realizada bajo la dirección de la DRA. ALICIA MASTRETTA YANES quedando integrado de la siguiente manera:

Presidente: DR. ALEJANDRO CASAS FERNÁNDEZ
Vocal: DR. SOL CRISTIANS NIZAWA
Vocal: DR. LUIS ENRIQUE EGUIARTE FRUNS
Vocal: DRA. ALEJANDRA VÁZQUEZ LOBO YURÉN
Secretario: DR. LEV ORLANDO JARDÓN BARBOLLA

Sin otro particular, me es grato enviarle un cordial saludo.

A T E N T A M E N T E
"POR MIRAZA HABLARÁ EL ESPÍRITU"
Ciudad Universitaria, Cd. Mx., a 21 de marzo de 2023

COORDINADOR DEL PROGRAMA



DR. ADOLFO GERARDO NAVARRO SIGÜENZA



Agradecimientos institucionales

Al Posgrado en Ciencias Biológicas, UNAM.

Este trabajo se financió por medio del Proyecto CONACYT No. 247730 "La variación genética de las plantas cultivadas en México: estrategias para enfrentar el cambio climático" y por el proyecto titulado "Genetic structure and mechanisms of drought adaptation in *Capsicum* " de la Agriculture and Food Research Initiative-USDA. Así como la beca nacional CONACYT no. 737066.

Finalmente le quiero agradecer a mis tutoras la Dra. Alicia Mastretta Yanes por su compromiso y apoyo tanto académico como emocional en este proyecto; así como a la Dra. Alejandra Moreno Letelier que me hizo parte de su grupo de trabajo en el IB y por todo su apoyo académico e interés mostrado en este trabajo. A mi tutor el Dr. Lev Jardón Barbolla pues gracias a él conocí la especie de este proyecto en campo y por todo el apoyo mostrado en la realización de esta tesis, además de incluirme en su grupo de trabajo con sus colaboradores de Ohio. Gracias a ustedes por mostrarme su amor a la evolución.

Agradecimientos a título personal

Quiero agradecer a *mis padres* cuyo apoyo me anima cada día a seguir luchando por mis metas. A mi mamá por ser mi cómplice en la vida, por ser mi soporte, tu apoyo constante y tu cariño incondicional fueron la fuente de inspiración que me impulsó a concluir esta tesis. Sin tu amor y tu respaldo, este logro no habría sido posible. A mi papá, quien siempre me brindó palabras de aliento y un apoyo constante, incluso en los momentos más difíciles. Agradezco su cariño incondicional y su alegría, que hicieron más llevadera la realización de este trabajo. A mis hermanos quienes siempre me impulsan a seguir adelante y perseguir mis sueños. A mis abuelas quienes siempre me han inculcado el amor a la naturaleza con sus historias de vida. A Balam, quien fue un gran apoyo emocional durante todo este proceso. Aunque no pueda leer estas palabras, quiero que sepa lo significativo que ha sido para mí su compañía. Kristov, quisiera agradecerte de manera especial por todo el apoyo que me brindaste durante este proceso, incluso cuando parecía imposible avanzar. Este trabajo no hubiera sido posible sin ti.

Agradezco profundamente a Alicia Mastretta por haber aceptado guiarme en este camino, ya que su ayuda académica en bioinformática y genómica fue fundamental para la realización de esta tesis. Además, quiero darte las gracias por tu apoyo emocional en los momentos difíciles y por darme las palabras precisas en el momento adecuado. A Ale Moreno, quien también me guió en este proceso, gracias por crear un ambiente de laboratorio enriquecedor y amistoso que fue fundamental en mi formación como estudiante, por todas las charlas de genómica que tuvimos y por cada seminario en el que siempre me contagiaste tu amor por la biología. A Lev Jardón, por todo el apoyo que me brindaste a lo largo de este camino académico. Te agradezco especialmente por haberme introducido en el estudio de *Capsicum* en campo y de forma teórica, por haberme incluido en tu grupo de trabajo con tus colegas de Ohio y por los seminarios a los que fui invitada.

A mis amigos de la facultad, especialmente a Tania y Cristóbal, quienes han sido mi apoyo incondicional y por todos los momentos buenos y malos que hemos compartido juntos. A mis amigas y amigos del posgrado, especialmente a Brenda, Lety y Emiliano, quienes compartimos el camino de la maestría y me contagiaron su amor por sus especies de estudio. A Lau, Erika, Noé y Yara, les quiero expresar mi gratitud por ser un apoyo incondicional en mi vida.

Quiero expresar mi más sincero agradecimiento al grupo de Genética de la Conservación. A los integrantes de Ale Moreno Emiliano, Niza, Rodrigo y Jorge les agradezco por hacer que mis días fueran amenos y divertidos, así como por su apoyo constante. También quiero agradecer a Mel, Oli, Duhya, Pame y Amanda, del grupo de Ana Wegier por esas tardes de charlas dentro y fuera del laboratorio. Al equipo de Alicia: Christian, Vero, Oli y Duhya por su

constante disposición para apoyarme en cuestiones técnicas y académicas, su colaboración fue fundamental para llevar a cabo este trabajo con éxito. A Ana Lau, por su pasión contagiosa por los chiles y por ser una excelente colega de laboratorio.

Al “Gorgojo Rojo”, proyecto estudiantil que nació un día aparentemente normal bajo los rayos del sol en el jardín espiga el Instituto de Biología, con alumnas y alumnos del laboratorio de Genética de la Conservación, que anhelaban formar una familia científica. A todos ellos, queridos gorgojos y gorgojas, mi agradecimiento profundo porque gracias a su pasión, mi estancia en el IB fue inolvidable.

Quiero expresar mi profundo agradecimiento al Dr. Luis Eguiarte, jefe del Laboratorio de Evolución Molecular y Experimental del Instituto de Ecología de la UNAM, por brindarme la oportunidad de llevar a cabo los experimentos realizados en esta tesis, así como por su generosidad al permitir el uso del invernadero donde germinaron algunos de los individuos estudiados en este trabajo. Y como esta tesis también es experimental gracias por todo su apoyo al denominado “equipo de extracción” que me apoyó incansablemente en el laboratorio. A Ana Lau, que me enseñaste a amar a los chiles, a Niza porque cuando llegaste todo era más ameno y agradable, a Emiliano que siempre me apoyó académica y emocionalmente, ¡son el mejor equipo de laboratorio!.

Quiero expresar mi agradecimiento al jurado que evaluó esta tesis, integrado por el Dr. Lev Jardón, el Dr. Luis Eguiarte, la Dra. Ale Vázquez, el Dr. Sol Cristians y el Dr. Alejandro Casas, pues sus valiosas observaciones enriquecieron académicamente esta tesis.

A Alexandra Elbakyan, cuyo compromiso con la ciencia ha desafiado el monopolio y las barreras que la restringen, permitiendo su acceso libre.

Esta tesis fue escrita en un tiempo desafiante para la humanidad y dedico estas palabras a todos aquellos que perdieron a un ser querido durante la pandemia COVID-19, así como a los científicos que, con su arduo trabajo y dedicación, lograron salvar millones de vidas a través de las vacunas, demostrando la importancia de la ciencia.

Para mi mamá y papá, por su amor y apoyo incondicional

-No -dijo el principito-. Yo busco amigos. ¿Qué significa “domesticar”?

-Es una cosa demasiado olvidada -dijo el zorro-. Eso significa “crear lazos”.

-Seguramente -dijo el zorro-. Tú todavía no eres para mí más que un niño parecido a cien mil niños. Y yo no tengo necesidad de ti. Y tú tampoco tienes necesidad de mí. Para ti no soy más que un zorro parecido a cien mil zorros. Pero si tú me domesticas, entonces tendremos necesidad el uno del otro. Serás para mí único en el mundo. Y yo seré para ti único en el mundo...

El zorro se calló, y miró largo rato al principito:

-¡Si quieres, domesticame! -dijo...

-¿Qué es lo que hay que hacer? -dijo el principito.

-Hace falta ser muy paciente -respondió el zorro-. Te sentarás primero un poco lejos de mí, de esta manera, en la hierba. Yo te miraré con el rabillo del ojo y tú no dirás nada. El lenguaje es fuente de malentendidos. Pero cada día podrás sentarte un poco más cerca...

Al día siguiente volvió el principito.

-Hubiera sido mejor volver a la misma hora -dijo el zorro-. Si tú vienes, por ejemplo, a las cuatro de la tarde, desde las tres empezaré a ser feliz. A medida que la hora se acerque me sentiré feliz ...

EL PRINCIPITO

Antoine de Saint-Exupéry

ÍNDICE

RESUMEN	1
ABSTRACT	3
INTRODUCCIÓN Y PLANTEAMIENTO DEL PROBLEMA	5
1. IMPORTANCIA DE LOS PARIENTES SILVESTRES DE VARIEDADES CULTIVADAS.....	5
2. ASPECTOS EVOLUTIVOS Y GENÉTICOS DE LA DOMESTICACIÓN.....	7
3. HISTORIA NATURAL, EVOLUTIVA, DOMESTICACIÓN Y GENÓMICA DEL GÉNERO CAPSICUM	12
4. EL CHILTEPÍN: BIOLOGÍA, FILOGENIA, DOMESTICACIÓN Y MANEJO	17
5. CONSERVACIÓN DE LA DIVERSIDAD GENÉTICA EN CAPSICUM	21
6. RETOS Y LIMITANTES DE ESTUDIOS MOLECULARES Y GENÉTICOS EN PARIENTES SILVESTRES	26
HIPÓTESIS	29
OBJETIVO GENERAL	30
OBJETIVOS PARTICULARES	30
METODOLOGÍA GENERAL	30
MATERIALES BIOLÓGICOS Y COLECTA.....	30
SECUENCIACIÓN	34
ANÁLISIS BIOINFORMÁTICOS	34
1.- <i>Preprocesamiento</i>	35
2.- <i>Alineamiento de lecturas</i>	36
3.- <i>Llamado de SNPs</i>	37
4.- <i>Análisis genómico-poblacionales</i>	38
RESULTADOS	41
MUESTREO.....	41
SECUENCIACIÓN Y GENOTIPADO	42
ANÁLISIS GENÓMICO-POBLACIONALES	46
DISCUSIÓN	60
PRE-PROCESAMIENTO GENÓMICO Y ANÁLISIS DE DIFERENTES FUENTES DE DNA DE CHILTEPÍN	60
MAPEO DE LAS LECTURAS O READS CONTRA EL GENOMA DE REFERENCIA.....	62
ANÁLISIS GENÓMICO-POBLACIONALES.....	64
DIVERSIDAD Y DIFERENCIACIÓN GENÓMICA.....	67
CONCLUSIONES	68
PERSPECTIVAS	69
REFERENCIAS	70
ANEXOS	89

RESUMEN

La alimentación humana recae mayoritariamente en una centena de cultivos, los cuales están sujetos a procesos tecnológicos con el fin de una mayor rendimiento y productividad, lo que evolutivamente se asocia de manera general, a la disminución en la variabilidad genética. Generalmente, las fuentes primarias de variación evolutivas son los parientes silvestres de los cultivos, los cuales no han pasado por el cuello de botella característico de la domesticación y pertenecen a poblaciones que se encuentran bajo selección natural, cuya diversidad genética, capacidad de adaptación y gran resiliencia a diferentes entornos se pueden aprovechar para el mejoramiento de los cultivos. La región de Mesoamérica es conocida por la gran cantidad de parientes silvestres de diversos cultivos entre los que se encuentra el pariente del chile (*Capsicum annum var. glabriusculum*), cuya distribución se ha visto seriamente amenazada. Nuestro estudio se llevó a cabo con ejemplares de chile de tres distintos estatus de domesticación: Silvestre, Semisilvestre y Domesticado, en México. Nuestro objetivo fue determinar como el grado de domesticación y la distancia geográfica influyen en las poblaciones del chile Silvestre y Semisilvestre. El estudio se realizó con muestras, obtenidas a partir de distintas fuentes de DNA: tejido seco, hoja congelada y fresca, fruto seco y congelado y muestras de herbario, con datos genómicos obtenidos por genotipado por secuenciación de GBS, por sus siglas en inglés. Tras aplicar filtros bioinformáticos de calidad de las secuencias, obtuvimos un total de 59,040 variantes genómicas o SNPs (polimorfismo de nucleótido único), presentes en 49 individuos. Nuestros resultados indican que se pueden obtener secuencias de manera exitosa a partir del fruto, que a nuestro conocimiento no había sido probado con anterioridad. Además, el porcentaje de mapeo al genoma de referencia no se encuentra asociado al estatus de domesticación, pero si a la fuente del DNA, de acuerdo al análisis de varianza (ANOVA). El análisis de componentes principales (PCA) nos mostró que hay cuatro grupos genómicos: Norte, Centro, Sur y Diverso; sin embargo, de acuerdo con los resultados FastStructure, solo se formaron 3 grupos genómicos: Norte, Centro y Sur, posiblemente están influidos por la latitud y longitud. El grupo con mayor número de alelos privados (23,579) y con mayor diversidad nucleotídica, (0.098) fue el Sur, mientras que el grupo con el menor número de alelos privados fue el Diverso (3,281) y el grupo con menor diversidad nucleotídica (0.037) fue el Norte. Este trabajo nos

permitió adentrarnos en la historia evolutiva del chile y de sus parientes silvestres desde un punto de vista genómico, el cual puede resultar útil para la conservación de la especie al apoyar la búsqueda de variedades resistentes a plagas u otras condiciones adversas. En este sentido, aunque consideramos que se debería profundizar en el estudio de la diversidad genética de todas las poblaciones del chile, una alternativa para dirigir los esfuerzos, podría ser la de centrarse en el grupo sur, para comprender los motivos de su alta diversidad genética y buscar genes de interés dentro de la misma.

ABSTRACT

Human feeding depends on a hundred crops, subjected to several technological processes to increase their yield and productivity, associated with a reduction in genetic variability. Generally speaking, the primary sources of evolutionary variation are the wild relatives of crops, which have not gone through the resulting bottleneck of domestication and belong to populations under natural selection, whose genetic diversity, adaptability, and great resilience to different environments can be used for crop improvement. The Mesoamerican region is known for its large number of wild relatives of diverse crops, such as the chiltepin (*Capsicum annuum* var. *glabriusculum*), the wild relative of the chili pepper, whose distribution has been seriously threatened. Our study was carried out with chili peppers specimens of three different domestication statuses: Wild, Semi-wild and Domesticated, in Mexico. Our objective was to determine the influence of the domestication degree and geographical distance in the populations of wild and semi-wild chili peppers. The study was carried out with samples obtained from different DNA sources: dry tissue, frozen and fresh leaves, dried and frozen fruits, and herbarium samples, with genomic data obtained by genotyping by sequencing (GBS). After performing the quality control of genomic reads, we obtained 59,040 single nucleotide polymorphisms (SNPs) from 49 individuals. Our results indicate that sequences can be successfully obtained from fruits, which had not been previously tested to our knowledge. According to the analysis of variance (ANOVA), the fraction of reads mapped to the reference genome was not associated with the domestication status, but rather with the DNA source. Principal component analysis (PCA) showed that there were four genomic groups: North, Central, South, and Diverse; however, according to the FastStructure results, only 3 genomic clusters were formed: North, Center, and South, possibly influenced by latitude and longitude. The group with the highest number of private alleles (23,579) and the greatest nucleotide diversity (0.098) was the South, while the group with the lowest number of private alleles was Diverse (3,281), and the group with the lowest nucleotide diversity (0.037) was the North. This work allowed us to delve into the evolutionary history of the chili pepper and its wild relatives from a genomic point of view, which may be useful for the conservation of the species by supporting the search for varieties resistant to pests or other adverse conditions. In this sense, although we believe that the study of the genetic diversity of all chili pepper populations should be deepened, one alternative to direct the efforts could be to focus

on the South group, to understand the reasons for its high genetic diversity and search for genes of interest within it.

Introducción y planteamiento del problema

1. Importancia de los parientes silvestres de variedades cultivadas

Actualmente la alimentación de la humanidad recae aproximadamente sobre una docena a una centena de cultivos de cerca de las 300,000 especies de angiospermas existentes (Casas y Parra, 2007; McCouch *et al.*, 2013). Es decir, que dependemos de un porcentaje mínimo de diversidad genética para la sobrevivencia humana, pues solo 30,000 especies de angiospermas son utilizadas para satisfacer alguna necesidad humana (McCouch *et al.*, 2013, Jardón-Barbolla, 2015). Bajo tal premisa, nos enfrentamos a nuevos retos marcados por un aumento en la demanda de los alimentos, el cambio climático, la carencia de agua dulce, la reducción en la disponibilidad de suelo debido a la urbanización, así como la erosión y salinización de los suelos, todo lo cual conduce a un rumbo insostenible para garantizar la alimentación humana (Godfray *et al.*, 2010; McCouch *et al.*, 2013).

La *diversidad biológica* de los *agroecosistemas* incluye a todas las especies vegetales con potencial importancia para el humano. Esto incluye a los parientes silvestres de cultivos, cuya característica principal es que no son poblaciones domesticadas, pero que han dado, o pueden dar origen a formas domesticadas o están relacionadas cercanamente a especies cultivadas, al tiempo que funcionan como reservorio génico de variación y en principio pueden cruzarse con los cultivos (Harlan y de Wet, 1971; Maxted *et al.*, 2006). Las especies domesticadas pasaron por un proceso de selección artificial regido por el ser humano, que en muchos casos resultó en que las especies domesticadas sean dependientes del ser humano, al grado de que necesitan de él para su propagación. Sin embargo, no todos los casos son tan extremos pues hay estadios intermedios en este proceso de domesticación (Harlan, 1992).

La pérdida de la agrobiodiversidad podría tener mayor impacto en regiones agrícolas ancestrales, que se erigen como centros de origen y diversificación de cultivos. Dado que generalmente los parientes silvestres y las variedades domesticadas nativas coexisten en el espacio y el tiempo, hay una mayor preocupación por la conservación y estudio de dichas zonas, como ocurre en Mesoamérica (Casas y Parra,

2007; McCouch *et al.*, 2013; Gepts, 2014; Mastretta-Yanes *et al.*, 2018). Una aproximación para la conservación de estos recursos botánicos es buscar una sinergia entre las ciencias sociales y biológicas, para mantener así la sostenibilidad y productividad de los recursos (Godfray *et al.*, 2010; McCouch *et al.*, 2013).

Las especies agrícolas pueden pasar por procesos tecnológicos para la producción de nuevos cultivos con mayor rendimiento y productividad (McCouch *et al.*, 2013). Generalmente, las fuentes primarias de variación evolutiva son los parientes silvestres de los cultivos, seguidos de las variedades locales de las plantas domesticadas. Esto se debe a que una característica que se ha supuesto intrínseca a la domesticación es la reducción de la diversidad genética en las especies domesticadas respecto a los parientes silvestres (McCouch *et al.*, 2013; Gepts, 2014; Mastretta-Yanes *et al.*, 2018). No obstante, esta supuesta reducción de la diversidad genética depende de la especie, el sistema de apareamiento, el entorno ecológico y sistema de manejo como monocultivo o policultivo (McCouch *et al.*, 2013). Un ejemplo de ello son los barridos selectivos, en los cuales hay un aumento de la frecuencia de los alelos que se seleccionaron positivamente y como resultado hay reducción significativa de la diversidad genética en sitios flanqueados por mutaciones seleccionadas positivamente (Meyer y Purugganan, 2013). En contraparte, después de un cuello de botella, la diversidad genética puede aumentar durante la fase de manejo, especialmente si este manejo se dirige a promover la diversificación y la introgresión con los parientes silvestres (Meyer y Purugganan, 2013).

Debido a su cercanía genética y evolutiva de los parientes silvestres y de las variedades locales, se puede aprovechar su diversidad génica, su capacidad de adaptación y gran resiliencia a diferentes entornos, aunado a que no presentan el cuello de botella típico de la domesticación y se encuentran bajo selección natural (Casas y Parra, 2007; McCouch *et al.*, 2013; Gepts, 2014; Mastretta-Yanes *et al.*, 2018; CONABIO *et al.*, 2019a; CONABIO *et al.*, 2019b); no obstante, gran parte de su variación sigue sin explorarse. Así si se presenta flujo génico de las poblaciones silvestres locales a las cultivadas, o hay preferencia de los humanos por ciertos cultivos, o existe adaptación local a diferentes condiciones ambientales en distintas regiones geográficas, entre otros procesos, se podría favorecer la conservación de altos niveles de diversidad génica (Cornille *et al.*, 2013).

Actualmente, la agricultura intensiva ha propiciado diversos cambios en el proceso de domesticación. Estos cambios conllevan a que la producción de semillas para siembra se concentre en unas pocas empresas agroindustriales (Matretta-Yanes *et al.*, 2018). Es por ello que la conservación de la diversidad depende del sistema de manejo que lleven a cabo los agricultores, quienes generalmente prefieren los policultivos, los cuales favorecen aspectos funcionales tales como retención de agua, el amortiguamiento de agentes erosivos y el manejo de un número considerable de variedades de la misma especie domesticadas, lo que resulta en una mayor diversidad biológica. En contraste, los agricultores marcados por las metas de la producción intensiva reducen drásticamente este espectro de variación (Casas y Vallejo, 2019). Por tal motivo la preservación de la diversidad genética de los cultivos recae sobre muchos pequeños agricultores que, al cultivar bajo diferentes condiciones ambientales y preferencias culturales, mantienen y gestionan la diversidad génica (Mastretta-Yanes *et al.*, 2018).

México se encuentra entre los países con mayor diversidad vegetal y Mesoamérica, que abarca gran parte del territorio mexicano y se extiende hasta Centroamérica, es una de las principales regiones en las que se encuentran distribuidos gran cantidad de parientes silvestres de plantas domesticadas (Casas y Parra, 2007; Hernández-Verdugo *et al.*, 2012; Castañeda-Álvarez *et al.*, 2016; CONABIO *et al.*, 2019a; CONABIO *et al.*, 2019b). Por ello, Mesoamérica se considera como centro de origen, diversidad y domesticación de algunos de los alimentos más destacados de la dieta mundial, entre los cuales destacan el chile, la calabaza, el maíz y el frijol (Vavilov, 1951; Harlan, 1971; CONABIO *et al.*, 2019a). Por lo tanto, Mesoamérica es una zona con especial importancia para la conservación de los parientes silvestres, en el contexto de los desafíos para adaptar los cultivos frente al cambio climático (Dempewolf *et al.*, 2014; Castañeda-Álvarez *et al.*, 2016).

2. Aspectos evolutivos y genéticos de la domesticación

Desde 1859 Darwin hace referencia de la selección consciente e inconsciente para el estudio de la domesticación (Ruse, 1975). La domesticación es un proceso histórico,

dinámico, con procesos evolutivos que operan en los parientes silvestres de plantas o animales, donde la selección es ejercida por el ser humano de forma consciente o inconsciente. En sentido evolutivo, la domesticación puede aumentar la adecuación bajo el cuidado del humano, y en unos casos disminuir sin el cuidado del mismo. De este modo, el mayor grado de domesticación ocurre cuando la supervivencia de la planta o animal depende enteramente del ser humano, o puede presentarse una domesticación incipiente, donde los organismos se reproducen sin la intervención humana (Meyer y Purugganan, 2013; Gepts, 2014; Jardón-Barbolla, 2015), por lo que el cultivo de las plantas es una se considera como una condición para que exista la domesticación. Además de la selección artificial ejercida por los humanos, las otras fuerzas evolutivas, como la deriva génica, la selección natural y el flujo génico (introgresión) también moldean las características sujetas a domesticación (Gepts, 2014).

El origen de la domesticación de plantas se remonta a hace al menos 12 mil años, momento en el que algunos parientes silvestres se comenzaron a usar como materia prima génica de la que se derivaron los cultivos (Doebley, 1989; Meyer y Purugganan, 2013; Gepts, 2014). De esta manera, la domesticación es un proceso continuo que actúa sobre plantas en fases incipientes o tempranas, también opera sobre otras especies semi-domesticadas y continúa operando aún sobre especies ya domesticadas (Casas *et al.*, 2007).

A diferencia de una población enteramente silvestre, una población semi-domesticada ha sido modificada significativamente por el manejo y la selección e intervención humana, por lo que el fenotipo promedio puede diferir del rango encontrado en las poblaciones silvestres para los rasgos bajo selección. El estudio de las interacciones de las personas con plantas que hoy se encuentran en las etapas de domesticación o semi-domesticación incipiente puede conducir a una mejor comprensión de los procesos que conducen a la domesticación en la prehistoria (Meyer y Purugganan, 2013; Casas *et al.*, 2016). De tal forma que para este trabajo consideraremos que las plantas silvestres que no tienen contacto directo con los humanos y las plantas semi-silvestres las catalogamos como aquellas fenotípicamente más cercanas a las silvestres, pero que se encuentran en un ambiente de manejo,

sujetas a selección artificial consciente, las cuales se ubican generalmente en traspatios (Meyer *et al.*, 2012)

Las plantas también pueden ser fomentadas en las que puede o no haber selección humana, cuyas características fenotípicas se asocian a condiciones favorables encontradas en el agrohábitat y en menor medida a la selección consciente de individuos con características morfológicas específicas (Rendón-Aguilar *et al.*, 2017). A pesar de que hay selección e intervención humana, el fenotipo tiende a ser similar a las poblaciones silvestres para las características seleccionadas ya que existe mucho flujo génico con las poblaciones silvestres o porque la selección humana no es lo suficientemente fuerte (Casas *et al.*, 2016).

Otro grupo estrechamente relacionado son las plantas arvenses, que tienen gran importancia en las milpas o huertos familiares, algunas de las cuales son toleradas deliberadamente en los campos de cultivo y algunas incluso son fomentadas. Las plantas arvenses crecen a partir de reclutamiento en ambientes perturbados de manera recurrente por los seres humanos, pero no dependen de éstos para sobrevivir o reproducirse. Las plantas arvenses poseen adaptaciones a ambientes antrópicos, y algunas especies son deliberadamente favorecidas o promovidas, para lo cual las semillas se guardan para el próximo ciclo agrícola, lo cual puede involucrar una selección incipiente (Rendón-Aguilar *et al.*, 2017).

Las características involucradas en los llamados síndromes de domesticación o los rasgos que diferencian a los cultivos domesticados de las especies silvestres pueden ser fisiológicos, bioquímicos y morfológicos (Gepts, 2014). Entre estas características destacan la reducción del hábito de crecimiento (pocas ramas y más cortas); pérdida de la latencia de semillas; tendencia a los sistemas de reproducción endogámicos como la propagación vegetativa; aumento del tamaño de las semillas y frutos; reducción de la toxicidad e incremento a la tolerancia a factores estresantes como la sequía, por mencionar algunos. Como consecuencia, la productividad de los parientes silvestres es menor en comparación con las plantas domesticadas en términos de superficie y tiempo de crecimiento de los cultivos (Meyer y Purugganan, 2013; Gepts, 2014).

Los síndromes de domesticación dependen de la historia de vida de los organismos (anuales *versus* perennes) y el sistema reproductivo (como autofecundación, reproducción vegetativa y heterocruza). Asimismo, los síndromes de domesticación alteran los rasgos fenológicos como el tipo y tiempo de floración, además de la época de crecimiento (Gepts, 2014). Otra característica común que resulta de la domesticación es la reducción de la diversidad genética en comparación con los parientes silvestres de los cultivos. Debido a lo anterior, muchos alelos que están relacionados con la domesticación se encuentran en frecuencias bajas o moderadas en los parientes silvestres (Meyer y Purugganan, 2013). Por lo tanto, generalmente se espera mayor variación genética en los parientes silvestres de los cultivos que en los cultivares, pues no han pasado por cuellos de botella característicos de la domesticación que reduzcan el tamaño efectivo de la población (Doebley, 1989; Moyers *et al.*, 2018). No obstante, los niveles de variación se pueden modificar posteriormente debido al flujo génico con los parientes silvestres. Como se mencionó anteriormente, la reducción de la variación genética también se encuentra ligada al sistema de reproducción, por ejemplo, en las especies con heterocruza predominante la diversidad se reduce en menor cantidad en comparación con las especies con autofecundación como resultado de la endogamia (Gepts, 2014).

Una consecuencia evolutiva de la domesticación es el denominado “costo de la domesticación”, un concepto propuesto por Lu *et al.* (2006) que consiste en que los procesos evolutivos que moldearon la domesticación dejaron huellas “negativas” en los linajes, pues la domesticación implica el aumento en la frecuencia de alelos deletéreos y la disminución de la heterocigosis. Esto ocurre por la baja eficiencia de la selección natural en comparación con la deriva génica en las poblaciones pequeñas durante el cuello de botella de la domesticación, ya que las mutaciones deletéreas que normalmente se eliminarían con tamaños efectivos poblacionales grandes o tasas de recombinación elevadas, no se purgan (Moyers *et al.*, 2018) lo cual es un resultado esperado por la teoría de genética de poblaciones. De tal modo que Peischl y colaboradores (2013) denominaron como “carga de expansión” a la acumulación de variación genética deletérea debida a una posterior expansión poblacional súbita y la migración a larga distancia, que son características de linajes domesticados después de los cuellos de botella sufridos en la domesticación original.

Las plantas anuales son frecuentes en climas áridos mientras que las perennes en climas áridos, semiáridos y húmedos. Los diferentes climas se encuentran asociados a síndromes de domesticación específicos. En las regiones áridas hay menor cambio en los frutos y la estrategia reproductiva, pero la morfología de la semilla tiende a diferir. En las plantas perennes la domesticación puede estar vinculada a la difusión de técnicas de propagación vegetativa y el aumento en la ploidía. Por otro lado, en los tubérculos y cultivos frutales perennes se presentan menos cambios asociados a los síndromes de domesticación que en los cultivos anuales de semilla, mientras que la domesticación suele ser más lenta ya que las generaciones sexuales están acotadas a cierto período de tiempo (Meyer *et al.*, 2012).

Los cuellos de botella asociados a la domesticación parecen ser menos severos en plantas perennes que en las anuales. De acuerdo con algunos estudios, la variación génica después de los cuellos de botella en los cultivos de frutas tales como manzana (Cornille *et al.*, 2014), cereza (Mariette *et al.*, 2010) y uva (Myles *et al.*, 2011) presentaron poca o nula pérdida de diversidad. Asociado a lo anterior, hay menor cantidad de cambios fenotípicos durante la domesticación de cultivos frutales perennes que en cultivos anuales (Gaut *et al.*, 2015). Gaut y colaboradores, (2015), se percataron de que la introgresión de las plantas silvestres a los cultivos es más importante en plantas perennes que en las anuales debido a dos razones. La primera razón es la similitud génica que se encuentra en poblaciones silvestres y cultivares, comúnmente en algunas plantas perennes de larga vida, como las aceitunas. La segunda es que la introgresión puede contribuir a la formación de cultivares locales, como sucede en las uvas y aceitunas. Sin embargo, puede resultar difícil discriminar la adaptación local debida a la introgresión de las plantas silvestres a los cultivos y la eventualidad del cruzamiento sexual.

El estudio de la domesticación en plantas desde una perspectiva genómica permite una mejor comprensión de los orígenes de la agricultura, el estudio de la diversidad genómica y los efectos del flujo génico en las distintas regiones genómicas (Gepts, 2014). Se han estudiado muchas especies de origen mesoamericano o mexicano, entre las que destacan el maíz (Arteaga *et al.*, 2016; Stitzer y Ross-Ibarra, 2018; Moreno-Letelier *et al.*, 2021), las calabazas (Barrera-Redondo *et al.*, 2020 y 2021), el frijol (Guerra-García *et al.*, 2017 y 2021) y otros cultivos (Mastretta-Yanes *et*

al, 2019). La mayoría de los estudios previos se han enfocado en describir la estructura genética, en realizar análisis de selección, describir el genoma e identificar las poblaciones a partir de las cuales se lleva a cabo la domesticación. Sin embargo, pocos estudios se han enfocado en examinar los gradientes de domesticación, por ejemplo, el estudio del genoma completo de 295 accesiones de Sudamérica y Mesoamérica de jitomate, permitió determinar que sus valores de diversidad nucleotídica (π) van de los 0.00003 a 0.00071. π en *S. lycopersicum* var. *cerasiforme* (jitomate semi-domesticado), que es de un orden de magnitud menor en comparación con el pariente silvestre *Solanum pimpinellifolium* ($\pi= 0.00066$). El jitomate cultivado *Solanum lycopersicum* var. *lycopersicum* presentó una mayor abundancia de alelos raros en comparación con *cerasiforme* y *pimpinellifolium* (Razifard *et al.* 2020). Además, aún menos estudios se han enfocado la reconstrucción de pangenomas, como es el caso de *Oryza sativa*, donde se estudiaron de 3,010 accesiones asiáticas y se determinó que los pangenomas más grandes reconstruidos y la mayor diversidad genómica se encuentran en el grupo Xian/Indica en comparación con el grupo Geng/Japonica. Es así que se concluye que las diferencias en la distribución ecogeográfica, tales como los ambientes más hostiles en dónde se distribuye Geng/Japonica, en comparación con ambientes menos hostiles y más diversos de Xian/Indica, resultan en la adaptación ambiental que se refleja en las funciones génicas predominantes de cada grupo (Wang *et al.*, 2018).

3. Historia natural, evolutiva, domesticación y genómica del género *Capsicum*

Los chiles pertenecen al género *Capsicum* (del griego *kapto*, “morder” por la sensación de picor), que tuvieron origen en el continente americano (Perry *et al.*, 2007; Kim *et al.*, 2014). *Capsicum* pertenece a la familia Solanaceae y comprende un número de especies que oscila entre las 29 a 36, dependiendo del autor, de las cuales cinco especies han sido domesticadas. Los problemas taxonómicos radican en la exclusión o fusión de especies del género, sumados a la variación intraespecífica y la dificultad en el reconocimiento de los parientes silvestres y las formas cultivadas (Hernández-Verdugo *et al.*, 1999).

Las especies de *Capsicum* generalmente son autocompatibles, la única excepción es *C. cardenasii*, con distribución en Perú y Bolivia (Pickersgill, 1997;

Moscone *et al.*, 2006). Morfológicamente, las especies de *Capsicum* tienen un número variable de flores, que va desde 1 a 3 por nudo dependiendo de la especie. La flor puede ser blanca o púrpura, a veces con el cáliz dentado según la especie. Respecto al color y textura de la semilla puede variar entre amarillo y lisa o negra-rugosa (Hernández-Verdugo *et al.*, 1999). Las flores poseen 5 pétalos y tienen cáliz. Desde la primera ramificación de la planta, en cada nodo hay estructuras florales (Bosland y Votava, 2012).

Son plantas herbáceas o trepadoras, que habitan bajo el dosel arbóreo, que les provee sombra y protección. Por este motivo son susceptibles a la pérdida de cobertura vegetal (Hernández-Verdugo *et al.*, 1999). En las regiones tropicales, las plantas pueden llegar a ser arbustos perennes de 2 metros de altura, y vivir hasta unos 10 años, aunque en climas más fríos las heladas los matan, por lo que solo crecen como anuales (Hernández-Verdugo *et al.*, 1999).

El género *Capsicum* se ha identificado como monofilético mediante datos que incluyen caracteres morfológicos, anatómicos, cariológicos, fitoquímicos y/o de ADN. Por otro lado, se ha reconocido al género *Lycianthes* como taxón hermano de *Capsicum* (Carrizo-García *et al.*, 2016). A partir de una filogenia con los marcadores de plástidos *matK* y el espaciador intergénico *psbA-trnH*, así como los exones 2 y 10 del gen nuclear *waxy* y calibrada con fósiles de la familia Solanaceae, se estimó que la divergencia entre los géneros *Solanum* y *Capsicum* ocurrió ca. 19 MA y entre *Capsicum* y *Lycianthes* ca.13 MA; estas fechas corresponden al Mioceno temprano-medio (Särkinen *et al.*, 2013; Qin *et al.*, 2014). Posteriormente, hace ca. 6 MA (Millones de años) en el Plioceno temprano, pudieron presentarse migraciones desde la región Sudamericana hacia el norte, dando lugar a eventos de especiación y favoreciendo la dispersión de *Capsicum* en las Américas (Carrizo-García *et al.*, 2016). La diversidad actual de especies de *Capsicum*, pudo haber ocurrido hace 2-5 MA, lo cual se relaciona con los cambios ambientales (ciclos glaciales) durante el Cuaternario (Särkinen *et al.*, 2013; Turchetto-Zolet *et al.*, 2013).

La mayoría de las especies del género *Capsicum* se distribuyen en Sudamérica (Pickersgill, 1984; Hernández-Verdugo *et al.*, 1999; Perry *et al.*, 2007; ; Raju *et al.*, 2010; Hernández-Verdugo, 2018). Todas las especies del género *Capsicum* se originaron en

los trópicos de América. Los países que presentan mayor número de especies de *Capsicum* son Brasil (11), Argentina y Bolivia (8 cada uno), Ecuador y Perú (7 cada uno), Colombia (6) y México (4). De acuerdo con estos datos al avanzar en su distribución a Norteamérica, el número de especies va disminuyendo (Pickersgill, 1971; 1984; 1997; Hernández-Verdugo *et al.*, 1999). Tanto los chiles silvestres como los domesticados se dividen en dos grupos que representan linajes evolutivos diferentes: el grupo de las flores blancas y el grupo de las flores púrpuras.

Hay tres teorías del origen del género *Capsicum*. La primera, plantea que al encontrarse un mayor número de especies en Sudamérica, se sugiere que el centro de origen de *Capsicum* fue la región de Bolivia, norte de Argentina y centro-sur de Brasil, que en conjunto estarían situados al oeste del Amazonas (Pickersgill, 1984; Hernández-Verdugo *et al.*, 1999). La segunda, propuesta por McLeod *et al.* (1982; 1983) sugiere que *C. chacoense*, o un ancestro de esta especie, dio origen a las flores blancas y púrpuras de *Capsicum*. De esta forma se discute que el ancestro del género pudo situarse en una “zona núcleo” de Bolivia sur-central con posterior dispersión a los Andes y tierras bajas del Amazonas. El grupo de flores blancas se habría dispersado hacia la cuenca del Amazonas, donde posteriormente surgiría *C. annuum*. Por otro lado, el grupo de las flores púrpuras se habría dispersado a las tierras altas de los Andes, donde se originó *C. cardenasii* y la planta domesticada *C. pubescens*. La tercera teoría, derivada de un estudio exhaustivo de reconstrucción filogenética de Carrizo-García *et al.*, (2016), proponen que la diversificación y expansión de *Capsicum* ocurrió de oeste-noroeste de América del Sur a través de las tierras bajas del Amazonas, hasta el centro-sureste de Brasil, posteriormente regresando al centro y oeste de Sudamérica para finalizar en el norte hasta Centroamérica, sugiriendo una rápida especiación que se ve reflejada en el total de especies del género.

Las plantas del género *Capsicum* se consideran entre las primeras en ser domesticadas en el continente americano y Mesoamérica, junto con el maíz, frijol y calabaza (Hernández-Verdugo *et al.*, 1999; Nuez, 2003; Kraft *et al.*, 2014). Actualmente, las especies de *Capsicum* se encuentran ampliamente cultivadas en todo el mundo, a donde fueron distribuidas después de la conquista española (Long-Solís, 1986; Kraft *et al.*, 2014, Aguirre Hernández y Muñoz-Ocotoero, 2015). Las cinco especies domesticadas del género *Capsicum* son *C. annuum* domesticada en México, *C.*

frutescens domesticada en Costa Rica, *C. chinense* domesticada en tierras de baja altitud del Amazonas, *C. baccatum* domesticada en Bolivia y *C. pubescens* domesticada en los Andes (Hernández-Verdugo *et al.*, 2012; Hernández-Verdugo, 2018). Las cuatro primeras especies pertenecen al grupo de flores blancas, y solo *C. pubescens* ha sido domesticada del grupo de flores púrpura (Hernández-Verdugo *et al.*, 1999; Hernández-Verdugo, 2018).

En *Capsicum*, los rasgos del síndrome de domesticación se pueden apreciar al comparar los frutos de los chiles silvestres, que rara vez sufren una abscisión y crecen erectos, mientras que los chiles domesticados son pendulares y sus frutos si presentan abscisión. En cuanto a las semillas, las de los chiles silvestres presentan latencia y las de los cultivados carecen de ella. La domesticación del chile no sólo afecta a los frutos, sino también su sistema reproductivo, pues debido a la domesticación se ha favorecido el aumento de la autopolinización y las estructuras florales tienen distintos tamaños siendo más grandes en cultivados que en silvestres (Hernández-Verdugo, 2018). En el caso de los chiles cultivados los estilos y las anteras pueden ser del mismo tamaño, haciendo posible la autofecundación, lo cual aumenta la endogamia. Aunque hay que tener en cuenta que si bien hay pérdida del potencial evolutivo en la variación que resulta de la consanguinidad, también hay un beneficio en términos de adaptación local por combinaciones de genes homocigos, sacrificando la variación dada por los polimorfismos dentro de los cuales el sistema genético puede alcanzar un equilibrio (Jain, 1976). En cambio, en los chiles silvestres el estilo es más largo que las anteras, lo que facilita la polinización cruzada (Figura 1). Dichas diferencias de la estructura floral en chiles silvestres y domesticados se pueden deber al efecto de la deriva génica, por el tamaño efectivo pequeño y/o como resultado de la selección artificial al trasladar a las plantas a áreas donde los polinizadores eran escasos (Hernández-Verdugo, 2018).

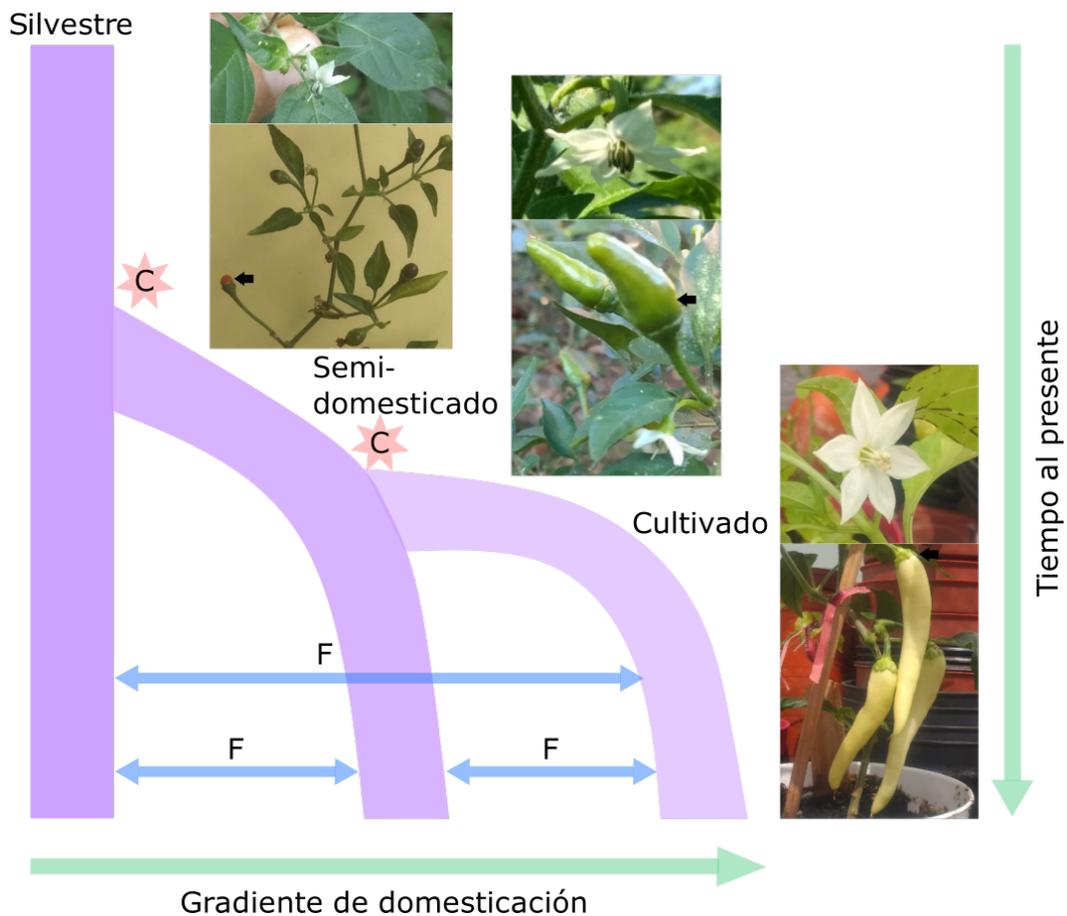


Figura 1. Representación del proceso de domesticación en *Capsicum annuum*. El gradiente de domesticación de *C. annuum* es el siguiente: Silvestre, Semi-domesticado y Cultivado. En las imágenes podemos observar el cambio en la domesticación del chile silvestre al cultivado. Las estrellas con la letra C representan los cuellos de botella característicos de la domesticación y las flechas azules con letra F, señalan el flujo génico entre las poblaciones bajo distintos gradientes de domesticación. El degradado de color morado representa la pérdida de variación genética al pasar por los cuellos de botella (Inspirado de Meyer y Purugganan, 2013).

El genoma de las especies de *Capsicum* (3.07-3.26 gigabases o Gb) es aparentemente unas cuatro veces mayor en tamaño que el del jitomate, que es una Solanaceae cercanamente relacionada con el chile (Qin et al. 2014). Se ha propuesto que el aumento en el tamaño del genoma del chile sucedió recientemente en la historia de la especie, hace unos 300 mil años, debido a una alta inserción de retrotransposones, lo que resultó en un alto porcentaje (81%) de secuencias repetitivas, lo que puede estar ligado a la diversificación del género *Capsicum* (Qin et al. 2014). Sin embargo, hay que tomar en cuenta que los autores del trabajo anterior (Qin et al.,

2014) solo comparan algunas especies de la familia de las solanáceas como la papa, el jitomate y el chile, por lo que sería conveniente tratar de evaluar nuevamente el tiempo de expansión en el tamaño del genoma del chile a partir de un muestreo más exhaustivo. Adicionalmente, existe variación en el número de cromosomas en el género *Capsicum*, por ejemplo *C. annuum*, que es la especie domesticada más emblemática, tiene 12 cromosomas mientras que aproximadamente un tercio de las especies de *Capsicum* cuentan con 13 (Hernández-Verdugo, 2018).

4. El chiltepín: biología, filogenia, domesticación y manejo

C. annuum var. *annuum* es la variedad especie domesticada de mayor importancia económica del género *Capsicum* y también la de mayor variación morfológica. Incluye a la mayoría de los chiles mexicanos. Por lo tanto, la mayoría de las variedades locales y los cultivares de chile corresponden a esta especie. *C. annuum* var. *annuum* crece con facilidad a tierras altas, pero es propenso al ataque de plagas y enfermedades (Pickersgill, 2016; Hernández-Verdugo, 2018), y su pariente silvestre más cercano es *C. annuum* var. *glabriusculum*, el cual recibe diversos nombres en México, incluyendo el de chiltepín. A partir de poblaciones de la variedad *glabriusculum*, *C. annuum* var. *annuum* fue domesticado varias veces (Aguilar-Meléndez *et al.*, 2009).

C. annuum var. *glabriusculum* se distribuye desde el sur de Estados Unidos hasta Colombia o noreste de Sudamérica (Pickersgill, 1991; Hernández-Verdugo *et al.*, 1999; Hernández, 2011; Hernández-Verdugo *et al.*, 2012; Kim *et al.*, 2014). Son plantas herbáceas o trepadoras, las cuales pueden llegar a alcanzar los 5 m. (Hernández-Verdugo *et al.*, 1999; Pickersgill, 2016). Son perennes o con ciclo de vida corto, con una flor por nudo o rara vez dos o tres (Hernández-Verdugo *et al.*, 1999). El cáliz tiene dientes ausentes o muy rudimentarios, con corola blanca, ocasionalmente verde, con hojas glabras (Hernández-Verdugo *et al.*, 1999; Pickersgill, 2016). Cada individuo produce numerosos frutos que son de color verde a negro o púrpura cuando son inmaduros y rojos cuando maduros. Los frutos caen al madurar, son redondeados y erectos, lo que los hace visibles para las aves, sus dispersoras, con un tamaño de 3-9 mm de diámetro y de 4-10 mm de largo, con 1-40 semillas por fruto (Pickersgill, 2016). Los frutos maduros del chiltepín son consumidos por aves, las cuales posteriormente

dispersan las semillas, pues a diferencia de los mamíferos ellas no tienen los receptores de dolor para percibir el picor de los chiles (Pickersgill, 2016). Por el contrario, a los mamíferos pequeños se les considera como depredadores de semillas, ya que se pierde por completo la capacidad de germinación de las semillas cuando las consumen, a diferencia de las aves que las dejan intactas, es por ello la pungencia de los chiles se considera como una forma de protección contra la depredación (Pickersgill, 2016).

El chiltepín habita en sitios no perturbados de la selva baja, en zonas montañosas cercanas a ríos y cañones, huertos o en sitios perturbados como potreros. Comúnmente, el chiltepín crece junto a árboles más grandes que le sirven de protección como plantas nodriza. Rara vez se encuentra por arriba de los 1000 msnm, aunque se ha registrado a más de 2000 msnm en la Sierra Gorda, Querétaro (Hernández-Verdugo *et al.*, 1999, 2012; Hernández-Verdugo, 2018; Pickersgill, 2016). Se ha registrado en el oeste de México desde Baja California hasta Chiapas y en el este de Tamaulipas a la península de Yucatán y en el sur de Estados Unidos en Arizona y Texas (Lizarde *et al.*, 2011; Pickersgill, 2016). Se han encontrado poblaciones de chiltepín con genes de resistencia a virus patógenos, hongos y nemátodos (Lizarde *et al.*, 2011; Pickersgill, 2016). En relación con lo anterior, se encontró que el riesgo de padecer la enfermedad estaba relacionado con el nivel de manejo humano, que a su vez estaba relacionado con la diversidad genética de la especie (Pagán *et al.*, 2012; Kim *et al.*, 2014).

A pesar de ser una especie silvestre, *C. annuum* var. *glabriusculum* es aprovechada a través de la recolección de unas 50 toneladas por año, en el Norte y centro de México (Votava *et al.*, 2002). Esto supone que se trata de un recurso genético importante para la industria alimenticia. Sin embargo, debido a la pérdida acelerada del hábitat del chiltepín y la germinación lenta e irregular de las semillas, la colocan como una planta silvestre de difícil cultivo. Además, los agricultores cortan ramas enteras de las plantas para evitar la depredación de las aves, para el posterior resguardo en sus casas, así contribuyendo así con el daño a las plantas silvestres (Laborde-Cancino y Pozo-Campadónico, 1982; Pickersgill, 2016).

Los esfuerzos por entender la diversidad del chiltepín en su área de distribución son varios, los cuales han utilizado marcadores moleculares diversos (Tabla 1). Por un lado, los estudios con marcadores moleculares han permitido encontrar: (i) el área

conocida como Mesa Central sea una barrera geográfica para el flujo génico de los chiles (Loaiza-Figueroa *et al.*, 1989; Aguilar-Meléndez *et al.*, 2009); (ii) evidencia de la elevada diversidad en las poblaciones del sur de México, especialmente en la península de Yucatán (Votava *et al.*, 2002; Aguilar-Meléndez *et al.*, 2009); (iii) existencia de flujo génico entre distintas variedades de chile (Guzmán *et al.*, 2005); (iv) mayor variación genética y riqueza alélica en las poblaciones silvestres de chiltepín, y menor en poblaciones semi-silvestres y cultivadas (Aguilar-Meléndez *et al.*, 2009); y mayor presencia de plasticidad fenotípica en chiltepines del noroeste de México que fueron expuestas a diferente disponibilidad de luz (Hernández-Verdugo *et al.*, 2015). Por otro lado, se han encontrado poblaciones de chiltepín con genes de resistencia a virus patógenos, hongos y nemátodos. Se encontró que el riesgo de padecer la enfermedad estaba relacionado con el nivel de manejo humano, que a su vez estaba relacionado con la diversidad genética de la especie (Pagán *et al.*, 2012; Kim *et al.*, 2014).

Tabla 1. Estudios genéticos de *Capsicum* y sus principales aportes

Especie	Marcador	Principales aportes	Referencia
186 accesiones de <i>Capsicum</i> en México, domesticadas, semi domesticadas y silvestres	Polimorfismo enzimático (aconitasa, isocitrato deshidrogenasa, malato deshidrogenasa, peroxidasa, entre otras)	La diversidad genética total para <i>Capsicum</i> : domesticadas=0.077 y silvestres=0.282. Se encontró diferenciación genética significativa entre poblaciones que difieren geográficamente. La estructura de las poblaciones de <i>Capsicum</i> en México consiste en genotipos predominantemente homocigos debido al sistema de reproducción de autopolinización y cuellos de botella en las poblaciones	Loaiza-Figueroa <i>et al.</i> , 1989
43 muestras de <i>C. annuum</i> var. <i>glabriusculum</i> de Arizona E.U., <i>in situ</i> y de México y Guatemala <i>ex situ</i>	RAPD (Random Amplified Polymorphic DNA)	Las poblaciones de <i>Capsicum</i> geográficamente cercanas son genéticamente más similares entre sí. Las poblaciones de Coronado National Forest en Arizona, pueden ser una población de chiles disyunta con fuertes presiones selectivas y/o deriva después de un efecto fundador	Votava <i>et al.</i> , 2002
74 muestras de <i>Capsicum</i> spp. de Guatemala: <i>C. annuum</i> , <i>C. chinense</i> , <i>C. frutescens</i> , <i>C. pubescens</i> , semicultivadas y cultivadas	AFLPs (Aplified Frafment Length Polymorphims)	Diversidad genética total H=0.280, diversidad genética de muestras semicultivadas H=2.48 y cultivadas H=2.86. La diversidad genética de los traspatios de Alta Verapaz es representativa del total de la diversidad génica de Guatemala	Guzmán <i>et al.</i> , 2005
<i>C. annuum</i> , número de muestras: 58 semisilvestres 22 domesticadas, en México	Loci nucleares, DHn, G3pdh y Waxy	Heterocigosidad Dhn=0%, G3pdh=25% y Waxy=10%. Diversidad promedio de los 3 loci en chiles semisilvestres $\theta_w=2.80 \times 10^{-3}$ y $\theta_w=2.55 \times 10^{-3}$ para domesticados. La diversidad total de haplotipos en chiles semisilvestres es del 74% y en domesticados del 52%, por lo tanto, la variación en muestras cultivadas es un subconjunto de la diversidad total de los chiles silvestres.	Aguilar-Meléndez <i>et al.</i> , 2009

661 individuos de 27 poblaciones de <i>C. annuum</i> var. <i>glabriusculum</i> silvestres y arvenses en México	9 microsatélites altamente polimórficos	La mayor variación genética se encontró en las poblaciones de Yucatán y la menor en poblaciones de Sonora. $F_{ST} = 0.430$ para todas las poblaciones y la prueba de Mantel mostró correlación entre la distancia genética y geográfica ($r=0.652$).	González-Jara <i>et al.</i> , 2011
--	---	---	------------------------------------

Además de estudios con marcadores moleculares, se han realizado diversos trabajos para analizar el genoma del chile. Estos trabajos incluyen la secuenciación genómica del chiltepín, que presenta un genoma de 3.07 Gb; la secuenciación de la variedad Zunla-1, que tiene una longitud total de 3.26 Gb (Qin *et al.*, 2014); y la de una variante del estado de Morelos, cuyo tamaño es de 3.48 Gb (Kim *et al.*, 2014). Estos estudios muestran que el tamaño del genoma del chile es de aproximadamente entre 3 y 3.5 Gb, influenciado por la acumulación de secuencias repetitivas en regiones heterocromáticas y eucromáticas (Kim *et al.*, 2014), así como a la presencia de una gran cantidad de elementos transponibles. A pesar de que el tamaño del genoma de *C. annuum* es similar al de la papa y el jitomate (Kim *et al.*, 2014; Qin *et al.*, 2014), durante la domesticación del chile hubo eventos de translocación de regiones cromosómicas que diferencian a los parientes silvestres de los cultivos (Qin *et al.*, 2014).

5. Conservación de la diversidad genética en *Capsicum*

Una de las principales razones para estudiar a los parientes silvestres de los cultivos, es que constituyen un reservorio genético que pueden aportar a solucionar problemas agrícolas, proporcionando diversidad genética para enfrentar problemas como plagas y enfermedades (Hernández-Verdugo *et al.*, 1998; Hernández-Verdugo, 2018).

Pero, ¿Por qué conservar la diversidad genética? Spielman y colaboradores (2004) proponen que a menor diversidad genética, aumenta la probabilidad de extinción de las especies, ya que la diversidad es la “base del cambio evolutivo” para la adaptación al cambio climático, a nuevos hábitats y para las cambiantes interacciones bióticas. A los parientes silvestres de los cultivos se les considera como reservas de

variación genética, al ser organismos de los que derivaron los organismos cultivados, con un gran potencial para contribuir con rasgos para la mejora de los cultivos para aumentar el rendimiento, la estabilidad y o la resistencias abióticas y bióticas (Maxted, *et al.*, 2006, 2012). En la gran mayoría de los parientes silvestres no hay información sobre la diversidad genética.

Un reto importante para la conservación de la biodiversidad es que generalmente no se comprende su valor actual ni su potencial para el futuro (Maxted *et al.*, 2012). Maxted y colaboradores (2012), estimaron que las 2500 especies de plantas que se cultivan en el mundo cuentan con aproximadamente 50,000 parientes silvestres (Lyuri, 2008), de los cuales 800 son de máxima prioridad para la conservación activa inmediata, debido a su relación con los cultivos alimentarios. Se espera que para el 2055, cerca del 16-22% de los parientes silvestres de cultivos se extingan, y aunque no se extinguieran, muchos de estos parientes silvestres con distribución fragmentada estarían gravemente amenazados (Jarvis *et al.*, 2008). Si bien algunas poblaciones de los parientes silvestres se encuentran resguardadas en áreas protegidas, especialmente cuando son declaradas de interés nacional para los países (Maxted *et al.*, 2012), hay que tomar en cuenta que el cambio climático afectará las condiciones agroambientales de las zonas en las que las especies se distribuyen.

Una posible herramienta de conservación es la llamada “translocación genética” o “migración asistida”, que se lleva a cabo en el área de distribución de la especie, en la cual puede haber flujo génico que resulte en la mejora o conservación de la adaptación de la especie. Esta herramienta puede ser útil en especies que pasaron por fragmentación del hábitat o en especies con un amplio rango de distribución con diferencias clinales que tengan bajo rendimiento, permitiendo en ambos casos, lograr una mayor adaptabilidad y resiliencia evolutiva (Sgrò *et al.*, 2011). En este sentido, una perspectiva evolutiva debe incorporarse al planear políticas de conservación y se vuelve especialmente relevante al experimentar el cambio climático y la fragmentación del hábitat (Sgrò *et al.*, 2011). En este sentido, las políticas de conservación deben incorporar una perspectiva evolutiva, y se vuelven de particular relevancia al hacer frente a problemáticas como el cambio climático y la ofragmentación del hábitat (Sgrò *et al.*, 2011).

Los patrones de distribución geográfica ayudan a entender cómo ocurren los cambios de la riqueza vegetal, para así hacer planes de manejo para su conservación (Hernández-Verdugo, 1998, 2018). Las principales formas de conservar la diversidad genética pueden dividirse en conservación *in situ* y conservación *ex situ* (Sgrò *et al.*, 2011; Hoban *et al.*, 2020). Por un lado, en la conservación *in situ* de la agrobiodiversidad, las plantas se conservan en su distribución original y continúan allí evolucionando. Para el caso de las especies silvestres, esto normalmente ocurre en sitios geográficamente determinados, aislados de la fragmentación del hábitat y bajo el supuesto de un clima estable, generalmente alejados de las tierras agrícolas o ciudades (Sgrò *et al.*, 2011; Hoban *et al.*, 2020). Sin embargo, algunos parientes silvestres también crecen en vegetación secundaria o zonas perturbadas, como es el caso de los algodones (Casas *et al.*, 2016; Hernández-Terán, 2020), por lo que su conservación no depende estrictamente solo de la vegetación primaria (Goettsch *et al.*, 2021).

Para el caso de las especies bajo manejo, la conservación *in situ* se refiere a que dichas plantas continúen siendo sembradas o manejadas, por lo que los campesinos y los pueblos indígenas son parte fundamental. En otras palabras, se conserva la diversidad genética al fomentar la disponibilidad de plantas útiles y protegerlas mejorando los ambientes locales. Por lo tanto, la conservación *in situ* puede conllevar procesos de selección artificial y domesticación incipiente. Tan solo en Mesoamérica se tiene registro de 600-700 especies botánicas bajo manejo *in situ* (Casas *et al.*, 2007; Casas y Parra, 2007).

Asimismo, existe la agricultura a pequeña escala, que se caracteriza en que parte de la producción se destina al autoconsumo, a diferencia del modelo de agricultura convencional que consiste en la modernización agrícola para maximizar la producción, lo que ha implicado impactos negativos en los ecosistemas y amenazas a la sostenibilidad. Específicamente, los pequeños agricultores incorporan variedades agrícolas modernas y tradicionales, lo que se traduce en niveles significativos de biodiversidad en las parcelas campesinas. Por ello, que la biodiversidad en las parcelas se relaciona con la resiliencia y el aumento de la seguridad alimentaria, manteniendo: cultivos tradicionales, plantas silvestres y especies animales (Altieri y Toledo, 2011; Altieri *et al.*, 2012; Chappell *et al.*, 2013). La agricultura a pequeña escala, o campesina contribuye a mantener y generar la diversidad genética, pues aunque las parcelas son

pequeñas, en total constituyen millones de hectáreas, donde miles de familias campesinas realizan el proceso de selección de semillas en ambientes y condiciones culturales distintas (Bellón *et al.*, 2018, Bellon *et al.*, 2021). Así, la soberanía alimentaria representa una nueva perspectiva hacia la conservación de la biodiversidad en asociación con científicos y economistas, por mencionar algunos sectores, para que los agricultores puedan obtener un mayor acceso a conocimiento agroecológico, así como a tierras, semillas y alternativas agroecológicas que se adapten a las necesidades de los productores (Altieri y Toledo, 2011; Altieri *et al.*, 2012).

Por otro lado, en la conservación *ex situ*, el germoplasma de las plantas que se colecta en su área de distribución original y se almacena en bancos de semillas, los cuales se centran en la conservación de la mayor cantidad de especies posibles y no en el rescate de la diversidad genética *per se* (Hoban *et al.*, 2020). Sin embargo, Walters y colaboradores (2013) advierten que entre el 10 a 20% de las especies de plantas (la mayoría tropicales), no pueden ser almacenadas en bancos de semillas. De tal forma que el éxito de la conservación *in situ* y *ex situ* dependerán de la planeación y las gestiones que conduzcan a favorecer la diversidad genética de las poblaciones (Sgrò *et al.*, 2011; Maxted *et al.*, 2012; ; Hoban *et al.*, 2020). Jardón-Barbolla (2015, 2016) analiza los esfuerzos de la conservación *ex situ* y la dependencia de unos cuantos cultivos que sostienen la alimentación humana. En este sentido, es importante considerar el esfuerzo de los agricultores y agricultoras locales que durante generaciones dieron origen a las variedades locales que hoy en día forman parte de esfuerzos de conservación *ex situ*, como sucede en la Bóveda de Semillas de Svalbard. Jardón-Barbolla (2016) nos muestra un panorama al que llama “cosificación de la diversidad genética”, en dónde enfatiza la necesidad de reflexionar sobre las causas de pérdida de diversidad antes de llegar a poner nuestras esperanzas en la conservación *ex situ*.

Si bien hoy en día tenemos noción de la importancia de la diversidad y conservación de las especies, Nicolai Vavilov fue un pionero en el tema, dando especial importancia a los parientes silvestres de los cultivos. Vavilov contaba con la inquietud de involucrar el conocimiento genético y evolutivo en el mejoramiento de los cultivos y con un fuerte compromiso social a evitar la hambruna en el mundo, mediante el estudio de la agrobiodiversidad (Vavilov, 1920, 1922, 1951). Adicionalmente, Vavilov tenía la

visión de entender la lingüística, la evidencia arqueológica y la historia como parte integral de la labor botánica para entender el origen histórico de la agrobiodiversidad como producto de la sociedad y la naturaleza (Jardón-Barbolla, 2021). Vavilov sentó las bases por el interés en la conservación de la agrobiodiversidad. Actualmente hay una creciente motivación a nivel mundial en desarrollar estrategias de conservación, entre las cuales se encuentra el Convenio sobre la Biodiversidad Biológica (CBD) que en 2010 puso en marcha las metas de Aichi para ser cumplidas en 2020. Específicamente, el objetivo 13 se centra en preservar la diversidad genética de plantas cultivadas y sus parientes silvestres (Convention on Biological Diversity (CBD), 2011).

En México los esfuerzos para la planificación de programas de conservación de la diversidad genética de cultivos y sus parientes silvestres buscan integrar los procesos evolutivos actuales, además de involucrar a los grupos de investigación con agricultores para proponer programas que favorezcan el desarrollo agrícola sustentable (Mastretta-Yanes *et al.*, 2018). Una metodología para incluir a cualquier taxón sin datos genéticos previos y asegurar su variación genética, son los *proxies* de diversidad genética (PDG). Los PDG son una representación de la diversidad genética con un enfoque en la integración de variables ambientales y filogeográficas (variables ecológicas y genéticas). Por tanto, se emplea la variabilidad ambiental y la diferenciación histórica guiada por patrones filogeográficos, lo cual es una opción cuando los estudios genéticos son inviables o los presupuestos para su estudio limitados (Tobón-Niedfeldt *et al.*, 2021).

La posición geográfica de México y las complejas historias climáticas y geológicas asociadas (sistemas montañosos tales como sierras y cinturones que van de los 1,000-5,000 m.s.n.m, además de desiertos), convierten a México en un país idóneo para el estudio de la evolución y diversificación de la biodiversidad vegetal (Sosa *et al.*, 2018). Adicionalmente la evolución de los cultivos guiada por los seres humanos, las prácticas culturales, sistemas de manejo ejercen un papel relevante como es el caso de los chiles (Jardón-Barbolla, comunicación personal, septiembre 2021). En relación con lo anterior, el trabajo genómico más reciente que incluye variedades locales y distintas especies de chiles fue realizado por Taitano y colaboradores (2018), quienes estudiaron accesiones de variedades locales mexicanas y accesiones globales de chiles, con el objetivo de caracterizar la diversidad genética para estudiar alelos

potencialmente útiles para el mejoramiento de cultivares modernos. Encontraron diferencias genéticas entre las variedades locales de chiles, las cuales presentaban diferentes usos, además de que la diversidad genética se relacionaba estrechamente con las técnicas de cultivo empleadas en las variedades locales. Los autores del estudio sugieren que, si bien hay diferencias genómicas en las variedades locales, éstas no se detectan al incluir la colección global. Adicionalmente encontraron señales de selección en las regiones cromosómicas adjuntas que se correlacionan con la morfología de la fruta.

El siguiente paso en el estudio de la diversidad genética de los chiles silvestres es estudiar su variación genética a nivel poblacional. Actualmente es posible generar datos genéticos de decenas o cientos de muestras mediante métodos de secuenciación reducida de genomas, mismos que pueden ensamblarse de *novο* o usando un genoma de referencia (Andrews *et al.*, 2016). Sin embargo, dichos métodos requieren DNA de alta calidad (Andrews *et al.*, 2016), lo que para parientes silvestres de cultivos puede representar un reto.

6. Retos y limitantes de estudios moleculares y genéticos en parientes silvestres

Los Polimorfismos Nucleótido Único (SNPs) son variaciones en solo una par de bases en una posición determinada del genoma que se presentan una porción lo suficientemente grande de la población (1% o más). Existen métodos de “representación reducida” de genomas que permiten obtener miles y hasta millones de SNPs. Estos métodos consisten en secuenciar sólo una fracción del genoma que preliminarmente ha sido digerido por enzimas de restricción, las cuales en sitios específicos del genoma (Andrews *et al.*, 2016). Entre los métodos de secuenciación reducida de genoma, se encuentran la secuenciación asociada a sitios de restricción o RADseq (*Restriction site-associated DNA sequencing*) el ddRad (*Double digest restriction-site associated*) y el GBS (*Genotyping by sequencing*) (Narum *et al.*, 2013; Andrews *et al.*, 2016). Todos estos métodos requieren una cantidad mínima de DNA de 20-50 nanogramos por microlitro de muestra, además de que el DNA debe estar poco fragmentado para que la extracción se considere lo suficientemente buena (Universidad de Wisconsin, Centro de Biotecnología, 2021).

La obtención de DNA de buena calidad es a veces un reto en parientes silvestres, ya que no presentan ningún síndrome de domesticación, por tal motivo su germinación y crecimiento son lentos (Casas *et al.*, 2007). Esto representa problemas, ya que el periodo de colecta en campo se acota a unos cuantos meses (para colectar hojas) o las semillas se tienen que germinar (Hernández-Verdugo, 2018). Casas y colaboradores (2007) encontraron en chiles un gradiente en el tiempo de germinación en individuos cultivados, semi-silvestres y silvestres, específicamente en plantas perennes de vida larga, con polinización cruzada, con y sin propagación vegetativa, plantas herbáceas con y sin propagación vegetativa, puesto que la intensidad en el manejo se encuentra vinculado con la germinación. Es así que generalmente a mayor grado de domesticación, es menor el tiempo de germinación.

Específicamente, en *C. annuum* el rasgo más claro de la domesticación es cambio en la velocidad de germinación, siendo más rápido en los chiles domesticados, tal vez debido al hecho de que las accesiones de frutos más pequeños pueden haber sido objeto de menos selección (Pickersgill, 2016; Bernau *et al.*, 2020). Bernau y colaboradores (2020), estudiaron las condiciones de germinación de semillas de chiles silvestres y variedades locales en respuesta al estrés hídrico. Sus resultados mostraron que la germinación reducida en condiciones de baja humedad puede ser adaptativa al nivel poblacional; la germinación más lenta y reducida bajo estrés por sequía puede ser un mecanismo de evitación adaptativo. Es así como el tiempo prolongado de germinación en parientes silvestres, el crecimiento lento de las plantas, la tasa de germinación baja y en diferentes tiempos, conllevan a mayores dificultades para la extracción del material genético (Casas *et al.*, 2007). En cuanto a los tejidos utilizados en la extracción de DNA, específicamente en *Capsicum*, se han reportado el uso de tejido foliar realizado de hojas pues se considera que la capsaicina interfiere el proceso de extracción, por lo tanto, no se ha probado la extracción de otras partes de la planta (Hernández-Verdugo, 2018).

Si se cuenta con DNA de estudios previos que no implementaron métodos genómicos, resulta desafiante obtener DNA genómico apto para su secuenciación, ya que el DNA se degrada con el paso del tiempo, pudiendo tener una concentración menor a 20 ng/ml requerida para GBS y/o el volumen de muestra no ser suficiente. Una opción para recuperar muestras de las cuales se cuenta con muy poco DNA, es aplicar

métodos de amplificación de genoma, como el kit REPLI-g de Qiagen (Qiagen 2021a; Qiagen 2021b) que utiliza la amplificación del genoma isotérmico, mejor conocida como amplificación de desplazamiento múltiple con la polimerasa Phi 29. Como resultado se obtiene una amplificación imparcial de genomas completos. El protocolo de REPLI-g consiste en la unión de hexámeros aleatorios al DNA desnaturalizado, posteriormente ocurre la síntesis por desplazamiento de cadena a temperatura constante. La DNA polimerasa no se disgrega de la cadena de DNA, por lo tanto, puede generar fragmentos de hasta 100kb, además de que la polimerasa tiene actividad exonucleasa 3' → 5' (Qiagen 2021a; Qiagen 2021b). Sin embargo se ha reportado (Ahsanuddin *et al.*, 2017) que el kit REPLI-g produce artefactos por dos vías: amplifica contaminantes de laboratorio, y amplifica de forma no aleatoria de las muestras de DNA. REPLI-g se ha empleado en estudios de metagenómicos de baja biomasa, o estudios de microbiomas como abejas (Herb *et al.*, 2012), en sangre humana de neonatos (Mendell *et al.*, 2012), en virus en pangolines (Liu *et al.*, 2019), en bacterias marinas no cultivadas (Stepanauskas y Sieracki, 2007), o en células individuales de sangre (Qn *et al.*, 2019), por mencionar algunos. No se había probado en plantas con genomas grandes, como es el caso del chile, por lo que es incierto si puede ser útil para recuperar muestras con baja concentración de DNA o si producirá artefactos.

Por ejemplo, el tener muestras con diferentes estatus de domesticación y tejidos de *Capsicum*, así como diversos métodos de extracción y recuperación, desconocemos si el DNA extraído sea de la misma calidad y si las digestiones son inconsistentes, lo cual puede llegar a ser problemático al hablar de datos perdidos y exceso de homocigosis (O'Leary *et al.*, 2018).

Otro de los grandes retos al estudiar los parientes silvestres, en contraste con las plantas cultivadas, es que las plantas silvestres generalmente tienen mayor diversidad genética, pero para realizar estudios genómicos los genomas de referencia disponibles son en su mayoría de variedades domesticadas. Tal es el caso de los parientes silvestres de cacahuates (Zhuang *et al.*, 2019), duraznos (Li *et al.*, 2019a; Li *et al.*, 2019b), jitomates (Lin *et al.*, 2014), berenjenas (Acquadro *et al.*, 2017), y vides (Marrano *et al.*, 2018). Actualmente, el estudio genómico de parientes silvestres se está inclinando por el ensamblado *de novo* como en calabazas pipianas (Barrera-Redondo *et al.*, 2020), vides (Badouin *et al.*, 2020), frijol de soya (Xie *et al.*, 2019) y jitomate

(Razali *et al.*, 2018). Asimismo, se ha optado por el ensamblado de pan-genomas, para obtener la mayor cantidad de variación en parientes silvestres y cultivos de arroz (Zhao *et al.*, 2018), y en parientes silvestres y cultivos de frijol de soya (Liu *et al.*, 2020). El genoma de referencia que se ocupa para mapear datos de secuenciación reducida de genomas es muy relevante, ya que permite ahorrar tiempo y mejorar la anotación ya que muchos genomas de referencia cuentan con la anotación de genes que sirven para proponer hipótesis, dependiendo del enfoque del trabajo.

El presente estudio busca aportar información acerca de la diversidad genómica de *C. annuum* var. *glabriusculum*, uno de los parientes silvestres con mayor relevancia para el resguardo de la seguridad alimentaria de uno de los cultivos de gran importancia económica y cultural, como lo es el chile. Como un primer paso para poder realizar estudios genómicos a escala poblacional, examinamos si es posible obtener DNA de calidad para secuenciación reducida de genoma mediante la técnica de genotipado por secuenciación (GBS por sus siglas en inglés; Andrews *et al.*, 2016) a partir de DNA de diversas fuentes.

Hipótesis

- I) Se espera mayor diferenciación genética entre las poblaciones de *C. annuum* var. *glabriusculum* silvestres y/o semi-silvestres respecto a las cultivadas, debido a que éstas últimas han perdido variabilidad genética por los procesos de domesticación ejercidos por el ser humano.

- II) La distribución espacial de los chiles favorecerá el agrupamiento de las poblaciones más cercanas geográficamente entre sí.

- III) La fuente de la que se obtuvo el DNA para la secuenciación: tejido seco, congelado, fresco; fruto seco y congelado; re-amplificación de DNA (REPLI-g), pueden influir la calidad de la secuenciación y causar artefactos que dificulten la determinación precisa de la de la estructura genética de *C. annuum* var. *glabriusculum*.

Objetivo general

Determinar cómo el grado de domesticación y la cercanía geográfica influyen en la diferenciación genómica de las poblaciones de chile silvestre y semi-domesticado a lo largo de su área de distribución en México.

Objetivos particulares

I.- Determinar la estructura genética a nivel genómico de *C. annuum* var. *glabriusculum* en México.

II.- Inferir si el grado de domesticación influye en la estructuración genómica de *C. annuum*.

III.- Dados los retos para la obtención de DNA de calidad en parientes silvestres, determinar si fuentes alternativas de tejido y la re-amplificación de DNA causan artefactos en la obtención de SNPs, que a su vez afecten la estimación de la estructura y variación en *C. annuum*.

Metodología general

Materiales biológicos y colecta

La mayoría de las colectas necesarias para este proyecto ya se habían realizado previamente (Pagán *et al.*, 2009; González-Jara *et al.*, 2011). Así, ya contábamos con DNA de baja concentración, al que tratamos con el kit de amplificación genómica total (REPLI-g de Qiagen), para aumentar su concentración.

Las muestras de *C. annuum* se complementaron mediante colectas en Querétaro, Chiapas y Oaxaca (Anexo 3). Debido a que algunas de las muestras de las localidades anteriormente mencionadas carecían de hoja y teníamos una colección de semillas, decidimos germinarlas probando las siguientes condiciones: 1) consistió en 200 microlitros de ácido giberélico a una concentración de 250 mg/L y 24 horas de reposo; 2) en la escarificación con ácido sulfúrico puro por 5-10 segundos, mas 24 horas

de incubación con ácido giberélico, 3) la incubación a 35°C en cajas de Petri, hasta que la semilla germinara. En todos los casos las semillas fueron puestas en charolas de germinación en el invernadero del Instituto de Ecología, UNAM.

Se utilizaron las siguientes fuentes de DNA: tejido seco, hoja congelada y fresca, fruto seco y congelado, REPLI-g y herbario (Tabla 2). Además, se consideraron tres distintos grados de domesticación: el primero o *muestras silvestres* sin modificaciones fenotípicas asociadas a la domesticación y que se encuentran en un ambiente silvestre, segundo *muestras semisilvestres* que definimos con modificaciones muy incipientes o nulas asociadas a la domesticación, pero que se encuentran fuera del ambiente silvestre (por ejemplo, traspatio) y por último *muestras domesticadas* o cuando ya presentan un síndrome de domesticación claro. El número de muestras de cada fuente de DNA y grado de domesticación se muestra en la Tabla 3, dando como resultado un total de sesenta y siete individuos secuenciados.

La calidad genómica de las muestras se evaluó con tres métodos. Primero con geles de agarosa al 0.8%, en TBE 1X a 90V por 40 minutos Segundo, con el equipo nanodrop, conservando para el análisis aquellas muestras con concentración mayor o igual a 30 ng/uL y absorbancia A260/280 (que indica la cantidad de RNA y contaminación con impurezas orgánicas e inorgánicas) de 1.80-2.0. El tercer método implicó una lectura en Qubit de concentraciones de al menos 20 ng/uL. Usando ggplot 2 (Wickham, 2016) en R (R Core Team, 2020), construimos un mapa con las coordenadas geográficas de cada muestra y las provincias biogeográficas asociadas al territorio mexicano.

Tabla 2. Fuentes de DNA de *C. annuum* que se utilizaron para la extracción de DNA

Fuente del DNA	Características de la fuente del DNA	Tiempo de almacenamiento
Tejido seco	Abarcó la hoja y el pecíolo, este tejido se secó previamente en gel de sílice y se mantuvo almacenado a temperatura ambiente	Tres a cuatro meses
Hoja congelada	Tejido almacenado a una temperatura de -20°C por lo que al momento de la extracción la hoja estaba hidratada	Un año - año y medio
Fruto seco	Tejido de frutos almacenados a temperatura ambiente en sobres de papel manila (solo se extrajo el pericarpo y la placenta)	Año y medio - dos años
Fruto congelado	Tejidos de frutos almacenados a -20°C y estaban hidratados. Solo se extrajo el pericarpo y la placenta)	Un año - año y medio
Hoja fresca	Se cortó al momento de la extracción de las hojas y pecíolo los chiles, previamente crecidos en invernadero	No aplica
Tejido de herbario	Se escogieron las hojas más grandes de herbario y su extracción fue posterior a la donación	Once años
Re-amplificación de DNA previamente extraído (Kit REPLI-g)	Estas muestras provienen del trabajo de González-Jara <i>et al.</i> , 2011, que fueron re-extraídas de hoja seca por la Dra. Alejandra Moreno en 2018. El DNA se almacenó a -20°C y se utilizó el kit REPLI-g para aumentar la concentración del material genómico	Año y medio

Tabla 3. Estatus de domesticación, fuente del DNA y número de muestras para el presente estudio

Estatus de domesticación	Fuente del DNA	Número de muestras
Silvestre n=37	REPLI-g	14
	Tejido seco	17
	Tejido fresco	2
	Hoja congelada	2
	Fruto seco	1
	Herbario	1
Semi-silvestres n=23	REPLI-g	3
	Tejido fresco	5
	Hoja congelada	11
	Fruto seco	2
	Fruto congelado	2
Domesticado n=7	REPLI-g	1
	Tejido fresco	5
	Fruto seco	1

Secuenciación

Se obtuvieron datos genómicos de las muestras mediante GBS (*Genotyping By Sequencing*), que es un método de representación reducida del genoma. GBS permite la identificación de miles a millones de SNPs a lo largo del genoma de cientos de muestras mediante la secuenciación de fragmentos de DNA cercanos a enzimas de restricción (Narum *et al*, 2013).

La construcción y secuenciación de las bibliotecas GBS (Narum *et al*, 2013), se realizó con la enzima de corte frecuente ApeKI, en el Centro de Biotecnología de la Universidad de Wisconsin-Madison, en la plataforma NovaSeq 6000, en una línea compartida. Se obtuvieron 250 millones de lecturas por la placa o bien 2.6 millones de lecturas por muestra, a partir de extremos pareados (*paired-end*, 2x150).

Análisis bioinformáticos

El flujo de trabajo se llevó a cabo en cuatro puntos, como se muestra en la Figura 2 y se explica detalladamente a continuación. Asimismo, todo el trabajo bioinformático de esta tesis está en la plataforma GitHub en el repositorio de código https://github.com/gabymandrade/Capsicum_GBS.

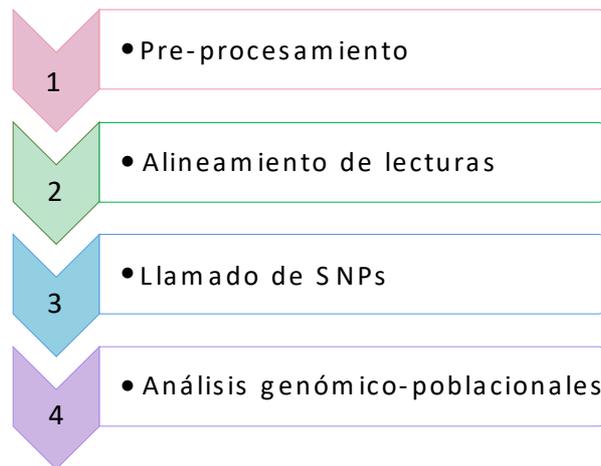


FIGURA 2. Resumen del flujo de trabajo que se dividió en cuatro puntos principales. Primero, el pre-procesamiento consistió en la eliminación de adaptadores y el demultiplexeo. Segundo, el alineamiento de las lecturas contra un genoma de referencia y el indexado del genoma. Tercero, llamado de SNPs y el ensamblado de los loci. Cuarto, los análisis genómico poblacionales, con los cálculos de estadísticos a nivel poblacional y gráficas de estructura genómica.

1.-Preprocesamiento

La calidad de las lecturas o reads de todas las muestras antes de demultiplexear se revisó en el informe MultiQC. La limpieza de las lecturas se realizó con Trimmomatic v. 3 (Bolger *et al.*, 2014), en modo *paired-end* y con ILLUMINACLIP TruSeq2-PE.fa:2:30:10, con una longitud mínima de corte de lectura de 140 (MINLEN:140). El “demultiplexeo” se llevó a cabo con Process radtags en Stacks v. 2.2 (Catchen *et al.*, 2013), en modo *paired-end*, contemplando que la enzima utilizada en la secuenciación fue ApeKI, y corriendo el script con los *flags* `--inline_null -r -c -q` (Fig. 3). Después del demultiplexeo y eliminación de adaptadores, se analizó la calidad de las muestras de forma independiente con el programa FastQC v0.10.1 (Andrews, 2010). Posteriormente graficamos en R v. 4.0.3 (R Core Team, 2020) con la paquetería ggplot2 (Wickham, 2016) las lecturas por muestra para cada una de las poblaciones estudiadas y el número de lecturas coloreadas por fuente de DNA.

I. Preprocesamiento

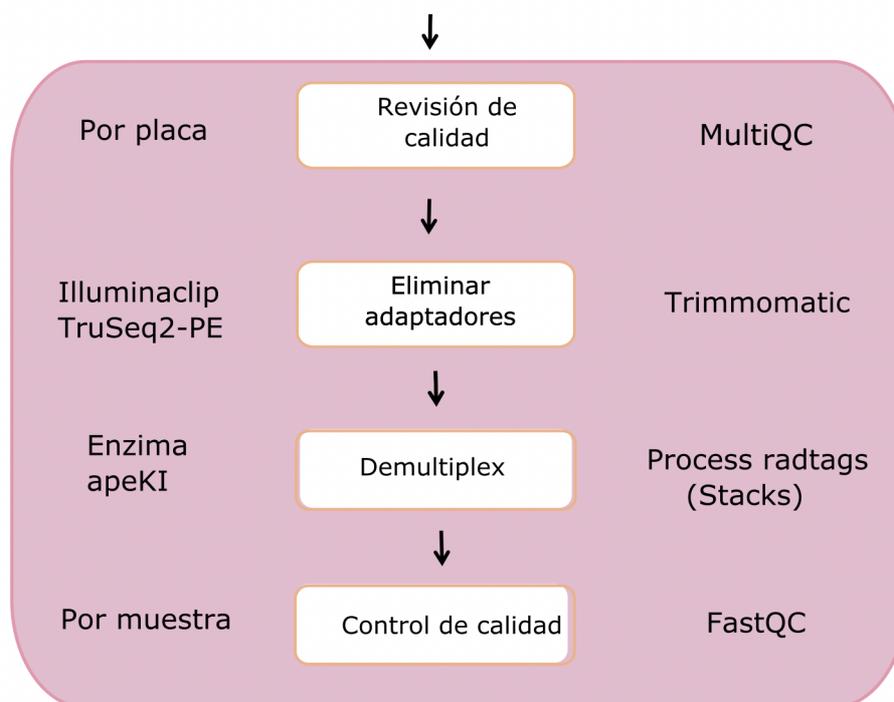


FIGURA 3. Flujo de trabajo del pre-procesamiento de los datos. En medio de muestra el paso (análisis de calidad conjunto, eliminar adaptadores, demultiplexeo y análisis de calidad por muestra), a la derecha se muestra el programa de cómputo utilizado en cada paso y a la izquierda especificaciones utilizadas.

2.- Alineamiento de lecturas

El alineamiento de lecturas se realizó utilizando como genoma de referencia el de una variedad local de *C. annuum*, CM334 (Kim *et al.*, 2014), proveniente de Morelos. El genoma de referencia se descargó desde GenBank con el código de acceso GCA_000512255.2, usando el vínculo ftp. El indexado del genoma se realizó utilizando el programa BWA v. 0.7.17 (Li y Durbin, 2009) con la función `bwa index` y el algoritmo `bwts` (para genomas grandes), pues el genoma de *C. annuum* var. *glabriusculum* es de aproximadamente 3.07 Gb (Qin *et al.*, 2014) (Fig. 4).

Para el alineamiento con el genoma de referencia, utilizamos el genoma previamente indexado. El alineamiento *per se* utilizó el algoritmo `bwa mem` implementado en BWA (Li y Durbin, 2009), dicho algoritmo se emplea para lecturas de alta calidad, es preciso y rápido, con un porcentaje de error menor al 3% para lecturas con longitud de 150 pb (Fig. 4) (Li y Durbin, 2009).

2. Alineamiento de lecturas

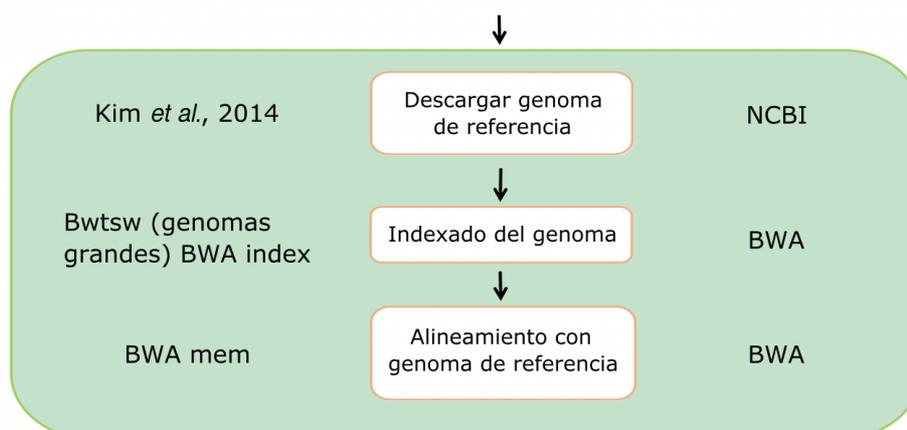


FIGURA 4. Flujo de trabajo del alineamiento de las lecturas, primero se descargó el genoma de referencia, posteriormente se indexó el genoma y finalmente alineamos las muestras de *Capsicum* con el genoma de referencia. Del lado derecho de la figura se especifica el parámetro utilizado o la referencia del genoma, en medio el objetivo y del lado izquierdo se indica el programa con el que se realizó cada análisis.

Para comparar el porcentaje de las lecturas que se alinearon exitosamente contra el genoma de referencia según el tipo de muestra (silvestre, semi-silvestres o cultivada) y el método de extracción de DNA (método de preservación del tejido y tipo de tejido de la planta), se visualizaron los datos con un diagrama de caja. La figura se hizo con el software R, con el paquete ggplot2 utilizando el output de *gstacks*. El diagrama de caja lo realizamos filtrando las muestras cuya fracción de mapeo de las lecturas contra el genoma de referencia fue menor a 0.60. Adicionalmente, se realizó una prueba de ANOVA de dos vías para verificar si el grado de domesticación, la fuente del DNA o la interacción entre ambas variables influía en la fracción de lecturas mapeadas contra el genoma de referencia, utilizando el paquete utilizando R.

3.- Llamado de SNPs

El llamado de SNPs por individuo se realizó con Stacks (Catchen et al., 2013), en una sola función (*ref_map.pl*) que engloba dos análisis: *gstacks* y *populations*. *Ref_map.pl*, trabaja con datos previamente alineados contra un genoma de referencia. Particularmente, *gstacks* realizó el llamado de SNPs y el ensamblado de los loci con las posiciones de alineamiento proporcionadas (Fig. 5) (Catchen et al., 2013). En este paso se utilizaron los valores por defecto de *gstacks*, incluido la cobertura mínima de 2.

Posteriormente los loci e individuos se filtraron modificando parámetros de *populations* conforme se indica en la siguiente sección.

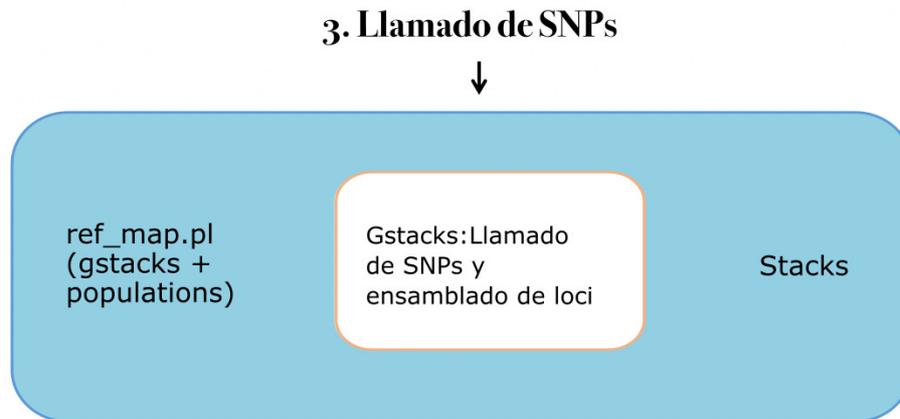


FIGURA 5. Flujo de trabajo del llamado de SNPs, que se realizó mediante el análisis `ref_map.pl` para el llamado de SNPs y el ensamblado de los loci. Del lado derecho de la figura se especifican los parámetros utilizados, en medio el objetivo y del lado izquierdo se indica el programa con el que se realizó el análisis.

4.-Análisis genómico-poblacionales

Los estadísticos resumidos a nivel poblacional los obtuvimos de *populations* del análisis `ref_map.pl`, descrito anteriormente. Inicialmente filtramos loci que no estaban en equilibrio Hardy-Weinberg. Este paso se realizó porque se detectaron regiones con un exceso de sitios variables por locus, lo cual crea un exceso de heterocigosis. Este exceso de heterocigosis en datos GBS frecuentemente es causado por regiones repetitivas o parálogos que se ensamblan como un solo locus (Mastretta-Yanes *et al.*, 2014).

Posteriormente se filtraron los datos acordes a: el número mínimo de poblaciones en las que debería estar presente un locus (*-p*). Este parámetro se estableció como -17 ya que es un poco más de la mitad del total de poblaciones (30). El porcentaje mínimo de individuos a nivel poblacional para que un locus se procese (*-r*) se estableció en 0.5 (50% de individuos por población). Se filtraron los loci cuya frecuencia mínima del alelo menor (*-MAF*) fuera <0.05; esta medida nos ayuda a

descartar loci que son poco polimórficos y que podrían ser producto de errores de secuenciación. También se estableció el número mínimo de alelos menores para procesar un SNP (*-min-mac*, parámetro que se aplica a nivel de población) en 2. Este parámetro sirve para evitar loci monomórficos. Por último, ya que FastStructure (Raj *et al.*, 2014) requiere SNPs que no estén ligados genéticamente, utilizamos la función *--write-single-snp* de Stacks para exportar solo un SNP por locus, lo que evita que el conjunto de datos tenga SNPs a pocos pares de bases entre sí, que estarían fuertemente ligados (Fig. 6).

Para explorar la estructura y detectar los conglomerados en los datos genómicos, así como los datos atípicos, realizamos un Análisis de Componentes Principales o (PCA, por sus siglas en inglés). Para realizar el PCA convertimos los datos a un archivo *.bed* con Plink v 1.9 (Purcell *et al.*, 2007), utilizando los datos genéticos con la función *-pca*, posteriormente se graficaron los dos primeros componentes en el paquete *ggplot2* (Wickham, 2016) en el programa (R Core Team, 2020). En el PCA se incluyó el grado de domesticación y se identificaron gráficamente las siguientes propiedades como análisis exploratorio: localidad geográfica, estados mexicanos, provincias biogeográficas, fuente de DNA (para descartar que tuviera un efecto sobre los agrupamientos) y latitud y longitud.

Posteriormente realizamos un Análisis Discriminante de Componentes Principales o DAPC por sus siglas en inglés, en Rstudio (R Core Team, 2020) con la paquetería *adegenet* (Jombart, 2008; Jombart y Ahmed, 2011). El DAPC es un enfoque estadístico multivariado, el cual requiere poblaciones definidas, de tal forma que maximiza la varianza en las poblaciones, la minimiza dentro del grupos y la divide en componentes inter e intrapoblacionales. Los datos se transforman mediante I) un análisis de componentes principales o PCA, y II) los grupos se identifican mediante el análisis discriminante DA. El DAPC, identifica grupos de individuos genéticamente relacionados y la contribución de alelos individuales a la estructuración de la población (Jombart *et al.*, 2010; Jombart y Ahmed, 2011; Jombart y Collins, 2015; Miller *et al.*, 2020).

Con el objetivo de inferir el número de poblaciones e individuos ancestrales, analizamos los datos filtrados con el programa FastStructure v.1 (Raj *et al.*, 2014),

probando de 1 ($K=1$) a 4 ($K=4$) grupos utilizando la función `structure.py`. Se probó ese número de grupos porque se encontraron tres grupos genéticos *principales* en el análisis de PCA. Posteriormente, con `chooseK.py` elegimos la K más adecuada para nuestros valores, por la complejidad del modelo que maximiza la probabilidad marginal y los componentes del modelo utilizados para explicar la estructura en los datos. Así, teniendo un valor de K que mejor se ajusta a nuestros SNPs que fue $k=3$, graficamos los datos con el paquete `Pophelper` (Francis, 2017) de Rstudio (R Core Team, 2020), ordenando los datos por individuo y población. Adicionalmente también graficamos $k=4$, para observar si se mantenían los cuatro grupos propuestos por los PCA. Aquí dibujamos las elipses manualmente, debido a que tenemos pocas muestras por población y al hacer las elipses de forma estadística no obteníamos resultados.

Con la finalidad de realizar un agrupamiento jerárquico UPGMA (método de grupos de pares no ponderados con media aritmética) de distancias genómicas, realizamos un dendrograma con R con la paquetería `ape` (Paradis *et al.*, 2004). En el dendrograma identificamos con distintos colores a cada agrupamiento que resultó del PCA.

Para evaluar si nuestras muestras de *C. annuum* obedecen a patrones geográficos, que provocan la estructura de las poblaciones, realizamos una prueba de Mantel por muestra, que consiste en la comparación de distancias genéticas y geográficas (Diniz-Filho *et al.*, 2013). La matriz de distancias genéticas la realizamos con R, con la paquetería `adegenet` (Jombart, 2008), que utiliza la distancia euclidiana entre vectores de frecuencias alélicas. La matriz de distancias geográficas pareadas en kilómetros la obtuvimos con Geographic Distance Matrix Generator v1.2.3. Ya con las matrices genéticas y geográficas, realizamos la prueba de Mantel en R con la paquetería `vegan` (Oksanen *et al.*, 2020). Los parámetros utilizados fueron: 1000 permutaciones y se empleó el coeficiente de correlación de Pearson.

Para obtener el número de sitios variables, la diversidad nucleotídica, la heterocigosis y la homocigosis por agrupamiento (de acuerdo con los resultados del PCA) empleamos `Stacks` (Catchen *et al.*, 2013) y para calcular los promedios Excel. De manera similar, calculamos los índices de diferenciación genética F_{ST} , con `VCFtools` v. 0.1.16 (Danecek *et al.*, 2011) para cada agrupamiento y para el promedio.

4. Análisis genómico-poblacionales

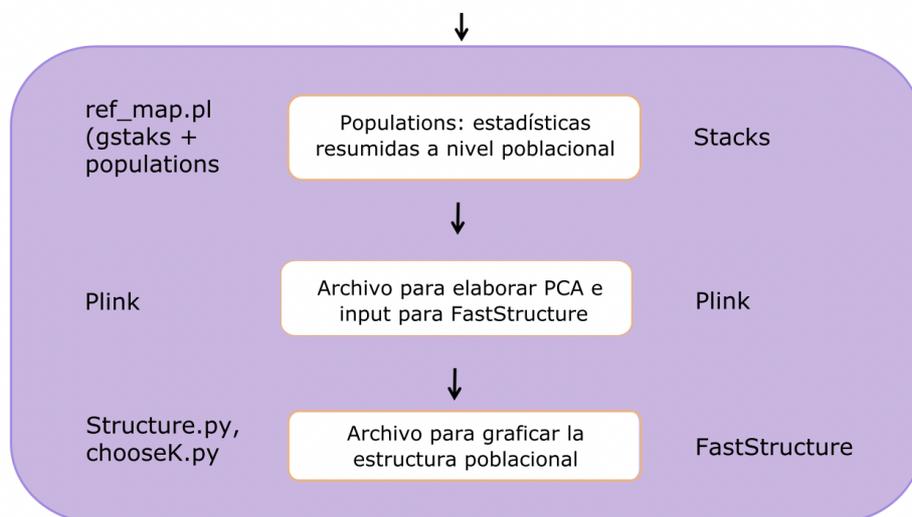


FIGURA 6. Flujo de trabajo de los análisis genómico-poblacionales. En esta parte nos enfocamos en las estadísticas a nivel poblacional de *populations* y en la elaboración de los PCA y FastStructure. Del lado derecho de la figura se especifica el parámetro, en medio el objetivo y del lado izquierdo se indican los programas con el que se realizó cada análisis.

Resultados

Muestreo

Obtuvimos sesenta y nueve muestras, treinta y siete silvestres y veintiséis semi-silvestres de *C. annuum* var. *glabriusculum* además de seis muestras cultivadas de *C. annuum* var. *annuum* distribuidas en el territorio mexicano (Fig. 7). Veinte provienen de la colecta de González-Jara *et al.* (2011); cuarenta y uno provienen de las colectas realizadas para este trabajo en colaboración con el Dr. Lev Jardón, que agrupa muestras colectadas en Querétaro, Chiapas y Oaxaca 2019, así como colectas de Oaxaca del año 2018; una muestra fue del herbario del Herbario Nacional del Instituto de Biología de la UNAM; las siete muestras restantes son de colectas realizadas en el año 2019. En la Tabla 3 también se ilustra el muestreo analizado, separado en muestras silvestres y semi-silvestres. Las muestras cultivadas se incluyeron como grupo externo.

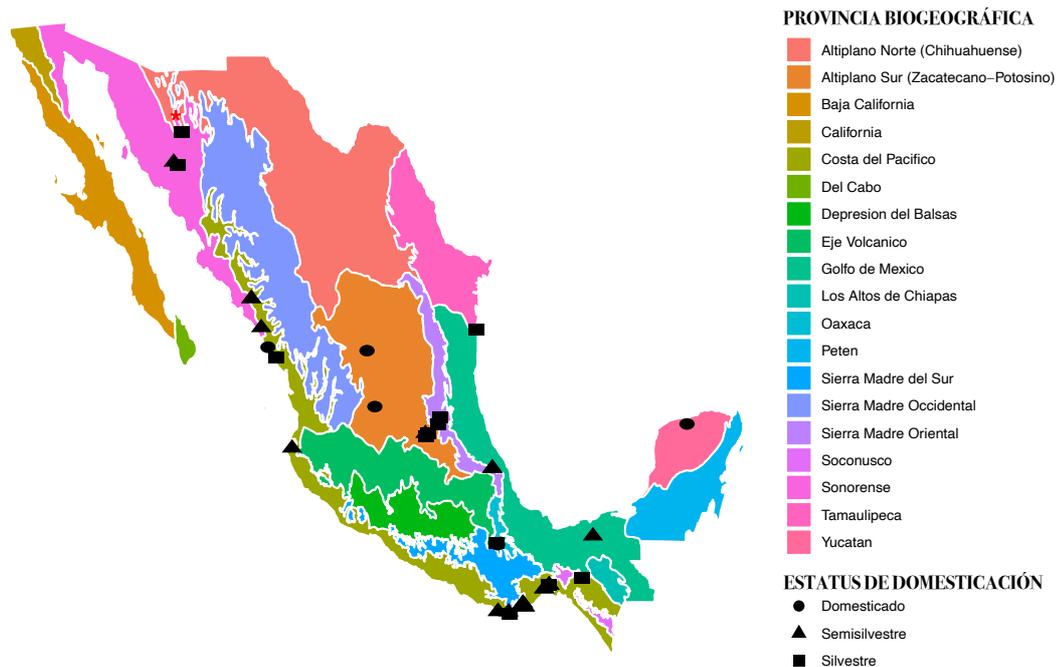


FIGURA 7. Mapa de colecta del chiltepín por provincias biogeográficas y estatus de domesticación (CONABIO, 1997). Las muestras silvestres se indican con un cuadro, las semisilvestres con un triángulo y las domesticadas con un círculo. *En el mapa la población de Moctezuma aparenta estar en Sonorense pero en realidad se encuentra en el Altiplano Norte.

Secuenciación y Genotipado

La extracción de *C. annuum*, se realizó probando diferentes fuentes de DNA: Fruto seco y congelado, tejido seco y fresco, Hula congelada, Herbario y el kit de recuperación de DNA REPLI-g, ver Anexo 1. Los individuos a los que les aplicamos el kit REPLI-g ya se encontraban extraídos, tenían aproximadamente un año a año y medio congelados a -70°C y su concentración de DNA era muy baja para estudio genómicos. Es por lo anterior que al aplicar el kit REPLI-g de Qiagen, obtuvimos una mayor concentración de DNA. Para ver los valores de la concentración antes y después de aplicar el REPLI-g (Anexo 2).

En todas las fuentes del DNA, obtuvimos muestras exitosas y no exitosas, como se puede observar en los geles. En algunos casos pudimos hacer más de una extracción por muestra en caso de que ésta no contara con la concentración adecuada, ya que teníamos el tejido suficiente de hoja y/o fruto. En cuanto al tejido de herbario, la extracción estaba limitada por la cantidad de tejido que nos fue donada por el Herbario

Nacional, es por eso que para la secuenciación solo incluimos la muestra de SOT1 (Soto la Marina, Tamaulipas). Los valores de concentración de DNA finales se muestran en el Anexo 3.

La calidad de las lecturas (*reads*), se encuentra por encima de valores Phred (Q) 30 (Fig. 8). Como la base de este indicador de calidad es logarítmica (i. e., $Q = -10 \log_{10} P$, donde P es la probabilidad de error), el valor de Phred de 30 indica que la probabilidad de que una base sea asignada incorrectamente es de 1/1000. Como se infiere de la Fig. 8, tanto la extracción como la secuenciación fueron de buena calidad para la totalidad de las muestras; los valores Phred comienzan a decaer a medida que llegamos al final de las lecturas, lo cual es esperado en la secuenciación GBS.



FIGURA 8. MultiQC de los resultados promedio de la secuenciación GBS de *C. annuum*. En el eje x se encuentra la posición en pares de bases y en el eje y los valores Phred. La zona rosa indica una calidad de secuenciación baja, la amarilla media y la verde alta.

El número de lecturas por muestra de *C. annuum* se ilustra en la Fig. 9. El control (primera barra) tiene pocas lecturas, con ello podemos descartar contaminación en la placa. El promedio de lecturas por muestra fue mayor a 2.5 millones, lo cual indica que

la extracción y la secuenciación de la placa fue exitosa. No se perciben diferencias sustanciales debido a la fuente del DNA en el número de lecturas por muestra (Fig. 9), a excepción de dos muestras de DNA reamplificado, que tienen pocas lecturas. La primera muestra es de la localidad de Temporal Sonora, la segunda de Sanalona, Sinaloa, ambas cuentan con 1.5 millones de lecturas, ver Fig. 9.

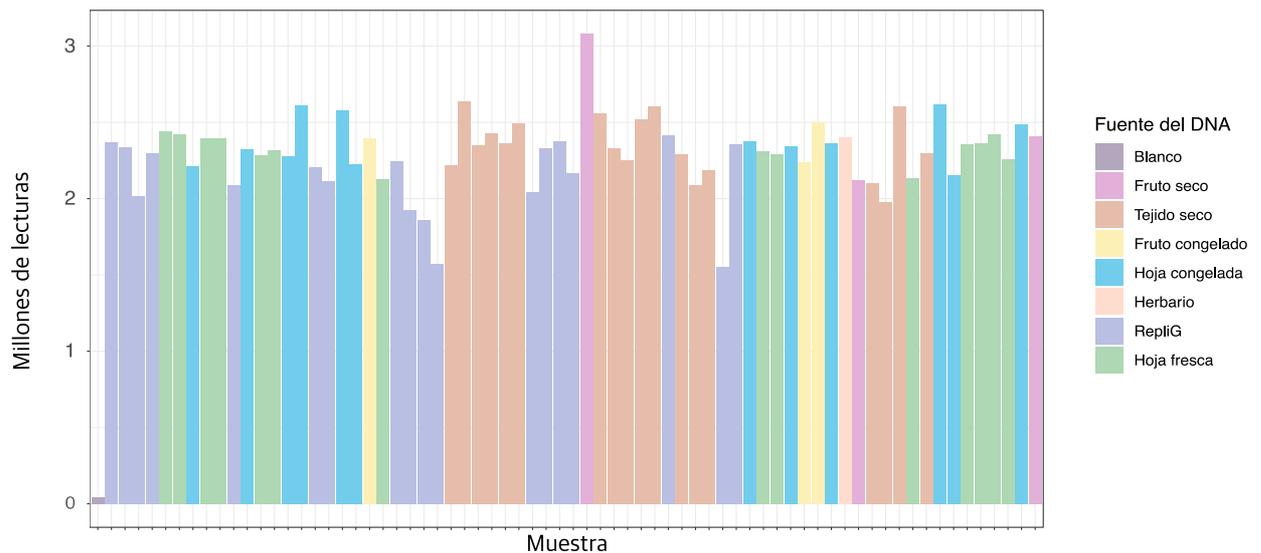


FIGURA 9. Gráfica de lecturas por muestra, para cada fuente del DNA.

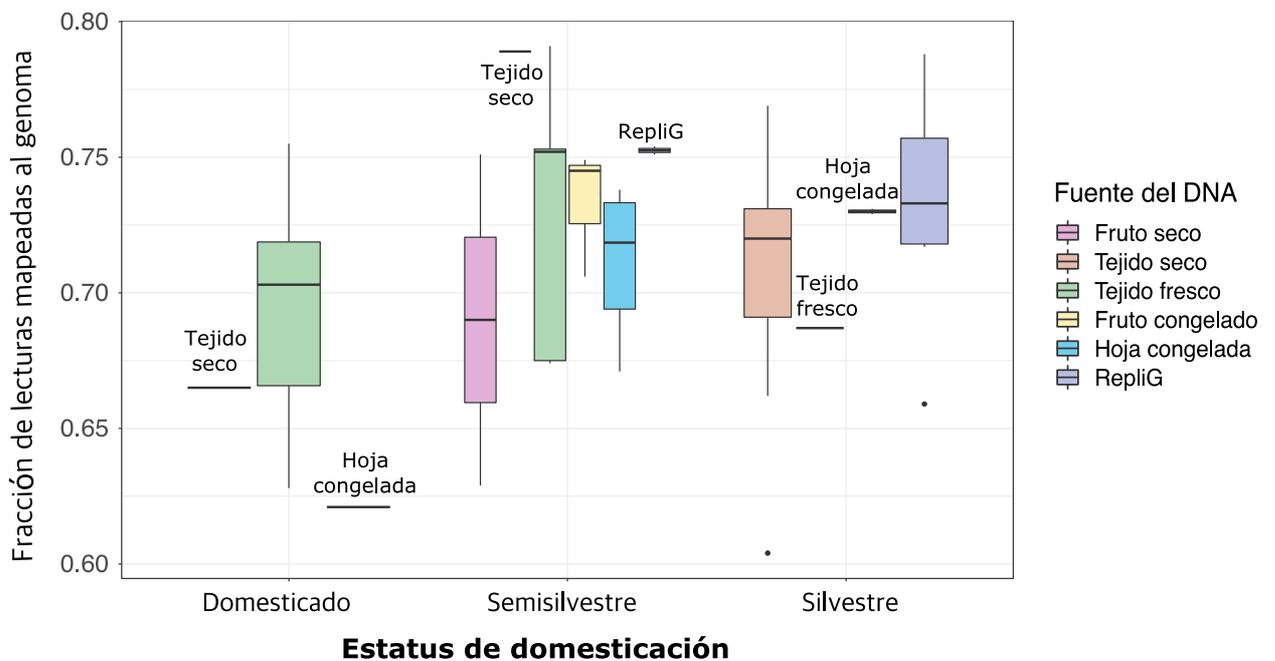


FIGURA 10. Diagrama de caja de la fracción mapeada al genoma de referencia con filtro de 0.60, graficado por estatus de domesticación y fuente de DNA.

Tabla 4. ANOVA de dos vías, para probar el efecto de la fuente de DNA y el estatus de domesticación en la fracción de lecturas mapeadas al genoma de referencia.

Variable	Grados de libertad	Suma de los cuadrados	Media cuadrática	Valor F	Significancia
Fuente del DNA	5	0.013	0.002	1.396	0.242
Estatus de domesticación	2	0.011	0.006	3.433	0.040
Interacción de la fuente de DNA y estatus de domesticación	5	0.0103	0.002	1.229	0.0310
Residuales	49	0.082	0.002		

Se utilizó el filtro 0.60 de lecturas mapeadas al genoma de referencia

La figura 10 muestra que el tejido fresco y seco tuvo una efectividad mayor al 65% para el mapeo del genoma de referencia para cada estatus de domesticación analizado: Silvestre, Semisilvestre y Domesticado. El fruto congelado y seco para los chiles Semisilvestres, también tienen una tasa de éxito mayor al 65% de mapeo y hay que considerar que esto es muy prometedor, ya que con ellos podemos evitar la germinación de las semillas y extraer material genético del fruto de Chile. Es importante mencionar que, entre el fruto seco y el congelado, éste último presentó los mejores porcentajes de mapeo, en nuestras muestras. El kit REPLI-g permitió analizar individuos que inicialmente no contaban con la concentración adecuada para su secuenciación, aunque al momento del filtrado a 0.60 se descartaron varias muestras y como se puede observar funciona para muestras Semisilvestres y Silvestres, esto es relevante ya que se puede aplicar en muestras con las que tenemos poco tejido, ya que las muestras silvestres tienen un tiempo muy largo de germinación y crecimiento en contraste con las plantas de Chile cultivadas. De acuerdo con nuestros resultados, la hoja congelada funciona mejor en chiles silvestres que en domesticados. El ANOVA

de dos vías mostró que la fracción mapeada al genoma de referencia no se encuentra asociada a la fuente del DNA, pero si al estatus de domesticación (Tabla 4, Anexo 4). Todos los análisis anteriores se realizaron con un total de 1,316,359 SNPs, los cuales aún no filtrábamos.

Análisis genómico-poblacionales

Los análisis a partir de esta sección se hicieron con los filtros $p=17$, $r=0.5$, $\text{min-maf}=0.05$, $\text{min-mac}=2$ y Hardy-Weinberg, descritos en el apartado 4 de metodología general. El número total de SNPs resultantes fueron 59,040 de 49 individuos (25 silvestres, 18 semi-silvestres, 6 cultivados). Los datos faltantes (*missing data*) tuvo un promedio de 13.78%, con una desviación estándar de 9.28.

En el PCA el primer componente principal contribuye con el 58.2 % y el segundo componente el 14.1% de la varianza (Fig. 11), que sumados incluyen más del 72% de la varianza genética total, por lo que centramos el análisis en esos dos componentes ya que el porcentaje de la varianza explicada para cada uno de los componentes decae a partir del componente 3.

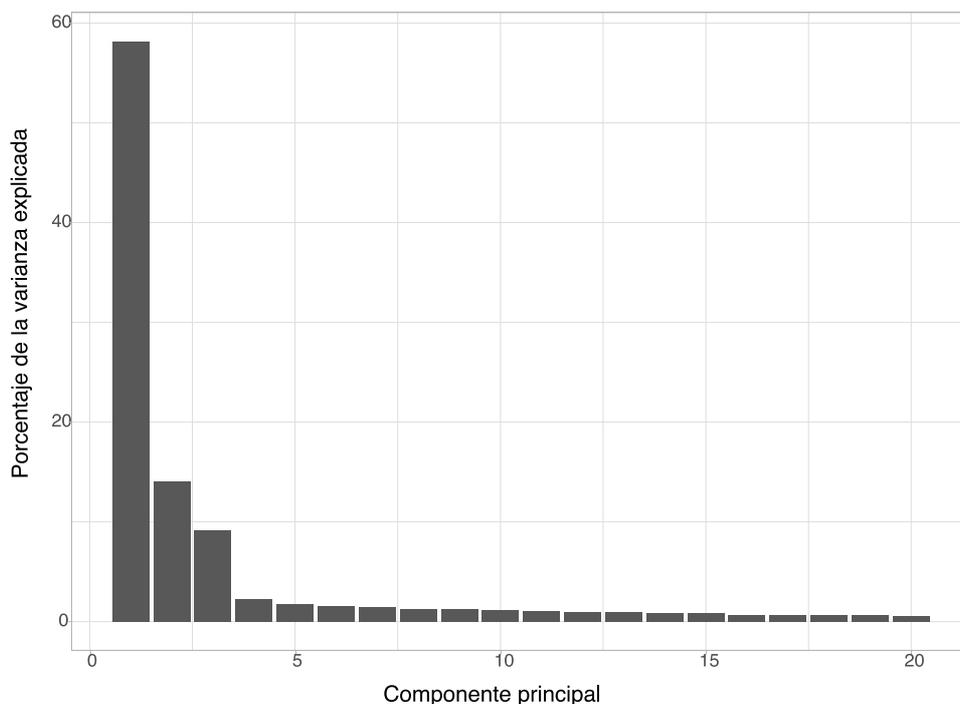


Figura 11. Porcentaje de varianza explicada para cada uno de los componentes principales de

59,040 SNPs.

Las poblaciones se agrupan en cuatro grupos genéticos (Fig. 12-14): NORTE) Sinaloa, Sonora; CENTRO) Querétaro; SUR) Oaxaca, Tabasco, Veracruz; DIVERSO) tres muestras domesticadas de Zacatecas, Oaxaca y Yucatán y dos provenientes de Veracruz y Oaxaca, semi-silvestre y silvestre respectivamente (Tabla 5), además debemos considerar que el componente principal 1 explica el 58.2% de la variación genética y el componente principal 2 el 14.1 y en conjunto el 72.3% de la variación genética total, lo cual es muy elevado. En la esquina superior izquierda de las Fig. 12-14, están las muestras de Querétaro, de tierras relativamente altas.

El mismo PCA, pero indicando las provincia biogeográfica y grado de domesticación se muestra en la Fig. 12. En este caso se formaron tres agrupamientos genéticos principales: 1. Altiplano Norte (Chihuahuense), Costa del Pacífico, Sonorense 2. Altiplano Sur (Zacatecano-Potosino) 3. Costa del Pacífico, Golfo de México. 4. Altiplano Sur (Zacatecano-Potosino), Oaxaca, Golfo de México y Yucatán. Es así que el estatus de domesticación no parece un factor importante para la formación de agrupamientos genéticos.

El PCA por origen de DNA y grado de domesticación (Fig. 13) formaron tres agrupamientos principales: 1) REPLI-g; 2) Hoja congelada, tejido de invernadero, fruto congelado y fruto seco; 3) Tejido seco, REPLI-g y un cuarto grupo, que contiene muestras de hoja fresca y congelada, aquí notamos que, si bien no teníamos para cada población cada origen del tejido analizado, podemos notar que el grupo 1 y 3 comparten REPLI-g. En la Fig. 14a se muestra una posible existencia de patrones latitudinales que corresponden a cada una de las muestras, mientras que en la Fig. 14b se aprecia que la distribución de las muestras también parece asociada a la longitud.

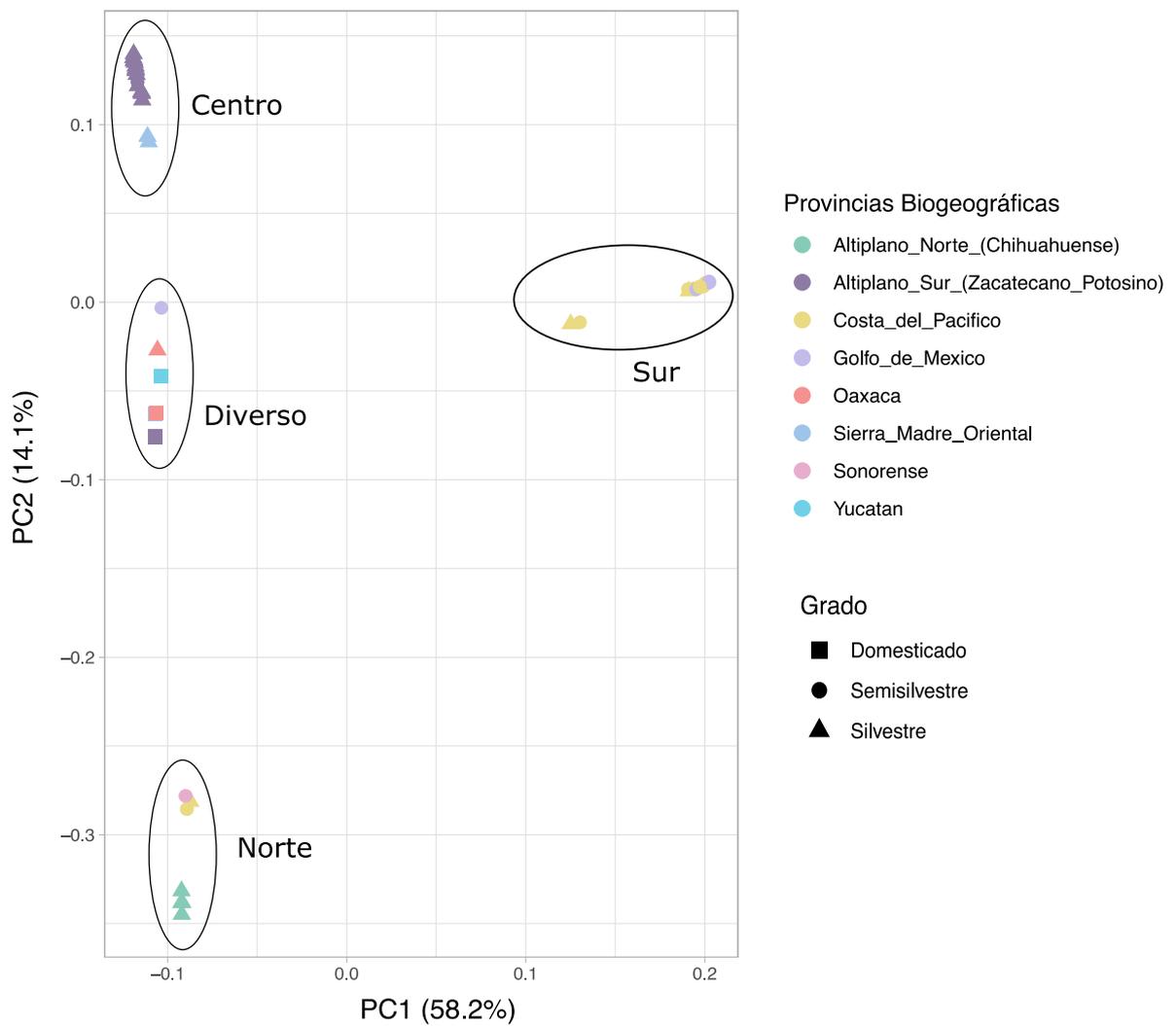


FIGURA 12. PCA de un total de 59,040 SNPs de los dos primeros componentes genéticos, indicando la provincia biogeográfica y grado de domesticación. Las elipses se agregaron manualmente para señalar de manera más clara cada uno de los grupos propuestos.

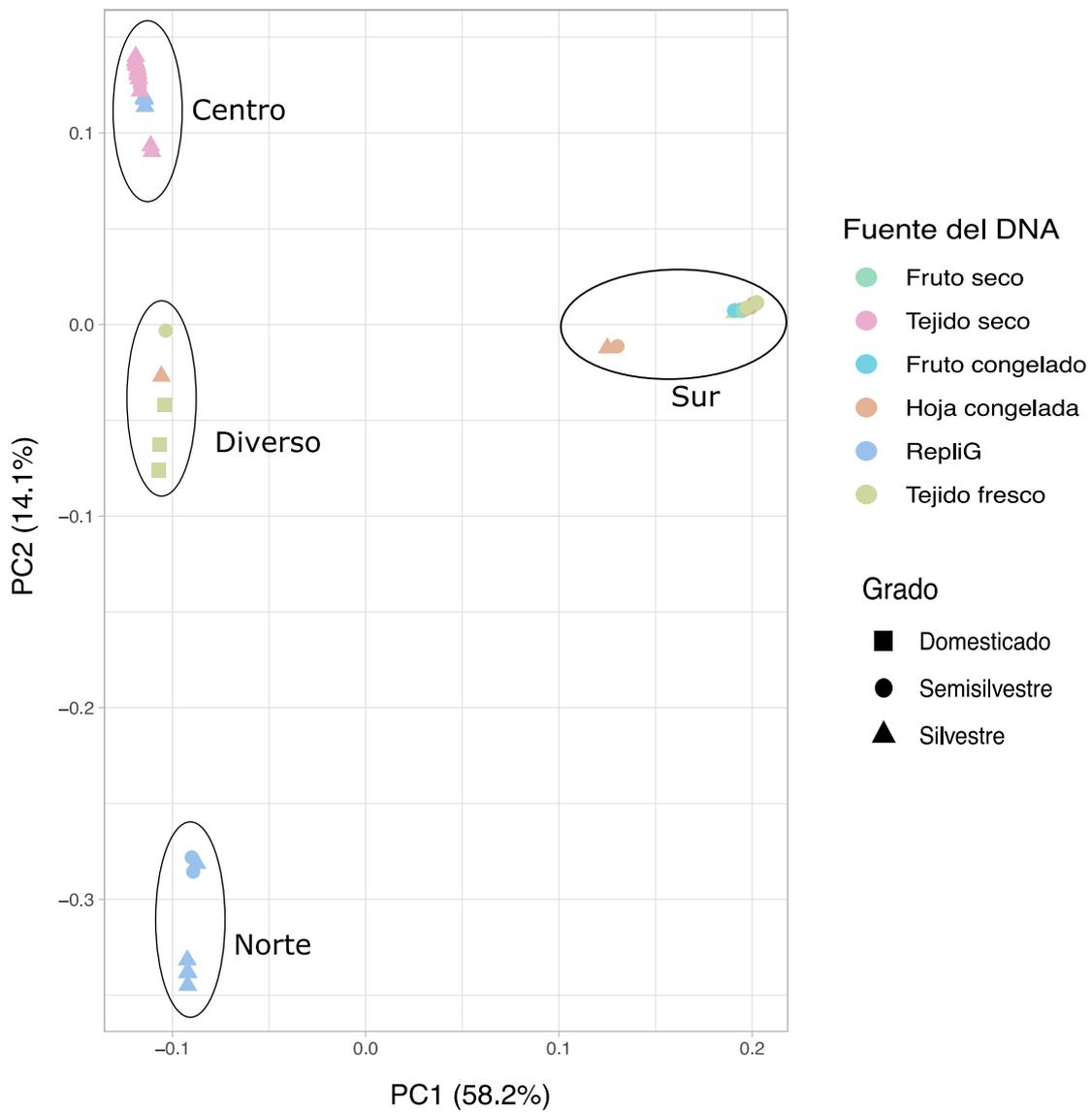


FIGURA 13. PCA de un total de 59,040 SNPs de los dos primeros componentes genéticos, indicando la por fuente de DNA y grado de domesticación. Las elipses se agregaron manualmente para señalar de manera más clara cada uno de los grupos propuestos.

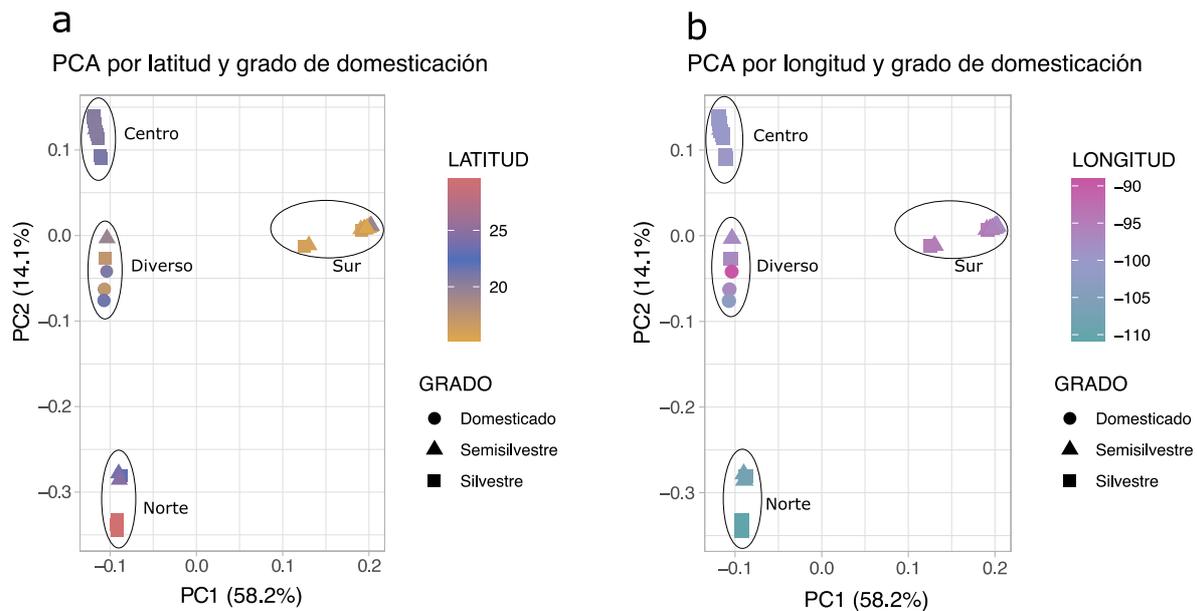


FIGURA 14. PCA de un total de 59,040 SNPs de los dos primeros componentes genéticos, por grado de domesticación y 13a. latitud y 13b. longitud. Las elipses se agregaron manualmente para señalar de manera más clara cada uno de los grupos propuestos.

Para complementar el análisis de PCA, realizamos un DAPC por agrupamiento tomando como base los grupos definidos por el PCA (Tabla 5). Los grupos propuestos por el DAPC fueron: Norte, Centro y Sur + Diverso (Fig. 15). Los agrupamientos Sur + Diverso no se diferencian entre sí. Las elipses inerciales alrededor de los grupos representan el espacio de variación y asignación de los individuos, esto se puede observar en el grupo Centro y Norte. Dichas elipses ilustran un supuesto de agrupamiento y los puntos (individuos), que se salen de ella no se ajustan a un patrón geográfico.

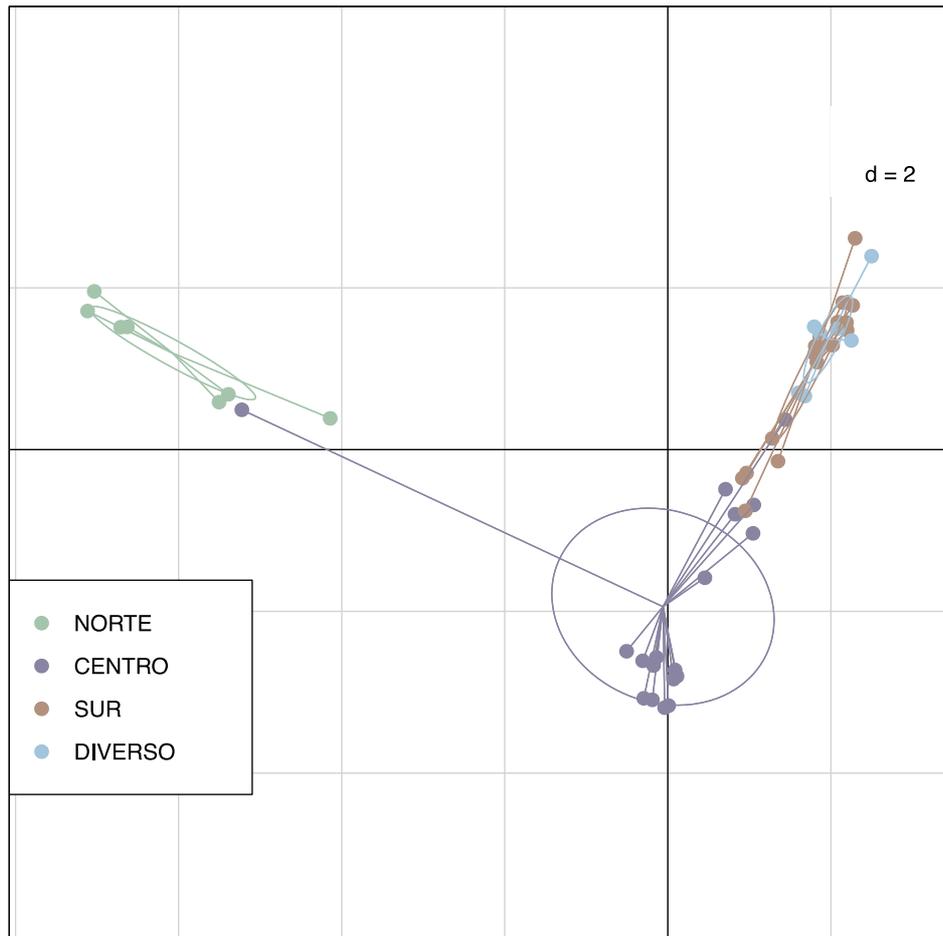


Figura 15. DAPC por agrupamiento: Norte, Centro, Sur, Diverso, para *C. annuum*.

Tabla 5. Muestra asociada a cada uno de los cuatros agrupamientos genómicos: Norte, Centro, Sur y Diverso. Se presenta el estado, localidad y grado de domesticación para cada muestra.

GRUPO	MUESTRA	ESTADO	LOCALIDAD	GRADO DE DOMESTICACIÓN
NORTE	CaMOC15	Sonora	Moctezuma	Silvestre
NORTE	CaMOC25	Sonora	Moctezuma	Silvestre
NORTE	CaMOC30	Sonora	Moctezuma	Silvestre
NORTE	CaMOC37	Sonora	Moctezuma	Silvestre
NORTE	CaSAN2	Sinaloa	Sanalona	Semi-silvestre
NORTE	CaELO15	Sinaloa	Elota	Semi-silvestre
NORTE	CaHUJ4	Sinaloa	El Huajote	Silvestre
CENTRO	CaRES5	Querétaro	Rancho la Espadañuela	Silvestre
CENTRO	CaRES6	Querétaro	Rancho la Espadañuela	Silvestre
CENTRO	CaBER21	Querétaro	Bernal	Silvestre
CENTRO	CaBER23	Querétaro	Bernal	Silvestre
CENTRO	CaBER33	Querétaro	Bernal	Silvestre
CENTRO	CaBER36	Querétaro	Bernal	Silvestre
CENTRO	CaTOL1	Querétaro	Tolimán	Semi-silvestre
CENTRO	CaMIR2	Querétaro	El Mirador	Silvestre
CENTRO	CaMIR6	Querétaro	El Mirador	Silvestre
CENTRO	CaMIR13	Querétaro	El Mirador	Silvestre
CENTRO	CaMIR15	Querétaro	El Mirador	Silvestre
CENTRO	CaMIR16	Querétaro	El Mirador	Silvestre
CENTRO	CaMIR18	Querétaro	El Mirador	Silvestre
CENTRO	CaPAT3	Querétaro	El Patol	Silvestre
CENTRO	CaPAT7	Querétaro	El Patol	Silvestre
CENTRO	CaPAT8	Querétaro	El Patol	Silvestre
CENTRO	CaPAT9	Querétaro	El Patol	Silvestre
CENTRO	CaPAT11	Querétaro	El Patol	Silvestre
SUR	CaJUC1	Oaxaca	Juchitán	Semi-silvestre
SUR	CaJUC2	Oaxaca	Juchitán	Semi-silvestre
SUR	CaJUC3	Oaxaca	Juchitán	Semi-silvestre
SUR	CaTOM1	Oaxaca	El Tomatal	Semi-silvestre
SUR	CaTOM2	Oaxaca	El Tomatal	Semi-silvestre
SUR	CaYEL1	Jalisco	Yelapa	Semi-silvestre

GRUPO	MUESTRA	ESTADO	LOCALIDAD	GRADO DE DOMESTICACIÓN
SUR	CaGVI3	Oaxaca	Guadalupe Victoria	Semi-silvestre
SUR	CaCOY2	Oaxaca	El Coyul	Semi-silvestre
SUR	CaCOY11	Oaxaca	El Coyul	Semi-silvestre
SUR	CaSDI19_2R	Oaxaca	Santo Domingo del Ingenio	Silvestre
SUR	CaSDI2	Oaxaca	Santo Domingo del Ingenio	Semi-silvestre
SUR	CaSDI3	Oaxaca	Santo Domingo del Ingenio	Silvestre
SUR	CaSDI21	Oaxaca	Santo Domingo del Ingenio	Semi-silvestre
SUR	CaSDI22	Oaxaca	Santo Domingo del Ingenio	Semi-silvestre
SUR	CaVPI1_1	Veracruz	Tlapacoyan	Semi-silvestre
SUR	CaVPI2_11	Veracruz	Tlapacoyan	Semi-silvestre
SUR	CaTab1	Tabasco	Río Viejo	Semi-silvestre
DIVERSO	CaVPI2_1	Veracruz	Tlapacoyan	Semi-silvestre
DIVERSO	CaVTR1	Oaxaca	Valerio Trujano	Silvestre
DIVERSO	CaDZI1_1	Yucatán	Dzibichaltún	Domesticado
DIVERSO	CaCHI4_2R	Oaxaca	El Chilar	Domesticado
DIVERSO	CaRES9	Querétaro	Rancho la Espadañuela	Domesticado
DIVERSO	CaVCO1_8	Zacatecas	Villa de Cos	Domesticado
DIVERSO	CaTEC2_3	Zacatecas	El Tecuán	Domesticado

Usando FastStructure se infirió que la estructura poblacional que más se ajustaba a nuestros datos de SNPs fue una $K=3$ (Fig. 16), considerando tanto la complejidad del modelo que maximiza la probabilidad marginal, como los componentes del modelo utilizados para explicar la estructura en los datos. El 100% del genoma de la mayoría de los individuos, corresponde a uno de estos tres grupos genómicos: 1. Norte de México (Sinaloa, Sonora); 2. Centro de México (Querétaro); 3. Sur de México + Diverso (Oaxaca, Tabasco, Veracruz, Yucatán y Zacatecas), que coinciden con el DAPC y son similares a los resultados del PCA.

En el muestreo hay estados que están representados por una sola muestra, como el caso de Jalisco, Tabasco, y en algunos estados como Yucatán y Zacatecas sólo tenemos muestras domesticadas (Tabla 5, Fig. 16). Esto es relevante porque a pesar de que el muestreo no es exhaustivo a nivel poblacional, se observa un patrón de diferenciación geográfica.

El patrón geográfico observado es independiente del grado de domesticación (silvestre y semisilvestre) y es más claro para el sur y norte de México. En el caso de los chiles cultivados el chile huacle (CaCHI4_2R), chile criollo (CaRES9), chile dulce (CaDZI1_1), chile guajillo (CaTEC2_3) y mirasol (CaVCO1_8), se agrupan en el grupo perteneciente al Norte de México, mientras que el único chile cultivado que se agrupa con las muestras del Sur de México es un piquín (CaCOY2).

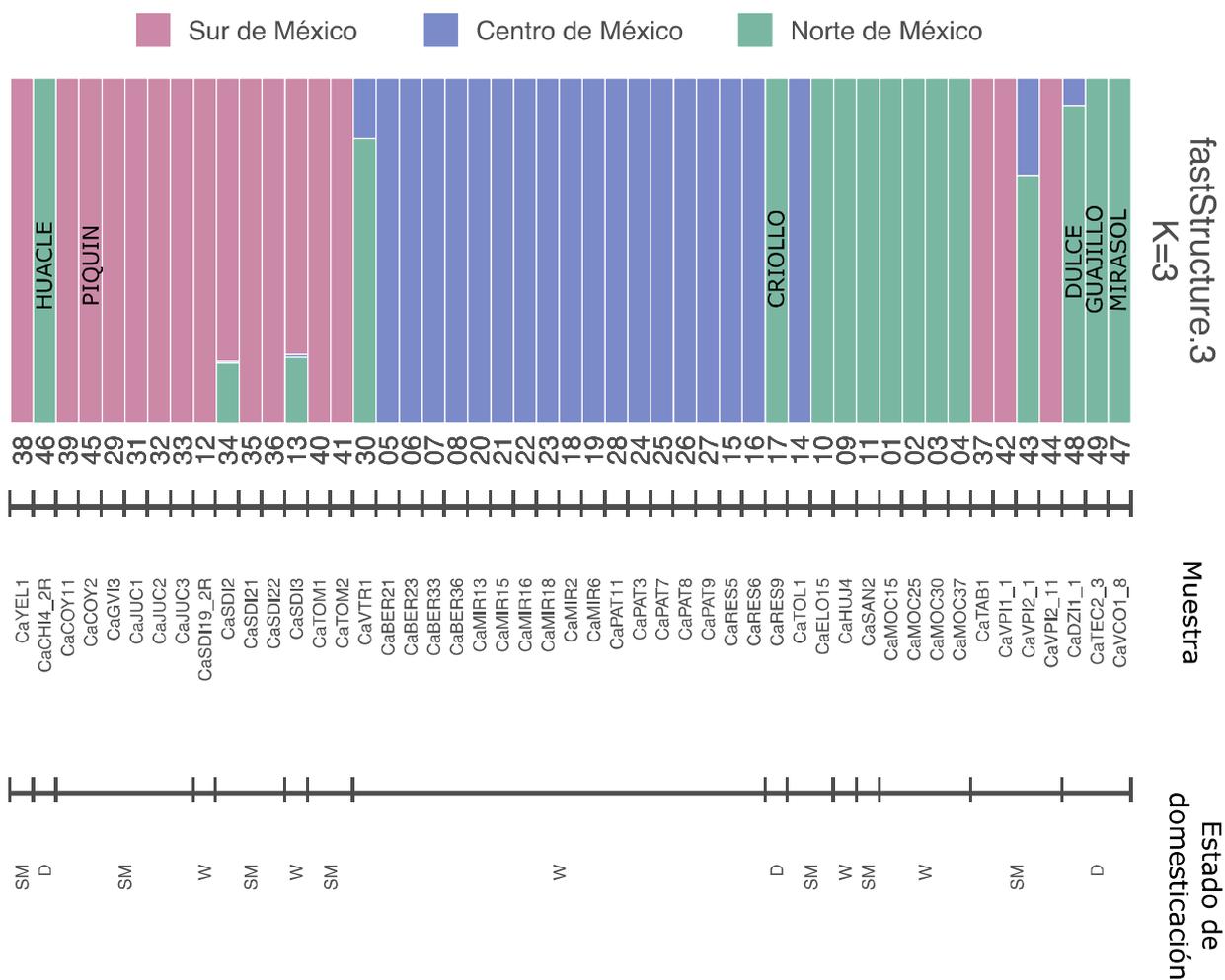


FIGURA 16. Resultados de FastStructure para K=3 en *Capsicum annum*. Las muestras están ordenadas según su procedencia de estados de México (ESTADO). Se indica el GRADO de domesticación si la muestra es de origen Silvestre (W), SemiSilvestres (SM), Domesticadas (D). De las muestras Domesticadas, se detalla el nombre común del chile y la muestra semisilvestre piquín. De cada muestra las barras de colores indican la proporción del genoma que corresponde a los tres grupos genéticos detectados, que corresponden al Norte, Centro y Sur de México.

Para probar si se establecían los cuatro grupos genéticos iniciales del análisis de componentes principales, decidimos graficar los resultados obtenidos con FastStructure para K=4 (Fig. 17). En este caso, a pesar de que el análisis tuvo una k de 4 solo se perciben tres grupos: Norte (morado), Centro (verde) y Sur de México (rosa), de tal manera que no se visualiza el agrupamiento número 3 (naranja) en este análisis de FastStructure.

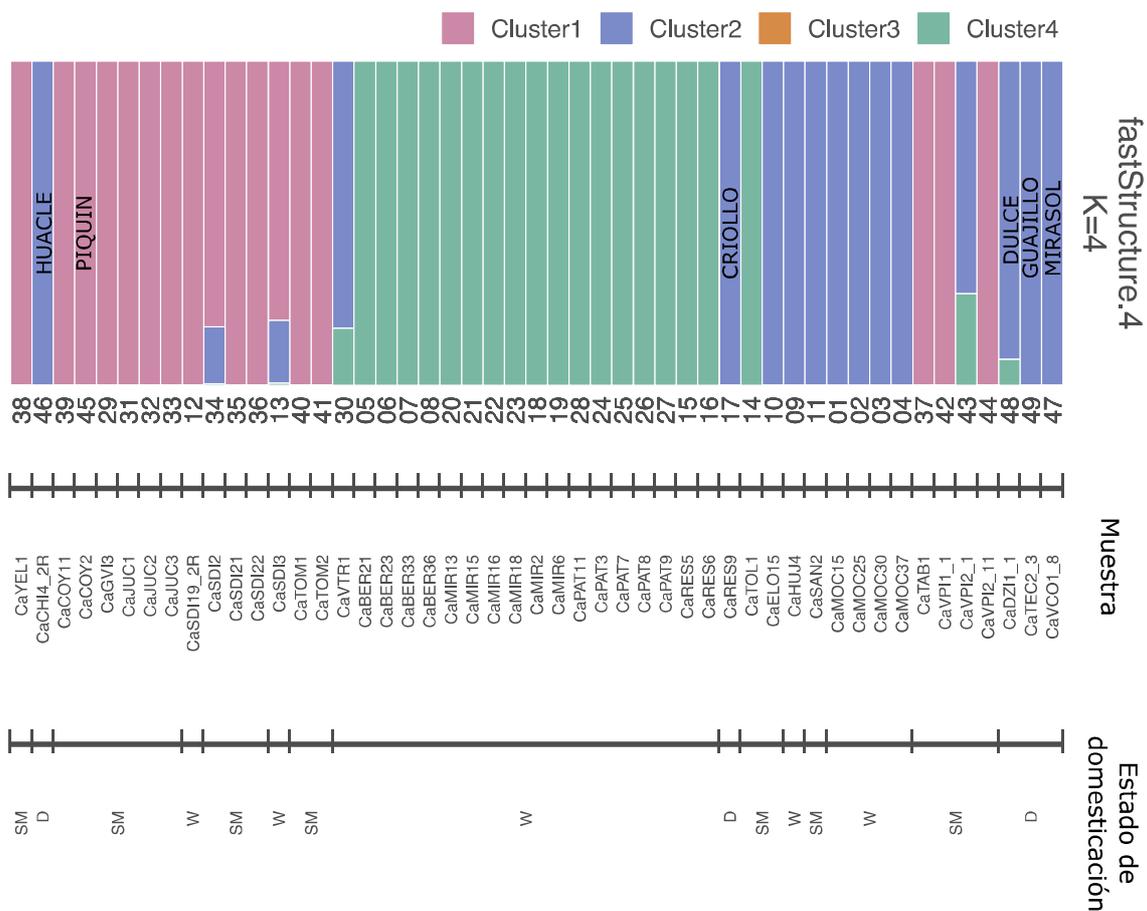


FIGURA 17. Resultados de FastStructure para K=4 en *C. annuum*. Las muestras están ordenadas según su procedencia de estados de México (ESTADO). Se indica el GRADO de domesticación si la muestra es de origen Silvestre (W), Semi-Silvestres (SM), Domesticadas (D). De las muestras domesticadas, se detalla el nombre común del chile y la muestra semisilvestre piquín. De cada muestra las barras de colores indican la proporción del genoma que corresponde a los cuatro grupos genéticos.

El dendrograma construido con las 49 muestras y los 59,040 SNPs (Fig. 18) es congruente con los resultados obtenidos en el PCA, en el que observan los mismos cuatro grupos: Norte, Centro, Sur y Diverso, la mayoría de las ramas tiene un soporte de 100 de acuerdo con el método de remuestreo.

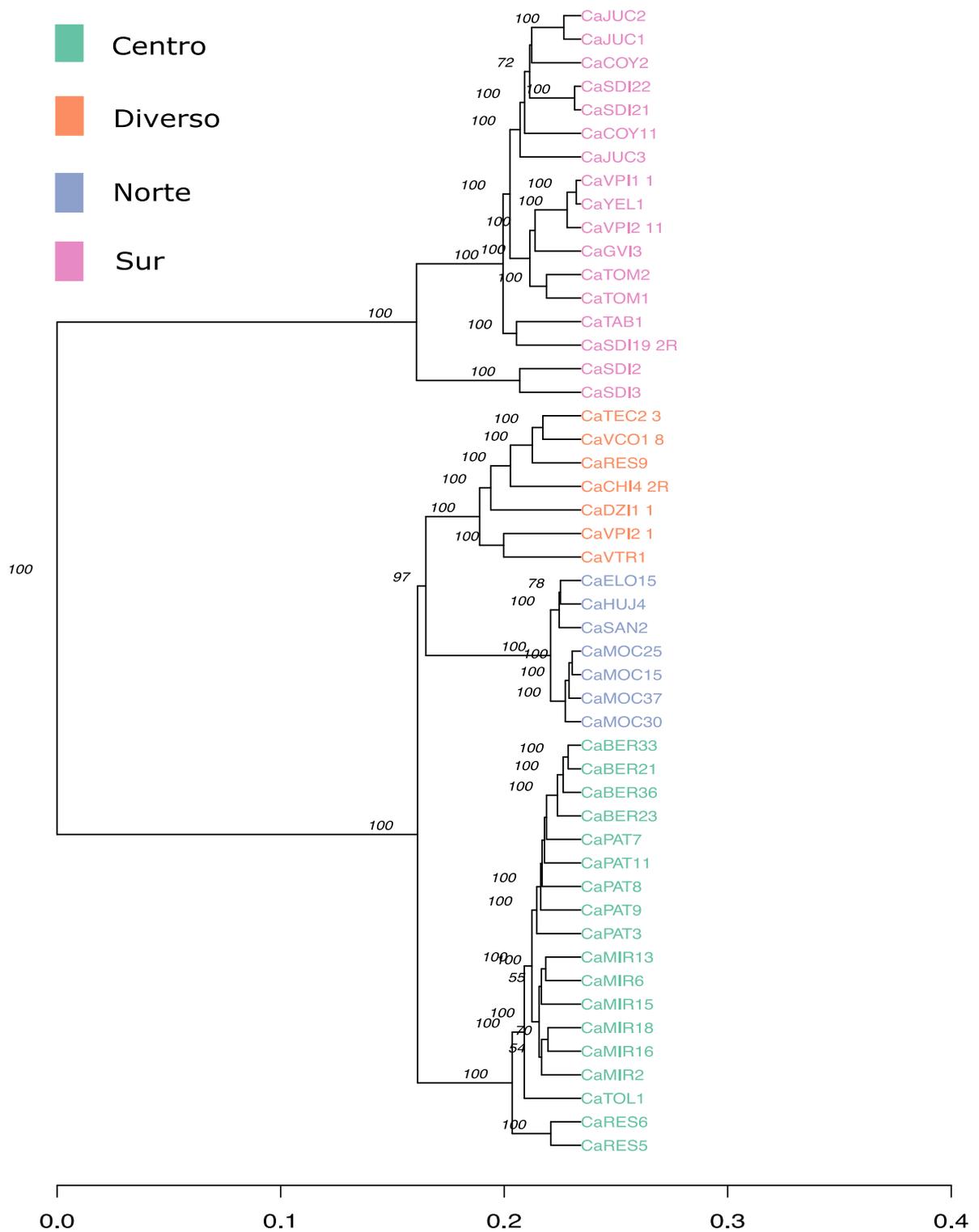


Figura 18. Dendrograma de distancias genómicas, por medio del método UPGMA. Los colores muestran los cuatro agrupamientos: Norte, Centro, Sur y Diverso.

La prueba de Mantel con datos genéticos y geográficos arrojó los siguientes resultados: $r = 0.447$ y un valor de $p = 0.001$, la relación entre la distancia geográfica y genética se ilustra en la Fig. 19, en la que se puede ver que, a mayor distancia genética, mayor es la distancia geográfica.

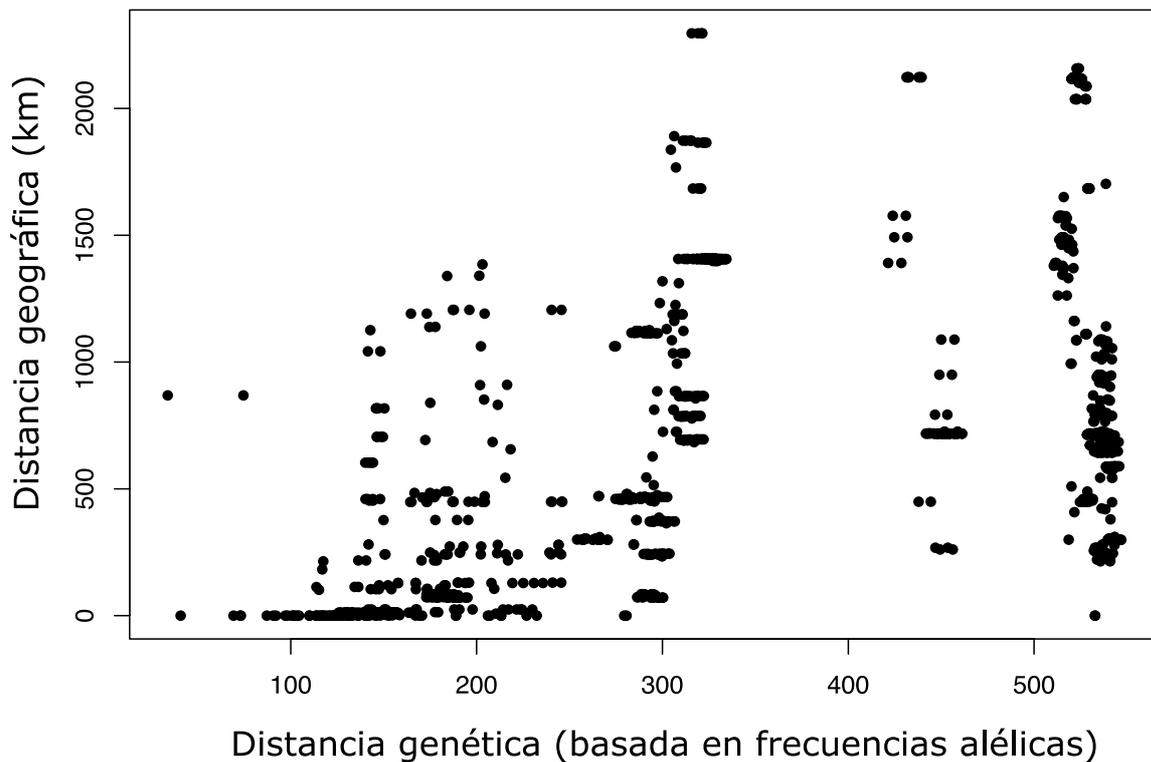


Figura 19. Relación entre la distancia genética y geográfica en *C. annuum*.

Teniendo en consideración que en algunas poblaciones no contamos con más que con un individuo, calculamos las medidas de diversidad y la F_{ST} para *C. annuum* por grupos (Tabla 5). El número de alelos privados fue menor en el grupo Diverso (3,281) y mayor en el Sur (23,579), resultado que explica las grandes diferencias que existen entre el grupo Sur y el resto de los agrupamientos. Asimismo, el mayor número de alelos privados en el Sur se vió reflejado en su diversidad nucleotídica (0.0098) y en su mayor heterocigosis, la cuales también fueron las mayores de todas. Los resultados se muestran a continuación (Tabla 6).

La diversidad nucleotídica (π) para el promedio de los grupos para todas las muestras fue de 0.071. La heterocigosidad total (He) fue de 0.031, y también se calculó para cada agrupamiento (Tabla 6) y fueron las siguientes:

Tabla 6. Índices de diversidad genómica para cada agrupamiento obtenido con el PCA.

	Norte	Centro	Sur	Diverso	Promedio
No. de alelos privados	3,486	5,787	23,579	3,281	9,033.25
Diversidad nucleotídica (π)	0.037 \pm 0.0006	0.066 \pm 0.0007	0.098 \pm 0.0007	0.084 \pm 0.0008	0.071
Heterocigosidad observada	0.012 \pm 0.0004	0.03 \pm 0.0004	0.047 \pm 0.0004	0.021 \pm 0.0004	0.027
Heterocigosidad esperada	0.034 \pm 0.0005	0.064 \pm 0.0006	0.094 \pm 0.0007	0.077 \pm 0.0007	0.067
Homocigosidad observada	0.988 \pm 0.0004	0.97 \pm 0.0004	0.953 \pm 0.0004	0.979 \pm 0.0004	0.972
Homocigosidad esperada	0.966 \pm 0.0005	0.936 \pm 0.0006	0.906 \pm 0.0007	0.923 \pm 0.0007	0.932

El \pm indica el error estándar para cada medida de diversidad.

El agrupamiento con la heterocigosidad más alta fue el grupo Sur, la más baja corresponde al agrupamiento Norte. El índice de diferenciación genético F_{ST} más alto fue la comparación Norte-Sur 0.735, el más bajo fue Centro-Diverso 0.419 (Tabla 7); mientras que el promedio fue de 0.596, valor inferior al de todas las comparaciones del grupo Sur, lo que señala que es el más divergente de todos.

Tabla 7. Índice de diferenciación genético F_{ST} por agrupamiento: Norte, Centro, Sur y Diverso en *C. annuum*.

GRUPO	NORTE	CENTRO	SUR	DIVERSO
NORTE				
CENTRO	0.512			
SUR	0.735	0.715		
DIVERSO	0.477	0.419	0.721	

Discusión

Pre-procesamiento genómico y análisis de diferentes fuentes de DNA de chiltepín

De acuerdo con el ANOVA, el mapeo contra el genoma de referencia está influenciado por el estatus de domesticación, pero no así por la fuente del DNA, esto se puede explicar porque nuestro genoma de referencia pertenecía a una variedad local de *C. annuum* (Kim et al., 2014), que probablemente mantenga más semejanzas con el chiltepín que con los chiles domesticados. Otros motivos por los que probamos la relación entre la fracción de genoma mapeado y las muestras fueron el de asegurarnos que las secuencias obtenidas pertenecieran a la especie de interés y no a contaminantes, procediera de muestras con material genético degradado o de errores en la secuenciación. De tal modo que recomendamos la extracción del material biológico disponible, como frutos secos o frescos, hojas congeladas u hojas secas e incluso se podría aumentar la concentración genómica de DNA previamente extraído con el kit REPLI-g, pues no afecta de manera detectable los análisis posteriores, lo cual es sumamente relevante, pues los chiltepinos son plantas de germinación y crecimiento difíciles a diferencia de los chiles cultivados (Hernández-Verdugo, 2018).

Las mejores fuentes de DNA para extracción fueron el tejido fresco y seco, así como los frutos congelados y secos, mientras que las menos prometedoras fueron el tejido de herbario y el DNA re-amplificado, aunque éste último presentó muestras con un alto porcentaje de mapeo. REPLI-g fue el método de recuperación de la concentración de DNA de muestras previamente extraídas sin fines genómicos. En el caso de la muestra de herbario, debemos considerar que solo extrajimos una muestra, por lo que este resultado no es concluyente. Las muestras restantes se extrajeron de diferentes tipos de tejido, entre los cuales los mejores resultados los dieron el tejido seco (4 meses almacenado en gel de sílice), los frutos secos (1-2 años almacenados en sobres de papel) y los frutos congelados (1-2 años a -20°C). Por lo tanto, nuestros resultados demuestran que es posible utilizar como fuente de DNA tejidos y muestras que previamente no hubieran sido consideradas en estudios de genómica poblacional. Resaltamos que los frutos secos y congelados de hasta dos años nos proporcionan excelentes resultados para extraer material genético, a pesar del gran contenido de capsaicina, compuesto que teóricamente degrada el DNA por la presencia de especies reactivas de oxígeno (Singh *et al.*, 2001; Yang *et al.*, 2009). Esto es relevante, ya que posiblemente es el primer trabajo en reportar que es factible la extracción de DNA de frutos de chiles. La extracción de frutos se realizó con éxito para individuos silvestres y semi-silvestres, tanto congelados como secos, lo que es conveniente ya que el tiempo de germinación y crecimiento de las plantas es prolongado para las plantas silvestre y semisilvestres (Casas *et al.*, 2007; Hernández-Verdugo, 2018). Por lo tanto, para futuros trabajos se recomienda que, si no hay disponibilidad de tejido, se puede optar por la extracción de fruto de chile, evitando así el tiempo prolongado de germinación de los chiles silvestres y semi-silvestres, siempre y cuando los frutos o el tejido no se sobrepase un lapso mayor a cinco años.

El tejido seco proporciona la ventaja de que se puede triturar con mayor facilidad, en comparación con el tejido fresco y se puede almacenar en gel de sílice para evitar la humedad y el crecimiento de hongos. El tejido de hoja congelada por un periodo de tiempo extenso (más de un año) es poco recomendable, por dos motivos: (I) al hacer el filtrado de las muestras a 0.60 de proporción de lecturas mapeadas al genoma de referencia descartamos varios individuos, y (II) la extracción de dicho tejido es compleja y toma mucho tiempo. El tejido de invernadero es una buena opción, aunque la desventaja principal es el prolongado y variable tiempo de germinación y crecimiento.

Hasta el momento se han publicado trabajos de la utilidad de REPLI-g en virus, bacterias, microbiomas, y secuenciación de células individuales de sangre humana (Herb *et al.*, 2012; Liu *et al.*, 2019; Mendell *et al.*, 2012; Qn *et al.*, 2019; Stepanauskas y Sieracki, 2007), pero no en genomas grandes de plantas, que suelen tener metabolitos secundarios inhibidores, como en *C. annuum*. Demostramos, cuando menos en unos casos, la efectividad del DNA re-amplificado con el kit REPLI-g en especies de plantas con genomas grandes. Nuestros análisis indican que su efectividad está dada por el material genético, perdimos algunas muestras por falta de calidad. El REPLI-g aumenta la concentración de DNA, pero es probable que si el DNA de origen se encuentra degradado en fragmentos pequeños, el DNA resultante se encuentre también degradado.

El tejido de herbario, a pesar de estar seco, conlleva más trabajo de laboratorio. También, la cantidad de tejido es limitada, y en nuestro caso, llevaba preservado por al menos quince años, lo que favorece la degradación del DNA. Kates y colaboradores (2021), analizaron muestras de herbario con secuenciación de lecturas cortas y encontraron que: (i) el herbario dónde se encuentran las muestras y el grupo taxonómico, al contener metabolitos secundarios que afectan el éxito de la secuenciación (ii) mientras más verde sea la hoja mayor, será el rendimiento de DNA, (iii) la extracción de DNA y la creación de bibliotecas influye en la calidad de secuenciación. Un punto relevante sobre la utilidad de las muestras de herbario radica en que hay plantas para las que su hábitat natural puede haber disminuido sensiblemente en los últimos años; además permite hacer muestreos geográficos, fenológicos, cronológicos entre otros, sin depender de muestreos adicionales.

Mapeo de las lecturas o reads contra el genoma de referencia

Las muestras con mayor porcentaje de mapeo respecto al genoma de referencia CM334 (Kim *et al.*, 2014), después de pasar los filtros de Stacks, llegaron hasta 78% de mapeo de lecturas. En otras solanáceas, como *Solanum pimpinellifolium* pariente silvestre del jitomate el porcentaje de mapeo fue de 84.4% (Celik *et al.*, 2017), y en la berenjena casi el 100% (Acquadro *et al.*, 2017), ambos con datos de GBS. Otros

cultivos, también con datos de GBS, como los frijoles *Phaseolus coccineus* y *P. dumosus*, tienen un porcentaje de mapeo superior al 80% en promedio (Guerra-García *et al.*, 2017; Guerra-García *et al.* en prep), vides el 88.2% (Laucou *et al.*, 2018) con genome-wide SNPs, en el durazno del 82.24 al 96.11% (Yu *et al.*, 2018; Vanderzande *et al.*, 2019) con genome-wide y SNPs y large-scale SNP.

El porcentaje de lecturas que se mapearon contra el genoma de referencia de este trabajo es relativamente bajo en comparación con las especies anteriormente mencionadas. Tales valores de mapeo podrían deberse a que *Capsicum* posee gran variabilidad genómica y a que el genoma de referencia de una variedad local de *Capsicum* no refleja, e incluso podría subestimar la variación total de la especie. Esto también se ha descrito en maíz y ajonjolí (Yang *et al.*, 2019; Wei *et al.*, 2016).

Otra posibilidad que podría influenciar el porcentaje de mapeo son las variantes estructurales (generalmente rearrreglos cromosómicos), las cuales podrían ser responsables de variación genética que hasta el momento no se ha detectado mediante los SNPs. Además de la gran cantidad de elementos transponibles que representan aproximadamente el 76-81% de secuencias repetitivas (Qin *et al.*, 2014; Kim *et al.*, 2014). Dicha variación se puede detectar usando métodos de secuenciación que produzcan fragmentos largos y la obtención de más genomas, y eventualmente el pangenoma de la especie (Tettelin *et al.*, 2005; Yang *et al.*, 2019).

Los estudios tipo pangenoma, al integrar la variación presente en diversos genomas incluyendo en este caso variedades cultivadas, locales y silvestres, contribuye a la integración de toda la diversidad genómica. Este término fue acuñado por Tettelin y colaboradores en el 2005. De tal forma que la implementación de los pangenomas posibilitará detallar la variación existente enmascarada por las variaciones estructurales (Golicz *et al.*, 2016).

Se recomienda específicamente para *Capsicum* la obtención de pangenomas para conocer las variaciones genómicas del acervo génico de la especie (Khan *et al.*, 2020). Los super-pangenomas, consisten en considerar especialmente la variación genómica de parientes silvestres, y de todas las especies de un género (Khan *et al.*, 2020), en particular considerando que nuestros datos no mapearon al genoma de

referencia en más de un 78%. Otro punto para considerar la implementación de pan y/o super-pangenomas es la extensa distribución de *C. annuum* var. *glabriusculum*, y el tamaño efectivo potencialmente grande de las formas cultivadas.

Es importante mencionar que si bien el mapeo se tiene un rango de 60% al 78% se eliminaron varias muestras que no presentaban buena calidad de mapeo. La mayoría de dichas muestras eran derivadas del tratamiento con REPLI-g de muestras de extracciones previas que no estaban dirigidas a estudios genómicos, ya que, aunque amplificaban el DNA, este posiblemente se encontraba demasiado degradado para ser útil para GBS.

Análisis genómico-poblacionales

El PCA considera que los componentes uno y dos explican la mayor parte de la varianza de los datos y encontramos un patrón claro, con tres grupos genéticos principales con base en la agrupación del PCA y otro grupo de origen heterogéneo. Cabe mencionar que al realizar un dendrograma con el método de UPGMA, se observaron los mismos cuatro grupos:

1. Norte de México: Sinaloa y Sonora
2. Centro de México: Querétaro
3. Sur de México: Oaxaca, Tabasco y Veracruz
4. Diverso: Una semisilvestre de Veracruz, una silvestre de Oaxaca, domesticadas: una de Yucatán, una de Oaxaca, una de Querétaro y dos de Zacatecas.

Los tres grupos principales muestran la existencia de diferenciación genómica en *C. annuum*. Aunque no se cubrió la totalidad de la distribución geográfica del chiltepín demostramos que existen al menos tres grupos: Norte, Centro y Sur, que pueden tomarse en cuenta para la conservación. Los resultados de la prueba de Mantel ($r = 0.447$, $p = 0.001$) parecen apoyar la diferenciación genética por factores geográficos, pues mostraron la relación entre las distancias geográfica y genética, por lo que podríamos asumir cierto grado de aislamiento por distancia, que refuerce las

diferencias originadas por otras variables como los factores bióticos y abióticos que definen las provincias biogeográficas. El aislamiento por distancia también se ha reportado en poblaciones silvestres y arvenses de *C. annuum* mediante microsatélites (González-Jara, 2011). La falta de un muestreo más extenso dentro y entre de las provincias no nos permite distinguir entre una diferenciación dada por procesos biogeográficos y un patrón de aislamiento por distancia. Aunque hay que recordar que la dispersión de las semillas de los chiltepines es por aves, y ellas en muchos casos no respetan los límites establecidos en las provincias biogeográficas.

De manera complementaria, el PCA coloreado por latitud, longitud y grado de domesticación, sugiere que la distribución de la diversidad genómica obedece a patrones latitudinales y longitudinales, en el que las muestras del grupo Sur presentan diferencias latitudinales y el Norte diferencias longitudinales. Por otro lado, el grupo más heterogéneo o Diverso agrupa varias muestras domesticadas, una semisilvestre y otra silvestre. Una posible explicación para la falta de un patrón geográfico dentro de este último grupo es la existencia de flujo génico promovido por las poblaciones humanas. Hernández-Verdugo (2018) propuso que entre los factores geográficos que influyen en la diferenciación de las poblaciones de chiles están la latitud, longitud y la distancia, lo que se ajusta a nuestros resultados. Existe una gran plasticidad fenotípica y diversidad morfológica y genética entre las poblaciones de *C. annuum* var. *glabriusculum*, lo que le permite proliferar en diferentes hábitats (Hernández-Verdugo, 2018).

Los tres grupos principales revelados por el PCA (Norte, Centro y Sur) son apoyados por el análisis de FastStructure (Figura 18), donde la K que mejor se ajustó a nuestros datos genómicos fue de 3. Es importante mencionar que, aunque corrimos el mismo análisis con $k = 4$ sólo se detectaron tres grupos, por lo que este parece el número más adecuado de grupos.

De manera complementaria, al análisis DAPC confirmó la presencia de los mismos tres grupos que el FastStructure: 1) Norte, 2) Centro, 3) Sur+Diverso. Específicamente el grupo Diverso agrupó a muestras mayoritariamente domesticadas que el DAPC. En el caso del grupo del Centro y Diverso, podemos observar que no todos los individuos se están ajustando a las elipses dibujadas en el DAPC. Es por ello que estos agrupamientos no sólo siguen un comportamiento geográfico, si no que

probablemente están influyendo otras fuerzas evolutivas, tales como el movimiento de semillas por seres humanos a diferentes regiones geográficas o comparten ancestría.

En los resultados del análisis FastStructure con $k=3$ las muestras de Chile fueron mayoritariamente homogéneas, pero algunos individuos mostraron mezcla genómica entre el Norte y el Centro de México. Podemos destacar que los chiles domesticados huacle, criollo, dulce, guajillo y mirasol se agruparon con las muestras silvestres y semisilvestres del norte de México, sin importar la localidad de colecta. Cabe mencionar que el Chile piquín (CaCOY2), al cual consideramos como semisilvestre se agrupa correctamente junto con otras muestras semisilvestres del grupo Sur.

Las muestras silvestres y semisilvestres se agruparon en su mayoría de acuerdo con cada grupo genómico: Norte, Centro y Sur de México, salvo por algunos casos. Las excepciones en el Sur fueron CaSDI2 (semisilvestre), CaSDI3 (silvestre) y CaVTR1 (silvestre). Las primeras dos muestras pertenecen a la localidad de Santo Domingo del Ingenio en Oaxaca y mostraron un ligero grado de mezcla genómica con el Norte, además de un mínimo porcentaje de mezcla con el Centro. Por otro lado, CaVTR1 exhibió una mayor composición genómica del grupo Norte y una menor del grupo Centro. En cuanto al Centro, la única excepción fue CaRES9, muestra domesticada del Rancho la Espadañuela en Querétaro, cuya totalidad de la mezcla genómica corresponde al Norte. Finalmente, en el Norte las muestras CaVPI2_1 (semisilvestre) y CaDZI1_1 (domesticado), de Tlapacoyan Veracruz y Dizibichaltún Yucatán, respectivamente, mostraron una mezcla genómica entre Norte y Centro, dominada por el Norte. Lo anterior puede ser indicativo de flujo génico reciente, posiblemente debido al manejo agrícola, aunque el patrón genómico claro sugiere que las muestras semisilvestres están en proceso de domesticación incipiente y retienen características fenotípicas similares a las silvestres y se agrupan genómicamente con ellas.

El análisis de FastStructure con $k=4$, mostró los mismos grupos genómicos que $k=3$, los cuales son: Norte, Centro y Sur de México. De tal forma que el cuarto grupo tiene tan poco porcentaje de mezcla genómica que no se percibe en la gráfica. Es así como los tres grupos del DAPC y FastStructure son los que contienen la mayoría de las mezclas genómicas, mientras que el grupo diverso que mostraron el PCA y el dendrograma sea un subconjunto del grupo Sur.

Para probar si se establecían los cuatro grupos genéticos iniciales del análisis de componentes principales, decidimos graficar los resultados obtenidos con FastStructure para $K=4$ (Fig. 21). En este caso, a pesar de que el análisis tuvo una k de 4 solo se perciben tres grupos: Norte (morado), Centro (verde) y Sur de México (rosa), de tal manera que no se visualiza el agrupamiento número 3 (naranja) en este análisis de FastStructure.

Diversidad y diferenciación genómica

El agrupamiento con un mayor número de alelos privados es el Sur (23,579,787), seguido del Centro (5,787), Norte (3,486) y por último el Diverso (3,281). El mayor número de alelos privados en el Sur parece estar relacionado con su bajo flujo génico (Slatkin, 1985), el cual se puede observar al comparar sus valores de F_{ST} (0.735-0.715) con el promedio (0.596). Así pues, mientras que las agrupaciones más similares fueron Centro y Diverso ($F_{ST} = 0.419$), las más de mayor estructura genómica fueron Norte y Sur ($F_{ST} = 0.735$). Éste análisis apoya los agrupamientos observados en el PCA y en el dendrograma, aunque no coinciden por completamente con el DPCA y el resultado de FastStructure, por lo que la ubicación del grupo Diverso no queda del todo clara, pero las diferencias entre Norte, Centro y Sur quedan reforzadas. La estructura genética de los cuatro grupos: Norte, Centro, Sur y Diverso, se puede explicar por factores que van desde los orográficos como las cadenas montañosas que se encuentran a lo largo de la distribución de *C. annuum*, la dispersión limitada de las aves y que las poblaciones de la planta se distribuyen de forma discontinua (González-Jara, *et al.*, 2011). Así mismo, el grupo con mayor diversidad nucleotídica y heterocigosis observada es también Sur (0.098 y 0.047 respectivamente) y el que presenta los valores menores es el Norte (0.037 y 0.012). De manera interesante, el grupo Diverso es el segundo en cuanto a diversidad nucleotídica (0.084) y heterocigosis observada (0.021), lo cual podría estar relacionado con su cercanía geográfica con el agrupamiento Sur.

Específicamente la heterocigosis total de *C. annuum* fue de 0.031 y entre los agrupamientos varió de 0.012-0.047. En otras especies como *Solanum melongena* o berengena nivel de heterocigosis en promedio fue 0.016 y en dos variedades de *S.*

melongena la heterocigosis fue <0.005 (Acquadro *et al.*, 2017). En *Solanum boliviense* o papa silvestre 0.053 (Bali *et al.*, 2018) y en *Solanum habrochaites* o jitomate silvestre fue 0.0006- 0.015 (Landis *et al.*, 2021). Comparando los chiles de este trabajo con el jitomate silvestre el nivel de heterocigosidad fue alto. El estadístico F_{ST} se ubicó en el rango de 0.419-0.735, lo que confirma la diferenciación genómica de los grupos.

CONCLUSIONES

1) La principal aportación del nuestro estudio fue la exploración de la estructuración y variación genómica en individuos con distintos grados de domesticación de *C. annuum* y nos permitió empezar a explorar la historia evolutiva de la especie desde el punto de vista genómico.

2) Estudiar a los parientes silvestres de los cultivos es crucial para la conservación de las plantas cultivadas, ya que los parientes silvestres contienen la mayor cantidad de la diversidad genética de la especie, suelen ser resistentes a plagas y a distintas condiciones ambientales.

3) Uno de los principales retos para el estudio genómico de los parientes silvestres es la obtención de buena calidad de DNA. La obtención del DNA en chiles sólo se había tratado en hojas de la planta, debido a que se considera que los frutos concentran grandes cantidades de capsaicina, lo que teóricamente las hacía inviables para la extracción de DNA, por la presencia de especies reactivas de oxígeno.

En este trabajo probamos diferentes tejidos: frescos, congelados y secos, que se pueden implementar para la obtención de DNA. Reportamos por primera vez que es posible la extracción de DNA para estudios genómicos de fruto de chile silvestre, que es relevante, ya que podríamos prescindir de los prolongados tiempos de germinación de las semillas para obtener hojas de las plantas recién germinadas.

4) Nuestros resultados demuestran la utilidad del kit de amplificación REPLI-g para genomas de plantas de más de tres gibabases de longitud, lo cual puede ser de ayuda para muestras en las que se tiene DNA escaso o para muestras de herbario con poco tejido disponible para la extracción.

5) En cuanto a los chiles silvestres y semi-silvestres, se detectaron tres grupos genómicos principales: Norte, Centro y Sur de México y uno adicional que agrupa a muestras domesticadas, una semi-silvestre y otra silvestre.

6) La mayor diversidad nucleotídica se encontró en el agrupamiento Sur y la menor en los grupos Diverso y Norte. Por otro lado se observó una gran diferenciación genética entre las poblaciones del Sur y Norte de acuerdo con los valores del índice F_{ST} .

PERSPECTIVAS

Como ya se mencionó, el porcentaje de alineamiento de nuestras secuencias contra el genoma de referencia de *C. annuum* fue bajo en comparación con el de otras solanáceas, por ello, a pesar de que se cuenta con un genoma de referencia (Kim et al., 2014), consideramos necesario secuenciar más genomas de chile para lograr la reconstrucción del pangenoma del chile silvestre, con lo que se podría integrar de manera más completa la diversidad y acervo genómico de la especie.

Por otro lado, consideramos que resultaría útil otros nuevos tejidos (como los tallos y las raíces) para la extracción de DNA genómico, con la finalidad de tener más alternativas para la obtención de de DNA que ayuden a evitar los largos tiempos de germinación y propagación de los ejemplares.

Asimismo, recomendamos ampliar el muestreo de poblaciones de chile a otros puntos de la República Mexicana, a otros estados que no fueron muestreados, al sur de Estados Unidos y hacia centro y Sudamérica, a fin de obtener un muestreo representativo a lo largo de toda su área de distribución. Con este muestreo, podríamos tener una valoración más fina de la influencia de las características bióticas y abióticas que determinan la variabilidad genética de sus poblaciones, así como detectar SNPs que se encuentren bajo selección natural que pueden resultar claves para la adaptación de la especie, que le permitan hacer frente a plagas o condiciones adversas de su medio, con lo que adicionalmente, se contribuiría con la seguridad alimentaria.

REFERENCIAS

- Acquadro, A., Barchi, L., Gramazio, P., Portis, E., Vilanova, S., Comino, C., ... Lanteri, S. (2017). Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLOS ONE*, 12(7), e0180774. doi: 10.1371/journal.pone.0180774.
- Aguilar-Meléndez, A., Morrell, P. L., Roose, M. L., y Kim, S. C. (2009). Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annum*; Solanaceae) from Mexico. *American Journal of Botany*, 96(6), 1190-1202.
- Aguirre Hernández, E., y Muñoz-Ocoter, V. (2015). El CHILE como alimento. *Ciencia*.
- Ahsanuddin, S., Afshinnekoo, E., Gandara, J., Hakyemezoğlu, M., Bezdán, D., Minot, S., ... Mason, C. E. (2017). Assessment of REPLI-g Multiple Displacement Whole Genome Amplification (WGA) Techniques for Metagenomic Applications. *Journal of Biomolecular Techniques: JBT*, 28(1), 46–55. doi: 10.7171/jbt.17-2801-008
- Altieri, M. A., Funes-Monzote, F. R., y Petersen, P. (2012). Agroecologically efficient agricultural systems for smallholder farmers: contributions to food sovereignty. *Agronomy for sustainable development*, 32(1), 1-13.
- Altieri, M. A., y Toledo, V. M. (2011). The agroecological revolution in Latin America: rescuing nature, ensuring food sovereignty and empowering peasants. *Journal of peasant studies*, 38(3), 587-612.
- Andrews, K., Good, J., Miller, M., Luikart, G., y Hohenlohe, P. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, 17. doi: 10.1038/nrg.2015.28
- Andrews, S. (2010). <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Anil Raj, Matthew Stephens y Jonathan K. Pritchard. *fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets*, (Genetics) Junio 2014 197:573-589

- Arteaga, M. C., Moreno-Letelier, A., Mastretta-Yanes, A., Vázquez-Lobo, A., Breña-Ochoa, A., Moreno-Estrada, A., Eguiarte L. E. y Piñero, D. (2016). Genomic variation in recently collected maize landraces from Mexico. *Genomics Data*, 7, 38-45.
- Badouin, H., Velt, A., Gindraud, F., Flutre, T., Dumas, V., Vautrin, S., ... Marais, G. A. B. (2020). The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. *Genome Biology*, 21(1), 223. doi: 10.1186/s13059-020-02131-y
- Bali, S., Robinson, B. R., Sathuvalli, V., Bamberg, J., y Goyer, A. (2018). Single nucleotide polymorphism (SNP) markers associated with high folate content in wild potato species. *PloS one*, 13(2), e0193415.
- Barrera-Redondo, J., Sanchez-de La Vega, G., Aguirre-Liguori, J. A., Castellanos-Morales, G., Gutiérrez-Guerrero, Y. T., Aguirre-Dugua, X., ... y Eguiarte, L. E. (2021). The domestication of *Cucurbita argyrosperma* as revealed by the genome of its wild relative. *Horticulture research*, 8.
- Barrera-Redondo, J., Vega, G. S. la, Aguirre-Liguori, J. A., Castellanos-Morales, G., Gutiérrez-Guerrero, Y. T., Aguirre-Dugua, X., ... Eguiarte, L. E. (2020). A wild *Cucurbita* genome reveals the role of structural variants and introgression in domestication. *BioRxiv*, 2020.10.15.341990. doi: 10.1101/2020.10.15.341990
- Bellon, M. R., Mastretta-Yanes, A., Ponce-Mendoza, A., Ortiz-Santamaría, D., Oliveros-Galindo, O., Perales, H., ... y Sarukhán, J. (2018). Evolutionary and food supply implications of ongoing maize domestication by Mexican campesinos. *Proceedings of the Royal Society B*, 285(1885), 20181049.
- Bellon, M. R., Mastretta-Yanes, A., Ponce-Mendoza, A., Ortiz-Santa María, D., Oliveros-Galindo, O., Perales, H., ... & Sarukhán, J. (2021). Beyond subsistence: the aggregate

- contribution of campesinos to the supply and conservation of native maize across Mexico. *Food Security*, 13(1), 39-53.
- Bernau, V. M., Jardón Barbolla, L., McHale, L. K., y Mercer, K. L. (2020). Germination response of diverse wild and landrace chile peppers (*Capsicum* spp.) under drought stress simulated with polyethylene glycol. *PloS one*, 15(11), e0236001.
- Bolger, A. M., Lohse, M., y Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.
- Bosland, P. W., y Votava, E. J. (2012). Peppers: Vegetable and Spice Capsicums. *Peppers: Vegetable and Spice Capsicums*, 2, 1–230.
- Carrizo García, C., Barfuss, M. H. J., Sehr, E. M., Barboza, G. E., Samuel, R., Moscone, E. A., y Ehrendorfer, F. (2016). Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, Solanaceae). *Annals of Botany*, 118(1), 35–51. doi: 10.1093/aob/mcw079
- Casas A, Blancas J, y Lira R. (2016). Mexican Ethnobotany: Interactions of People and Plants in Mesoamerica. En *Ethnobotany of Mexico interactions of people and plants in Mesoamerica*. (pp. 1-20). New York: Springer-Verlag.
- Casas, A., Otero-Arnaiz, A., Pérez-Negrón, E., y Valiente-Banuet, A. (2007). In situ Management and Domestication of Plants in Mesoamerica. *Annals of Botany*, 100(5), 1101–1115. doi: 10.1093/aob/mcm126
- Casas, A., y Parra, F. (2007). Agrobiodiversidad, parientes silvestres y cultura. Retrieved March 31, 2021, from Leisa website: <http://www.leisa-al.org/web/index.php/volumen-23-numero-2/1799-agrobiodiversidad-parientes-silvestres-y-cultura>
- Casas, A., y Vallejo, M. (2019). Agroecología y agrobiodiversidad. *Crisis ambiental en México*, 103.

- Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., ... Toll, J. (2016). Global conservation priorities for crop wild relatives. *Nature Plants*, 2(4), 1–6. doi: 10.1038/nplants.2016.22
- Catchen, J., Hohenlohe P., Bassham S., Amores A, y Cresko W. 2013. *Stacks: an analysis tool set for population genomics*. *Molecular Ecology*.
- Celik, I., Gurbuz, N., Uncu, A. T., Frary, A., y Doganlar, S. (2017). Genome-wide SNP discovery and QTL mapping for fruit quality traits in inbred backcross lines (IBLs) of *Solanum pimpinellifolium* using genotyping by sequencing. *BMC genomics*, 18(1), 1.
- Chappell, M. J., Wittman, H., Bacon, C. M., Ferguson, B. G., Barrios, L. G., Barrios, R. G., ... y Perfecto, I. (2013). Food sovereignty: an alternative paradigm for poverty reduction and biodiversity conservation in Latin America. *F1000Research*, 2.
- Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). (1997). 'Provincias biogeográficas de México'. Escala 1:4 000 000. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México, D. F.
- CONABIO, INIFAP, ICTA, DIBIO-MIAMBIENTE, Universidad de Birmingham y UICN. 2019a. Parientes silvestres de cultivos de Mesoamérica. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México.
- CONABIO, INIFAP, ICTA, DIBIO-MIAMBIENTE, Universidad de Birmingham y UICN. 2019b. 2019. Salvaguardar los parientes silvestres de cultivos mesoamericanos: Síntesis ejecutiva. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad-Unión Internacional para la Conservación de la Naturaleza, México.
- Convention on Biological Diversity (CBD). 2011. <https://www.cbd.int/sp/targets/rationale/target-13/>
- Cornille, A., Giraud, T., Smulders, M. J., Roldán-Ruiz, I., y Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57-65.

- Cornille, A., Gladieux, P., y Giraud, T. (2013). Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications*, 6(5), 737–748.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... y 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.
- Dempewolf, H., Eastwood, R. J., Guarino, L., Khoury, C. K., Müller, J. V., y Toll, J. (2014). Adapting Agriculture to Climate Change: A Global Initiative to Collect, Conserve, and Use Crop Wild Relatives. *Agroecology and Sustainable Food Systems*, 38(4), 369–377. doi: 10.1080/21683565.2013.870629
- Diniz-Filho, J. A. F., Soares, T. N., Lima, J. S., Dobrovolski, R., Landeiro, V. L., Telles, M. P. D. C., Rangel T. y Bini, L. M. (2013). Mantel test in population genetics. *Genetics and molecular biology*, 36, 475-485.
- Doebley, J. (1989). Isozymic Evidence and the Evolution of Crop Plants. In D. E. Soltis, P. S. Soltis, & T. R. Dudley (Eds.), *Isozymes in Plant Biology* (pp. 165–191). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-009-1840-5_9
- Dray, S., y Dufour, A.-B. (2007). Theade4Package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4). <https://doi.org/10.18637/jss.v022.i04>.
- Ersts, P.J. [Internet] Geographic Distance Matrix Generator(version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Available from http://biodiversityinformatics.amnh.org/open_source/gdmg.
- Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17: 27–32. doi:10.1111/1755-0998.12509

- Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the contrasting dynamics of annual and perennial domestication. *Trends in genetics*, 31(12), 709-719.
- Gepts, P. (2014). Domestication of Plants. In *Encyclopedia of Agriculture and Food Systems* (pp. 474–486). doi: 10.1016/B978-0-444-52512-3.00231-X
- Gepts, P. (2014). The contribution of genetic and genomic approaches to plant domestication studies. *Current opinion in plant biology*, 18, 51-59.
- Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., ... Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science*, 327(5967), 812–818. doi: 10.1126/science.1185383
- Goettsch, B., Urquiza-Haas, T., Koleff, P., Acevedo Gasman, F., Aguilar-Meléndez, A., Alavez, V., ... y Jenkins, R. K. (2021). Extinction risk of Mesoamerican crop wild relatives. *Plants, People, Planet*.
- Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., ... y Paterson, A. H. (2016). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature communications*, 7(1), 1-8.
- González-Jara, P., Moreno-Letelier, A., Fraile, A., Piñero, D., y García-Arenal, F. (2011). Impact of Human Management on the Genetic Variation of Wild Pepper, *Capsicum annum* var. *glabriusculum*. *PLoS ONE*, 6(12), e28715. doi: 10.1371/journal.pone.0028715
- Guerra-García, A., Rojas-Barrera, I. C., Ross-Ibarra, J., Papa, R., y Piñero, D. (2021). The genomic signature of wild-to-crop introgression during the domestication of scarlet runner bean (*Phaseolus coccineus* L.). *BioRxiv*.
- Guerra-García, A., Suárez-Atilano, M., Mastretta-Yanes, A., Delgado-Salinas, A., y Piñero, D. (2017). Domestication genomics of the open-pollinated scarlet runner bean (*Phaseolus coccineus* L.). *Frontiers in Plant Science*, 8, 1891.

- Guzmán, F. A., Ayala, H., Azurdia, C., Duque, M. C., & De Vicente, M. C. (2005). AFLP assessment of genetic diversity of *Capsicum* genetic resources in Guatemala: Home gardens as an option for conservation. *Crop Science*, 45(1), 363-370.
- Guzmán, F. A., Ayala, H., Azurdia, C., Duque, M. C., y Vicente, M. C. de. (2005). AFLP Assessment of Genetic Diversity of *Capsicum* Genetic Resources in Guatemala. *Crop Science*, 45(1), cropsoci2005.0363. doi: <https://doi.org/10.2135/cropsci2005.0363>
- Harlan, J. R. (1992). Crops and Man. 2nd ed.1992. By Jack R. Harlan. American Society of Agronomy, 677 S. Segoe Road, Madison, WI 53711. 284 pp. *American Journal of Alternative Agriculture*, 8(1), 47–48. doi: 10.1017/S0889189300004938
- Harlan, J. R., y de Wet, J. M. (1971). Toward a Rational Classification of Cultivated Plants. *TAXON*, 20(4), 509–517. doi: <https://doi.org/10.2307/1218252>
- Harlan, Jack R. (1971). Agricultural Origins: Centers and Noncenters. *Science*, 174(4008), 468–474. doi: 10.1126/science.174.4008.468
- Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., ... Feinberg, A. P. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neuroscience*, 15(10), 1371–1373. doi: 10.1038/nn.3218
- Hernández-Terán, A., Navarro-Díaz, M., Benítez, M., Lira, R., Wegier, A., y Escalante, A. E. (2020). Host genotype explains rhizospheric microbial community composition: the case of wild cotton metapopulations (*Gossypium hirsutum* L.) in Mexico. *FEMS Microbiology Ecology*, 96(8), fiae109.
- Hernández-Verdugo S. (2018). El chile silvestre. Ecología, evolución y genética. *Universidad Autónoma de Sinaloa*.
- Hernández-Verdugo, S., Dávila, P., y Oyama, K. (1999). Síntesis del conocimiento taxonómico, origen y domesticación del género *Capsicum*. *Botanical Sciences*, 65. doi: 10.17129/botsci.1583

- Hernández-Verdugo, S., González-Sánchez, R. A., Porras, F., Parra-Terraza, S., Valdez-Ortiz, A., Pacheco-Olvera, A., y López-España, R. G. (2015). Plasticidad fenotípica de poblaciones de chile silvestre (*Capsicum annum* var. *glabriusculum*) en respuesta a disponibilidad de luz. *Botanical Sciences*, 93(2). doi: 10.17129/botsci.237
- Hernández-Verdugo, S., Guevara-Gonzalez, R., Rivera-Bustamante, R., Vázquez-Yañes, C., y Oyama, K. (1998). Los parientes silvestres del chile (*Capsicum* spp.) como recursos genéticos. *Boletín de La Sociedad Botánica de México*, 62, 171–181. doi: 10.17129/botsci.1559
- Hernández-Verdugo, S., Porras, F., Pacheco-Olvera, A., López-España, R. G., Villarreal-Romero, M., Parra-Terraza, S., y Osuna Enciso, T. (2012). Caracterización y variación ecogeográfica de poblaciones de chile (*Capsicum annum* var. *glabriusculum*) silvestre del noroeste de México. *Polibotánica*, (33), 175-191.
- Hernández, S. (2011). El chile. Protagonista de la Independencia y la Revolución. Fundación Herdez.
- Hoban, S., Bruford, M., D'Urban Jackson, J., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., ... Laikre, L. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, 248, 108654. doi: 10.1016/j.biocon.2020.108654
- Hothorn, T., Bretz, F., y Westfall, P. (2008). Simultaneous Inference in General Parametric Models. In *Biometrical Journal* (Vol. 50, Issue 3, pp. 346–363).
- Jain, S. K. (1976). The evolution of inbreeding in plants. *Annual review of ecology and systematics*, 7(1), 469-495.
- Jain, S. K. (1976). The evolution of inbreeding in plants. *Annual review of ecology and systematics*, 7(1), 469-495.

- Jardón Barbolla, L. (2015). De la evolución al valor de uso, ida y vuelta: exploraciones en la domesticación y diversificación de plantas. *INTERdisciplina*, 3(5).
- Jardón Barbolla, L. (2021). Líneas de conflicto en la apropiación de los recursos agrobiológicos: las contrastantes expediciones de Vavilov y Brücher. *Inter disciplina*, 9(24), 73-95.
- Jarvis, A., Lane, A., y Hijmans, R. J. (2008). The effect of climate change on crop wild relatives. *Agriculture, Ecosystems & Environment*, 126(1), 13–23. doi: 10.1016/j.agee.2008.01.013
- Jombart T. (2008). “adegenet: a R package for the multivariate analysis of genetic markers.” *Bioinformatics*, **24**, 1403-1405.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405.
- Jombart, T., Devillard, S., y Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, 11(1), 1-15.
- Jombart, T., y Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070-3071.
- Jombart, T., y Collins, C. (2015). A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0. *London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling*.
- Kates, H. R., Doby, J. R., Siniscalchi, C. M., LaFrance, R., Soltis, D. E., Soltis, P. S., ... y Folk, R. A. (2021). The Effects of Herbarium Specimen Characteristics on Short-Read NGS Sequencing Success in Nearly 8000 Specimens: Old, Degraded Samples Have Lower DNA Yields but Consistent Sequencing Success. *Frontiers in plant science*, 12, 1076.

- Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D., y Varshney, R. K. (2020). Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends in plant science*, 25(2), 148-158.
- Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J. M., Lee, H.-A., ... Choi, D. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature Genetics*, 46(3), 270–278. doi: 10.1038/ng.2877
- Kraft, K. H., Brown, C. H., Nabhan, G. P., Luedeling, E., Ruiz, J. de J. L., d'Eeckenbrugge, G. C., ... Gepts, P. (2014). Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annum*, in Mexico. *Proceedings of the National Academy of Sciences*, 111(17), 6165–6170. doi: 10.1073/pnas.1308933111
- Laborde Cancino JA y Pozo Campodónico O. *Present and past of chile in Mexico*.(1982). Ministry of Agriculture and Water Resources, National Agricultural Research Institute. Mexico, DF.
- Landis, J. B., Miller, C. M., Broz, A. K., Bennett, A. A., Carrasquilla-Garcia, N., Cook, D. R., ... y Moghe, G. D. (2021). Migration through a major Andean ecogeographic disruption as a driver of genetic and phenotypic diversity in a wild tomato species. *Molecular biology and evolution*, 38(8), 3202-3219.
- Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A. F., Bérard, A., ... y Le Paslier, M. C. (2018). Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS one*, 13(2), e0192540.
- Li H. y Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-60.
- Li, Q., Guo, R., Li, Y., Hartman, W. H., Li, S., Zhang, Z., ... Wang, H. (2019a). Insight into the Bacterial Endophytic Communities of Peach Cultivars Related to Crown Gall Disease

- Resistance. *Applied and Environmental Microbiology*, 85(9), e02931-18. doi: 10.1128/AEM.02931-18
- Li, Y., Cao, K., Zhu, G., Fang, W., Chen, C., Wang, X., ... y Wang, L. (2019b). Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome biology*, 20(1), 1-18.
- Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., ... Huang, S. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nature Genetics*, 46(11), 1220–1226. doi: 10.1038/ng.3117
- Lira, R., Casas, A., y Blancas, J. (Eds.). (2016). *Ethnobotany of Mexico: Interactions of People and Plants in Mesoamerica*. New York: Springer-Verlag. doi: 10.1007/978-1-4614-6669-7
- Liu, P., Chen, W., y Chen, J. P. (2019). Viral metagenomics revealed Sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*). *Viruses*, 11(11), 979. doi.org/10.3390/v11110979
- Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., ... Tian, Z. (2020). Pan-Genome of Wild and Cultivated Soybeans. *Cell*, 182(1), 162-176.e13. doi: 10.1016/j.cell.2020.05.023
- Lizarde A., N., Lizarde A., E., y Martínez, J. G. (2011). Evaluación de la germinación y crecimiento de Plántula de Chiltepín (*Capsicum annuum* L variedad *glabriusculum*) en invernadero. *Revista Colombiana de Biotecnología*, 13(2), 170–175.
- Loaiza-Figueroa, F., Ritland, K., Cancino, J. A. L., & Tanksley, S. D. (1989). Patterns of genetic variation of the genus *Capsicum* (Solanaceae) in Mexico. *Plant Systematics and Evolution*, 165(3), 159-188.
- Loaiza-Figueroa, F., Ritland, K., Cancino, J. A. L., y Tanksley, S. D. (1989). Patterns of genetic variation of the genus *Capsicum* (Solanaceae) in Mexico. *Plant Systematics and Evolution*, 165(3/4), 159–188.

- Lu, J., Tang, T., Tang, H., Huang, J., Shi, S., y Wu, C.-I. (2006). The accumulation of deleterious mutations in rice genomes: A hypothesis on the cost of domestication. *Trends in Genetics*, 22(3), 126–131. doi: 10.1016/j.tig.2006.01.004
- Lyuri, D., & Lands, A. A. (2008). A Short History of Agriculture. *Encyclopedia of Ecology*, 76.
- Mariette, S., Tavaud, M., Arunyawat, U., Capdeville, G., Millan, M., y Salin, F. (2010). Population structure and genetic bottleneck in sweet cherry estimated with SSRs and the gametophytic self-incompatibility locus. *Bmc Genetics*, 11, 1-13.
- Marrano, A., Micheletti, D., Lorenzi, S., Neale, D., y Grandó, M. S. (2018). Genomic signatures of different adaptations to environmental stimuli between wild and cultivated *Vitis vinifera* L. *Horticulture Research*, 5(1), 1–12. doi: 10.1038/s41438-018-0041-2
- Mastretta-Yanes, A., Acevedo Gasman, F., Burgeff, C., Cano Ramírez, M., Piñero, D., y Sarukhán, J. (2018). An Initiative for the Study and Use of Genetic Diversity of Domesticated Plants and Their Wild Relatives. *Frontiers in Plant Science*, 9. doi: 10.3389/fpls.2018.00209
- Mastretta-Yanes, A., Bellon, M. R., Acevedo, F., Burgeff, C., Piñero, D., y Sarukhán, J. (2019). Un programa para México de conservación y uso de la diversidad genética de las plantas domesticadas y sus parientes silvestres. *Revista fitotecnia mexicana*, 42(4), 321-334.
- Maxted, N., Ford-Lloyd, B. V., Jury, S., Kell, S., y Scholten, M. (2006). Towards a definition of a crop wild relative. *Biodiversity & Conservation*, 15(8), 2673–2685. doi: 10.1007/s10531-005-5409-6
- Maxted, N., Kell, S., Ford-Lloyd, B., Dulloo, E., y Toledo, Á. (2012). Toward the Systematic Conservation of Global Crop Wild Relative Diversity. *Crop Science*, 52(2), 774–785. doi: <https://doi.org/10.2135/cropsci2011.08.0415>

- McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., ... Zamir, D. (2013). Feeding the future. *Nature*, 499(7456), 23–24. doi: 10.1038/499023a
- McLeod, M. J., Guttman, S. I., y Eshbaugh, W. H. (1982). Early evolution of chili peppers (*Capsicum*). *Economic Botany*, 36(4), 361–368. doi: 10.1007/BF02862689
- McLeod, Michael J., Guttman, S. I., Eshbaugh, W. H., y Rayle, R. E. (1983). An Electrophoretic Study of Evolution in *Capsicum* (Solanaceae). *Evolution*, 37(3), 562–574. doi: <https://doi.org/10.1111/j.1558-5646.1983.tb05573.x>
- Mendell, J. R., Shilling, C., Leslie, N. D., Flanigan, K. M., al-Dahhak, R., Gastier-Foster, J., ... y Weiss, R. B. (2012). Evidence-based path to newborn screening for Duchenne muscular dystrophy. *Annals of neurology*, 71(3), 304-313.
- Meyer, R. S., DuVal, A. E., y Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist*, 196(1), 29-48.
- Meyer, R. S., y Purugganan, M. D. (2013). Evolution of crop species: Genetics of domestication and diversification. *Nature Reviews Genetics*, 14(12), 840–852. doi: 10.1038/nrg3605
- Miller, J. M., Cullingham, C. I., y Peery, R. M. (2020). The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity*, 125(5), 269-280.
- Moreno-Letelier, A., Aguirre-Liguori, J. A., Piñero, D., Vázquez-Lobo, A., y Eguiarte, L. E. (2020). The relevance of gene flow with wild relatives in understanding the domestication process. *Royal Society open science*, 7(4), 191545
- Moscone, E. A., Scaldaferrò, M., Grabielle, M., Cecchini, N., Sánchez García, Y., Jarret, R., ... Ehrendorfer, F. (2006). The evolution of Chili Peppers (*Capsicum*-Solanaceae): A cytogenetic perspective. *Acta Horticulturae*. doi: 10.17660/ActaHortic.2007.745.5

- Moyers, B. T., Morrell, P. L., y McKay, J. K. (2018). Genetic Costs of Domestication and Improvement. *Journal of Heredity*, 109(2), 103–116. doi: 10.1093/jhered/esx069
- Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., Prins B., Reynolds A., Jer-Ming C., Ware D., Bustamente C. y Buckler, E. S. (2011). Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences*, 108(9), 3530-3535.
- Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., y Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular ecology*, 22(11), 2841.
- Nuez, F., Ortega, G., y Costa, R. (1996). *El cultivo de pimientos, chiles y ajíes* (No. 633.84 N8897 Ej. 1 010957). Mundi Prensa,.
- O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., y Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists.
- Oksanen J., Blanchet F, Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P., O'Hara R., Simpson G., Solymos P., Stevens H., Szoecs E y Wagner H. (2020). vegan: Community Ecology Package. R package version 2.5-7. <https://CRAN.R-project.org/package=vegan>.
- Pagán, I., González-Jara, P., Moreno-Letelier, A., Rodelo, M., Fraile, A., Piñero, D., y García-Arenal, F. (2012). Effect of Biodiversity Changes in Disease Risk: Exploring Disease Emergence in a Plant-Virus System. *PLoS Pathogens*, 8, e1002796. doi: 10.1371/journal.ppat.1002796
- Paradis E., Claude J. y Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289-290.

- Peischl, S., Dupanloup, I., Kirkpatrick, M., y Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, 22(24), 5972–5982. doi: <https://doi.org/10.1111/mec.12524>
- Perry, L., Dickau, R., Zarrillo, S., Holst, I., Pearsall, D. M., Piperno, D. R., ... Zeidler, J. A. (2007). Starch Fossils and the Domestication and Dispersal of Chili Peppers (*Capsicum* spp. L.) in the Americas. *Science*, 315(5814), 986–988. doi: 10.1126/science.1136914
- Pickersgill B. (1991). Cytogenetics and evolution of *Capsicum* L. En Tsuchiya T. y Gupta P. K. (Eds.), *Chromosome engineering in plants, part B* (pp. 139-160). Elsevier.
- Pickersgill B. (2016). Chile Peppers (*Capsicum* spp.). En *Ethnobotany of Mexico interactions of people and plants in Mesoamerica*. (pp. 1-20). New York: Springer-Verlag.
- Pickersgill, B. (1971). Relationships Between Weedy and Cultivated Forms in Some Species of Chili Peppers (Genus *Capsicum*). *Evolution*, 25(4), 683–691. doi: 10.2307/2406949
- Pickersgill, B. (1984). Migration of Chili Peppers, *Capsicum* spp, in the Americas. *Pre-Columbian Plant Migration*. Retrieved from <https://ci.nii.ac.jp/naid/10018105522/>
- Pickersgill, B. (1997). Genetic resources and breeding of *Capsicum* spp. *Euphytica*, 96(1), 129–133. doi: 10.1023/A:1002913228101
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... y Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), 559-575.
- Qiagen. 2021a. <https://www.qiagen.com/mx/service-and-support/learning-hub/technologies-and-research-topics/wga/overview-on-wga/>
- Qiagen. 2021b. <https://www.qiagen.com/mx/service-and-support/learning-hub/technologies-and-research-topics/wga/ngs/>
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., ... Zhang, Z. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum*

- domestication and specialization. *Proceedings of the National Academy of Sciences*, 111(14), 5135–5140. doi: 10.1073/pnas.1400975111
- Qn, X., Q, S., Jy, Z., Yl, Z., L, L., Xl, L., y Ct, L. (2019). Evaluating the Efficiency of REPLI-g® Single Cell Kit for Trace DNA Amplification. *Fa Yi Xue Za Zhi*, 35(2), 210–215. doi: 10.12116/j.issn.1004-5619.2019.02.015
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org/index.html>.
- Raju, P. S., Chauhan, O. P., y Bawa, A. S. (2010). Chili Flavor. In *Handbook of Fruit and Vegetable Flavors* (pp. 775–801). John Wiley & Sons, Ltd. doi: 10.1002/9780470622834.ch41
- Razali, R., Bougouffa, S., Morton, M. J. L., Lightfoot, D. J., Alam, I., Essack, M., ... Negrão, S. (2018). The Genome Sequence of the Wild Tomato *Solanum pimpinellifolium* Provides Insights Into Salinity Tolerance. *Frontiers in Plant Science*, 9. doi: 10.3389/fpls.2018.01402
- Razifard, H., Ramos, A., Della Valle, A. L., Bodary, C., Goetz, E., Manser, E. J., Li X, Zhang L. Visa S., Tieman D, van der Knaap E y Caicedo, A. L. (2020). Genomic evidence for complex domestication history of the cultivated tomato in Latin America. *Molecular biology and evolution*, 37(4), 1118-1132.
- Rendón-Aguilar, B., Bernal-Ramírez, L. A., y Sánchez-Reyes., G. A. (2017). Las plantas arvenses: Más que hierbas del campo. *Oikos*, 17:30-34.
- Ruse, M. (1975). Charles Darwin and artificial selection. *Journal of the History of Ideas*, 36(2), 339-350.
- Särkinen, T., Bohs, L., Olmstead, R. G., y Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evolutionary Biology*, 13(1), 214. doi: 10.1186/1471-2148-13-214

- Sgrò, C. M., Lowe, A. J., y Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, 4(2), 326–337. doi: <https://doi.org/10.1111/j.1752-4571.2010.00157.x>
- Singh, S., Asad, S. F., Ahmad, A., Khan, N. U., y Hadi, S. M. (2001). Oxidative DNA damage by capsaicin and dihydrocapsaicin in the presence of Cu (II). *Cancer letters*, 169(2), 139-146.
- Slatkin, M. (1985). Rare alleles as indicators of gene flow. *Evolution*, 39(1), 53-65.
- Sosa, V., De-Nova, J. A., y Vásquez-Cruz, M. (2018). Evolutionary history of the flora of Mexico: Dry forests cradles and museums of endemism. *Journal of Systematics and Evolution*, 56(5), 523–536. doi: <https://doi.org/10.1111/jse.12416>
- Spielman, D., Brook, B. W., y Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences*, 101(42), 15261–15264. doi: 10.1073/pnas.0403809101
- Stepanauskas, R., y Sieracki, M. E. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proceedings of the National Academy of Sciences*, 104(21), 9052–9057. doi: 10.1073/pnas.0700496104
- Stitzer, M. C., y Ross-Ibarra, J. (2018). Maize domestication and gene interaction. *New phytologist*, 220(2), 395-408.
- Taitano, N., Bernau, V., Jardón-Barbolla, L., Leckie, B., Mazourek, M., Mercer, K., ... y van der Knaap, E. (2019). Genome-wide genotyping of a novel mexican chile pepper collection illuminates the history of landrace differentiation after *Capsicum annum* L. domestication. *Evolutionary applications*, 12(1), 78-92.
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., ... y DeBoy, R. T. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus*

agalactiae: implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences*, 102(39), 13950-13955.

Tobón-Niedfeldt, W., Mastretta-Yanes, A., Urquiza-Haas, T., Goettsch, B., Cuervo-Robayo, A. P., Urquiza-Haas, E., ... Koleff, P. (2021). Incorporating evolutionary and threat processes into crop wild relatives conservation. *BioRxiv*, 2021.06.15.448560. doi: 10.1101/2021.06.15.448560

Tsuchiya T. y Gupta P. K. (1991). Chromosome engineering in plants, part B

Turchetto-Zolet, A. C., Pinheiro, F., Salgueiro, F., y Palma-Silva, C. (2013). Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, 22(5), 1193–1213. doi: <https://doi.org/10.1111/mec.12164>

Universidad de Wisconsin, Centro de Biotecnología.
2021. <https://www.biotech.wisc.edu/services/dnaseq>

Universidad de Wisconsin, Centro de Biotecnología.
2021. <https://www.biotech.wisc.edu/services/dnaseq>

Vanderzande, S., Howard, N. P., Cai, L., Da Silva Linge, C., Antanaviciute, L., Bink, M. C., ... y Van de Weg, E. (2019). High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. *PloS one*, 14(6), e0210928.

Vavilov N. 1920. The law of homologous series in variation. Proceedings of the III All-Russian

Vavilov N. 1922. The law of homologous series in variation. *Journal of Genetics* 12: 47–89.

Vavilov Nicolai I. The origin, variation, immunity and breeding of cultivated plants. EUA: *Chronica Botanica*, 1951.

Votava, E. J., Nabhan, G. P., & Bosland, P. W. (2002). Genetic diversity and similarity revealed via molecular analysis among and within an *in situ* population and ex situ accessions of chiltepin (*Capsicum annuum* var. *glabriusculum*). *Conservation Genetics*, 3(2), 123-129.

- Votava, E. J., Nabhan, G. P., y Bosland, P. W. (2002). Genetic diversity and similarity revealed via molecular analysis among and within an *in situ* population and ex situ accessions of chiltepín (*Capsicum annuum* var. *glabriusculum*). *Conservation Genetics*, 3(2), 123–129. doi: 10.1023/A:1015216504565
- Walters, C., Berjak, P., Pammenter, N., Kennedy, K., y Raven, P. (2013). Preservation of Recalcitrant Seeds. *Science*, 339(6122), 915–916. doi: 10.1126/science.1230935
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., ... & Leung, H. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*, 557(7703), 43-49.
- Wei, X., Zhu, X., Yu, J., Wang, L., Zhang, Y., Li, D., ... y Zhang, X. (2016). Identification of sesame genomic variations from genome comparison of landrace and variety. *Frontiers in plant science*, 7, 1169.
- Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Xie, M., Chung, C. Y.-L., Li, M.-W., Wong, F.-L., Wang, X., Liu, A., ... Lam, H.-M. (2019). A reference-grade wild soybean genome. *Nature Communications*, 10(1), 1216. doi: 10.1038/s41467-019-09142-9
- Yang, K., Pyo, J., Kim, G. Y., Yu, R., Ju, S., Kim, W., y Kim, B. S. (2009). Capsaicin induces apoptosis by generating reactive oxygen species and disrupting mitochondrial transmembrane potential in human colon cancer cell lines. *Cellular and Molecular Biology Letters*, 14(3), 497-510.
- Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., ... y Wang, Y. (2019). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nature genetics*, 51(6), 1052-1059.

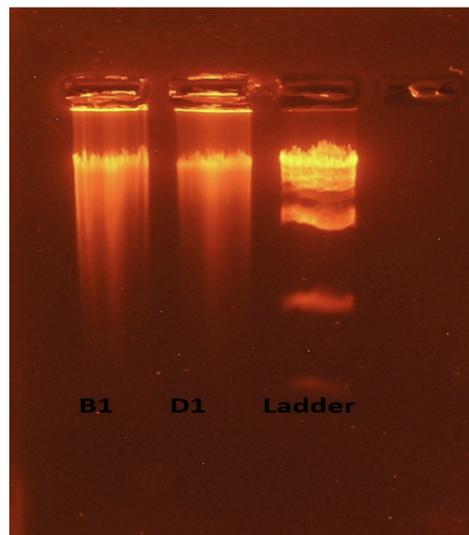
Yu, Y., Fu, J., Xu, Y., Zhang, J., Ren, F., Zhao, H., ... y Jiang, D. (2018). Genome re-sequencing reveals the evolutionary history of peach fruit edibility. *Nature communications*, 9(1), 1-13.

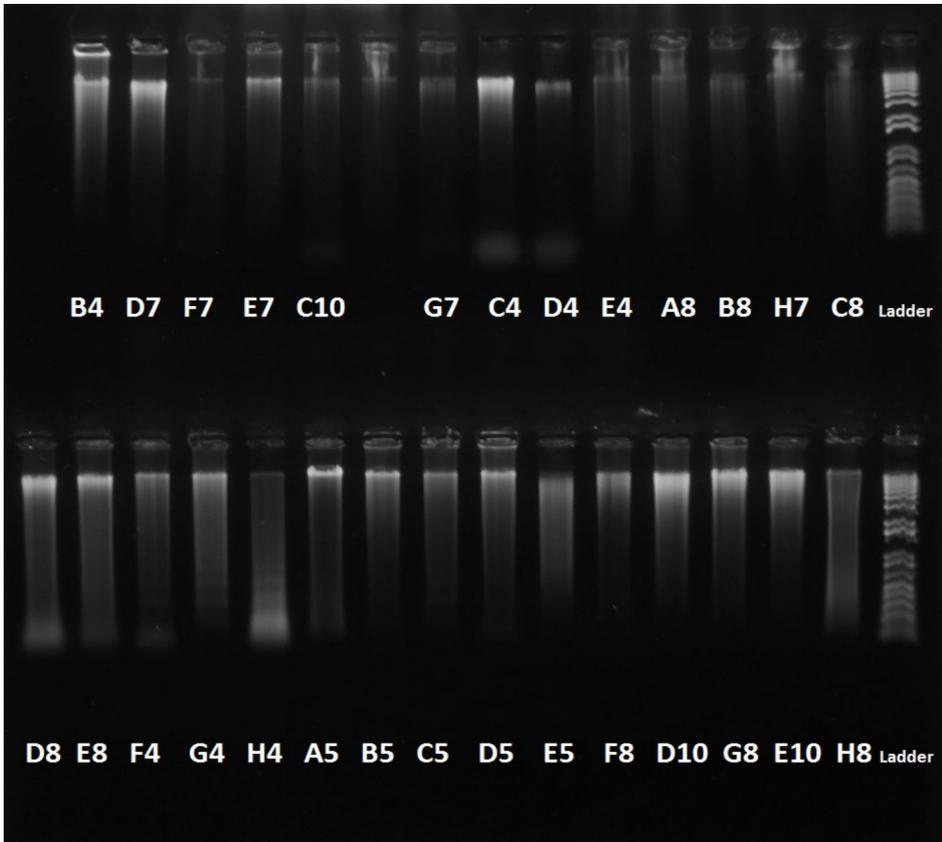
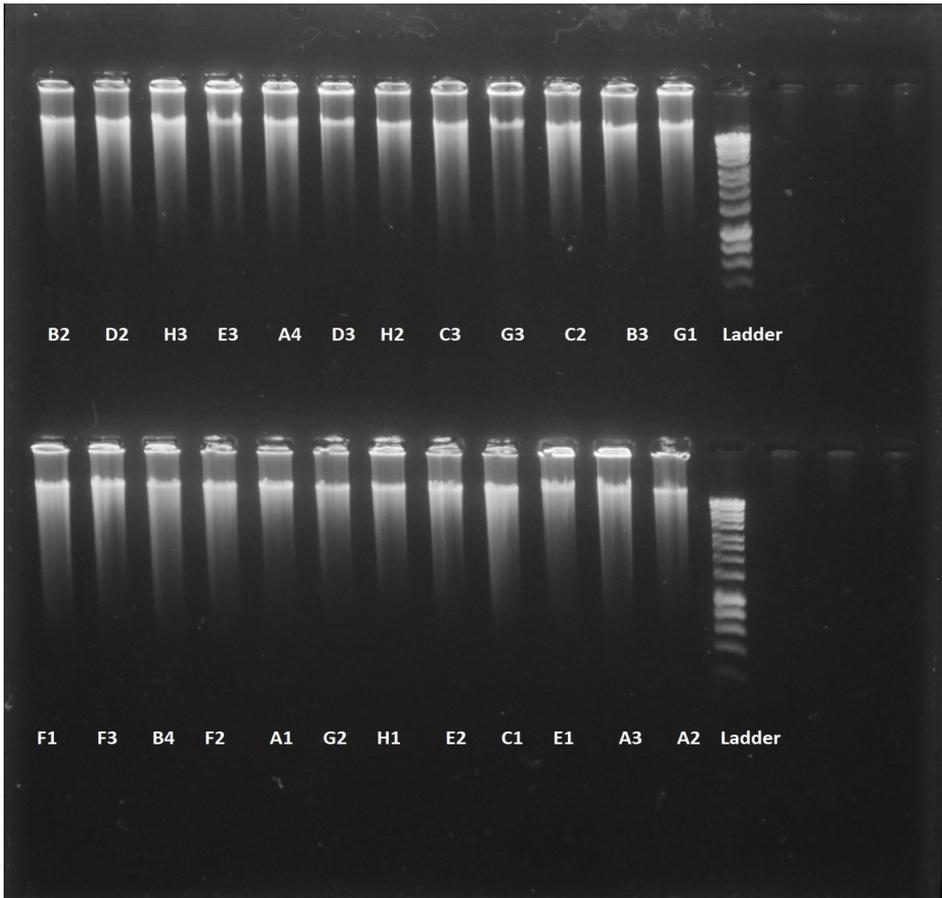
Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., ... Huang, X. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics*, 50(2), 278–284. doi: 10.1038/s41588-018-0041-z.

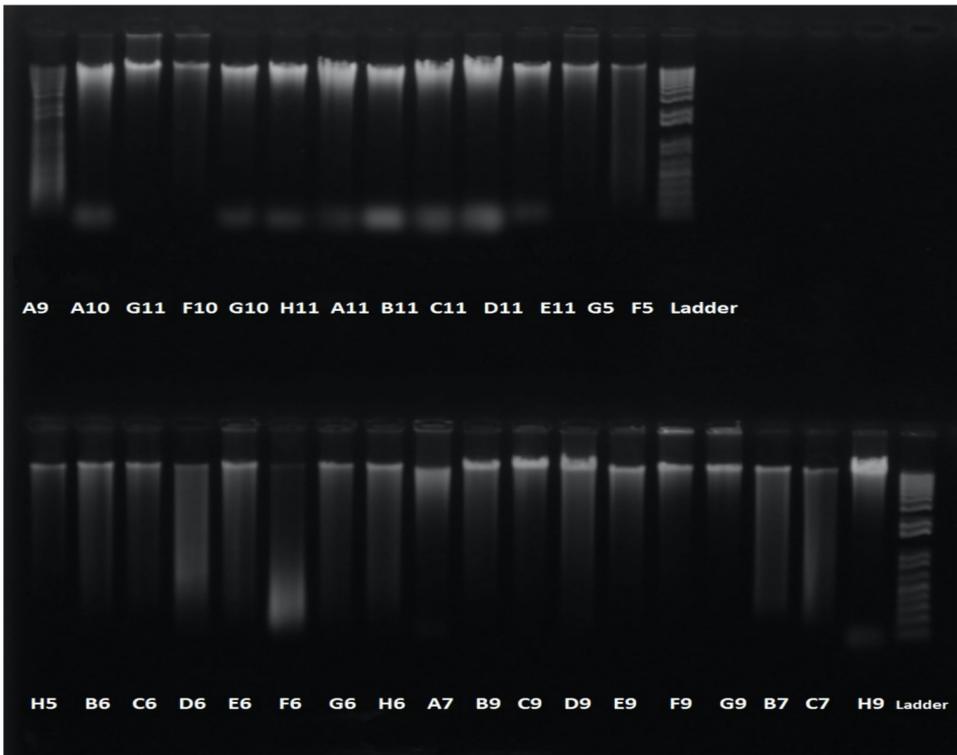
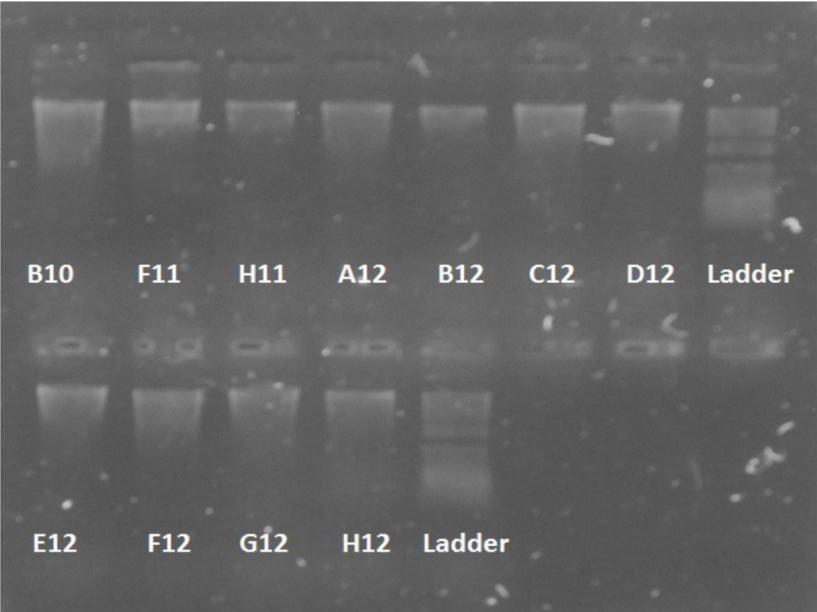
Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., ... Varshney, R. K. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics*, 51(5), 865–876. doi: 10.1038/s41588-019-0402-2.

ANEXOS

ANEXO 1. Geles de agarosa de los individuos secuenciados de *Capsicum annuum*, en los cuales se indica el individuo de acuerdo con cómo fueron cargados en la placa de secuenciación.







ANEXO 2. Muestras a las que se les aplicó REPLI-g, como método de recuperación de DNA. Así como la cuantificación de Qubit antes y después de REPLI-g. En la columna de estatus de domesticación la S=Silvestre, A=Semi-silvestres, D=Domesticado.

Posición en la placa	Muestra	Estado	Localidad	Estatus de domesticación	Concentración (ng/μl)	
					Antes de REPLIg	Después de REPLIg
A1	CaMOC15	Sonora	Moctezuma	S	18.2	97.9
B1	CaMOC25	Sonora	Moctezuma	S	3.34	35.9
C1	CaMOC30	Sonora	Moctezuma	S	14.2	69.6
D1	CaMOC37	Sonora	Moctezuma	S	7.86	29.3
E1	CaMAU7	Sonora	Los_Mautos	S	6.5	106
F1	CaMAU9	Sonora	Los_Mautos	S	17.8	39.9
G1	CaMAU14	Sonora	Los_Mautos	S	6.98	106
H1	CaMAU25	Sonora	Los_Mautos	S	5.06	44.1
A2	CaBER21	Querétaro	Bernal	S	3.24	53.4
B2	CaBER23	Querétaro	Bernal	S	Muy bajo	40.5
C2	CaBER33	Querétaro	Bernal	S	2.44	80.3
D2	CaBER36	Querétaro	Bernal	S	6.26	72.4
E2	CaHUU4	Sinaloa	El_Huajote	S	16.4	32.8
F2	CaHUU6	Sinaloa	El_Huajote	S	8.44	87.4
G2	CaELO15	Sinaloa	Elota	A	3.08	86.6
H2	CaPOT9	Sinaloa	El_Potrero	D	3.4	47.6

A3	CaSAN1	Sinaloa	Sanalona	A	11.6	127
B3	CaSAN2	Sinaloa	Sanalona	A	2.86	30.2

Anexo 3. Valores de la medición de la concentración de DNA realizada con Qubit en ng/μl para cada muestra. En la columna de estatus de domesticación la S=Silvestre, A=Semi-silvestre, D=Domesticado.

Posición placa	Muestra	Estado	Localidad	Estatus	Origen del DNA	Concentración (ng/μl)
A1	CaMOC15	Sonora	Moctezuma	S	REPLIg	97.9
B1	CaMOC25	Sonora	Moctezuma	S	REPLIg	35.9
C1	CaMOC30	Sonora	Moctezuma	S	REPLIg	69.6
D1	CaMOC37	Sonora	Moctezuma	S	REPLIg	29.3
E1	CaMAU7	Sonora	Los Mautos	S	REPLIg	106
F1	CaMAU9	Sonora	Los Mautos	S	REPLIg	39.9
G1	CaMAU14	Sonora	Los Mautos	S	REPLIg	106
H1	CaMAU25	Sonora	Los Mautos	S	REPLIg	44.1
A2	CaBER21	Querétaro	Bernal	S	REPLIg	53.4
B2	CaBER23	Querétaro	Bernal	S	REPLIg	40.5
C2	CaBER33	Querétaro	Bernal	S	REPLIg	80.3
D2	CaBER36	Querétaro	Bernal	S	REPLIg	72.4
E2	CaHUJ4	Sinaloa	El Huajote	S	REPLIg	32.8
F2	CaHUJ6	Sinaloa	El Huajote	S	REPLIg	87.4
G2	CaELO15	Sinaloa	Elota	A	REPLIg	86.6
H2	CaPOT9	Sinaloa	El Potrero	D	REPLIg	47.6
A3	CaSAN1	Sinaloa	Sanalona	A	REPLIg	127
B3	CaSAN2	Sinaloa	Sanalona	A	REPLIg	30.2
C3	CaTEM1	Sonora	Temporal	D	REPLIg	29.6
D3	CaTEM6	Sonora	Temporal	D	REPLIg	97.9
E3	CaTEM7	Sonora	Temporal	D	REPLIg	157
F3	CaTEM8	Sonora	Temporal	D	REPLIg	88.1
G3	CaTEM9	Sonora	Temporal	D	REPLIg	46.3
H3	CaTEM11	Sonora	Temporal	D	REPLIg	150
A4	CaTEM16	Sonora	Temporal	D	REPLIg	112
B4	CaCHS1	Chihuahua	Chihuahua	A	Invernadero	164
C4	CaSDI19_1 R	Oaxaca	Santo Domingo del Ingenio	S	Invernadero	53.9
D4	CaSDI19_2 R	Oaxaca	Santo Domingo del Ingenio	S	Invernadero	80.2

E4	CaSDI3	Oaxaca	Santo Domingo del Ingenio	S	Hoja congelada	23.7
F4	CaTAN8	Querétaro	Tancamá	S	Tejido seco	22.7
G4	CaTAN10	Querétaro	Tancamá	S	Tejido seco	35.5
H4	CaTAN13	Querétaro	Tancamá	S	Tejido seco	26.1
A5	CaTOL1	Querétaro	Tolimán	A	Tejido seco	24.6
B5	CaRES5	Querétaro	Rancho la Espadañuela	S	Tejido seco	30.1
C5	CaRES6	Querétaro	Rancho la Espadañuela	S	Tejido seco	23.2
D5	CaRES9	Querétaro	Rancho la Espadañuela	D	Tejido seco	28.4
E5	CaSOT1	Tamaulipas	Soto la Marina	S	Herbario	21.5
F5	CaMIR2	Querétaro	El Mirador	S	Tejido seco	24.8
G5	CaMIR6	Querétaro	El Mirador	S	Tejido seco	28.7
H5	CaMIR13	Querétaro	El Mirador	S	Tejido seco	23.5
A6	CaMIR15	Querétaro	El Mirador	S	Tejido seco	43.7
B6	CaMIR16	Querétaro	El Mirador	S	Tejido seco	26.7
C6	CaMIR18	Querétaro	El Mirador	S	Tejido seco	29.2
D6	CaPAT3	Querétaro	El Patol	S	Tejido seco	50.7
E6	CaPAT7	Querétaro	El Patol	S	Tejido seco	45.2
F6	CaPAT8	Querétaro	El Patol	S	Tejido seco	52.5
G6	CaPAT9	Querétaro	El Patol	S	Tejido seco	28.1
H6	CaPAT11	Querétaro	El Patol	S	Tejido seco	43.7
A7	CaJUD2	Oaxaca	Juan_Diegal	S	Invernadero	27.9
B7	CaGVI3	Oaxaca	Guadalupe Victoria	A	Hoja congelada	38.8
C7	CaVTR1	Oaxaca	Valerio Trujano	S	Hoja congelada	34.7
D7	CaJUC1	Oaxaca	Juchitán	A	Hoja congelada	26.3
E7	CaJUC2	Oaxaca	Juchitán	A	Hoja congelada	26.3
F7	CaJUC3	Oaxaca	Juchitán	A	Fruto congelado	31.6
G7	CaGUZ2	Oaxaca	Guzmán	A	Hoja congelada	22.4

H7	CaSDI1	Oaxaca	Santo Domingo del Ingenio	A	Hoja congelada	20.4
A8	CaSDI2	Oaxaca	Santo Domingo del Ingenio	A	Hoja congelada	34.1
B8	CaSDI21	Oaxaca	Santo Domingo del Ingenio	A	Fruto congelado	27.5
C8	CaSDI22	Oaxaca	Santo Domingo del Ingenio	A	Fruto congelado	37.8
D8	CaTAB1	Tabasco	Río Viejo	A	Fruto seco	41.2
E8	CaYEL1	Jalisco	Yelapa	A	Fruto seco	30
F8	CaOCZ1	Chiapas	Ocozocuautila	S	Fruto seco	25.4
G8	CaCOY11	Oaxaca	El Coyul	A	Hoja congelada	45.6
H8	CaTOM1	Oaxaca	El Tomatal	A	Hoja congelada	26.8
A9	CaTOM2	Oaxaca	El Tomatal	A	Hoja congelada	38.8
B9	CaVPI1_1	Veracruz	Tlapacoyan	A	Invernadero	52.3
C9	CaVPI2_1	Veracruz	Tlapacoyan	A	Invernadero	36.3
D9	CaVPI2_11	Veracruz	Tlapacoyan	A	Invernadero	112
E9	CaGVI3_R	Chiapas	Guadalupe Victoria	A	Invernadero	36.2
F9	CaGVI4	Chiapas	Guadalupe Victoria	A	Hoja congelada	26.9
G9	CaGVI5	Chiapas	Guadalupe Victoria	A	Hoja congelada	21.6
H9	CaCOY2	Oaxaca	El Coyul	A	Invernadero	20
A10	CaCHI4_1R	Oaxaca	El Chilar	D	Invernadero	79
B10	CaCHI4_2R	Oaxaca	El Chilar	D	Invernadero	61.1
C10	CaGUZ1	Oaxaca	Guzmán	D	Hoja congelada	23.2
D10	CaZAA3	Oaxaca	Zaachila	D	Hoja congelada	46
E10	CaCOY13	Oaxaca	El Coyul	D	Hoja congelada	30.3
F10	CaMAT6	Zacatecas	Matancillas	D	Invernadero	23.2
G10	CaVCO1_8	Zacatecas	Villa_de_Cos	D	Invernadero	36.3
H10	CaDZI1_1	Yucatán	Dzibichaltún	D	Invernadero	48.3

A11	CaSRA4_2	Zacatecas	Rancho San Ramón	D	Invernadero	22
B11	CaSRA4_7	Zacatecas	Rancho San Ramón	D	Invernadero	24.1
C11	CaSRA1_2	Zacatecas	Rancho San Ramón	D	Invernadero	31.4
D11	CaSRA1_6	Zacatecas	Rancho San Ramón	D	Invernadero	51.6
E11	CaSRA1_14	Zacatecas	Rancho San Ramón	D	Invernadero	21.5
F11	CaTEC1_5	Zacatecas	El Tecuán	D	Invernadero	28.5
G11	CaTEC2_3	Zacatecas	El Tecuán	D	Invernadero	37.3
H11	CaTEC6_2	Zacatecas	El Tecuán	D	Invernadero	26.5
A12	BLANCO	BLANCO	BLANCO	BLANCO	BLANCO	0
B12	CaZOR1	Zacatecas	El Zorrillo	D	Invernadero	43.8
C12	CaZOR2_1R	Zacatecas	El Zorrillo	D	Invernadero	25.1
D12	CaZOR2_3R	Zacatecas	El Zorrillo	D	Invernadero	31.3
E12	CaZOR7	Zacatecas	El Zorrillo	D	Invernadero	53.2
F12	CaZOR10	Zacatecas	El Zorrillo	D	Invernadero	29.3
G12	CaZOR11	Zacatecas	El Zorrillo	D	Invernadero	25.1
H12	CaZOR13_3	Zacatecas	El Zorrillo	D	Invernadero	35

Anexo 4. Datos de cada muestra de *C. annuum* y su correspondiente fracción mapeada al genoma de referencia. Complemento de la Figura 12.

MUESTRAS	ESTADO	LOCALIDAD	FRACCIÓN MAPEADA AL GENOMA
CaMOC15	Sonora	Moctezuma	0.731
CaMOC25	Sonora	Moctezuma	0.733
CaMOC30	Sonora	Moctezuma	0.757
CaMOC37	Sonora	Moctezuma	0.717
CaMAU7	Sonora	Los Mautos	0.028
CaMAU9	Sonora	Los Mautos	0.205
CaMAU14	Sonora	Los Mautos	0.034
CaMAU25	Sonora	Los Mautos	0.057
CaBER21	Querétaro	Bernal	0.718
CaBER23	Querétaro	Bernal	0.659
CaBER33	Querétaro	Bernal	0.74
CaBER36	Querétaro	Bernal	0.788
CaHUJ4	Sinaloa	El Huajote	0.767
CaHUJ6	Sinaloa	El Huajote	0.536
CaELO15	Sinaloa	Elota	0.754
CaPOT9	Sinaloa	El Potrero	0.472
CaSAN1	Sinaloa	Sanalona	0.101
CaSAN2	Sinaloa	Sanalona	0.751
CaTEM1	Sonora	Temporal	0.12
CaTEM6	Sonora	Temporal	0.086
CaTEM7	Sonora	Temporal	0.053
CaTEM8	Sonora	Temporal	0.022
CaTEM9	Sonora	Temporal	0.113
CaTEM11	Sonora	Temporal	0.206
CaTEM16	Sonora	Temporal	0.004
CaCHS1	Chihuahua	Chihuahua	0.124
CaSDI19_1R	Oaxaca	Santo Domingo del Ingenio	0.593
CaSDI19_2R	Oaxaca	Santo Domingo del Ingenio	0.687
CaSDI3	Oaxaca	Santo Domingo del	0.731

			Ingenio
CaTAN8	Querétaro	Tancamá	0.351
CaTAN10	Querétaro	Tancamá	0.373
CaTAN13	Querétaro	Tancamá	0.589
CaTOL1	Querétaro	Tolimán	0.789
CaRES5	Querétaro	Rancho la Espadañuela	0.604
CaRES6	Querétaro	Rancho la Espadañuela	0.662
CaRES9	Querétaro	Rancho la Espadañuela	0.665
CaSOT1	Tamaulipas	Soto la Marina	0.434
CaMIR2	Querétaro	El Mirador	0.755
CaMIR6	Querétaro	El Mirador	0.757
CaMIR13	Querétaro	El Mirador	0.721
CaMIR15	Querétaro	El Mirador	0.769
CaMIR16	Querétaro	El Mirador	0.731
CaMIR18	Querétaro	El Mirador	0.673
CaPAT3	Querétaro	El Patol	0.704
CaPAT7	Querétaro	El Patol	0.72
CaPAT8	Querétaro	El Patol	0.702
CaPAT9	Querétaro	El Patol	0.691
CaPAT11	Querétaro	El Patol	0.724
CaJUD2	Oaxaca	Juan Diegal	0.584
CaGVI3	Oaxaca	Guadalupe Victoria	0.674
CaVTR1	Oaxaca	Valerio Trujano	0.729
CaJUC1	Oaxaca	Juchitán	0.671
CaJUC2	Oaxaca	Juchitán	0.722
CaJUC3	Oaxaca	Juchitán	0.706
CaGUZ2	Oaxaca	Guzmán	0.476
CaSDI1	Oaxaca	Santo Domingo del Ingenio	0.334
CaSDI2	Oaxaca	Santo Domingo del Ingenio	0.737
CaSDI21	Oaxaca	Santo Domingo del Ingenio	0.745
CaSDI22	Oaxaca	Santo Domingo del Ingenio	0.749

CaTAB1	Tabasco	Río Viejo	0.629
CaYEL1	Jalisco	Yelapa	0.751
CaOCZ1	Chiapas	Ocozocuautla	0
CaCOY11	Oaxaca	El Coyul	0.687
CaTOM1	Oaxaca	El Tomatal	0.738
CaTOM2	Oaxaca	El Tomatal	0.715
CaVPI2_1	Veracruz	Tlapacoyan	0.791
CaVPI2_11	Veracruz	Tlapacoyan	0.753
CaGVI3_R	Chiapas	Guadalupe Victoria	0.402
CaGVI4	Chiapas	Guadalupe Victoria	0.381
CaGVI5	Chiapas	Guadalupe Victoria	0.288
CaCOY2	Oaxaca	El Coyul	0.675
CaCHI4_1R	Oaxaca	El Chilar	0.639
CaCHI4_2R	Oaxaca	El Chilar	0.699
CaGUZ1	Oaxaca	Guzmán	0.288
CaZAA3	Oaxaca	Zaachila	0.562
CaCOY13	Oaxaca	El Coyul	0.621
CaMAT6	Zacatecas	Matancillas	0.755
CaVCO1_8	Zacatecas	Villa de Cos	0.653
CaDZI1_1	Yucatán	Dzibichaltún	0.742
CaSRA4_2	Zacatecas	Rancho San Ramón	0.746
CaSRA4_7	Zacatecas	Rancho San Ramón	0.67
CaSRA1_2	Zacatecas	Rancho San Ramón	0.746
CaSRA1_6	Zacatecas	Rancho San Ramón	0.703
CaSRA1_14	Zacatecas	Rancho San Ramón	0.703
CaTEC1_5	Zacatecas	El Tecuán	0.683
CaTEC2_3	Zacatecas	El Tecuán	0.706
CaTEC6_2	Zacatecas	El Tecuán	0.58
CaZOR1	Zacatecas	El Zorrillo	0.583
CaZOR2_1R	Zacatecas	El Zorrillo	0.628
CaZOR2_3R	Zacatecas	El Zorrillo	0.573
CaZOR7	Zacatecas	El Zorrillo	0.647
CaZOR10	Zacatecas	El Zorrillo	0.597
CaZOR11	Zacatecas	El Zorrillo	0.707

CaZOR13_3

Zacatecas

El Zorrillo

0.711
