
# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

## POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE ECOLOGÍA
BIOLOGÍA EVOLUTIVA

## ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE *Cucurbita argyrosperma* Huber (CUCURBITACEAE) EN MÉXICO.

# TESIS

QUE PARA OPTAR POR EL GRADO DE:

**DOCTOR EN CIENCIAS**

PRESENTA:

**GUILLERMO SÁNCHEZ DE LA VEGA**

**TUTOR PRINCIPAL DE TESIS:**

DR. LUIS ENRIQUE EGUIARTE FRUNS INSTITUTO DE ECOLOGÍA

**COMITÉ TUTOR:**

DR. JUAN PABLO JARAMILLO CORREA INSTITUTO DE ECOLOGÍA

DR. RAFAEL LIRA SAADE FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

**CIUDAD UNIVERSITARIA, CD.MX., 2022**


# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

## POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE ECOLOGÍA
BIOLOGÍA EVOLUTIVA

## ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE *Cucurbita argyrosperma* Huber (CUCURBITACEAE) EN MÉXICO.

# TESIS

QUE PARA OPTAR POR EL GRADO DE:

**DOCTOR EN CIENCIAS**

PRESENTA:

**GUILLERMO SÁNCHEZ DE LA VEGA**

**TUTOR PRINCIPAL DE TESIS:**

DR. LUIS ENRIQUE EGUIARTE FRUNS INSTITUTO DE ECOLOGÍA

**COMITÉ TUTOR:**

DR. JUAN PABLO JARAMILLO CORREA INSTITUTO DE ECOLOGÍA

DR. RAFAEL LIRA SAADE FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

**CIUDAD UNIVERSITARIA, CD.MX., 2022**

**COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS**
**ENTIDAD INSTITUTO DE ECOLOGÍA**
**OFICIO CPCB/536/2022**
**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, en su reunión ordinaria del día 06 de diciembre de 2021, aprobó el siguiente jurado para el examen de grado de **DOCTOR EN CIENCIAS** del estudiante: **SÁNCHEZ DE LA VEGA GUILLERMO,** con número de cuenta **88072926,** con la tesis titulada: **"ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE *Cucurbita argyrosperma* HUBER (CUCURBITACEAE) EN MÉXICO",** bajo la dirección del **DR. LUIS ENRIQUE EGUIARTE FRUNS**, Tutor Principal, quedando integrado de la siguiente manera**:**

| | |
|---|---|
| Presidente: | DR. DANIEL IGNACIO PIÑERO DALMAU |
| Vocal: | DR. ALEJANDRO CASAS FERNÁNDEZ |
| Vocal: | DRA. ALEJANDRA CITLALLI MORENO LETELIER |
| Vocal: | DRA. ERIA ALAIDE REBOLLAR CAUDILLO |
| Secretario: | DR. RAFAEL LIRA SAADE |

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

**A T E N T A M E N T E**
**"POR MI RAZA HABLARÁ EL ESPÍRITU"**
Ciudad Universitaria, Cd. Mx., a 06 de junio de 2022

**COORDINADOR DEL PROGRAMA**

POSGRADO EN CIENCIAS BIOLOGICAS
COORDINACIÓN

**DR. ADOLFO GERARDO NAVARRO SIGÜENZA**

## Agradecimientos institucionales

Agradezco al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México, por permitirme llevar a cabo mis estudios de doctorado en el Instituto de Ecología de la UNAM en Ciudad Universitaria. Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico durante la realización de mis estudios de doctorado con la Beca no. 292164.

Asimismo, agradezco al financiamiento otorgado por la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) con los proyectos KE004 "Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de *Cucurbita*" y PE001 "Diversidad genética de las especies de *Cucurbita* en México. Fase II. Genómica evolutiva y de poblaciones, recursos genéticos y domesticación".

Al proyecto de Ciencia Básica CONACyT 2011.167826 "Genómica de poblaciones: estudios en el maíz silvestre, el teosinte (*Zea mays* ssp. *parviglumis* y *Zea mays* ssp. *mexicana*) y al proyecto de Problemas Nacionales (CONACyT) 247730, Por su apoyo en la realización de esta tesis.

Agradezco a mi comité tutor, integrado por el **Dr. Luis E. Eguiarte Fruns** del Instituto de Ecología UNAM, el **Dr. Rafael Lira Saade** de la UBIPRO de la Facultad de Estudios Superiores Iztacala UNAM, y al **Dr. Juan Pablo Jaramillo Correa** del Instituto de Ecología UNAM. Gracias por su tiempo, apoyo y dedicación invertidos en mi formación académica. Gracias por compartir sus Conocimientos y por todos sus por sus consejos.

## Agradecimientos personales

Agradezco al Dr. Rafael Lira y al Dr. Luis Eguiarte por permitirme ser parte de este proyecto y por todo su apoyo y enseñanzas durante este tiempo. Asimismo, agradezco a la Dra. Valeria Souza y a Luis por aceptarme en el laboratorio, por su apoyo personal y académico durante estos años. Un agradecimiento especial a la Dra. Erika Aguirre, quien con su magia logística hace que todo fluya en el laboratorio. Gracias Eri por tu amistad y apoyo en todo momento.

A todos los miembros del Team *Cucurbita*. A la Dra. Gabriela Castellanos, por encaminar para bien este proyecto y por compartir siempre sus conocimientos y amistad, Gracias Gaby. A Helena Hernández, por ser mi cómplice en este proyecto casi desde el inicio, por su amistad y por su apoyo en este viaje, gracias por todo Helena. A Josué Barrera por compartir tiempo, conocimientos y amistad este camino de estudiar a *C. argyrosperma*. A todas las calabacitas tiernas del Team Cucurbita: Leslie, Paulina, Karen, Fernando, Carmina y Brenda, por aportar esa dosis de energía, entusiasmo y trabajo en equipo que impulso siempre el proyecto. Asimismo, Agradezco a los técnicos y posdocs (además de Gaby): Dr. Enrique Scheinvar, Dr. Jaime Gasca y Dra. Xitlali Aguirre por su constante aporte para mejorar el proyecto *Cucurbita* y por su ayuda y amistad.

A todos los miembros del Instituto de Ecología y del LEMyE, con quienes compartí durante todo este tiempo amistad, aprendizaje, salidas a campo y momentos personales: Manuel Rosas, Yocelyn Gutiérrez, Enrique Scheinvar, Valerie, Alejandra, Niza, Jonás, Laura Espinosa, Roberto, Jorge, Mirna, Santiago, Gabriel Merino, Ricardo, Manuel Ulloa, Mariette, Diego, Itzi. Agradecimiento especial a la Sra. Silvia Barrientos, por imponer la disciplina en el laboratorio, por su apoyo y amistad durante este tiempo.

A los miembros de la mesa del Jurado: Dr. Daniel Piñero Dalmau, Dr. Alejandro Casas Fernández, Dra. Alejandra Citlalli Moreno Letelier, Dra. Eria Alaide Rebollar Caudillo y al Dr. Rafael Lira Saade, por su tiempo, apoyo y sus comentarios en esta etapa final.

Agradezco el soporte técnico y ayuda brindada por parte del personal del Instituto de Ecología: a la Lic. Erika Rodríguez Reyes por su invaluable apoyo y ayuda en todos los trámites durante mi doctorado. Al Biól. Marco Tulio Solano de la Cruz responsable del Laboratorio de Genética Molecular y al Ing. Rodrigo García Herrera, técnico académico a cargo de la unidad de cómputo en el LANCIS.

Agradezco a mi familia su apoyo incondicional. A mis padres a quienes todo debo y quienes siempre han estado a mi lado, aún en los momentos difíciles. A mi hija Natalie, por ser el motor que me impulsa a seguir siempre adelante.

# INDICE

## RESUMEN

*Cucurbita argyrosperma* (Cucurbitaceae) --la calabaza pipiana o mixta-- se compone de poblaciones cultivadas, silvestres y en algunos casos escapadas o ferales. Cuenta con dos subespecies, la subespecie *sororia,* la cual es silvestre y la subespecie *argyrosperma,* que es cultivada. Se distribuyen a lo largo de la costa del océano Pacífico y en algunas zonas de la costa del Golfo de México; *C. argyrosperma* además es cultivada en la península de Yucatán y en estados del centro y del norte de México. *Cucurbita argyrosperma* ha sido una de las especies de calabazas domesticadas menos estudiadas, a pesar de ser un recurso fitogenético de gran valor económico. Es de suma importancia describir la diversidad genética de las razas locales cultivadas y la de las poblaciones silvestres, documentar la presencia de las variedades cultivadas, estudiar su adaptación local y la forma en que se distribuye esta variación en México. El objetivo de esta tesis fue caracterizar la diversidad, estructura y flujo génico de variedades locales estudiando 19 poblaciones de la subespecie domesticada *C. argyrosperma* y 7 poblaciones de la subespecie silvestre *C. sororia,* utilizando como marcadores moleculares nueve loci de microsatélites nucleares y 2,861 SNPs neutros secuenciados con librerías de tGBS (*tunable genotyping-by-sequencing*) de todo el genoma. Además se utilizó un modelo coalescente que permitiera ubicar y entender el proceso de domesticación de esta especie. La diversidad genética fue similar entre subespecies ($H_E$=0.428 para *sororia* y $H_E$=0.410 para *argyrosperma*) con microsatélites y con SNPs ($H_E$=0.096 para *sororia* y $H_E$=0.094 para *argyrosperma*). Ambas subespecies están bien diferenciadas y los valores de $F_{ST}$ entre poblaciones oscilan entre 0.152 a 0.652 con microsatélites y entre 0.0565 a 0.0751 con SNPs. Los valores de endogamia estimados a partir de los microsatélites ($F_{IS}$= 0.077 para *sororia* y $F_{IS}$= 0.033 para *argyrosperma*) y de los SNPs ($F_{IS}$= 0.011 para *sororia* y $F_{IS}$= 0.034 para *argyrosperma*) fueron bajos en general. Los resultados muestran un origen monofilético de la especie y de las plantas domesticadas (*argyrosperma*) y que ambas subespecies están bien diferenciadas genéticamente. Dentro de cada subespecie se encontró que hay estructura y grupos genéticos. Detectamos que existe flujo entre subespecies. Nuestros modelos de distribución indican que las características ambientales en la región de Jalisco son las de un centro de domesticación potencial, lo cual es confirmado con nuestro análisis coalescente de domesticación a partir de los SNPs. En general, los patrones de estructura genética de *C. argyrosperma* son coherentes con la evidencia arqueológica de la migración humana temprana a lo largo de Mesoamérica. Los análisis permiten comprender mejor la historia evolutiva de *C. argyrosperma* y sus relaciones filogeográficas. Es necesario establecer una estrategia de conservación *in situ* y *ex situ*, tanto de las poblaciones silvestres como domesticadas, principalmente de las variedades locales, ya que consideramos que esto es importante para mantener la diversidad genética de *C. argyrosperma*, una especie que representa un relevante recurso fitogenético para México y para el mundo.

## ABSTRACT

*Cucurbita argyrosperma* (Cucurbitaceae) --the pipiana or mixta squash-- consists of cultivated, wild, and in some cases escaped or feral populations. It has two subspecies, the wild *sororia* subspecies and the cultivated *argyrosperma* subspecies. They are distributed along the coast of the Pacific Ocean and in some areas of the coast of the Gulf of Mexico; *C. argyrosperma* is also cultivated in the Yucatan peninsula and in the central and northern states of Mexico. Cucurbita argyrosperma has been one of the least studied species of domesticated gourds, despite being a plant genetic resource of great economic value. It is extremely important to describe the genetic diversity of local cultivated breeds and that of wild populations, to document the presence of cultivated varieties, to study their local adaptation and the way in which this variation is distributed in Mexico. The aim of this thesis was to characterize the diversity, structure, and gene flow of landraces by studying 19 populations of the domesticated subspecies *C. argyrosperma* and 7 populations of the wild subspecies *C. sororia*, using nine nuclear microsatellite loci and 2,861 neutral SNPs sequenced with tGBS (tunable genotyping-by-sequencing) libraries across the entire genome, as molecular markers. In addition, a coalescent model was used to locate and understand the domestication process of this species. Genetic diversity was similar between subspecies ($H_E$=0.428 for sororia and $H_E$=0.410 for *argyrosperma*) with microsatellites and with SNPs ($H_E$=0.096 for sororia and $H_E$=0.094 for argyrosperma). Both subspecies are well differentiated and $F_{ST}$ values between populations range from 0.152 to 0.652 with microsatellites and from 0.0565 to 0.0751 with SNPs. Inbreeding values estimated from microsatellites ($F_{IS}$= 0.077 for *sororia* and $F_{IS}$= 0.033 for *argyrosperma*) and SNPs ($F_{IS}$= 0.011 for *sororia* and $F_{IS}$= 0.034 for *argyrosperma*) were generally low. The results show a monophyletic origin of the species and domesticated plants (*argyrosperma*) and that both subspecies are genetically well differentiated. Within each subspecies, genetic structure and different genetic groups were found. We detected gene flow between subspecies. Our distribution models indicate that the environmental characteristics in the Jalisco region are those of a potential domestication center, which is confirmed by our coalescent analysis of domestication from SNPs. In general, the genetic structure patterns of *C. argyrosperma* are consistent with archaeological evidence for early human migration throughout Mesoamerica. The analyzes allow a better understanding of the evolutionary history of *C. argyrosperma* and its phylogeographic relationships. It is necessary to establish an in situ and ex situ conservation strategy, both for wild and domesticated populations, mainly local varieties, since we consider that this is important to maintain the genetic diversity of *C. argyrosperma*, a species that represents a relevant plant genetic resource for Mexico and for the world.

# INTRODUCCIÓN GENERAL

## La variación genética de las plantas y su domesticación

En México se han registrado alrededor de 7,000 especies de plantas útiles; de las cuales alrededor de 200 se encuentran en estado avanzado de domesticación, incluyendo algunas especies nativas de importancia mundial como el maíz, los frijoles, los chiles, las calabazas, el cacao, algodón, los agaves y el amaranto, además de otras especies de importancia nacional o regional (Casas *et al.,* 2007; Perales y Aguirre, 2008).

Las evidencias arqueológicas y moleculares muestran que la agricultura en México se inició hace al menos 10 mil años (Zizumbo-Villarreal y Colunga-GarcíaMarin, 2008; Piperno, 2011), con la selección de variedades silvestres por parte de los cazadores recolectores, hasta llegar a producir poblaciones domesticadas genética y fenotípicamente distintas de sus parientes silvestres (Meyer y Purugganan, 2013). Cabe mencionar que algunas de las especies domesticadas conviven en ciertas regiones del país con dichos parientes silvestres (Goettsch *et al.,* 2021), lo que evolutivamente sugiere un modelo de aislamiento en presencia de flujo génico.

La domesticación implica una transformación de las poblaciones silvestres en individuos con fenotipos deseables a través de la selección, que puede estar relacionada con ciertas adaptaciones a condiciones ecológicas y preferencias específicas. Las características en una especie que son seleccionadas durante la domesticación engloban una amplia serie de cambios evolutivos en las especies que se denomina *síndrome de domesticación* (Fuller, 2012; Meyer y Purugganan, 2013). Aunque las características de dicho síndrome varían entre especies, estas pueden incluir cambios en la arquitectura de los tallos, el aumento en tamaño de frutos, semillas y partes vegetativas, la pérdida de la habilidad de dispersión y de la dormancia en las semillas, así como la pérdida de la protección química o mecánica contra los herbívoros (Wilcox, 2012; Meyer y Purugganan, 2013). Todo esto puede disminuir la capacidad de los individuos para sobrevivir en la naturaleza (adecuación), pero la puede aumentar cuando estas plantas dependen de cuidados humanos.

En los últimos veinte años los avances tecnológicos y metodológicos en diferentes campos -- que van de la biología molecular a la arqueología-- han revolucionado los estudios de plantas domesticadas, mejorando nuestra comprensión sobre el origen y la diversidad genética que poseen (Piñero, 2008a; Perales y Aguirre, 2008). Se puede considerar que la diversidad genética es uno de elementos básicos de la biodiversidad y se refiere al número total de características genéticas heredables que ocurren entre los individuos de una población y entre las poblaciones de cada especie (Rimieri, 2017). La diversidad genética es la base sobre la que actúan las diferentes fuerzas evolutivas y es necesaria para la adaptación de las especies a ambientes nuevos (Hedrick, 2011), incluyendo los efectos de la selección antrópica. Cuando no hay variación genética o sus niveles son bajos, existe un elevado riesgo de que no resistan esos cambios ambientales. Niveles bajos de variación genética pueden deberse a factores como la reducción en el tamaño efectivo de las poblaciones y el aumento de apareamientos consanguíneos (Hedrick, 2011). La variación genética en un momento dado es la combinación de diversidad genética preexistente con diversidad nueva, surgida por mutación, más la que llega por flujo génico de otras poblaciones.

La importancia de los parientes silvestres de las especies domesticadas es principalmente debida a su relevancia potencial para la alimentación, ya que son depósitos genéticos que pueden ser útiles en programas de mejoramiento en las especies cultivadas. Recientemente se ha incrementado el interés por estudiar a estos parientes silvestre las plantas cultivadas (Doebley *et al.,* 2006; Zeder *et al.,* 2006; Bar-Yosef y Price *et al.,* 2011, Larson *et al.,* 2014; Lira *et al.,* 2016; Contreras *et al.,* 2019, Khoury *et al.,* 2020). Por eso es importante determinar cómo está organizada su diversidad genética y cuál es su dinámica cuando hay intercambios genéticos con poblaciones domesticadas. Además, es necesario interpretar los patrones temporales en los que se dan dichos procesos (Abbo *et al.,* 2010; Innan y Kim, 2004) y los factores humanos y ecológicos que lo promueven (Gepts *et al.,* 2012; Bar-Yosef y Price *et al.,* 2011).

Parámetros como la heterocigocidad, la proporción de loci polimórficos o la diversidad nucleotídica entre otros, son útiles para cuantificar y comparar los niveles de diversidad genética (Piñero, 2008a; Hedrick, 2011).

## El género *Cucurbita* en México

El género *Cucurbita* es reconocido por sus plantas domesticadas, las cuales son comúnmente conocidas con nombres como "calabazas", o "zapallos" en los países de habla hispana, o mediante numerosos nombres en lenguas indígenas (Lira, 1995). El uso más importante al que se han destinado las plantas domesticadas de *Cucurbita* es el alimenticio, no solo en Latinoamérica, sino también en muchas otras regiones del mundo; por lo que constituye un recurso fitogenético de gran importancia (FAO, 2010).

Los frutos (inmaduros y/o maduros) y las semillas de *Cucurbita* son las partes más comúnmente empleadas como alimento, y representan una fuente importante de proteínas y pueden almacenarse por periodos prolongados de tiempo (Villanueva, 2007; OECD, 2012). Adicionalmente también se utilizan, en menor escala, sus flores (generalmente las masculinas o estaminadas) y las partes tiernas de los tallos, conocidas en varios países como “puntas de las guías" (Lira *et al.,* 2009a).

El género *Cucurbita* es un género estrictamente americano, pues todos sus miembros crecen de forma natural o fueron domesticados en América. Se considera que fueron las primeras plantas cultivada en Mesoamérica, formando parte de la dieta de los pueblos de México desde hace más de 10,000 años (Lira, 1995; Smith, 1997; Piperno, 2011). El género incluye 15 especies de 20 taxa distintos, de los cuales 15 de ellos crecen espontáneamente o se cultivan en México; lo que hace de este país el más importante centro de origen, domesticación y diversificación de *Cucurbita* (Lira, 1995; Lira *et al.,* 2009a; OECD, 2012). Las especies cultivadas y conocidas como calabazas son: *C. pepo* (calabaza de castilla o calabacita ), *C. ficifolia* (chilacayote), *C. moschata* (calabaza de castilla)*; C. maxima* (calabaza kabosha, zapallo) y *C. argyrosperma* (calabaza criolla, pipiana o mixta) (Villanueva, 2007; OECD, 2012).

Las plantas del género *Cucurbita* son anuales o perennes, monoicas, y rastreras a trepadoras (subarbustivas en algunos cultivares comerciales). Sus flores son gamopétalas, con corolas tubular-campanuladas y muy vistosas, de color amarillo pálido a amarillo-anaranjado brillante; las flores de ambos sexos son solitarias (Lira, 1995). Las flores de *Cucurbita* abren muy temprano por la mañana y son polinizadas por especies de abejas solitarias de los géneros *Peponapis* y

*Xenoglossa*. Los frutos son del tipo pepo y, en las plantas domesticadas, se producen en una gran diversidad de formas, tamaños, colores, y tipos de superficies, entre otras características. Mientras que en las plantas silvestres, los frutos son relativamente uniformes en cuanto a su forma (globosos, ovoides o raramente piriformes); además, su superficie es generalmente lisa o sin ornamentaciones, con coloración blanca, amarillenta o verde, con o sin manchas y/o franjas y de tamaño comparativamente pequeño (Lira, 1995; Villanueva, 2007). Todas las especies de *Cucurbita* tienen 20 pares de cromosomas, y se ha propuesto que el género es un antiguo tetraploide (Lira *et al.,* 2009a, Barrera-Redondo, *et al.,* 2019).

Durante siglos, los agricultores en México han desarrollado empíricamente una elevada diversidad morfológica de calabazas, principalmente asociadas al agroecosistema conocido como *milpa* (Zizumbo-Villarreal y Colunga-GarcíaMarin, 2008). Sin embargo, en las últimas décadas la diversidad de variedades locales se ha visto aparentemente reducida, favoreciendo el uso de las variedades mejoradas de calabaza en agroecosistemas tradicionales, debido a que son supuestamente más productivas (Lira y Montes, 1994; OECD, 2012). Para las diferentes especies domesticadas de *Cucurbita*, la mayoría de sus variedades locales o tradicionales han sido poco estudiadas en México y requieren una caracterización morfológica, ecológica y sobre todo genética; por ejemplo, para conocer cómo se relacionan entre ellas y para establecer criterios de conservación y aprovechamiento.

## El grupo Argyrosperma

El Grupo Argyrosperma corresponde a la especie biológica *Cucurbita argyrosperma* Huber, la cual incluye dos subespecies: *Cucurbita argyrosperma* subsp. *sororia* (L.H. Bailey) Merrick & D.M. Bates*,* que es la forma silvestre, y la subespecie cultivada *C. argyrosperma* Huber subsp. *Argyrosperma*, en la que se ubican todas las variedades domesticadas del grupo (Lira, 1995; OECD, 2012). Las plantas de ambas subespecies pueden ser herbáceas, rastreras y trepadoras vigorosas; son anuales, silvestres o cultivadas, e incluso espontaneas o ferales. Se distribuyen desde el suroeste de los Estados Unidos hasta Nicaragua y Panamá; en regiones desde nivel del mar, hasta los 1800 m s. n. m. (Lira y Montes-Hernández, 1994; Nee, 1993; Lira, 1995; Montes-Hernández y Eguiarte, 2002).

La subespecie *C. a.* subsp. *sororia*, comprende a las formas silvestres o espontáneas y corresponde al ancestro del cual partió la domesticación (Lira *et al.,* 2009). Este presenta frutos de forma generalmente ovoide, de color verde, con franjas y manchas longitudinales. El intervalo altitudinal de esta subespecie es muy similar a la de la subespecie cultivada, creciendo desde el nivel de mar, hasta aproximadamente los 1800 msnm. Crece cercana a cultivos, en agrosistemas, o en bosques, y matorrales; generalmente en climas cálidos y algo secos (Lira y Montes-Hernández, 1994; Nee, 1993; Lira, 1995), principalmente formando parte de selvas bajas caducifolias, matorrales, vegetación secundaria, ruderal y riparia (Nee, 1993; Lira, 1995). Está presente en México, a lo largo de la costa entre Sonora y Chiapas, y en los estados de Morelos, Tamaulipas, Veracruz; asimismo también se ha reportado en América Central (lira, 1995).

La subespecie cultivada *argyrosperma*, es conocida comúnmente como arota, pipiana, patipona, mixta, *cushaw*, *silver seed gourd*, entre otros nombres (Lira, 1995: OCDE, 2016). Se caracteriza por poseer una gran variación en su fruto, que es de tamaño variable, dependiendo de la región. La parte externa de los mismo es de textura mayormente lisa, con formas redondas, aplanadas dorsiventralmente o alargadas: los patrones de franjas y rayas longitudinales son características de esta subespecie (Lira, 1995; Villanueva, 2007). Aunque se consumen sus hojas, tallos, flores, y en algunas regiones la pulpa, la semilla es el principal producto de esta especie. Las semillas tienen una gran variedad morfológica, siendo posiblemente las de mayor diversidad dentro de todas las especies de *Cucurbita*. Se ha propuesto que su domesticación se inició en la cuenca del río Balsas, hace más de 8,500 años (Ranere *et al.,* 2009); por lo que variedades cultivadas se pueden encontrar en diferentes regiones del país. El potencial de hibridación entre ambas subespecies es alto, dada su co-existencia y cercana relación evolutiva (Nabhan, 1984; Lira, 1995; Montes-Hernández y Eguiarte, 2002; OCDE, 2016). Los híbridos se pueden dar en ambas direcciones, y su descendencia tiene una alta fertilidad (Lira, 1995; Montes-Hernández y Eguiarte, 2002).

## Estudios sobre la diversidad genética en *C. argyrosperma*

En los últimos 20 años diferentes estudios biosistemáticos y etnobotánicos han producido información acerca de la diversidad morfológica y genética de las especies del género *Cucurbita,*

así como de sus relaciones con otras especies del género (Sanjur *et al.,* 2002; Montes-Hernández y Eguiarte, 2002; Villanueva, 2007; Cerón *et al.,* 2010; Lira *et al.,* 2016; OCDE, 2012). Más recientemente, los estudios de sistemática y diversidad genética se han multiplicado con el uso de marcadores y herramientas genómicas de nueva generación, permitiendo caracterizar y actualizar la información existente (Castellanos-Morales *et al.,* 2018, 2019; Hernández-Rosales *et al.,* 2020; Martínez-González *et al.,* 2021).

Los estudios publicados para *C. argyrosperma* han sido mucho menores a los que se han publicado para otras especies, posiblemente porque no se cultiva globalmente (Lira *et al.,* 2016). En general, su cultivo se destina en mayor proporción a autoconsumo y a un comercio regional. Entre los primeros estudios de diversidad genética para esta especie se encuentra el de Wilson (1990), que analizó 46 poblaciones con 9 loci isoenzimáticos; estas incluyeron variedades criollas de *C. moschata* y *C. argyrosperma* subsp. *argyrosperma* y poblaciones silvestres de *C. argyrosperma* subsp. *sororia* de varias regiones de México. Se encontró un extenso flujo genético entre las subespecies *argyrosperma* y *sororia*, pero menor variación genética total que la registrada en *C. moschata*.

Por otro lado, Montes-Hernández y Eguiarte (2002), utilizando 12 loci isoenzimaticos, estudiaron 11 poblaciones de plantas domesticadas y silvestres del Grupo Argyrosperma, cuatro de *C. moschata* y una de *C. pepo* como grupo externo; todas las poblaciones eran simpátricas en la región de Autlán, Jalisco. Encontraron altos niveles de polimorfismo (0.96) y de heterocigosis esperada ($H_e$= 0.407), particularmente en la subsp. *sororia* ($H_e$= 0.411 – 0.433), y poca diferenciación genética entre poblaciones ($F_{st}$ = 0.087), sugiriendo elevado flujo génico entre poblaciones.

Cerón *et al.* (2010) analizaron la diversidad genética entre y dentro de cuatro especies de *Cucurbita: argyrosperma* (20 localidades)*, moschata* (20 localidades)*, pepo* (3 localidades) y *ficifolia* (7 localidades), con 100 semillas por localidad. Las semillas eran provenientes de los estados de Morelos, Nayarit, Oaxaca, Veracruz, Yucatán y Estado de México de México. Se utilizaron marcadores moleculares tipo RAPD y probaron 60 iniciadores para un total de 185 loci. Se encontró que *C. argyrosperma* tenía una diversidad ($H_e$=0.068) mayor que *C. ficifolia* ($H_e$ =

0.058) y menor que *C. pepo* ($H_e$ =0.104) y *C. moschata* ($H_e$ =0.116). El porcentaje de loci polimórficos *P* entre especies fue de 90.6 %, pero se presentó reducida variabilidad genética dentro de cada especie (*C. argyrosperma P* = 14.7 %; *C. ficifolia P* = 14 %; *C. pepo P* = 20.8 % y en *C. moschata*= *P* 37 %). Además reportan un flujo genético bajo para todas las especies (*Nm* = 0.14) dado su alto nivel de diferenciación ($G_{st}$ = 0.77). Además, *C. argyrosperma* y *C. moschata* fueron las especies más relacionadas, con un coeficiente de identidad de 0.79. Por otro lado, *C. pepo* tuvo un menor coeficiente de identidad con *C. argyrosperma* (0.63) y *C. moschata* (0.69); la especie más alejada fue *C. ficifolia*.

Balvino-Olvera *et al.* (2017) describieron la estructura poblacional de 61 poblaciones de *C. argyrosperma* subsp. *sororia* en siete estados de la vertiente del Pacifico de México, utilizando 14 microsatélites, seis de los cuales eran específicos de la especie. El objetivo era encontrar centros de diversidad genética con fines de conservación. Estimaron una diversidad genética elevada ($H_e$ = 0.49-0.76); además encontraron una estructura latitudinal, que consistía en tres grupos: Oaxaca-Guerrero; Michoacán-Colima junto a algunas poblaciones de Oaxaca y Guerrero; y Nayarit, Sinaloa, Jalisco, Guerrero y Oaxaca. Además, encontraron cierta diferenciación entre los grupos ($F_{ST}$= 0.018 – 0.214) y sugiere que las poblaciones del sur (Guerrero $H_e$= 0.71; Oaxaca $H_e$= 0.76) pueden ser centros importantes de diversidad.

Priori *et al.* (2018) evaluaron los niveles de diversidad genética y flujo génico en nueve poblaciones de *C. argyrosperma* del estado de Jalisco con condiciones agroecológicas contrastantes. Estas incluían poblaciones riparias, arvenses y ruderales, en sitios abandonados o perturbados y una cultivada. Utilizando 20 microsatélites, estimaron que la diversidad genética es alta ($H_e$ = 0.37 – 0.65, media = 0.50), siendo mayor en las poblaciones de la subsp. *sororia* espontáneas en traspatio ($H_e$= 0.65) y en las domesticadas ($H_e$=0.63). Además, mencionan que esta mayor diversidad genética puede haber sido debido al alto flujo de genes entre estas poblaciones, promovido por los polinizadores y la selección humana. Se encontraron dos y tres grupos genéticos, correspondientes a cada subespecie, aunque estos tenían un alto flujo de genes, lo que destaca la importancia de conservar y mantener estos recursos genéticos.

Adicionalmente, existen estudios que muestran evidencia de flujo génico de *C. argyrosperma* con otras especies y subespecies de *Cucurbita* (Nabhan, 1984; Decker-Walters *et al.,* 1990; Wilson, 1990; Wilson *et al.,* 1994; Montes-Hernández y Eguiarte, 2002), incluso reportes de flujo genético ante posibles cultivos transgénicos de calabaza (Cruz-Reyes *et al.,* 2015).

## Justificación

Aunque existen trabajos que aportan información sobre la diversidad genética de *C. argyrosperma* en algunas regiones de México no se tenía previamente una perspectiva nacional del estado actual de su diversidad genética. Es importante conocer la forma en que se distribuye dicha diversidad genética en el país, para inferir el centro de origen y diversificación de la especie. Es necesario mediante el uso de marcadores moleculares, ampliar a todo el territorio la descripción de la especie, tanto de las plantas silvestres como de las cultivadas, para llenar vacíos de información existentes, sobre todo a nivel de las razas locales de *C. argyrosperma*.

Con este proyecto se pretende describir el estado actual tanto de las poblaciones de plantas silvestres como de las cultivadas, a lo largo del país e identificando condiciones que pudieran promover la pérdida de variedades locales de la subespecie cultivada *argyrosperma* y de su pariente silvestre, la subespecie *sororia,* y así contribuir a la conservación de este importante recurso fitogenético.

El presente proyecto de doctorado surgió a partir de los proyectos "Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de *Cucurbita*" (KE004), y su continuación "Diversidad genética de las especies de *Cucurbita* en México. Fase II. Genómica evolutiva y de poblaciones, recursos genéticos y domesticación" (PE001), los cuales fueron financiados por CONABIO entre 2013 y 2020, coordinados por los doctores Rafael Lira, Luis E. Eguiarte y Salvador Montes. Al inicio del proyecto se identificaron una serie de importantes vacíos de información en aspectos básicos de las diferentes especies que era necesario cubrir o actualizar, como el caso de *C. argyrosperma*.

Esta tesis está formada por tres capítulos, en los que se describe la diversidad genética y la filogeografía de *Cucurbita argyrosperma* en México. En el primer capítulo se resalta la importancia de estudiar a las calabazas como recurso fitogenético, se describe su domesticación

a través de vestigios arqueológicos en diferentes cuevas de México. Se detallan además las características de las especies de calabaza y el manejo que se hace de ellas. Finalmente, se menciona la forma en que pueden ser estudiadas usando genética de poblaciones, dando especial relevancia al proyecto del cual se desprende esta tesis.

En el segundo capítulo se describe la diversidad genética de ambas subespecies de *C. argyrosperma* en México. Se incluyen análisis de filogeografía, estructura genética y modelos de distribución potencial. El tercer capítulo estudia la domesticación de *Cucurbita argyrosperma* desde un enfoque de genómica de poblaciones, además se describe un modelo probable de domesticación para la especie. Se incluyen análisis de diversidad, de estructura genética, y de coalescencia de 26 poblaciones de ambas subespecies, utilizando marcadores genómicos (SNPs). En conjunto esta tesis muestra un panorama general del estado actual de la diversidad Genética de *Cucurbita argyrosperma* en México.

# OBJETIVOS

## Objetivo general

- Analizar la diversidad genética de *Cucurbita argyrosperma* Huber en México, tanto en sus poblaciones silvestres como en las cultivadas.

### Objetivos específicos

1. Generar mapas distribución espacial y modelos de distribución potencial y detectar sitios en los que se requiera hacer trabajo de recolección a partir de la base de datos del banco nacional de germoplasma (INIFAP-Celaya) y de la CONABIO para
2. Buscar y colectar poblaciones en las regiones predichas por los modelos de distribución potencial.
3. Determinar la variación genética de la especie y entre poblaciones mediante el uso de marcadores moleculares.
4. Mapear y describir los patrones filogeográficos y evolutivos de la diversidad genética documentada.

## CAPITULO 1

# DE LA CUEVA A LA MESA, Y AHORA AL LABORATORIO GENÓMICO: LA DIVERSIDAD DE CALABAZAS DE MÉXICO.

Publicado en **OIKOS= No. 17 en marzo de 2017.**

*Artículo*

# De la cueva a la mesa, y ahora al laboratorio genómico: la diversidad de calabazas de México

Guillermo Sánchez de la Vega

**Desde las cuevas**

Durante miles de años las cuevas fueron refugio y hogar de diferentes grupos humanos en México y en el mundo. En las décadas de 1950 y 1960 se descubrieron vestigios arqueológicos en diferentes partes de nuestro país, con lo que fue posible entender cómo las tribus de cazadores-recolectores fueron cambiando sus costumbres y comenzaron a cultivar sus alimentos (domesticándose especies como las calabazas, el maíz y el frijol), evento que permitió que las tribus permanecieran en un solo lugar. Esto fue lo que dio inicio a la formación de los primeros asentamientos humanos.

Durante los últimos 60 años, investigadores como Richard S. Macneish y Kent Flannery han liderado trabajos multidisciplinarios en las cuevas de Tamaulipas, Tehuacán y Guilá Naquitz Oaxaca (Figura 1). Estas investigaciones revelaron el hallazgo de los restos más antiguos de plantas como la calabaza, maíz, frijol y guajes, entre otras especies, y también de algunos parientes silvestres de éstas. Lo anterior ha sido crucial para comprender el origen de la agricultura y la domesticación de las plantas que son la base de nuestra alimentación. La domesticación transformó la cultura y diferentes aspectos de la biología del ser humano, iniciando el camino hacia nuestro modo de vida moderno.

El hallazgo de restos de calabazas (semillas y fragmentos de fruto) en la cueva *Guilá Naquitz* en 1966 ha permitido determinar que la calabaza es la primera especie domesticada en nuestro continente, incluso varios miles de años antes que el maíz y el frijol. En esa cueva se encontraron restos de *Cucurbita pepo* de 8,000 a.C. Además existen registros de domesticación de otras calabazas de hasta 6,500 a.C., como los de *C. argyrosperma* (calabaza criolla) en el norte de Guerrero, y algunos otros, más tardíos y en menor número, de *C. moschata* y *C. ficifolia* en diferentes cuevas. La ausencia de registros de otras especies de calabazas en Oaxaca ha sido explicada debido a que

Figura 1. Entrada a la cueva de Guila Naquitz, Oaxaca. (Fotografía: Alejandro Linares Garcia CC BY-SA 4.0-3.0-2.5-2.0-1.0 vía *Wikipedia Commons*).

estas plantas no toleran las temperaturas frías y secas de esta región. Hasta ahora, con los datos que han podido obtenerse se sabe que las calabazas *C. pepo* aparecieron en las cuevas de la región de Oaxaca y la depresión del Balsas (8,000 a.C.) y después se desplazaron al norte pasando por Tehuacán (5,900 a.C.) y Tamaulipas (4,300 a.C.) hasta llegar al suroeste de Estados Unidos (1,500 a.C.).

Aunque las cuevas de Tamaulipas (Valenzuela y Romero), Tehuacán (Coxcatlán y San Marcos) y Oaxaca (Guilá Naquitz) solo representan tres puntos separados de un gran número de sitios posibles que podrían aportar información, en estos lugares

Figura 2. Calabazas de la región de Ures, Sonora. Fotografía: Guillermo Sánchez de la Vega.

se han encontrado los registros de domesticación más antiguos de los tres cultivos más importantes en Mesoámerica. Por esta razón la fama de las cuevas ha llegado incluso a nivel mundial: la cueva Guilá Naquitz, en conjunto con algunas otras cuevas en las inmedaciones de Mitla y Yagul, han sido incluídas en la categoría de Paisaje Cultural, dentro de la lista de patrimonios de la humanidad de la UNESCO desde 2010.

**Llegando a la mesa**

En Mesoamérica la agricultura se inició hace unos 10 mil años. Los primeros agricultores utilizaron plantas silvestres en sus cultivos, y conforme las fueron seleccionando, se generaron las especies domesticadas, las cuales resultan genética y morfológicamente distintas a sus parientes silvestres. La domesticación de la calabaza tomó unos 6 mil años, y tuvo que pasar todavía más tiempo para que los primeros asentamientos humanos se formaran. La dieta de estos asentamientos dependía principalmente del cultivo de diversas especies domesticadas. Hoy en día en México tenemos alrededor de 200 especies nativas en estado avanzado de domesticación, incluyendo algunas especies de plantas de importancia mundial tales como la calabaza, maíz, jitomate, frijol, chile, cacao, agaves, entre muchas otras.

Con el paso del tiempo, el manejo y cultivo de las calabazas se fue perfeccionando hasta predominar como medio de subsistencia. Así, por miles de años, las calabazas se han cultivado en casi todas las regiones agrícolas de México bajo el sistema tradicional de cultivo denominado *milpa*, y como consecuencia de ello los agricultores han logrado desarrollar una elevada diversidad genética (Figura 2). Por esta razón México es reconocido como el centro de origen, domesticación y diversidad de calabazas (Figura 3).

Como consecuencia de este manejo ancestral y actual, existen una gran cantidad de razas y variedades nativas de calabaza, que son nombradas de forma diferente de acuerdo al país o a la región de la que provienen: calabaza de castilla, de casco duro, calabacitas, arotas, criollas, mixtas, tamalayotas, pipianas, chompas, kaboshas y chilacayotes. Las cinco especies de calabazas han sido parte de la dieta de los pueblos de Latinoamérica y posteriormente de otras zonas del mundo, y hoy en día se utiliza prácticamente cada una de las partes de la planta, desde sus "guías" o brotes tiernos y flores, que son consumidos en diferentes guisos, también los frutos inmaduros que se usan como verdura, hasta los frutos maduros utilizados en postres y dulces; en la gastronomía mexicana, las semillas o pepitas (que son una fuente de proteínas que es posible almacenar por periodos prolongados de tiempo) se consumen crudas, tostadas y/o saladas, y son base de toda una gama de moles y pipianes; también cabe resaltar el uso de los frutos en forma de contenedores y artesanías, o aprovechados como forraje (Figura 4).

Las calabazas son plantas generalmente anuales y forman parte de la familia Cucurbitaceae, a la cual también pertenecen la sandía, el melón, el pepino, el chayote y el estropajo, entre unas 965 especies. Dentro de esta familia, las calabazas conforman al género *Cucurbita* que se reconoce como un grupo de plantas americano, que incluye a 20 especies de las cuales 15 crecen espontáneamente o se cultivan en México (Figura 5). Las especies de ca-


Figura 3. A) Calabaza *C. argyrosperma* ssp. *sororia* creciendo sobre *Zea mays* ssp. *parviglumis*, en Jalisco. Ambas subespecies son parientes silvestres de especies de calabaza y maíz domesticadas. B) Productor de la variedad cultivada *C. argyrosperma* ssp. *argyrosperma*, de la región de Autlán, Jalisco. Fotografías: Guillermo Sánchez de la Vega.

Figura 4. Diversidad de calabazas de la península de Yucatán. Fotografía: Guillermo Sánchez de la Vega.

labazas domesticadas son *Cucurbita argyrosperma*, conocida como calabaza criolla o pipiana (Figura 6); *C. ficifolia*, el chilacayote; *C. moschata*, la calabaza de castilla (Figura 7); *C. pepo* y *C. maxima*, kabosha, la calabacita. En los cultivos de ciertas zonas del país, estas especies conviven con sus parientes silvestres, que en algunos casos son endémicos de pequeñas regiones.

Las diferentes especies están adaptadas a distintas condiciones ambientales. Por ejemplo, *C. argyrosperma* es una especie cultivada (y también puede encontrarse en estado silvestre), que vive generalmente en zonas por debajo de los 1,800 m con climas cálidos y algo secos. *C. ficifolia*, sólo conocida en estado cultivado, prefiere altitudes mayores, de unos 1,300 m o más, en zonas templadas a frías. *C. moschata* se cultiva en casi todo el continente americano en lugares cálidos, con altitud menor de 1,000 m sobre el nivel del mar. *C. pepo* crece en lugares con altitud superior a 1,000 m y debido a su capacidad de hibridizar con otras especies tiene gran importancia y uso en fitomejoradores. Finalmente, *C. maxima* se cultiva en lugares con clima templado a seco, y se considera que se domesticó en América del Sur. Esta especie no se cultivaba en nuestro país, pero en los últimos años se ha introducido en el noroeste de México, y su producción ha sido principalmente para los mercados de exportación.

En las últimas décadas se ha visto reducida la diversidad genética de las variedades locales de calabaza. Una de las causas ha sido el abandono de su cultivo, ya que las calabazas locales han sido desplazadas por variedades nuevas y más productivas; en consecuencia, algunas variedades locales se han mantenido aisladas, y otras aún requieren selección y mejora de aquellas características sobresalientes y deseables para los agricultores y consumidores. El noreste de México es un ejemplo del desplazamiento hacia variedades nuevas: el cultivo de calabazas para exportación se ha incrementado considerablemente en las últimas décadas, porque representa una mejor opción de producción y obtención de divisas para los productores mexicanos.

En México el cultivo de calabazas comerciales inició hace aproximadamente 25 años en las regiones agrícolas de Sonora y Sinaloa principalmente. Algunas variedades comerciales de *C. pepo* han sido introducidas en el mercado de Latinoamérica, donde se usa como un sustituto de las tradicionales; y otras variedades de esta misma especie se siembran para exportación, como la calabaza *butternut* y *spaghetti*, cuya producción se destina a los Estados Unidos y Canadá con motivo del día de Acción de Gracias. Toda la producción de la calabaza *kabocha* (más de 50,000 toneladas), también conocida como *calabaza japonesa*, se destina al Japón, donde tiene el mayor consumo per capita a nivel mundial, y es un alimento tradicional, particularmente durante los festivales asociados al solsticio de invierno.

Figura 5. *C. lundelliana* especie silvestre creciendo en Calakmul, Campeche. Fotografía: Guillermo Sánchez de la Vega.

**Trabajando en la genética y la genómica**

A pesar de la importancia biológica y cultural de las calabazas, han sido pocos los esfuerzos encaminados a mostrar un panorama de la diversidad genética de las calabazas (con excepción de las variedades que han sido mejoradas para la producción de verdura (*C. pepo*)), si se compara con otras especies de la milpa. La mayoría de los estudios de diversidad genética se han enfocado en cuantificar atributos morfológicos, por lo que el nivel de conocimiento actual de la variación genética en México es todavía muy limitado. Los estudios genéticos y genómicos son necesarios para aportar información sobre dónde y cómo se domesticaron las calabazas, y cómo fue su posterior dispersión a diferentes regiones, en las que estas plantas se fueron adaptando a las prácticas de manejo y clima de cada localidad.

Recientemente, se ha renovado el interés para estudiar a los taxa silvestres de especies domesticadas, porque las plantas silvestres poseen una mayor diversidad genética (razón por la cual se les considera depósitos genéticos), que puede utilizarse para programas de mejoramiento de cultivos. Es importante determinar cómo está organizada la diversidad genética dentro y entre las poblaciones de estos parientes silvestres, buscando patrones genéticos a lo largo de su área de distribución geográfica, que pueden ser consecuencia de diversos procesos evolutivos tales como flujo génico, deriva genética, mutación y selección, entre otros. Con las herramientas genómicas que se han desarrollado a gran escala, y gracias a su relativo bajo costo, ha sido posible realizar proyectos multidisciplinarios de evaluación de la diversidad genética y la domesticación, que nuestro equipo de trabajo también está realizando. Los datos moleculares que hemos obtenido recientemente para mi tesis de doctorado, y en otros estudios que se realizan en el Instituto de Ecología y la fES Iztacala, UNAM, bajo la dirección de los doctores Luis Eguiarte y Rafael Lira, respectivamente, confirman una gran diversidad existente en las calabazas. Hemos encontrado que la diversidad genética presenta un patrón de "aislamiento por distancia" (mientras más lejanas las poblaciones, menos se parecen genéticamente), lo que sugiere que procesos como el intercambio de genes y la hibridación son muy importantes para las variedades domesticadas y sus parientes silvestres. A corto plazo, evaluaremos la diversidad genética para cada especie, con el fin de determinar cómo las prácticas de manejo agrícola de cada región influyen en esta diversidad.

Así como las cuevas han sido refugio y una fuente de información sobre la domesticación y el origen de las calabazas y la agricultura, los laboratorios e instituciones especializados en ciencias genéticas y genómicas se han convertido en modernos refugios de germoplasma, y refugios de la diversidad de estas es-

Figura 6. *C. moschata* (izquierda) y *C. argyrosperma* de la zona maya de Xbec, Yucatán. Fotografía: Guillermo Sánchez de la Vega.

Figura 7. *C. moschata* del sur de Quintana Roo. Fotografía: Guillermo Sánchez de la Vega.

pecies: permiten llenar los vacíos de información sobre aspectos genéticos, genómicos y etnobotánicos, ya sea documentando la presencia de las especies domesticadas y sus parientes silvestres a lo largo del país, o evaluando la diversidad genética que poseen y cómo está organizada. Además, permiten el estudio de genes particulares relacionados con la domesticación.

Conocer la diversidad genética y la forma en que se adaptan las calabazas en diferentes regiones, constituye un elemento de gran importancia para preservar áreas donde coexisten variedades cultivadas y sus parientes silvestres, debido a que estos últimos pueden ser la base para futuros programas de mejoramiento de cultivos actuales. La diversidad está estrechamente relacionada con las condiciones ambientales y las diferentes formas de usos de la calabaza. Para determinar las estrategias de conservación y aprovechamiento de estas plantas, es importante reconocer la variación morfológica y genética, así como la relación que esta variación guarda con el ambiente, y el impacto social que trae consigo. Por ello las estrategias de conservación deben estar vinculadas a su manejo, tanto tecnificado como tradicional, y a una política de conservación *in situ* y *ex situ* en bancos de germoplasma.

Para conservar la diversidad de las calabazas tanto cultivadas y silvestres, la conservación, evaluación, documentación, y la fácil disponibilidad de los recursos genéticos, son estrategias que se están volviendo cada vez más indispensables. Todo lo anterior requiere proyectos con financiamiento concreto, que definan claramente una serie de acciones interrelacionadas de conservación, adicionados con estudios especializados.

**Agradecimientos**

El presente trabajo se realizó con apoyo del Posgrado en Ciencias Biológicas de la UNAM y el programa de Becas para Estudios del Posgrado del CONACYT (Número de becario 292164), además de CONABIO a través del proyecto "Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de *Cucurbita*" (clave KE004).

**Guillermo Sánchez de la Vega.** Estudió la Licenciatura en Biología en la FES Iztacala y la Maestría en Ciencias en el Posgrado de Ciencias Biológicas de la UNAM. Fue Profesor Investigador de la UMAR en la costa de Oaxaca durante varios años y cuenta con experiencia en proyectos de manejo y conservación de recursos naturales. Actualmente estudia el Doctorado en Ciencias en el Laboratorio de Evolución Molecular y Experimental del Instituto de Ecología, en el proyecto *Diversidad Genética de* Cucurbita argyrosperma *en México*, bajo la dirección del Dr. Luis E. Eguiarte.

**Para saber más**


- Bellón, M.R., *et al*. 2009. *Diversidad y conservación de recursos genéticos en plantas cultivadas.* Pp. 355-382, *en: Capital natural de México*. Vol. II: Estado de conservación y tendencias de cambio. CONABIO, México.
- Flannery, K.V. 2002. "Había gigantes en aquellos días". Richard Stockton MacNeish, 1918-2001. *Arqueología*, (27), 113-127.
- Flannery, K.V. 1999. Los orígenes de la agricultura en Oaxaca. *Cuadernos del Sur*, 5:5-14.
- Lira, R., Casas, A. y J., Blancas. (eds.). 2016. *Ethnobotany of Mexico: Interactions of People and Plants in Mesoamerica*. Springer.
- Zeder, M.A. 2006. *Documenting domestication: New Genetic and Archaeological Paradigms*. University of California Press.

# CAPITULO 2

# DIVERSIDAD GENÉTICA Y FILOGEOGRAFÍA DE *Cucurbita argyrosperma* Huber EN MÉXICO CON MICROSATÉLITES

**Artículo:**

Genetic Resources in the “Calabaza Pipiana” Squash *Cucurbita argyrosperma*) in Mexico: Genetic Diversity, Genetic Differentiation and Distribution Models.

Publicado en 2018 en: *Frontiers in Plant Science*, vol. 9, p. 400.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 29 March 2018
doi: 10.3389/fpls.2018.00400

# Genetic Resources in the "Calabaza Pipiana" Squash (*Cucurbita argyrosperma)* in Mexico: Genetic Diversity, Genetic Differentiation and Distribution Models

*Guillermo Sánchez-de la Vega[1], Gabriela Castellanos-Morales[1,2,3], Niza Gámez[1], Helena S. Hernández-Rosales[1], Alejandra Vázquez-Lobo[1,4], Erika Aguirre-Planter[1], Juan P. Jaramillo-Correa[1], Salvador Montes-Hernández[5], Rafael Lira-Saade[2]* and Luis E. Eguiarte[1]**

*[1] Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico, Mexico, [2] Unidad de Biotecnología y Prototipos, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Mexico, Mexico, [3] Departamento de Conservación de la Biodiversidad, El Colegio de la Frontera Sur, Villahermosa, Mexico, [4] Centro de Investigación en Biodiversidad y Conservación, Universidad Autónoma del Estado de Morelos, Cuernavaca, Mexico, [5] Campo Experimental Bajío, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Celaya, Mexico*

OPEN ACCESS

*Edited by:*
*Charles Roland Clement, National Institute of Amazonian Research, Brazil*

*Reviewed by:*
*Yong Liu, Hunan Academy of Agricultural Sciences (CAAS), China*
*Maria Isabel Chacon Sanchez, Universidad Nacional de Colombia, Colombia*

**Correspondence:*
*Rafael Lira-Saade*
*rlira@unam.mx*
*Luis E. Eguiarte*
*fruns@unam.mx*

*Specialty section:*
*This article was submitted to Agroecology and Land Use Systems, a section of the journal Frontiers in Plant Science*

*Received: 17 August 2017*
*Accepted: 13 March 2018*
*Published: 29 March 2018*

*Citation:*
*Sánchez-de la Vega G, Castellanos-Morales G, Gámez N, Hernández-Rosales HS, Vázquez-Lobo A, Aguirre-Planter E, Jaramillo-Correa JP, Montes-Hernández S, Lira-Saade R and Eguiarte LE (2018) Genetic Resources in the "Calabaza Pipiana" Squash (Cucurbita argyrosperma) in Mexico: Genetic Diversity, Genetic Differentiation and Distribution Models. Front. Plant Sci. 9:400. doi: 10.3389/fpls.2018.00400*

Analyses of genetic variation allow understanding the origin, diversification and genetic resources of cultivated plants. Domesticated taxa and their wild relatives are ideal systems for studying genetic processes of plant domestication and their joint is important to evaluate the distribution of their genetic resources. Such is the case of the domesticated subspecies *C. argyrosperma* ssp. *argyrosperma,* known in Mexico as *calabaza pipiana*, and its wild relative *C. argyrosperma* ssp. *sororia*. The main aim of this study was to use molecular data (microsatellites) to assess the levels of genetic variation and genetic differentiation within and among populations of domesticated *argyrosperma* across its distribution in Mexico in comparison to its wild relative, *sororia*, and to identify environmental suitability in previously proposed centers of domestication. We analyzed nine unlinked nuclear microsatellite loci to assess levels of diversity and distribution of genetic variation within and among populations in 440 individuals from 19 populations of cultivated landraces of *argyrosperma* and from six wild populations of *sororia*, in order to conduct a first systematic analysis of their genetic resources. We also used species distribution models (SDMs) for *sororia* to identify changes in this wild subspecies' distribution from the Holocene (~6,000 years ago) to the present, and to assess the presence of suitable environmental conditions in previously proposed domestication sites. Genetic variation was similar among subspecies ($H_E$ = 0.428 in *sororia*, and $H_E$ = 0.410 in *argyrosperma*). Nine *argyrosperma* populations showed significant levels of inbreeding. Both subspecies are well differentiated, and genetic differentiation ($F_{ST}$) among populations within each subspecies ranged from 0.152 to 0.652. Within *argyrosperma* we found three genetic groups (Northern Mexico, Yucatan Peninsula, including Michoacan and Veracruz, and Pacific coast plus Durango). We detected low levels of gene flow among populations at a regional scale (<0.01), except

for the Yucatan Peninsula, and the northern portion of the Pacific Coast. Our analyses suggested that the Isthmus of Tehuantepec is an effective barrier isolating southern populations. Our SDM results indicate that environmental characteristics in the Balsas-Jalisco region, a potential center of domestication, were suitable for the presence of *sororia* during the Holocene.

Keywords: *Cucurbita*, cultivated squash, genetic diversity, genetic structure, nuclear microsatellites, species distribution models

## INTRODUCTION

Domestication is an ideal model to study evolution because it is usually fast and gradual (Purugganan and Fuller, 2011; Meyer and Purugganan, 2013; Gaut, 2015; Gaut et al., 2015). Population genetics studies allow to analyze the dynamics of the domestication process and to make inferences about the origins and histories of crops (Meyer and Purugganan, 2013; Aguirre-Liguori et al., 2016).

Sometimes the ancestral wild populations can still be studied along with the domesticated forms and varieties, allowing paired comparisons between populations under different selection processes in the same environment (Aguirre-Liguori et al., 2016). Also, the coexistence and possibility of hybridization of domesticated taxa and their wild relatives allows having a source of genetic variation during domestication, increasing genetic diversity and the presence of alleles of agronomic value (Warschefsky et al., 2014). Nevertheless, the possibility of hybridization raises questions, such as: (1) How do domesticated and wild relatives remain genetically differentiated? and (2) How frequent is introgression among wild and domesticated relatives? It is also important to mention that signals of domestication may be confused by long-distance human-mediated dispersal and by intermittent crosses between domesticated and wild taxa, sometimes making it difficult to disentangle the history of domestication (Besnard et al., 2013; Meyer and Purugganan, 2013). As domestication and crop improvement involve genetic bottlenecks (Gaut et al., 2015), they can lead to a reduction of genetic diversity and increased inbreeding. During domestication, crops are transported from their center of domestication to new environments, which may lead to new local adaptation that in some cases can be achieved through introgression with their wild relatives or other domesticated relatives (Gaut et al., 2015).

The mechanisms for the maintenance of the genetic differentiation among domesticated populations and their wild relatives have seldom been studied. It has been proposed that in some species, such as *Cucurbita argyrosperma* and *Zea mays*, gene flow is asymmetric, being more frequent from the wild to the domesticated taxa (Montes, 2002; Hufford et al., 2013). Moreover, Hufford et al. (2013) found resistance to gene flow from domesticated maize into wild teosinte, which could be explained by low gene flow rates, by the fact that domesticated genes are not advantageous for wild taxa, or strong selection by humans against hybrids. Cruz-Reyes et al. (2015) observed that domesticated-wild hybrids of *Cucurbita* showed lower reproductive output. Hufford et al. (2013) found that many alleles that characterize domesticated varieties are found at lower frequencies in their wild relatives, suggesting that the attributes associated with domestication are not produced by *de novo* mutations, but constitute part of the standing genetic variation of wild taxa (Doebley et al., 2006).

Surveys of genetic variation of wild populations and their cultivated relatives is a first step for the description of genetic resources, such as analyzing how much genetic variation is still found in domesticated taxa compared to their wild relatives, their degree of differentiation, and evaluating how much ancestral and ongoing gene flow (hybridization) exists among wild and domesticated taxa (Warschefsky et al., 2014). These topics are relevant for the management of domesticated populations and for the future preservation of genetic resources (Warschefsky et al., 2014; Govindaraj et al., 2015). Moreover, these comparative studies are the first step toward understanding the origin and diversification of domesticated plant taxa. Current molecular tools, along with population genetics and modern phylogeographic approaches, allow understanding the distribution of genetic variation from an evolutionary perspective (Eguiarte et al., 2013; Aguirre-Dugua and González-Rodríguez, 2016; Aguirre-Liguori et al., 2016).

The study of crop origins has traditionally involved identifying geographic areas of high diversity and sampling populations of wild progenitor species (Kraft et al., 2014). Linking genes, crops, and landscapes through a geographical analysis of genetic data is one important way to achieve multilevel integration (Van Etten and Hijmans, 2010; Hufford et al., 2012; Besnard et al., 2013). Furthermore, species distribution modeling, projected into past conditions, offers a view of the potential geographic pattern of taxa during the domestication process (Hufford et al., 2012; Kraft et al., 2014).

The genus *Cucurbita* (pumpkins, squashes, and gourds), with 20 taxa of perennial or annual plants, is native to the Americas. Mexico is considered its center of origin and diversification (Lira-Saade, 1995; Mapes and Basurto, 2016). *Cucurbita* represents an interesting system for the study of domestication (Lira et al., 2016b) with five different domesticated species: *C. pepo*, *C. moschata*, *C. ficifolia*, *C. maxima*, and *C. argyrosperma* (Wilson et al., 1992; Sanjur et al., 2002; Kocyan et al., 2007; Gong et al., 2013; Zheng et al., 2013; Kates et al., 2017). Cucurbits were some of the first plants domesticated in the Americas, ca. 10,000 years ago (Smith, 1997; Zizumbo-Villarreal et al., 2016). Within the genus *Cucurbita*, each domestication event occurred independently, sometimes on more than one occasion

(Lira et al., 2016b). Today, domesticated cucurbits still have a fundamental role in the diet of people in Mexico, Central and South America, and in many other regions of the world, and they are considered an essential phytogenetic resource (FAO, 2010).

Among domesticated cucurbits, *C. argyrosperma*, known in Mexico as *calabaza pipiana* or *calabaza mixta*, is highly appreciated for its seeds, which are used in Mexican gastronomy. Also, fruits are medicinal, commercial, and food resources (Lira-Saade, 1995; Villanueva, 2007). It is a species with cultural and economic importance both locally and worldwide. The oldest evidence of domestication for this species is ∼8,600 years old from the Xihuatoxtla shelter, in the state of Guerrero (Rannere et al., 2009). This is a highly diverse species in form, color and size of its seeds and fruits (Lira-Saade, 1995; **Figure 1**). *C. argyrosperma* is currently divided into two subspecies: the domesticated *C. argyrosperma* ssp. *argyrosperma* (*argyrosperma* hereafter) and its wild relative *C. argyrosperma* ssp. *sororia* (*sororia* hereafter; **Figure 1**) (Organization for Economic Co-operation and Development [OECD], 2012; Gong et al., 2013; Zheng et al., 2013; Kates et al., 2017). Both wild and cultivated subspecies can be found in tropical and semi desert regions from the Southeastern United States through Mexico and northern Central America, reaching Nicaragua, from sea level to 1,700 m above sea level (Villanueva, 2007; Lira et al., 2016b). These subspecies have a sympatric distribution in most of their range, except for the Yucatan peninsula, where the wild subspecies is absent (Lira-Saade, 1995; Organization for Economic Co-operation and Development [OECD], 2012; **Figure 2**).

*Cucurbita argyrosperma* is an important crop in local agriculture systems in Mexico and in other countries in the Americas. It is grown and selected in traditional ways. It is commonly found as a seasonal crop, but irrigation is used in some areas. A large amount of its production is not reported because it is used in subsistence agriculture in Mexico and Central and South America (Montes, 1991, 2002; Villanueva, 2007; Organization for Economic Co-operation and Development [OECD], 2012). In other regions of the world it is not extensively cultivated because of the low quality of its flesh (Lira-Saade, 1995; Organization for Economic Co-operation and Development [OECD], 2012), but there are records of some genetically improved cultivars grown in the United States and Canada. Some improved lines show differences in fruit and seed size, shape, and color, such as "Green Striped Cushaw," "White Cushaw," "Magdalena Striped," "Papago," "Japanese Pie," "Hopi," "Taos," "Parral Cushaw," "Veracruz Pepita," and "Silver Seed Gourd" (Organization for Economic Co-operation and Development [OECD], 2012).

Nevertheless, there are few studies focused on analyzing the genetic resources of cucurbits and covering most of their distributions (Bellon et al., 2009; Lira et al., 2016b). Only a few studies have analyzed the genetic variation of *C. argyrosperma*, including an analysis at a local scale (a region in the state of Jalisco) using isozymes (Montes-Hernández and Eguiarte, 2002), which found that *argyrosperma* has less genetic variation ($H_E$ = 0.35–0.41) than its wild relative ($H_E$ = 0.433), and low levels of genetic differentiation among populations ($F_{ST}$ = 0.077). Two studies, one based on isozymes in commercial cultivars (Decker-Walters et al., 1990), and another with accessions using RAPDs (Cerón et al., 2010), found lower genetic diversity in *argyrosperma* ($H$ = 0.039 and 0.063 for isozymes and RAPDs, respectively) than in other domesticated taxa of the genus, such as *C. moschata* ($H$ = 0.052 and 0.11 for isozymes and RAPDs, respectively) and in *C. pepo* ($H$ = 0.068 and 0.104 for isozymes and RAPDs, respectively) (Decker-Walters et al., 1990; Cerón et al., 2010). Recently, Balvino-Olvera et al. (2017) studied wild populations of *sororia* along the Pacific coast of Mexico with microsatellites, reporting high levels of genetic variation ($H_E$ = 0.756) and higher genetic diversity and heterogeneity among southern populations in the states of Guerrero and Oaxaca. Clearly, there is a lack of population-based genetic diversity analyses that include both the cultivated and wild *C. argyrosperma* throughout its range.

The main aim of this study was to use molecular data (microsatellites) to assess the levels of genetic variation within and among populations of domesticated *argyrosperma* across its distribution in Mexico. We also analyzed populations of its wild relative, *sororia*, to compare levels of genetic variation and differentiation. Additionally, we estimated the levels of recent gene flow among populations and subspecies, and performed projections of the wild subspecies' distribution area in the mid-Holocene (∼6,000 years ago), in order to identify environmental suitability in previously proposed domestication centers, such as the Balsas-Jalisco region based on archeological records (Sanjur et al., 2002; Piperno et al., 2009; Rannere et al., 2009). We expected to find lower genetic diversity and higher levels of inbreeding in cultivated *argyrosperma* than in its wild relative *sororia*, in accordance with a previous study (Montes-Hernández and Eguiarte, 2002). In addition, we expected to find lower genetic differentiation among populations in the same geographic area than in more distant areas, and signals of on-going gene flow among subspecies, as reported by Montes (2002) and Montes-Hernández and Eguiarte (2002).

# MATERIALS AND METHODS

## Sampling and DNA Extraction

We obtained seeds from at least 3 fruits (one fruit from each different plant) from 19 populations of cultivated landraces of *argyrosperma* and 6 wild populations of *sororia* representative of the species distribution in Mexico (**Figure 2** and Supplementary Table S1). Ten populations were obtained from the germplasm collection of the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Campo Experimental Bajío, in 2014 (BG in Supplementary Table S1). Fruits from 15 additional populations were collected in the field between 2013 and 2015. Sampled wild populations were located close to cultivars of *argyrosperma* to assess levels of gene flow among subspecies. All the collected fruits were stored in greenhouse conditions at the Institute of Ecology, UNAM, until they became ripe. Between 5 and 20 seeds from each collected fruit were grown in commercial substrate under greenhouse conditions (35°C in average) for 40 days, and young leaves were collected for DNA extraction.

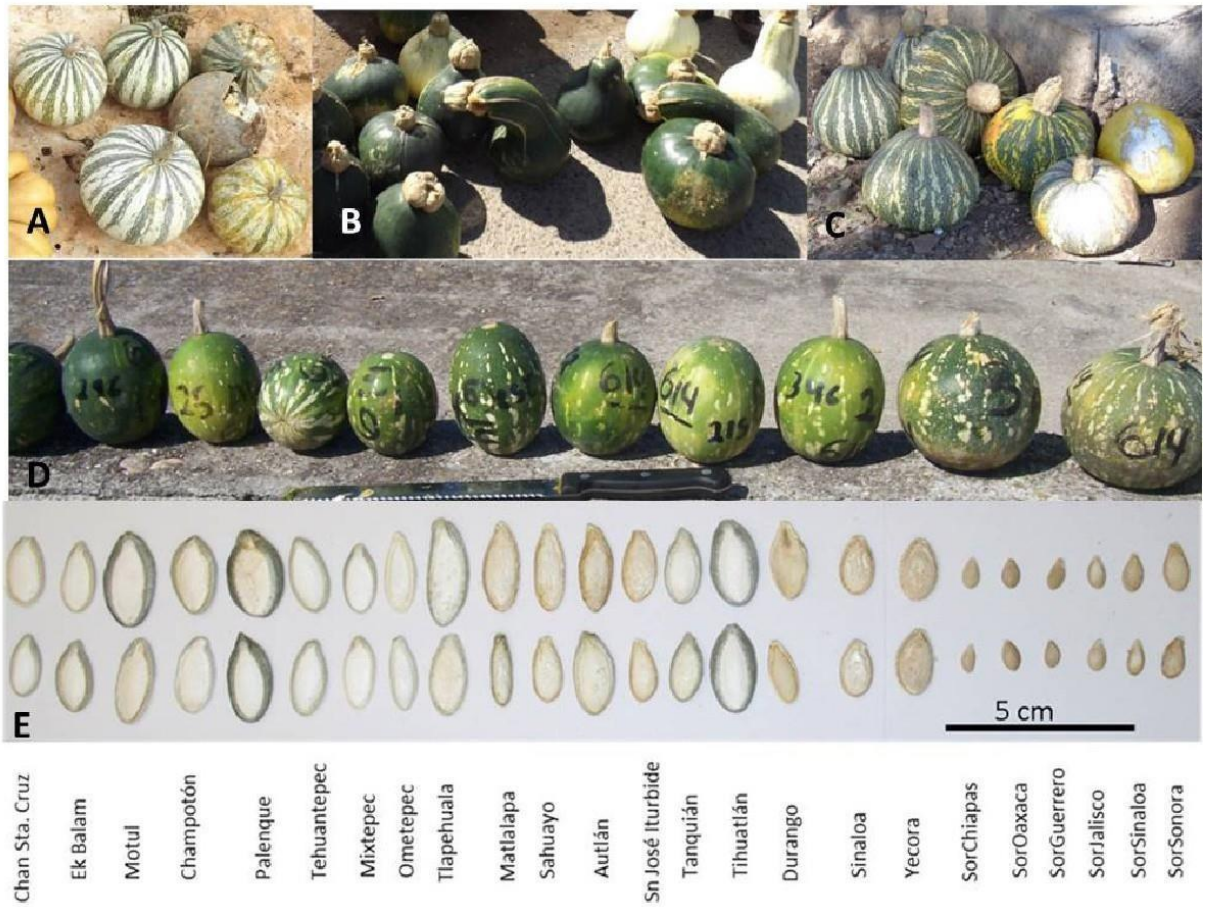



FIGURE 1 | Morphological characteristics of subspecies *C. argyrosperma* ssp. *argyrosperma* from Yucatán (A), Sonora (B) and Jalisco (C), and *C. argyrosperma* ssp. *sororia* (D) from Jalisco. Seeds from each population are shown at the bottom (E). Population ID is shown in Table 1.

DNA extraction was performed using a modified CTAB protocol (Doyle and Doyle, 1987). For nuclear microsatellite loci, we genotyped a total of 440 individuals, 327 of which were attributed to *argyrosperma* and 113 to *sororia*.

## Microsatellite Analyses

We amplified 12 of the nuclear microsatellite loci reported by Gong et al. (2008) for *C. pepo* (Supplementary Table S2). Loci were selected from different chromosomes to improve genome coverage and to reduce the probability of linkage disequilibrium. For better results, we selected only highly variable dinucleotide loci. We used a multiplex approach for microsatellite amplification in a 15 µl final volume, consisting of $1\times$ Buffer, 1.2 mM $MgCl_2$, 0.2 mM of dNTPs, 0.13 µM of each primer (six primers per multiplex, forward primers were marked with one of the following fluorescent dyes: 6-FAM, HEX and VIC), 1 µl of Taq polymerase (PROMEGA) and 10 ng of genomic DNA. Amplification reactions were performed in a Veriti 96-well Thermal cycler (Applied Biosystems) with the following program: 95°C for 5min, followed by 35 cycles of 95°C for 40s, 60°C (Ta) for 40s, 75°C for 55s, and a final step of 72°C for 5min followed by 4°C. To control for possible contamination, we used blank controls for each reaction. All products were verified in 2 % agarose gels and PCR products were sent to the Roy J. Carver Biotechnology Center at the University of Illinois, United States for genotyping[1] . Electropherograms were analyzed with PeakScanner (Applied Biosystems) to build a matrix with the genotypes of each individual.

## Null Alleles and Measures of Genetic Diversity

We conducted a null allele analysis using the method proposed by Chakraborty et al. (1992) implemented in the Microchecker v2.2.3 (Van Oosterhout et al., 2004). In addition, we performed a Hardy-Weinberg exact test and a linkage disequilibrium test using Arlequin v. 3.0 (Excoffier et al., 2005). We obtained allele frequencies by direct estimation using Arlequin v. 3.0, and determined the number of private alleles for each population and subspecies by direct count from the allele frequencies data. We also obtained descriptive statistics, such as the proportion

[1] www.biotech.illinois.edu

of polymorphic loci per population (*P*), allelic richness (*A*), and the expected ($H_E$) and observed ($H_O$) heterozygosities with the same software, and estimated the inbreeding coefficient ($F_{IS}$) for each population using Genepop 4.0 (Rousset, 2008). In addition, we obtained the rarefied allelic richness with ADZE 1.0 (Szpiech et al., 2008) accounting for the lowest population size of six individuals.

## Genetic Differentiation and Genetic Structure

To assess the genetic structure among subspecies and among populations we used Structure v 2.3.4 (Pritchard et al., 2000). This program uses Bayesian probability to assign individuals to different genetic clusters (*K*) based on allele frequencies without considering the population of origin. We performed previous runs to assess the best combination of priors to be used for the analysis and the length of the Markov Chain Monte Carlo (MCMC) chains. Accordingly, we performed a final run with admixture and correlated allele frequencies as priors, and without considering the putative population of origin of each individual. We used a burn-in of 500,000 chains and 1,000,000 MCMC chains, and tested values of *K* from 1 to 10, and 10 repetitions for each *K*. The results were run through Structure Harvester v 0.6.93 (Earl and vonHoldt, 2012), and the results from the Evanno test (Evanno et al., 2005) were considered to determine the value of *K* that showed fit to our data. We performed an analysis of molecular variance (AMOVA; Excoffier et al., 1992) considering the genetic clusters obtained with Structure.

As an additional test to identify the number of genetic groups formed by our data, we used the adegenet library to perform discriminant analysis of principal components (DAPC; Jombart et al., 2010). DAPC is a multivariate analysis that summarizes the genetic differentiation between groups. This analysis identifies genetically related individuals by partitioning the within group and among group genetic variation (Jombart et al., 2010). We performed two independent DAPC analyses. The first analysis included all individuals to assess the relationship among subspecies. We conducted a cross-validation test to determine the number of PCs to be retained. Accordingly, we retained 40 PCs and two discriminant functions. For the second analysis, we excluded the populations that showed high genetic differentiation to allow depicting the relationship among populations within *argyrosperma*. We retained 25 PCs and two discriminant functions in accordance to the cross-validation test.

TABLE 1 | Genetic diversity obtained for 9 nuclear microsatellite loci for *C. argyrosperma* ssp. *argyrosperma* and *C. argyrosperma* ssp. *sororia*.

| State | Population | Population ID | *n* | *P* | *A* | *RA* | *Pa* | $H_O$ | $H_E$ | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Quintana Roo | Chan_Sta_Cruz | (1) Chan | 21 | 0.44 | 4.0 (2.1) | 1.6 (0.76) | 2 | 0.318 (0.21) | 0.524 (0.19) | 0.39* |
| Yucatán | Ek_Balam | (2) Ek | 22 | 0.77 | 2.1 (0.37) | 1.4 (0.16) | 0 | 0.311 (0.29) | 0.250 (0.16) | −0.253* |
| Yucatán | Motul | (3) Mot | 22 | 0.88 | 2.75 (0.70) | 1.6 (0.28) | 0 | 0.278 (0.21) | 0.303 (0.20) | 0.08 |
| Campeche | Champoton | (4) Champ | 22 | 0.77 | 3 (2.2) | 1.7 (0.78) | 1 | 0.246 (0.19) | 0.342 (0.25) | 0.29* |
| Chiapas | Palenque | (5) Pal | 20 | 0.88 | 2.62 (0.73) | 1.8 (0.34) | 2 | 0.437 (0.29) | 0.369 (0.21) | −0.18* |
| Oaxaca | Tehuantepec | (6) Teh | 20 | 0.88 | 3.3 (1.9) | 1.8 (0.27) | 4 | 0.316 (0.17) | 0.348 (0.16) | 0.08 |
| Oaxaca | Mixtepec | (7) Mix | 21 | 0.77 | 2.4 (0.535) | 1.6 (0.23) | 1 | 0.292 (0.22) | 0.331 (0.18) | 0.11 |
| Guerrero | Ometepec | (8) Ome | 17 | 0.88 | 2.6 (1) | 1.8 (0.32) | 0 | 0.525 (0.31) | 0.420 (0.18) | −0.26* |
| Guerrero | Tlapehuala | (9) Tla | 10 | 0.77 | 2.8 (0.9) | 1.8 (0.66) | 0 | 0.428 (0.33) | 0.462 (0.23) | 0.078 |
| Guerrero | Matlalapa | (10) Mtp | 18 | 0.88 | 2.5 (0.75) | 1.8 (0.28) | 0 | 0.494 (0.27) | 0.422 (0.17) | −0.173* |
| Michoacán | Sahuayo | (11) Sah | 22 | 0.66 | 2.6 (0.81) | 1.6 (0.31) | 0 | 0.454 (0.25) | 0.374 (0.18) | −0.22* |
| Jalisco | Autlán | (12) Aut | 21 | 0.66 | 4.3 (2.8) | 1.9 (1.1) | 2 | 0.365 (0.24) | 0.449 (0.28) | 0.19* |
| Guanajuato | San José Iturbide | (13) SJI | 6 | 0.66 | 2.1 (0.40) | 1.6 (0.22) | 0 | 0.472 (0.28) | 0.482 (0.12) | 0.022 |
| San Luis Potosí | Tanquian | (14) Tan | 22 | 0.77 | 2.7 (1.4) | 1.7 (0.49) | 0 | 0.285 (0.17) | 0.356 (0.18) | 0.2* |
| Veracruz | Tihuatlán | (15) Tih | 23 | 0.88 | 3.27 (1.0) | 1.71 (0.16) | 5 | 0.228 (0.16) | 0.315 (0.11) | 0.28* |
| Durango | Durango | (16) Dgo | 6 | 0.88 | 2.1 (0.35) | 1.36 (0.05) | 1 | 0.395 (0.26) | 0.464 (0.13) | 0.15* |
| Coahuila | Cuatrocienegas | (17) CCC | 22 | 0.88 | 2.7 (0.88) | NA | 3 | 0.315 (0.31) | 0.458 (0.17) | 0.31* |
| Sinaloa | Sinaloa | (18) SinalP | 6 | 0.66 | 2.1 (0.40) | 1.6 (0.76) | 0 | 0.388 (0.32) | 0.588 (0.10) | 0.39* |
| Sonora | Yecora | (19) Yec | 6 | 0.77 | 3 (1) | 1.4 (0.16) | 2 | 0.380 (0.27) | 0.547 (0.2) | 0.32* |
| *C. argyrosperma* | ssp. *argyrosperma* | | 327 | 0.775 (0.39) | 2.786 (0.76) | 1.6 (0.28) | 23 | 0.364 (0.25) | 0.410 (0.16) | 0.033 (0.06) |
| Chiapas | Sor_Chis | (20) SChis | 20 | 0.55 | 2.8 (0.83) | 1.7 (0.78) | 1 | 0.39 (0.28) | 0.425 (0.10) | 0.084 |
| Oaxaca | Mixtepec | (21) Soax | 21 | 0.55 | 3.4 (1.4) | 1.8 (0.34) | 3 | 0.352 (0.32) | 0.502 (0.14) | 0.303* |
| Guerrero | Ometepec | (22) Sgro | 22 | 0.77 | 2.8 (1.8) | 1.8 (0.27) | 0 | 0.233 (0.25) | 0.300 (0.21) | 0.225* |
| Jalisco | Autlán | (23) Sjal | 22 | 0.66 | 3 (1.2) | 1.6 (0.23) | 1 | 0.431 (0.28) | 0.447 (0.16) | 0.035 |
| Sinaloa | Culiacán | (24) SoSin | 13 | 0.77 | 3.2 (1.3) | 1.8 (0.32) | 2 | 0.527 (0.35) | 0.501 (0.21) | −0.053 |
| Sonora | Alamos | (25) SoSon | 15 | 0.55 | 2.4 (0.5) | 1.8 (0.66) | 1 | 0.400 (0.16) | 0.394 (0.13) | −0.013 |
| *C. argyrosperma* | ssp. *sororia* | | 113 | 0.641 (0.11) | 2.93 (1.17) | 1.8 (0.28) | 8 | 0.388 (0.27) | 0.428 (0.15) | 0.077 (0.16)* |

*The table shows the State and Population ID of collected samples, number of individuals (n), proportion of polymorphic loci (P), allelic richness (A), corrected allelic richness (RA), Number of private alleles (Pa), observed heterozygosity ($H_O$), expected heterozygosity ($H_E$), and inbreeding coefficient ($F_{IS}$), (*) significant at $p < 0.05$. Standard deviation (SD) is shown in parenthesis. NA could not be estimated by the program due to the presence of missing data.*

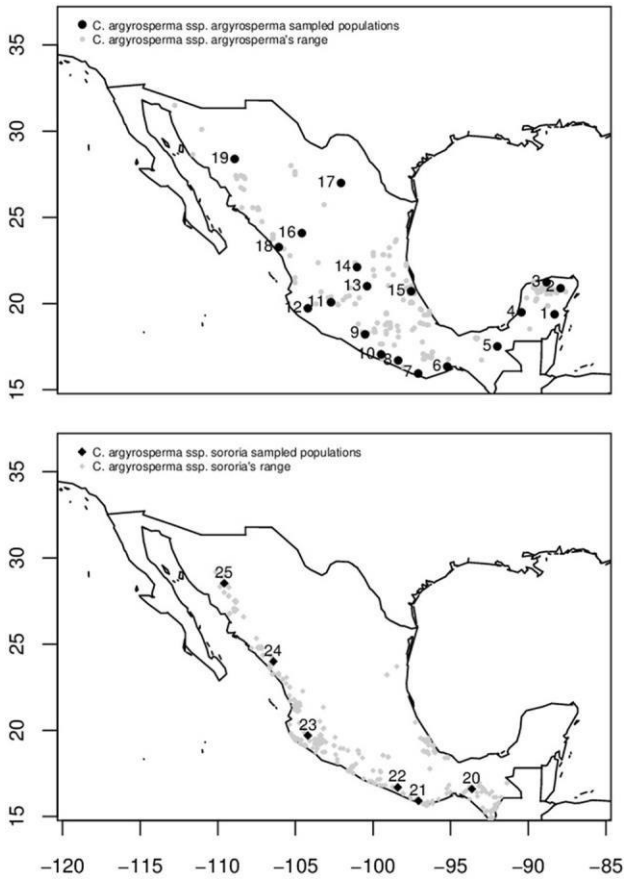



FIGURE 2 | Distribution area in Mexico (gray symbols) and sampled populations of *C. argyrosperma* ssp. *argyrosperma* (Top, black points) and *C. argyrosperma* ssp. *sororia* (Bottom, black diamonds). Population ID is shown in Table 1.

We estimated the genetic differentiation among populations through pairwise $F_{ST}$ using adegenet (Jombart, 2008; Jombart and Ahmed, 2011) for R v.1.4.2 (R Core Team, 2016). To depict the genetic relationships among populations, we used the pairwise $F_{ST}$ matrix to construct a dendrogram with the complete agglomeration method using the *hclust* function in the package ape (Paradis et al., 2004) for R. To determine the degree of statistical support for internal nodes we made an UPGMA dendrogram with R v.3.2.0, and evaluated 1000 trees constructed from bootstrap resampling of the loci with this same library.

To test for isolation by distance in each subspecies, we used ade4 (Dray and Dufour, 2007) for R to perform a Mantel test with 999 permutations. For this test, we first used the Geographic Distance Matrix Generator (Ersts, 2011) to transform sample coordinates into a geographic distance matrix. We also performed an AMOVA (Excoffier et al., 1992) testing different scenarios to determine whether subspecies, or genetic clusters provide a better explanation of the genetic variance in the species *C. argyrosperma*.

## Gene Flow

To obtain estimates of the migration rates among populations and subspecies, we used BayesAss v.3.0.4 (Wilson and Rannala, 2003). This program, based on Bayesian probability, detects immigrant ancestors up to two generations in the past, even if overall genetic differentiation is low. An advantage of this approximation is that it does not assume that populations have reached equilibrium (Rannala and Mountain, 1997), which may be the case for species that have undergone rapid demographic expansion, such as domesticated taxa. We performed several runs to determine the best number of MCMC, to tune the priors and to check for convergence. Accordingly, we performed 30,000,000 MCMC iterations, with a burn-in of 3,000,000 and a sampling frequency of 2,000. We set the parameters as follows: deltaA = 0.70 (mixing parameter for allele frequencies), deltaF = 0.90 (mixing parameter for inbreeding coefficient), deltam = 0.05 (mixing parameter for migration rates) to obtain an acceptance rate between 0.2 and 0.6, as suggested by Rannala (2007). We obtained a trace file to check for convergence with Tracer v.1.5[2] (Rambaut and Drummond, 2009).

To detect barriers to gene flow, we used the Monmonier algorithm (Monmonier, 1973; Manni et al., 2004) implemented in adegenet for R, considering both subspecies and for each subspecies separately. The Monmonier algorithm conducts a heuristic search used to define barriers based on dissimilarity scores. First, the genetic distance between contiguous populations is computed and the two populations with the highest level of differentiation are used to specify the starting boundary of the barrier. Then, the barrier is followed to both ends until either end reaches the edge or a barrier. These steps are repeated until the within-group sum of squares indicates that regional subdivision has progressed considerably (Monmonier, 1973). We used the *optimize.monmonier* function, which uses different starting points to find the solution that better explains genetic distances among populations based on the largest sum of local distances. We used values of pairwise $F_{ST}$ as the distance matrix to perform the analysis, and the number of starting points set to 10 for *argyrosperma* and to three for *sororia* (i.e., half of the number of populations).

## Species Distribution Models

To assess environmental suitability in possible areas of domestication we used species distribution models (SDMs) projections into the mid-Holocene (6,000 years ago) for the wild relative *sororia*. We constructed a database with geographic coordinates of collected and known *sororia* populations. Points from Central America were downloaded from GBIF[3], and 699 points from Mexico were obtained from Salvador Montes-Hernández, for a total of 720 occurrence. This database was purged to eliminate duplicated pixels. In addition, to ensure that all points fell within the species distribution, we estimated Mahalanobis distances using previously selected environmental variables (see below for selection methodology).

[2] http://beast.bio.ed.ac.uk

[3] www.gbif.org

The points deviating by two standard deviations or more from the mean were mapped, checked, and discarded if they fell outside the species range. We finally retained 273 occurrence points to perform the SDMs (Supplementary Figure S1).

To reduce the uncertainty associated with SDMs, it is necessary to select only the more informative and uncorrelated climatic variables (Hirzel et al., 2002). To do so, we downloaded the set of 19 bioclimatic variables taken from the worldwide temperature and rainfall data within the WorldClim 1.4 dataset (Hijmans et al., 2005). To determine which climatic variables to use, we analyzed the 273 occurrence points in a principal component analysis (PCA) and a Spearman correlation matrix. For the PCA, we considered as informative the components that, taken together, represented 87% of the variance associated with the data. For the Spearman correlation matrix, we defined an uncorrelated model by using a threshold of $r < 0.85$ (Booth et al., 1994). Nine bioclimatic layers were selected: Mean Temperature of Warmest Quarter, Mean Temperature of Coldest Quarter, Isothermality (BIO2/BIO7) (*100), Maximum Temperature of Warmest Month, Precipitation of Wettest Month, Precipitation of Driest Month, Precipitation Seasonality, Precipitation of Warmest Quarter, and Precipitation of Coldest Quarter.

We generated SDMs for current and past climate conditions with MaxEnt 3.3.3 (Phillips et al., 2004, 2006), using the 273 occurrence points and nine bioclimatic variables. We limited the analysis by cropping all climate layers to the distribution of *sororia* (9.51003°N to 34.18357°N and −116.1351°W to −70.80147° W). MaxEnt was executed using a 20% random test rate, 30 replicates, replicated bootstraps, 1000 maximum iterations and a convergence threshold of 0.00001, with extrapolation and clamping turned off. The distribution model was derived from the average model and evaluated using the score of the area under the curve (AUC; Elith et al., 2006).

For SDM projecting to mid-Holocene climate conditions, we downloaded the layers corresponding to atmospheric-ocean general circulation models (AOGCM) based on the Community Climate System Model CCSM4 (Collins et al., 2006), which incorporates dynamics of atmospheric processes, including radiation, convection, condensation and evaporation. This AOGCM has already been used in the reconstruction of past distributional models in the region (Waltari et al., 2007; Peterson and Nyári, 2008; Waltari and Guralnick, 2009; Holmgren et al., 2010; Gámez et al., 2014; Scheinvar et al., 2017). All environmental analyses were performed at a resolution of 30 arcsec (~1 km$^2$).

In order to create a presence/absence map, we used the 95th percentile value of observed sample points as a threshold for the logistic model. This value assumes that up to 5% of the records used for generating the model are subject to error. For current and mid-Holocene times, we generated presence/absence maps for *sororia*. To identify areas with suitable environmental conditions for the species under current and past climate conditions, we performed a sum of the SDMs, thus highlighting the areas with potential stability conditions from the mid-Holocene to the present.

# RESULTS

## Genetic Diversity

Three microsatellites (CMTp175, CMTm187, and CMTm144) showed evidence of null alleles and were therefore excluded from further analyses. As expected, there was no evidence of linkage disequilibrium among loci.

All loci showed significant deviations from Hardy-Weinberg equilibrium (HWE) in at least one population (Supplementary Table S3). Nevertheless, we performed multiple comparisons, which show that loci with significant deviations from HWE are different among populations.

We obtained a total of 84 alleles for the nine analyzed loci. At least one locus was monomorphic in each population (**Table 1**). We found higher levels of polymorphism per population ($p = 0.02$) in the cultivated populations of *argyrosperma* ($P = 0.775 \pm 0.39$ SD; **Table 1**) than in the wild *sororia* ($P = 0.641 \pm 0.11$ SD; **Table 1**). For *argyrosperma*, the proportion of polymorphic loci ranged between 0.44 and 0.88. For *sororia*, the proportion of polymorphic loci ranged between 0.55 and 0.77.

Forty alleles were private to *argyrosperma* and eleven were found only in *sororia*. Mean number of private alleles per population was 1.21 in *argyrosperma* and 1.33 in *sororia*. Within *argyrosperma*, populations Tih and Teh (full name of geographic locations are shown in **Table 1**) showed the highest number of private alleles. Within *sororia*, populations Soax and SoSin showed the highest number of private alleles. Overall, the populations from Oaxaca showed the highest proportion of private alleles in both subspecies. Allelic richness was similar ($p = 0.23$) in cultivated *argyrosperma* ($A = 2.786 \pm 0.76$ SD) and in wild *sororia* ($A = 2.93 \pm 1.17$ SD). For *argyrosperma*, rarefied allelic richness ranged from 1.9 in Aut to 1.36 in Dgo (**Table 1**). For *sororia*, rarefied allelic richness ranged from 1.8 in four populations to 1.6 in Aut (**Table 1**).

Mean observed and expected heterozygosity were similar among subspecies ($H_O = 0.388$ and $H_E = 0.428$ in *sororia*, and $H_O = 0.36$ and $H_E = 0.410$ in *argyrosperma*; $H_O$ $p = 0.484$; $H_E$ $p = 0.656$). For *argyrosperma*, genetic diversity ($H_E$) ranged from 0.588 in SinalP to 0.25 in Ek. For *sororia*, mean genetic diversity ($H_E$) was 0.428, with the highest value in Soax (0.502) and the lowest in Sgro (0.3) (**Table 1**).

Both subspecies showed similar overall inbreeding coefficients ($F_{IS} = 0.033 \pm 0.069$ SD, $p = 0.34$ in *argyrosperma* and $F_{IS} = 0.077 \pm 0.16$ SD, $p = 0.33$ in *sororia*, and were not statistically different $p = 0.656$). Within *argyrosperma*, five populations exhibited heterozygosity excess, while nine showed heterozygote deficiency and the rest were in HWE (**Table 1**). The highest values for heterozygosity deficiency were found in Chan and SinalP ($F_{IS} = 0.39$) and Yec ($F_{IS} = 0.32$) and for heterozygosity excess the lowest values were found in Ek and Ome ($F_{IS} = -0.253$ and $-0.26$, respectively; **Table 1**). Within *sororia*, two populations exhibited heterozygosity deficiency,

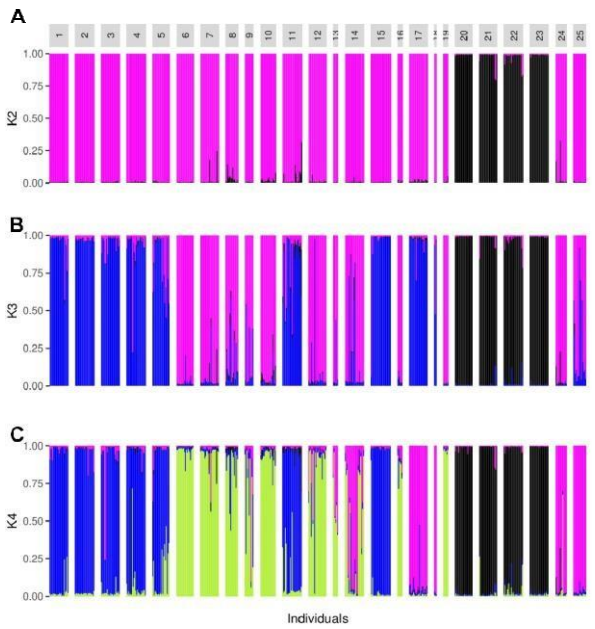



FIGURE 3 | Barplot from the Structure analysis. (A–C) $K$ = 2, $K$ = 3, and $K$ = 4. Population ID are as shown in Table 1.

while the rest were in HWE (**Table 1**). The highest $F_{IS}$ value was found in Soax (0.303) (**Table 1**).

## Population Structure

The analysis performed with Structure suggested a value of $K$ = 2, followed by $K$ = 4 (**Figure 3**). For $K$ = 2 there was a clear genetic differentiation between subspecies (**Figure 3A**), except for the *sororia* populations SoSin and SoSon, that were more similar to *argyrosperma*. The Structure barplot for $K$ = 3 shows that within *argyrosperma*, the populations from the Yucatan peninsula, Chiapas, Veracruz, Michoacán and CCC constituted a cluster, while the populations from northern and central Mexico formed another cluster (**Figure 3B**). Finally, for $K$ = 4 the clusters largely corresponded to subspecies' geographic distributions (**Figures 3**, **4**). The first cluster consisted of four *sororia* populations: Schis, Soax, Sgro, Sjal (black in **Figure 4**). The other two *sororia* populations (SoSin and SoSon) were assigned to a second cluster with *argyrosperma* populations CCC, and SinalP located in northern Mexico (pink in **Figure 4**). The third cluster consisted of *argyrosperma* populations from Ek, Mot, Chan, Champ and Pal, from the Yucatan Peninsula and populations from Michoacán (Sah) and Veracruz (Tih) (blue in **Figure 4**). The fourth cluster was constituted by populations Teh, Mix, Ome, Tla, Aut, and Yec, from the Pacific coast and Durango (green in **Figure 4**). The results from this analysis showed some degree of admixture among populations, particularly within *argyrosperma* (populations Tan, SJI and Mtp in **Figure 3C**), but the populations from *sororia* had low levels of admixture with the domesticated subspecies.

The results from the DAPC analysis were consistent with the results from Structure (**Figure 5**). Four *sororia* populations were clearly differentiated from *argyrosperma*, while two populations clustered within *argyrosperma*. All *argyrosperma* populations were grouped together, except for Tih from the state of Veracruz, which seemed in this analysis to be well differentiated from the other populations (**Figure 5**). In this DAPC 97.9% of the variance is explained by 40 PCs. A DAPC analysis considering only *argyrosperma* populations, except Tih (**Figure 6**), showed that Mtp and Tehuantepec from the states of Guerrero and Oaxaca, respectively, were well differentiated. Some populations formed very cohesive clusters, i.e., populations from the Yucatan Peninsula and populations from the Pacific Coast. In this second DAPC 94.6% of the variance is explained by 25 PCs.

Overall genetic structure was higher for wild *sororia* ($F_{ST}$ = 0.492, $R_{ST}$ = 0.610) than for domesticated *argyrosperma* ($F_{ST}$ = 0.264, $R_{ST}$ = 0.4). Genetic differentiation among populations (pairwise $F_{ST}$) of *argyrosperma* was moderate to high ($F_{ST}$ = 0.031–0.515), while genetic differentiation among populations of *sororia* was higher in general ($F_{ST}$ = 0.171–0.639). Genetic differentiation among populations of the different subspecies was medium to high ($F_{ST}$ = 0.152–0.652; **Figure 7** and Supplementary Table S5). When we estimated genetic differentiation among populations of wild *sororia* without escaped populations (SoSin and SoSon) values were moderate ($F_{ST}$ = 0.181–0.352).

The dendrogram built using pairwise $F_{ST}$ values (**Figure 7**) showed two well-defined groups: one group including only *sororia* populations from different states along the Pacific coast in Mexico (Oaxaca, Guerrero, Chiapas and Jalisco), and another group of *argyrosperma* populations, including the two *sororia* populations (SoSin and SoSon) mentioned above. Bootstrap values are in general low, as is usually found in intraspecific studies (due to both gene flow and recent common ancestry). Low bootstrap values within *argyrosperma* could also be due to homoplasy and care should be taken with their interpretation. Nevertheless, the two groups had higher support values (above 50%) (**Figure 7**).

An AMOVA that considered each subspecies, explained 20.05% of the genetic variance between subspecies, and most of the variance was found within individuals (50.3%), followed by among populations within subspecies (26.2%), and only 3% of the variance was found among individuals within populations. Given that two putatively wild populations (SoSin and SoSon) were identified as belonging to *argyrosperma* in all genetic structure analyses, and the seeds show an intermediate morphology for size and color (**Figure 1**), we also performed an AMOVA analysis considering these populations as *argyrosperma*. This analysis showed that a higher percentage (30.3%) of the genetic variance was allocated between subspecies; most variance was still found within individuals (46.1%), and less among populations within subspecies (21.4%), finally, 2.3% of the genetic variance is found among individuals within populations. On the other hand, an AMOVA analysis considering the partition suggested by the Structure analysis, $K$ = 4, explained 24.5% of the genetic variance among clusters, the variance among populations within clusters was 30.8%, and most of the variance was found within populations (44.6%).

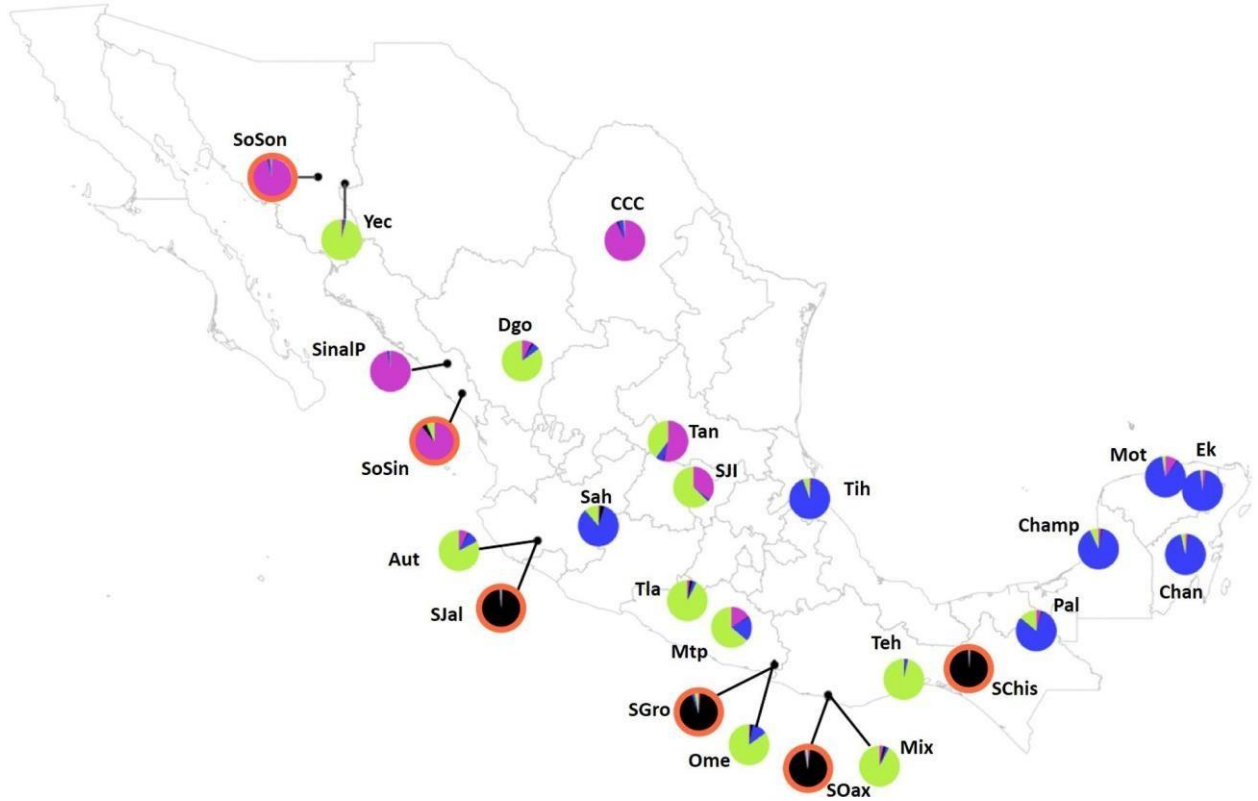



FIGURE 4 | Population map depicting pie graphs corresponding to the proportion of individuals per population assigned to each genetic cluster obtained with STRUCTURE for $K = 4$. Populations with graph pie outlined in orange highlight *sororia* populations. The black cluster consists of four *sororia* populations: Schis, Soax, Sgro, Sjal. The other two *sororia* populations (SoSin and SoSon) were assigned to a second cluster with *argyrosperma* populations CCC, and SinalP located in Northern Mexico (pink). The third cluster (blue) consists of populations Ek, Mot, Chan, Champ and Pal, from the Yucatan Peninsula and populations from Michoacán (Sah) and Veracruz (Tih). The fourth cluster (green) is constituted by populations Teh, Mix, Ome, Tla, Aut, and Yec, from the Pacific coast and Durango (Dgo). Population ID are as shown in Table 1.

Mantel tests were significant for both subspecies, indicating spatial structure due to isolation by distance. We found that geographically closer populations are genetically more similar than expected by chance (**Figure 8**).

## Estimates of Gene Flow

Estimates of recent gene flow suggest that the total proportion of migrants for each population was from 17 to 33% (Supplementary Table S4). Nevertheless, in general the proportion of migrants among pairs of populations was low $\leq 0.01$; the exceptions were between some populations in the Yucatan Peninsula and in Chiapas, the Pacific coast, and in the northern portion of the Pacific coast, for *argyrosperma*; and in the southern-central portion of the Pacific coast for *sororia* (Supplementary Table S4). The only case where gene flow between cultivated *argyrosperma* and wild *sororia* populations was detected involved the two Northern *sororia* populations (SoSin and SoSon) and a Northern population in San Luis Potosi state, Tan; other analyses strongly suggest that SoSin and SoSon are *argyrosperma* populations escaped from cultivation. This suggests that gene flow between cultivated and truly wild populations is low.

The Monmonier analysis indicated that for *argyrosperma* (**Figures 9A,D**), the northern part of the Sierra Madre Occidental may function as an effective barrier to gene flow, isolating the SinalP and Yec populations. This contrasts with results from the DAPC and structure analyses, where these populations do not seem to be isolated. This can be due to differences in the methodologies, in which Monmonier analysis takes spatial distances into account. For *sororia*, the southern portion of the Sierra Madre Occidental also functions as an effective barrier to gene flow, and isolates population Sjal (**Figure 9B**). When both subspecies were analyzed together, we observed that the main barrier is located in the region of the Isthmus of Tehuantepec, isolating the populations from the Yucatan Peninsula (**Figure 9C**).

## Species Distribution Models

The SDM for the wild subspecies, *sororia* (**Figure 10**), showed stability and good support (AUC: 0.96). The SDM projection to the mid-Holocene ($\sim$6000 years ago) suggests that the distribution area of *sororia* has been more or less stable since domestication (**Figure 10**). Nevertheless, the analysis also suggests that *sororia* may have been present in the Yucatan

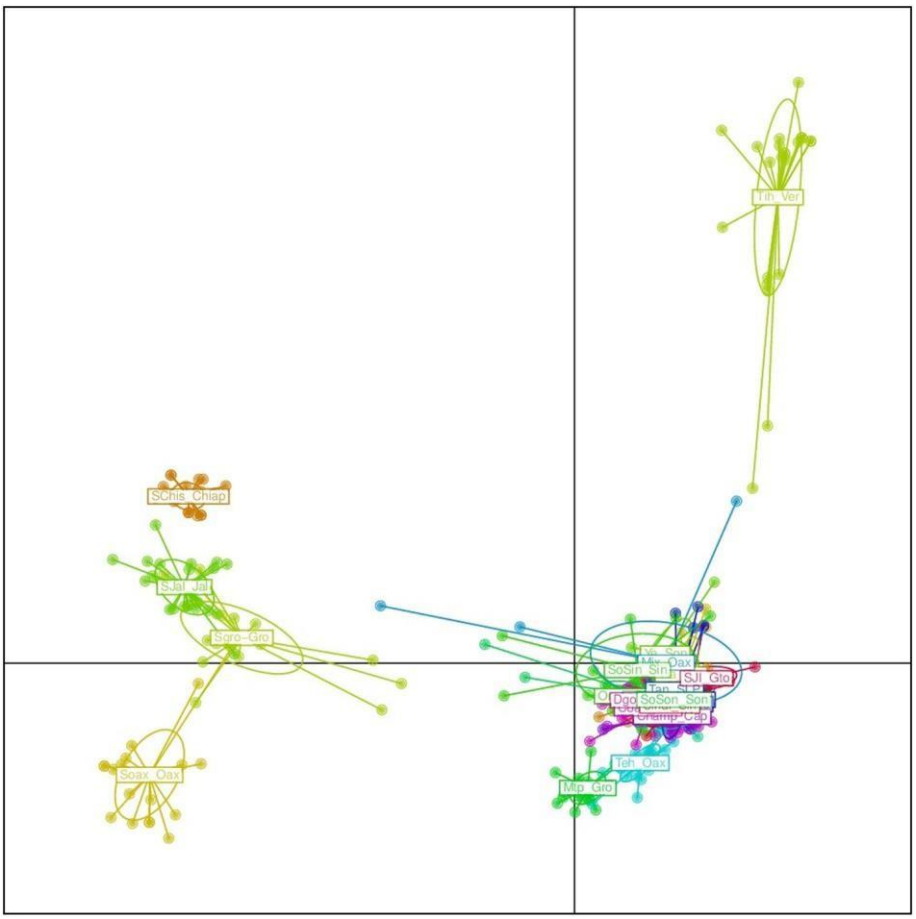



FIGURE 5 | Discriminant analysis of principal components (DAPC) for 19 *argyrosperma* and 6 *sororia* populations. Names starting with S or So correspond to *sororia* populations. Four *sororia* populations are well differentiated from the *argyrosperma* populations.

Peninsula during the mid-Holocene and its distribution in the regions of Oaxaca and Guerrero, where the most ancient archeological remains have been found, may have been more continuous than today (**Figure 10**). In addition, the distribution of *sororia* in Central America may have been wider and more continuous from Guatemala and Honduras to the northern area of Nicaragua.

## DISCUSSION

The present study represents the first wide range analysis of the genetic variation, genetic structure and gene flow of *C. argyrosperma*, covering the cultivated *argyrosperma* distribution in Mexico, and including populations of the wild *sororia* distribution in the Pacific Coast from Northern Mexico (Sonora) to Southern Mexico (Chiapas). Our analyses show similar levels of genetic variation in the cultivated populations and in its wild ancestors. Genetic differentiation is higher in wild *sororia* ($F_{ST}$ = 0.492) than in domesticated *argyrosperma* ($F_{ST}$ = 0.264), but this estimate is probably the product of including two escaped populations (SoSin and SoSon) that were misclassified and analyzed as *sororia*. When we remove these populations, differentiation in wild *sororia* ($F_{ST}$ = 0.243) became even lower than the differentiation found in cultivated *argyrosperma.* Gene flow at a regional level is associated to movement of pollen by *Cucurbita* pollinators and to human cultural practices, such as seed exchange among populations (Montes-Hernández et al., 2005; Organization for Economic Co-operation and Development [OECD], 2012). Some patterns of gene flow detected in *argyrosperma* may be the result of these seed exchanges, but these hypotheses should be tested with ethnobotanical data in future analyses.

### Genetic Variation and Inbreeding

Priori et al. (2013) reported an 85% transferability for microsatellites designed for *C. pepo* to cultivated *C. argyrosperma.* Accordingly, nine of twelve microsatellite loci used in this study were adequate for *C. argyrosperma*, while we discarded the additional three because of a high number of null alleles.

Cultivated species often show low levels of genetic variation (Gaut et al., 2015). Surprisingly, both subspecies showed similar levels of genetic diversity (**Table 1**). Comparable values of

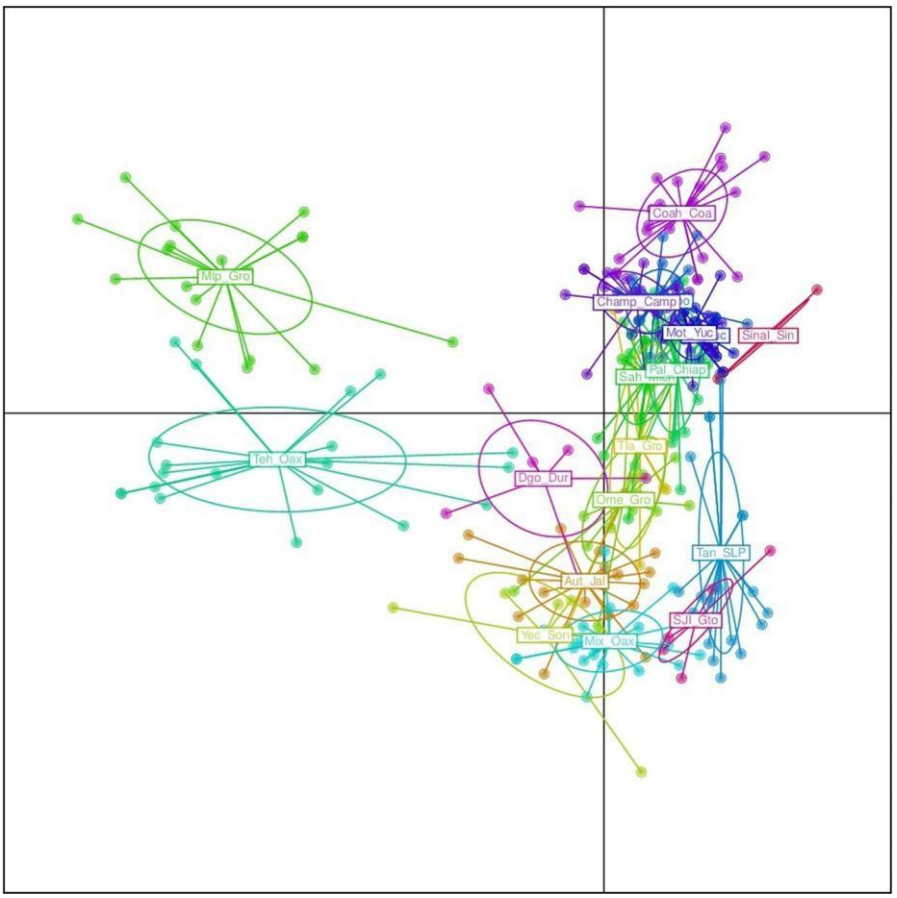



FIGURE 6 | Discriminant analysis of principal components (DAPC) for only 18 *argyrosperma* populations (excluding Tih). Mtp and Teh from the states of Guerrero and Oaxaca, respectively, are well differentiated. Population ID are as shown in Table 1.

polymorphic loci, allelic richness and genetic diversity among *argyrosperma* and *sororia* suggests that the subspecies have had similar effective population sizes, and that the theoretical bottleneck associated with domestication was either mild and/or of short duration, followed by a rapid population expansion (Hedrick, 2011).

Genetic variation was similar to what has been reported for other annual plants in microsatellite studies ($H_E$ = 0.46; Nybom, 2004), but lower when compared to other outcrossing species ($H_E$ = 0.65; Nybom, 2004). Also, the mean allele number in *C. argyrosperma* was lower than those reported for *C. pepo* using nuclear microsatellite loci ($A$ = 3.2–5.6; Formisano et al., 2012; Gong et al., 2012, 2013; Priori et al., 2013; Ntuli et al., 2015) and lower than those reported by Balvino-Olvera et al. (2017) in wild *sororia* populations along the West coast of Mexico ($A$ = 12.3). Low levels of allelic richness and genetic diversity in *sororia* may suggest that this species has undergone one or several bottlenecks due to ecological shifts during the Pleistocene, followed by rapid population expansion, as suggested by Kistler et al. (2016). Nevertheless, these comparisons should be taken with caution because analyses were performed with different sets of microsatellite loci, and these hypotheses should be investigated in future studies.

Certain aspects of agricultural management, such as seed exchange, may also affect the levels of genetic variation in *C. argyrosperma* (Montes-Hernández et al., 2005) and in *C. pepo* (Enríquez Cotton, 2017; Enríquez et al., 2017). In particular, the *milpa* system, which predominates in the central and southern portions of Mexico (Lira et al., 2016a), is a form of polyculture (i.e., growing several *Cucurbita* species in the same area) and seed exchange, that can reduce inbreeding at the local level. Nevertheless, it is advisable to perform similar analyses in other wild and domesticated cucurbits to gain further insight into the amount of genetic variation present in *Cucurbita*.

For *argyrosperma*, populations located in the extremes of its distribution (SinalP in Sinaloa and Chan in Quintana Roo) showed the highest levels of genetic variation, while the populations from the Yucatan Peninsula (except Chan) showed the lowest levels of genetic variation. These results, together with the barriers analysis, suggest that the cultivated populations from the Yucatan Peninsula are isolated genetically (Zizumbo-Villarreal and Colunga-GarcíaMarín, 2010; Moreno-Estrada et al., 2014). Moreover, the wild subspecies, *sororia*, is not

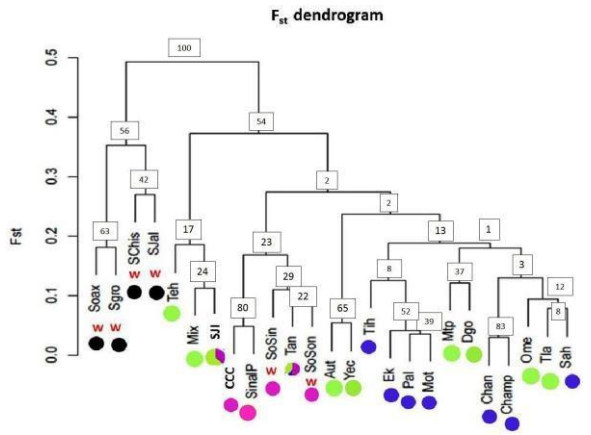



FIGURE 7 | Pairwise $F_{ST}$ dendrogram for 19 *argyrosperma* populations and 6 *sororia* populations. Populations of the wild subspecies (*sororia*) are indicated (w). Dot colors agree with the pie graphs corresponding to the proportion of individuals per population assigned to each genetic cluster obtained with STRUCTURE for $K$ = 4 shown in Figure 6. Population ID is shown in Table 1. Node values represent bootstrap support.

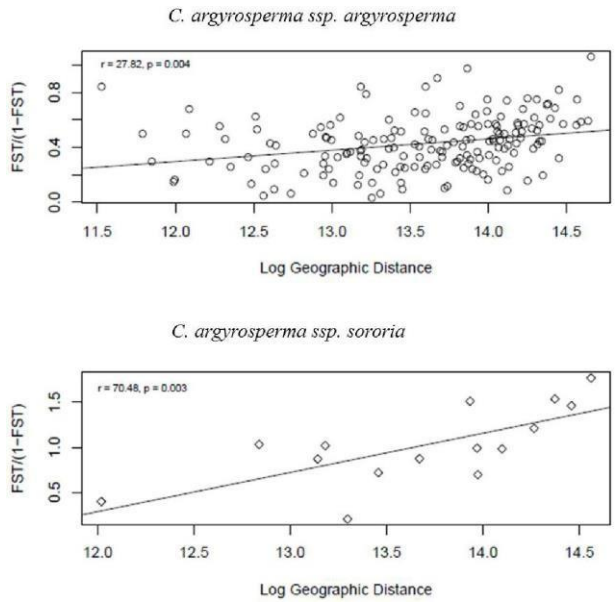



FIGURE 8 | Test of isolation by distance (IBD) obtained through a Mantel test. Results were significant, showing a strong positive correlation between geographic and genetic distances for both *argyrosperma* ($r$ = 27.8, $p$ = 0.004), and *sororia* ($r$ = 70.04, $p$ = 0.003).

distributed in the Yucatan Peninsula, thus affecting the potential gene flow among subspecies in this area. In subspecies *sororia* we did not find a geographic pattern for the distribution of its genetic diversity, and Oaxaca was the population that showed the highest genetic variation.

Cultivated species often show high levels of inbreeding (Gaut et al., 2015). Estimates of inbreeding coefficients ($F_{IS}$) in *argyrosperma* were highly variable (**Table 1**), with some populations showing heterozygote deficiency (9 populations), as may be expected in a domesticated species, and other populations showing heterozygote excess (5 populations), as previously reported by Montes-Hernández and Eguiarte (2002), may be related to the type of agriculture and management (Cerón et al., 2010), as well as pollinator availability and home range. Heterozygote deficiency could be the result of the short flight capacity of the bees that pollinate these species (Montes, 2002), or to the fact that in traditional subsistence agriculture only a few fruits are selected to plant the next generation (thus, within a field all individuals are highly related; Montes, 2002), while in northern populations the use of improved inbred genetic lines (Servicio de Información Agroalimentaria y Pesquera [SIAP], 2016) could be the cause of heterozygosity deficiency. Negative $F_{IS}$ values found in some cases suggests that seed exchange is frequent at a local level (i.e., among neighbor populations) promoting outbreeding, but at a regional level (i.e., among extremes of the distribution) gene flow is low. It will be important to conduct detailed ethnobotanic studies in different regions of the country, along with genetic analysis, to test the effect of agricultural management on the genetic variation of this crop (Montes-Hernández et al., 2005; Bellon et al., 2009).

## Genetic Structure

When populations are isolated, genetic drift promotes the random fixation of alleles, thus the number of private alleles among populations can be used as a reference for population connectivity (Hedrick, 2011). We found a high number of private alleles among subspecies (40 in *argyrosperma* and 11 in *sororia*), while the mean number of private alleles per population was similar within subspecies (1.31 in *sororia* and 1.21 in *argyrosperma*).

The number of private alleles found in each subspecies suggests that overall levels of gene flow among subspecies have been low, thus promoting their divergence since the domestication of *argyrosperma*∼8,600 years ago (Rannere et al., 2009). A coalescent based approach such as those implemented in Approximate Bayesian Computation (ABC) analyses, together with a genome-wide approach (thousands of SNPs) will be conducted in the future to test whether these patterns relate to incomplete lineage sorting, ancestral introgression or current introgression.

In *argyrosperma*, we found a high number of private alleles in the populations from the states of Veracruz (Tih) and Oaxaca (Teh). Tih is geographically distant from other sampled populations, and its private alleles may be present in other populations from the Gulf of Mexico; thus, it is advisable to include more populations from this area in further analyses. The population Teh from Oaxaca is located in the area of the Isthmus of Tehuantepec that has been previously identified as an important barrier for the Mexican biota (Ornelas et al., 2013). Moreover, for *sororia*, the population with highest number of private alleles is located in the same area (Soax), further supporting the Isthmus of Tehuantepec as an important biogeographical barrier. Furthermore, seed morphology is distinctive in the populations of *argyrosperma* of Southeastern Mexico, where seeds show clear gray margins

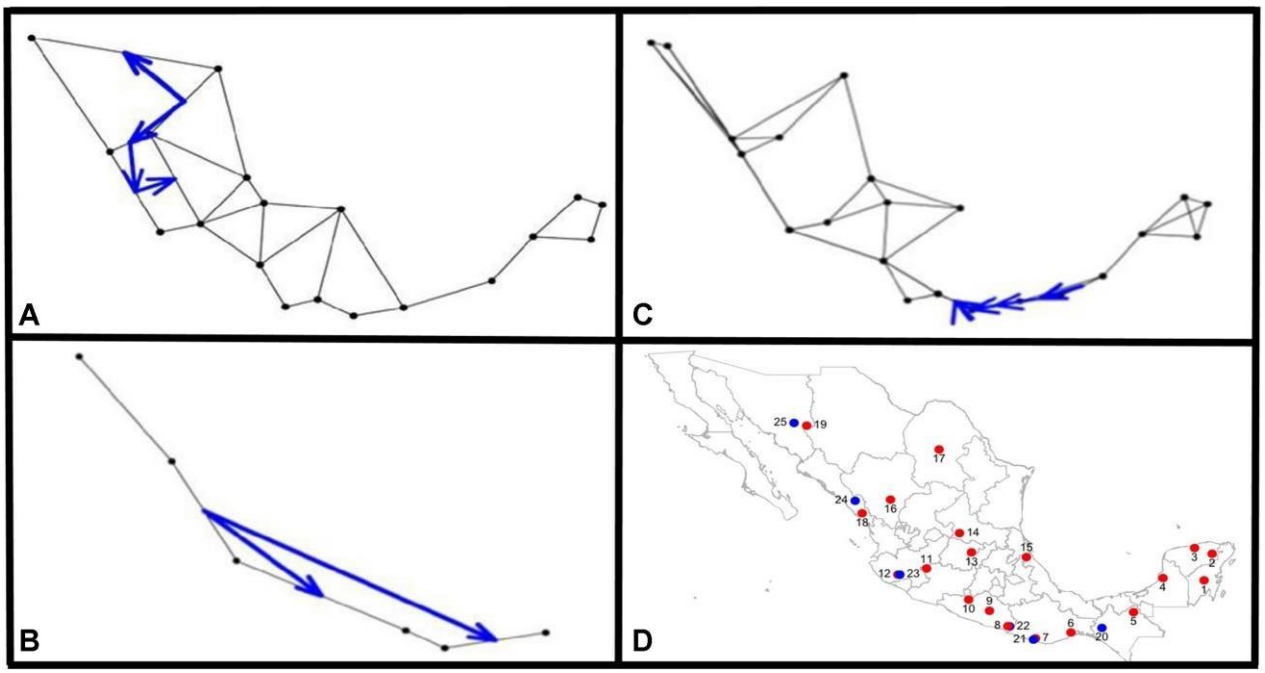



FIGURE 9 | Results from the Monmonier analysis performed with adegenet Black dots represent sampled populations and blue arrows depict significant barriers to gene flow, for (A) *C. argyrosperma* ssp. *argyrosperma*, (B) *C. argyrosperma* ssp. *sororia*, (C) both subspecies. (D) Location of populations on a map, blue dots corresponding to *C. argyrosperma* ssp. *sororia* and red dots are *C. argyrosperma ssp. argyrosperma*. Population ID is shown in Table 1.

in contrast to the golden color found in northern populations (**Figure 1**). People in southeastern Mexico have a strong preference for local varieties (G. Sánchez de la Vega, personal observation). In addition, this may suggest that the high number of private alleles in this area may be related to strong selection pressures associated with seed morphology, as has been reported in *C. pepo* commercial varieties (Formisano et al., 2012). Selection for morphological characters promotes selective sweeps that results in allele fixation in neutral sites of the genome (Meyer and Purugganan, 2013). Alternatively, a high number of private alleles could relate to isolation of these populations. Therefore, we need to conduct genomic and morphologically detailed analyses to test these hypotheses.

The results from the Structure and DAPC analyses show clear genetic differentiation among subspecies (**Figures 3**, **5**). Within *argyrosperma*, Structure analyses (**Figure 3**) show geographically associated groups: (1) a northern group; (2) Yucatan Peninsula; and (3) Pacific coast (**Figure 4**). It is interesting that these genetic groups roughly correspond to the genetic groups reported by Moreno-Estrada et al. (2014) for human Native American populations, which suggest that cultural aspects may be important in determining the genetic structure of this crop. Further analyses should test for the correlations between genetic clusters in domesticated taxa and human Native American populations.

Values of genetic differentiation ($F_{ST}$) were variable among populations of both subspecies. *Sororia* showed similar levels of genetic differentiation ($F_{ST}$ = 0.243, $R_{ST}$ = 0.3) as *argyrosperma* ($F_{ST}$ = 0.264, $R_{ST}$ = 0.4). In addition, our Mantel test results show that geographically close populations of both subspecies are genetically more similar than expected by chance. Both subspecies have wide distributions (**Figure 2**) that cover a distance of over 1,000 km, thus promoting genetic differentiation among extreme populations. Genetic differentiation in *sororia* could also be related to its patchy distribution and the limited movement (~0.7 km) and local low densities of its main pollinators from the genera *Peponapis* and *Xenoglossa* (Kohn and Casper, 1992; Montes, 2002; Enríquez et al., 2015). Levels of genetic differentiation among *argyrosperma* populations are similar to those reported for other outcrossing plants ($F_{ST}$ = 0.22; Nybom, 2004). It is advisable to include more *sororia* populations in future analyses to determine fine patterns of genetic differentiation.

$F_{ST}$ pairwise values are directly related to the degree of phenotypical resemblance among populations and provide insights into their demographic history (Holsinger and Weir, 2009). Our pairwise $F_{ST}$ analyses, along all our results of genetic differentiation suggest that both subspecies are genetically well-differentiated.

Also, it is worth noticing that all the analyses of genetic differentiation consistently suggest that the *sororia* populations SoSon and SoSin from the states of Sonora and Sinaloa are more like subspecies *argyrosperma* than *sororia*, including some morphological characteristics of the seeds (**Figure 1**). Our results suggest that these may be escaped populations of *argyrosperma*, as suggested by Merrick and Bates (1989) and Villanueva (2007), who mentioned that individuals from *argyrosperma* are capable of surviving without agricultural management, but that these

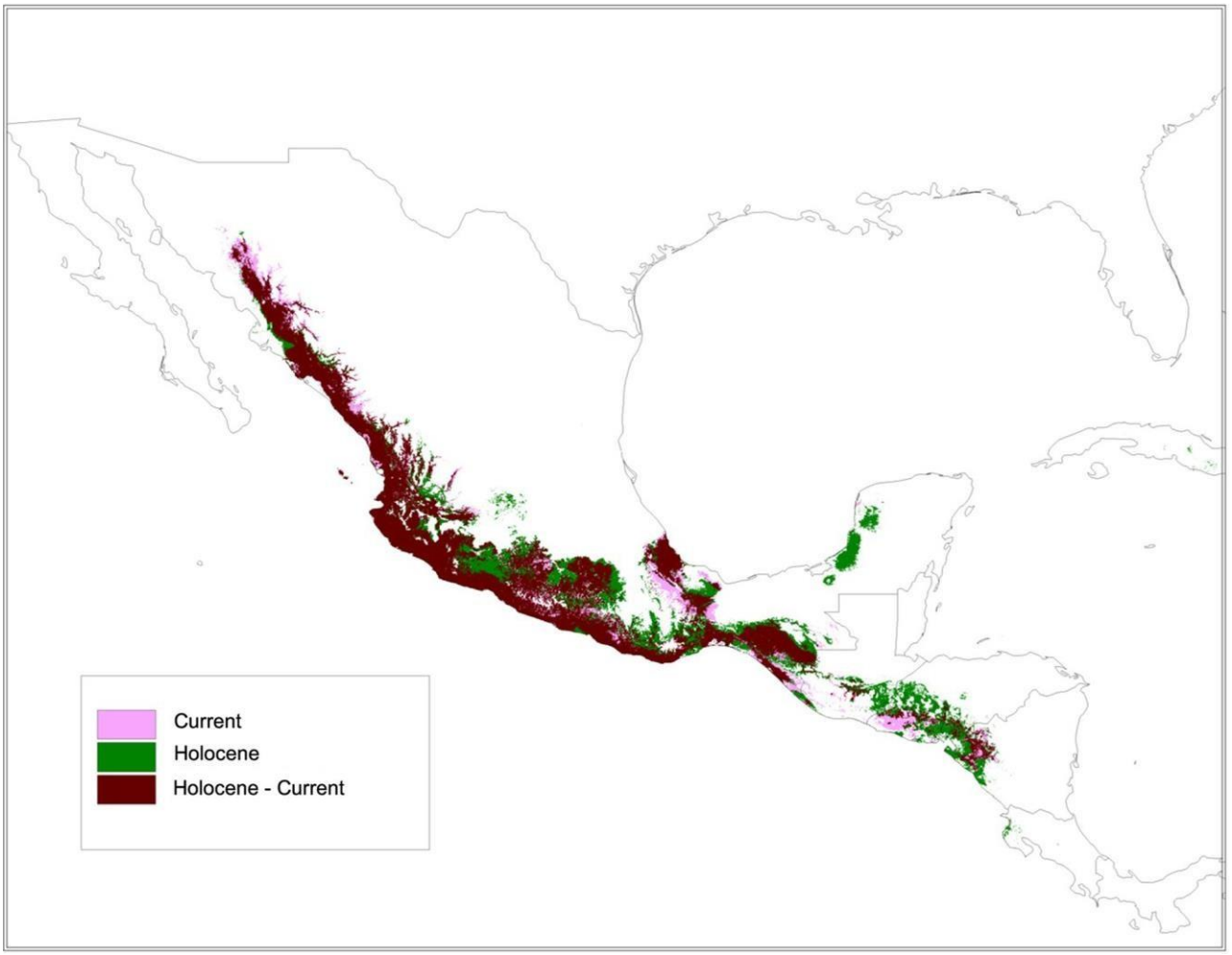



FIGURE 10 | The species distribution model (SDM) for ssp. *sororia* showed stability and good support (AUC: 0.96). Persistence during the mid-Holocene (~6000 years ago) and present time in light persistence only in the Holocene in purple; and persistence only in the present in green.

individuals show a reduction in seed size. Reports suggest that cultivars of *C. pepo* and *C. moschata* in Tamaulipas, Mexico are also capable of surviving and producing fruits in semi-wild or in extreme environmental conditions (Hanselka, 2010).

Gene flow is frequent among subspecies of *Cucurbita* and with other taxa at local levels (Montes-Hernández and Eguiarte, 2002; Lira et al., 2016b). Our analysis suggests that gene flow is less frequent at the regional level, with a few exceptions: (1) in the Yucatan Peninsula, and Chiapas, people mentioned that they usually exchange seeds among neighbors and family members and sell seeds for cultivation, which is consistent with estimated levels of gene flow in this area, and (2) the northern portion of the Pacific Coast, that apparently acts as a genetic corridor that has been previously reported for other crops (Zizumbo-Villarreal and Colunga-GarcíaMarín, 2010).

The results from the barrier analysis are consistent with this idea, where the northern portion of the Sierra Madre Occidental isolates the populations located along the Pacific coast. A pattern of isolation of populations located in Jalisco has also been reported for *Zea mays* ssp. *parviglumis*, the wild relative of maize (Aguirre-Liguori et al., 2017). Finally, when both subspecies were included in the analysis, the Isthmus of Tehuantepec appears as an important barrier, as has been previously reported for many wild taxa (Ornelas et al., 2013).

## Species Distribution Models

Species distribution models of wild relatives of domesticated taxa are a useful tool to corroborate hypotheses of possible domestication sites and environmental suitability for the presence of the wild species (Hufford et al., 2012; Besnard et al., 2013). The SDM for *sororia* suggests that its range has been more or less stable since the mid-Holocene, with possible presence in the Yucatan Peninsula and a more continuous range in Oaxaca and Guerrero during the mid-Holocene (~6,000 years ago). Many domestication events occurred during this time because of environmental changes and vegetation transitions associated with the end of the Last Glacial Maximum-Holocene, and with the impact of anthropogenic activities (Flannery,

1986; Piperno et al., 2007; Aguirre-Liguori et al., 2016). For the region of Guerrero, Piperno et al. (2007) and Rannere et al. (2009) proposed that the end of the Pleistocene was cold, and as the Holocene advanced this area became warmer, promoting a transition from temperate arboreal elements to tropical forests, environment conditions associated to principal events of domestication in the Balsas basin.

Previous genetic, molecular, biogeographic, and archeological analyses suggest that *argyrosperma* was domesticated in the Balsas-Jalisco region, approximately 9,000 years ago (Sanjur et al., 2002; Piperno et al., 2009; Rannere et al., 2009; Lira et al., 2016b). The oldest archeological remains are from caves located in the Balsas region (Guerrero) from 6,100 to 8,500 years ago. Initial domestication (before 7,000 years ago) was followed by early diversification (Lira et al., 2016b). Our SDM results indicate that environmental characteristics were suitable for the presence of *sororia* in this area during this period.

## CONCLUSION

Our analyses describe broad patterns of genetic variation, genetic differentiation and gene flow among domesticated and wild *C. argyrosperma*. The levels of genetic variation and genetic differentiation were similar for *sororia* and *argyrosperma*. These could relate to their demographic histories, but further analyses should be conducted to test different demographic hypotheses. Isolation by distance and gene flow analyses suggest that gene flow is more common at a local scale than at a regional scale, perhaps because of pollen movement by specialized pollinators and to human cultural practices, such as seed exchange among populations, but these hypotheses should be tested with ethnobotanical data in future analyses. *Sororia*'s distribution has been relatively stable since the mid-Holocene and suggests the presence of this subspecies in previously described domestication centers based on archeological records. Future analyses should gather information about agricultural management, morphological variation and the behavior of pollinators, along with a wider sampling of the wild populations and the use of massive sequencing data to expand our knowledge of squash domestication.

## AUTHOR CONTRIBUTIONS

GS-dlV and GC-M contributed to fieldwork, lab work, molecular and population genetics analysis, drafting the manuscript, and final approval of the version to be published. NG contributed to analysis of species distribution models (SDM), drafting the manuscript, and final approval of the version to be published. HH-R contributed to fieldwork, lab work, molecular and data analyses, and final approval of the version to be published. AV-L contributed to laboratory work, logistics, correcting the manuscript, and final approval of the version to be published. EA-P contributed to laboratory work, logistics and molecular analysis, correcting the manuscript, and final approval of the version to be published. JJ-C contributed to correcting the manuscript and final approval of the version to be published. SM-H project leader, contributed to fieldwork, germplasm collection, generated database for project design, and final approval of the version to be published. RL-S project leader, contributed to logistics and final approval of the version to be published. LE project leader, designed and coordinated the project, logistics, drafted and corrected the manuscript, and final approval of the version to be published.

## FUNDING

This work was funded by CONABIO KE 004, Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de *Cucurbita*, by CONACYT Investigación Científica Básica 2011.167826 (clave de identificación oficial CB2011/167826), Genómica de poblaciones: estudios en el maíz silvestre, el teosinte (*Zea mays* ssp. *parviglumis* y *Zea mays* ssp. *mexicana*) and by CONACYT Problemas Nacionales through the grant number 247730. GS-dlV was supported by grant number 292164 of the Consejo Nacional de Ciencia y Tecnología (CONACYT). The sabbatical leave of LEE at the Department of Plant and Microbial Biology, University of Minnesota, was supported by the program PASPA-DGAPA, UNAM.

## ACKNOWLEDGMENTS

This manuscript is presented in partial fulfillment for the requirements to obtain a Ph.D. degree by GS-dlV in the Posgrado en Ciencias Biológicas, Universidad Nacional Autonóma de México. We acknowledge the Posgrado en Ciencias Biológicas for the support provided during the development of this project. Special thanks to Dr. Valeria Souza and Dr. Daniel Piñero for supporting this research. We thank the Laboratorio de Evolución Molecular y Experimental, the Instituto de Ecología, and Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, and Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Campo Experimental Bajío. We also thank for help in laboratory, computer analyses, and fieldwork Enrique Scheinvar, Laura Espinosa Asuar, Gabriel Manuel Rosas, Leslie Mariel Paredes, Paulina Hernández, Josué Barrera, Dulce C. Hernández, Silvia Barrientos, Karen Y. Ruíz Mondragón, Jonás A. Aguirre-Liguori, Talitha E. Legaspi, and the technicians at Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Campo Experimental Bajío José Manuel Escutia Ponce and Miguel Ángel Mora Martínez. This paper was written during a sabbatical leave of LEE in the Department of Plant and Microbial Biology, University of Minnesota in Dr. Peter Tiffin's laboratory.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00400/full#supplementary-material

## REFERENCES


Aguirre-Dugua, X., and González-Rodríguez, A. (2016). "Phylogeographical approaches to the study of plant domestication, with special emphasis on perennial plants," in *Ethnobotany of Mexico: Interactions of People and Plants in Mesoamerica*, eds R. Lira, A. Casas, and J. Blancas (New York, NY: Springer International Publishing), 319–366. doi: 10.1007/978-1-4614-6669-7_13

Aguirre-Liguori, J., Tenaillon, M., Vázquez-Lobo, A., Gaut, B., Jaramillo-Correa, J. P., Montes-Hernandez, S., et al. (2017). Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Mol. Ecol.* 26, 4226–4240. doi: 10.1111/mec.14203

Aguirre-Liguori, J. A., Aguirre-Planter, E., and Eguiarte, L. E. (2016). "Genetics and ecology of wild and cultivated maize: domestication and introgression," in *Ethnobotany of Mexico: Interactions of People and Plants in Mesoamerica*, eds R. Lira, A. Casas, and J. Blancas (New York, NY: Springer International Publishing), 403–416. doi: 10.1007/978-1-4614-6669-7_16

Balvino-Olvera, F. J., Sánchez-Gómez, K. F., Lobo, J. A., Avila-Sakar, G., Cruz-Reyes, R., Sánchez-Montoya, G., et al. (2017). Latitudinal structured populations of the Mexican wild squash *Cucurbita argyrosperma* subsp. *sororia* revealed by microsatellite markers. *Crop Pasture Sci.* 68, 850–858. doi: 10.1071/CP17341

Bellon, M. R., Barrientos-Priego, A. F., Colunga-García Marín, P., Perales, H., Reyes Agüero, J. A., Rosales-Serna, R., et al. (2009). Diversidad y conservación de recursos genéticos en plantas cultivadas. *Cap. Nat. México* 2, 355–382.

Besnard, G., Khandari, B., Navascués, M., Fernánez-Mazuecos, M., El Bakkali, A., Arrigo, N., et al. (2013). The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proc. R. Soc. B* 280:201228333. doi: 10.1098/rspb.2012.2833

Booth, G. D., Niccolucci, M. J., and Schuster, E. G. (1994). "Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation," in *Proceedings of the Research Paper INT-470, United States Department of Agriculture Forest Service*, Ogden, UT.

Cerón, G. L., Legaría, S. J., Villanueva, V. C., and Sahagún Castellanos, J. (2010). Diversidad genética en cuatro especies mexicanas de calabaza (*Cucurbita* spp.). *Rev. Fitotec. Mex.* 33, 189–196.

Chakraborty, R., Andrade, M. D., Daiger, S. P., and Budowle, B. (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Ann. Hum. Genet.* 56, 45–47. doi: 10.1111/j.1469-1809.1992.tb01128.x

Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., et al. (2006). The community climate system model version 3 (CCSM3). *J. Climatol.* 19, 2122–2143. doi: 10.1175/JCLI3761.1

Cruz-Reyes, R., Ávila-Sakar, G., Sánchez-Montoya, G., and Quesada, M. (2015). Experimental assessment of gene flow between transgenic squash and a wild relative in the center of origin of Cucurbits. *Ecosphere* 6, 1–13. doi: 10.1890/ES15-00304.1

Decker-Walters, D. S., Walters, T. W., Poluszny, U., and Kevan, P. G. (1990). Genealogy and gene flow among annual domesticated species of *Cucurbita*. *Can. J. Bot.* 68, 782–789. doi: 10.1139/b90-104

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Dray, S., and Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1–20. doi: 10.18637/jss.v022.i04

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv.Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Eguiarte, L. E., Aguirre-Liguori, J. A., Jardón-Barbolla, L., Aguirre-Planter, E., and Souza, V. (2013). Genómica de poblaciones: nada en evolución va a tener sentido si no es a la luz de la genómica, y nada en genómica tendrá sentido si no es a la luz de la evolución. *TIP* 16, 42–56. doi: 10.1016/S1405-888X(13)72077-1

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151. doi: 10.1111/j.2006.0906-7590.04596.x

Enríquez, E., Ayala, R., Gonzalez, V. H., and Núnez-Farfán, J. (2015). Alpha and beta diversity of bees and their pollination role on *Cucurbita pepo* L. (Cucurbitaceae) in the Guatemalan cloud forest. *Pan-Pac. Entomol.* 91, 211–222. doi: 10.3956/2015-91.3.211

Enríquez, E., Landaverde-González, P., Lima-Cordón, R., Solórzano-Ortíz, E., Tapia-López, R., and Nuñez-Farfán, J. (2017). Population genetics of traditional landraces of *Cucurbita pepo* L., 1753 in the cloud forest in Baja Verapaz, Guatemala. *Genet. Resour. Crop Evol.* 65, 979–991. doi: 10.1007/s10722-017-0589-y

Enríquez Cotton, M. E. (2017). *Efecto de la Estructura del Paisaje Sobre la Diversidad de Polinizadores, y Genética Poblacional de Cucurbita pepo, en un Bosque de Niebla de Guatemala*. Ph.D. thesis, Universidad Nacional Autonoma de Mexico, Mexico.

Ersts, P. J. (2011). *Geographic Distance Matrix Generator (version 1.2.3). Museum of Natural History. Center for Biodiversity and Conservation*. Available at: http://biodiversityinformatics.amnh.org/open_source/gdmg

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform.* 1, 47–50. doi: 10.1177/117693430500100003

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial restriction data. *Genetics* 131, 479–491.

FAO (2010). *El segundo Informe sobre el Estado de los Recursos Fitogenéticos para la Alimentación y la Agricultura en el Mundo*. Available at: http://www.fao.org/docrep/015/i2624s/i2624s00.pdf

Flannery, K. V. (ed.). (1986). *Guila Naquitz: Archaic Foraging and Early Agriculture in Oaxaca, Mexico*. Orlando, FL: Academic Press.

Formisano, G., Roig, C., Esteras, C., Ercolano, M. R., Nuez, F., Monformte, A. J., et al. (2012). Genetic diversity of Spanish *Cucurbita pepo* landraces: an unexploited resource for summer squash breeding. *Genet. Resour Crop Evol.* 59, 1169–1184. doi: 10.1007/s10722-011-9753-y

Gámez, N., Escalante, T., Espinosa, D., Eguiarte, L. E., and Morrone, J. J. (2014). Temporal dynamics of areas of endemism under climate change: a case study of Mexican Bursera (Burseraceae). *J. Biogeogr.* 41, 871–881. doi: 10.1111/jbi.12249

Gaut, B. S. (2015). Evolution is an experiment: assessing parallelism in crop domestication and experimental evolution: (Nei Lecture, SMBE 2014, Puerto Rico). *Mol. Biol. Evol.* 32, 1661–1671. doi: 10.1093/molbev/msv105

Gaut, B. S., Díez, C. M., and Morrell, P. L. (2015). Genomics and the contrasting dynamics of annual and perennial domestication. *Trends Genet.* 31, 709–719. doi: 10.1016/j.tig.2015.10.002

Gong, L., Paris, H. S., Nee, M. H., Stift, G., Pachner, M., Vollmann, J., et al. (2012). Genetic relationships and evolution in *Cucurbita pepo* (pumpkin, squash, gourd) as revealed by simple sequence repeat polymorphisms. *Theor. Appl. Genet.* 124, 875–891. doi: 10.1007/s00122-011-1752-z

Gong, L., Paris, H. S., Stift, G., Pachner, M., Vollmann, J., and Lelley, T. (2013). Genetic relationships and evolution in *Cucurbita* as viewed with simple sequence repeat polymorphisms: the centrality of *C. okeechobeensis*. *Genet. Resour. Crop Evol.* 60, 1531–1544. doi: 10.1007/s10722-012-9940-5

Gong, L., Stift, G., Kofler, R., Pachner, M., and Lelley, T. (2008). Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L. *Theor. Appl. Genet.* 117, 37–48. doi: 10.1007/s00122-008-0750-2

Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet. Res. Int.* 2015:431487. doi: 10.1155/2015/431487

Hanselka, J. K. (2010). Informal planting of squashes and gourds by rural farmers in southwestern Tamaulipas, Mexico, and implications for the local adoption of food production in prehistory. *J. Ethnobiol.* 30, 31–51. doi: 10.2993/0278-0771-30.1.31

Hedrick, P. W. (2011). *Genetics of Populations*. Burlington, MA: Jones & Bartlett Learning.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Hirzel, A. H., Hausser, J., Chessel, D., and Perrin, N. (2002). Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data?

*Ecology* 83, 2027–2036. doi: 10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2

Holmgren, C. A., Betancourt, J. L., and Rylander, K. A. (2010). A long-term vegetation history of the Mojave-Colorado Desert ecotone at Joshua Tree National Park. *J. Quat. Sci*. 25, 222–236. doi: 10.1002/jqs.1313

Holsinger, K. E., and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat. Rev.* 10, 639–650. doi: 10.1038/nrg2611

Hufford, M. B., Lubinksy, P., Pyhäjärvi, T., Devengenzo, M. T., Ellstrand, N. C., and Ross-Ibarra, J. (2013). The genomic signature of crop-wild introgression in maize. *PLoS Genet.* 9:e1003477. doi: 10.1371/annotation/2eef7b5b-29b2-412f-8472-8fd7f9bd65ab

Hufford, M. B., Martínez-Meyer, E., Gaut, B. S., Eguiarte, L. E., and Tenaillon, M. I. (2012). Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PLoS One* 7:e47659. doi: 10.1371/journal.pone.0047659

Jombart, T. (2008) *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129

Jombart, T., and Ahmed, I. (2011). *adegenet* 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94

Kates, H. R., Soltis, P. S., and Soltis, D. E. (2017). Evolutionary and domestication history of *Cucurbita* (pumpkin and squash) species inferred from 44 nuclear loci. *Mol. Phylogenet. Evol*. 111, 98–109. doi: 10.1016/j.ympev.2017.03.002

Kistler, L., Newsom, L. A., Ryan, T. M., Clarke, A. C., Smith, B. D., and Perry, G. H. (2016). Gourds and squashes (*Cucurbita* ssp.) adapted to megafaunal extinction and ecological anachronism through domestication. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15107–15112. doi: 10.1073/pnas.1516109112

Kocyan, A., Zhang, L. B., Schaefer, H., and Renner, S. S. (2007). A multi-locus chloroplast phylogeny for the Cucurbitaceae and its implications for character evolution and classification. *Mol. Phylogenet. Evol.* 44, 553–577. doi: 10.1016/j.ympev.2006.12.022

Kohn, J. R., and Casper, B. B. (1992). Pollen-mediated gene flow in *Cucurbita foetidissima* (Cucurbitaceae). *Am. J. Bot.* 79, 57–62. doi: 10.2307/2445197

Kraft, K. H., Brown, C. H., Nabhan, G. P., Luedeling, E., Luna Ruiz, J. J., Coppens, G., et al. (2014). Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6165–6170. doi: 10.1073/pnas.1308933111

Lira, R., Casas, A., and Blancas, J. (eds). (2016a). *Ethnobotany of Mexico: Interactions of People and Plants in Mesoamerica*. New York, NY: Springer.

Lira, R., Eguiarte, L., Montes, S., Zizumbo-Villarreal, D., Marín, P. C. G., and Quesada, M. (2016b). "*Homo sapiens*–Cucurbita interaction in mesoamerica: domestication, dissemination, and diversification," in *Ethnobotany of Mexico*, eds R. Lira, A. Casas, and J. Blancas (New York, NY: Springer), 389–401.

Lira-Saade, R. (1995). *Estudios Taxonómicos y Ecogeográficos de las Cucurbitáceas Latinoamericanas de Importancia Económica*. Rome: IPGRI.

Manni, F., Guerard, E., and Heyer, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected using Monmonier's algorithm. *Hum. Biol.* 6, 173–190. doi: 10.1353/hub.2004.0034

Mapes, C., and Basurto, F. (2016). "Biodiversity and edible plants of Mexico," in *Ethnobotany of Mexico*, eds R. Lira, A. Casas, and J. Blancas (New York, NY: Springer), 83–131.

Merrick, L. C., and Bates, D. M. (1989). Classification and nomenclature of *Cucurbita argyrosperma* Huber. *Baileya* 23, 94–102.

Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605

Monmonier, M. S. (1973). Maximum-difference barriers: an alternative numerical regionalization method. *Geogr. Anal*. 5, 245–261. doi: 10.1111/j.1538-4632.1973.tb01011.x

Montes, H. S. (1991). "Calabazas (Cucurbita spp.)," in *Avances en el Estudio de los Recursos Fitogenéticos de Mexico*, eds P. R. Ortega, H. G. Palomino, G. F. Castillo, and V. Gonzáles (Chapingo: Sociedad Mexicana de Fitogenética), 239–250.

Montes, H. S. (2002). *Flujo Génico en Calabaza (Cucurbita spp.) Dentro del Sistema Milpa en el Occidente de México*. Ph.D. thesis, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico.

Montes-Hernández, S., and Eguiarte, L. E. (2002). Genetic structure and indirect estimates of gene flow in three taxa of *Cucurbita* (Cucurbitaceae) in western Mexico. *Am. J. Bot.* 89, 1156–1163. doi: 10.3732/ajb.89.7.1156

Montes-Hernández, S., Merrick, L. C., and Eguiarte, L. E. (2005). Maintenance of squash (*Cucurbita* spp.) landrace diversity by farmers' activities in Mexico. *Genet. Resour. Crop Evol.* 52, 697–707. doi: 10.1007/s10722-003-6018-4

Moreno-Estrada, A., Gignoux, C. R., Fernández-López, J. C., Zakharia, F., Sikora, M., Contreras, A. V., et al. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 314, 1280–1285. doi: 10.1126/science.1251688

Ntuli, N. R., Tongoona, P. B., and Zobolo, A. M. (2015). Genetic diversity in *Cucurbita pepo* landraces revealed by RAPD and SSR markers. *Sci. Hortic.* 189, 192–200. doi: 10.1016/j.scienta.2015.03.020

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* 13, 1143–1155. doi: 10.1111/j.1365-294X.2004.02141.x

Organization for Economic Co-operation and Development [OECD] (2012). *Consensus Document on the Biology of Cucurbita L. (Squashes, Pumpkins, Zucchinis, and gourds). Series on Harmonization of Regulatory Oversight in Biotechnology (53)*. Paris: OECD.

Ornelas, J. F., Sosa, V., Soltis, D. E., Daza, J. M., González, C., Soltis, P. S., et al. (2013). Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. *PLoS One* 8:e56283. doi: 10.1371/journal.pone.0056283

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Peterson, A. T., and Nyári, A. (2008). Ecological niche conservatism and Pleistocene refugia in the thrush-like mourner, Schiffornis spp., *in the Neotropics*. *Evolution* 62, 173–183. doi: 10.1111/j.1558-5646.2007.00258.x

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model*. 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., and Schapire, R. E. (2004). "A maximum entropy approach to species distribution modeling," in *Proceedings of the Twenty-First International Conference of Machine Learning* (New York, NY: ACM), 83. doi: 10.1145/1015330.1015412

Piperno, D. R., Moreno, J. E., Iriarte, J., Holst, I., Lachniet, M., Jones, J. G., et al. (2007). Late Pleistocene and Holocene environmental history of the Iguala Valley, central Balsas watershed of Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11874–11881. doi: 10.1073/pnas.0703442104

Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J., and Dickau, R. (2009). Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5019–5024. doi: 10.1073/pnas.0812525106

Priori, D., Barbieri, R. L., Castro, C. M., de Oliveira, A. C., Vilela, J. C. B., and Mistura, C. C. (2013). Diversidade genética de *Cucurbita pepo, C. argyrosperma e C. ficifolia* empregando marcadores microssatélites. *Hortic. Bras.* 31, 361–368. doi: 10.1590/S0102-05362013000300004

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Purugganan, M. D., and Fuller, D. Q. (2011). Archaeological data reveal slow rates of evolution during plant domestication. *Evolution* 65, 171–183. doi: 10.1111/j.1558-5646.2010.01093.x

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Rambaut, A., and Drummond, A. J. (2009). *Tracer: MCMC Trace Analysis Tool, Version 1.5*. Available at: http://tree.bio.ed.ac.uk/software/tracer/

Rannala, B. (2007). *BayesAss Edition 3.0 User's Manual*. Davis, CA: University of California at Davis.

Rannala, B., and Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9197–9201. doi: 10.1073/pnas.94.17.9197

Rannere, A. J., Piperno, D. R., Holst, I., Dickau, R., and Iriarte, J. (2009). The cultural and chronological context of early Holocene maize and squash

domestication in the Central Balsas River Valley, Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5014–5018. doi: 10.1073/pnas.0812590106

Rousset, F. (2008). Genepop'007: a complete reimplementation of the genepop software for Windows and Linux. *Mol. Ecol. Notes* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Sanjur, O. I., Piperno, D. R., Andres, T. C., and Wessel-Beaver, L. (2002). Phylogenetic relationships among domesticated and wild species of Cucurbita (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. *Proc. Natl. Acad. Sci. U.S.A.* 99, 535–540. doi: 10.1073/pnas.012577299

Scheinvar, E., Gámez, N., Castellanos-Morales, G., Aguirre-Planter, E., and Eguiarte, L. E. (2017). Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan Desert. *J. Biogeogr.* 44, 322–334. doi: 10.1111/jbi.12851

Servicio de Información Agroalimentaria y Pesquera [SIAP] (2016). *Anuario Estadístico de la Producción Agrícola*. Available at: http://nube.siap.gob.mx/cierre_agricola

Smith, B. D. (1997). The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science* 276, 932–934. doi: 10.1126/science.276.5314.932

Szpiech, Z. A., Jackobsson, M., and Rosenberg, N. A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24, 2498–2504. doi: 10.1093/bioinformatics/btn478

Van Etten, J., and Hijmans, R. J. (2010). A geospatial modeling approach integrating archaeobotany and genetics to trace the dispersal of domesticated plants. *PLoS One* 5:e12060. doi: 10.1371/journal.pone.0012060

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., and Shipley, P. (2004). MICROCHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4, 535–538. doi: 10.1111/j.1471-8286.2004.00684.x

Villanueva, V. C. (2007). *Calabazas Cultivadas. Identificación de Especies, Caracterización y Descripción Varietal*. Mexico: Universidad Autónoma Chapingo.

Waltari, E., and Guralnick, R. P. (2009). Ecological niche modelling of montane mammals in the Great Basin, North America: examining past and present connectivity of species across basins and ranges. *J. Biogeogr.* 36, 148–161. doi: 10.1111/j.1365-2699.2008.01959.x

Waltari, E., Hijmans, R. J., Peterson, A. T., Nyári, A. S., Perkin, S. L., and Guralnick, R. P. (2007). Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS One* 2:e563. doi: 10.1371/journal.pone.0000563

Warschefsky, E., Penmetsa, R. V., Cook, D. R., and von Wettberg, E. J. (2014). Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am. J. Bot.* 101, 1791–1800. doi: 10.3732/ajb.1400116

Wilson, G. A., and Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163, 1177–1191.

Wilson, H. D., Doebley, J., and Duvall, M. (1992). Chloroplast DNA diversity among wild and cultivated members of Cucurbita (Cucurbitaceae). *Theor. Appl. Genet.* 84, 859–865. doi: 10.1007/BF00227397

Zheng, Y. H., Alverson, A. J., Wang, Q. F., and Palmer, J. D. (2013). Chloroplast phylogeny of Cucurbita: evolution of the domesticated and wild species. *J. Syst. Evol.* 51, 326–334. doi: 10.1111/jse.12006

Zizumbo-Villarreal, D., and Colunga-GarcíaMarín, P. (2010). Origin of agriculture and plant domestication in West Mesoamerica. *Genet. Resour. Crop Evol.* 57, 813–825. doi: 10.1007/s10722-009-9521-4

Zizumbo-Villarreal, D., Colunga-GarcíaMarín, P., and Flores-Silva, A. (2016). "Pre-Columbian food system in West Mesoamerica," in *Ethnobotany of Mexico: Interactions of People and Plants in Mesoamerica*, eds R. Lira, A. Casas, and J. Blancas (New York, NY: Springer), 67–82. doi: 10.1007/978-1-4614-6669-7_4

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Genetic resources in the "calabaza pipiana" squash (*Cucurbita argyrosperma*) in Mexico: Genetic diversity, genetic differentiation and distribution models.

Guillermo Sanchez de la Vega[1], Gabriela Castellanos-Morales[1,2], Niza Gámez[1], Helena S. Hernández-Rosales[1], Alejandra Vázquez-Lobo[1,3], Erika Aguirre-Planter[1], Salvador Montes-Hernández[4], Juan P. Jaramillo-Correa, Rafael Lira-Saade[2*], Luis E. Eguiarte[1*]

**Supplementary Table S1.** State and population identification of collected samples, number of individuals (n) and decimal coordinates for a) *C. argyrosperma* ssp. *argyrosperma* and b) *C. argyrosperma* ssp. *sororia.* Populations obtained from the germplasm collection of the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Campo Experimental Bajío (BG).

*a) C. argyrosperma* ssp. *argyrosperma*

| State | Population | n | Population ID | Coordinates | |
|---|---|---|---|---|---|
| | | | | W | N |
| Quintana Roo | Chan_Santa_Cruz | 21 | Chan | -88.33289167 | 19.36706389 |
| Yucatán | Ek_Balam | 22 | Ek | -87.91666667 | 20.91666667 |
| Yucatán | Motul | 22 | Mot | -88.82540556 | 21.24834444 |
| Campeche | Champoton | 22 | Champ | -90.46137778 | 19.50115556 |
| Chiapas | Palenque | 20 | Pal | -91.98776111 | 17.51281389 |
| Oaxaca | Tehuantepec | 20 | Teh | -95.23303611 | 16.33283056 |
| Oaxaca | Mixtepec | 21 | Mixt | -97.08491667 | 15.95875 |
| Guerrero | Ometepec | 17 | Ome | -98.40631944 | 16.69614722 |
| Guerrero | Matlalapa (BG) | 18 | Tla | -100.5347222 | 18.24166667 |
| Guerrero | Tlapehuala (BG) | 10 | Mtp | -99.59026111 | 36.36916667 |
| Michoacán | Sahuayo | 22 | Sah | -102.7162333 | 20.05871944 |
| Jalisco | Autlán | 21 | Aut | -104.2043139 | 19.70195833 |
| Guanajuato | San José Iturbide (BG) | 6 | SJI | -100.387222 | 20.998889 |
| San Luis Potosí | Tanquian (BG) | 22 | Tan | -101.0093306 | 22.11548611 |
| Veracruz | Tihuatlán (BG) | 23 | Tih | -97.53950278 | 20.72006667 |
| Durango | Durango (BG) | 6 | Dgo | 104.583333 | 24.066667 |
| Coahuila | Cuatrocienegas | 22 | CCC | -102.066389 | 26.986111 |
| Sinaloa | Sinaloa (BG) | 6 | SinalP | -106.0555556 | 23.27388889 |
| Sonora | Yecora (BG) | 6 | Yec | -108.9269861 | 28.37204167 |

*b) C. argyrosperma* ssp. *sororia*

| State | Population | n | Population ID | Coordinates W | Coordinates N |
|---|---|---|---|---|---|
| | | | ***Cucurbita argyrosperma ssp. sororia*** | | |
| Chiapas | Chiapas | 20 | 20. SChis | -93.626878 | 16.597486 |
| Oaxaca | Oaxaca | 21 | 21.SOax | -97.076308 | 15.918225 |
| Guerrero | Guerrero | 22 | 22. SGro | -98.406319 | 16.688139 |
| Jalisco | Jalisco | 22 | 23.SJal | -104.2043139 | 19.70195833 |
| Sinaloa | Sinaloa (BG) | 13 | 24.SoSin | -106.416667 | 24 |
| Sonora | Sonora (BG) | 15 | 25.SoSon | -109.583333 | 28.533333 |

**Supplementary Table S2.** List of nuclear microsatellite loci (Gong et al., 2008) used in the current analysis of *Cucurbita argyrosperma*.

| Marker name | Alleles | Motif | No.of Repeats | Forward primer | Reverse primer | Expected size (bp) | Annealing temperature (°C) |
|---|---|---|---|---|---|---|---|
| CMTp17 | 4 | CT | 16 | ACTGCTCAATAAGGCAAGGA | AAACAAGAGTGCACAAACAGG | 84 | 58 |
| CMTp88 | 5 | TC | 12 | ACCTACCGTCACACCCACAT | CCACCTGAAAACAGGGCTAA | 167 | 60 |
| CMTp129 | 5 | AG | 19 | CTCTTGCTCATCTTCCTTGTTG | CCCACCCATTTACCCTCTAGT | 147 | 59 |
| CMTp175 | 5 | GA | 11 | TCCAATGCACAACCTTGC | GCCTCGGTTTTTGTCAAGAT | 155 | 59 |
| CMTp193 | 5 | GA | 18 | GGTGACGGCAAGAAAAGCTA | GCTGACCCTCTCTCCCTCTC | 186 | 60 |
| CMTm11 | 5 | AG | 14 | TGGAAGGATTCTCCCACAGT | TACAATTTGACGTCCGCAAG | 108 | 59 |
| CMTm54 | 5 | CT | 15 | GTGTGGATGCAAATGGTGAG | GGGAATCGAGGGTTTTGAAT | 143 | 60 |
| CMTm120 | 4 | CT | 13 | GCCAAAGGTTCCAAATGACA | TGATTTGCGCACAaACAAAC | 121 | 60 |
| CMTm144 | 4 | AG | 11 | ACATGGGCATACCTCGAATC | CACCTGGCTGTTTTGTCTGA | 150 | 60 |
| CMTm187 | 4 | AG | 19 | ATCGGTGAGTCCCAAAAATG | ATCACAAAGCGGGAAAACAC | 128 | 60 |
| CMTm221 | 4 | CT | 11 | CAATAAGATAGCTCTCACGTTGC | TGCCTAGTTATCGCGACTTC | 108 | 58 |
| CMTm261 | 4 | A | 21 | GGTGGCCTCTGAACAATTTC | ACCTAACCAATGGGCATGAG | 228 | 60 |

**Supplementary Table S3.** *Cucurbita argyrosperma* populations sampled and loci with significant departures from Hardy–Weinberg (HW) equilibrium (* $p < 0.05$ and ** $p < 0.001$). Values correspond to $F_{IS}$ estimates for each locus and each population. Population ID from Supplementary Table S1.

| Population ID | CMTp17 | CMTm120 | CMTm261 | CMTm11 | CMTp129 | CMTp193 | CMTm221 | CMTm54 | CMTp88 | Multi-locus |
|---|---|---|---|---|---|---|---|---|---|---|
| Chan | 0.068 | -0.2 | --- | --- | 0.647** | --- | --- | --- | 0.786** | 0.374** |
| Ek | -0.629** | -0.235 | --- | -0.105 | -0.105 | -0.235 | --- | 0 | 1* | -0.254* |
| Mot | 0.05 | 0.35 | --- | -0.077 | -0.286 | 0 | -0.053 | -0.11 | 0.661* | 0.052 |
| Champ | -0.105 | 0.045 | --- | --- | 0.469** | -0.135 | -0.05 | 0 | 1** | 0.277** |
| Pal | -0.67** | -0.103 | 0 | -0.132 | 0.13 | -0.166 | --- | -0.187 | -0.027 | -0.189* |
| Teh | -0.299 | -0.09 | -0.056 | 1* | 0.013 | 0.13 | 0.159 | 0.172 | --- | 0.038 |
| Mix | -0.316 | 0.588* | --- | -0.053 | -0.101 | -0.212 | 0.355* | 1* | --- | 0.12* |
| Ome | -0.387 | -0.439 | --- | -0.867* | 0.222 | -0.481 | 0.65 | -0.077 | -0.12 | -0.26 |
| Tla | 0.196 | 0.273 | --- | -0.346 | 0.386 | -0.286 | --- | 0 | --- | 0.018 |
| Mtp | -0.172 | 0.15 | --- | -0.259 | -0.684* | -0.417 | -0.083 | 0 | 0.382 | -0.174* |
| Sah | -0.11 | 0.129 | --- | -0.312 | -0.503* | -0.448 | -0.024 | --- | --- | -0.26** |
| Aut | 0.096 | 0.355 | --- | -0.176 | 0.395** | -0.073 | --- | --- | 0 | 0.131* |
| SJI | -0.091 | -0.25 | 0.333 | -0.667 | --- | --- | -0.429 | --- | 1 | -0.067 |
| Tan | 0.173 | 0.332* | 1** | -0.135 | -0.077 | -0.312 | --- | -0.14 | --- | 0.203* |
| Tih | 0.509* | 0.333 | --- | -0.193 | 0.134 | -0.294 | 0.355* | 0.662* | 1** | 0.238** |
| Dgo | 1* | -0.25 | --- | -0.25 | 0 | -0.111 | 0.062 | -0.429 | -0.25 | 0.04* |
| CCC | -0.279 | -0.188 | 0.894** | 1* | 0 | -0.293 | --- | -0.07 | 0.878** | 0.128** |
| SinalP | 0.5 | --- | -1 | --- | --- | -0.333 | --- | -0.333 | 1 | 0.059 |
| Yec | -0.667 | 0.259 | --- | 0.167 | 0.4 | 1* | --- | -0.333 | 0 | 0.185* |
| SChis | 0.823* | 0.237 | 0.202 | --- | --- | -0.691* | --- | -0.267 | --- | 0.085 |
| SOax | 0.2 | -0.135 | --- | --- | 1** | -0.6 | --- | 0.707** | --- | 0.304** |
| SGro | 1 | 0.36* | --- | -0.024 | 0.131 | -0.294 | -0.05 | 1** | --- | 0.225* |
| SJal | 1* | 0.148 | --- | -0.448 | 0.601** | -0.273 | --- | -0.333* | --- | 0.035 |
| SoSin | --- | -0.352* | -0.037 | -0.251 | 0.783** | -0.333 | -0.043 | -0.143 | --- | -0.107 |
| SoSon | --- | 0.248 | -0.077 | 0.012 | -0.077 | -0.125 | --- | --- | --- | -0.013 |
| Overall | 0.017 | 0.068 | 0.148* | -0.235* | 0.271** | -0.226** | 0.06** | -0.019 | 0.599** | 0.043** |

**Supplementary Table S4.** Matrix of estimates of recent migration rates among populations and subspecies of *Cucurbita argyrosperma* obtained with BayesAss based on 12 nuclear microsatellite loci. The proportion of non-migrants is shown in the main diagonal (in bold). The migration matrix is presented in three parts: a) populations 1 – 10 of *C. argyrosperma* ssp. *argyrosperma*; b) populations 11 – 19 of *C. argyrosperma* ssp. *argyrosperma*; and c) populations 20 – 25 correspond to *C. argyrosperma* ssp. *sororia* (shaded in light gray). Population ID from Supplementary Table S1.

*a)* Populations 1 – 10 of *C. argyrosperma* ssp. *argyrosperma*

| Population ID | 1. Chan | 2. Ek | 3. Mot | 4. Champ | 5.Pal | 6. Teh | 7. Mixt | 8.Ome | 9.Tla | 10.Mtp |
|---|---|---|---|---|---|---|---|---|---|---|
| **1. Chan** | **0.8175 (0.0251)** | 0.0159 (0.0135) | 0.0070 (0.0069) | 0.0073 (0.0070) | 0.0069 (0.0068) | 0.0069 (0.0070) | 0.0071 (0.0070) | 0.0071 (0.0070) | 0.0070 (0.0069) | 0.0075 (0.0076) |
| **2. Ek** | 0.0087 (0.0086) | **0.8260 (0.0249)** | 0.0069 (0.0068) | 0.0072 (0.0072) | 0.0072 (0.0070) | 0.0071 (0.0072) | 0.0072 (0.0072) | 0.0068 (0.0068) | 0.0072 (0.0068) | 0.0068 (0.0068) |
| **3. Mot** | 0.0101 (0.0094) | 0.1527 (0.0233) | **0.6740 (0.0072)** | 0.0071 (0.0070) | 0.0072 (0.0070) | 0.0071 (0.0070) | 0.0074 (0.0072) | 0.0071 (0.0070) | 0.0073 (0.0071) | 0.0073 (0.0072) |
| **4. Champ** | 0.1385 (0.0259) | 0.0288 (0.0166) | 0.0072 (0.0070) | **0.6739 (0.0071)** | 0.0072 (0.0071) | 0.0072 (0.0071) | 0.0072 (0.0071) | 0.0071 (0.0070) | 0.0071 (0.0069) | 0.0074 (0.0072) |
| **5.Pal** | 0.0095 (0.0090) | 0.1548 (0.0252) | 0.0073 (0.0072) | 0.0073 (0.0072) | **0.6741 (0.0072)** | 0.0073 (0.0072) | 0.0073 (0.0071) | 0.0074 (0.0072) | 0.0075 (0.0073) | 0.0073 (0.0070) |
| **6. Teh** | 0.0222 (0.0147) | 0.0140 (0.0124) | 0.0073 (0.0072) | 0.0076 (0.0075) | 0.0074 (0.0074) | **0.7841 (0.0265)** | 0.0095 (0.0093) | 0.0076 (0.0073) | 0.0075 (0.0074) | 0.0150 (0.0130) |
| **7. Mixt** | 0.0074 (0.0072) | 0.0082 (0.0076) | 0.0069 (0.0065) | 0.0071 (0.0069) | 0.0074 (0.0071) | 0.0088 (0.0086) | **0.8135 (0.0251)** | 0.0072 (0.0072) | 0.0075 (0.0073) | 0.0075 (0.0072) |
| **8.Ome** | 0.0084 (0.0081) | 0.0083 (0.0080) | 0.0082 (0.0081) | 0.0083 (0.0083) | 0.0085 (0.0083) | 0.0082 (0.0080) | 0.0087 (0.0085) | **0.6750 (0.0081)** | 0.0083 (0.0081) | 0.0083 (0.0083) |
| **9.Tla** | 0.0160 (0.0138) | 0.0221 (0.0181) | 0.0095 (0.0091) | 0.0096 (0.0095) | 0.0096 (0.0093) | 0.0095 (0.0091) | 0.0116 (0.0109) | 0.0095 (0.0092) | **0.6762 (0.0093)** | 0.0095 (0.0092) |
| **10.Mtp** | 0.0079 (0.0077) | 0.0093 (0.0088) | 0.0080 (0.0078) | 0.0079 (0.0075) | 0.0079 (0.0077) | 0.0096 (0.0095) | 0.0082 (0.0077) | 0.0076 (0.0074) | 0.0075 (0.0077) | **0.8002 (0.0263)** |
| **11.Sah** | 0.0158 (0.0110) | 0.1477 (0.0243) | 0.0073 (0.0071) | 0.0072 (0.0071) | 0.0071 (0.0071) | 0.0073 (0.0071) | 0.0076 (0.0074) | 0.0072 (0.0069) | 0.0069 (0.0069) | 0.0070 (0.0068) |
| **12.Aut** | 0.0073 (0.0074) | 0.0090 (0.0087) | 0.0072 (0.0069) | 0.0073 (0.0072) | 0.0073 (0.0069) | 0.0083 (0.0084) | 0.0092 (0.0088) | 0.0073 (0.0071) | 0.0071 (0.0073) | 0.0079 (0.0078) |
| **13.SJI** | 0.0106 (0.0102) | 0.0108 (0.0104) | 0.0106 (0.0102) | 0.0109 (0.0105) | 0.0109 (0.0105) | 0.0105 (0.0102) | 0.0758 (0.0248) | 0.0107 (0.0104) | 0.0108 (0.0104) | 0.0108 (0.0104) |

| Population ID | 1. Chan | 2. Ek | 3. Mot | 4. Champ | 5.Pal | 6. Teh | 7. Mixt | 8.Ome | 9.Tla | 10.Mtp |
|---|---|---|---|---|---|---|---|---|---|---|
| **14.Tan** | 0.0073 (0.0072) | 0.0092 (0.0090) | 0.0069 (0.0069) | 0.0072 (0.0070) | 0.0071 (0.0071) | 0.0070 (0.0068) | 0.0691 (0.0234) | 0.0073 (0.0071) | 0.0070 (0.0067) | 0.0072 (0.0072) |
| **15.Tih** | 0.0069 (0.0067) | 0.0072 (0.0068) | 0.0072 (0.0067) | 0.0070 (0.0068) | 0.0069 (0.0065) | 0.0072 (0.0070) | 0.0073 (0.0074) | 0.0070 (0.0070) | 0.0072 (0.0070) | 0.0072 (0.0068) |
| **16.Dgo** | 0.0108 (0.0106) | 0.0108 (0.0104) | 0.0107 (0.0106) | 0.0107 (0.0104) | 0.0109 (0.0105) | 0.0108 (0.0105) | 0.0109 (0.0107) | 0.0107 (0.0104) | 0.0108 (0.0104) | 0.0110 (0.0107) |
| **17.CCC** | 0.0076 (0.0074) | 0.0070 (0.0070) | 0.0072 (0.0070) | 0.0073 (0.0073) | 0.0073 (0.0073) | 0.0071 (0.0068) | 0.0072 (0.0072) | 0.0076 (0.0073) | 0.0074 (0.0070) | 0.0073 (0.0072) |
| **18.SinalP** | 0.0119 (0.0115) | 0.0119 (0.0116) | 0.0119 (0.0116) | 0.0121 (0.0116) | 0.0118 (0.0115) | 0.0118 (0.0113) | 0.0119 (0.0116) | 0.0119 (0.0115) | 0.0118 (0.0114) | 0.0122 (0.0116) |
| **19. Yec** | 0.0107 (0.0103) | 0.0109 (0.0105) | 0.0105 (0.0103) | 0.0106 (0.0104) | 0.0106 (0.0103) | 0.0107 (0.0103) | 0.0113 (0.0107) | 0.0106 (0.0104) | 0.0106 (0.0102) | 0.0107 (0.0105) |
| **20. SChis** | 0.0073 (0.0072) | 0.0073 (0.0073) | 0.0071 (0.0069) | 0.0073 (0.0072) | 0.0074 (0.0072) | 0.0074 (0.0074) | 0.0075 (0.0072) | 0.0078 (0.0079) | 0.0075 (0.0076) | 0.0073 (0.0072) |
| **21.SOax** | 0.0073 (0.0071) | 0.0072 (0.0069) | 0.0072 (0.0072) | 0.0073 (0.0069) | 0.0073 (0.0070) | 0.0074 (0.0072) | 0.0076 (0.0077) | 0.0076 (0.0073) | 0.0072 (0.0072) | 0.0074 (0.0074) |
| **22. SGro** | 0.0072 (0.0069) | 0.0071 (0.0070) | 0.0069 (0.0067) | 0.0068 (0.0068) | 0.0074 (0.0069) | 0.0070 (0.0068) | 0.0071 (0.0067) | 0.0072 (0.0071) | 0.0070 (0.0071) | 0.0072 (0.0070) |
| **23.SJal** | 0.0073 (0.0071) | 0.0070 (0.0068) | 0.0068 (0.0066) | 0.0070 (0.0069) | 0.0074 (0.0073) | 0.0070 (0.0066) | 0.0071 (0.0068) | 0.0072 (0.0069) | 0.0072 (0.0069) | 0.0069 (0.0068) |
| **24.SoSin** | 0.0088 (0.0086) | 0.0088 (0.0085) | 0.0085 (0.0083) | 0.0089 (0.0086) | 0.0088 (0.0085) | 0.0088 (0.0085) | 0.0087 (0.0085) | 0.0088 (0.0086) | 0.0088 (0.0085) | 0.0093 (0.0093) |
| **25.SoSon** | 0.0084 (0.0083) | 0.0085 (0.0084) | 0.0084 (0.0082) | 0.0083 (0.0080) | 0.0082 (0.0081) | 0.0083 (0.0082) | 0.0084 (0.0081) | 0.0084 (0.0081) | 0.0085 (0.0082) | 0.0083 (0.0081) |

*b)* Populations 11 – 19 of *C. argyrosperma* ssp. *argyrosperma*

| Population ID | 11.Sah | 12.Aut | 13.SJI | 14.Tan | 15.Tih | 16.Dgo | 17.CCC | 18.SinalP | 19. Yec |
|---|---|---|---|---|---|---|---|---|---|
| **1. Chan** | 0.0080 (0.0076) | 0.0072 (0.0070) | 0.0070 (0.0069) | 0.0091 (0.0083) | 0.0072 (0.0073) | 0.0073 (0.0072) | 0.0075 (0.0075) | 0.0074 (0.0071) | 0.0069 (0.0067) |
| **2. Ek** | 0.0089 (0.0084) | 0.0070 (0.0070) | 0.0072 (0.0069) | 0.0076 (0.0077) | 0.0070 (0.0068) | 0.0070 (0.0071) | 0.0070 (0.0072) | 0.0072 (0.0070) | 0.0071 (0.0069) |
| **3. Mot** | 0.0072 (0.0070) | 0.0072 (0.0071) | 0.0071 (0.0069) | 0.0119 (0.0103) | 0.0073 (0.0070) | 0.0072 (0.0071) | 0.0073 (0.0072) | 0.0072 (0.0072) | 0.0072 (0.0071) |
| **4. Champ** | 0.0088 (0.0084) | 0.0072 (0.0070) | 0.0070 (0.0068) | 0.0070 (0.0068) | 0.0072 (0.0071) | 0.0071 (0.0070) | 0.0071 (0.0070) | 0.0072 (0.0069) | 0.0071 (0.0070) |
| **5.Pal** | 0.0076 (0.0075) | 0.0074 (0.0072) | 0.0074 (0.0073) | 0.0074 (0.0073) | 0.0075 (0.0073) | 0.0074 (0.0071) | 0.0072 (0.0070) | 0.0074 (0.0072) | 0.0072 (0.0071) |
| **6. Teh** | 0.0092 (0.0090) | 0.0117 (0.0099) | 0.0076 (0.0071) | 0.0082 (0.0077) | 0.0073 (0.0072) | 0.0073 (0.0070) | 0.0075 (0.0075) | 0.0074 (0.0073) | 0.0073 (0.0071) |
| **7. Mixt** | 0.0097 (0.0093) | 0.0138 (0.0115) | 0.0074 (0.0071) | 0.0071 (0.0072) | 0.0074 (0.0071) | 0.0076 (0.0072) | 0.0074 (0.0071) | 0.0064 (0.0065) | 0.0074 (0.0073) |
| **8.Ome** | 0.1329 (0.0254) | 0.0083 (0.0082) | 0.0084 (0.0082) | 0.0084 (0.0082) | 0.0083 (0.0081) | 0.0084 (0.0081) | 0.0084 (0.0083) | 0.0084 (0.0082) | 0.0083 (0.0081) |
| **9.Tla** | 0.0450 (0.0222) | 0.0342 (0.0201) | 0.0095 (0.0093) | 0.0232 (0.0160) | 0.0095 (0.0093) | 0.0097 (0.0092) | 0.0097 (0.0092) | 0.0096 (0.0092) | 0.0095 (0.0093) |
| **10.Mtp** | 0.0083 (0.0079) | 0.0115 (0.0112) | 0.0077 (0.0076) | 0.0125 (0.0102) | 0.0080 (0.0079) | 0.0078 (0.0074) | 0.0077 (0.0076) | 0.0077 (0.0076) | 0.0077 (0.0075) |
| **11.Sah** | **0.6762 (0.0092)** | 0.0081 (0.0080) | 0.0072 (0.0070) | 0.0081 (0.0077) | 0.0076 (0.0074) | 0.0072 (0.0070) | 0.0072 (0.0070) | 0.0071 (0.0069) | 0.0072 (0.0071) |
| **12.Aut** | 0.0088 (0.0087) | **0.8174 (0.0252)** | 0.0073 (0.0074) | 0.0090 (0.0086) | 0.0076 (0.0076) | 0.0068 (0.0066) | 0.0073 (0.0073) | 0.0073 (0.0071) | 0.0074 (0.0073) |
| **13.SJI** | 0.0106 (0.0102) | 0.0107 (0.0103) | **0.6774 (0.0104)** | 0.0106 (0.0103) | 0.0109 (0.0104) | 0.0108 (0.0105) | 0.0107 (0.0103) | 0.0104 (0.0101) | 0.0108 (0.0106) |
| **14.Tan** | 0.0075 (0.0072) | 0.0111 (0.0107) | 0.0070 (0.0067) | **0.7606 (0.0249)** | 0.0074 (0.0070) | 0.0072 (0.0070) | 0.0071 (0.0071) | 0.0071 (0.0069) | 0.0071 (0.0068) |
| **15.Tih** | 0.0071 (0.0069) | 0.0128 (0.0095) | 0.0071 (0.0071) | 0.0069 (0.0066) | **0.8262 (0.0237)** | 0.0067 (0.0065) | 0.0068 (0.0066) | 0.0068 (0.0067) | 0.0070 (0.0071) |
| **16.Dgo** | 0.0109 (0.0105) | 0.0111 (0.0105) | 0.0109 (0.0104) | 0.0110 (0.0107) | 0.0109 (0.0105) | **0.6776 (0.0106)** | 0.0735 (0.0251) | 0.0107 (0.0104) | 0.0107 (0.0102) |

| Population ID | 11.Sah | 12.Aut | 13.SJI | 14.Tan | 15.Tih | 16.Dgo | 17.CCC | 18.SinalP | 19. Yec |
|---|---|---|---|---|---|---|---|---|---|
| **17.CCC** | 0.0072 (0.0069) | 0.0072 (0.0071) | 0.0074 (0.0073) | 0.0073 (0.0071) | 0.0075 (0.0073) | 0.0071 (0.0072) | **0.8253 (0.0245)** | 0.0072 (0.0070) | 0.0072 (0.0071) |
| **18.SinalP** | 0.0118 (0.0113) | 0.0119 (0.0114) | 0.0120 (0.0117) | 0.0337 (0.0195) | 0.0130 (0.0127) | 0.0119 (0.0116) | 0.0248 (0.0171) | **0.6785 (0.0115)** | 0.0119 (0.0115) |
| **19. Yec** | 0.0107 (0.0105) | 0.0766 (0.0253) | 0.0106 (0.0103) | 0.0107 (0.0104) | 0.0105 (0.0104) | 0.0106 (0.0103) | 0.0106 (0.0104) | 0.0107 (0.0101) | **0.6774 (0.0105)** |
| **20. SChis** | 0.0073 (0.0071) | 0.0073 (0.0070) | 0.0076 (0.0072) | 0.0079 (0.0076) | 0.0076 (0.0073) | 0.0073 (0.0072) | 0.0078 (0.0076) | 0.0073 (0.0070) | 0.0076 (0.0071) |
| **21.SOax** | 0.0074 (0.0069) | 0.0071 (0.0070) | 0.0074 (0.0070) | 0.0069 (0.0068) | 0.0074 (0.0072) | 0.0070 (0.0068) | 0.0068 (0.0067) | 0.0072 (0.0070) | 0.0071 (0.0070) |
| **22. SGro** | 0.0070 (0.0069) | 0.0072 (0.0069) | 0.0073 (0.0070) | 0.0069 (0.0066) | 0.0072 (0.0071) | 0.0072 (0.0071) | 0.0070 (0.0069) | 0.0073 (0.0071) | 0.0072 (0.0068) |
| **23.SJal** | 0.0067 (0.0067) | 0.0070 (0.0070) | 0.0072 (0.0073) | 0.0068 (0.0068) | 0.0073 (0.0070) | 0.0073 (0.0073) | 0.0074 (0.0071) | 0.0071 (0.0070) | 0.0074 (0.0072) |
| **24.SoSin** | 0.0086 (0.0085) | 0.0099 (0.0097) | 0.0088 (0.0085) | 0.1221 (0.0261) | 0.0088 (0.0084) | 0.0086 (0.0084) | 0.0086 (0.0085) | 0.0088 (0.0085) | 0.0087 (0.0085) |
| **25.SoSon** | 0.0085 (0.0082) | 0.0083 (0.0081) | 0.0083 (0.0080) | 0.1324 (0.0247) | 0.0084 (0.0080) | 0.0083 (0.0081) | 0.0082 (0.0079) | 0.0085 (0.0082) | 0.0085 (0.0082) |

*c)* Populations 20 – 25 correspond to *C. argyrosperma* ssp. *sororia*

| Population ID | 20. SChis | 21.SOax | 22. SGro | 23.SJal | 24.SoSin | 25.SoSon |
|---|---|---|---|---|---|---|
| **1. Chan** | 0.0068 (0.0067) | 0.0071 (0.0069) | 0.0071 (0.0068) | 0.0074 (0.0069) | 0.0068 (0.0069) | 0.0070 (0.0070) |
| **2. Ek** | 0.0072 (0.0068) | 0.0076 (0.0074) | 0.0072 (0.0070) | 0.0072 (0.0071) | 0.0072 (0.0071) | 0.0069 (0.0067) |
| **3. Mot** | 0.0072 (0.0070) | 0.0072 (0.0071) | 0.0073 (0.0071) | 0.0073 (0.0071) | 0.0071 (0.0069) | 0.0071 (0.0069) |
| **4. Champ** | 0.0072 (0.0070) | 0.0072 (0.0068) | 0.0070 (0.0068) | 0.0071 (0.0068) | 0.0071 (0.0069) | 0.0070 (0.0069) |
| **5.Pal** | 0.0074 (0.0072) | 0.0073 (0.0072) | 0.0072 (0.0072) | 0.0072 (0.0070) | 0.0073 (0.0070) | 0.0074 (0.0072) |
| **6. Teh** | 0.0074 (0.0071) | 0.0075 (0.0075) | 0.0071 (0.0072) | 0.0075 (0.0072) | 0.0073 (0.0072) | 0.0075 (0.0072) |

| Population ID | 20. SChis | 21.SOax | 22. SGro | 23.SJal | 24.SoSin | 25.SoSon |
|---|---|---|---|---|---|---|
| **7. Mixt** | 0.0075 (0.0071) | 0.0075 (0.0070) | 0.0073 (0.0071) | 0.0074 (0.0073) | 0.0073 (0.0072) | 0.0073 (0.0073) |
| **8.Ome** | 0.0084 (0.0082) | 0.0083 (0.0081) | 0.0084 (0.0082) | 0.0084 (0.0082) | 0.0082 (0.0082) | 0.0082 (0.0080) |
| **9.Tla** | 0.0095 (0.0092) | 0.0095 (0.0092) | 0.0095 (0.0093) | 0.0097 (0.0093) | 0.0096 (0.0094) | 0.0095 (0.0092) |
| **10.Mtp** | 0.0079 (0.0078) | 0.0078 (0.0075) | 0.0079 (0.0078) | 0.0078 (0.0076) | 0.0078 (0.0074) | 0.0078 (0.0078) |
| **11.Sah** | 0.0071 (0.0070) | 0.0072 (0.0069) | 0.0072 (0.0071) | 0.0072 (0.0069) | 0.0072 (0.0071) | 0.0072 (0.0069) |
| **12.Aut** | 0.0075 (0.0071) | 0.0068 (0.0067) | 0.0075 (0.0075) | 0.0070 (0.0068) | 0.0072 (0.0072) | 0.0072 (0.0072) |
| **13.SJI** | 0.0106 (0.0102) | 0.0108 (0.0106) | 0.0109 (0.0107) | 0.0109 (0.0104) | 0.0106 (0.0103) | 0.0109 (0.0106) |
| **14.Tan** | 0.0072 (0.0069) | 0.0071 (0.0069) | 0.0071 (0.0069) | 0.0072 (0.0069) | 0.0072 (0.0070) | 0.0070 (0.0069) |
| **15.Tih** | 0.0068 (0.0067) | 0.0072 (0.0069) | 0.0067 (0.0064) | 0.0069 (0.0068) | 0.0069 (0.0068) | 0.0070 (0.0064) |
| **16.Dgo** | 0.0108 (0.0104) | 0.0108 (0.0103) | 0.0108 (0.0104) | 0.0109 (0.0106) | 0.0108 (0.0104) | 0.0108 (0.0105) |
| **17.CCC** | 0.0073 (0.0071) | 0.0073 (0.0070) | 0.0078 (0.0075) | 0.0071 (0.0072) | 0.0073 (0.0070) | 0.0069 (0.0068) |
| **18.SinalP** | 0.0117 (0.0114) | 0.0121 (0.0115) | 0.0116 (0.0113) | 0.0120 (0.0116) | 0.0120 (0.0117) | 0.0120 (0.0115) |
| **19. Yec** | 0.0108 (0.0105) | 0.0108 (0.0106) | 0.0106 (0.0104) | 0.0109 (0.0106) | 0.0106 (0.0104) | 0.0107 (0.0103) |
| **20. SChis** | **0.8219 (0.0246)** | 0.0074 (0.0073) | 0.0074 (0.0070) | 0.0076 (0.0075) | 0.0073 (0.0071) | 0.0070 (0.0068) |
| **21.SOax** | 0.0069 (0.0072) | **0.8138 (0.0245)** | 0.0198 (0.0120) | 0.0073 (0.0070) | 0.0070 (0.0070) | 0.0073 (0.0073) |
| **22. SGro** | 0.0071 (0.0069) | 0.0084 (0.0081) | **0.8280 (0.0238)** | 0.0073 (0.0069) | 0.0071 (0.0072) | 0.0070 (0.0068) |
| **23.SJal** | 0.0076 (0.0072) | 0.0073 (0.0073) | 0.0202 (0.0122) | **0.8163 (0.0241)** | 0.0068 (0.0070) | 0.0068 (0.0066) |

| Population ID | 20. SChis | 21.SOax | 22. SGro | 23.SJal | 24.SoSin | 25.SoSon |
|---|---|---|---|---|---|---|
| **24.SoSin** | 0.0087 (0.0084) | 0.0088 (0.0087) | 0.0087 (0.0084) | 0.0087 (0.0084) | **0.6753 (0.0084)** | 0.0088 (0.0085) |
| **25.SoSon** | 0.0085 (0.0082) | 0.0082 (0.0081) | 0.0082 (0.0081) | 0.0086 (0.0083) | 0.0084 (0.0082) | **0.6751 (0.0083)** |

**Supplementary Table S5.** Matrix of genetic differentiation among populations (pairwise $F_{ST}$) among populations and subspecies of *Cucurbita argyrosperma* obtained with Arlequin based on 12 nuclear microsatellite loci. The pairwise $F_{ST}$ matrix is presented in three parts: a) populations 1 – 10 of *C. argyrosperma* ssp. *argyrosperma*; b) populations 11 – 19 of *C. argyrosperma* ssp. *argyrosperma*; and c) populations 20 – 25 correspond to *C. argyrosperma* ssp. *sororia* (shaded in light gray). Population ID from Supplementary Table S1.

*a)* Populations 1 – 10 of *C. argyrosperma* ssp. *Argyrosperma*

| Population ID | 1. Chan | 2. Ek | 3. Mot | 4. Champ | 5.Pal | 6. Teh | 7. Mixt | 8.Ome | 9.Tla | 10.Mtp |
|---|---|---|---|---|---|---|---|---|---|---|
| **1. Chan** | 0 | | | | | | | | | |
| **2. Ek** | 0.02741 | 0 | | | | | | | | |
| **3. Mot** | 0.0592 | 0.03482 | 0 | | | | | | | |
| **4. Champ** | 0.02577 | 0.0494 | 0.08443 | 0 | | | | | | |
| **5.Pal** | 0.0926 | 0.03331 | 0.00885 | 0.08776 | 0 | | | | | |
| **6. Teh** | 0.17231 | 0.1687 | 0.10328 | 0.14972 | 0.06534 | 0 | | | | |
| **7. Mixt** | 0.05978 | 0.0012 | 0.04653 | 0.08369 | 0.0341 | 0.13485 | 0 | | | |
| **8.Ome** | 0.08978 | 0.05533 | 0.06342 | 0.04811 | 0.00681 | 0.05657 | 0.07044 | 0 | | |
| **9.Tla** | 0.35304 | 0.25026 | 0.31241 | 0.30027 | 0.16971 | 0.22946 | 0.27528 | 0.10875 | 0 | |
| **10.Mtp** | 0.09725 | 0.03096 | 0.09175 | 0.14515 | 0.05503 | 0.15308 | 0.0207 | 0.07152 | 0.22706 | 0 |
| **11.Sah** | 0.14006 | 0.17401 | 0.24273 | 0.08925 | 0.17739 | 0.18588 | 0.20095 | 0.07981 | 0.27858 | 0.21564 |
| **12.Aut** | 0.28378 | 0.16019 | 0.2522 | 0.25804 | 0.12347 | 0.22838 | 0.15126 | 0.12199 | 0.07888 | 0.14556 |
| **13.SJI** | 0.16032 | 0.03345 | 0.05536 | 0.21116 | 0.04814 | 0.14047 | 0.00814 | 0.14816 | 0.36232 | 0.0852 |
| **14.Tan** | 0.1242 | 0.03497 | 0.04902 | 0.13945 | 0.0075 | 0.12757 | 0.01113 | 0.06928 | 0.27529 | 0.07116 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **15.Tih** | 0.82554 | 0.78641 | 0.76164 | 0.83157 | 0.71457 | 0.73646 | 0.79548 | 0.76551 | 0.77942 | 0.78992 |
| **16.Dgo** | 0.30611 | 0.16352 | 0.11862 | 0.34433 | 0.02655 | 0.10667 | 0.16113 | 0.1507 | 0.31989 | 0.16051 |
| **17.CCC** | 0.35035 | 0.21154 | 0.13722 | 0.31787 | 0.03218 | 0.13116 | 0.24259 | 0.12782 | 0.27555 | 0.2543 |
| **18.SinalP** | 0.30243 | 0.09777 | 0.14611 | 0.27624 | 0.01099 | 0.14903 | 0.12108 | 0.05725 | 0.05024 | 0.10219 |
| **19. Yec** | 0.38352 | 0.34345 | 0.29131 | 0.37018 | 0.16951 | 0.0749 | 0.32098 | 0.147 | 0.17083 | 0.30615 |
| **20. SChis** | 0.85536 | 0.83268 | 0.77236 | 0.84621 | 0.6997 | 0.61162 | 0.84615 | 0.72858 | 0.80024 | 0.84191 |
| **21.SOax** | 0.79956 | 0.78146 | 0.72838 | 0.78967 | 0.66561 | 0.58479 | 0.78649 | 0.67619 | 0.72777 | 0.7793 |
| **22. SGro** | 0.22541 | 0.13578 | 0.07133 | 0.21882 | 0.02522 | 0.12885 | 0.14881 | 0.07901 | 0.26653 | 0.13635 |
| **23.SJal** | 0.66459 | 0.6214 | 0.54596 | 0.65173 | 0.45466 | 0.39877 | 0.62101 | 0.49116 | 0.54373 | 0.61042 |
| **24.SoSin** | 0.37773 | 0.21982 | 0.27705 | 0.34866 | 0.14276 | 0.28051 | 0.26515 | 0.15354 | 0.07521 | 0.2123 |
| **25.SoSon** | 0.23619 | 0.06986 | 0.14734 | 0.2423 | 0.05967 | 0.22991 | 0.08849 | 0.11763 | 0.2109 | 0.08823 |

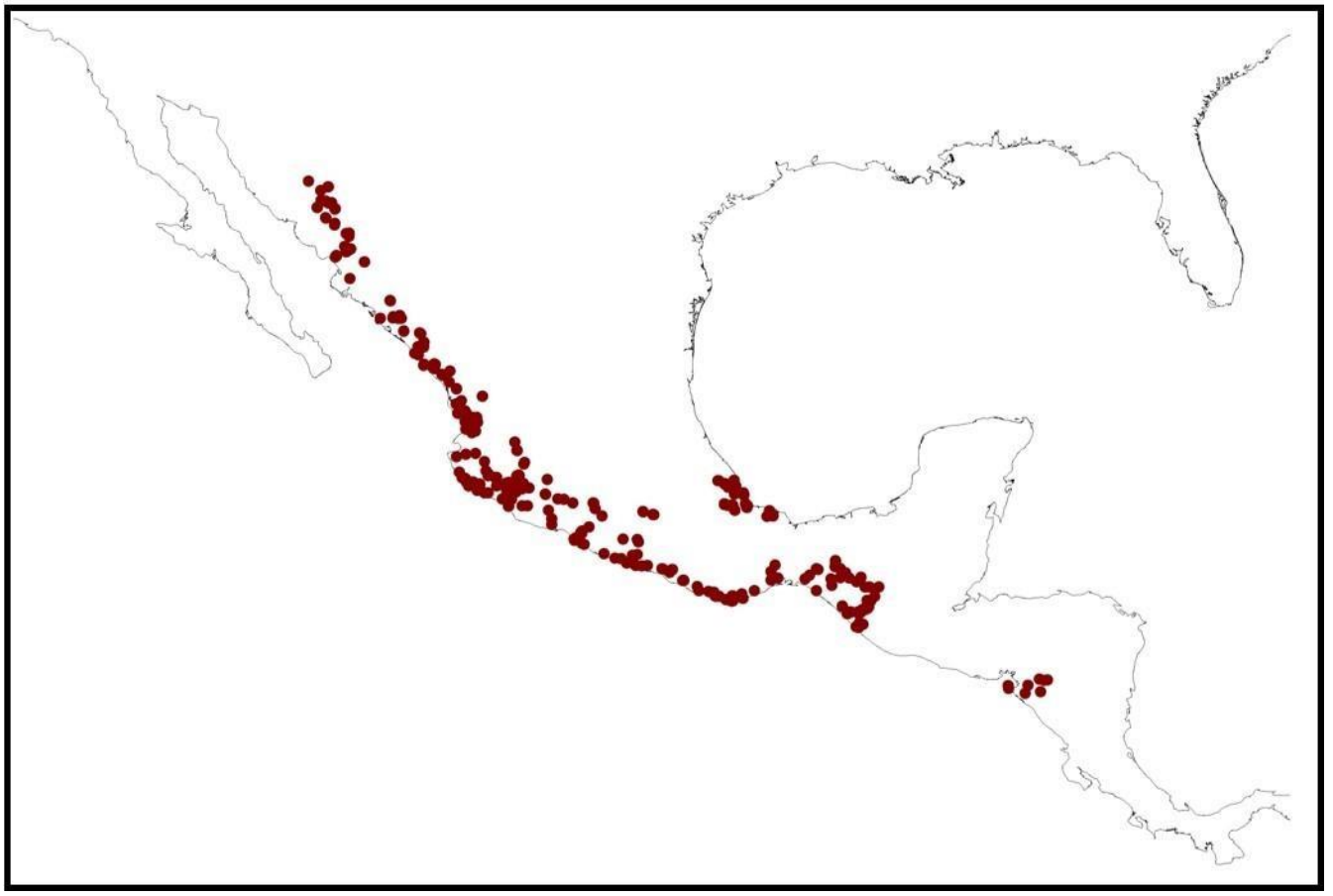
*b)* Populations 11 – 19 of *C. argyrosperma* ssp. *Argyrosperma*

| **Population ID** | **11.Sah** | **12.Aut** | **13.SJI** | **14.Tan** | **15.Tih** | **16.Dgo** | **17.CCC** | **18.SinalP** | **19. Yec** |
|---|---|---|---|---|---|---|---|---|---|
| **11.Sah** | 0 | | | | | | | | |
| **12.Aut** | 0.25626 | 0 | | | | | | | |
| **13.SJI** | 0.42804 | 0.28132 | 0 | | | | | | |
| **14.Tan** | 0.25444 | 0.13708 | 0.11845 | 0 | | | | | |
| **15.Tih** | 0.86769 | 0.7898 | 0.74086 | 0.79732 | 0 | | | | |
| **16.Dgo** | 0.46416 | 0.27599 | 0.22785 | 0.18008 | 0.59151 | 0 | | | |
| **17.CCC** | 0.45107 | 0.30276 | 0.33911 | 0.21569 | 0.68731 | 0.23085 | 0 | | |
| **18.SinalP** | 0.42557 | 0.02366 | 0.57814 | 0.10579 | 0.7563 | 0.25044 | 0.14034 | 0 | |
| **19. Yec** | 0.35721 | 0.26056 | 0.28004 | 0.28458 | 0.74306 | 0.14127 | 0.27448 | 0.15186 | 0 |
| **20. SChis** | 0.8793 | 0.84727 | 0.90702 | 0.84225 | 0.90218 | 0.83282 | 0.7193 | 0.88424 | 0.6082 |
| **21.SOax** | 0.8149 | 0.78817 | 0.78989 | 0.77844 | 0.86421 | 0.71462 | 0.67158 | 0.74916 | 0.5002 |
| **22. SGro** | 0.3515 | 0.26178 | 0.1763 | 0.13839 | 0.73825 | 0.1465 | 0.06933 | 0.09438 | 0.27937 |
| **23.SJal** | 0.69549 | 0.61771 | 0.60556 | 0.59612 | 0.75152 | 0.47444 | 0.37922 | 0.51322 | 0.3159 |
| **24.SoSin** | 0.43157 | 0.13943 | 0.4114 | 0.25751 | 0.77505 | 0.36876 | 0.22962 | -0.04492 | 0.32879 |
| **25.SoSon** | 0.36797 | 0.09552 | 0.23904 | 0.0776 | 0.78815 | 0.2746 | 0.25734 | -0.03218 | 0.33728 |

*c)* Populations 20 – 25 correspond to *C. argyrosperma* ssp. *sororia*

| POPULATION | 20. SChis | 21.SOax | 22. SGro | 23.SJal | 24.SoSin | 25.SoSon |
|---|---|---|---|---|---|---|
| **1. Chan** | | | | | | |
| **2. Ek** | | | | | | |
| **3. Mot** | | | | | | |
| **4. Champ** | | | | | | |
| **5.Pal** | | | | | | |
| **6. Teh** | | | | | | |
| **7. Mixt** | | | | | | |
| **8.Ome** | | | | | | |
| **9.Tla** | | | | | | |
| **10.Mtp** | | | | | | |
| **11.Sah** | | | | | | |
| **12.Aut** | | | | | | |
| **13.SJI** | | | | | | |
| **14.Tan** | | | | | | |
| **15.Tih** | | | | | | |
| **16.Dgo** | | | | | | |
| **17.CCC** | | | | | | |
| **18.SinalP** | | | | | | |
| **19. Yec** | | | | | | |
| **20. SChis** | 0 | | | | | |
| **21.SOax** | 0.06682 | 0 | | | | |
| **22. SGro** | 0.72271 | 0.68373 | 0 | | | |
| **23.SJal** | 0.21812 | 0.25969 | 0.4521 | 0 | | |
| **24.SoSin** | 0.85007 | 0.77283 | 0.19354 | 0.57232 | 0 | |
| **25.SoSon** | 0.87895 | 0.79829 | 0.15187 | 0.61453 | 0.57232 | 0 |

**Supplementary Figure S1.** Distribution of 273 occurrence points of *C. argyrosperma* ssp. *sororia* used for the species distribution model (SDM).



# CAPITULO 3

## ORIGEN DE LA DOMESTICACIÓN Y DIVERSIDAD GENÓMICA DE *Cucurbita argyrosperma* Huber EN MÉXICO.

**Artículo:**

The domestication of *Cucurbita argyrosperma* as revealed by the genome of its wild relative.

Publicado en 2021 en: *Horticulture Research*, 8(1), 1-14.

Horticulture Research
www.nature.com/hortres

ARTICLE Open Access

# The domestication of *Cucurbita argyrosperma* as revealed by the genome of its wild relative

Josué Barrera-Redondo[1], Guillermo Sánchez-de la Vega[1], Jonás A. Aguirre-Liguori[2], Gabriela Castellanos-Morales[3], Yocelyn T. Gutiérrez-Guerrero[1], Xitlali Aguirre-Dugua[1], Erika Aguirre-Planter[1], Maud I. Tenaillon[4], Rafael Lira-Saade[5] and Luis E. Eguiarte[1]

## Abstract

Despite their economic importance and well-characterized domestication syndrome, the genomic impact of domestication and the identification of variants underlying the domestication traits in *Cucurbita* species (pumpkins and squashes) is currently lacking. *Cucurbita argyrosperma*, also known as cushaw pumpkin or silver-seed gourd, is a Mexican crop consumed primarily for its seeds rather than fruit flesh. This makes it a good model to study *Cucurbita* domestication, as seeds were an essential component of early Mesoamerican diet and likely the first targets of human-guided selection in pumpkins and squashes. We obtained population-level data using tunable Genotype by Sequencing libraries for 192 individuals of the wild and domesticated subspecies of *C. argyrosperma* across Mexico. We also assembled the first high-quality wild *Cucurbita* genome. Comparative genomic analyses revealed several structural variants and presence/absence of genes related to domestication. Our results indicate a monophyletic origin of this domesticated crop in the lowlands of Jalisco. We found evidence of gene flow between the domesticated and wild subspecies, which likely alleviated the effects of the domestication bottleneck. We uncovered candidate domestication genes that are involved in the regulation of growth hormones, plant defense mechanisms, seed development, and germination. The presence of shared selected alleles with the closely related species *Cucurbita moschata* suggests domestication-related introgression between both taxa.

## Introduction

Domestication is an evolutionary process where human societies select, modify and eventually assume control over the reproduction of useful organisms. A mutualistic relationship emerges from this interaction, where humans exploit a particular resource of interest, while the domesticated organism benefits from increased fitness and extended geographical range[1,2]. This is well illustrated in *Cucurbita* L. (pumpkins, squashes, and some gourds), where human-guided domestication and breeding have considerably extended their distribution despite the extinction of their natural dispersers (e.g., mastodons and similar megafauna)[3]. Today, *Cucurbita* stand as successful crops grown and consumed worldwide, with a global annual production of ~24 million tons[4].

With ca. 21 taxa, the *Cucurbita* genus has experienced independent domestication events in five species[5,6]. Each *Cucurbita* crop experienced a unique selection for specific traits, predominantly defined by the nutritional and cultural needs of early human populations in America[7]. However, many domestication traits are common to domesticated *Cucurbita*, including the loss of bitter compounds (cucurbitacins), the loss of physical defense mechanisms (e.g., urticating trichomes), the loss of seed

Correspondence: Josué. Barrera-Redondo (josue_barrera@comunidad.unam.mx) or Rafael Lira-Saade (rlira@unam.mx) or Luis E. Eguiarte (fruns@unam.mx)
[1]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Circuito Exterior s/n Anexo al Jardín Botánico, 04510 Ciudad de México, México
[2]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA
Full list of author information is available at the end of the article
These authors contributed equally: Josué Barrera-Redondo, Guillermo Sánchez-de la Vega

dormancy, the enlargement of fruits and seeds, and the diversification of fruit morphology[4,8].

The initial steps of *Cucurbita* domestication were most likely directed towards seed rather than flesh consumption[9]. Seeds are rich in both carbohydrates and fatty acids, and cucurbitacins can be removed through boiling and washing; processes that are still employed for the consumption of wild *Cucurbita* seeds in Western Mexico[7]. Because the cultivation of *C. argyrosperma* (pipiana squash, cushaw pumpkin, or silver-seed gourd) is directed towards seed production rather than fruit flesh, it stands as an excellent model to investigate the early steps of *Cucurbita* domestication. *Cucurbita argyrosperma* subsp. *argyrosperma* (*argyrosperma* hereafter) was domesticated in Mesoamerica from its wild relative *Cucurbita argyrosperma* subsp. *sororia* (*sororia* hereafter), according to archaeological and genetic evidence[6,10,11]. *Argyrosperma* exhibits morphological differences from *sororia*, including larger fruits, larger seeds, and lack of urticating trichomes (Fig. 1). The earliest archaeological record of *argyrosperma* is presumed to be from 8700-year-old phytoliths in the Central Balsas Valley (Guerrero), although its taxonomic identity remains uncertain[10]. *C. argyrosperma* is a monoecious outcrossing species and gene flow has been previously described between the domesticated and wild subspecies[12]. Both subspecies are sympatric throughout the Pacific Coast of Mexico and Central America, with a few populations scattered in the coast of the Gulf of Mexico[11,13]. The domesticated taxon is also distributed in the Yucatan Peninsula, where its wild counterpart is absent[11].

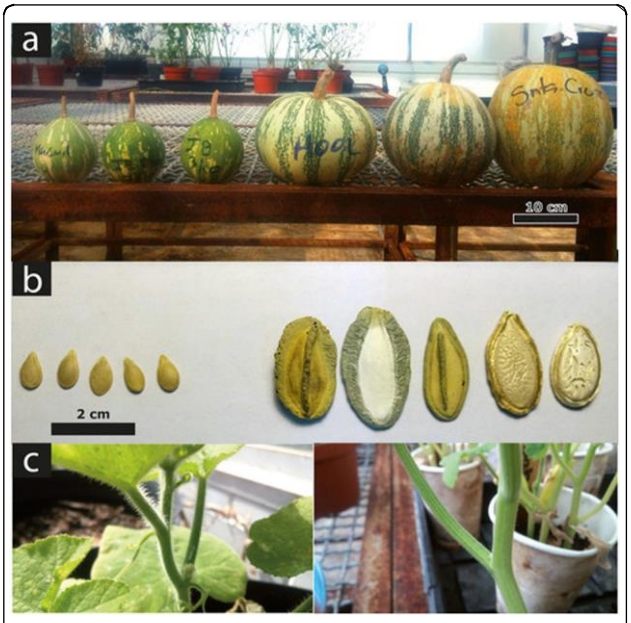



Fig. 1 Some morphological differences between *C. argyrosperma* subsp. *sororia* (left) and *C. argyrosperma* subsp. *argyrosperma* (right). Differences in a fruit size, b seed size and shape, c and in the presence of urticating trichomes

Despite the economic importance and the growing genomic resources in *Cucurbita* species[8,13], studies aimed at understanding their domestication are still lacking. To start filling this gap, we report here the first genome assembly of the wild relative *sororia*, which complements the existing assembly of *argyrosperma*[14]. The comparison between the genomes allowed us to find genomic structural variants between both subspecies. We characterized a large sample of *argyrosperma* landraces (117 individuals from 19 locations) and *sororia* accessions (50 individuals from 4 locations) using genome-wide data to investigate their demographic history and propose a domestication scenario. We also performed selection scans throughout the genome of *C. argyrosperma* to detect candidate regions associated with the domestication of this species.

## Results

### Genome assembly of *C. argyrosperma* subsp. *sororia*

We sequenced the genome of a wild individual of subspecies *sororia* using Illumina HiSeq4000 (213x coverage) and PacBio Sequel (75.4x coverage). The genome was assembled in 828 contigs with an N50 contig size of 1.3 Mbp and an L50 of 58 contigs (Table S1). A BUSCO analysis[15] against the *embryophyte odb9* database detected 92.8% of complete BUSCOs, 1.2% fragmented BUSCOs, and 6.0% missing BUSCOs within the genome assembly, similarly to other *Cucurbita* genome assemblies[14,16]. We predicted 30,592 protein-coding genes within the genome assembly using BRAKER2[17]. The BUSCO completeness of the gene predictions (92.5% complete BUSCOs, 3.1% fragmented BUSCOs) is comparable to that of the genome assembly and that of other genome annotations[14,16], despite using RNA-seq data of a different individual from which the genome was assembled (see "Methods"). Around 35.8% of our *sororia* genome assembly is composed of transposable elements (TEs), slightly higher than the 34.1% of TEs found in a previous *argyrosperma* assembly[14].

The genome of *argyrosperma* was previously assembled in 920 scaffolds[14], so we aimed at improving the assemblies for both the *sororia* and the *argyrosperma* genomes using a reference-guided scaffolding step against the genome assembly of *C. moschata*[16]. We anchored 99.97% of the *argyrosperma* genome assembly and 98.8% of the *sororia* genome assembly into 20 pseudomolecules using RaGOO[18], which corresponds to the haploid chromosome number in *Cucurbita*[19]. Both assemblies show high synteny conservation across the genus (Fig. S1) and confirm a previously reported inversion in chromosome four that is shared with *C. moschata*[16].

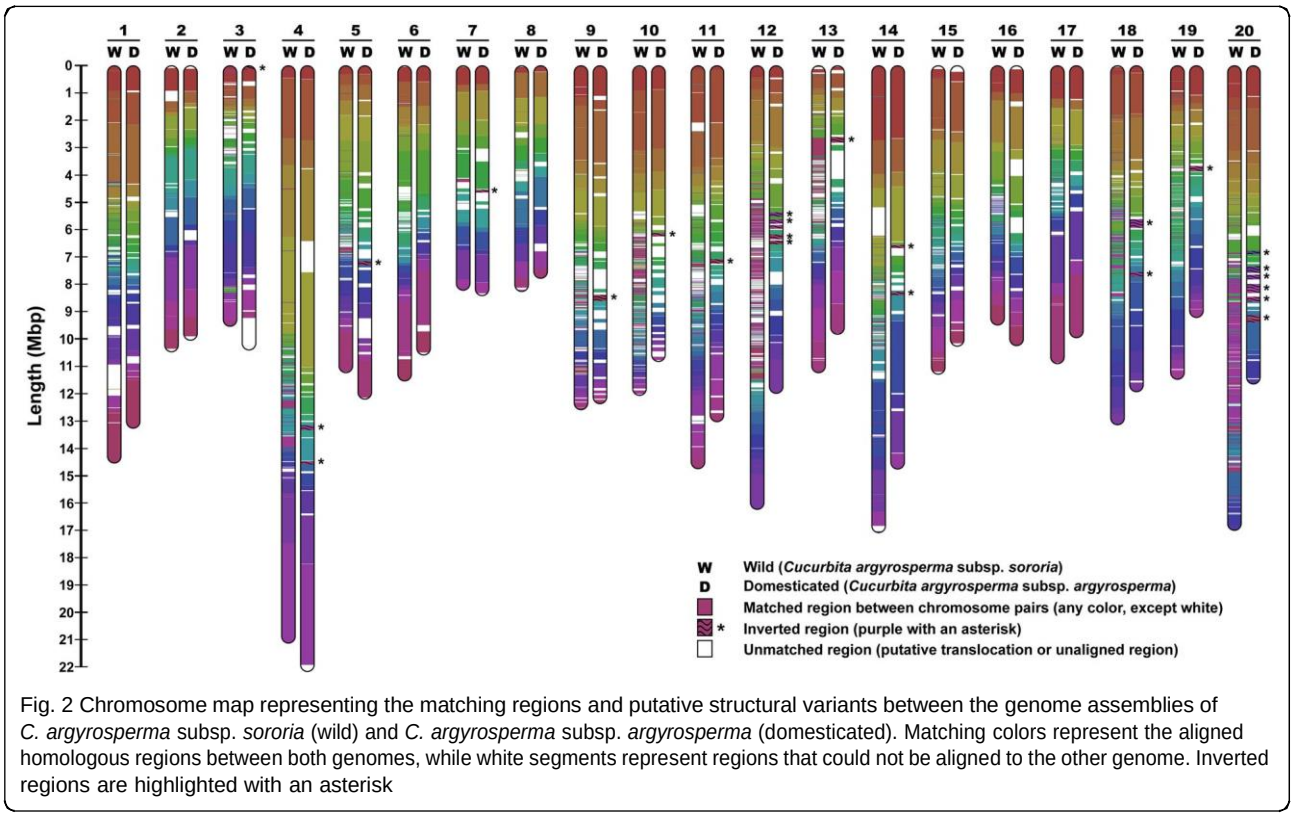



Fig. 2 Chromosome map representing the matching regions and putative structural variants between the genome assemblies of *C. argyrosperma* subsp. *sororia* (wild) and *C. argyrosperma* subsp. *argyrosperma* (domesticated). Matching colors represent the aligned homologous regions between both genomes, while white segments represent regions that could not be aligned to the other genome. Inverted regions are highlighted with an asterisk

## Structural variants between the wild and domesticated genomes

We compared the genome of *argyrosperma* against the genome of *sororia* (Fig. 2). Some of the centromeres in the genome of *sororia* were larger than in *argyrosperma*, possibly due to a better assembly of the repetitive regions. We found 443 high-confidence structural variants (SVs) such as copy-number variants (CNVs), inversions, and translocations between the wild and the domesticated genomes (Table S2). We also found several regions that could not be aligned between both genomes (Table S2). The size of the *sororia* genome assembly is ~254 Mbp, 9.23% larger than the genome assembly of *argyrosperma*[14], which could be partially explained by these SVs. We found that two copies of thaumatin-like protein 1a (*TL1*) and auxin-responsive protein *SAUR32* were disrupted by inversions in the *argyrosperma* genome. We also found a CNV loss spanning the transcription factor *PIF1* and a translocation containing the gene LONG AFTER FAR-RED 3 (*LAF3*). The genomes of *argyrosperma* and *sororia* share some common genes within their unaligned regions, such as microtubule-associated proteins and genes related to tryptophan biosynthesis (Table S2), suggesting that those regions contain highly divergent sequences and are not limited to presence/absence variants. However, other unaligned regions contain more genes in *sororia* than in *argyrosperma*, including the major pollen allergen Ole e 6 (*OLE6*), some proteolytic enzymes and sucrose biosynthetic genes that are absent in the *argyrosperma* genome (Table S2), suggesting that presence/absence variants are also included within the unaligned regions.

## Population data and SNP genotyping

We used samples previously collected throughout Mexico[11] corresponding to 117 individuals of *argyrosperma*, 50 individuals of *sororia*, 19 feral individuals of *argyrosperma* previously reported to have a semi-wild phenotype and a cultivated genotype based on microsatellite data[11] and 6 individuals of *C. moschata* (the domesticated sister species of *C. argyrosperma*), that were used as outgroups (Table S3). The samples were sequenced using the tunable Genotype by Sequencing (tGBS) method[20] to obtain genome-wide genetic information of the *C. argyrosperma* populations. The reads were quality-filtered and mapped against the genome assemblies of *argyrosperma* and *sororia* to predict single nucleotide polymorphisms (SNPs) and assess possible reference biases in the SNP prediction. Using the reference genome of *argyrosperma*, we obtained an initial dataset consisting of 12,813 biallelic SNPs with a mean read depth of 50 reads per SNP and a minor allele frequency (MAF) of at least 1% (13k dataset). We also mapped the whole-genome Illumina reads of

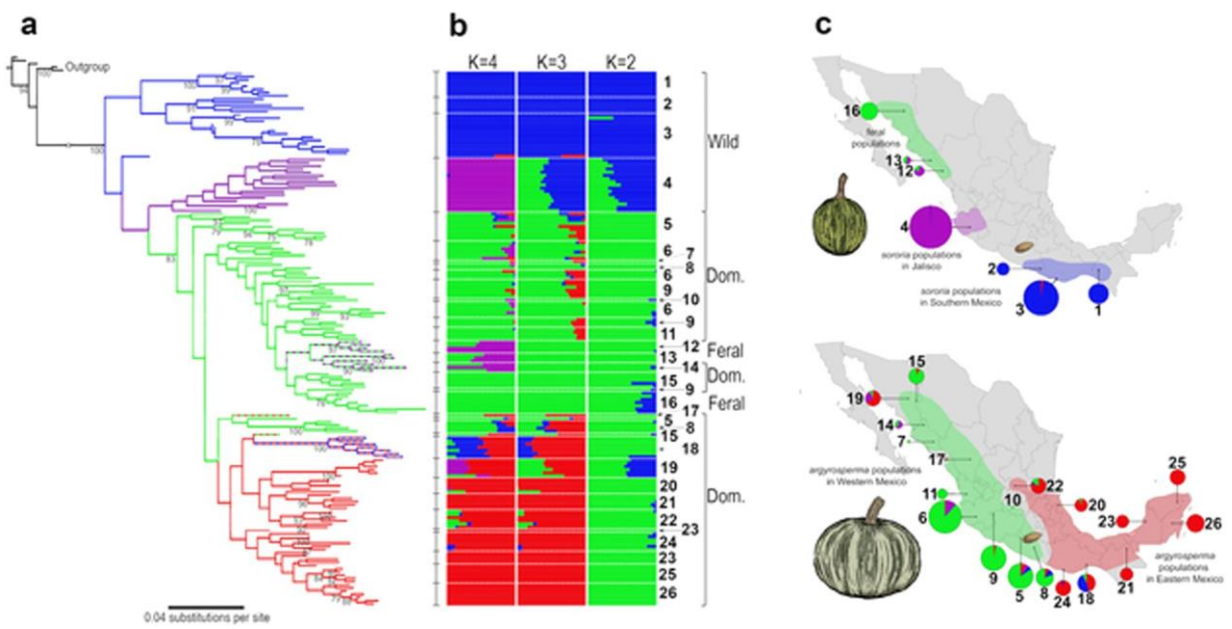



Fig. 3 Genetic structure and phylogenetic relationships between the wild and domesticated populations of *Cucurbita argyrosperma* based on 2861 SNPs. a Maximum-Likelihood tree with 100 bootstraps (only bootstrap values > 70 shown). b ADMIXTURE analysis using *K* values ranging from 2 to 4. c Geographic distribution of the wild (top) and domesticated (bottom) populations, with pie chart colors representing the ADMIXTURE assignation of the individuals in *K* = 4 genetic groups (size of pie charts proportional to sample size). The seed in the maps represents the earliest archaeological record of *argyrosperma* from Xihuatoxtla, Guerrero (dated 8700 years BP)[10]

*argyrosperma* and *sororia*, as well as the whole-genome sequencing of a *C. moschata* individual and a *C. okeechobeensis* subsp. *martinezii* individual (a closely related wild *Cucurbita* species), against the reference genome of *argyrosperma* to obtain a dataset of 11,498,421 oriented biallelic variants (SNPs and indels) across the genome that was used to assess introgression and incomplete lineage sorting.

### Demographic history of *C. argyrosperma* during its domestication

We eliminated the SNPs that deviated from Hardy–Weinberg equilibrium (exact test with $p < 0.01$) and pruned nearby SNPs under linkage disequilibrium (LD with an $r^2 > 0.25$ in 100 kbp sliding windows) from the 13k dataset to retrieve a set of 2861 independent SNPs that could be used for demographic analyses.

We found similar genetic variation in *sororia* and *argyrosperma*, regardless of the reference genome used (average nucleotide diversity $\pi$ range 0.095–0.098 for both taxa, see Table S4). At a population scale, the wild population in Jalisco had the highest genetic diversity within *sororia*, while the highest diversity in *argyrosperma* was found in the Pacific Coast of Mexico (Table S5). The domesticated and wild populations of *C. argyrosperma* displayed low genetic differentiation ($F_{ST}$ = 0.0646; 95% confidence interval from 0.0565 to 0.0751), while feral populations were more closely related to *argyrosperma* ($F_{ST}$ = 0.0479) than with *sororia* ($F_{ST}$ = 0.1006).

We used SNPhylo[21] and ADMIXTURE[22] to evaluate the genealogical relationships and genetic structure among the wild and domesticated populations of *C. argyrosperma*. We confirmed the genetic differentiation between *sororia* and *argyrosperma*, as detected by the $F_{ST}$ analyses. Our Maximum-Likelihood (ML) tree groups all the *argyrosperma* populations in a single monophyletic clade (Fig. 3a). We found additional genetic differentiation between the *sororia* populations in Southern Mexico (populations 1–3) and the *sororia* populations in Jalisco (population 4), in both the ADMIXTURE assignations (Fig. 3b) and their positions in the ML tree (Fig. 3a). The *sororia* populations of Jalisco are genetically closer to *argyrosperma*, as shown by their paraphyletic position in the ML tree (Fig. 3a). Consistent with a domestication in the lowlands of Western Mexico, the *argyrosperma* populations of Guerrero and Jalisco represent the basal branches of the *argyrosperma* clade (Fig. 3a), all showing instances of genetic similarity to the *sororia* populations in Jalisco in the four genetic groups (*K*) of ADMIXTURE (Fig. 3b). The *argyrosperma* populations in Western Mexico (populations 5–17) are genetically differentiated from the Eastern populations (populations 18–26), with a possible recent anthropogenic dispersion event of Eastern populations into Onavas, Sonora (population 19; Fig. 3b, c). These four

genetic groups are also retrieved in a principal component analysis (PCA; Fig. S2). The ADMIXTURE results (Fig. 3b) uncovered admixture events between some *sororia* and *argyrosperma* populations. This pattern is evident in the populations of Oaxaca and Sinaloa (populations 14 and 18; Fig. 3c). The feral populations are consistently grouped alongside their sympatric domesticated populations within *argyrosperma* (Fig. 3a, b), indicating that these populations diverged recently from nearby domesticated populations. These demographic patterns are robust to different SNP filters, such as different thresholds for missing data, or filtering for significant deviations from Hardy–Weinberg equilibrium (Fig. S3).

We used Fastsimcoal 2[23] to explicitly test whether *argyrosperma* was domesticated from a *sororia* population in Southern Mexico or from a *sororia* population in Jalisco. Given that gene flow has been previously observed between *argyrosperma* and *sororia*[12], we compared three different gene flow models (continuous gene flow, secondary contact, or no gene flow) for each scenario (Fig. 4a). A comparison between models using the Akaike Information Criterion indicates that the Jalisco domestication model with secondary contact (*i.e.*, extant gene flow after initial genetic isolation between subspecies) is the most likely of the domestication scenarios assayed. We were able to discard the other unrealistic domestication scenarios (i.e., domestication in southern Mexico and lack of gene flow between subspecies), further supporting a domestication event in Jalisco.

### Selection scans in *C. argyrosperma*

In order to perform the tests to detect selective signals associated with the domestication of *C. argyrosperma*, we removed from the 13k dataset the *C. moschata* individuals as well as the feral individuals of *C. argyrosperma*. We used the 1% MAF threshold for this subset, obtaining a 10,617 SNP dataset suitable to detect selective signals, with a marker density of 44 SNPs per Mb. LD was limited within the dataset, with a mean pairwise $r^2$ of 0.1, and with *argyrosperma* showing a faster LD decay than *sororia* (Fig. S4). We performed three outlier tests as implemented in BayeScEnv[24], PCAdapt[25], and LFMM 2[26,27] to detect selective signals between the domesticated and the wild populations of *C. argyrosperma* (Fig. 5). BayeScEnv is an $F_{ST}$-based method that tests correlation with other variables, in our case the wild or domesticated nature of each population (coded as 0 and 1, respectively). PCAdapt does not require an a priori grouping of individuals into wild/domesticated, as we used the two principal components of a PCA to control for the underlying genetic structure between subspecies (Fig. S2). LFMM 2 identifies the number of latent factors in the populations through least square estimates to find correlations between genetic variants and the domesticated or wild phenotypes in the

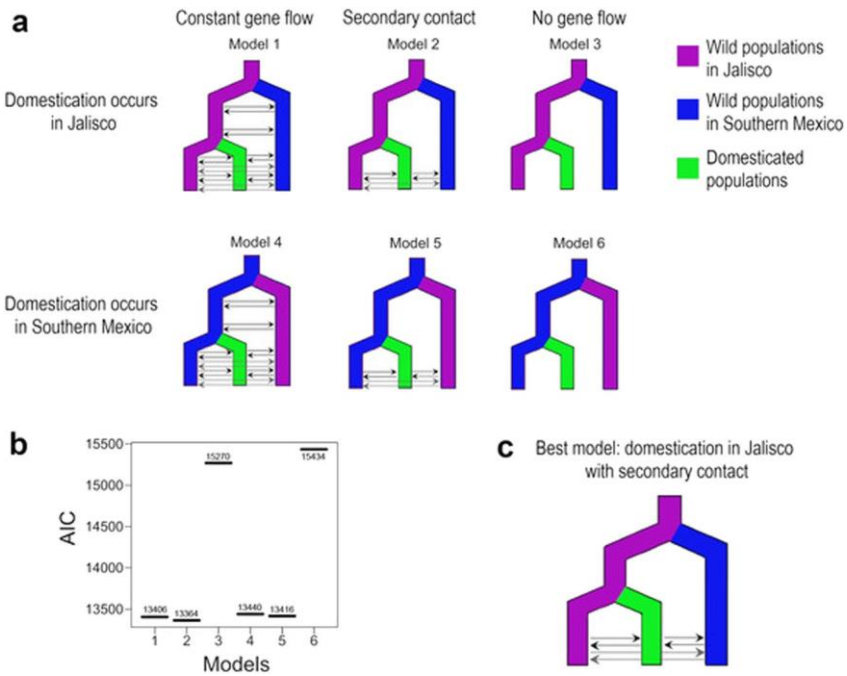



Fig. 4 Coalescent simulations and most likely domestication scenario of *Cucurbita argyrosperma*. a The six models assessed against the unfolded multidimensional Site Frequency Spectrum of our data. b Comparison of the Akaike Information Criterion (AIC) of all the models. c The domestication model that best fits the data

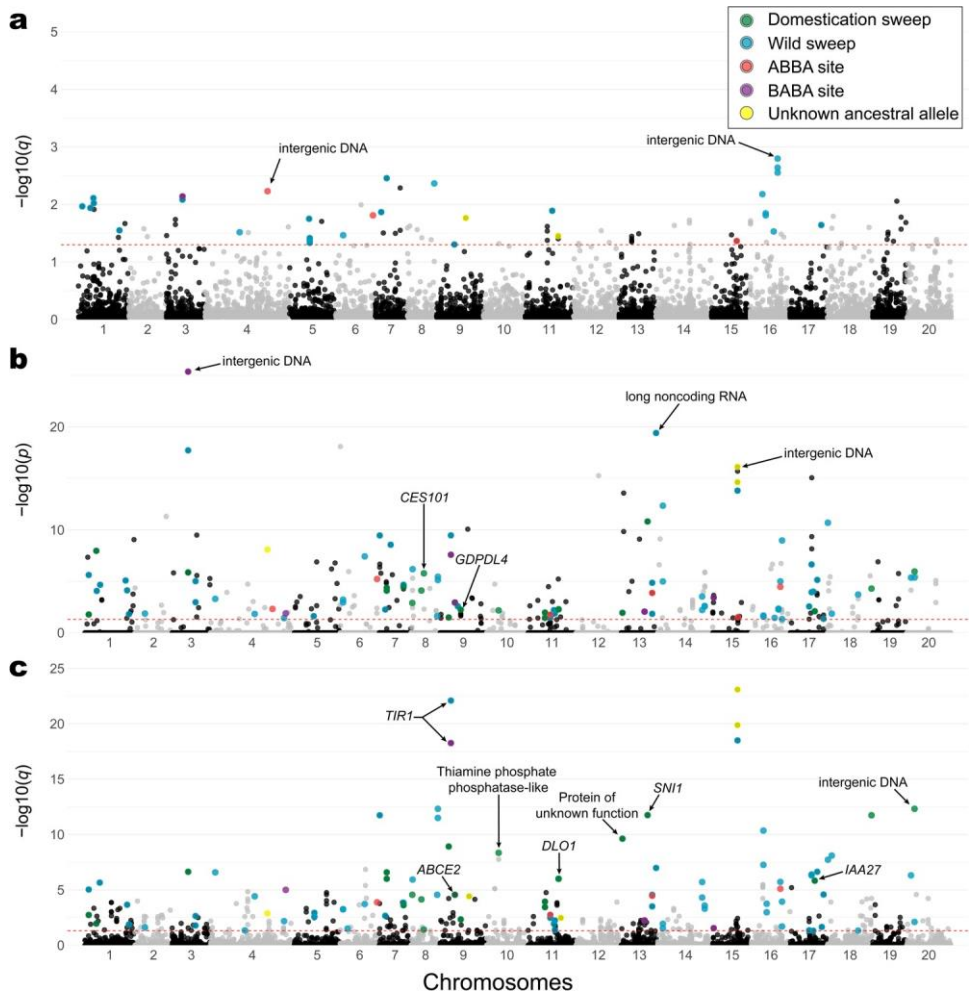



Fig. 5 Putative footprints of selection associated with the domestication of *Cucurbita argyrosperma*. Manhattan plots representing the a BayeScEnv, b PCAdapt, and c LFMM 2 tests in each chromosome of the genome. The red dotted line indicates the cutoff value (*q*-value or Bonferroni-corrected *p*-value < 0.05) to determine a candidate locus. Only 110 loci were retrieved as candidate SNPs by more than one test, which are highlighted depending on whether the putative signal of positive selection corresponds to *argyrosperma* (green), *sororia* (blue), an ABBA site (red) a BABA site (purple) or an unknown selective direction (yellow). The arrows indicate the position of some of the discovered candidate genes

dataset[26]. In order to reduce the number of false positives, we only analyzed the SNPs that were detected as outliers by at least two of the tests.

We discovered 110 SNPs that converged as candidates in at least two tests (Fig. 5). We used *C. moschata* and *C. okeechobeensis* subsp. *martinezii* as outgroups to determine the direction of the putative selective pressures for each candidate SNP, as well as determining possible introgression or incomplete lineage sorting between *sororia*, *argyrosperma* and *C. moschata*. We found that 22 of the candidate SNPs corresponded to selective signals in *argyrosperma*, while 70 candidate SNPs corresponded to selective signals in *sororia* (Fig. 5). We could not determine the direction of selection for 5 candidate SNPs, and 13 showed signals of introgression or incomplete lineage sorting.

We identified several instances of either genetic introgression or incomplete lineage sorting between *C. moschata* and both *argyrosperma* (ABBA sites) or *sororia* (BABA sites) (Fig. S5). We performed a genome-wide *D*-statistic analysis and found significantly more instances of shared derived variants between *argyrosperma* and *C. moschata* than between *sororia* and *C. moschata* (block-jackknife *p*-value = 0.0014), obtaining an overall admixture fraction $f_G$ of 0.01 (Table S6). From the 13 candidate SNPs with signals of introgression, 6 correspond to ABBA sites and 7 correspond to BABA sites (Fig. 5).

We found 45 protein-coding genes and one long-noncoding RNA including candidate SNPs within their structure (i.e., introns, exons, UTRs), which were assigned according to the observed direction of the putative

selective signals (Table S7). Among the genes under putative selection in *sororia* were four serine/threonine-protein kinases, including *PBL10* and *PBL23* (Table S7). Among the genes under putative selection in *argyrosperma*, we found glycerophosphodiester phosphodiesterase *GDPDL4*, auxin-responsive protein *IAA27*, ABC transporter E family member 2 (*ABCE2*), *DMR6*-like oxygenase 1 (*DLO1*), and serine/threonine-protein kinase *CES101* (Table S7). We also found several genes under putative selection overlapping ABBA and BABA sites (Table S7). We found a transport inhibitor response 1 (*TIR1*) homolog under putative selection in both *sororia* and as a BABA site. Curiously, the gene *MKP1* was found under selection in both *argyrosperma* and in *sororia*.

## Discussion

The genome assembly of *sororia* represents the first high-quality sequenced genome of a wild *Cucurbita*, which allowed us to detect structural and functional differences with the *argyrosperma* genome. The genome assembly of *argyrosperma* was smaller than the *sororia* assembly, which is possibly caused by the loss of structural variants during its domestication, as has been reported in pan-genome studies[28]. Many of these unaligned regions contain entire genes in *sororia*, making this wild taxon a reservoir of potentially adaptive presence/absence variants. However, the extant genetic diversity of *argyrosperma* is similar to that of *sororia*, which suggests that the effects of the domestication bottleneck were alleviated by the current gene flow between both subspecies as suggested by our coalescent simulations and by the results of the previous studies[12]. The fast LD decay in *argyrosperma* further supports the limited effect of the domestication bottleneck. This gene flow may be related to the sympatric distribution of the wild and domesticated populations of *C. argyrosperma* throughout the Pacific Coast of Mexico[11], where their coevolved pollinator bees *Peponapsis* spp. and *Xenoglossa* spp. are found[29]. Traditional agricultural practices are another fundamental force that maintains the diversity of crop species[30]. Since *argyrosperma* is a traditional crop cultivated for both self-supply and local markets where it has a specialized gastronomic niche[13], the genetic diversity in *argyrosperma* is also maintained by the conservation of local landrace varieties at local scales[31,32].

Our demographic analyses suggest that the extant populations from Jalisco are the closest modern relatives of the initial population of *sororia* from which *argyrosperma* originated. The genetic relatedness between the *sororia* populations from Western Mexico and *argyrosperma* was also observed with mitochondrial markers[6]. The domestication of *C. argyrosperma* likely started around 8,700 years ago, as suggested by the earliest, albeit taxonomically ambiguous, archaeological record of *argyrosperma*[10,33]. Since crop domestication in Mesoamerica is linked to migration patterns and cultural development of early human populations in America[7,34], we expected *argyrosperma* to share historical demography patterns with human history. Our data shows that the *argyrosperma* populations found in Guerrero and Jalisco are the most closely related to the *sororia* populations from Jalisco. This means that even if the closest wild relatives of *argyrosperma* are currently found in Jalisco, the domestication process may have occurred throughout the lowlands of Jalisco and the Balsas basin[10]. Ancient human migration events have been proposed to occur along the river basins in Southwestern Mexico, which may explain the genetic cohesiveness among the *argyrosperma* populations of that area that represent the first fully domesticated lineage of the species[34]. Previous studies based on 8,700 years old phytoliths found in Xihuatoxtla, Guerrero, suggest the co-occurrence of *Zea mays* and *C. argyrosperma* in the Balsas region within this time period[10,33]. Overall, the patterns of genetic structure for *C. argyrosperma* are coherent with the archaeological evidence of early human migration throughout Mesoamerica[35,36] (Fig. 6).

We found several signals of putative selection between *argyrosperma* and *sororia*, even though tGBS sequencing has a limited capacity to detect selective signals across the genome[37]. The SNP density for our selection tests was of 44 SNPs per Mb, which is one order of magnitude denser than other studies using reduced-representation genome sequencing to detect selective signals[37]. This is a consequence of the relatively small genome size of *C. argyrosperma*[14]. Nonetheless, the LD in *C. argyrosperma* decays at a shorter length than our SNP density, so several signals of selection might be missing from our scans, and some signals might not correspond to the actual gene under selection, but rather to a neighboring region in the genome[37]. Therefore, our genome scans should be interpreted as a partial representation of the selective signals associated with the domestication process[37].

Most of the putative selective signals were attributed to *sororia*, probably because wild taxa are subject to many natural selective pressures. Many of the SNPs that were retrieved as outliers were found on genes involved in biotic and abiotic plant defense responses. For example, *PBL10* and *PBL23* have been suggested to be involved in plant defense pathways due to their similarity to other serine/threonine-protein kinases[38].

However, we also found some genes under putative selection in *argyrosperma* that could be attributed to defense mechanisms such as *DLO1*, which is involved in pathogenic defense responses[39]. We found *GDPDL4* under putative selection, which is involved in trichome differentiation[40], a morphological characteristic that differentiates *argyrosperma* from *sororia* (Fig. 1). We also

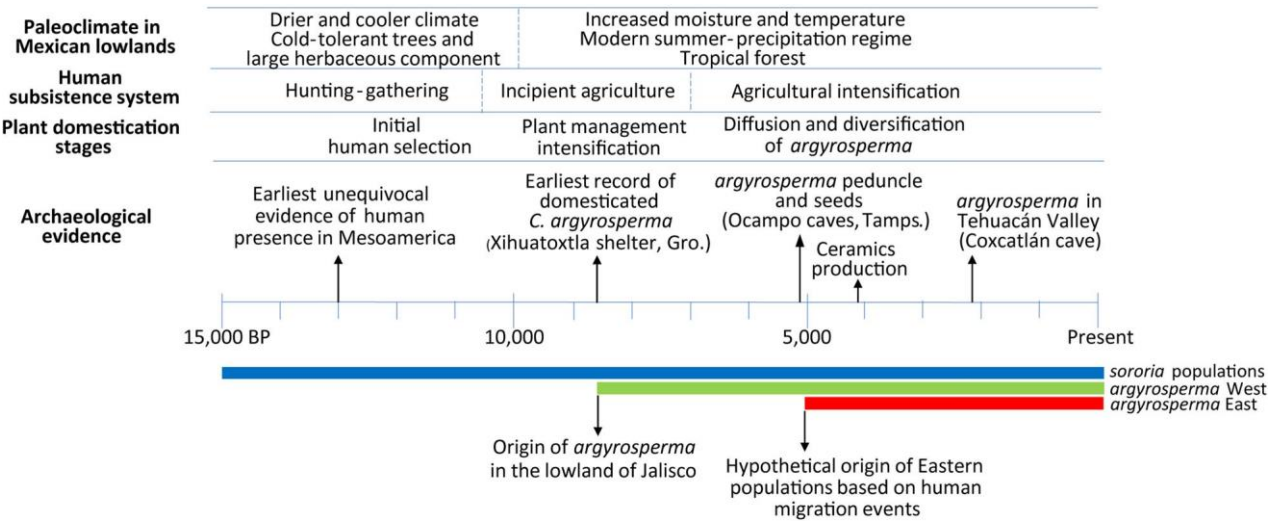



Fig. 6 Model of *Cucurbita argyrosperma* domestication based on archaeological[6,33–35] and genetic data. The colored lines represent the presence of each *C. argyrosperma* genetic cluster in Mesoamerica (blue = *sororia* populations, green = *argyrosperma* populations in Western Mexico, red = *argyrosperma* populations in Eastern Mexico)

found two disrupted copies of *TL1* and the absence of *OLE6* within the *argyrosperma* genome, both of which cause allergic reactions in humans[41,42]. This is concordant with previous studies showing that selective pressures during domestication actively purge these defense mechanisms, as the products of these responses are usually unpleasant or harmful to humans when the plant is consumed[43]. This is particularly important for breeding programs, since wild *Cucurbita* such as *sororia* harbor loci associated with disease resistance that their domesticated counterparts have lost[4]. The selective pressures found in *MKP1* suggest a disruptive selection regime between *sororia* and *argyrosperma*. Since *MKP1* modulates defense responses[44], it is possible that both subspecies adapted to differential environmental pressures as domestication took place.

We found several candidate genes involved in the regulation of ABA. The alteration of growth hormones may play an important role in *C. argyrosperma* domestication. ABA is involved in a myriad of functions, such as the regulation of plant growth, plant development, seed dormancy, and response to biotic/abiotic stress[45]. In this sense, the lack of dormancy in seeds and gigantism are both common domestication changes that are present in domesticated cucurbits that may be caused by changes in the regulation of ABA and brassinosteroids[8,46]. We found *SAUR32*, *PIF1*, and *LAF3* within the high-confidence SVs in *argyrosperma*, all of which act as inhibitors of seed germination under dark conditions[47–49] and may explain the lack of seed dormancy in *argyrosperma*. Likewise, we found *IAA27* under selection in *argyrosperma*, which is involved in plant growth and development[50].

Selection over seed size has been particularly important in the domestication of *C. argyrosperma*[9]. We found *ABCE2* under selection in *argyrosperma*. Some ABC transporters are involved in the transmembrane transport of ABA-GE, an ABA conjugate that is usually attributed to the plant response against water stress[51]. However, previous studies in *Hordeum vulgare* suggest that the transport of ABA-GE may play a role in seed development alongside *de novo* ABA synthesis within the developing seed[52], suggesting a role of ABC transporters in the seed development of *C. argyrosperma*. Previous studies have also identified an association between variants in ABC transporter proteins and seed size in *Cucurbita maxima*[53] and *Linum usitatissimum*[54], further suggesting that ABA may be deregulated via selective pressures on ABC transporters to enhance seed size in *C. argyrosperma* during its domestication. Variants in a serine/threonine-protein kinase, as the one we found in our selective scans, have also been associated with seed size in *C. maxima*[53].

We found shared derived variants between *C. moschata* and *C. argyrosperma* under putative selection. This suggests that, given the close relationship between *C. argyrosperma* and *C. moschata*, both species may share domestication loci involved in common domestication traits. However, we also found *TIR1* under selection in *sororia* and as a BABA site, which is an auxin receptor involved in ethylene signaling and antibacterial resistance in roots[55]. The introgression of wild alleles into domesticated *Cucurbita* has been previously reported as an effective method to improve the resistance of domesticated crops[56]. Along this line, our results suggest that *sororia* is an important source of adaptive alleles for

*C. moschata*. ABBA and BABA sites under selection may be shared with *C. moschata* either due to incomplete lineage sorting or by adaptive introgression with the wild and domesticated populations of *C. argyrosperma*. The incorporation of domesticated loci between *C. argyrosperma* and *C. moschata* through introgression may have been an effective way for Mesoamerican cultures to domesticate multiple *Cucurbita* taxa. This hypothesis is supported by the significant amount of ABBA sites shared between the genomes of *C. argyrosperma* and *C. moschata*. However, this hypothesis needs to be further addressed using population-level data of *C. moschata* with other wild and domesticated *Cucurbita* species.

## Materials and methods

### Genome assembly and annotation of *Cucurbita argyrosperma* subsp. *sororia*

We sequenced and assembled *de novo* the genome of a *sororia* individual collected in Puerto Escondido (Oaxaca, Mexico). Its DNA was extracted from leaf tissue and sequenced using PacBio Sequel at the University of Washington PacBio Sequencing Services and using Illumina HiSeq4000 at the Vincent J. Coates Genomics Sequencing Laboratory in UC Berkeley (NIH S10 Instrumentation Grants S10RR029668 and S10RR027303). We filtered the Illumina sequences using the qualityControl.py script (https://github.com/Czh3/NGSTools) to retain the reads with a PHRED quality ≥ 30 in 85% of the sequence and an average PHRED quality ≥ 25. The Illumina adapters were removed using SeqPrep (https://github.com/jstjohn/SeqPrep) and the paired reads that showed overlap were merged. The chloroplast genome was assembled using NOVOplasty[57] and the organellar reads were filtered using Hisat2[58] against the chloroplast genome of *argyrosperma*[14] and the mitochondrial genome of *C. pepo*[59]. We assembled the nuclear genome into small contigs using the Illumina reads and the Platanus assembler[60]. The Platanus contigs were assembled into larger contigs using the PacBio Sequel reads and DBG2OLC[61]. We performed two iterations of minimap2 and racon[62] to obtain a consensus genome assembly by mapping the PacBio reads and the Platanus contigs against the DBG2OLC backbone. We performed three additional polishing steps using PILON[63] by mapping the Illumina reads against the consensus genome assembly with BWA *mem*[64].

The genome annotation processes were performed using the GenSAS v6.0 online platform[65]. The transposable elements within the genome were predicted and masked using RepeatModeler (http://www.repeatmasker.org/RepeatModeler/). We downloaded five RNA-seq libraries of *C. argyrosperma* available on the Sequence Read Archive (accessions SRR7685400, SRR7685404–SRR7685407) to use them as RNA-seq evidence for the gene prediction. We performed the same quality filters described above for the RNA-seq data and aligned the high-quality reads against the masked genome of *C. argyrosperma* subsp. *sororia* using STAR v2.7[66]. We used filterBAM from the Augustus repository[67] to filter low-quality alignments and used the remaining alignments as RNA-seq evidence to predict the gene models using BRAKER2[17]. The gene predictions were functionally annotated using InterProScan[68] and by aligning the gene models against the SwissProt database[69] using BLASTp[70] with an *e*-value < $1e^{-6}$. We performed a manual assessment of the predicted gene models to eliminate annotation artifacts.

### Anchoring the reference genomes into pseudomolecules

We aimed to improve the genome assemblies of *argyrosperma* and *sororia*, which were assembled in 920 scaffolds[14] and 817 contigs, respectively. Thus, we generated PacBio corrected reads from the published PacBio RSII reads of *argyrosperma* (NCBI SRA accession SRR7685401) and the PacBio Sequel reads of *sororia* (sequenced for this study at the University of Washington PacBio Sequencing Services) using CANU[71]. The macrosynteny of *Cucurbita* genomes is largely conserved between species[72], so we performed a reference-guided scaffolding step using RaGOO[18] to anchor the genome assemblies of *argyrosperma*[14] and *sororia* into pseudomolecules. We used the genome assembly of *C. moschata*[16] as the reference genome for RaGOO[18] and we used the PacBio corrected reads of each taxon to detect and correct misassemblies, using a gap size of 2600 bp for chromosome padding (i.e., we filled the gaps between contigs with 2600 stretches of Ns, corresponding to the average gap length of the *argyrosperma* genome assembly). The chromosome numbers in both assemblies were assigned in correspondence to the genome assembly of *C. moschata*[16].

### Structural variant analysis

We evaluated the synteny between *Cucurbita* genomes using Synmap2[73]. We predicted the SVs between *argyrosperma* and *sororia* using SyRI[74] alongside nucmer[75] with a minimum cluster length of 500 bp, an alignment extension length of 500 bp, a minimum match length of 100 bp and a minimum alignment identity of 90%. We also used Sniffles alongside NGMLR[76] as an additional SV predictor, by aligning the *argyrosperma* PacBio reads against the genome assembly of *sororia* (only the SVs with a minimum support of 6 reads were retained). We only analyzed the SVs that were independently predicted by SyRI and Sniffles. Only the SVs that overlapped within a range of ±100 bp at the start and end positions of each prediction were retained as high-confidence SVs. Due to the limitations of read-mapping techniques to predict presence–absence variants, we only analyzed the unaligned regions predicted by SyRI for this type of variants.

The gene content associated with each type of structural variant was considered either as the overlap between genes and variants (inversions and translocations) or as the genes contained entirely within the structural variants (CNVs and unaligned regions). We performed a Gene Ontology enrichment analysis using topGO and the *weight01* algorithm[77] to find enriched biological functions associated with each type of structural variant. We determined the significantly enriched biological functions by performing Fisher's exact test ($p$-value < 0.05). We plotted the genome rearrangements and SVs between *sororia* and *argyrosperma* using Smash[78] with a minimum block size of 100,000 bp, a threshold of 1.9, and a context of 28.

### Data filtering and SNP genotyping

We used previously collected seeds from 19 populations of *argyrosperma* landraces, four populations of *sororia*, and three feral populations[11], covering most of the reported distribution of this species throughout Mexico[5] (Table S3). Each of the collected seeds came from a different maternal plant, in order to avoid signals of inbreeding. The seeds were germinated in a greenhouse and total DNA was extracted from fresh leaves using a DNeasy Plant MiniKit (Qiagen) of 192 individuals across the collected populations (Table S3), including five individuals of *C. moschata* to be used as outgroup. All 192 individuals were sequenced by Data2Bio LLC using the tunable Genotyping by Sequencing (tGBS) method[20] with an Ion Proton instrument and two restriction enzymes (Sau3AI/BfuCI and NspI). The wild and domesticated populations were randomly assigned to the plate wells before library preparation to avoid sequencing biases.

The raw reads of the tGBS sequencing were trimmed using LUCY2[79], removing bases with PHRED quality scores < 15 using overlapping sliding windows of 10 bp. Trimmed reads shorter than 30 bp were discarded. The trimmed reads were mapped against the genome assembly of *argyrosperma* using segemehl[80], since empirical studies suggest this read-mapping software outperforms others for Ion Torrent reads[81]. We only retained the reads that mapped uniquely to one site of the reference genome for subsequent analyses.

We used BCFtools[82] for an initial variant calling step, retaining variants with at least 6 mapped reads per individual per site where the reads had a minimum PHRED quality score of 20 in the called base and a minimum mapping quality score of 20[83]. We used plink[84] to perform additional filters, such as retaining only biallelic SNPs, retaining SNPs with no more than 50% of missing data, individuals with no more than 50% of missing data and sites with a minor allele frequency (MAF) of at least 1% (13k dataset). After eliminating individuals with missing data, 109 individuals of *argyrosperma*, 44 individuals of *sororia*, 14 feral individuals and 5 individuals of *C. moschata* remained for the subsequent analyses. We repeated the SNP prediction using the reference genome of *sororia* to evaluate potential reference biases. We found a similar number of SNPs (10,990) and comparable estimates of genetic diversity (see Table S4), suggesting that reference bias does not have a meaningful impact on our results. Thus, we employed the domesticated genome as the reference for the rest of the population analyses. We repeated our analyses using a separate filtering step of missing data for the domesticated and the wild populations, retrieving 84.18% of the SNPs from the 13k dataset and obtaining the same results (Fig. S3).

In order to obtain an adequate SNP dataset to infer the demographic history of *C. argyrosperma*, we performed additional filters to the 13k dataset with plink[84], including (i) the elimination of all the SNPs that diverged significantly ($p$ < 0.01) from the Hardy–Weinberg equilibrium exact test[85] to remove potential allelic dropouts, (ii) the elimination of adjacent SNPs with a squared correlation coefficient ($r^2$) larger than 0.25 within 100 kbp sliding windows with a step size of 100 bp. We repeated the demographic analyses without filtering the SNPs with significant deviations from Hardy–Weinberg equilibrium, obtaining the same results (Fig. S3).

We also generated a SNP dataset to detect selective signals associated with the domestication of *C. argyrosperma* by eliminating all the feral individuals of *C. argyrosperma*, which could not be assigned to either a wild or a domesticated population, as well as the five individuals of *C. moschata*. We also eliminated the SNP sites with more than 50% missing data and performed a MAF filter of 1% after reducing the number of individuals in the 13k dataset. The SNP density was calculated with VCFtools[86] and the LD decay was calculated using plink[84] with a minimum $r^2$ threshold of 0.001.

We also sequenced the genome of a *C. moschata* individual from Chiapas (Mexico) and the genome of a *C. okeechobeensis* subsp. *martinezii* individual from Coatepec (Veracruz, Mexico) using the Illumina HiSeq4000 platform in UC Berkeley, to evaluate possible introgression and incomplete lineage sorting with *C. argyrosperma*. We downloaded the genome sequences of *argyrosperma*[14] from the Sequence Read Archive (accessions SRR7685402 and SRR7685403). The Illumina whole-genome sequences were filtered using the same quality parameters as the ones used in the genome assembly of *sororia* (see above) and were aligned against the genome assembly of *argyrosperma* using BWA *mem*[64]. We only retained the reads that mapped uniquely to one site of the reference genome and retained only the biallelic sites with a sequencing depth >=10 reads per genome.

### Population structure

We used diveRsity[87] to calculate the pairwise $F_{ST}$ statistics, using 100 bootstraps to calculate the 95% confidence intervals. We estimated the genetic variation in the wild, domesticated and feral populations with STACKS[88]. Using ADMIXTURE[22], we evaluated the genetic structure among the *sororia* and *argyrosperma* populations, evaluating their individual assignment into one (CV error = 0.26205), two (CV error = 0.25587), three (CV error = 0.25806) and four (CV error = 0.26658) genetic groups (*K*). We reconstructed a maximum-likelihood tree with SNPhylo[21], based on substitutions per site between all the individuals with 100 bootstraps to assess the reliability of the tree topology. We used plink[84] to perform a principal component analysis (PCA) using 10 principal components.

### Coalescent simulations

We used Fastsimcoal 2[23] to determine the parameters that maximize the composite likelihood of each model given the unfolded multidimensional SFS. The unfolded multidimensional SFS was obtained with DADI[89], using 17 *sororia* individuals of Jalisco, 27 *sororia* individuals of Southern Mexico, 109 *argyrosperma* individuals and 5 *C. moschata* individuals as an outgroup to unfold the SFS. We ran 100,000 simulations with 20 replicates for each model (two divergence scenarios and three gene flow scenarios) using the following settings: a parameter estimation by Maximum Likelihood with a stopping criterion of 0.001 difference between runs, a minimum SFS count of 1, a maximum of 40 loops to estimate the SFS parameters and a maximum of 200,000 simulations to estimate the SFS parameters. We also selected log-uniform priors for parameter estimations, setting times of divergence between 1000 and 200,000 generations (domestication times are expected to fall within this interval, given that the presumed most ancient evidence of the human presence in America is 33,000 years old[90]; while the split between the wild populations is expected to coincide with the extinction of the megafauna in America, which acted as the natural dispersers of wild *Cucurbita* during the Pleistocene[3]), effective population sizes ($N_e$) between 100 and 60,000 individuals and migration rates (m) between 0.0001 and 0.5. The constant gene flow scenarios calculated a migration rate matrix throughout the simulation, from the present back to the common ancestral population of all lineages, with independent migration rates for each possible direction of gene flow. The secondary contact scenarios simulated a migration matrix only at the start of the simulation, before the coalescence event between the wild and domesticated lineages (see Data S1–S6 for detailed model parameters). We also constrained the times of divergence in all scenarios by forcing the domesticated taxa to diverge after the wild one. Each generation in the model corresponds to one calendar year, as this species displays an annual life cycle[13]. The 20 replicates of each model converged to similar likelihoods, indicating that the simulations performed well. After corroborating that all replicates converged to similar likelihoods, we combined all replicates and retained all outputs that were above 95% of the likelihood distribution. We found that the Jalisco model of divergence with secondary contact had the lowest Akaike Information Criterion values for all the tested models.

### Tests to detect selective signals and introgression

We used BayeScEnv[24] to detect putative regions under selection that were differentiated between *sororia* and *argyrosperma*. For the "environmental" values used by BayeScEnv, we assigned each population as either wild (0) or domesticated (1). We ran two independent MCMC analyses with 20 initial pilot runs with a length of 10,000 generations and a main run with an initial burn-in of 100,000 generations and a subsequent sampling step for 100,000 generations sampling every 20 generations. We confirmed the convergence between both chains using the Gelman and Rubin statistic[91]. The SNPs with *q*-values < 0.05 were regarded as candidate loci under selection.

The Mahalanobis distances implemented in PCAdapt[25] were used to detect candidate SNPs after controlling for the first two principal components in our dataset, which correspond to the subspecies and geographical differentiation observed among the populations (see Fig. S2). We performed Bonferroni corrections to adjust the *p*-values obtained from PCAdapt and the SNPs with Bonferroni-corrected *p*-values < 0.05 were regarded as candidate loci under selection.

We used LFMM 2[26] to identify candidate SNPs differentiating the wild and domesticated phenotypes of *C. argyrosperma*. We tested *K* number of latent factors from 1 to 10 using sNMF[27] and determined an optimal $K = 6$. We used 6 latent factors and a ridge penalty to identify significant associations between the response (wild or domesticated phenotypes) and the genetic variants. Finally, we performed FDR corrections to obtain *q*-values, with *q*-values < 0.05 being regarded as candidate loci under selection.

We performed an ABBA-BABA test using Dsuit[92] against the 11,498,421 whole-genome variants to evaluate signals of introgression or incomplete lineage sorting between *argyrosperma*, *sororia* and *C. moschata*, while using *C. okeechobeensis* subsp. *martinezii* as an outgroup. We calculated a global *D*-statistic by performing a SNP-by-SNP analysis to determine the amount of ABBA and BABA sites throughout the entire genome. We also calculated local *D*-statistics within 500 SNP windows with a step size of 250 SNPs. We used the five tGBS data of *C. moschata*, as well as the whole-genome sequences

of *C. moschata* and *C. okeechobeensis* subsp. *martinezii*, to define the ancestral state of each candidate locus from the selection scans and determine the direction of the selective signals or whether they corresponded to ABBA or BABA sites. We used snpEff[93] to associate the candidate loci found in at least two tests with the genome annotation of *argyrosperma*[14].

Acknowledgements
This work was funded by Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) KE004 "Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de *Cucurbita*" and CONABIO PE001 "Diversidad genética de las especies de *Cucurbita* en México. Fase II. Genómica evolutiva y de poblaciones, recursos genéticos y domesticación" (both awarded to R.L-S. and L.E.E.). J.B-R. and G.S-V. are doctoral students from Programa de Doctorado en Ciencias Biomédicas and Programa de Doctorado en Ciencias Biológicas, Universidad Nacional Autónoma de México, and received fellowships 583146 and 292164 from CONACyT. We acknowledge the Doctorado en Ciencias Biológicas and the Doctorado en Ciencias Biomédicas for the support provided during the development of this project. We thank Jodi Lynn Humann for her technical support while using the GenSAS web server. Special thanks to Laura Espinosa-Asuar and Silvia Barrientos for their technical support. We thank Rodrigo García Herrera, head of the Scientific Computing Department at LANCIS, Instituto de Ecología UNAM, for running the HTC infrastructure we used for the genome assembly and chromosome anchoring. The population genomic analyses were carried out using CONABIO's computing cluster, which was partially funded by SEMARNAT through the grant "Contribución de la Biodiversidad para el Cambio Climático" to CONABIO. We acknowledge the technical support of MSc Emanuel Villafán and the resources for high-performance computing that the Institute of Ecology (INECOL) made available for conducting the functional gene annotations.

Author details
[1]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Circuito Exterior s/n Anexo al Jardín Botánico, 04510 Ciudad de México, México. [2]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA. [3]Departamento de Conservación de la Biodiversidad, El Colegio de la Frontera Sur, Villahermosa, Carretera Villahermosa-Reforma km 15.5 Ranchería El Guineo 29 sección, 86280 Villahermosa, Tabasco, México. [4]Génétique Quantitative et Evolution - Le Moulon, Université Paris-Saclay, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, Centre National de la Recherche Scientifique, AgroParisTech, Gif-sur-Yvette 91190, France. [5]UBIPRO, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Av. de los Barrios #1, Col. Los Reyes Iztacala, Tlalnepantla, Edo. de Mex 54090, México

Author contributions
L.E.E., M.I.T., G.S-V., and J.B-R. designed the research; G.S-V. and G.C-M. collected the biological material; G.S-V. extracted DNA for the tGBS data; G.C-M. and G.S-V. coordinated the sequencing procedures of the tGBS data; J.B-R. and X.A-D. extracted DNA for the reference genome and coordinated its sequencing procedures; E.A-P. coordinated the administrative and lab work; J-B-R., J.A.A-L. and Y.T.G-G. performed the bioinformatic analyses; G.S-V., J.B-R., J.A.A-L., M.I.T., and L.E.E. analyzed the results; J.B-R., G.S-V., and L.E.E. drafted the manuscript; R.L-S. and L.E.E. obtained the funding. All authors revised the final version of the manuscript.

Data availability
All the raw sequencing data and genome assemblies are available in the National Center of Biotechnology Information under the BioProject accession PRJNA485527 (genome accesions SDJN00000000 and JAGKQH010000000; SRA accesions available in Table S3). The genome assemblies are also available in Figshare (https://doi.org/10.6084/m9.figshare.14370584.v1) and at the Cucurbit Genomics Database[94].

Conflict of interest
The authors declare no competing interest.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41438-021-00544-9.

Received: 9 November 2020 Revised: 3 March 2021 Accepted: 14 March 2021
Published online: 01 May 2021

References

1. Meyer, R. S. & Purugganan, M. D. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840-852 (2013).
2. Zeder, M. A. Core questions in domestication research. *Proc. Natl Acad. Sci. USA* 112, 3191-3198 (2015).
3. Kistler, L. et al. Gourds and squashes (*Cucurbita* spp.) adapted to megafaunal extinction and ecological anachronism through domestication. *Proc. Natl Acad. Sci. USA* 112, 15107-15112 (2015).
4. Paris, H. S. in Genetics and Genomics of Cucurbitaceae 111-154 (Springer International Publishing, 2016). https://doi.org/10.1007/7397_2016_3.
5. Castellanos-Morales, G. et al. Historical biogeography and phylogeny of *Cucurbita*: Insights from ancestral area reconstruction and niche evolution. *Mol. Phylogenet Evol.* 128, 38–54 (2018).
6. Sanjur, O. I., Piperno, D. R., Andres, T. C. & Wessel-Beaver, L. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin. *Proc. Natl Acad. Sci. USA* 99, 535-540 (2002).
7. Zizumbo-Villarreal, D., Flores-Silva, A. & Marín, P. C.-G. The Archaic Diet in mesoamerica: incentive for milpa development and species domestication. *Economic Bot.* 66, 328-343 (2012).
8. Chomicki, G., Schaefer, H. & Renner, S. S. Origin and domestication of Cucurbitaceae crops: insights from phylogenies, genomics and archaeology. *New Phytol.* 226, 1240-1255 (2019).
9. Whitaker, T. W.& Cutler, H. C. Cucurbits and cultures in the Americas. *Economic Bot.* 19, 344-349 (1965).
10. Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J. & Dickau, R. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl Acad. Sci. USA* 106, 5019-5024 (2009).
11. Sánchez-de la Vega, G. et al. Genetic resources in the Calabaza Pipiana Squash (*Cucurbita argyrosperma*) in Mexico: genetic diversity, genetic differentiation and distribution models. *Front Plant Sci.* 9, 400 (2018).
12. Montes-Hernandez, S. & Eguiarte, L. E. Genetic structure and indirect estimates of gene flow in three taxa of *Cucurbita* (Cucurbitaceae) in western Mexico. *Am. J. Bot.* 89, 1156-1163 (2002).
13. Lira, R. et al. in *Ethnobotany of Mexico* 389-401 (Springer New York, 2016). https://doi.org/10.1007/978-1-4614-6669-7_15.
14. Barrera-Redondo, J. et al. The genome of *Cucurbita argyrosperma* (Silver-Seed Gourd) reveals faster rates of protein-coding gene and long noncoding RNA turnover and neofunctionalization within *Cucurbita*. *Mol. Plant* 12, 506-520 (2019).
15. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V.& Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210-3212 (2015).
16. Sun, H. et al. Karyotype stability and unbiased fractionation in the Paleo-Allotetraploid *Cucurbita* genomes. *Mol. Plant* 10, 1293-1306 (2017).
17. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32, 767-769 (2016).
18. Alonge, M. et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.* 20, 224 (2019).
19. Whitaker, T. W. & Bemis, W. P. Origin and evolution of the cultivated *Cucurbita*. *Bull. Torrey Bot. Club* 102, 362-368 (1975).
20. Ott, A. et al. tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. *Nucleic Acids Res.* 45, e178 (2017).
21. Lee, T. H., Guo, H., Wang, X., Kim, C. & Paterson, A. H. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15, 162 (2014).
22. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655-1664 (2009).

23. Excoffier, L. & Foll, M. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27, 1332-1334 (2011).
24. de Villemereuil, P. & Gaggiotti, O. E. A new $F_{ST}$-based method to uncover local adaptation using environmental variables. *Methods Ecol. Evol.* 6, 1248-1258 (2015).
25. Luu, K., Bazin, E. & Blum, M. G. PCAdapt: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* 17, 67-77 (2017).
26. Caye, K., Jumentier, B., Lepeule, J. & François, O. LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Mol. Biol. Evol.* 36, 852-860 (2019).
27. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196, 973-983 (2014).
28. Khan, A. W. et al. Super-Pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci.* 25, 148-158 (2020).
29. Wilson, H. D. Gene flow in squash species. *BioScience* 40, 449-455 (1990).
30. Jarvis, D. I. et al. A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. *Proc. Natl Acad. Sci. USA* 105, 5326-5331 (2008).
31. Montes-Hernández, S., Merrick, L. C. & Eguiarte, L. E. Maintenance of squash (*Cucurbita* spp.) landrace diversity by farmers activities in Mexico. *Genet. Resour. Crop Evol.* 52, 697-707 (2005).
32. Barrera-Redondo, J. et al. Landrace diversity and local selection criteria of domesticated squashes and gourds (*Cucurbita*) in the central Andean mountain range of Peru: Tomayquichua, Huánuco. *Bot. Sci.* 98, 101-116 (2020).
33. Ranere, A. J., Piperno, D. R., Holst, I., Dickau, R. & Iriarte, J. The cultural and chronological context of early Holocene maize and squash domestication in the Central Balsas River Valley, Mexico. *Proc. Natl Acad. Sci. USA* 106, 5014-5018 (2009).
34. Zizumbo-Villarreal, D. & Colunga-GarcíaMarín, P. Origin of agriculture and plant domestication in West Mesoamerica. *Genet. Resour. Crop Evol.* 57, 813-825 (2010).
35. Stinnesbeck, W. et al. The earliest settlers of Mesoamerica date back to the late Pleistocene. *PLoS ONE* 12, e0183345 (2017).
36. Piperno, D. R. The origins of plant cultivation and domestication in the new world tropics. *Curr. Anthropol.* 52, S453–S470 (2011).
37. Lowry, D. B. et al. Responsible RAD: striving for best practices in population genomic studies of adaptation. *Mol. Ecol. Resour.* 17, 366-369 (2017).
38. Zhang, J. et al. Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. *Cell Host Microbe* 7, 290-301 (2010).
39. Zeilmaker, T. et al. DOWNY MILDEW RESISTANT 6 and DMR6-LIKE OXYGENASE 1 are partially redundant but distinct suppressors of immunity in *Arabidopsis*. *Plant J.* 81, 210-222 (2015).
40. Hayashi, S. et al. The glycerophosphoryl diester phosphodiesterase-like proteins SHV3 and its homologs play important roles in cell wall organization. *Plant Cell Physiol.* 49, 1522-1535 (2008).
41. de Jesús-Pires, C. et al. Plant thaumatin-like proteins: function, evolution and biotechnological applications. *Curr. Protein Pept. Sci.* 21, 36-51 (2020).
42. Batanero, E., Ledesma, A., Villalba, M. & Rodríguez, R. Purification, amino acid sequence and immunological characterization of Ole e 6, a cysteine-enriched allergen from olive tree pollen. *FEBS Lett.* 410, 293-296 (1997).
43. Moreira, X., Abdala-Roberts, L., Gols, R. & Francisco, M. Plant domestication decreases both constitutive and induced chemical defenses by direct selection against defensive traits. *Sci. Rep.* 8, 12678 (2018).
44. Ulm, R. et al. Distinct regulation of salinity and genotoxic stress responses by *Arabidopsis* MAP kinase phosphatase 1. *EMBO J.* 21, 6483-6493 (2002).
45. Chen, K. et al. Abscisic acid dynamics, signaling, and functions in plants. *J. Integr. Plant Biol.* 62, 25-54 (2020).
46. Martínez, A. B. et al. Differences in seed dormancy associated with the domestication of *Cucurbita maxima*: elucidation of some mechanisms behind this response. *Seed Sci. Res.* 28, 1-7 (2017).
47. Park, J. E., Kim, Y. S., Yoon, H. K. & Park, C. M. Functional characterization of a small auxin-up RNA gene in apical hook development in *Arabidopsis*. *Plant Sci.* 172, 150-157 (2007).
48. Yang, L., Jiang, Z., Jing, Y. & Lin, R. PIF1 and RVE1 form a transcriptional feedback loop to control light-mediated seed germination in *Arabidopsis*. *J. Integr. plant Biol.* 62, 1372-1384 (2020).
49. Hare, P. D., Møller, S. G., Huang, L. F.& Chua, N. H. LAF3, a novel factor required for normal phytochrome A signaling. *Plant Physiol.* 133, 1592-1604 (2003).
50. Liscum, E. & Reed, J. W. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol. Biol.* 49, 387–400 (2002).
51. Burla, B. et al. Vacuolar transport of abscisic acid glucosyl ester is mediated by ATP-binding cassette and proton-antiport mechanisms in *Arabidopsis*. *Plant Physiol.* 163, 1446-1458 (2013).
52. Seiler, C. et al. ABA biosynthesis and degradation contributing to ABA homeostasis during barley seed development under control and terminal drought-stress conditions. *J. Exp. Bot.* 62, 2615-2632 (2011).
53. Wang, Y. et al. Construction of a high-density genetic map and analysis of seed-related traits using specific length amplified fragment sequencing for *Cucurbita maxima*. *Front Plant Sci.* 10, 1782 (2019).
54. Guo, D. et al. Resequencing 200 flax cultivated accessions identifies candidate genes related to seed size and weight and reveals signatures of artificial selection. *Front Plant Sci.* 10, 1682 (2019).
55. Navarro, L. et al. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* 312, 436-439 (2006).
56. Holdsworth, W. L., LaPlant, K. E., Bell, D. C., Jahn, M. M. & Mazourek, M. Cultivar-based introgression mapping reveals wild species-derived Pm-0, the major powdery mildew resistance locus in squash. *PloS ONE* 11, e0167715 (2016).
57. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18 (2017).
58. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907-915 (2019).
59. Alverson, A. J. et al. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.* 27, 1436-1448 (2010).
60. Kajitani, R. et al. Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 24, 1384-1395 (2014).
61. Ye, C., Hill, C. M., Wu, S., Ruan, J. & Ma, Z. S. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Sci. Rep.* 6, 31900 (2016).
62. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094-3100 (2018).
63. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9, e112963 (2014).
64. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589-595 (2010).
65. Humann, J. L., Lee, T., Ficklin, S.& Main, D. Structural and functional annotation of eukaryotic genomes with GenSAS. *Methods Mol. Biol.* 1962, 29–51 (2019).
66. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21 (2013).
67. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* 7, 62 (2006).
68. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236-1240 (2014).
69. Schneider, M. et al. The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. *J. Proteom.* 72, 567-573 (2009).
70. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* 10, 421 (2009).
71. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722-736 (2017).
72. Gong, L., Pachner, M., Kalai, K. & Lelley, T. SSR-based genetic linkage map of *Cucurbita moschata* and its synteny with *Cucurbita pepo*. *Genome* 51, 878-887 (2008).
73. Haug-Baltzell, A., Stephens, S. A., Davey, S., Scheidegger, C. E. & Lyons, E. SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics* 33, 2197-2198 (2017).
74. Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 20, 277 (2019).
75. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12 (2004).
76. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* 15, 461-468 (2018).

77. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22, 1600-1607 (2006).
78. Pratas, D., Silva, R. M., Pinho, A. J. & Ferreira, P. J. An alignment-free method to find and visualize rearrangements between pairs of DNA sequences. *Sci. Rep.* 5, 10203 (2015).
79. Li, S. & Chou, H. H. LUCY2: an interactive DNA sequence quality trimming and vector removal tool. *Bioinformatics* 20, 2865-2866 (2004).
80. Hoffmann, S. et al. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol.* 5, e1000502 (2009).
81. Caboche, S., Audebert, C., Lemoine, Y. & Hot, D. Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data. *BMC Genomics* 15, 264 (2014).
82. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079 (2009).
83. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851-1858 (2008).
84. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559-575 (2007).
85. Wigginton, J. E., Cutler, D. J. & Abecasis, G. R. A note on exact tests of Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* 76, 887-893 (2005).
86. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* 27, 2156-2158 (2011).
87. Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W. & Prodöhl, P. A. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evolution* 4, 782-788 (2013).
88. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A.& Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124-3140 (2013).
89. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multi-dimensional SNP frequency data. *PLoS Genet.* 5, e1000695 (2009).
90. Ardelean, C. F. et al. Evidence of human occupation in Mexico around the Last Glacial Maximum. *Nature* 548, 87-92 (2020).
91. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457-472 (1992).
92. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite—fast *D*-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* (2020).
93. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80-92 (2012).
94. Zheng, Y. et al. Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* 47, D1128-D1136 (2019).

Supplementary information for:

# The domestication of *Cucurbita argyrosperma* as revealed by the genome of its wild relative

Josué Barrera-Redondo, Guillermo Sánchez-de la Vega, Jonás A. Aguirre-Liguori, Gabriela Castellanos-Morales, Yocelyn T. Gutiérrez-Guerrero, Xitlali Aguirre-Dugua, Erika Aguirre-Planter, Maud I. Tenaillon, Rafael Lira-Saade, Luis E. Eguiarte

Contact: josue_barrera@comunidad.unam.mx, rlira@unam.mx and fruns@unam.mx

**This file includes:**

---

**Supplementary Tables**



**Supplementary Figures**



**Supplementary Data**

## Supplementary Tables

**Table S1. Assembly metrics of the genome of *Cucurbita argyrosperma* subsp. *sororia* and *C. argyrosperma* subsp. *argyrosperma*, before and after RaGOO scaffolding.**

| Metrics | *C. argyrosperma* subsp. *sororia* | | *C. argyrosperma* subsp. *argyrosperma* | |
|---|---|---|---|---|
| | **Before RaGOO** | **After RaGOO** | **Before RaGOO** | **After RaGOO** |
| Assembly size (bp) | 252,760,152 | 255,194,784 | 228,814,150 | 231,583,150 |
| No. of contigs | 817 | 959 | 1,481 | 1,653 |
| No. of scaffolds | 817 | 72 | 920 | 27 |
| Longest contig (bp) | 4,976,248 | 4,922,130 | 2,172,140 | 2,172,140 |
| Longest scaffold (bp) | 4,976,248 | 21,187,142 | 2,746,581 | 22,228,640 |
| Contig N50 (bp) | 1,323,288 | 1,205,533 | 463,388 | 447,042 |
| Scaffold N50 (bp) | 1,323,288 | 12,094,557 | 620,880 | 11,677,370 |
| Contig L50 | 58 contigs | 60 contigs | 132 contigs | 134 contigs |
| Scaffold L50 | 58 scaffolds | 9 scaffolds | 103 scaffolds | 9 scaffolds |
| No. of contigs > 1 kb | 817 (100.0%) | 959 (100.0%) | 1,481 (100.0%) | 1,625 (98.3%) |
| No. of contigs > 10 kb | 799 (97.8%) | 941 (98.1%) | 1,417 (95.7%) | 1,527 (92.4%) |
| No. of contigs > 100 kb | 289 (35.4%) | 325 (33.9%) | 493 (33.3%) | 492 (29.8%) |
| No. of scaffolds > 10kb | 799 (97.8%) | 62 (86.1%) | 903 (98.2%) | 25 (92.6%) |
| No. of scaffolds > 100kb | 289 (35.4%) | 25 (34.7%) | 455 (49.5%) | 20 (74.1%) |
| No. of scaffolds > 1Mb | 81 (9.9%) | 20 (27.8%) | 51 (5.5%) | 20 (74.1%) |
| CG content | 36.54% | | 36.45% | |
| Illumina read coverage | 213x | | 120x | |
| PacBio read coverage | 75.4x | | 31x | |
| No. of genes | 30,592 | | 27,998 | |
| Average gene size (bp) | 3,318 | | 3,476 | |
| Complete BUSCOs | 92.8% | | 93.2% | |
| Fragmented BUSCOs | 1.2% | | 0.9% | |
| Missing BUSCOs | 6.0% | | 5.9% | |

**Table S2. High-confidence structural variants found between the genomes of *C. argyrosperma* subsp. *sororia* (wild genome) and *C. argyrosperma* subsp. *argyrosperma* (domesticated genome), as predicted by both SyRI and Sniffles.**

| Structural variants | Number of variants | Cumulative size of variants (bp) | Genes within variants | Enriched GO biological functions (*p*-value) |
|---|---|---|---|---|
| * Copy-gain variants | 159 | 162,758 | 3 | DNA replication (0.0088) |
| * Copy-loss variants | 259 | 454,056 | 12 | Proton transmembrane transport (5.5 $e^{-5}$) |
| Translocations | 22 | 73,002 | 23 | DNA topological change (0.0065)<br>Fatty acid biosynthetic process (0.0409) |
| Inversions | 3 | 19,271 | 4 | DNA replication (0.017)<br>Response to auxin (0.025) |
| ** Unaligned regions in domesticated genome | 2479 | 18,979,604 | 149 | Cytoskeleton organization (0.0045)<br>NADP biosynthetic process (0.0182)<br>Tryptophan biosynthetic process (0.0212)<br>Negative regulation of translation (0.0272)<br>Protein ubiquitination (0.0275)<br>Lysine biosynthetic process via diaminopimelate (0.0302)<br>Inositol phosphate dephosphorylation (0.0391)<br>Translational termination (0.0450) |
| ** Unaligned regions in wild genome | 3846 | 28,437,839 | 637 | Proteolysis (0.0005)<br>Tryptophan biosynthetic process (0.0012)<br>DNA replication initiation (0.0050)<br>Phosphorelay signal transduction system (0.0123)<br>ATP synthesis coupled proton transport (0.0130)<br>Negative regulation of DNA helicase activity (0.0134)<br>Glucosylceramide catabolic process (0.0266)<br>Sucrose biosynthetic process (0.0266)<br>Cytoskeleton organization (0.0307)<br>L-phenylalanine biosynthetic process (0.0460)<br>Cytochrome complex assembly (0.0460) |

* Copy-number variants are considered gains or losses with respect to the domesticated genome.
** The unaligned regions were only predicted by SyRI.

**Table S3. Information of 192 individuals sequenced using tGBS libraries, including population name, geographical coordinates and SRA accession. (within population names: W = wild, D = domesticated)**

| Individual ID | Population name | Population number | Taxon | Latitude | Longitude | SRA accession |
|---|---|---|---|---|---|---|
| M_SON1 | Alamos, Sonora (Outgroup) | 0 | *Cucurbita moschata* (Outgroup) | 27.02694 | -108.93659 | SRR12937368 |
| M_SON2 | Alamos, Sonora (Outgroup) | 0 | *Cucurbita moschata* (Outgroup) | 27.02694 | -108.93659 | SRR12937367 |
| M_SON3 | Alamos, Sonora (Outgroup) | 0 | *Cucurbita moschata* (Outgroup) | 27.02694 | -108.93659 | SRR12937288 |
| M_SON4 | Alamos, Sonora (Outgroup) | 0 | *Cucurbita moschata* (Outgroup) | 27.02694 | -108.93659 | SRR12937277 |
| M_SON5 | Alamos, Sonora (Outgroup) | 0 | *Cucurbita moschata* (Outgroup) | 27.02694 | -108.93659 | SRR12937266 |
| M_SON6 | Alamos, Sonora (Outgroup) | 0 | *Cucurbita moschata* (Outgroup) | 27.02694 | -108.93659 | SRR12937255 |
| S_CHIS1 | Jiquipilas, Chiapas (W) | 1 | *Cucurbita argyrosperma* subsp. *sororia* | 16.597486 | -93.626878 | SRR12937244 |
| S_CHIS2 | Jiquipilas, Chiapas (W) | 1 | *Cucurbita argyrosperma* subsp. *sororia* | 16.597486 | -93.626878 | SRR12937233 |
| S_CHIS3 | Jiquipilas, Chiapas (W) | 1 | *Cucurbita argyrosperma* subsp. *sororia* | 16.597486 | -93.626878 | SRR12937222 |
| S_CHIS4 | Jiquipilas, Chiapas (W) | 1 | *Cucurbita argyrosperma* subsp. *sororia* | 16.597486 | -93.626878 | SRR12937211 |
| S_CHIS5 | Jiquipilas, Chiapas (W) | 1 | *Cucurbita argyrosperma* subsp. *sororia* | 16.597486 | -93.626878 | SRR12937366 |
| S_CHIS6 | Jiquipilas, Chiapas (W) | 1 | *Cucurbita argyrosperma* subsp. *sororia* | 16.597486 | -93.626878 | SRR12937355 |
| S_CHIS7 | Jiquipilas, Chiapas (W) | 1 | *Cucurbita argyrosperma* subsp. *sororia* | 16.597486 | -93.626878 | SRR12937200 |
| S_CHIS8 | Jiquipilas, Chiapas (W) | 1 | *Cucurbita argyrosperma* subsp. *sororia* | 16.597486 | -93.626878 | SRR12937189 |
| S_GRO1 | Ometepec, Guerrero (W) | 2 | *Cucurbita argyrosperma* subsp. *sororia* | 16.688139 | -98.406319 | SRR12937178 |
| S_GRO2 | Ometepec, Guerrero (W) | 2 | *Cucurbita argyrosperma* subsp. *sororia* | 16.688139 | -98.406319 | SRR12937335 |
| S_GRO3 | Ometepec, Guerrero (W) | 2 | *Cucurbita argyrosperma* subsp. *sororia* | 16.688139 | -98.406319 | SRR12937324 |
| S_GRO4 | Ometepec, Guerrero (W) | 2 | *Cucurbita argyrosperma* subsp. *sororia* | 16.688139 | -98.406319 | SRR12937313 |
| S_GRO5 | Ometepec, Guerrero (W) | 2 | *Cucurbita argyrosperma* subsp. *sororia* | 16.688139 | -98.406319 | SRR12937302 |

| | | | | | | |
|---|---|---|---|---|---|---|
| S_OAX1 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937291 |
| S_OAX2 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937287 |
| S_OAX3 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937286 |
| S_OAX4 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937285 |
| S_OAX5 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937284 |
| S_OAX6 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937283 |
| S_OAX7 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937282 |
| S_OAX8 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937281 |
| S_OAX9 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937280 |
| S_OAX10 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.918225 | -97.076308 | SRR12937279 |
| S_OAX11 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.9528333 | -97.0772778 | SRR12937278 |

| | | | | | | |
|---|---|---|---|---|---|---|
| S_OAX12 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.9528333 | -97.0772778 | SRR12937276 |
| S_OAX13 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.9528333 | -97.0772778 | SRR12937275 |
| S_OAX14 | Puerto Escondido, Oaxaca (W) | 3 | *Cucurbita argyrosperma* subsp. *sororia* | 15.9528333 | -97.0772778 | SRR12937274 |
| S_JAL1 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.682 | -104.333278 | SRR12937273 |
| S_JAL2 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.682 | -104.333278 | SRR12937272 |
| S_JAL3 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.7019583 | -104.204314 | SRR12937271 |
| S_JAL4 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9005833 | -104.160222 | SRR12937270 |
| S_JAL5 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9005833 | -104.160222 | SRR12937269 |
| S_JAL6 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9005833 | -104.160222 | SRR12937268 |
| S_JAL7 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9005833 | -104.160222 | SRR12937267 |
| S_JAL8 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.6997222 | -104.203056 | SRR12937265 |
| S_JAL9 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.6997222 | -104.203056 | SRR12937264 |
| S_JAL10 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.6997222 | -104.203056 | SRR12937263 |
| S_JAL11 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.8753889 | -104.072333 | SRR12937262 |
| S_JAL12 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9123056 | -104.116833 | SRR12937261 |
| S_JAL13 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9123056 | -104.116833 | SRR12937260 |
| S_JAL14 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9123056 | -104.116833 | SRR12937259 |
| S_JAL15 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9123056 | -104.116833 | SRR12937258 |
| S_JAL16 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9600833 | -104.037472 | SRR12937257 |
| S_JAL17 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.9574722 | -103.988389 | SRR12937256 |

| | | | | | | |
|---|---|---|---|---|---|---|
| S_JAL18 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.8753889 | -104.072333 | SRR12937254 |
| S_JAL19 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.8356111 | -104.081417 | SRR12937253 |
| S_JAL20 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.8356111 | -104.081417 | SRR12937252 |
| S_JAL21 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.6909722 | -104.360833 | SRR12937251 |
| S_JAL22 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.6909722 | -104.360833 | SRR12937250 |
| S_JAL23 | Jalisco (W) | 4 | *Cucurbita argyrosperma* subsp. *sororia* | 19.6909722 | -104.360833 | SRR12937249 |
| A_TLAP1 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937248 |
| A_TLAP2 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937247 |
| A_TLAP3 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937246 |
| A_TLAP4 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937245 |
| A_TLAP5 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937243 |
| A_TLAP6 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937242 |
| A_TLAP7 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937241 |
| A_TLAP8 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937240 |
| A_TLAP9 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937239 |
| A_TLAP10 | Tlapehuala, Guerrero (D) | 5 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 18.2416667 | -100.534722 | SRR12937238 |
| A_JAL1 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.7019583 | -104.204314 | SRR12937237 |
| A_JAL2 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.7019583 | -104.204314 | SRR12937236 |
| A_JAL3 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.6997222 | -104.203056 | SRR12937235 |
| A_JAL4 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.6997222 | -104.203056 | SRR12937234 |
| A_JAL5 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8753889 | -104.072333 | SRR12937232 |
| A_JAL6 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8753889 | -104.072333 | SRR12937231 |
| A_JAL7 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.9600833 | -104.037472 | SRR12937230 |

| | | | | | | |
|---|---|---|---|---|---|---|
| A_JAL8 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8714722 | -104.217333 | SRR12937229 |
| A_JAL9 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8714722 | -104.217333 | SRR12937228 |
| A_JAL10 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.9574722 | -103.988389 | SRR12937227 |
| A_JAL11 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8356111 | -104.081417 | SRR12937226 |
| A_JAL12 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8356111 | -104.081417 | SRR12937225 |
| A_JAL13 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8356111 | -104.081417 | SRR12937224 |
| A_JAL14 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8356111 | -104.081417 | SRR12937223 |
| A_JAL15 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8356111 | -104.081417 | SRR12937221 |
| A_JAL16 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8356111 | -104.081417 | SRR12937220 |
| A_JAL17 | Jalisco (D) | 6 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.8356111 | -104.081417 | SRR12937219 |
| A_BAD1 | Badiraguato, Sinaloa (D) | 7 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 25.359173 | -107.558408 | SRR12937218 |
| A_MTP1 | Matlalapa, Guerrero (D) | 8 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5971639 | -99.4579917 | SRR12937217 |
| A_MTP2 | Matlalapa, Guerrero (D) | 8 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5971639 | -99.4579917 | SRR12937216 |
| A_MTP3 | Matlalapa, Guerrero (D) | 8 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5971639 | -99.4579917 | SRR12937215 |
| A_MTP4 | Matlalapa, Guerrero (D) | 8 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5971639 | -99.4579917 | SRR12937214 |
| A_MTP5 | Matlalapa, Guerrero (D) | 8 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5971639 | -99.4579917 | SRR12937213 |
| A_MTP6 | Matlalapa, Guerrero (D) | 8 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5971639 | -99.4579917 | SRR12937212 |
| A_MTP7 | Matlalapa, Guerrero (D) | 8 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5971639 | -99.4579917 | SRR12937210 |
| A_SAH1 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937209 |
| A_SAH2 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937208 |
| A_SAH3 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937207 |
| A_SAH4 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937206 |
| A_SAH5 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937205 |

| | | | | | | |
|---|---|---|---|---|---|---|
| A_SAH6 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937204 |
| A_SAH7 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937203 |
| A_SAH8 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937202 |
| A_SAH9 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937201 |
| A_SAH10 | Sahuayo, Michoacán (D) | 9 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.0587194 | -102.716233 | SRR12937365 |
| A_SAL1 | Salamanca, Guanajuato (D) | 10 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.5205778 | -101.190992 | SRR12937364 |
| A_SJI1 | San José Iturbide, Guanajuato (D) | 10 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.9988889 | -100.385 | SRR12937363 |
| A_NAY1 | Tepic, Nayarit (D) | 11 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 21.519956 | -104.893423 | SRR12937362 |
| A_NAY2 | Tepic, Nayarit (D) | 11 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 21.519956 | -104.893423 | SRR12937361 |
| A_NAY3 | Tepic, Nayarit (D) | 11 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 21.519956 | -104.893423 | SRR12937360 |
| A_NAY4 | Tepic, Nayarit (D) | 11 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 21.519956 | -104.893423 | SRR12937359 |
| F_PLAT1 | El Platanar, Sinaloa (feral) | 12 | feral individual | 24.0303667 | -106.432561 | SRR12937358 |
| F_PLAT2 | El Platanar, Sinaloa (feral) | 12 | feral individual | 24.0303667 | -106.432561 | SRR12937357 |
| F_PLAT3 | El Platanar, Sinaloa (feral) | 12 | feral individual | 24.0303667 | -106.432561 | SRR12937356 |
| F_PLAT4 | El Platanar, Sinaloa (feral) | 12 | feral individual | 24.0303667 | -106.432561 | SRR12937354 |
| F_PLAT5 | El Platanar, Sinaloa (feral) | 12 | feral individual | 24.0303667 | -106.432561 | SRR12937353 |
| F_PLAT6 | El Platanar, Sinaloa (feral) | 12 | feral individual | 24.0303667 | -106.432561 | SRR12937352 |
| F_PLAT7 | El Platanar, Sinaloa (feral) | 12 | feral individual | 24.0303667 | -106.432561 | SRR12937351 |
| F_CUL1 | Culiacán, Sinaloa (feral) | 13 | feral individual | 24.817335 | -107.416667 | SRR12937350 |
| F_CUL2 | Culiacán, Sinaloa (feral) | 13 | feral individual | 24.817335 | -107.416667 | SRR12937349 |
| F_CUL3 | Culiacán, Sinaloa (feral) | 13 | feral individual | 24.817335 | -107.416667 | SRR12937348 |
| F_CUL4 | Culiacán, Sinaloa (feral) | 13 | feral individual | 24.817335 | -107.416667 | SRR12937347 |

| | | | | | | |
|---|---|---|---|---|---|---|
| F_CUL5 | Culiacán, Sinaloa (feral) | 13 | feral individual | 24.817335 | -107.416667 | SRR12937346 |
| A_CHOI1 | Choix, Sinaloa (D) | 14 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 26.5967833 | -108.335581 | SRR12937345 |
| A_CHOI2 | Choix, Sinaloa (D) | 14 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 26.5967833 | -108.335581 | SRR12937199 |
| A_CHOI3 | Choix, Sinaloa (D) | 14 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 26.5967833 | -108.335581 | SRR12937198 |
| A_CHOI4 | Choix, Sinaloa (D) | 14 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 26.5967833 | -108.335581 | SRR12937197 |
| A_CHOI5 | Choix, Sinaloa (D) | 14 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 26.5967833 | -108.335581 | SRR12937196 |
| A_YEC1 | Yecora, Sonora (D) | 15 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.3720417 | -108.926986 | SRR12937195 |
| A_YEC2 | Yecora, Sonora (D) | 15 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.3720417 | -108.926986 | SRR12937194 |
| A_YEC3 | Yecora, Sonora (D) | 15 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.3720417 | -108.926986 | SRR12937193 |
| A_YEC4 | Yecora, Sonora (D) | 15 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.3720417 | -108.926986 | SRR12937192 |
| A_YEC5 | Yecora, Sonora (D) | 15 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.3720417 | -108.926986 | SRR12937191 |
| A_YEC6 | Yecora, Sonora (D) | 15 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.3720417 | -108.926986 | SRR12937190 |
| F_ONAV1 | Onavas, Sonora (feral) | 16 | feral individual | 28.533333 | -109.583333 | SRR12937188 |
| F_ONAV2 | Onavas, Sonora (feral) | 16 | feral individual | 28.533333 | -109.583333 | SRR12937187 |
| F_ONAV3 | Onavas, Sonora (feral) | 16 | feral individual | 28.533333 | -109.583333 | SRR12937186 |
| F_ONAV4 | Onavas, Sonora (feral) | 16 | feral individual | 28.533333 | -109.583333 | SRR12937185 |
| F_ONAV5 | Onavas, Sonora (feral) | 16 | feral individual | 28.533333 | -109.583333 | SRR12937184 |
| F_ONAV6 | Onavas, Sonora (feral) | 16 | feral individual | 28.533333 | -109.583333 | SRR12937183 |
| F_ONAV7 | Onavas, Sonora (feral) | 16 | feral individual | 28.533333 | -109.583333 | SRR12937182 |
| A_DGO1 | Durango, Durango (D) | 17 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 24.066667 | -104.583333 | SRR12937181 |
| A_TEH1 | Tehuantepec, Oaxaca (D) | 18 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 16.3328306 | -95.2330361 | SRR12937180 |
| A_TEH2 | Tehuantepec, Oaxaca (D) | 18 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 16.3328306 | -95.2330361 | SRR12937179 |
| A_TEH3 | Tehuantepec, Oaxaca (D) | 18 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 16.3328306 | -95.2330361 | SRR12937177 |

| | | | | | | |
|---|---|---|---|---|---|---|
| A_TEH4 | Tehuantepec, Oaxaca (D) | 18 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 16.3328306 | -95.2330361 | SRR12937344 |
| A_TEH5 | Tehuantepec, Oaxaca (D) | 18 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 16.3328306 | -95.2330361 | SRR12937343 |
| A_TEH6 | Tehuantepec, Oaxaca (D) | 18 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 16.3328306 | -95.2330361 | SRR12937342 |
| A_TEH7 | Tehuantepec, Oaxaca (D) | 18 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 16.3328306 | -95.2330361 | SRR12937341 |
| A_ONAV1 | Onavas, Sonora (D) | 19 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.533333 | -109.583333 | SRR12937340 |
| A_ONAV2 | Onavas, Sonora (D) | 19 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.533333 | -109.583333 | SRR12937339 |
| A_ONAV3 | Onavas, Sonora (D) | 19 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.533333 | -109.583333 | SRR12937338 |
| A_ONAV4 | Onavas, Sonora (D) | 19 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.533333 | -109.583333 | SRR12937337 |
| A_ONAV5 | Onavas, Sonora (D) | 19 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.533333 | -109.583333 | SRR12937336 |
| A_ONAV6 | Onavas, Sonora (D) | 19 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 28.533333 | -109.583333 | SRR12937334 |
| A_VER1 | Tihuatlán, Veracruz (D) | 20 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.7200667 | -97.5395028 | SRR12937333 |
| A_VER2 | Tihuatlán, Veracruz (D) | 20 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.7200667 | -97.5395028 | SRR12937332 |
| A_VER3 | Tihuatlán, Veracruz (D) | 20 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.7200667 | -97.5395028 | SRR12937331 |
| A_VER4 | Tihuatlán, Veracruz (D) | 20 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.7200667 | -97.5395028 | SRR12937330 |
| A_VER5 | Tihuatlán, Veracruz (D) | 20 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.7200667 | -97.5395028 | SRR12937329 |
| A_PAL1 | Palenque, Chiapas (D) | 21 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5128139 | -91.9877611 | SRR12937328 |
| A_PAL2 | Palenque, Chiapas (D) | 21 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5128139 | -91.9877611 | SRR12937327 |
| A_PAL3 | Palenque, Chiapas (D) | 21 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5128139 | -91.9877611 | SRR12937326 |
| A_PAL4 | Palenque, Chiapas (D) | 21 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5128139 | -91.9877611 | SRR12937325 |
| A_PAL5 | Palenque, Chiapas (D) | 21 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5128139 | -91.9877611 | SRR12937323 |
| A_PAL6 | Palenque, Chiapas (D) | 21 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 17.5128139 | -91.9877611 | SRR12937322 |
| A_SLP1 | Tanquián, San Luis Potosí (D) | 22 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 22.1154861 | -101.009331 | SRR12937321 |

| | | | | | | |
|---|---|---|---|---|---|---|
| A_SLP2 | Tanquián, San Luis Potosí (D) | 22 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 22.1154861 | -101.009331 | SRR12937320 |
| A_SLP3 | Tanquián, San Luis Potosí (D) | 22 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 22.1154861 | -101.009331 | SRR12937319 |
| A_SLP4 | Tanquián, San Luis Potosí (D) | 22 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 22.1154861 | -101.009331 | SRR12937318 |
| A_SLP5 | Tanquián, San Luis Potosí (D) | 22 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 22.1154861 | -101.009331 | SRR12937317 |
| A_SLP6 | Tanquián, San Luis Potosí (D) | 22 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 22.1154861 | -101.009331 | SRR12937316 |
| A_CHAMP1 | Champotón, Campeche (D) | 23 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.5011556 | -90.4613778 | SRR12937315 |
| A_CHAMP2 | Champotón, Campeche (D) | 23 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.5011556 | -90.4613778 | SRR12937314 |
| A_CHAMP3 | Champotón, Campeche (D) | 23 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.5011556 | -90.4613778 | SRR12937312 |
| A_CHAMP4 | Champotón, Campeche (D) | 23 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.5011556 | -90.4613778 | SRR12937311 |
| A_CHAMP5 | Champotón, Campeche (D) | 23 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.5011556 | -90.4613778 | SRR12937310 |
| A_MIXT1 | Mixtepec, Oaxaca (D) | 24 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 15.95875 | -97.0849167 | SRR12937309 |
| A_MIXT2 | Mixtepec, Oaxaca (D) | 24 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 15.95875 | -97.0849167 | SRR12937308 |
| A_MIXT3 | Mixtepec, Oaxaca (D) | 24 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 15.95875 | -97.0849167 | SRR12937307 |
| A_MIXT4 | Mixtepec, Oaxaca (D) | 24 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 15.95875 | -97.0849167 | SRR12937306 |
| A_MIXT5 | Mixtepec, Oaxaca (D) | 24 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 15.95875 | -97.0849167 | SRR12937305 |
| A_MIXT6 | Mixtepec, Oaxaca (D) | 24 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 15.95875 | -97.0849167 | SRR12937304 |
| A_EK1 | Ek Balam, Yucatan (D) | 25 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.9166667 | -87.9166667 | SRR12937303 |
| A_EK2 | Ek Balam, Yucatan (D) | 25 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.9166667 | -87.9166667 | SRR12937301 |

| | | | | | | |
|---|---|---|---|---|---|---|
| A_EK3 | Ek Balam, Yucatan (D) | 25 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.9166667 | -87.9166667 | SRR12937300 |
| A_EK4 | Ek Balam, Yucatan (D) | 25 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.9166667 | -87.9166667 | SRR12937299 |
| A_EK5 | Ek Balam, Yucatan (D) | 25 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.9166667 | -87.9166667 | SRR12937298 |
| A_EK6 | Ek Balam, Yucatan (D) | 25 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 20.9166667 | -87.9166667 | SRR12937297 |
| A_CHAN1 | Chan Santa Cruz, Quintana Roo (D) | 26 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.3670639 | -88.3328917 | SRR12937296 |
| A_CHAN2 | Chan Santa Cruz, Quintana Roo (D) | 26 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.3670639 | -88.3328917 | SRR12937295 |
| A_CHAN3 | Chan Santa Cruz, Quintana Roo (D) | 26 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.3670639 | -88.3328917 | SRR12937294 |
| A_CHAN4 | Chan Santa Cruz, Quintana Roo (D) | 26 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.3670639 | -88.3328917 | SRR12937293 |
| A_CHAN5 | Chan Santa Cruz, Quintana Roo (D) | 26 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.3670639 | -88.3328917 | SRR12937292 |
| A_CHAN6 | Chan Santa Cruz, Quintana Roo (D) | 26 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.3670639 | -88.3328917 | SRR12937290 |
| A_CHAN7 | Chan Santa Cruz, Quintana Roo (D) | 26 | *Cucurbita argyrosperma* subsp. *argyrosperma* | 19.3670639 | -88.3328917 | SRR12937289 |

**Table S4. Average genetic diversity of the wild, domesticated and feral populations of *Cucurbita argyrosperma* using 2,861 unlinked SNPs predicted with the domesticated (D) reference genome, and 1,771 unlinked SNPs predicted with the wild (W) reference genome.**

| Taxon | $N_{ind}$ | $N_{pop}$ | $H_O$ (Var) | | $H_E$ (Var) | | $\pi$ (Var) | | $F_{IS}$ (Var) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | D | W | D | W | D | W | D | W |
| *Cucurbita argyrosperma* subsp. *sororia* | 44 | 4 | 0.098 (0.017) | 0.089 (0.016) | 0.096 (0.014) | 0.094 (0.015) | 0.098 (0.015) | 0.096 (0.015) | 0.011 (0.027) | 0.044 (0.037) |
| *Cucurbita argyrosperma* subsp. *argyrosperma* | 109 | 19 | 0.094 (0.012) | 0.095 (0.015) | 0.094 (0.010) | 0.097 (0.012) | 0.095 (0.010) | 0.098 (0.012) | 0.034 (0.030) | 0.067 (0.038) |
| feral populations | 14 | 3 | 0.102 (0.032) | 0.102 (0.035) | 0.088 (0.019) | 0.089 (0.022) | 0.094 (0.023) | 0.096 (0.026) | -0.015 (0.022) | -0.011 (0.026) |
| *Cucurbita moschata* (outgroup) | 5 | 1 | 0.077 (0.042) | 0.080 (0.047) | 0.058 (0.019) | 0.058 (0.021) | 0.068 (0.029) | 0.069 (0.032) | -0.017 (0.017) | -0.020 (0.019) |

($N_{ind}$ = number of individuals, $N_{pop}$ = number of populations, $H_O$ = observed heterozygosity, $H_E$ = expected heterozygosity, $\pi$ = nucleotide diversity, $F_{IS}$ = inbreeding coefficient, Var = variance)

**Table S5. Genetic diversity of each wild, domesticated and feral population of *Cucurbita argyrosperma* using 2,861 unlinked SNPs (r2 < 0.25, MAF > 1%).**

| Population name | Population number | *N* | $H_O$ (Var) | $H_E$ (Var) | $\pi$ (Var) | $F_{IS}$ (Var) |
|---|---|---|---|---|---|---|
| *Cucurbita moschata* (Outgroup) | 0 | 5 | 0.077 (0.042) | 0.058 (0.019) | 0.068 (0.029) | -0.018 (0.017) |
| Jiquipilas, Chiapas (W) | 1 | 8 | 0.091 (0.039) | 0.073 (0.020) | 0.082 (0.027) | -0.020 (0.017) |
| Ometepec, Guerrero (W) | 2 | 5 | 0.089 (0.038) | 0.076 (0.021) | 0.088 (0.030) | 0.001 (0.027) |
| Puerto Escondido, Oaxaca (W) | 3 | 14 | 0.094 (0.031) | 0.079 (0.018) | 0.085 (0.022) | -0.021 (0.017) |
| Jalisco (W) | 4 | 17 | 0.115 (0.033) | 0.099 (0.020) | 0.106 (0.024) | -0.020 (0.020) |
| Tlapehuala, Guerrero (D) | 5 | 10 | 0.098 (0.029) | 0.086 (0.018) | 0.093 (0.022) | -0.010 (0.023) |
| Jalisco (D) | 6 | 13 | 0.099 (0.025) | 0.090 (0.017) | 0.096 (0.020) | -0.006 (0.022) |
| Badiraguato, Sinaloa (D) | 7 | 1 | 0.095 (0.086) | 0.047 (0.021) | 0.095 (0.086) | 0.000 (0.000) |
| Matlalapa, Guerrero (D) | 8 | 7 | 0.092 (0.029) | 0.081 (0.018) | 0.090 (0.023) | -0.006 (0.017) |
| Sahuayo, Michoacán (D) | 9 | 10 | 0.101 (0.027) | 0.091 (0.018) | 0.097 (0.021) | -0.008 (0.022) |
| Salamanca, Guanajuato (D) | 10 | 1 | 0.087 (0.079) | 0.043 (0.020) | 0.087 (0.079) | 0.000 (0.000) |
| Tepic, Nayarit (D) | 11 | 4 | 0.089 (0.044) | 0.067 (0.020) | 0.083 (0.034) | -0.012 (0.012) |
| El Platanar, Sinaloa (feral) | 12 | 4 | 0.107 (0.062) | 0.074 (0.024) | 0.097 (0.046) | -0.019 (0.018) |
| Culiacán, Sinaloa (feral) | 13 | 3 | 0.099 (0.064) | 0.066 (0.023) | 0.094 (0.053) | -0.010 (0.014) |
| Choix, Sinaloa (D) | 14 | 3 | 0.101 (0.060) | 0.069 (0.023) | 0.096 (0.050) | -0.008 (0.014) |
| Yecora, Sonora (D) | 15 | 6 | 0.104 (0.040) | 0.081 (0.020) | 0.092 (0.027) | -0.025 (0.013) |
| Onavas, Sonora (feral) | 16 | 7 | 0.101 (0.044) | 0.077 (0.022) | 0.088 (0.029) | -0.026 (0.015) |
| Durango, Durango (D) | 17 | 1 | 0.076 (0.071) | 0.038 (0.017) | 0.076 (0.071) | 0.000 (0.000) |
| Tehuantepec, Oaxaca (D) | 18 | 7 | 0.085 (0.032) | 0.071 (0.018) | 0.079 (0.023) | -0.014 (0.014) |
| Onavas, Sonora (D) | 19 | 6 | 0.089 (0.040) | 0.067 (0.018) | 0.078 (0.026) | -0.025 (0.013) |
| Tihuatlán, Veracruz (D) | 20 | 5 | 0.088 (0.041) | 0.069 (0.020) | 0.082 (0.030) | -0.011 (0.018) |
| Palenque, Chiapas (D) | 21 | 5 | 0.089 (0.035) | 0.072 (0.019) | 0.083 (0.026) | -0.013 (0.015) |
| Tanquián, San Luis Potosí (D) | 22 | 6 | 0.090 (0.033) | 0.074 (0.019) | 0.084 (0.025) | -0.013 (0.013) |
| Champotón, Campeche (D) | 23 | 5 | 0.096 (0.039) | 0.076 (0.020) | 0.090 (0.030) | -0.011 (0.016) |
| Mixtepec, Oaxaca (D) | 24 | 6 | 0.101 (0.038) | 0.081 (0.020) | 0.092 (0.027) | -0.020 (0.016) |
| Ek Balam, Yucatan (D) | 25 | 6 | 0.098 (0.041) | 0.075 (0.020) | 0.086 (0.027) | -0.026 (0.014) |
| Chan Santa Cruz, Quintana Roo (D) | 26 | 7 | 0.089 (0.032) | 0.073 (0.018) | 0.081 (0.022) | -0.018 (0.016) |

(*N* = sample size, $H_O$ = observed heterozygosity, $H_E$ = expected heterozygosity, $\pi$ = nucleotide diversity, $F_{IS}$ = inbreeding coefficient, Var = variance) (within population names: W = wild, D = domesticated)

**Table S6. Results of ABBA-BABA test to detect introgression using 11,498,421 variants between *C. argyrosperma* subsp. *sororia* (P1), *C. argyrosperma* subp. *argyrosperma* (P2) and *C. moschata* (P3), while using *C. okeechobeensis* subsp. *martinezii* as an outgroup.**

| P1 | P2 | P3 | AABB sites | ABBA sites | BABA sites | *D*-statistic | *p*-value | $f_G$ |
|---|---|---|---|---|---|---|---|---|
| *C. argyrosperma* subsp. *sororia* | *C. argyrosperma* subsp. *argyrosperma* | *C. moschata* | 652029 | 90253.8 | 81267 | 0.0523945 | 0.001414 | 0.0106867 |

**Table S7. Candidate genes containing at least one outlier SNP (predicted by at least two tests) within their inner structure (introns, exons, UTRs). The direction of selection was inferred according to the ancestral and derived allelic state for each outlier SNP. (AED = Annotation Edit Distance; GO ID = Gene Ontology ID)**

| Direction of selection | Gene ID (domesticated genome) | Gene ID (wild genome) | Chromosome location | Functional annotation against SwissProt | AED | GO ID |
|---|---|---|---|---|---|---|
| domesticated populations | Carg02896 | Csor.00g176320 | Chr09 | Similar to *AKT1* Potassium channel *AKT1* (*Arabidopsis thaliana*) | 0.14 | GO:0005216, GO:0006811, GO:0016020, GO:0055085 |
| domesticated populations | Carg04908 | Csor.00g009170 | Chr08 | Similar to *RPS15AE* 40S ribosomal protein S15a-5 (*Arabidopsis thaliana*) | 0.11 | GO:0003735, GO:0005840, GO:0006412 |
| domesticated populations | Carg07327 | Csor.00g196260 | Chr01 | Similar to *dusA* tRNA-dihydrouridine(20/20a) synthase (*Vibrio vulnificus*) | 0.27 | GO:0008033, GO:0017150, GO:0050660, GO:0055114 |
| domesticated populations | Carg09511 | Csor.00g083590 | Chr17 | Similar to *IAA27* Auxin-responsive protein *IAA27* (*Arabidopsis thaliana*) | 0.14 | NA |
| domesticated populations | Carg12845 | Csor.00g037030 | Chr09 | Similar to *GDPDL4* Glycerophosphodiester phosphodiesterase *GDPDL4* (*Arabidopsis thaliana*) | 0.1 | GO:0006629, GO:0008081 |

| | | | | | | |
|---|---|---|---|---|---|---|
| domesticated populations | Carg13010 | Csor.00g214890 | Chr10 | Similar to At4g29530 Thiamine phosphate phosphatase-like protein (*Arabidopsis thaliana*) | 0.11 | GO:0016791 |
| domesticated populations | Carg13432 | Csor.00g005000 | Chr08 | Similar to *efr3b* Protein *EFR3* homolog B (*Danio rerio*) | 0.16 | NA |
| domesticated populations | Carg14512 | Csor.00g157750 | Chr18 | Similar to *MKP1* Protein-tyrosine-phosphatase *MKP1* (*Arabidopsis thaliana*) | 0.01 | GO:0008138, GO:0016311 |
| domesticated populations | Carg20078 | Csor.00g227200 | Chr09 | Similar to *ABCE2* ABC transporter E family member 2 (*Arabidopsis thaliana*) | 0.12 | GO:0005524, GO:0016887 |
| domesticated populations | Carg20623 | Csor.00g016670 | Chr13 | Protein of unknown function | 0.03 | GO:0005515 |
| domesticated populations | Carg21521 | Csor.00g160720 | Chr13 | Similar to *SNI1* Negative regulator of systemic acquired resistance *SNI1* (*Arabidopsis thaliana*) | 0.19 | NA |
| domesticated populations | Carg23167 | Csor.00g003720 | Chr08 | Similar to *CES101* G-type lectin S-receptor-like serine/threonine-protein kinase *CES101* (*Arabidopsis thaliana*) | 0.04 | GO:0004672, GO:0004674, GO:0005524, GO:0006468 |
| domesticated populations | Carg26378 | Csor.00g260430 | Chr11 | Similar to *DLO1* Protein *DMR6*-LIKE OXYGENASE 1 (*Arabidopsis thaliana*) | 0.16 | GO:0016491, GO:0055114 |

| | | | | | | |
|---|---|---|---|---|---|---|
| wild populations | Carg00678 | Csor.00g059490 | Chr03 | Similar to *CSI1* Protein CELLULOSE SYNTHASE INTERACTIVE 1 (*Arabidopsis thaliana*) | 0.06 | GO:0005515 |
| wild populations | Carg00942 | Csor.00g292650 | Chr06 | Similar to *CTN* Cactin (*Arabidopsis thaliana*) | 0.14 | GO:0005515 |
| wild populations | Carg01177 | Csor.00g247900 | Chr06 | Protein of unknown function | 0.13 | GO:0071816 |
| wild populations | Carg01823 | Csor.00g164870 | Chr04 | Similar to *PP2AA2* Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform (*Arabidopsis thaliana*) | 0.19 | GO:0005515 |
| wild populations | Carg02996 | NA | Chr09 | Similar to At5g49980 Transport inhibitor response 1-like protein (*Arabidopsis thaliana*) | 0.02 | GO:0005515 |
| wild populations | Carg04587 | Csor.00g123000 | Chr04 | Similar to At3g53190 Probable pectate lyase 12 (*Arabidopsis thaliana*) | 0.06 | NA |
| wild populations | Carg04921 | Csor.00g009020 | Chr08 | Similar to *ATX2* Histone-lysine N-methyltransferase *ATX2* (*Arabidopsis thaliana*) | 0.23 | GO:0005515, GO:0005634 |
| wild populations | Carg05727 | Csor.00g172100 | Chr16 | Similar to *KIN7G* Kinesin-like protein *KIN-7G* (*Arabidopsis thaliana*) | 0.11 | GO:0003777, GO:0005524, GO:0007018, GO:0008017 |

| | | | | | | |
|---|---|---|---|---|---|---|
| wild populations | Carg06997 | Csor.00g080170 | Chr01 | Similar to *IREH1* Probable serine/threonine protein kinase *IREH1* (*Arabidopsis thaliana*) | 0.08 | GO:0004672, GO:0005524, GO:0006468 |
| wild populations | Carg07232 | Csor.00g265760 | Chr17 | Similar to *HULK3* Protein *HUA2*-LIKE 3 (*Arabidopsis thaliana*) | 0.22 | NA |
| wild populations | Carg07327 | Csor.00g196260 | Chr01 | Similar to *dusA* tRNA-dihydrouridine(20/20a) synthase (*Vibrio vulnificus*) | 0.27 | GO:0008033, GO:0017150, GO:0050660, GO:0055114 |
| wild populations | Carg07889 | Csor.00g113710 | Chr07 | Similar to *SFH9* Phosphatidylinositol/phosphatidylcholine transfer protein *SFH9* (*Arabidopsis thaliana*) | 0.14 | NA |
| wild populations | Carg09452 | Csor.00g084200 | Chr17 | Similar to *ALA4* Probable phospholipid-transporting ATPase 4 (*Arabidopsis thaliana*) | 0.09 | GO:0000166, GO:0000287, GO:0005524, GO:0015914, GO:0016021, GO:0140326 |
| wild populations | Carg10718 | Csor.00g080500 | Chr01 | Similar to *FBL15* F-box/LRR-repeat protein 15 (*Arabidopsis thaliana*) | 0.16 | GO:0005515 |
| wild populations | Carg11621 | Csor.00g072250 | Chr02 | Similar to Zeaxanthin epoxidase, chloroplastic (*Prunus armeniaca*) | 0.12 | GO:0005515, GO:0009507, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | GO:0009540, GO:0009688, GO:0016020, GO:0055114, GO:0071949 |
| wild populations | Carg12374 | Csor.00g192590 | Chr01 | Similar to *PBL10* Probable serine/threonine-protein kinase *PBL10* (*Arabidopsis thaliana*) | 0.19 | GO:0004672, GO:0006468 |
| wild populations | Carg14512 | Csor.00g157750 | Chr18 | Similar to *MKP1* Protein-tyrosine-phosphatase *MKP1* (*Arabidopsis thaliana*) | 0.01 | GO:0008138, GO:0016311 |
| wild populations | Carg14932 | Csor.00g116960 | Chr07 | Similar to *VPS54* Vacuolar protein sorting-associated protein 54, chloroplastic (*Arabidopsis thaliana*) | 0.26 | GO:0005515, GO:0008080, GO:0042147 |
| wild populations | Carg15904 | Csor.00g236940 | Chr08 | Similar to *SGS3* Protein SUPPRESSOR OF GENE SILENCING 3 homolog (*Oryza sativa* subsp. *indica*) | 0.13 | GO:0031047 |
| wild populations | Carg15929 | Csor.00g237260 | Chr08 | Similar to At5g19025 Uncharacterized protein At5g19025 (*Arabidopsis thaliana*) | 0.28 | NA |
| wild populations | Carg18944 | Csor.00g084780 | Chr17 | Similar to *ATL3* RING-H2 finger protein *ATL3* (*Arabidopsis thaliana*) | 0.01 | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| wild populations | Carg22232 | Csor.00g242160 | Chr16 | Similar to *POP1* Ribonucleases P/MRP protein subunit *POP1* (*Homo sapiens*) | 0.1 | NA |
| wild populations | Carg23772 | Csor.00g220140 | Chr04 | Similar to *SAC3A SAC3* family protein A (*Arabidopsis thaliana*) | 0.14 | NA |
| wild populations | Carg23802 | NA | Chr14 | Similar to At1g06840 Probable LRR receptor-like serine/threonine-protein kinase At1g06840 (*Arabidopsis thaliana*) | 0.19 | GO:0005515 |
| wild populations | Carg24812 | NA | Chr16 | Similar to *PBL23* Probable serine/threonine-protein kinase *PBL23* (*Arabidopsis thaliana*) | 0.07 | GO:0004672, GO:0005524, GO:0006468 |
| wild populations | Carg25546 | Csor.00g002770 | Chr17 | Similar to *TMKL1* Putative kinase-like protein *TMKL1* (*Arabidopsis thaliana*) | 0.3 | GO:0004672, GO:0005515, GO:0006468 |
| wild populations | Carg25639 | Csor.00g029150 | Chr01 | Similar to *OVA7* Serine--tRNA ligase, chloroplastic/mitochondrial (*Arabidopsis thaliana*) | 0.12 | GO:0000166, GO:0004812, GO:0004828, GO:0005524, GO:0006418, GO:0006434 |
| wild populations | Carg26784 | NA | Chr15 | Protein of unknown function | 0.02 | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| wild populations | Carg26826 | Csor.00g030720 | Chr16 | Similar to *PA200* Proteasome activator subunit 4 (*Arabidopsis thaliana*) | 0.14 | NA |
| wild populations | Carg27622 | Csor.00g103900 | Chr20 | Similar to *SUMO2* Small ubiquitin-related modifier 2 (*Arabidopsis thaliana*) | 0.32 | NA |
| wild populations | Carg_TCONS_00026631 | NA | Chr13 | Long noncoding RNA | NA | NA |
| ABBA sites | Carg07674 | Csor.00g064770 | Chr13 | Similar to *apaG* Protein *ApaG* (*Magnetospirillum magneticum*) | 0.21 | GO:0005515 |
| ABBA sites | Carg26784 | NA | Chr15 | Protein of unknown function | 0.02 | NA |
| ABBA sites | Carg26826 | Csor.00g030720 | Chr16 | Similar to *PA200* Proteasome activator subunit 4 (*Arabidopsis thaliana*) | 0.14 | NA |
| BABA sites | Carg04520 | Csor.00g122340 | Chr04 | Similar to *RBCMT* Ribulose-1,5 bisphosphate carboxylase/oxygenase large subunit N-methyltransferase, chloroplastic (*Nicotiana tabacum*) | 0.23 | GO:0005515 |
| BABA sites | Carg20078 | Csor.00g227200 | Chr09 | Similar to *ABCE2* ABC transporter E family member 2 (*Arabidopsis thaliana*) | 0.12 | GO:0005524, GO:0016887 |
| BABA sites | Carg21397 | Csor.00g161930 | Chr13 | Similar to *AGD12* ADP-ribosylation factor GTPase-activating protein AGD12 (*Arabidopsis thaliana*) | 0.16 | GO:0005096 |
| BABA sites | Carg22875 | Csor.00g267800 | Chr15 | Similar to *SPBC3E7*.09 Uncharacterized protein slp1 (Schizosaccharomyces pombe) | 0.11 | NA |

## Supplementary Figures

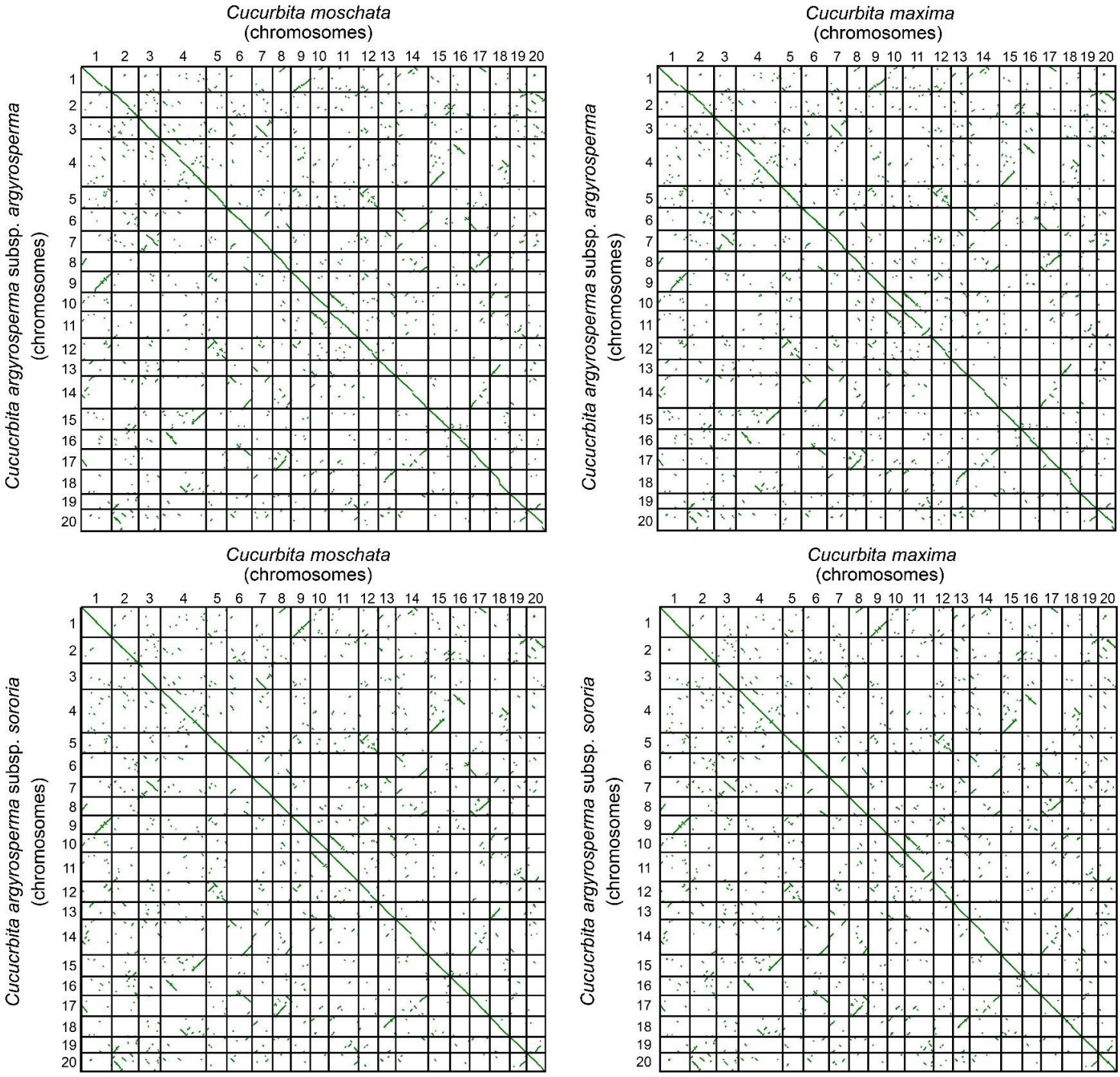



**Fig. S1. Synteny dot plots between the chromosome-level genome assemblies of *Cucurbita argyrosperma* subsp. *argyrosperma* and *C. argyrosperma* subsp. *sororia* against the reference genomes of *C. moschata* and *C. maxima* (Sun *et al*., 2017).** Most of the chromosomes show chromosome-wide homoeologous pairs within the genome assembly, which have been previously attributed to a whole-genome duplication event in the *Cucurbita* genus.

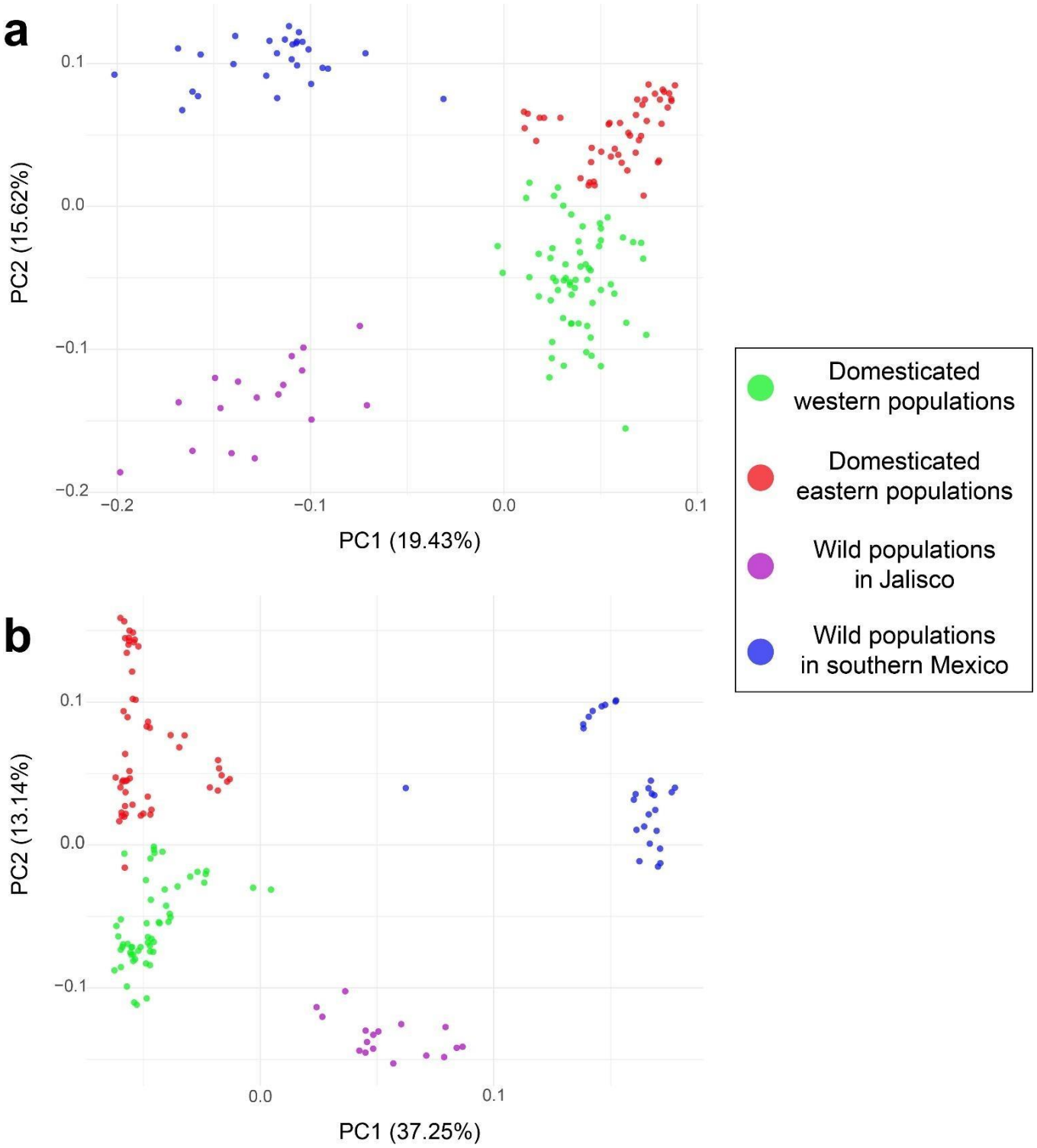



**Fig. S2. Principal Component Analysis (PCA) plots from data of *C. argyrosperma* subsp. *argyrosperma* (domesticated) and *C. argyrosperma* subsp. *sororia* (wild) populations.** The first two components were plotted for **a** the SNPs dataset used for demographic analyses (2,861 SNPs,153 individuals) and for **b** the SNP dataset used for the selection scans (10,617 SNPs, 153 individuals).

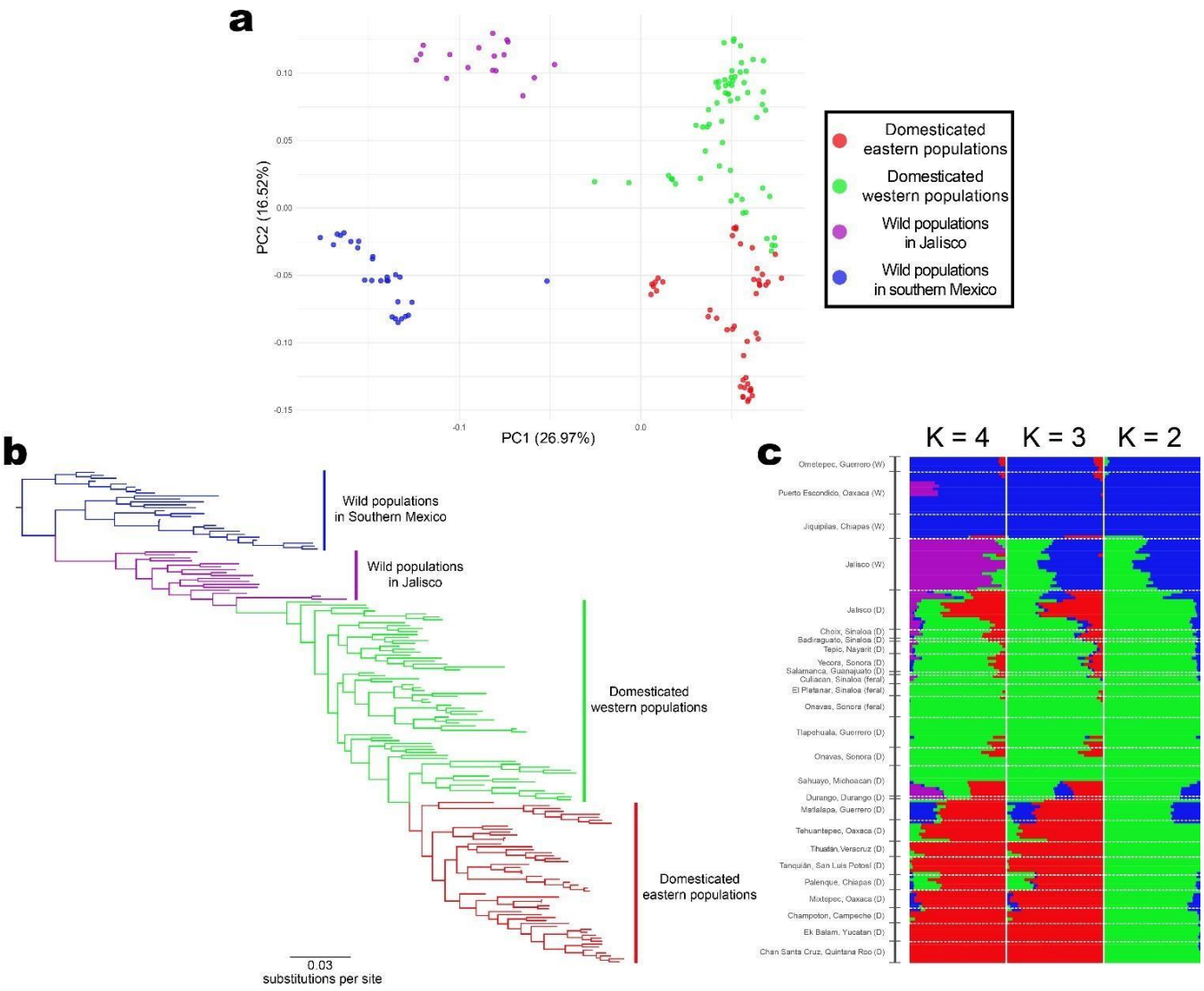



**Fig. S3. Genetic structure and phylogenetic relationships between the wild and domesticated populations of *Cucurbita argyrosperma* retrieved from an alternative SNP dataset.** The filtering step of 50% missing data was applied separately for the domesticated and wild populations. We also omitted the filter of the Hardy-Weinberg equilibrium exact test. **a** Principal component analysis. **b** Maximum Likelihood tree. **c** ADMIXTURE analysis using K values ranging from 2 to 4.

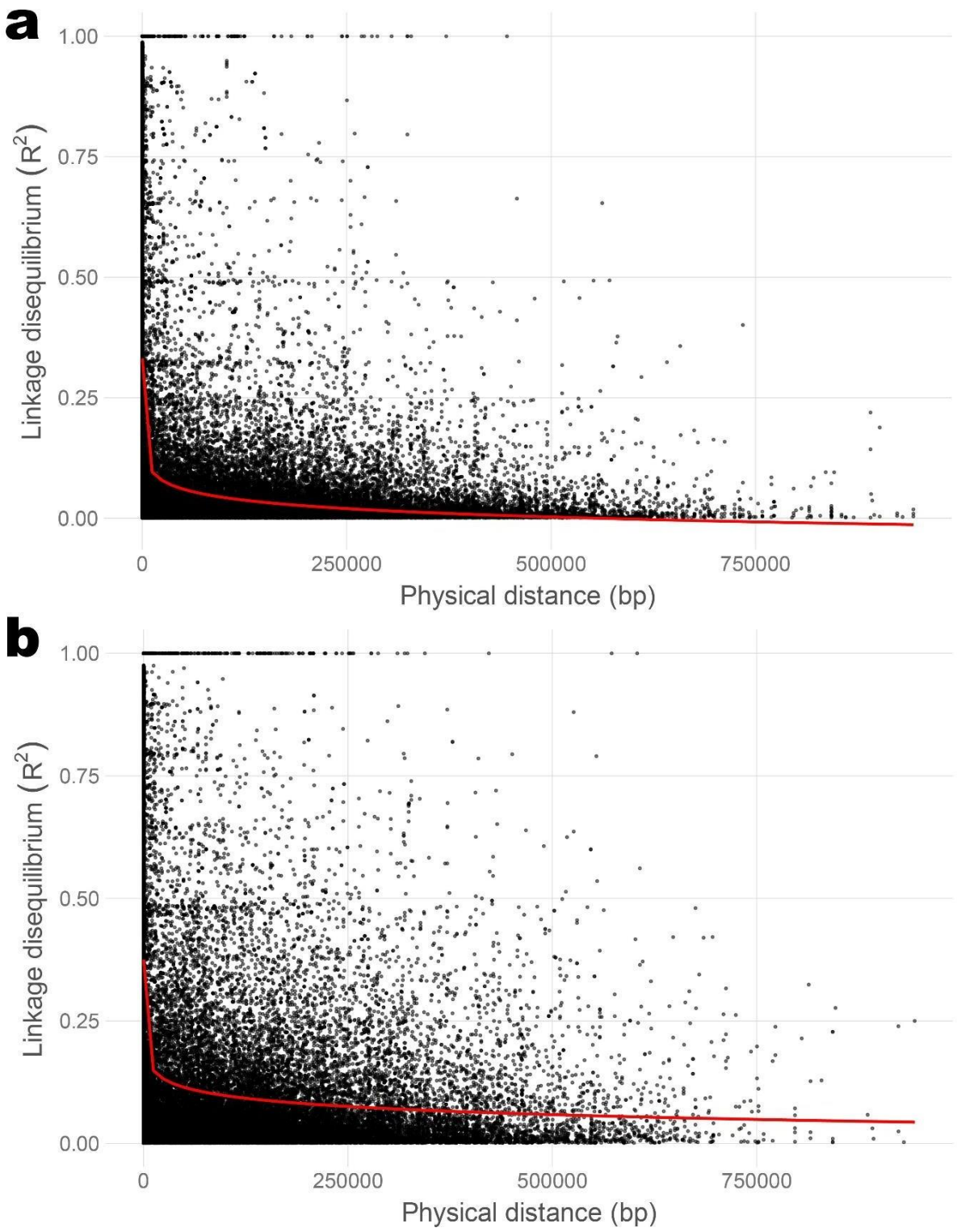



**Fig. S4. Linkage disequilibrium (LD) decay between the 10,617 SNPs used to perform the selective scans in *C. argyrosperma*. a** LD decay in *C. argyrosperma* subsp. *argyrosperma*. **b** LD decay in *C. argyrosperma* subsp. *sororia*.

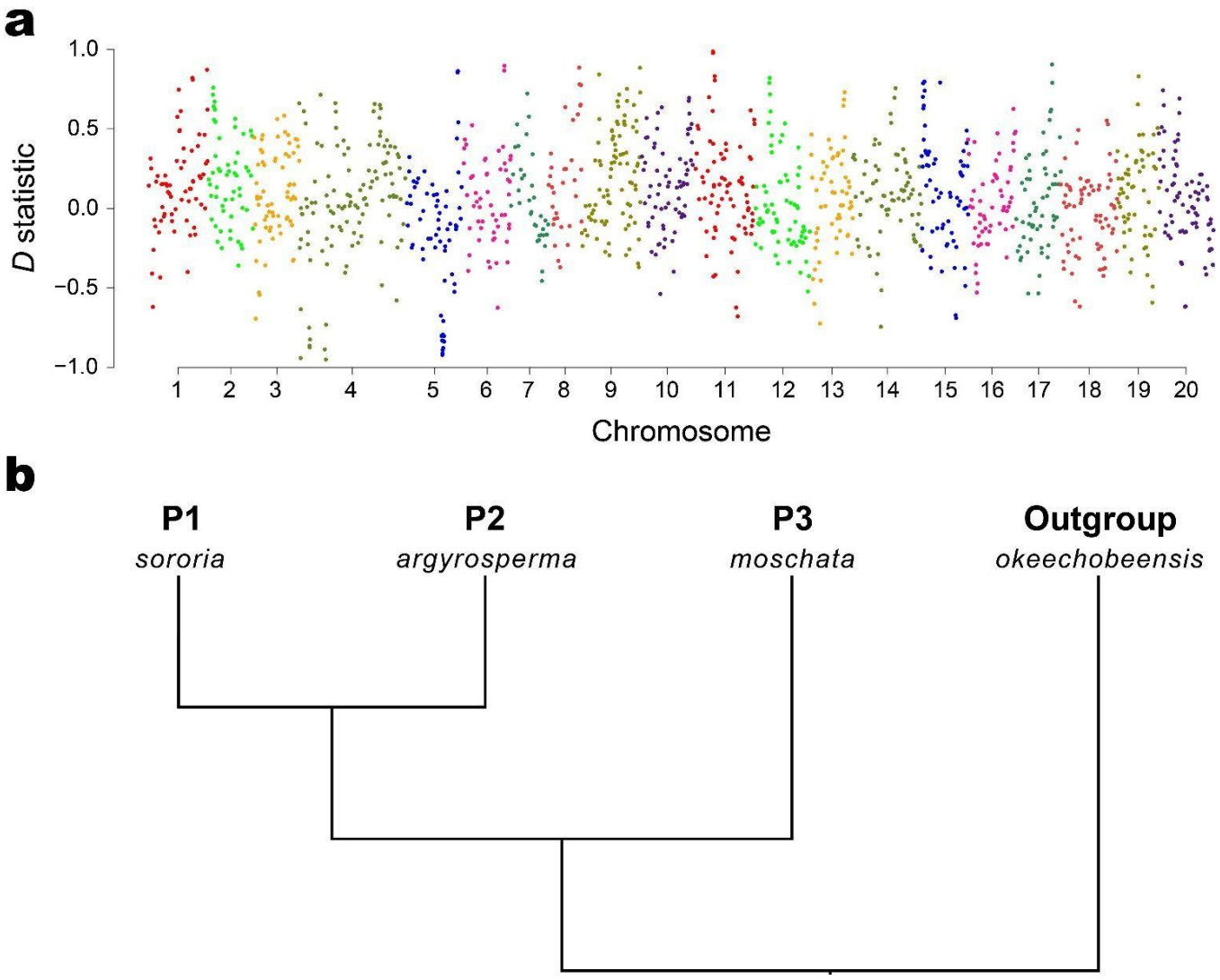



**Fig. S5. ABBA-BABA test to detect introgression using 11,498,421 variants between *C. argyrosperma* subsp. *sororia* (P1), *C. argyrosperma* subp. *argyrosperma* (P2) and *C. moschata* (P3), while using *C. okeechobeensis* subsp. *martinezii* as an outgroup.** **a** Manhattan plot of the *D*-statistic throughout the genome using 500 SNP windows with a step size of 250 SNPs. **b** Phylogenetic topology used to perform the ABBA-BABA test, as reconstructed by Dsuite.

## Supplementary Data

**Data S1: Coalescent simulation parameters of domestication in Jalisco with constant gene flow (Model 1).**

*//Number of population samples (demes)*
*3 samples to simulate :*
*//Population effective sizes (number of genes)*
*WILD_JALISCO_NPOP*
*WILD_SOUTH_NPOP*
*DOMESTICATED_NPOP*
*//Sample sizes*
*12*
*26*
*26*
*//Growth rates : negative growth implies population expansion*
*0*
*0*
*0*
*//Number of migration matrices: 0 implies no migration between demes*
*3*
*//Migration 0*
*0.0000 MIG10 MIG20*
*MIG01 0.0000 MIG21*
*MIG02 MIG12 0.0000*
*//Migration1*
*0.000 MIG10 0.000*
*MIG01 0.000 0.000*
*0.000 0.000 0.000*
*//Migration2*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*//historical event: time, source, sink, migrants, new size, growth rate, migr. matrix*
*2 historical event*
*TDOMESTICATED 2 0 1 RESDOM 0 1*
*TWILDSOUTH 1 0 1 RESWILDS 0 2*
*//Number of independent loci*
*1 0*
*//Per chromosome: Number of linkage blocks*
*1*
*//per Block: data type, num loci, rec. rate and mut rate + optional parameters*
*FREQ 1 0 2.5e-8 OUTEXP*

**Data S2: Coalescent simulation parameters of domestication in Jalisco with secondary contact (Model 2).**

*//Number of population samples (demes)*
*3 samples to simulate :*
*//Population effective sizes (number of genes)*
*WILD_JALISCO_NPOP*
*WILD_SOUTH_NPOP*
*DOMESTICATED_NPOP*
*//Sample sizes*
*12*
*26*
*26*
*//Growth rates : negative growth implies population expansion*
*0*
*0*
*0*
*//Number of migration matrices: 0 implies no migration between demes*
*3*
*//Migration 0*
*0.0000 MIG10 MIG20*
*MIG01 0.0000 MIG21*
*MIG02 MIG12 0.0000*
*//Migration1*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*//Migration2*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*//historical event: time, source, sink, migrants, new size, growth rate, migr. matrix*
*2 historical event*
*TDOMESTICATED 2 0 1 RESDOM 0 1*
*TWILDSOUTH 1 0 1 RESWILDS 0 2*
*//Number of independent loci*
*1 0*
*//Per chromosome: Number of linkage blocks*
*1*
*//per Block: data type, num loci, rec. rate and mut rate + optional parameters*
*FREQ 1 0 2.5e-8 OUTEXP*

**Data S3: Coalescent simulation parameters of domestication in Jalisco with no gene flow (Model 3).**

*//Number of population samples (demes)*
*3 samples to simulate :*
*//Population effective sizes (number of genes)*
*WILD_JALISCO_NPOP*
*WILD_SOUTH_NPOP*
*DOMESTICATED_NPOP*
*//Sample sizes*
*12*
*26*
*26*
*//Growth rates : negative growth implies population expansion*
*0*
*0*
*0*
*//Number of migration matrices: 0 implies no migration between demes*
*3*
*//Migration 0*
*0.0000 0.0000 0.0000*
*0.0000 0.0000 0.0000*
*0.0000 0.0000 0.0000*
*//Migration1*
*0.0000 0.0000 0.0000*
*0.0000 0.0000 0.0000*
*0.0000 0.0000 0.0000*
*//Migration2*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*//historical event: time, source, sink, migrants, new size, growth rate, migr. matrix*
*2 historical event*
*TDOMESTICATED 2 0 1 RESDOM 0 1*
*TWILDSOUTH 1 0 1 RESWILDS 0 2*
*//Number of independent loci*
*1 0*
*//Per chromosome: Number of linkage blocks*
*1*
*//per Block: data type, num loci, rec. rate and mut rate + optional parameters*
*FREQ 1 0 2.5e-8 OUTEXP*

**Data S4: Coalescent simulation parameters of domestication in southern Mexico with constant gene flow (Model 4).**

*//Number of population samples (demes)*
*3 samples to simulate :*
*//Population effective sizes (number of genes)*
*WILD_SOUTH_NPOP*
*WILD_JALISCO_NPOP*
*DOMESTICATED_NPOP*
*//Sample sizes*
*12*
*26*
*26*
*//Growth rates : negative growth implies population expansion*
*0*
*0*
*0*
*//Number of migration matrices: 0 implies no migration between demes*
*3*
*//Migration 0*
*0.0000 MIG10 MIG20*
*MIG01 0.0000 MIG21*
*MIG02 MIG12 0.0000*
*//Migration1*
*0.000 MIG10 0.000*
*MIG01 0.000 0.000*
*0.000 0.000 0.000*
*//Migration2*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*//historical event: time, source, sink, migrants, new size, growth rate, migr. matrix*
*2 historical event*
*TDOMESTICATED 2 0 1 RESDOM 0 1*
*TWILDJALISCO 1 0 1 RESWILDJ 0 2*
*//Number of independent loci*
*1 0*
*//Per chromosome: Number of linkage blocks*
*1*
*//per Block: data type, num loci, rec. rate and mut rate + optional parameters*
*FREQ 1 0 2.5e-8 OUTEXP*

**Data S5: Coalescent simulation parameters of domestication in southern Mexico with secondary contact (Model 5).**

*//Number of population samples (demes)*
*3 samples to simulate :*
*//Population effective sizes (number of genes)*
*WILD_SOUTH_NPOP*
*WILD_JALISCO_NPOP*
*DOMESTICATED_NPOP*
*//Sample sizes*
*12*
*26*
*26*
*//Growth rates : negative growth implies population expansion*
*0*
*0*
*0*
*//Number of migration matrices: 0 implies no migration between demes*
*3*
*//Migration 0*
*0.0000 MIG10 MIG20*
*MIG01 0.0000 MIG21*
*MIG02 MIG12 0.0000*
*//Migration1*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*//Migration2*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*//historical event: time, source, sink, migrants, new size, growth rate, migr. matrix*
*2 historical event*
*TDOMESTICATED 2 0 1 RESDOM 0 1*
*TWILDJALISCO 1 0 1 RESWILDJ 0 2*
*//Number of independent loci*
*1 0*
*//Per chromosome: Number of linkage blocks*
*1*
*//per Block: data type, num loci, rec. rate and mut rate + optional parameters*
*FREQ 1 0 2.5e-8 OUTEXP*

**Data S6: Coalescent simulation parameters of domestication in southern Mexico with no gene flow (Model 6).**

*//Number of population samples (demes)*
*3 samples to simulate :*
*//Population effective sizes (number of genes)*
*WILD_SOUTH_NPOP*
*WILD_JALISCO_NPOP*
*DOMESTICATED_NPOP*
*//Sample sizes*
*12*
*26*
*26*
*//Growth rates : negative growth implies population expansion*
*0*
*0*
*0*
*//Number of migration matrices: 0 implies no migration between demes*
*3*
*//Migration 0*
*0.0000 0.0000 0.0000*
*0.0000 0.0000 0.0000*
*0.0000 0.0000 0.0000*
*//Migration1*
*0.0000 0.0000 0.0000*
*0.0000 0.0000 0.0000*
*0.0000 0.0000 0.0000*
*//Migration2*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*0.000 0.000 0.000*
*//historical event: time, source, sink, migrants, new size, growth rate, migr. matrix*
*2 historical event*
*TDOMESTICATED 2 0 1 RESDOM 0 1*
*TWILDJALISCO 1 0 1 RESWILDJ 0 2*
*//Number of independent loci*
*1 0*
*//Per chromosome: Number of linkage blocks*
*1*
*//per Block: data type, num loci, rec. rate and mut rate + optional parameters*
*FREQ 1 0 2.5e-8 OUTEXP*

# DISCUSIÓN GENERAL

El presente estudio representa el primer análisis de variación genética, estructura genética y flujo génico que cuenta con poblaciones de ambas subespecies de *C. argyrosperma* en toda su área de distribución en México. Los análisis realizados nos muestran cómo integrar la información genómica, arqueológica y ecológica disponible para describir un escenario concreto de cómo pudo ocurrir la domesticación en Mesoamérica (Zeder *et al.,* 2006; Barrera-Redondo *et al.,* 2021).

Este proyecto de tesis tenía el objetivo de describir el estado actual de las poblaciones y llenar de la mejor manera posible los vacíos de información sobre la diversidad genética de una de las especies menos conocidas del género, *C. argyrosperma* y como se distribuye en México.

Al comienzo se planificó la solicitud y posterior análisis de diferentes accesiones de semillas de ambas subespecies de *C. argyrosperma* al Banco de Germoplasma INIFAP-Bajío y que habían sido utilizados en otros trabajos previos (Montes-Hernández y Eguiarte, 2002), pero debido a la baja viabilidad que mostraron las semillas de las accesiones se tomó la decisión de colectar frutos para tener una mejor representación de la distribución de la especie. De esta forma, se actualizaron las colectas de semillas en diferentes estados y poblaciones de las dos subespecies de *C. argyrosperma.*

Para esta tesis se colectaron en 14 estados del país un total de más de 60 accesiones para ambas subespecies. Este muestreo a nivel nacional es relevante, ya que los demás estudios que han analizado el germoplasma de la especie en México solo utilizaron colectas a nivel local o regional (Montes-Hernández y Eguiarte, 2002; Cerón *et al.*, 2010; Priori *et al.,* 2018) o solo en unos casos analizaron muestras de varios estados (Cruz-Reyes *et al.,* 2015; Balvino-Olvera *et al.,* 2017).

Cabe resaltar que varias de nuestras nuevas colectas de *C. argyrosperma* ya se han utilizado en otros tres estudios además de las publicaciones incluidas en esta tesis (Barrera-Redondo *et al.,* 2019; Tapia, 2021; Kates *et al.,* 2021). Consideramos que las colectas que

describimos en esta tesis representan un notable esfuerzo para conocer a los taxa y para su futura conservación *ex situ,* en particular considerando que para las colecciones previas del germoplasma de *C. argyrosperma*, la viabilidad de semillas es desconocida y se han enfocado principalmente en la subespecie cultivada.

## Baja diversidad genética en marcadores de cloroplasto y mitocondria en *C. argyrosperma*

Inicialmente, buscando analizar la diversidad presente en *C. argyrosperma* mediante el uso de marcadores de cloroplasto, amplificamos un subconjunto de individuos de diferentes poblaciones para evaluar la variación genética en cuatro regiones de cloroplasto: *trnL-trnF* (Taberlet *et al.,* 1991), *rpl20-rps12* (Hamilton, 1999), *psbJ-petA y psbD-trnT* (Shaw *et al.,* 2007). Al realizar un análisis BLAST en GenBank para corroborar que las secuencias obtenidas correspondían a las regiones esperadas del cloroplasto, los resultados nos sugirieron que la región amplificada de *trnL-trnF* correspondía a una región duplicada e invertida en el genoma mitocondrial de *Cucurbita* (Alverson *et al.,* 2010;), mientras que las secuencias de *rpl20-rps12*, *psbJ-petA* y *psbD-trnT* si correspondían a secuencias de cloroplasto de *Cucurbita*, como se describe en Aguirre-Dugua *et al.* (2019).

Los análisis de las diferentes regiones que si se encontraban codificadas en el genoma del cloroplasto mostraron niveles muy bajos de variación en ambas subespecies, con solo tres sitios segregantes en un total de 2,442 pb en 57 individuos. Tampoco la región mitocondrial duplicada *trnL-trmF* fue muy variable, ya que se detectó un solo sitio segregante en 977 pb en 25 individuos analizados. Por lo tanto, se decidió no usar estas regiones y continuar los análisis solo con microsatélites y con marcadores genómicos (GBS) tipo SNPs. Por otra parte, recomendamos que estas regiones del cloroplasto y mitocondria sean utilizadas con precaución en análisis filogeográficos y filogenéticos en *Cucurbita,* debido a la presencia de parálogos mitocondriales de origen plastídico, que pueden tener altas tasas de sustitución de nucleótidos, junto a regiones mitocondriales con bajas tasas de mutación (Aguirre-Dugua *et al.,* 2019).

## Estado de la diversidad genética de *C. argyrosperma*

En términos generales, los análisis tanto con microsatélites como con SNPs, muestran niveles similares de variación genética en ambas subespecies; aunque es común que las especies cultivadas presenten valores más bajos de variación genética (Gaut *et al.,* 2015). Lo anterior sugiere que es posible que ambas subespecies hayan mantenido tamaños de población efectivos similares y que el cuello de botella asociado teóricamente con la domesticación fue leve o de corta duración, seguido de una rápida expansión de las poblaciones (Hedrick, 2011). Adicionalmente, diferentes estudios han reportado flujo génico entre las dos subespecies a nivel local y con otras especies del género en poblaciones simpátricas (Nabhan, 1984; Decker-Walters *et al.,* 1990; Wilson, 1990; Wilson *et al.,* 1994; Montes-Hernández y Eguiarte, 2002). El flujo génico durante largos periodos también fue sugerido por nuestras simulaciones de coalescencia, como se presenta en el Capítulo 3 de esta tesis.

Nuestros análisis de $F_{ST}$ por pares, junto con los resultados de diferenciación genética (Capítulos 2 y 3), sugieren que ambas subespecies están bien diferenciadas genéticamente. Esto es relevante para estas subespecies, debido a que los valores de $F_{ST}$ deben estar inversamente relacionados con el grado de semejanza fenotípica presente entre sus poblaciones y proporcionan información sobre su historia demográfica (Holsinger y Weir, 2009).

Lo valores de diversidad genética obtenidos no mostraron un patrón claro, pues poblaciones con diversidad alta o baja se encuentran en diferentes latitudes, regiones geográficas o elevaciones, aunque algunos de los valores mayores ($H_e$, *Pi* ) se encuentran en poblaciones de Jalisco, Michoacán y la parte oeste de Guerrero, en las regiones de influencia de los ríos Lerma y Balsas, donde conviven poblaciones domesticadas de *argyrosperma* con las silvestre de *sororia*. Cabe resaltar que las poblaciones de la Península de Yucatán tuvieron valores generalmente intermedios, a pesar de que no conviven con poblaciones de *sororia*, pero si con las especies cultivadas *C. moschata* y *C. pepo,* donde se ha reportado alta diversidad (Hernández-Rosales *et al.,* 2020; Castellanos-Morales *et al.,*

2019; Martínez-González *et al.,* 2021) y con quienes se ha documentado la existencia de flujo génico (Wilson *et al.*, 1994; Hernández y Eguiarte, 2002).

La variación genética estimada con microsatélites fue similar a la reportada en otras plantas anuales, pero menor en comparación con otras especies con reproducción cruzada (Nybom, 2004). Además, el número medio de alelos en *C. argyrosperma* fue menor que los reportados usando microsatélites nucleares en *C. pepo* (Formisano *et al.,* 2012; Gong *et al.,* 2012, 2013; Priori *et al.,* 2018; Ntuli *et al.,* 2015; Castellanos-Morales *et al.,* 2019) y en *C. moschata* (Hernández-Rosales *et al.,* 2020). También fueron menores a los reportados en poblaciones silvestres de *sororia* a lo largo de la costa oeste de México (Balvino-Olvera *et al.,* 2017) y en Jalisco con ambas subespecies (Priori *et al.,* 2018).

Los valores altos de diversidad en *C. pepo* y en *C. moschata* han sido asociados también a un elevado flujo de genes entre poblaciones, debido a la actividad de sus polinizadores especializados, las abejas de los géneros *Peponapsis* y *Xenoglossa* (Hurd *et al.,* 1971; López-Uribe *et al.,* 2016). Además, ciertos aspectos del manejo agrícola--principalmente los asociados al sistema de policultivo "milpa" -- permiten que convivan diferentes especies de calabazas (Lira *et al.,* 2016). También existe un intercambio de germoplasma muy activo por agricultores en la región (Montes-Hernández *et al.,* 2005). Algunos aspectos del manejo agrícola, como el intercambio de semillas adicionalmente podrían afectar los niveles de variación genética en *C. argyrosperma* (Montes-Hernández *et al.,* 2005), como ocurre en *C. pepo* (Enríquez-Cotton, 2017; Enríquez *et al.,* 2017).

Es importante resaltar que se ha propuesto que las prácticas agrícolas tradicionales son otra fuerza fundamental que mantiene la diversidad de especies de cultivos (Jarvis *et al.,* 2008), particularmente en México, debido a que fomentan los policultivos y que bajo diferentes estrategias pueden tolerar y fomentar diferentes especies que son útiles. *Cucurbita argyrosperma* subsp. *argyrosperma* es un cultivo tradicional tanto para el autoconsumo como para comerciar en los mercados locales, principalmente las semillas, donde tiene un nicho gastronómico especializado (Lira *et al.,* 2016). Por ello, la diversidad

genética de *argyrosperma* se mantiene mediante la conservación de variedades locales (Montes-Hernández *et al.,* 2005; Barrera-Redondo *et al.,* 2020).

## Distribución de la diversidad genética de *C. argyrosperma*

Al evaluar la estructura genética y las relaciones filogenéticas entre ambas subespecies de *C. argyrosperma,* se detectó que existe diferenciación genética entre ellas con los análisis de $F_{ST}$, tanto de microsatélites como SNPs, y que además la especie tiene un origen monofilético, y aparentemente solo fue domesticada una vez, siendo la subespecie *sororia* el ancestro silvestre. Debido a que las poblaciones de las dos subespecies tienen una distribución amplia, es probable que la deriva génica este promoviendo la fijación de alelos en poblaciones que tengan cierto grado de aislamiento, como lo muestra la presencia de barreras geográficas en el análisis realizado con microsatélites. Es importante remarcar el papel de las barreras geográficas existentes como el istmo de Tehuantepec (Ornelas *et al.,* 2013) y la Sierra Madre Oriental (Aguirre-Liguori *et al.,* 2017), que regulan la dispersión de especies como *C. argyrosperma*.

Los grupos genéticos que se detectan tanto con microsatélites como con SNPs, corresponden en general con regiones geográficas bien definidas, de forma mayormente latitudinal, como ya había sido descrito anteriormente para la subespecie *sororia* (Balvino-Olvera *et al.,* 2017). Esta última subespecie en su distribución a lo largo de la costa del Pacífico por más de 1,000 km., se dividide entre el grupo de las poblaciones del Occidente (Jalisco y Nayarit) y el grupo con poblaciones del Sur en Guerrero, Oaxaca y Chiapas. La gran distancia que existe en la distribución natural de *C.a. sororia* a lo largo de la costa del Pacifico ha promovido su diferenciación. Las poblaciones de Sinaloa y Sonora que inicialmente se consideraba eran *sororia*, de acuerdo con nuestros análisis con ambas estrategias moleculares (microsatélites y SNPs) resultaron ser ferales de *argyrosperma.* Probablemente estas plantas ferales son descendientes de poblaciones cultivadas que tal vez han comenzado a aislarse de las poblaciones cultivadas, como lo muestran los valores de diferenciación que son los más elevados.

Es interesante que los grupos genéticos, particularmente de la subsp. *argyrosperma,* corresponden con algunos de los grupos genéticos reportados para grupos humanos nativos de México (Moreno-Estrada *et al.,* 2014). Adicionalmente, los patrones de dispersión de grupos humanos en los últimos 10,000 años en México son similares a los propuestos para la dispersión gradual de otros cultivos como maíz, frijol y otras especies de calabaza (Zizumbo-Villarreal y Colunga-GarcíaMarín, 2010). Lo anterior sugiere que aspectos culturales pueden ser determinantes para la estructura genética de este cultivo, por lo que futuros análisis deben encaminarse a buscar correlaciones entre grupos humanos y grupos genéticos de cultivos como *C. argyrosperma.*

## La domesticación de *Cucurbita argyrosperma* subsp. *argyrosperma*

Los modelos de distribución (SDM) de parientes silvestres de taxones domesticados son una herramienta útil para corroborar hipótesis de posibles sitios de domesticación y adecuación ambiental en el pasado (Hufford *et al.,* 2012; Besnard *et al.,* 2013). Nuestros resultados de SDM indican que las características ambientales en la región de Balsas-Jalisco, como potencial centro de domesticación, fueron adecuadas para la presencia de *sororia* durante el Holoceno y que han estado relativamente estables desde entonces.

Nuestros análisis demográficos y de coalescencia nos permitieron inferir algunos aspectos del proceso de domesticación de *C argyrosperma*, con variables evolutivas relevantes. Se encontró que las poblaciones existentes de Jalisco son los parientes modernos más cercanos a la población ancestral de *sororia* de la que se originó *argyrosperma* y a partir de las cuales se desarrollaron los demás grupos genéticos. La relación genética entre las poblaciones de *sororia* del occidente de México y *argyrosperma* también se ha sido reportada anteriormente con marcadores mitocondriales (Sanjur *et al.,* 2002) y con microsatélites (Sánchez-de la Vega *et al.,* 2018).

Nuestros resultados apoyan que la domesticación de *C. argyrosperma* probablemente comenzó hace poco más de 8,700 años en el occidente de México, en la región Jalisco-Balsas, como sugiere el registro arqueológico de *argyrosperma,* que aunque es taxonómicamente ambiguo, nos proporciona una buena idea de cómo pudo ser el

proceso, junto con el fechamiento de otros cultivos como maíz, con tal vez mayor antigüedad (Piperno *et al.,* 2009, Ranere *et al.,* 2009).

Al mismo tiempo se ha propuesto que ocurrieron eventos de migración humana en el pasado a lo largo de las cuencas de los ríos en el Occidente y el Suroeste de México, a manera de corredores bioculturales (Zizumbo-Villarreal y Colunga-GarcíaMarín, 2010). Aunado a lo anterior, los patrones de estructura genética de *C. argyrosperma* son coherentes con la evidencia arqueológica de la migración humana temprana a lo largo de Mesoamérica (Stinnesbeck *et al.,* 2017; Piperno, 2011). Lo anterior puede explicar la cohesión genética entre las poblaciones de *argyrosperma* de esa área, que representan el primer linaje completamente domesticado de la especie, y que el flujo génico, aunque está presente, ha ido disminuyendo a nivel regional, promoviendo su divergencia desde que comenzó la domesticación.

## Consideraciones finales

Aunado a recientes estudios producidos en el laboratorio con esta especie (Barrera-Redondo., *et al*., 2019; Tapia, 2021), el presente trabajo constituye un aporte importante para entender la dinámica evolutiva y los recursos genéticos de *C. argyrosperma* en México. Nos permite visualizar como se distribuye su diversidad genética y nos ayuda a comprender como pudo haber ocurrido su proceso de domesticación. Todo esto es relevante dada la importancia alimenticia y comercial de esta especie para muchas comunidades de México.

En conjunto, los resultados muestran que la diversidad genética en ambas subespecies es muy similar. Por otra parte, se encontró que existe estructura en sus poblaciones, formándose diferentes grupos y se confirma un origen monofilético de esta especie (Sánchez-de la Vega *et al.*, 2018; Barrera-Redondo *et al*., 2021). Adicionalmente, se propone que la subespecie *argyrosperma* pudo haber sido domesticada en las tierras bajas de Jalisco, basándonos en modelos de coalescencia (Barrera-Redondo *et al*., 2021). También se encontró evidencia de flujo génico entre las subespecies *argyrosperma* y *sororia*, lo que probablemente alivió los efectos del cuello de botella de la domesticación, esto es de suma

importancia tomando en cuenta que estos procesos demográficos han ocurrido de manera relativamente reciente (Purugganan y Fuller, 2009; Sánchez-de la Vega *et al.*, 2018).

A futuro y para aumentar el conocimiento sobre *C. argyrosperma*, es necesario llenar los vacíos en la colecta de semillas en aquellos sitios de los que se realizó parcialmente, como Michoacán y Guerrero, entre otros estados. Además es importante mantener en buen estado los bancos de germoplasma, que permitan mejorar y conservar los recursos fitogenéticos a nivel regional.

Considero que son necesarios análisis a nivel local que generen información sobre flujo génico entre poblaciones domesticadas y silvestres. Además serían importantes análisis de tipo GWAS de frutos y particularmente los análisis de pangenomas deben ser los siguientes en los estudios de domesticación en calabazas con enfoques de genómica poblacional. Todo lo anterior con fines de conservación y mejoramiento de este importante recurso fitogénetico.

## CONCLUSIONES

- El presente trabajo es el primero que analiza la variación genética de ambas subespecies de *Cucurbita argyrosperma: Cucurbita argyrosperma* subsp. sororia (silvestre) y *Cucurbita argyrosperma* subesp. a*rgyrosperma* (cultivada en toda su distribución en México utilizando microsatélites y datos genómicos (SNPs derivados de GBS).

- La diversidad genética *C. argyrosperma,* medida tanto con microsatélites como con SNPs, es menor a reportada en otras especies del género.

- Ambas subespecies tienen una diversidad genética similar, presentan flujo génico a nivel local entre ellas y están diferenciadas genéticamente. Los valores de diversidad genética no tienen un patrón geográfico ni altitudinal claro.

- Existen grupos genéticos para cada subespecie. En la subespecie *sororia* encontramos dos grupos genéticos, uno de poblaciones al sur de Jalisco y otro con poblaciones al norte a lo largo de su distribución en la costa del Pacífico. La subespecie a*rgyrosperma* tiene también dos principales grupos genéticos: uno al sur-sureste de Jalisco y otros al occidente-norte del país.

- Las poblaciones cultivadas tienen un patrón geográfico similar a algunos grupos humanos en México. De igual forma los patrones históricos de dispersión son similares a los que se desarrollaron en grupos humanos durante los últimos 8,000 años.

- Las poblaciones de la subespecie *sororia* en la región Jalisco-Balsas son las más cercanas genéticamente a la subespecie cultivada *argyrosperma*, por lo que posiblemente esa región fue el centro de domesticación de la especie.

- Es necesario establecer estrategias de conservación *in situ y ex situ*, tanto de las poblaciones silvestres como domesticadas, principalmente de las variedades locales. Esto es vital para mantener la diversidad genética de *C. argyrosperma*, que constituye un recurso fitogenético de gran importancia.

## REFERENCIAS BIBLIOGRÁFICAS


1. Abbo, S., Lev-Yadun, S., Gopher, A. 2010. Agricultural origins: Centers and noncenters; a Near Eastern reappraisal. *Crit. Rev, Plant Sci* ,29(5):317–328.

2. Alverson, A. J., Wei, X., Rice, D. W., Stern, D. B., Barry, K., & Palmer, J. D. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution*, 27(6), 1436-1448.

3. Aguirre-Dugua, X., Castellanos-Morales, G., Paredes-Torres, L. M., Hernández-Rosales, H. S., Barrera-Redondo, J., Sánchez-de la Vega, G., ... & Eguiarte, L. E. (2019). Evolutionary dynamics of transferred sequences between organellar genomes in *Cucurbita*. *Journal of Molecular Evolution*, 87(9), 327-342.

4. Aguirre-Liguori, J. A., M. I. Tenaillon, A. Vázquez-Lobo, B. S. Gaut, J. P. Jaramillo-Correa, S. Montes-Hernandez, V. Souza, y L. E. Eguiarte. 2017. Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Molecular Ecology*. 26: 4226-4240

5. Balvino-Olvera, F. J., Sánchez-Gómez, K. F., Lobo, J. A., Avila-Sakar, G., Cruz-Reyes, R., Sánchez-Montoya, G., ... & Quesada, M. 2017. Latitudinal structured populations of the Mexican wild squash *Cucurbita argyrosperma* subsp. *sororia* revealed by microsatellite markers. *Crop and Pasture Science*, 68(9), 850-858.

6. Barrera-Redondo, J., Ibarra-Laclette, E., Vázquez-Lobo, A., Gutiérrez-Guerrero, Y. T., Sánchez-de la Vega, G., Piñero, D., ... & Eguiarte, L. E. 2019. The genome of *Cucurbita argyrosperma* (silver-seed gourd) reveals faster rates of protein-coding gene and long noncoding RNA turnover and neofunctionalization within *Cucurbita. Molecular Plant*, 12(4), 506-520.

7. Barrera-Redondo, J., H. S. Hernández-Rosales, V. Cañedo-Torres, K. Aréstegui-Alegría, J. Torres-Guevara, F. Parra, I. Torres-García, y A. Casas. 2020. Landrace diversity and local selection criteria of domesticated squashes and gourds (*Cucurbita*) in the Central Andean mountain range of Peru: Tomayquichua, Huánuco. *Botanical Sciences*. 98: 101-116.

8. Barrera-Redondo, J., Sanchez-de La Vega, G., Aguirre-Liguori, J. A., Castellanos-Morales, G., Gutiérrez-Guerrero, Y. T., Aguirre-Dugua, X., ... & Eguiarte, L. E. 2021. The domestication of *Cucurbita argyrosperma* as revealed by the genome of its wild relative. *Horticulture Research,* 8(1), 1-14.

9. Bar-Yosef O. and Price D, eds. 2011. The Beginnings of Agriculture: New Data, New Ideas. *Current Anthropology* 52 (Supp 4). Wenner-Gren Symposium Series; Univ of Chicago Press, Chicago.

10. Besnard, G., Khandari, B., Navascués, M., Fernánez-Mazuecos, M., El Bakkali, A., Arrigo, N., et al. 2013. The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proc. R. Soc. B* 280:201228333.

11. Castellanos-Morales, G., Paredes-Torres, L. M., Gámez, N., Hernández-Rosales, H. S., Sánchez-de la Vega, G., Barrera-Redondo, J., et al. 2018. Historical biogeography and phylogeny of *Cucurbita*: insights from ancestral area reconstruction and niche evolution. *Mol. Phylogenet. Evol.* 128, 38–54.

12. Castellanos-Morales, G., Ruiz-Mondragón, K. Y., Hernández-Rosales, H. S., Sánchez-de la Vega, G., Gámez, N., Aguirre-Planter, E., *et al.* 2019. Tracing back the origin of pumpkins (*Cucurbita pepo ssp. pepo* L.) in Mexico. *Proc. R. Soc. B Biol. Sci.* 286:20191440.

13. Cerón-González L., Legaria-Solano J.P., Villanueva-Verduzco C., Sahagún-Castellanos, J. 2010. Diversidad genética en cuatro especies mexicanas de calabaza (*Cucurbita spp.). Rev Fitotec Mex*. 33:189–96.

14. Contreras-Toledo, A. R., Cortés-Cruz, M., Costich, D. E., de Lourdes Rico-Arce, M., Brehm, J. M., & Maxted, N. 2019. Diversity and conservation priorities of crop wild relatives in Mexico. *Plant Genetic Resources*, 17(2), 140-150.

15. Cruz-Reyes R, Ávila Sakar G, Sanchez-Montoya G, Quesada M. 2015. Experimental assessment of gene flow between transgenic squash and a wild relative in the center of origin of Cucurbits. *Ecosphere*. 6(12): art248.

16. Decker-Walters, D.S., T.W. Walters, U. Poluszny, P.G. Kevan. 1990. Genealogy and gene flow among annual domesticated species of *Cucurbita. Can. J. Bot*. 68:782-789.

17. Doebley, J., Gaut, B., Smith, B.D. 2006. The molecular genetics of crop domestication. Cell 127, 1309–1321.

18. Enríquez, E., Landaverde-González, P., Lima-Cordón, R., Solórzano-Ortíz, E., Tapia-López, R., and Nuñez-Farfán, J. 2017. Population genetics of traditional landraces of *Cucurbita pepo* L., 1753 in the cloud forest in Baja Verapaz, Guatemala. *Genet. Resour. Crop Evol*. 65, 979–991.

19. Enríquez Cotton, M. E. 2017. Efecto de la Estructura del Paisaje Sobre la Diversidad de Polinizadores, y Genética Poblacional de *Cucurbita pepo*, en un Bosque de Niebla de Guatemala. Tesis Doctorado, Universidad Nacional Autónoma de México, México.

20. Formisano, G., C. Roig, C. Esteras, M. R. Ercolano, F. Nuez, A. J. Monforte, y M. B. Picó. 2012. Genetic diversity of Spanish Cucurbita pepo landraces: An unexploited resource for summer squash breeding. *Genetic Resources and Crop Evolution*. 59: 1169-1184.

21. FAO. 2010. El Segundo informe sobre el estado de los recursos fitogenéticos para la alimentación y la agricultura en el mundo. Roma, Italia.

22. Fuller, D. Q. 2012. New Archaeobotanical Information on Plant Domestication from Macro-Remains: Tracking the Evolution of Domestication Syndrome Traits. In: Harlan, J. R., Gepts, P., Famula, T. R., Bettinger, R. L., Brush, S. B., Damania, A. B., ... & Qualset, C. O. (Eds.). Biodiversity in Agriculture: Domestication, Evolution, and Sustainability. Cambridge University Press, New York. pp. 110-135.

23. Gaut, B. S., C. M. Díez, y P. L. Morrell. 2015. Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*. 31: 709-719.

24. Gepts, P., Famula, T. R., Bettinger, R. L., Brush, S. B., Damania, A. B., McGuire, P. E., & Qualset, C. O. (Eds.). (2012). *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*. Cambridge University Press.

25. Gong, L., H. S. Paris, G. Stift, M. Pachner, J. Vollmann, y T. Lelley. 2013. Genetic relationships and evolution in *Cucurbita* as viewed with simple sequence repeat polymorphisms: the centrality of *C. okeechobeensis*. *Genetic Resources and Crop Evolution*. 60: 1531-1546.

26. Gong, L., H. S. Paris, M. H. Nee, G. Stift, M. Pachner, J. Vollmann, y T. Lelley. 2012. Genetic relationships and evolution in *Cucurbita pepo* (pumpkin, squash, gourd) as revealed by simple sequence repeat polymorphisms. *Theoretical and Applied Genetics*. 124: 875-891.

27. Hamilton, M. B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology*, 8(3), 521-523.

28. Hedrick, P. 2011. *Genetics of populations*. Jones & Bartlett Learning. Sudbury, MA. USA.

29. Hernández-Rosales, H. S., Castellanos-Morales, G., Sánchez-de la Vega, G., Aguirre-Planter, E., Montes-Hernández, S., Lira-Saade, R., et al. 2020. Phylogeographic and population genetics analyses of *Cucurbita moschata* reveal divergence of two mitochondrial lineages linked to an elevational gradient. *Am. J. Bot.* 107, 510–525.

30. Holsinger, K. E., & Weir, B. S. 2009. Genetics in geographically structured populations: defining, estimating and interpreting $F_{ST}$. *Nature Reviews Genetics*, 10(9), 639-650.

31. Hurd, P.D., Linsley, E.G., Whitaker TW. 1971. Squash and gourd bees (*Peponapis*, *Xenoglossa*) and the origin of the cultivated *Cucurbita*. *Evolution* 25, 218– 234.

32. Hufford, M. B., Martínez-Meyer, E., Gaut, B. S., Eguiarte, L. E., and Tenaillon, M. I. 2012. Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PLoS One* 7:e47659.

33. Innan, H., Kim, Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci* USA 101(29):10667–10672.

34. Kates, H. R., Anido, F. L., Sánchez-de la Vega, G., Eguiarte, L. E., Soltis, P. S., & Soltis, D. E. 2021. Targeted sequencing suggests wild-crop gene flow is central to different genetic consequences of two independent pumpkin domestications. *Frontiers in Ecology and Evolution*, 405.

35. Khoury, C. K., Carver, D., Kates, H. R., Achicanoy, H. A., van Zonneveld, M., Thomas, E., ... & Greene, S. L. 2020. Distributions, conservation status, and abiotic stress tolerance potential of wild cucurbits (*Cucurbita* L.). *Plants, People, Planet* 2(3), 269-283.

36. Larson, G., Piperno, D.R., Allaby, R.G., Purugganan, M.D., Andersson, L., Arroyo-Kalin, M. and Fuller, D.Q. 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences*, 111(17), 6139-6146.

37. Lira, R., Eguiarte, L., Montes, S., Zizumbo-Villarreal, D., Marín, P. C. G., & Quesada, M. 2016. Homo sapiens–*Cucurbita* interaction in mesoamerica: domestication, dissemination, and diversification. In: *Ethnobotany of Mexico* (pp. 389-401). Springer, New York, NY.

38. Lira, S. D. y Montes, S. H. 1994. Cucurbits. En: Bermejo, J. E. H., & León, J. (Eds.). *Neglected crops: 1492 from a different perspective* pp 63-79 (Vol. 26). FAO.

39. Lira, S. R., T. C. Andres y M. Nee, 2009a. Cucurbitaceae. En: Davidse, G., M. Sousa S., S. Knapp y F. Chiang (eds.). Cucurbitaceae a Polemoniaceae. *Flora Mesoamericana* Vol. 4(1). Instituto de Biología, Universidad Nacional Autónoma de México, Missouri Botanical Garden, The Natural History Museum. México, D.F.

40. Lira, Saade. R. 1995. Estudios taxonómicos ecogeográficos de las Cucurbitaceae Latinoamericanas de importancia económica. Roma, Italia: International Plant Genetic Resources Institute. Roma, Italia. 281 p.

41. López-Uribe, M.M., Cane, J.H., Minckley, R.L., Danforth, B.N. 2016. Crop domestication facilitated rapid geographical expansion of a specialist pollinator, the squash bee *Peponapis pruinosa*. *Proc. R. Soc. B* 283: 20160443.

42. Martínez-González C., Castellanos-Morales G., Barrera-Redondo J., Sánchez-de la Vega G., Hernández-Rosales H.S., Gasca-Pineda J., Aguirre-Planter E., Moreno-Letelier A., Escalante A.E., Montes-Hernández S., Lira-Saade R. and Eguiarte L. 2021. Recent and Historical Gene Flow in Cultivars, Landraces, and a Wild Taxon of *Cucurbita pepo* in Mexico. *Frontiers in Ecology and Evolution* 9: 656051.

43. Meyer, R. S., y M. D. Purugganan. 2013. Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics*. 14: 840-852.

44. Montes-Hernández, S., y Eguiarte, L. E. 2002. Genetic structure and indirect estimates of gene flow in three taxa of *Cucurbita* (Cucurbitaceae) in western Mexico. *American Journal of Botany* 89: 1156-1163.

45. Montes-Hernández, S., Merrick, L. C., & Eguiarte, L. E. 2005. Maintenance of squash (*Cucurbita* spp.) landrace diversity by farmers' activities in Mexico. *Genetic Resources and Crop Evolution*, 52(6), 697-707.

46. Moreno-Estrada, A., Gignoux, C. R., Fernández-López, J. C., Zakharia, F., Sikora, M., Contreras, A. V., et al. 2014. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 314, 1280–1285.

47. Nabhan, G. P. 1984. Evidence of gene flow between cultivated *Cucurbita mixta* and a field edge population of wild *Cucurbita* at Onavas, Sonora. *Cucurbit Genetics Cooperative*, 7, 76-77.

48. Nee M. 1990. The domestication of *Cucurbita*. *Economic Botany* 44: 56–68.

49. Nee, M., 1993. Cucurbitaceae I. En: V. Sosa, L. Cabrera R., T. Duncan, M. T. Mejía-Saulés, N. P. Moreno, M. Nee, L. I. Nevling, J. Rzedowski, B. G. Schubert, A. Gómez-Pompa (eds.). *Flora de Veracruz*. Fascículo 74. Instituto de Ecología, Xalapa, Veracruz, México.

50. Ntuli, N. R., Tongoona, P. B., & Zobolo, A. M. 2015. Genetic diversity in Cucurbita pepo landraces revealed by RAPD and SSR markers. *Scientia Horticulturae*, 189, 192-200.

51. Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, 13(5), 1143-1155.

52. OECD. 2012. Squashes, pumkins, zucchinis and gourds (Cucurbita species). In Safety Assessment of transgenic organisms in the environment (Vol. 5.) OECD Consensus Documents, OECD Publishing, Paris.

53. Ornelas, J. F., Sosa, V., Soltis, D. E., Daza, J. M., González, C., Soltis, P. S., ... & Ruiz-Sanchez, E. 2013. Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. *PloS one*, 8(2), e56283.

54. Perales, H.R., y J.R. Aguirre. 2008. Biodiversidad humanizada, en Capital natural de México, vol. I: Conocimiento actual de la biodiversidad. Conabio, México, pp. 565-603.

55. Piñero, D., Barahona, A., Eguiarte, L., Rocha, A., & Salas, R. 2008a. La variabilidad genética de las especies: aspectos conceptuales y sus aplicaciones y perspectivas en México. Conabio; 1 (Supt 1): 415-435.

56. Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J., & Dickau, R. 2009. Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences*, 106(13), 5019-5024.

57. Piperno, D. R. 2011. The origins of plant cultivation and domestication in the New World tropics: patterns, process, and new developments. *Current anthropology*, 52(S4), S453-S470.

58. Priori, D., Zizumbo-Villarreal, D., Ek, V. M. J. C., Limones-Briones, V., & Barbieri, R. L. 2018. Evolutionary dynamics of *Cucurbita argyrosperma* from the Mesoamerican domestication center using SSR molecular markers. *Pesquisa Agropecuaria Brasileira*, 53, 287-297.

59. Ranere, A. J., D. R. Piperno, I. Holst, R. Dickau, and J. Iriarte. 2009. The cultural and chronological context of early Holocene maize and squash domestication in the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences*, 106(13), 5014-5018.

60. Rimieri, P. 2017. La diversidad genética y la variabilidad genética: dos conceptos diferentes asociados al germoplasma y al mejoramiento genético vegetal. *BAG. Journal of Basic and Applied Genetics*, *28*(2), 7-13.

61. Sánchez-de la Vega, G., Castellanos-Morales, G., Gámez, N., Hernández-Rosales, H. S., Vázquez-Lobo, A., Aguirre-Planter, E., ... & Eguiarte, L. E. 2018. Genetic resources in the "Calabaza Pipiana" squash (*Cucurbita argyrosperma*) in Mexico: genetic diversity, genetic differentiation and distribution models. *Frontiers in Plant Science*, *9*, 400.

62. Sanjur, O. I., D. R. Piperno, T. C. Andres, y L. Wessel-Beaver. 2002. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. *Proceedings of the National Academy of Sciences of the United States of America*. 99: 535-540.

63. Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American journal of botany*, 94(3), 275-288.

64. Stinnesbeck, W., Becker, J., Hering, F., Frey, E., González, A. G., Fohlmeister, J., ... & Deininger, M. 2017. The earliest settlers of Mesoamerica date back to the late Pleistocene. *PLoS One,* 12(8), e0183345.

65. Taberlet, P., L. Gielly, G. Pautou, y J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*. 17: 1105-1109.

66. Tapia Aguirre, F. 2021. Filogeografía y diversidad genética de *Cucurbita argyrosperma* ssp. *sororia*. Tesis de Licenciatura, Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Iztacala. Estado de México.

67. Villanueva, V. C. 2007. Calabazas Cultivadas. Identificación de Especies, Caracterización y Descripción Varietal. Universidad Autónoma Chapingo. Chapingo, Edo. de México. 123 p.

68. Willcox, G. 2012. Pre-domestic cultivation during the late Pleistocene and early Holocene in the northern Levant. In: Harlan, J. R., Gepts, P., Famula, T. R., Bettinger, R. L., Brush, S. B., Damania, A. B., ... & Qualset, C. O. (Eds.). Biodiversity in Agriculture: Domestication, Evolution, and Sustainability. Cambridge University Press, New York. pp. 92-109.

*69.* Wilson, H.D., Lira, R., Rodríguez, I. 1994. Crop/weed gene flow: *Cucurbita argyrosperma* Huber and *C. fraterna* L.H. Bailey. *Econ Bot*. 48:293–300.

70. Wilson, H. D. 1990. Gene flow in squash species. *BioScience,* 40(6), 449-455.

71. Zeder, M.A., Emshwiller, E., Smith, B.D., and Bradley, D.G. 2006. Documenting domestication: the intersection of genetics and archaeology. *TRENDS in Genetics*, 22(3), 139-155.

72. Zizumbo-Villarreal, D. y Colunga G, P. 2008. El origen de la agricultura, la domesticación de plantas y el establecimiento de corredores biológico-culturales en Mesoamérica. *Rev. Geogr. Agrícola* 41, 85–113.