



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
INSTITUTO DE ECOLOGÍA

DIVERSIDAD GENÉTICA Y CONSERVACIÓN DE *GOSSYPIUM HIRSUTUM*
SILVESTRE Y CULTIVADO EN MÉXICO

T E S I S
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

PRESENTA:
ANA LAURA WEGIER BRIUOLO

TUTOR
DR. DANIEL PIÑERO DALMAU
INSTITUTO DE ECOLOGÍA

MIEMBROS DEL COMITÉ TUTOR
DR. ALEJANDRO CASAS
CENTRO DE INVESTIGACIONES EN ECOSISTEMAS
DR. RAFAEL LIRA SAADE
FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

MÉXICO, D. F. ENERO 2013



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Agradecimientos

Quiero agradecer a la Universidad Nacional Autónoma de México y a toda su comunidad por incubarme desde la licenciatura y dejarme tantas enseñanzas.

Agradezco el apoyo financiero y las facilidades otorgadas a la CONABIO a través de sus proyectos GEF-CIBIOGEM 00013572, PNUD-CIBIOGEM 00051868, para la recopilación de información sobre *Gossypium hirsutum* y sus parientes silvestres; y DGSPRNR “*Conocer el estado de las especies prioritarias de las que México es centro de origen y diversidad genética, en el marco del artículo 86 de la Ley de Bioseguridad de los Organismos Genéticamente Modificados*”.

Agradezco al Posgrado de Ciencias Biomédicas, al Instituto de Ecología de la UNAM y al CONACYT (numero de becario 172630) por darme la oportunidad de realizar mis estudios de doctorado, los cuales espero retribuir con la misma paciencia.

El Dr. Daniel Piñero es un buen formador de biólogos, maestros, doctores e instituciones. Sin lugar a dudas, estoy agradecida por todo lo que me ha enseñado. Has sido un padre académico excelente para mí.

Agradezco al comité tutorial que me acompañó en este doctorado, por su apoyo y dedicación: Dr. Rafael Lira y Dr. Alejandro Casas. A Dr. Richard Abbott y Dra. Amanda Gálvez que me asesoraron en mis estancias. Al comité sinodal: Dr. José Sarukhán, Dr. Antonio Turrent, Dra. Francisca Acevedo, Dra. Ana Elena Escalante y Dr. Daniel Piñero por realizar tan valiosos comentarios. Oswaldo Oliveros y Daniel Ortiz por su apoyo en la presentación de la información geográfica. A Valeria Alavez, Brian Urbano, Adriana Uscanga y Sandra Petrone, que me leyeron y esperaron con tanta paciencia.

Quiero agradecer a Valeria Alavez y Alma Piñeyro-Nelson por su apoyo y dedicación. A los coautores del artículo por embargarse en la aventura y seguir a flote: Dra. Elena Álvarez-Buylla, Dra. Amanda Gálvez Mariscal, Jesús Alarcón; y a quiénes remararon durante el proceso: José Casar, René Cerritos y Victoria Sork.

Yo tengo mucho que agradecer a muchas personas en el IE, las que colaboraron conmigo durante todo este tiempo, desde la licenciatura hasta el doctorado; desde el campo, las aulas y el laboratorio. A todos los que me enseñaron, apoyaron y aconsejaron. A mis compañeros del Laboratorio de genética y evolución y anexos: AleVL, Vale, Brian, Rodo, Lev, Argelia, AnaE, AleM, AleO, Ariadna, Mariana, Erick, Mastretta, Santiniss, Israel, Eguiartes y Valerios. A mis amigas Lluvis, Alma, Andrea, Lupis, Lau, a los investigadores Jaramillo, Mely, Alicia, Eguiarte, Valeria, Fornoni. A todos los que hicieron mi estancia tan grata que casi y no me voy... Soco, Ernesto, Oscar, Paty, Virgilio, América, Daniel, Gaby y a quienes no dejé en paz ni en vacaciones y siempre me recibieron con sonrisas; Paty vigilancia, Laura mi tocalla, Don José, Doña Silvia, Lupita, Anabel, todos ¡GRACIAS!

Índice

Resumen

Abstract

1. Introducción

- 1.1 El flujo génico
- 1.2 Genética del paisaje
- 1.3 Genética de la conservación
- 1.4 Modelo de estudio: *Gossypium hirsutum*

2. Historia evolutiva del género *Gossypium*

- 2.1 La tribu *Gossypieae*
- 2.2 El género *Gossypium*
 - 2.2.1 *Las especies australianas*
 - 2.2.2 *Especies Afro-Asiáticas*
 - 2.2.3 *Especies Americanas*
 - 2.2.4 *Los Gossypium alotetraploides (AADD)*
- 2.3 Tricomas de las semillas

3. Origen y distribución del subgénero *Houzingenia* Fryxell, del género *Gossypium* L. en México

- 3.1 Origen y distribución de especies diploides del género *Gossypium*
- 3.2 Distribución actual y áreas de distribución potencial de *Gossypium* diploides que habitan en México
 - 3.2.2 *Métodos y resultados*
- 3.3 Discusión y conclusiones

4. Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin

- 4.1 Semblanza

4.2 Síntesis del artículo “*Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (Gossypium hirsutum) at its centre of origin*” (omitiendo las figuras, metodología, resultados y materiales suplementarios)

4.3 Discusión

4.4 Metapoblaciones de algodón: distribución actual, la dinámica ecológica y el cambio de uso del suelo

4.5 Flujo de genes entre metapoblaciones y cultivares de algodón

4.6 Los transgenes en metapoblaciones de algodón silvestre

4.7 Conclusiones

Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin

5. Determinación de los centros de diversidad genética de *Gossypium hirsutum* en México

5.1 Introducción

5.2 Centros de diversidad genética

5.3 Unidades de conservación

5.3.1 Unidades evolutivamente significativas (ESU)

5.3.2 Unidades de manejo (MU)

5.4 Monitoreo de la diversidad genética

5.5 Métodos

5.5.1 Información genética utilizada

5.5.2 Establecimiento de los centros de diversidad genética de Gossypium hirsutum

5.5.3 Determinación de la tasa de dispersión entre grupos de poblaciones

5.5.4 Monitoreo de la diversidad genética

5.6 Resultados del análisis bioinformático: Identificación de los centros de diversidad genética en *Gossypium hirsutum*

5.6.1 Estructura genética

5.6.2 Determinación de las unidades de manejo

5.6.3 *Monitoreo de la diversidad genética dentro de cada unidad de manejo*

5.7 *Discusión*

5.7.1 *Estructura genética encontrada en las metapoblaciones de Gossypium hirsutum*

5.7.2 *Identificación de los centros de diversidad genética en Gossypium hirsutum y medidas necesarias para su conservación*

5.8 *Conclusiones*

6. Diversidad genética y conservación de *Gossypium hirsutum* silvestre y cultivado en México: Conclusiones y perspectivas

6.1 *Conclusiones y perspectivas*

6.1.1 *Biología evolutiva*

6.1.1.1 *Perspectivas biología evolutiva*

6.1.2 *Flujo génico*

6.1.3. *Genética del paisaje*

6.1.3.1. *Perspectivas genética del paisaje*

6.1.4. *Genética de la conservación*

6.2 *Estrategias de bioseguridad y medidas de protección para la conservación de la diversidad genética de *Gossypium hirsutum* en México*

6.2.1 *Estrategias de bioseguridad*

6.2.2 *Medidas de protección*

6.2.3 *Perspectivas generales*

7. Glosario

8. Literatura citada

Resumen

México es el centro de origen y diversidad genética del algodón, *Gossypium hirsutum* L. Es, además, donde se encuentra la mayor cantidad de poblaciones silvestres y donde también se cultivan plantas genéticamente modificadas (GM) de la especie. Para sugerir estrategias a largo plazo para la conservación y el manejo del algodón en México, fue necesario realizar un análisis integral para conocer tanto la historia evolutiva, como la estructura genética y geográfica de la especie. Se hizo especial énfasis en el flujo génico ancestral y actual, se propuso una estrategia para delimitar los centros de diversidad genética, así como las medidas de bioseguridad necesarias para evitar la introgresión entre algodones GM y las poblaciones silvestres y otros parientes.

La diversidad genética de las poblaciones silvestres encontrada utilizando microsatélites de cloroplasto fue alta ($H_e = 0.894 \pm 0.01$), mientras que en plantas cultivadas no se encontró variación. Se observó una estructura genética que consiste de ocho grupos (usando un análisis bayesiano BAPS, *Bayesian Analysis of Population Structure*) que coincide con la estructura geográfica (obtenida con un análisis genético, GARP, *Genetic Algorithm for Rule-set Production*) y ecológica de ocho metapoblaciones. El flujo génico histórico tiene un patrón de flujo a larga distancia, el cual continúa hasta la actualidad de acuerdo a los datos de flujo de transgenes liberados desde 1996 en el norte de México. Los análisis sobre los centros de diversidad genética indican que es necesaria la conservación de seis unidades, las cuales comprenden a ocho metapoblaciones. Esto se debe a que los fenómenos que las afectan pueden afectar a las demás debido, por un lado, al bajo número de individuos encontrados y por el otro a la alta tasa de migración reportada. Por lo anterior es necesario evitar el flujo génico con los parientes cultivados, debido a que son genéticamente homogéneos, pueden contener transgenes (los cuales ya se encuentran en cuatro de las ocho metapoblaciones) y aún desconocemos las consecuencias que pudieran tener sobre la especie y las interacciones que ésta sostiene en los ecosistemas en que habita (dunas costeras y selva baja caducifolia).

Abstract

Mexico is the center of origin and genetic diversity of cotton, *Gossypium hirsutum* L. It is also home to the largest number of wild populations of the species and one of the countries where genetically modified (GM) plants are cultivated. To suggest conservation and long-term management strategies for *G. hirsutum*, it was necessary to perform a comprehensive analysis to learn about the evolutionary history of the species, as well as about its genetic and geographic structure. Special emphasis was made on current and ancestral gene flow, on strategies for defining centers of genetic diversity and on biosafety measures required to prevent introgression between GM cotton and its wild relatives.

The genetic diversity found in wild plants using chloroplast microsatellites was high ($H_e = 0.894 \pm 0.01$), while in cultivated plants variation was not found. We observed a genetic structure that consisted in eight groups (using Bayesian analysis, BAPS, Bayesian Analysis of Population Structure) that agreed with the geographical structure (obtained with a genetic analysis, GARP, Genetic Algorithm for Rule-set Production) and with the ecological approach that resulted in eight metapopulations.

Historical gene flow presents a long-distance pattern, which continues to this day according to transgene flow data and given that GMOs were released since 1996 in northern Mexico. Analyses of centers of genetic diversity indicate that conservation of six units, which comprise all eight metapopulations, is necessary since the phenomena that could affect one of them may affect the others. This is due, on one hand, to the low number of individuals encountered and, on the other, to the high migration rate reported. Therefore, it is necessary to prevent gene flow with cultivated relatives, because they are genetically homogeneous, may contain transgenes (which are already in four of the eight metapopulations) and the consequences they might have on *G. hirsutum* and the interactions that this species sustains in the ecosystems in which it inhabits (coastal dunes and tropical deciduous forest) are not yet fully understood.

1. Introducción

Los procesos evolutivos y antropogénicos que dieron origen a cientos de cultivos importantes para la humanidad ocurrieron en áreas megadiversas biológica y culturalmente como Mesoamérica (Gepts 2004). En esta región aún encontramos poblaciones silvestres, semidomesticadas y distintas variedades domesticadas de estos cultivos. Las poblaciones silvestres forman parte de la dinámica ecológica y de los procesos evolutivos de los ecosistemas en los que viven. Conocer las características evolutivas en las que se originaron el género y la especie, determinar la diversidad y estructura genética, así como los procesos que la mantienen y entender la dinámica ecológica, ayudará a generar estrategias integrales para su adecuado aprovechamiento y conservación a largo plazo. Son datos que deben ser considerados en los análisis de riesgo y medidas de bioseguridad (Engels *et al.*, 2006). Asimismo deben de incorporarse en estos estudios los contextos culturales, sociales y económicos que los rodean (Brush 2001; Altieri 2004).

Este trabajo pretende aportar información que permita realizar estrategias para conservación y manejo a largo plazo de la especie *Gossypium hirsutum* L., la especie de algodón de la cual México es el centro de origen y de diversidad genética (Brubaker y Wendel 1994), donde actualmente se encuentra la mayor cantidad de poblaciones silvestres y donde también se cultivan plantas genéticamente modificadas (GM) de la especie. El objetivo general, es inferir la diversidad y estructura genética de las poblaciones silvestres de algodón para conocer los patrones históricos que la moldearon e influyen en la actualidad. Los objetivos particulares son: *i)* Revisar la biología evolutiva del género *Gossypium* con énfasis en la dispersión y especiación; *ii)* Modelar el nicho ecológico e inferir la estructura geográfica y ecológica de las poblaciones de *Gossypium hirsutum* L. en México; *iii)* Analizar el flujo génico interpoblacional y con sus parientes GM cultivados; y *iv)* Utilizar las herramientas de la genética de la conservación para delimitar las áreas de diversidad genética que pueden ser consideradas como centros de diversidad genética, así como sugerir las estrategias para su conservación.

1.1 El flujo génico

Las poblaciones de una especie, generalmente intercambian genes entre ellas en mayor o menor medida. En este proceso, que se denomina flujo génico, los genes

pueden ser transportados por individuos (como en la mayoría de los animales) o por gametos (como el polen o las gametas de muchos organismos marinos). Los migrantes que tienen éxito en reproducirse en su nueva población y dejan descendencia en la siguiente generación son aquellos que contribuyen al flujo génico (Futuyma, 2009).

El flujo génico es un componente principal de la estructura poblacional, ya que determina hasta qué punto cada población local de una especie es una unidad evolutiva independiente. Si existe una gran cantidad de flujo génico entre poblaciones locales, entonces todas ellas evolucionan juntas; pero si hay poco flujo génico, cada población evoluciona en forma casi independiente (Slatkin, 1995). Por lo tanto, el flujo génico homogeniza las poblaciones de una especie, es decir, las conduce a tener las mismas frecuencias alélicas, a menos que exista suficiente oposición de otras fuerzas divergentes como la deriva génica y la selección natural (Futuyma, 2009).

Las características de una especie afectan mayormente su capacidad de dispersión y flujo génico, razón por la cual es muy importante conocer a fondo su biología y estructura genética cuando se quieren realizar estrategias para la conservación de las especies o estudios de análisis de riesgo y/o diseñar medidas de bioseguridad (Bao-Rong y Snow, 2005, Meyer 2011).

El proceso de hibridación se puede definir, desde un punto de vista evolutivo, como la cruce entre individuos pertenecientes a poblaciones genéticamente diferentes. La hibridación frecuentemente conduce a un proceso conocido como introgresión, el cual consiste en la incorporación de genes de un taxón a otro o a más poblaciones (Ellstrand, 2003). La hibridación y la introgresión entre diferentes especies son subconjuntos del flujo génico, definido por Futuyma (2009) como “la incorporación de genes al pool genético de una población a partir de otra población u otras más”. Las consecuencias evolutivas de la hibridación entre cultivos y sus parientes silvestres pueden ser profundas, por ejemplo, las plantas cultivadas suelen tener poca diversidad genética debido a los cuellos de botella que pasaron durante la domesticación (Gepts 2004). En estos casos, el “genotipo cultivado” tiene una mayor probabilidad de estar presente en la siguiente generación simplemente por estar más representado que los silvestres que pueden ser todos distintos. Si hay una alta tasa de flujo génico y transcurren varias generaciones, entonces habrá pérdida de la diversidad genética de las poblaciones silvestres (Slatkin, 1987). Otra consecuencia del flujo génico es que se restrinja la adaptación de las poblaciones a las condiciones locales (Slatkin y

Hudson, 1991). La velocidad a la que ocurra dependerá principalmente de la variación genética inicial de las poblaciones, el número de individuos y la tasa de migración, aunque la selección natural y la deriva génica están siempre presentes y desconocemos como afectarán el resultado. Otros problemas potenciales son la evolución de supermalezas (Ellstrand *et al.*, 1999; Ellstrand, 2003) y la introgresión de transgenes a las poblaciones naturales, que es el impacto ambiental potencial más frecuentemente discutido en el ámbito de la biotecnología de plantas y que no solo debe ser medido sobre la diversidad de las plantas si no en todo el ecosistema (Wegier *et al.*, en preparación).

1.2 Genética del paisaje

Las herramientas de la genética del paisaje ofrecen en la actualidad la posibilidad de observar la influencia de los procesos ecológicos que modelaron la variación genética en relación a la distribución espacial (Manel *et al.* 2003; Storfer *et al.* 2007; Sork y Waits. 2010). Por lo tanto, se puede hacer explícita la variación genética respecto al ambiente y podemos utilizar entonces aproximaciones de esta disciplina para comprender procesos evolutivos históricos y contemporáneos (Manel *et al.* 2003). Entender cómo el paisaje geográfico influencia la conectividad de las poblaciones es un tema ampliamente discutido (Spielman & Smouse 1976; Sokal *et al.* 1991; Taylor *et al.* 1993; Baer 1998; en Dyer *et al.* 2010), pero la forma en la que se entiende y se mide la magnitud del flujo génico ha cambiado con el uso de marcadores moleculares neutrales (Dyer *et al.* 2010), que pueden mutar rápidamente, como los microsatélites del cloroplasto de las angiospermas dicotiledóneas (como el algodón).

Las redes gráficas de poblaciones permiten realizar comparaciones entre las varianzas y covarianzas intra e interpoblacionales. Además, podemos saber si las distancias genéticas corresponden con lo esperado por la distancia geográfica que separa a las poblaciones, así discriminar si las poblaciones más cercanas geográficamente son más parecidas genéticamente y si la diferenciación entre ellas aumenta con la distancia (modelo tradicional de aislamiento por distancia o IBD por sus siglas en inglés) o si el patrón observado es independiente de la distancia y por lo tanto podemos tener, tanto poblaciones cercanas muy diferentes, como también, poblaciones lejanas con y sin diferenciación genética entre ellas (modelo de flujo a larga distancia). Esto es muy importante en términos de genética de la conservación

ya que permite tomar decisiones adecuadas respecto a la distribución genética y geográfica de las poblaciones (Dyer *et al.*, 2010).

1.3 Genética de la conservación

El objetivo principal de la genética de la conservación es proteger a la diversidad biológica entendiendo a los procesos evolutivos que la mantienen (Moritz, 2002), y por lo tanto ayudando a minimizar las extinciones, evitando los problemas relacionados con tamaños efectivos pequeños, como el efecto deletéreo de la endogamia, la pérdida de diversidad y la habilidad para evolucionar en respuesta a los cambios ambientales, así como los efectos deletéreos que ocurren por la cruce entre individuos muy distintos (depresión por exogamia; Amos y Balmford, 2001; Frankham *et al.*, 2002). Los análisis genéticos también permiten estudiar el efecto de la fragmentación y la reducción del flujo génico en poblaciones estructuradas y el efecto de la acumulación y pérdida de mutaciones deletéreas (Rocha y Gasca 2007).

Un concepto ampliamente utilizado en el área de la Genética de la conservación es el de las *unidades de conservación* (Geist y Kuehn, 2005). Una vez que las unidades son delimitadas se puede discernir entre diferentes estrategias para la conservación de las poblaciones, así como identificar los procesos que las vulneran.

Las unidades de conservación pueden dividirse en *unidades evolutivamente significativas* y en *unidades de manejo*. Las unidades evolutivamente significativas sirven para separar grupos de poblaciones que han divergido hace mucho tiempo, mientras que las unidades de manejo sirven para dividir grupos de poblaciones que han divergido más recientemente y donde la tasa de dispersión entre grupos es menor al 10%, que representa la proporción de individuos que van de una población a otra (Moritz, 1994; Palsbøll *et al.*, 2007). En este trabajo definimos a las *unidades de manejo* como centros de diversidad genética, ya que el aislamiento entre poblaciones es el que nos permite diferenciar grupos que evolucionan independientemente. De acuerdo a Palsbøll *et al.* 2007, lo más importante para definir unidades de manejo es la tasa de dispersión y no el flujo génico. La diferencia entre la tasa de dispersión y el flujo génico es que en la tasa de dispersión lo único que importa es el movimiento de los individuos entre poblaciones. Mientras que en el flujo génico (a diferencia de la tasa de dispersión) se requiere que haya una contribución a la poza genética de otra población.

1.4 Modelo de estudio: *Gossypium hirsutum*

El algodón ha sido fundamental desde el origen de varias civilizaciones y continúa siendo una de las especies más importantes para la humanidad en nuestros días (Fryxell 1979; Wendel *et al.*, 2010). Es la fibra natural más utilizada y la tercera fuente de aceite vegetal (FAOSTAT 2010). Ocupa el sexto lugar mundial en superficie cultivada y el algodón genéticamente modificado se coloca en el tercer lugar de los cultivos biotecnológicos más sembrados. Es importante hacer notar, que dentro de los quince cultivos más importantes del mundo, el algodón es el único que no adquirió su valor por formar parte de la base de la alimentación (Wendel *et al.*, 2010).

De las cuatro especies de algodón domesticadas, *Gossypium hirsutum* ocupa el 95% de la producción actual y la mayoría de las poblaciones silvestres de esta especie habitan en México. Aunque se han realizado vastas investigaciones sobre la biología, ecología y genética de la especie, la mayoría se han basado en plantas domesticadas y fuera de su distribución natural por lo que en realidad se conoce poco de la especie, ya que después del proceso de domesticación, se conserva sólo una parte de la variación que pudiera encontrarse en las poblaciones silvestres y por lo tanto las investigaciones sobre las plantas domesticadas deben utilizarse con precaución y sin extrapolar al resto de la especie.

Debido a que el algodón (*Gossypium* L.) es un género tan importante en la economía mundial, ha capturado la atención de científicos agrícolas, taxónomos y biólogos evolutivos. Especialmente en las últimas décadas, las tecnologías moleculares se han aplicado para contestar las preguntas clásicas como el origen de la poliploidía de las especies, las relaciones filogenéticas entre las especies del género y los orígenes de las plantas domesticadas a partir de sus progenitores silvestres (Wendel *et al.*, 2011).

Quizás el aspecto más llamativo de su historia es que debido a su amplia distribución geográfica, se involucró con antiguas culturas en distintos continentes, lo cual dio lugar a un proceso de domesticación convergente o paralelo a partir de ancestros silvestres divergentes y geográficamente aislados. Esta domesticación paralela involucra a cuatro especies: dos americanas, *G. hirsutum* y *G. barbadense*, y dos africanas y asiáticas, *G. arboreum* y *G. herbaceum*. Los habitantes de distintas

regiones del mundo descubrieron miles de años atrás y de manera independiente, que las propiedades únicas de las fibras de las cuatro especies de algodón las hacían útiles para la elaboración de cuerdas, textiles y otras aplicaciones, por lo tanto, cada una de éstas especies posee una historia única de domesticación, diversificación y utilización (Fryxel 1979).

Esta rica historia involucra modelaje de su estructura genética, por manejo y selección artificial de la variación originada por los procesos evolutivos a lo largo de millones de años, incluyendo en los últimos años, la obtención de plantas genéticamente modificadas (GM), a través de ingeniería genética.

Actualmente, se han liberado al ambiente plantas genéticamente modificadas de algodón en 14 países, incluido el norte de México desde 1996. En donde se cultivaron, 190,000 h. en 2011, con dos objetivos principalmente; la resistencia al ataque de lepidópteros y la tolerancia a herbicidas (CERA 2011, FAO 2009).

2. Historia evolutiva del género *Gossypium*

Resumen

El objetivo general de este capítulo es documentar los patrones históricos de especiación y dispersión de los parientes cercanos a la especie *Gossypium hirsutum*, para estudiarlo tomando en cuenta su contexto evolutivo. Para ello se revisa el origen del género *Gossypium* y su diversidad, en particular sobre el origen y diversidad del subgénero *Houzenia* que contiene a las especies diploides americanas así como también las evidencias del origen de las especies tetraploides.

2.1 La tribu *Gossypieae*

El género del algodón pertenece a una pequeña tribu monofilética de la familia Malvaceae, originada hace 20 millones de años, llamada *Gossypieae* (LaDuke y Doble 1995; Seelanan *et al.*, 1997). Esta tribu incluye a ocho géneros que se caracterizan, entre otras cosas, por presentar glándulas de gossypol en toda la planta. Cuatro de estos géneros son pequeños con distribuciones geográficas restringidas, incluyendo *Lebronnecia* (Islas Marquesas), *Cephalohibiscus* (Nueva Guinea, Islas Salomón), *Gossypoides* (este de África, Madagascar) y *Kokia* (Hawaii). La tribu también incluye cuatro géneros de tamaño mediano con un intervalo geográfico mayor: *Hampea*, con 21 especies neotropicales; *Cienfuegosia*, un género diverso con 25 especies de los neotrópicos y parte de África; *Thespesia*, con 17 especies tropicales; y por último *Gossypium*, el género más numeroso y ampliamente distribuido de la tribu con más de 50 especies (Fryxell 1965, 1968; Wendel *et al.*, 2010).

Dentro de la tribu, los parientes más cercanos del género *Gossypium*, son el género malgache-africano *Gossypoides* y el género hawaiano endémico *Kokia*. Usando el reloj molecular se sabe que esta divergencia ocurrió hace 12.5 millones de años aproximadamente, lo cual se ha confirmado en estudios posteriores (Wendel 2011). La dispersión de la tribu *Gossypiae* despertó gran curiosidad en el siglo pasado (figura 1.1) y la conclusión es que la dispersión a larga distancia es transoceánica en la mayoría de los casos; esto sumado a los estudios fisiológicos realizados en los que se demuestra la tolerancia a la salinidad (Stephens, 1958, 1966; Fryxell, 1979; Wendel y Percival, 1990; Wendel y Percy, 1990; DeJooe y Wendel, 1992; Wendel y

Albert, 1992; Gossett *et al.*, 1992, 1994).

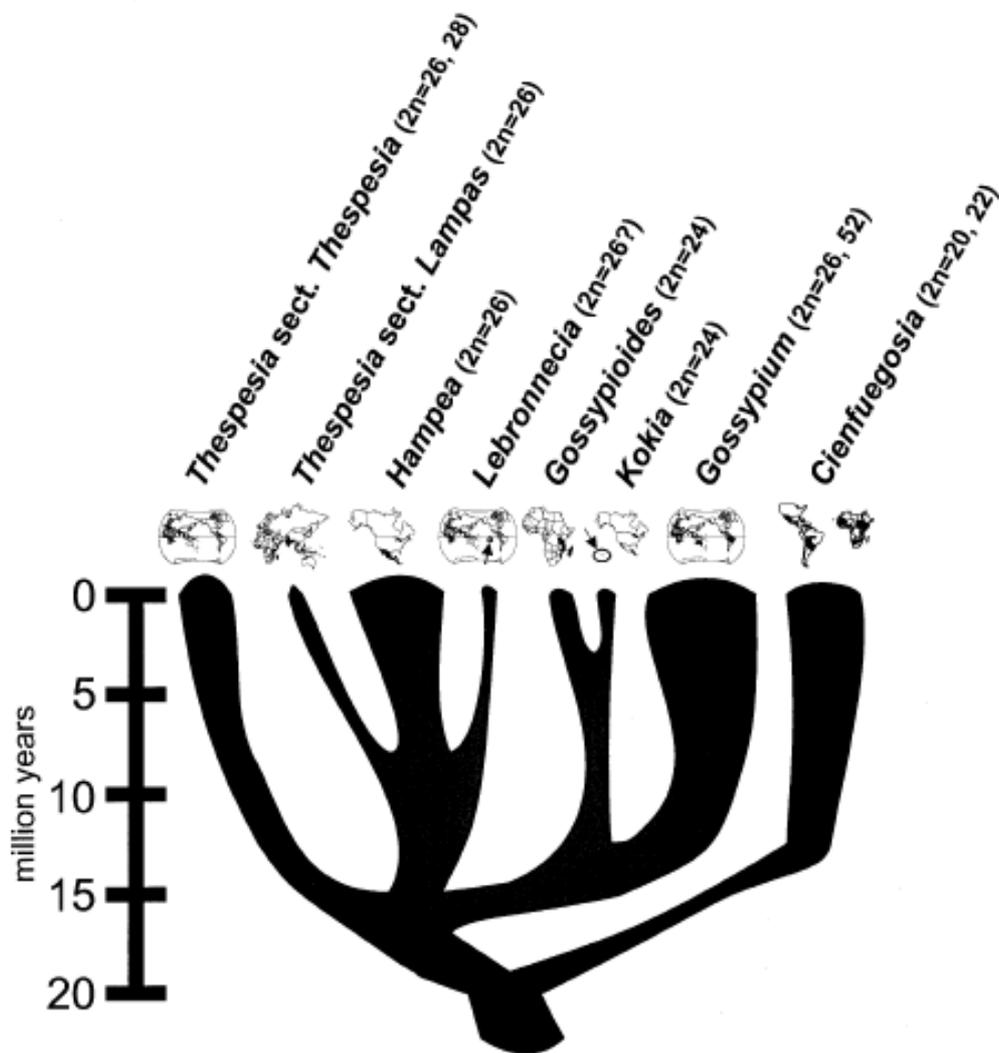


Figura 1.1 Relaciones filogenéticas de la tribu *Gossypiae*. Los tiempos de divergencia fueron estimados por reloj molecular (Seelanan *et al.*, 1997), el grosor de las ramas ejemplifica el número de especies por las que está representado cada uno de los géneros en la actualidad y sobre ellas se muestra el mapa de su distribución (figura tomada de Wendel *et al.*, 2011).

2.2 El género *Gossypium*

Gossypium parece haber divergido de sus parientes más cercanos durante el Mioceno, expandiéndose subsecuentemente alrededor del mundo por medio de dispersión transoceánica adquiriendo su distribución geográfica actual (Seelanan *et al.*, 1997).

La taxonomía del género ha sido bien estudiada. Las especies están agrupadas en 4 subgéneros y 8 secciones. Este sistema de clasificación está basado primordialmente en evidencia morfológica y geográfica, aunque la mayoría de las clasificaciones infragenéricas son congruentes con los datos citogenéticos y moleculares (Wendel *et al.*, 2010). Los centros de diversidad del género están definidos por ser ricos en el número de especies e incluyen Australia, especialmente la región de Kimberley al noroeste de ésta; el Cuerno de África (Somalia, Yibuti, Eritrea y Etiopía); el sur de la península Arábiga y la parte occidental del centro y sur de México. El reconocimiento de estos grupos de especies relacionadas y de sus atributos individuales refleja un profundo conocimiento acumulado que emergió de la exploración básica de las plantas y del estudio taxonómico y evolutivo (Watt, 1907; Hutchinson *et al.*, 1947; Saunders, 1961; Fryxell, 1979, 1992, Wendel *et al.*, 2011).

Actualmente, *Gossypium* incluye alrededor de 50 especies pero, notablemente, nuevas especies continúan siendo descubiertas (Wendel *et al.*, 2010). El género es extraordinariamente diverso; la morfología de las especies oscila entre adaptaciones al fuego, herbáceas perennes en el noroeste de Australia a árboles al suroeste de México los cuales escapan a la temporada de secas dejando caer sus hojas. Los colores de la corola abarcan un espectro que va del azul al púrpura (*G. triphyllum*), malvas y rosas (*G. sturtianum*, el emblema oficial del Territorio Norte de Australia), blancos y amarillos pálidos (noroeste de Australia, México y África-Arabia) e inclusive un profundo amarillo azufre (*G. tomentosum* de Hawái). Las cubiertas de las semillas van de casi glabro (*G. klotzschianum* y *G. davidsonii*), a presentar cerdas café, densas y firmes que ayudan en la dispersión por viento, hasta finas hebras blancas que caracterizan a formas altamente mejoradas de las cuatro especies cultivadas. Existen inclusive semillas que producen depósitos de grasa para facilitar la dispersión por hormigas (Fryxell, 1979; Wendel *et al.*, 2010).

Conforme el género se diversificó y se expandió, experimentó una extensa evolución cromosómica. Aunque todas las especies diploides comparten el mismo número de cromosomas ($n = 13$), existe más del triple de variación en el contenido de ADN por genoma. La morfología de los cromosomas es similar entre especies cercanamente emparentadas, y esto está reflejado en la habilidad de las especies para formar híbridos que muestran emparejamiento normal de cromosomas durante la meiosis y en ocasiones una alta fertilidad en F_1 . En contraste, las cruas entre parientes más distantes son difíciles de efectuar y aquellas que son exitosas suelen

presentar anormalidades durante la meiosis. Las observaciones colectivas del comportamiento de emparejamiento, el tamaño de los cromosomas y la relativa fertilidad en híbridos interespecíficos condujo a la designación de cada genoma con símbolos de una sola letra para definir los grupos de especies. Actualmente se reconocen 8 grupos de genomas diploides (desde el A hasta el G, más el K, Seelanan *et al.*, 1997, Figura 1.2).

Múltiples trabajos filogenéticos muestran que los linajes genealógicos de las especies son consistentes con la designación de los genomas, por consiguiente, cada grupo genómico corresponde a un único linaje natural y, en la mayoría de los casos, estos linajes son también geográficamente cohesivos (Cronn *et al.*, 1996; Seelanan *et al.*, 1997; Wendel *et al.*, 2010).

Existen cuatro linajes principales de especies diploides correspondientes a tres continentes: Australia (genomas C, G y K), América (genoma D) y África-Arabia (dos linajes: uno incluyendo el genoma A, B y F y otro que contiene el genoma E). El evento de divergencia más temprano ocurrió hace 10 millones de años y separó el genoma D americano de su ancestro en todos los otros taxa. Los siguientes eventos de divergencia se pueden observar en la figura 1.3, de la cual cabe resaltar que se trata de la última filogenia publicada (Wendel *et al.*, 2010), ya que en el pasado, algunas de las tricotomías presentes en la figura actual se mostraban resueltas. Del cuadro evolutivo se infiere que hubo una rápida radiación global temprana en la historia del género, con eventos de divergencia temporalmente cercanos.

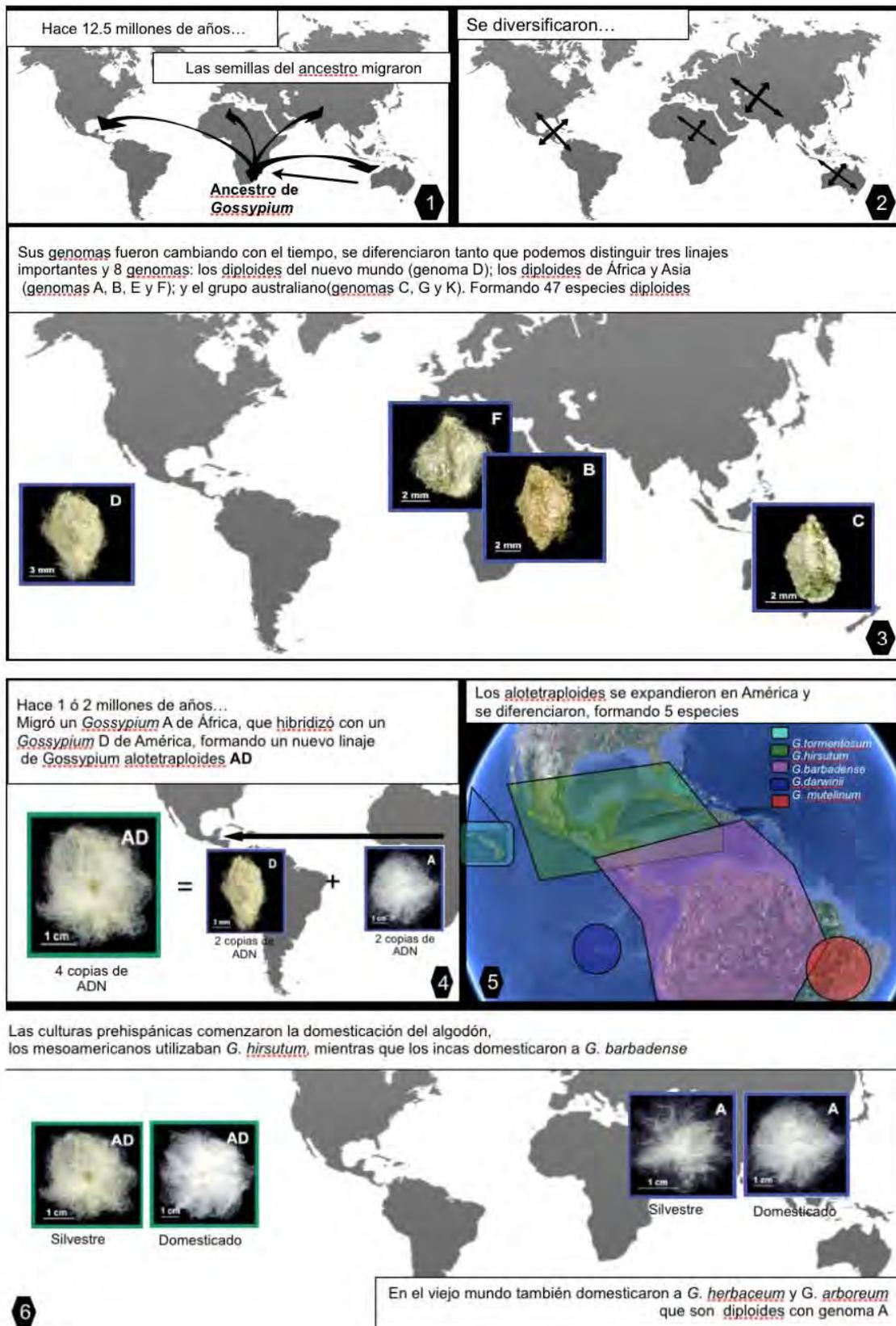


Figura 1.2 Esquema de la dispersión geográfica del género *Gossypium* (lo voy a mejorar).

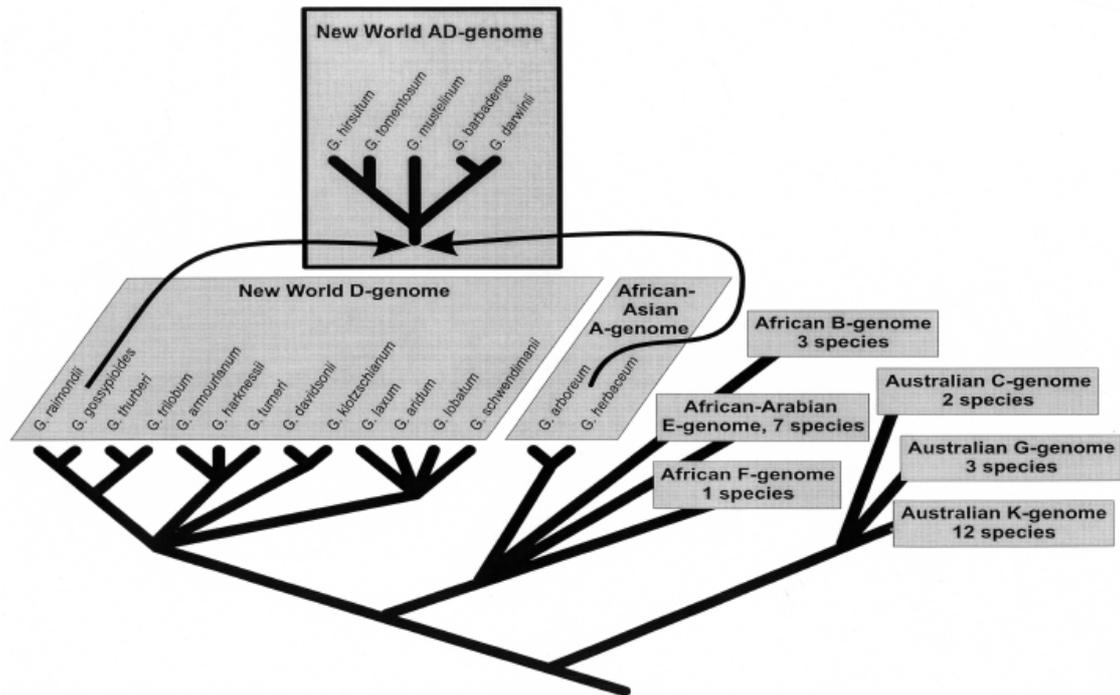


Figura 1.3 Historia evolutiva del género *Gossypium* (tomado de Wendel *et al.*, 2010).

2.2.1 Las especies australianas

Los algodones australianos (subgénero *Sturtia*) comprenden 16 especies descritas (Wendel *et al.*, 2010) y un número de nuevas especies cuya descripción está por publicarse (Wendel *et al.*, 2010). Colectivamente, estas taxas incluyen los grupos genómicos C, D y K, con 2, 3 y 12 especies, respectivamente. Estos tres grupos de especies, de acuerdo a sus secuencias de ADN, son un linaje natural, lo cual es consistente con alineamientos anteriores definiendo las secciones *Sturtia* (genoma C), *Hibiscoidea* (genoma G) y *Grandicalyx* (genoma K). No obstante, las relaciones entre los tres grupos son aún confusas. Algunos datos colocan a *G. robinsonii* como la base de las especies australianas (Wendel y Albert, 1992), sugiriendo que la radiación de *Gossypium* sucedió de oeste a este, partiendo de la región más occidental del continente pero se mantiene como una pregunta abierta si la posición basal de *G. robinsonii* se mantendrá con nuevos datos (Wendel *et al.*, 2010).

En relación a la taxonomía dentro de los tres grupos de genomas australianos, existe muy poca incertidumbre sobre los genomas C y G ya que éstos están muy bien representados en colecciones y han sido estudiados a detalle. La taxonomía de las especies genoma K, las cuales se colocan en la sección *Grandicalyx*, se mantiene como una interrogante, ya que es un grupo poco representado en las colecciones,

aunque colectas realizadas en el área de Kimberley han aumentado el conocimiento de la diversidad dentro del grupo y han resultado en el descubrimiento de al menos 7 especies (Fryxell *et al.*, 1992). Estas especies inusuales poseen una geografía, morfología y ecología distintivas, y exhiben características propias de adaptaciones al fuego. En particular, son plantas herbáceas, perennes con un patrón de crecimiento bi-temporal donde no hay crecimiento vegetativo durante la temporada seca, cuando mantienen rizomas subterráneos. Estos comenzarán un nuevo ciclo de crecimiento con el inicio de una nueva temporada de lluvias o después de un incendio. La especie de la sección *Grandicalyx* poseen flores que se encuentran erectas cuando abren pero que tras la polinización se convierten en péndulas. En la madurez las cápsulas liberan semillas que presentan escasas cerdas y poseen cuerpos de grasa que atraen a las hormigas que favorecen la dispersión (Fryxell *et al.*, 1992; Wendel *et al.*, 2010).

2.2.2 Especies Afro-Asiáticas

Catorce especies africanas y árabes se reconocen en el tratado taxonómico de Fryxell (1992), que es el más reciente del género. Ellas se agrupan en el subgénero *Gossypium*. La sección taxonómica *Gossypium* contiene cuatro subsecciones mientras que la sección *Serrata* contiene únicamente a *G. trifurcatum* encontrado en el área desértica del este de Somalia. La presencia de hojas dentadas eleva la posibilidad de que éste no pertenezca a *Gossypium*, pero trabajos moleculares recientes han establecido claramente esta entidad pobremente conocida como una especie inusual de algodón (Wendel *et al.*, 2010). Este último ejemplo enfatiza la naturaleza provisional de mucha de la taxonomía de las especies africanas y árabes de *Gossypium* las cuales requieren de exploración básica y estudios sistemáticos. Dentro de la sección *Pseudopambak*, el reconocimiento y definición de especies está basada en limitado material de herbario (por ejemplo *G. benadirensis*, *G. bricchettii*, *G. vollesenii*) y no hay germoplasma colectado (Wendel *et al.*, 2010).

Desde un punto de vista citogenético, las especies de África y Arabia exhiben una diversidad considerable, comprendiendo cuatro de los ocho grupos genómicos (A, B, E y F). El genoma A incluye dos algodones cultivados de la subsección *Gossypium*, *G. arboreum* y *G. herbaceum*. Tres especies africanas de la subsección *Anomala* (*G. anomalum*, *G. captis-viridis* y *G. triphyllum*) comprenden el genoma B. *Gossypium trifurcatum* podría también pertenecer al genoma B. La única especie del genoma F, *G. longicalyx*, es citogenéticamente distinta (Phillips y Strickland, 1966),

aislada morfológicamente (Fryxell, 1971; Valiček, 1978; Vollesen, 1987) y quizás adaptada a ambientes con una humedad moderada en comparación con otras especies diploides de *Gossypium* (Wendel *et al.*, 2010).

2.2.3 Especies Americanas

Los algodones diploides americanos (*Gossypium* L., subgénero *Houzingenia* Fryxell), forman un grupo monofilético (genoma D; Small y Wendell, 2000) de trece especies, distribuidas principalmente en México occidental, extendiéndose hasta Arizona y Baja California, con dos especies disyuntas, una en Perú y otra en las islas Galápagos.

En los estudios del grupo la especie *Gossypium gossypioides*, es la más conflictiva, ya que las evidencias morfológicas (Fryxell, 1979, 1992), citogenéticas (Brown y Menzel, 1952) y moleculares (Wendel *et al.* 1995b) se contraponen acerca de su afinidad filogenética con otros algodones americanos. Se han evaluado 16.4 kb de secuencias de ADN (ADN de Cloroplasto, 7.3 kb; de ITS núcleo-ribosoma, 0.7 kb y de genes nucleares únicos, 8.4 kb; Cronn *et al.*, 2002). Estos datos, junto con algunas evidencias previas de fracciones repetidas del genoma, muestran la compleja historia de *G. gossypioides*, que posiblemente incluye algunos eventos separados de introgresión de algodones divergentes, restringidos hoy a distintos hemisferios. La reconstrucción más reciente da como resultado un proceso de especiación ocasionado por la introgresión y recombinación de varios genomas (Cronn *et al.*, 2003, figura 1.4). Esta especie sufrió una introgresión nuclear de una especie africana, poco tiempo después de la divergencia con el resto del grupo de algodones americanos. Recientemente, parece que a partir de la hibridación con una especie mexicana, sufrió una introgresión en ADN de cloroplasto y, posiblemente, otra introgresión nuclear críptica. La especie *G. gossypioides*, brinda un ejemplo de la naturaleza quimérica de los genomas de algunas plantas y de las complicaciones filogenéticas que ésta produce. *Gossypium gossypioides*, es también la única especie diploide americana que muestra evidencias de contacto ancestral con el genoma A y por lo tanto, se piensa que es un ancestro de los *Gossypium* tetraploides (Cronn *et al.* 2003).

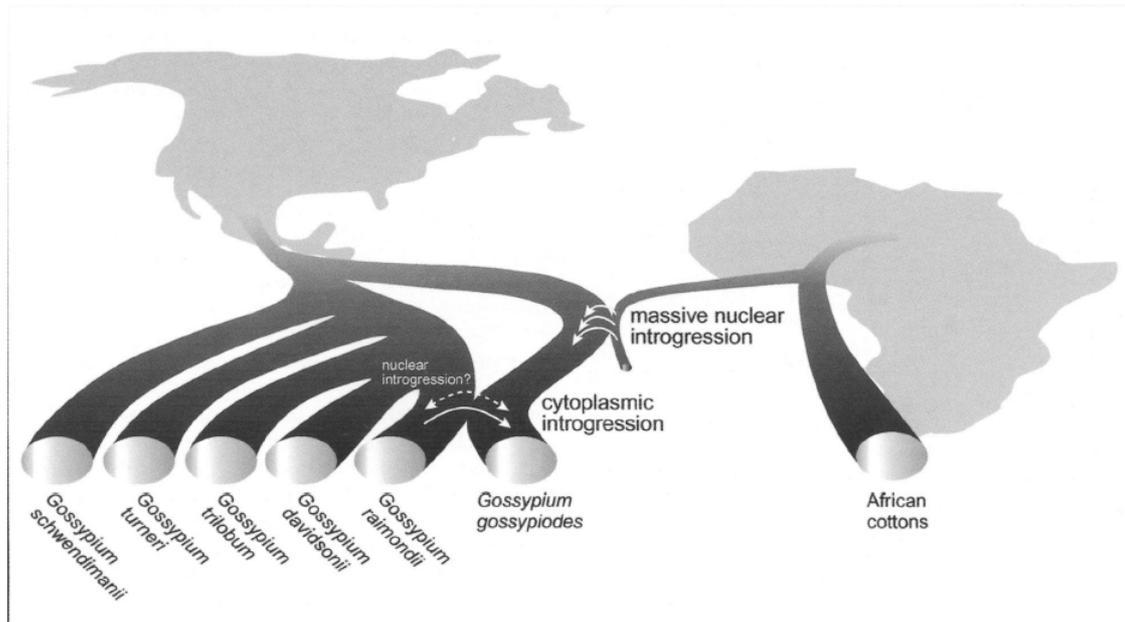


Figura 1.4 Hipótesis de las relaciones filogenéticas del genoma D (subgénero *Houzingenia*) propuesto por Cronn *et al.*, 2003.

2.2.4 Los *Gossypium* alotetraploides (AADD)

Los algodones alotetraploides son resultado de la unión de dos genomas, A y D, los cuales surgieron de la división más temprana del género, evolucionaron en hemisferios diferentes y divergieron por millones de años, aislados uno del otro (Wendel *et al.*, 2010). Los algodones aloploiploides entonces contienen copias duplicadas y ligeramente divergentes de la mayoría de sus genes. En promedio, la divergencia de las secuencias entre copias de estos genes es del 3% al 4%, aunque existe una varianza importante en esta media. Además, se ha estimado en algunos genes del genoma D que la tasa de mutación es 1.4 veces mayor que en el genoma A (Liu *et al.* 2001).

Al considerar la filogenia en un contexto temporal y a la luz de la historia de las placas tectónicas, se infieren múltiples dispersiones intercontinentales y otros episodios de trayectoria transoceánica durante la historia evolutiva de *Gossypium*. Estas incluyen al menos una dispersión entre Australia y África, otra hacia América (probablemente México) conduciendo a la evolución de los diploides del genoma D y una segunda, muy posterior, colonización del Nuevo Mundo por un ancestro del genoma A que dio lugar a los aloploiploides con genoma AD. La dispersión a larga distancia jugó un papel muy importante, no sólo en la diversificación de las principales líneas evolutivas sino también en la especiación dentro de los grupos

genómicos de *Gossypium*. Los ejemplos incluyen las dispersiones del sur de México a Perú (*G. raimondii*, Stephens 1966; Álvarez, *et al.* 2005), del norte de México a las Islas Galápagos (*G. klotzschianum*, Stephens 1966; Álvarez, *et al.* 2005), de América del Sur a las Islas Galápagos (*G. darwinii*, Stephens 1966), de África a las Islas de Cabo Verde (*G. capitata-viridis*, Stephens 1966), y de los geotrópicos a las Islas Hawaianas (*G. tomentosum*, DeJooe y Wendel 1992).

Un misterio clásico de la botánica durante medio siglo fue el origen de la alopoliploidía de los *Gossypium* americanos, pero ahora se ha podido responder gracias al uso de herramientas moleculares. Con respecto a cuándo se generó la alopoliploidía, datos de secuencias genéticas demuestran convincentemente que se originó en el Pleistoceno Medio, entre uno y dos millones de años antes de los primeros registros de *Homo sapiens*, y que por lo tanto, es poco probable que *Homo sapiens* interviniera en el proceso de hibridación. Por otra parte, ahora se conoce que ambas especies existentes del genoma A (*G. arboreum* y *G. herbaceum*) son igualmente divergentes del genoma A de las especies alopoliploides; ambas especies pasaron por un proceso de domesticación y no se conocen poblaciones silvestres de ellas. A nivel genómico, difieren de los genomas alotetraploides A, translocaciones recíprocas de los brazos cromosómicos (Brown y Menzel, 1950; Gerstel, 1953; Menzel y Brown, 1954), aunque por mucho tiempo se pensó que *G. herbaceum* se asemejaba más al genoma A donador que *G. arboreum*.

El pariente vivo más cercano del donador del genoma D progenitor es *G. gossypoides* (Wendel *et al.*, 1995a, Cronn *et al.*, 2003, Wendel *et al.*, 2010), aunque se tuvo otro candidato: *G. raimondii* (Endrizzi *et al.*, 1985). Un aspecto de la historia de los algodones poliploides que se ha aclarado es que todos contienen un citoplasma del genoma A y muy probablemente de una única fuente. Los estudios que emplean genes nucleares (heredados biparentalmente) conducen a la misma conclusión. Así, la evidencia indica que todos los algodones alopoliploides provienen del mismo ancestro (Wendel *et al.*, 2010).

Considerando un origen pleistocénico de las especies de algodones alopoliploides, se podría inferir que su diversificación morfológica y su expansión debieron de haber sucedido rápidamente. Actualmente se reconocen cinco especies alopoliploides: *G. darwinii* es nativa de las Islas Galápagos, donde forma poblaciones extensas y abundantes en algunas áreas. *G. tomentosum*, de las Islas Hawaianas presenta una estructura poblacional mucho más difusa, encontrándose en su mayoría

como individuos dispersos y poblaciones pequeñas en varias islas. Un tercer aloploiploide, *G. mustelinum*, es una especie poco común restringida a una región relativamente pequeña al norte de Brasil (Wendel *et al.* 1994). *G. barbadense*, por su parte, presenta una distribución natural sureña, concentrada en el tercio septentrional de América del Sur pero con una amplia región de superposición con *G. hirsutum* en el Caribe. Finalmente, *G. hirsutum*, posee una amplia distribución natural, colectivamente abarcando una riqueza morfológica que cubre el continuo de silvestre a domesticado. *G. hirsutum* se distribuye silvestre en las dunas y selvas bajas de Mesoamérica e inclusive se ha reportado en islas distantes del Pacífico como las Islas Salomón o las Marquesas. Estas dos últimas especies fueron domesticadas de forma independiente por culturas prehispánicas (Fryxell, 1979).

La distribución de las especies aloploiploides sugiere que la poliploidía condujo a la invasión de un nuevo nicho ecológico. Fryxell (1979) notó que, en contraste con la mayoría de las especies diploides, las especies aloploiploides se localizan en hábitats costeros, al menos aquellas formas que posiblemente son en verdad silvestres. Dos especies, ambas endémicas de islas (*G. darwinii* y *G. tomentosum*), se encuentran restringidas a zonas cercanas a la línea costera y otras dos especies (*G. barbadense* y *G. hirsutum*) se encuentran en hábitats litorales, por lo que pensaba en que la adaptación a los ambientes litorales de los nuevos aloploiploides les permitió sobrevivir a los fluctuantes niveles oceánicos que caracterizaron el Pleistoceno, facilitando el establecimiento del nuevo linaje poliploide, que además facilitaba la rápida dispersión de las semillas tolerantes al agua salada. Por los tiempos de divergencia ahora sabemos que esto ocurrió en poco tiempo, lo cual explica por qué hay pocas diferencias en 5000 pares de bases del ADN del cloroplasto entre los aloploiploides (Wendel, 1989; Wendel and Albert, 1992; Wendel *et al.*, 2010).

Wendel y Albert (1992) plantearon la posibilidad de una radiación previa a la domesticación del genoma A en Asia, seguida por una migración transpacífica, en lugar de transatlántica. Esta posibilidad es apoyada por la biogeografía de las especies de D-genoma, siguiendo la hipótesis de que se originaron en el oeste de México.

2.3 Tricomias de las semillas

Las semillas y las fibras de su cubierta son extraordinariamente diversas en *Gossypium* por lo que se han realizado varias revisiones al respecto (Basra y Malik, 1984; Ryser, 1985; DeLanghe, 1986; Kosmidou-Dimitropoulou, 1986). Algunos datos relevantes se mencionan a continuación.

Algunas especies diploides con genoma D (*G. thurberi*, *G. trilobum*, *G. davidsonii* y *G. klotzschianum*) no poseen fibras aparentes, no obstante sí están presentes aunque como estructuras reprimidas durante el desarrollo. De manera similar, tres especies diploides del genoma D de la subsección *Cauducibracteolata* parecen carecer de fibras, pero de hecho poseen estos filamentos firmemente presionados contra la semilla.

Análisis filogenéticos de las tasas de crecimiento muestran que la innovación genética de “prolongar la elongación de la fibra” surgió en el linaje de los genomas F y A, lo cual pudo haber fomentado la domesticación original de los algodones del genoma A. Este rasgo de elongación prolongada fue heredada a los algodones aloploidos, lo cual fue un atributo clave para su eventual domesticación (Wendel *et al.*, 2010). Sobre las bases genéticas y del desarrollo que dieron lugar a tan diversas morfologías se sabe relativamente poco. Aumentar el conocimiento sobre los cambios ocurridos durante la evolución y la domesticación del género puede contribuir al mejoramiento de este y otros cultivos (Wendel *et al.*, 2010).

En la naturaleza, la dispersión de semillas de *Gossypium* a menudo sigue un modelo de “sacudido”, donde las cápsulas maduras y erectas abren a lo largo de las suturas y las semillas se distribuyen cerca de la planta parental mientras el viento sacude sus ramas (Fryxel, 1979). Éste es posiblemente el método ancestral de dispersión de semillas del género, ya que ocurre en todas las especies de los genomas B, C, E, F y parte del D y en una especie del genoma G (*G. bickii*). Las especies de los genomas A, AD, G, K y algunas del genoma D, han evolucionado otro mecanismo de dispersión de semillas, que es tal vez una de las innovaciones más significativas: los cuerpos de grasa para que las hormigas atraídas por éstos dispersen las semillas. El desarrollo de “fibras trenzables” aparentemente ocurrió una única vez en la historia de *Gossypium*, en el ancestro de las dos especies del genoma A que se convirtieron en progenitor del subgenoma A de los algodones tetraploides (Hovav *et al.*, 2008, Chaudhary *et al.*, 2009).

La mayoría de los estudios sobre fisiología del algodón se han llevado a cabo comparando a uno o pocos individuos silvestres tomados de la Península de Yucatán

o líneas cultivadas con poco mejoramiento, por lo que los datos que se describen a continuación deben ser tomados con reserva, ya que la diversidad contenida en los silvestres es mucho mayor que en los cultivados. Nosotros hemos observado gran variedad de formas y colores de fibra en la población de la que provienen las semillas con las que se compara por lo general a los cultivados. Sin embargo, después de hacer la aclaración precautoria, es relevante revisar la información que se ha generado sobre la diversidad de las fibras y es claro que estudios futuros debiesen enfocarse al estudio morfológico de las poblaciones silvestres.

Las fibras de las plantas cultivadas son compuestas por una única célula de casi pura celulosa, la cual se alarga durante el desarrollo hasta alcanzar de dos a seis centímetros, dependiendo de la especie. En un estudio comparado entre silvestres y domesticados se encontró que semillas provenientes de la Península de Yucatán están compuestas principalmente de celulosa y suberina, y se alargan, como máximo, menos de un centímetro debido a un retardo en el tiempo de inicio de la síntesis de la pared secundaria (Booth, 1968). Estos datos se contraponen a otros realizados principalmente con intereses agronómicos en los que se explica la importancia de la temperatura tropical durante las noches, ya que eso favorece la síntesis de celulosa y aumenta los rendimientos en los cultivares, y otros en donde se explica que las propiedades de la fibra se ven fuertemente afectadas por la calidad del sustrato, enfermedades y plagas en las plantas, lo cual al final es más importante que el potencial genético de los cultivares (Booth, 1968, Taliercio y Haigler, 2011). Esta información resulta muy relevante para comprender por qué es importante realizar nuevos estudios sobre el síndrome de domesticación de *G. hirsutum*, comparando poblaciones silvestres, plantas escapadas de cultivo y domesticadas, ya que se ha descrito el porcentaje de fibra en los frutos como las principales características para diferenciar a los silvestres de los domesticados, pero no se ha descrito cómo cambiarían al crecer sin los cuidados de los cultivares o en las condiciones ambientales que habitan las plantas silvestres, tales como menor cantidad de agua, suelos altos en sales y con pocos nutrientes, con competencia con otras plantas y en interacción con muchos otros organismos, etcétera. Se ha propuesto que las plantas que viven asociadas a entornos modificados por el humano son semi-domesticadas, mientras las que tienen menor influencia son las verdaderas silvestres, sin embargo este argumento es poco convincente si se considera que, por un lado, los ecosistemas de las silvestres son habitados por comunidades humanas (Rapp *et al.*, 2010) y por otro,

las dinámicas ecológicas desde su especiación parecen indicar alta capacidad de adaptarse a vivir en ambientes dinámicos, donde producen menos fibra bajo condiciones de estrés.

Se ha especulado sobre las presiones selectivas que propiciaron la fijación de las fibras epidérmicas de las semillas o tricomas. Fryxell (1979), sugirió que las fibras alargadas fomentaban la dispersión por aves o Pterodáctilos (aunque ahora sabemos que estos últimos eran carnívoros y se extinguieron 50 millones de años antes, aproximadamente). La dispersión por aves gana credibilidad por la observación de un nido de ave al noroeste de Puerto Rico que contenía numerosas semillas de *G. hirsutum* silvestre, así como una colección de *G. darwinii* en un nido de pinzón en las Islas Galápagos. En México hemos observado en varias poblaciones, aves jalando fibras con semillas y después volar grandes distancias. Se puede también especular que las fibras sirven para inhibir la germinación mientras no haya suficiente humedad para saturar las fibras, ya que si la germinación ocurriera después de una lluvia ligera podría no haber suficiente agua para la supervivencia de las plántulas. A este respecto, la cubierta grasosa de las fibras podría repeler el agua en cierta medida y por lo tanto prevenir la germinación prematura. Otra posibilidad relacionada es que las fibras funcionan como “incubadoras biológicas” para facilitar la germinación sólo cuando las condiciones ecológicas son adecuadas por medio del reclutamiento de comunidades microbianas bajo regímenes de humedad apropiadas (Fryxel, 1979). Finalmente, dos especies del genoma G (*G. austral* y *G. nelsonii*) presentan fibras erectas y rígidas que facilitan la dispersión por el viento ya que la firmeza de los tricomas ayuda a forzar la salida de las semillas de los lóculos de las cápsulas (Fryxel, 1979, Wendel *et al.*, 2011). Sin lugar a dudas, las fibras permiten que las semillas sean dispersadas grandes distancias por viento y agua, pero sigue siendo necesario estudiar si las diferencias en porcentajes de fibra son hereditarias y confieren ventajas en su reproducción.

3. Origen y distribución del subgénero *Houzingenia* Fryxell, del género *Gossypium* L. en México

Resumen

Las especies diploides americanas del género *Gossypium* L., comprenden un ensamble monofilético y citogenético, conocido taxonómicamente como el subgénero *Houzingenia* Fryxell. Este grupo incluye 11 especies distribuidas en México, una en Perú y otra en las Islas Galápagos. Dentro de este grupo podemos encontrar a uno de los linajes parentales de los algodones alotetraploides cultivados, *Gossypium hirsutum* y *G. barbadense* (Álvarez *et al.*, 2005).

El objetivo de este capítulo es documentar el origen y distribución de los parientes cercanos a *Gossypium hirsutum* que habitan en México. Para ello se realizó una revisión de sus relaciones filogenéticas, posteriormente se revisaron los ejemplares del herbario nacional MEXU y se seleccionaron los registros para utilizarlos en la modelación del nicho ecológico utilizando los siguientes criterios: identificado o colectado por las autoridades taxonómicas del grupo y con coordenadas geográficas consistentes con la descripción de la localidad. Los resultados de la modelación del área de distribución potencial de nueve de las once especies (dos no tuvieron registros suficientes) se presentan gráficamente y aportan evidencia para sostener a México como el centro de origen y diversidad subgénero *Houzingenia*. La mayoría de las especies muestra una distribución restringida y bajas densidades poblacionales, además de habitar en regiones donde las actividades antropogénicas pudieran tener un alto impacto sobre la vegetación. Se recomienda verificar en campo la presencia de las especies en las áreas predichas e incluirlas en la revisión de la Norma Oficial Mexicana 059-SEMARNAT-2010, así como iniciar un programa activo para la conservación de las especies del género *Gossypium* en México.

3.1 Origen y distribución de especies diploides del género *Gossypium*

El género del algodón (*Gossypium*), incluye aproximadamente 52 especies distribuidas en regiones áridas y semiáridas de los trópicos y subtropicos (Fryxell, 1979, nueva especie Ulloa *et al.*, 2006, Wendel *et al.*, 2009). Las especies de este género presentan una gran variedad morfológica, citogenética y genómica, debida a la radiación global del género que llevó a la evolución de ocho grupos de especies diploides (genomas A-G y K). La historia evolutiva del género (ver detalles en el capítulo 2) incluye eventos consecutivos de dispersiones trans-oceánicas, invasión de nichos ecológicos, y una frecuencia sorprendentemente alta de hibridación entre los linajes. Los datos indican que el origen del género *Gossypium* se remonta a hace 10-15 millones de años y que tuvo una diversificación temprana y rápida de la mayoría de los grupos genómicos (Wendel y Cronn, 2003).

Existen cuatro linajes dentro del género, presentes en tres continentes: Australia (los del genoma C, G y K), América (genoma D) y África-Arabia (los del genoma A, B, F y E). Las especies con genoma D son trece: 11 especies distribuidas en México, una en Perú y otra en las Islas Galápagos. El centro de diversidad de este grupo es México y se sugiere que en esta región se originó y diversificó el subgénero *Houzingenia* (Wendel y Cronn, 2003), así como el linaje de los algodones alotetraploides (genoma AD; Álvarez *et al.*, 2005).

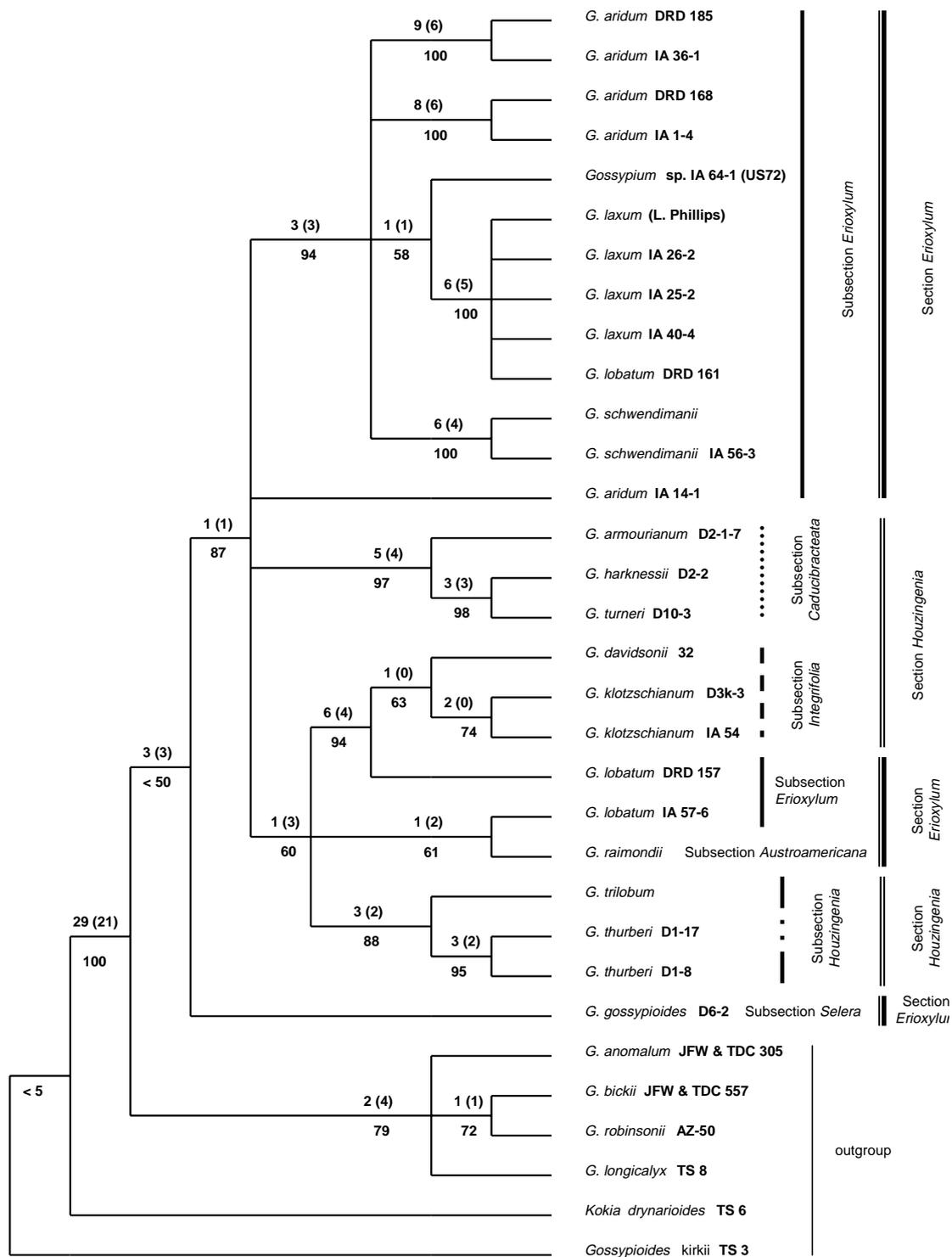


Figura 3.1. Tomada de Álvarez *et al.* (2005), árbol consenso estricto usando *AdhC*, (valores de *bootstrap* bajo las ramas), que muestra las especies incluidas en las secciones y subsecciones del grupo. Aunque trabajos posteriores con diferentes marcadores muestran más cercana la relación entre *Gossypium gossypoides* y *Gossypium raimondii* (Wendel *et al.*, 2011), los grupos se mantienen definidos.

Álvarez y colaboradores (2005), resolvieron las relaciones filogenéticas de este grupo usando tres genes nucleares: *A1341*, *CesA1b* y *AdhC*, misma que muestra que estas especies pertenecen a un clado único, que se agrupa en varias subsecciones (Fig. 3.1; *AdhC*). Los resultados del árbol consenso y la distribución conocida hasta el momento sugieren un escenario biogeográfico complejo, donde la divergencia inicial de la subsección *Selera* (Ulbrich) Fryxell, que se encuentra restringida o limitada a las zonas continentales de México (Oaxaca y Guerrero). Mientras que los miembros de la subsección *Erioxylum* (Rose & Standley) Prokhanov, cubren mayor área del oeste y suroeste de México (Sinaloa, Puebla, Guerrero, Michoacán, Colima y Jalisco), los miembros de la subsección *Caducibracteolata* Mauer, habitan en Sonora y Baja California (Álvarez *et al.*, 2005).

El patrón geográfico junto con la evidencia filogenética, sugieren una radiación rápida de este linaje diploide. Esta radiación pudo tener lugar en alguna región de México hace aproximadamente 6.7 millones de años (Senchina *et al.*, 2003; Álvarez *et al.*, 2005).

La historia evolutiva del género *Gossypium*, sugiere que los algodones alotetraploides se formaron hace 1.5 millones de años aproximadamente, luego de la divergencia de los progenitores diploides. Entre las especies existentes, *G. herbaceum* y *G. arboreum* son los parientes más cercanos al progenitor del genoma A, mientras que un ancestro de *G. raimondii* y *G. gossypioides* se ha propuesto como progenitor del genoma D (Liu *et al.*, 2001; Senchina *et al.*, 2003; Wendel y Cronn, 2003; Álvarez *et al.*, 2005), ya que aunque las investigaciones anteriores presentaban conclusiones opuestas, los trabajos más recientes parecen concluir que se trató de un ancestro de ambas especies (figura 2.3; Capítulo 2).

La taxonomía e historia evolutiva de las especies del subgénero *Houzingenia* está bien documentada (Fryxell, 1965b, 1967, 1979; Fryxell y Koch, 1987; Endrizzi *et al.*, 1985; DeJoode, 1992; Wendel y Albert, 1992; Wendel *et al.*, 1992, 2011; Liu *et al.*, 2001; Senchina *et al.*, 2003; Wendel y Cronn, 2003; Cronn *et al.*, 2002; Álvarez *et al.*, 2005). Sin embargo, aún es necesario conocer la distribución y el nicho ecológico que ocupan estas especies para que la necesidad de estrategias de conservación y manejo de sus poblaciones pueda ser analizada. La modelación del área de distribución potencial permitirá a su vez, ubicar a las especies en campo y realizar

estudios futuros sobre la diversidad genética y dinámicas ecológicas, que también son necesarios para proponer las unidades de manejo (ver capítulo 1).

3.2 Distribución actual y áreas de distribución potencial de *Gossypium* diploides que habitan en México

3.2.2 Métodos y resultados

Para conocer la distribución puntual y potencial de las especies diploides mexicanas, se revisó la base de datos de CONABIO (registros de la Red Mundial de Información sobre Biodiversidad (REMIB) y del Sistema de información de organismos vivos modificados (SIOVM), los archivos digitales de los ejemplares de la especie de los Herbarios de la Universidad de Arizona (ARIZ) y Nueva York (NYBG), además de los ejemplares del Herbario Nacional (MEXU) y Herbario del Instituto de Ecología A.C. (XAL). Los ejemplares de MEXU se analizaron y seleccionaron siguiendo los siguientes criterios: colectados o identificados por una autoridad taxonómica (principalmente P.A. Fryxell), presencia de coordenadas geográficas y concordancia de estas últimas con la descripción de la localidad.

Las predicciones de nicho ecológico se realizaron con GARP (por sus siglas en inglés *Genetic Algorithm for Rule-set Production*, algoritmo genético basado en reglas). GARP crea un modelo de nicho ecológico para una especie que representa las condiciones ambientales donde dicha especie sería capaz de mantener sus poblaciones (manual GarpDesktop: <http://www.nhm.ku.edu/desktopgarp/UsersManual.html>).

Como resultado se obtuvieron los mapas de distribución potencial de 9 de las once especies que habitan en México, los cuales se muestran en las figuras 3.2 a 3.11. (en ArcGis.9.3, referencia espacial WGS 1984 y sobre el mapa CONABIO 2003). Los registros obtenidos para *G. trilobum* (DC.) Skovsted y *G. turneri* Fryxell son insuficientes para modelar el área de distribución potencial, por ello la figura (3.11) contiene únicamente las localidades puntuales encontradas (en el caso de *G. trilobum* los ejemplares pertenecen a ARIZ).

Área de distribución potencial de *Gossypium aridum* (Rose y Standl.) Skovst.

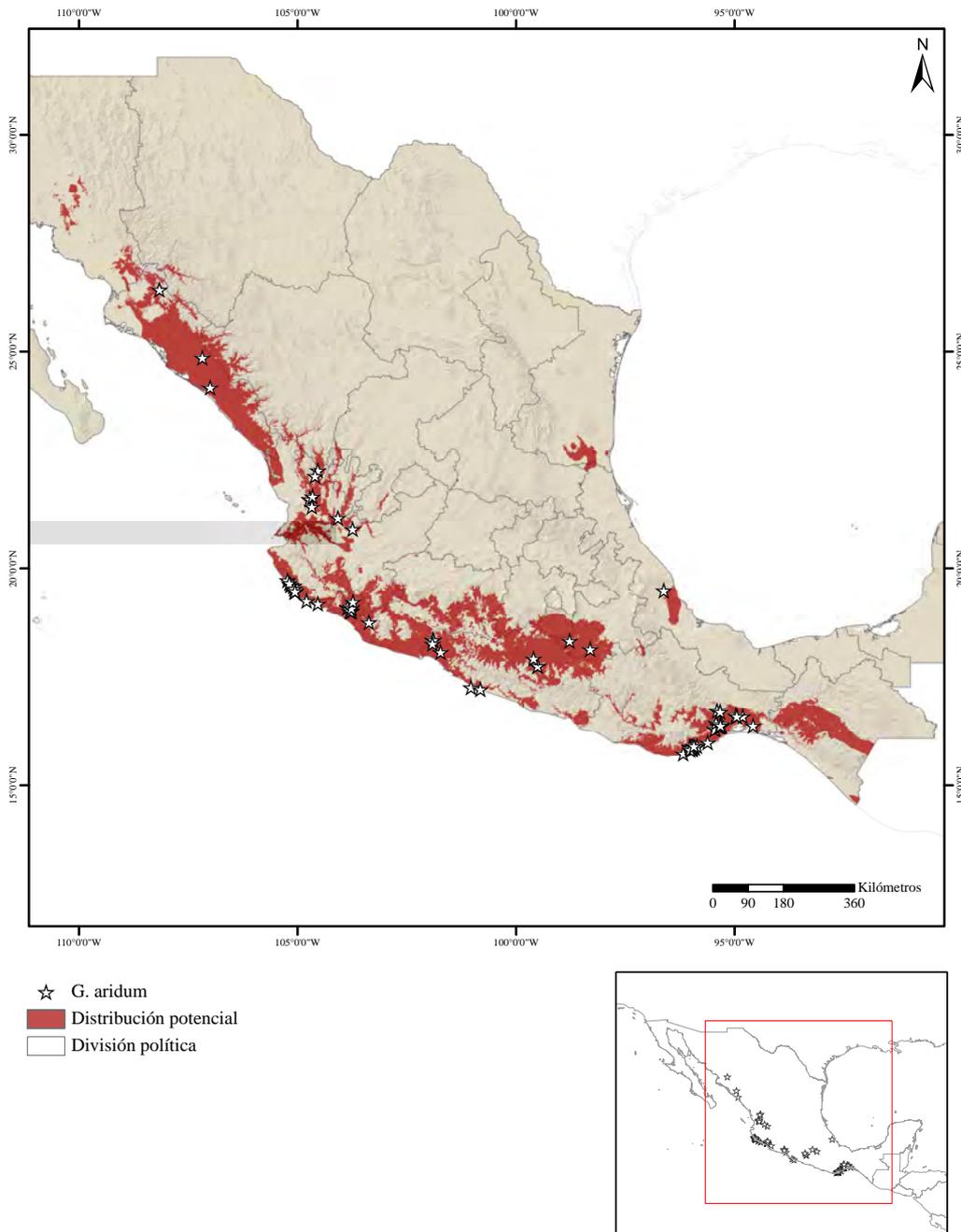


Figura 3.2. Distribución puntual y área de distribución potencial de *G. aridum*. Abajo: el recuadro indica el área ampliada en el mapa de arriba. Las estrellas indican los registros del MEXU seleccionados para el análisis.

Área de distribución potencial de *Gossypium armourianum* Kearney

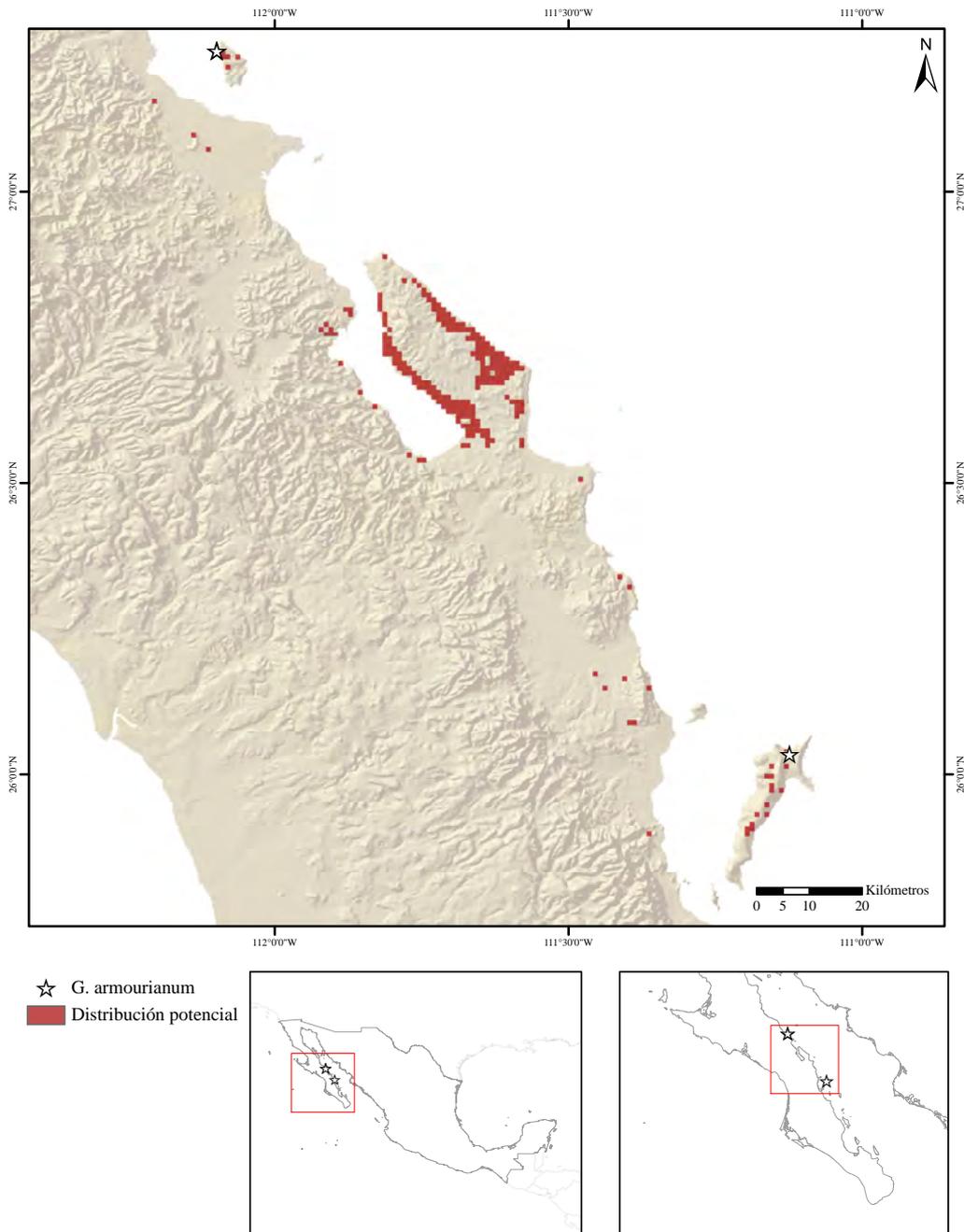


Figura 3.3. Distribución puntual y área de distribución potencial de *G. armourianum*. Abajo izquierda: el recuadro indica el área ampliada en el mapa de la derecha. Abajo derecha: el recuadro indica el área ampliada en el mapa de arriba. Las estrellas indican los registros del MEXU seleccionados para el análisis.

Área de distribución potencial de *Gossypium davidsonii* Kellogg

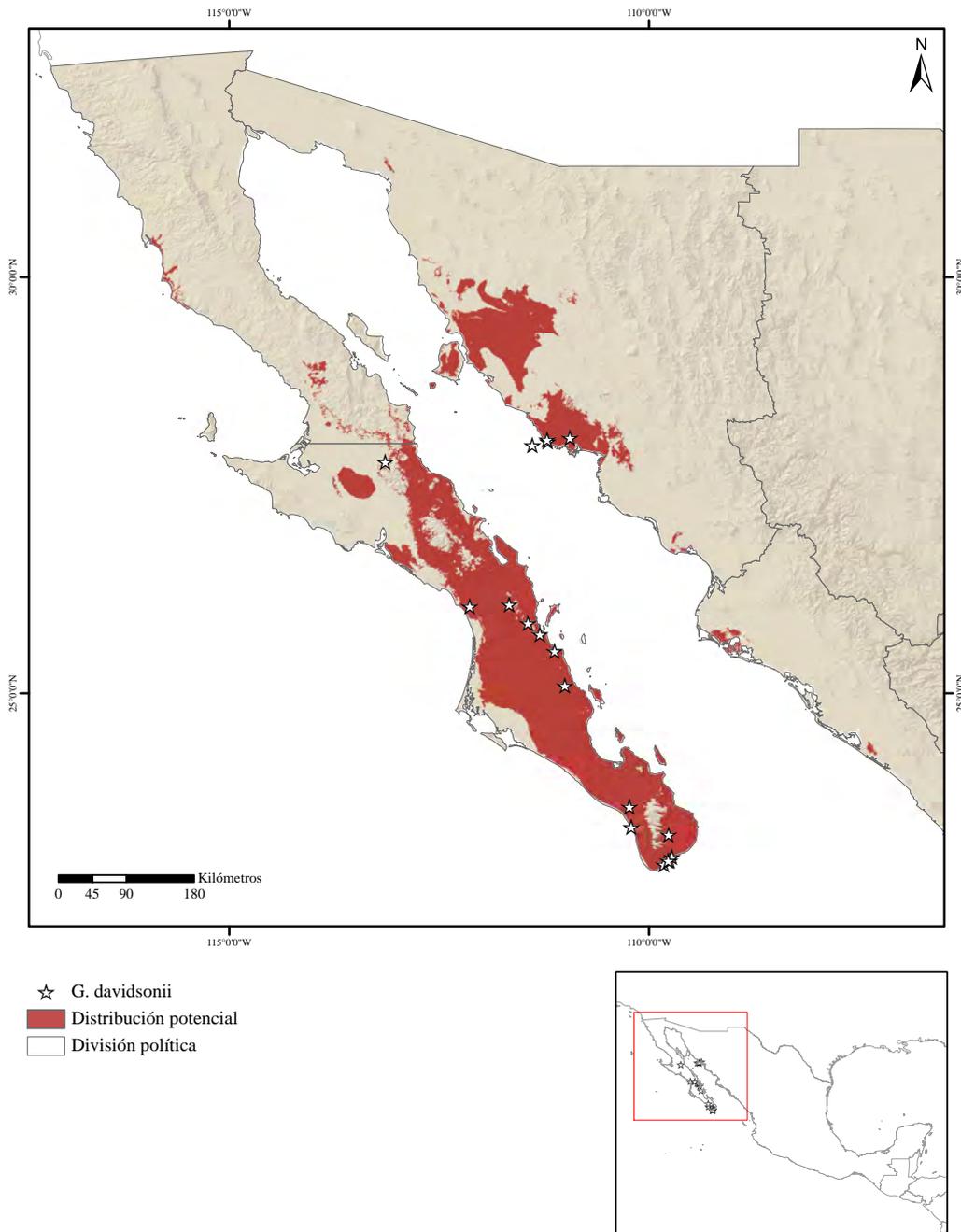


Figura 3.4. Distribución puntual y área de distribución potencial de *G. davidsonii*. Abajo: el recuadro indica el área ampliada en el mapa de arriba. Las estrellas indican los registros del MEXU seleccionados para el análisis.

Área de distribución potencial de *Gossypium gossypioides* (Ulbr.) Standl.

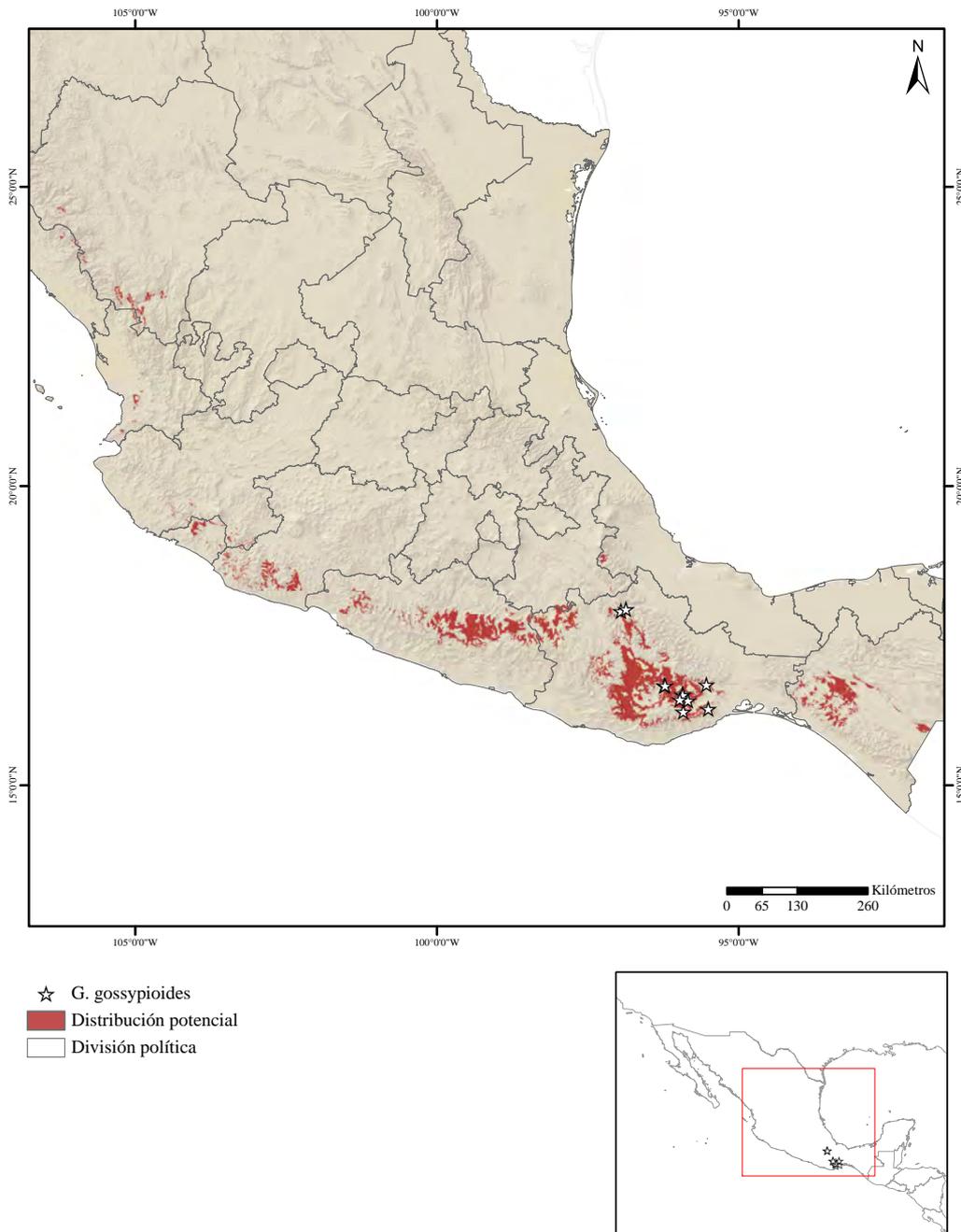


Figura 3.5. Distribución puntual y área de distribución potencial de *G. gossypioides*. Abajo: el recuadro indica el área ampliada en el mapa de arriba. Las estrellas indican los registros del MEXU seleccionados para el análisis.

Área de distribución potencial de *Gossypium harknessii* Brandege

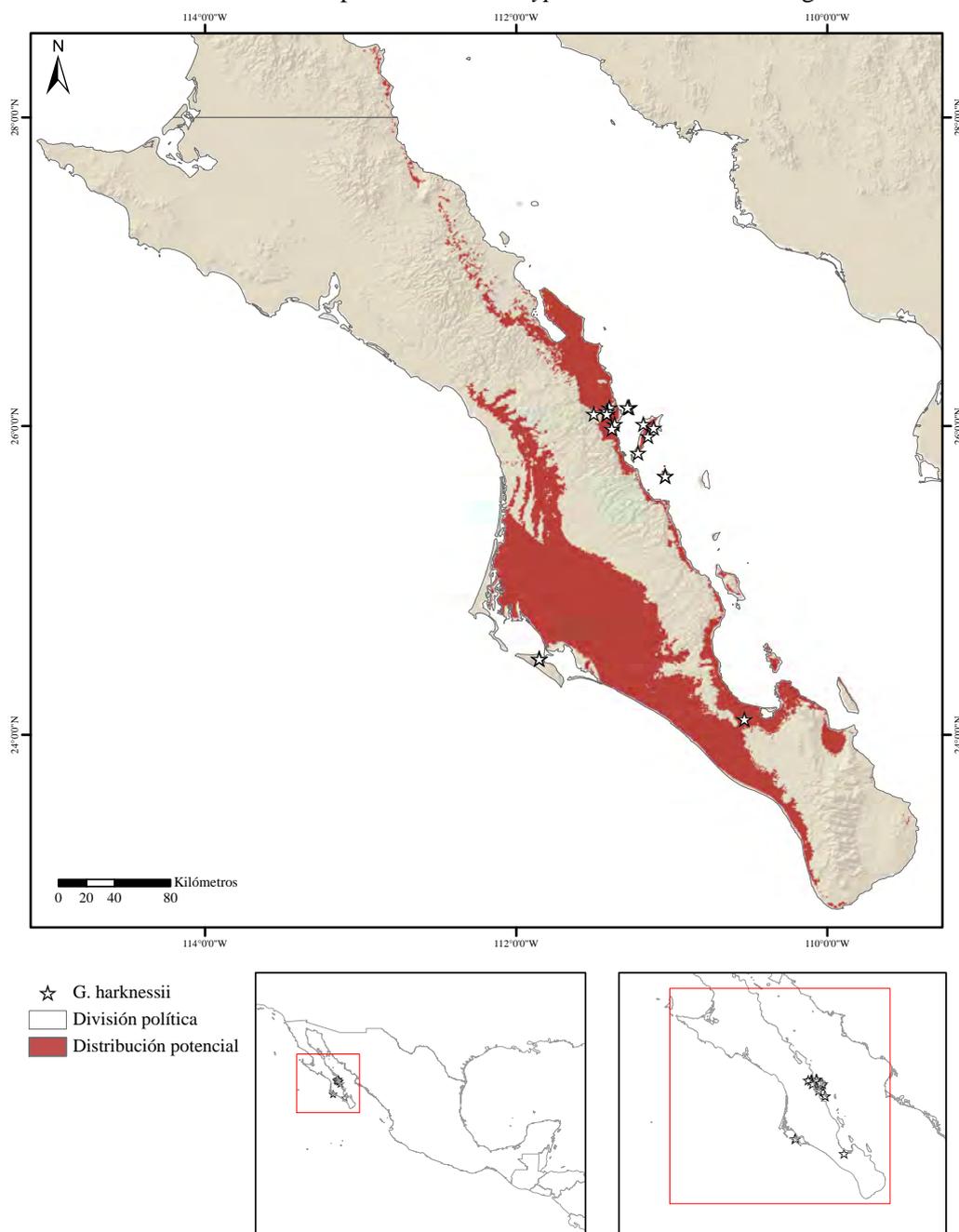


Figura 3.6. Distribución puntual y área de distribución potencial de *G. harknessii*. Abajo: el recuadro indica el área ampliada en el mapa de arriba. Las estrellas indican los registros del MEXU seleccionados para el análisis.

Área de distribución potencial de *Gossypium laxum* L.Ll. Phillips

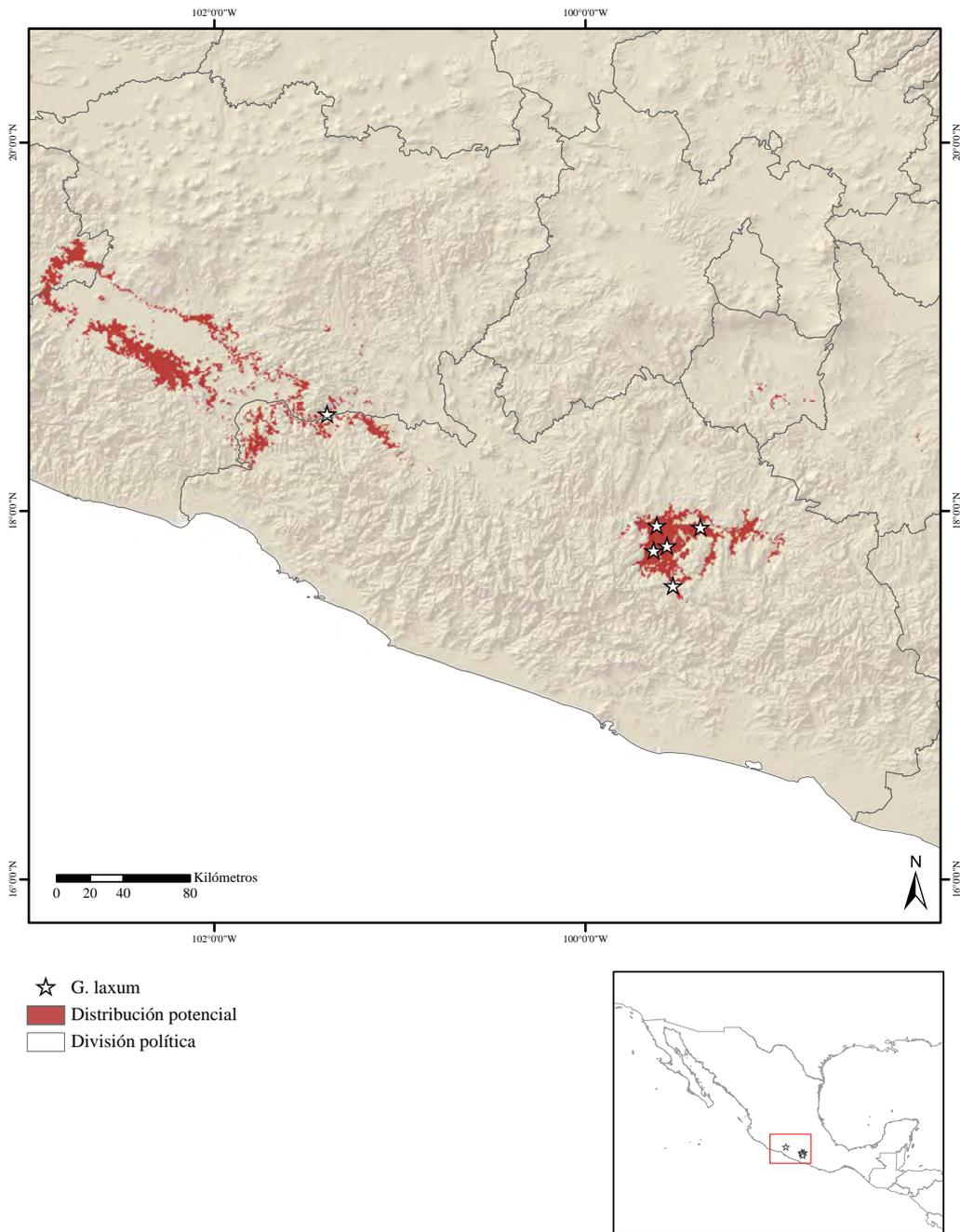


Figura 3.7. Distribución puntual y área de distribución potencial de *G. laxum*. Abajo: el recuadro indica el área ampliada en el mapa de arriba. Las estrellas indican los registros del MEXU seleccionados para el análisis.

Área de distribución potencial de *Gossypium lobatum* Gentry

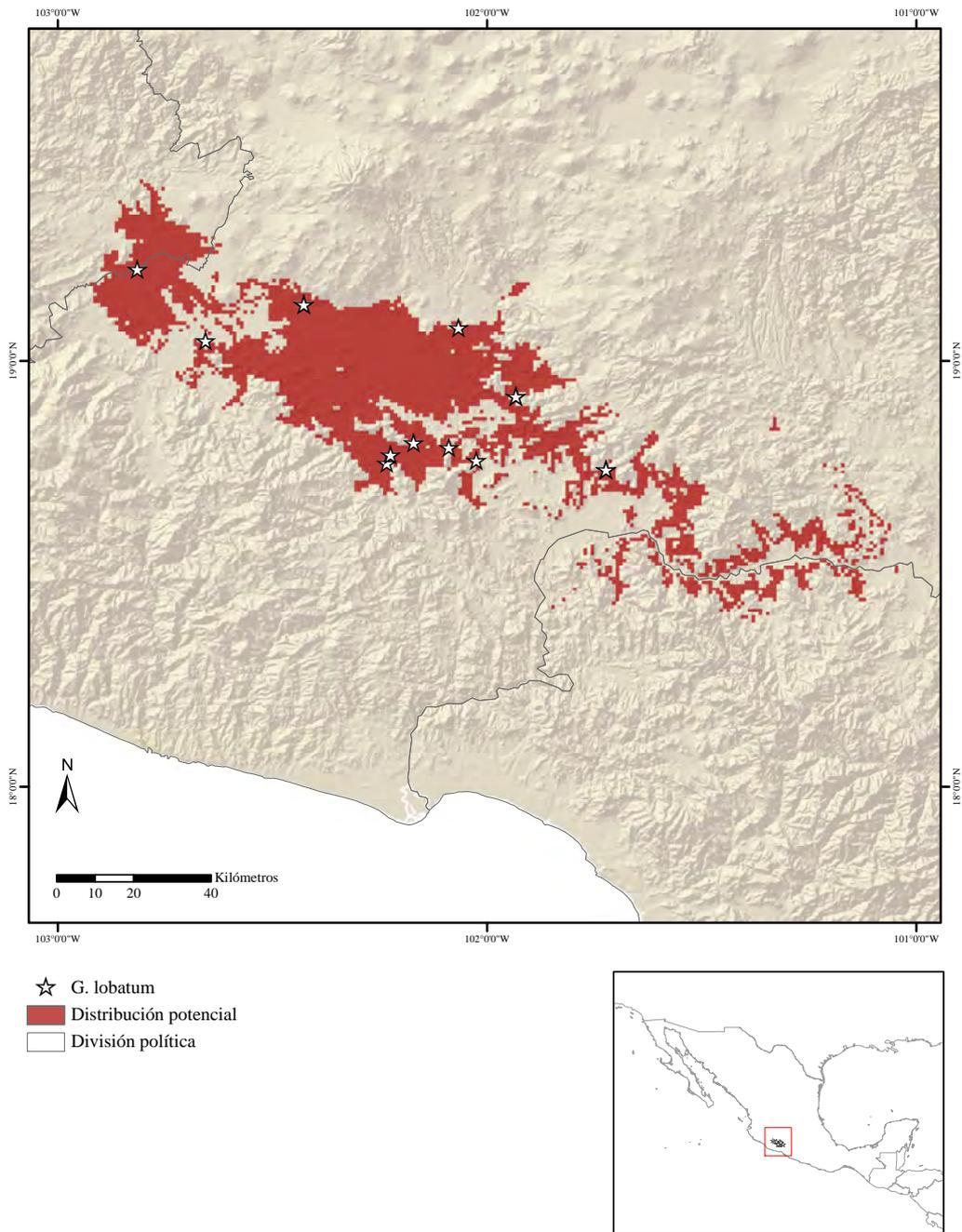


Figura 3.8. Distribución puntual y área de distribución potencial de *G. lobatum*. Abajo: el recuadro indica el área ampliada en el mapa de arriba. Las estrellas indican los registros del MEXU seleccionados para el análisis.

Área de distribución potencial de *Gossypium schwendimanii* Fryxell

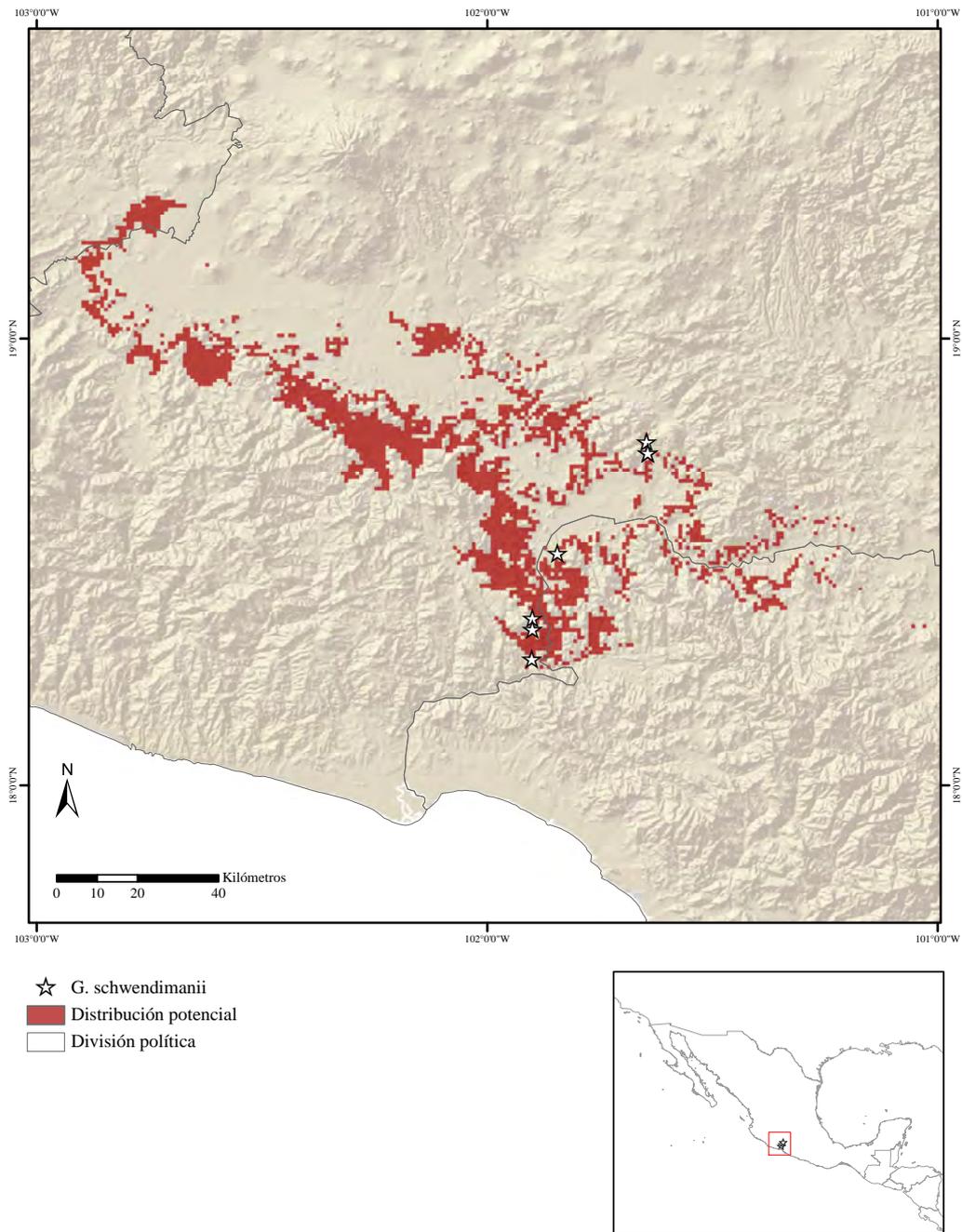


Figura 3.9. Distribución puntual y área de distribución potencial de *G. schwendimanii*. Abajo: el recuadro indica el área ampliada en el mapa de arriba, las estrellas indican los registros del MEXU seleccionados para el análisis.

Área de distribución potencial de *Gossypium thurberi* Tod.

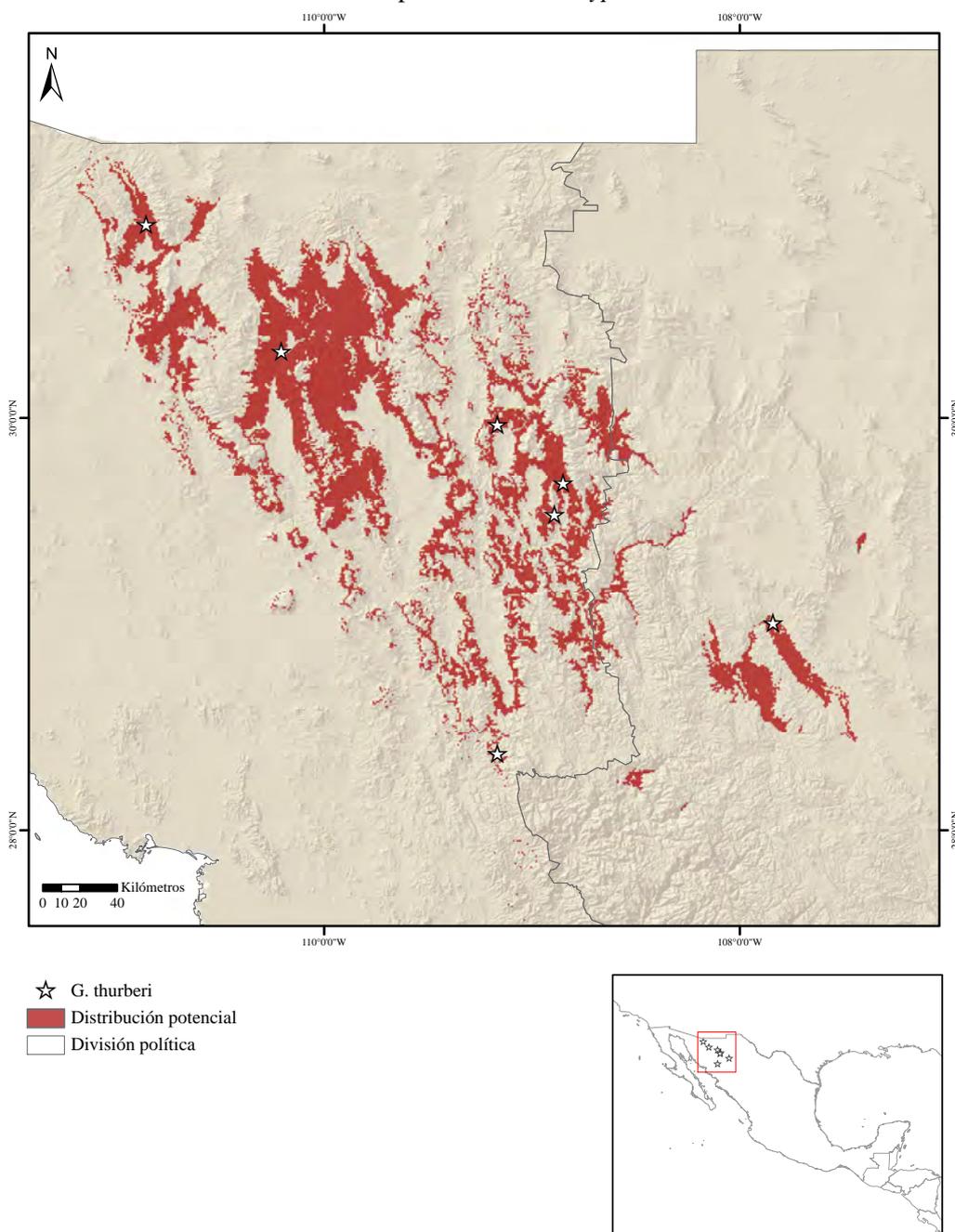


Figura 3.10. Distribución puntual y área de distribución potencial de *G. thurberi*. Abajo: el recuadro indica el área ampliada en el mapa de arriba. Las estrellas indican los registros del MEXU seleccionados para el análisis.

Registros puntuales de las especies *Gossypium turneri* Fryxell y *Gossypium trilobum* Skovst.

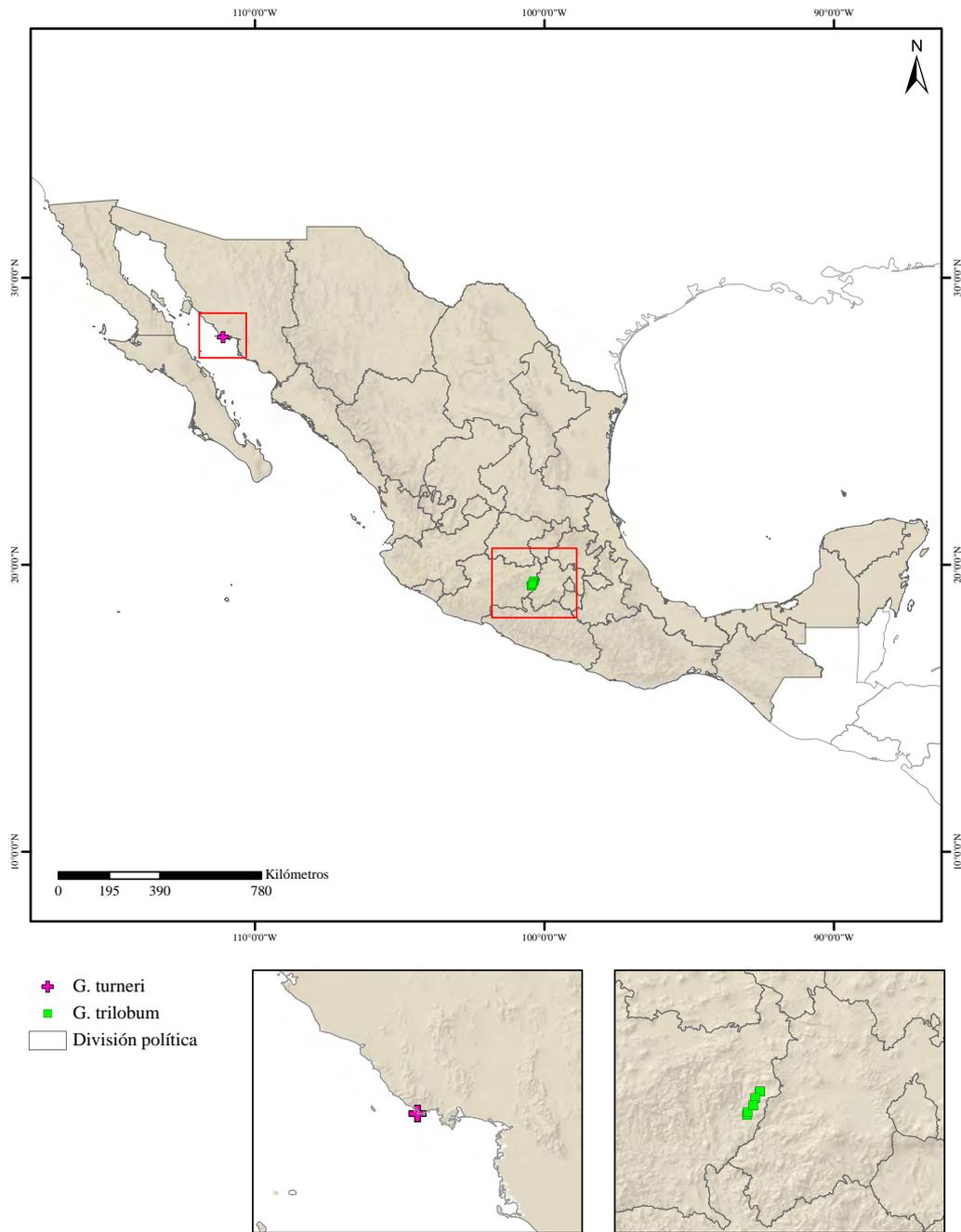


Figura 3.11. Registros seleccionados de *G. trilobum* y *G. turneri*. Arriba: los recuadros indican el área ampliada en los mapas de abajo. Los cuadros indican los registros de ARIZ seleccionados y la cruz indica el registro de MEXU seleccionado.

3.3 Discusión y conclusiones

Las áreas de distribución potencial de las especies mostraron en todos los casos áreas en donde existen las condiciones ambientales adecuadas para su existencia pero no se encontraron registros en los herbarios que documentaran su presencia. La modelación del nicho ecológico es una herramienta para conocer la distribución de las especies, pero ésta debe verificarse en campo (ver capítulo 4).

La información actual permite asegurar que el subgénero *Houzingenia* presenta distribuciones restringidas (a excepción de *G. aridum*, que incluso está presente en el Golfo de México al igual que *Gossypium hirsutum*). El número de registros en los herbarios y las descripciones de los ejemplares nos indican que las especies pueden ser poco abundantes además de restringidas, y las fechas de colecta nos muestran que es necesario realizar nuevas búsquedas de estas especies en campo, ya que algunas se localizan en áreas con alto impacto antropogénico.

Las subsecciones del subgénero fueron descritas originalmente con observaciones morfológicas y ecológicas, pero se agrupan en los estudios filogenéticos (aunque entre grupos sí ha cambiado la posición a medida que se obtiene más información; Álvarez *et al.*; 2005, Wendel *et al.*, 2011) y concuerdan con las distribuciones obtenidas en los modelos de predicción de nicho utilizados.

En el capítulo 2, se documentó la historia evolutiva del género la cual muestra una fuerte estructura genética y geográfica. Podemos inferir con los resultados del presente capítulo, que las condiciones ambientales también tienen una gran influencia sobre la estructura genética del subgénero *Houzingenia*, y como veremos en el capítulo siguiente, *Gossypium hirsutum*, también presenta una estructura genética que es consistente con la estructura geográfica y ecológica (Wegier *et al.*, 2011), generalidad importante para comprender el contexto evolutivo del grupo y diseñar una estrategia integral del mismo en México.

Legalmente, se recomienda incluir a estas especies en la siguiente revisión de la Norma Oficial Mexicana 059-SEMARNAT-2010. Así como iniciar un programa activo para la conservación de las especies del género *Gossypium* en México. La ley de bioseguridad de organismos genéticamente modificados, en el artículo 87, indica que uno de los criterios que se debe tomar en cuenta “para la determinación de los centros de origen y de diversidad” (en particular los centros de diversidad genética), son las regiones que actualmente albergan poblaciones de los parientes silvestres del

OGM de que se trate (en este caso es *Gossypium hirsutum*), incluyendo diferentes razas o variedades del mismo, porque constituyen una reserva genética del material.

Agradecimientos

A Oswaldo Oliveros y Daniel Ortiz su apoyo en el análisis y presentación de la información geográfica. Al Dr. Daniel Piñero, la Dra. Nadia Santini, y Biól. Jesús Alarcón, por su colaboración en el proyecto. Al comité sinodal de esta tesis, Valeria Alavez Gómez y Brian Urbano por sus comentarios al manuscrito. A la CARB-CONABIO por entender e interesarse en la información y difusión sobre *Gossypium hirsutum* y sus parientes, así como apoyar y coordinar el proyecto “Validación de información de registros biológicos y de mapas de distribución puntual y de los modelos de áreas de distribución potencial de las especies del género *Gossypium* en México” financiado en el proyecto marco “Continuación de la creación de capacidades institucionales y técnicas para la toma de decisiones en materia de bioseguridad (0051868)”.

Agradecemos el apoyo del Herbario Nacional (MEXU) y al Instituto de Biología, UNAM, por las facilidades otorgadas para la realización de esta investigación, a los herbarios ARIZ y New York Botanical Garden, por permitir la consulta de sus imágenes.

4. Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin

Resumen

Este capítulo engloba la historia desde la publicación del artículo “*Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (Gossypium hirsutum) at its centre of origin*”, publicado en la revista *Molecular Ecology*, en donde se concentran los resultados más relevantes de la tesis: la descripción geográfica y ecológica de las poblaciones de *Gossypium hirsutum* L.; los análisis sobre la diversidad y estructura genética de las poblaciones, así como los patrones históricos que la formaron; la confirmación de los resultados históricos usando a los transgenes de las plantas genéticamente modificadas como marcadores moleculares recientes; la discusión sobre las implicaciones de los resultados para la conservación de la especie; y finalmente, las conclusiones y perspectivas para continuar la línea de investigación.

4.1 Semblanza

El artículo “*Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (Gossypium hirsutum) at its centre of origin*”, fue publicado el día el 8 de septiembre de 2011 e impreso en octubre del mismo año en la revista *Molecular Ecology* (20(19):4182-94). Esta revista está catalogada como la quinta mejor revista de cuarenta y cinco en el área de biología evolutiva) y su factor de impacto es 5.52 ([http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1365-294X](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-294X)).

Los resultados del trabajo han sido presentados en los siguientes foros: *III Congreso Mexicano de Ecología* organizado por la Sociedad Científica Mexicana de Ecología, AC, en Boca del Río Veracruz, en abril de 2011; Curso *Organismos genéticamente modificados y bioseguridad* en Santiago de Cali, Colombia, en marzo de 2012 organizado por la Corporación Colombiana de Investigación Agropecuaria (CORPOICA), *Empresa Brasileira de Pesquisa Agropecuária* (EMBRAPA), el Instituto Alexander Von Humboldt y la Universidad de Cali y financiado por el

proyecto GEF Lac Biosafety; en el Congreso *Evolution 2012* organizado por *the American Society of Naturalists (ASN)*, *the Canadian Society for Ecology and Evolution (CSEE)*, *the European Society for Evolutionary Biology (ESEB)*, *the Society for the Study of Evolution (SSE)*, y *the Society of Systematic Biologists (SSB)* realizado en julio de 2012 en Ottawa, Canadá; y durante la COP/MOP 6 (Sexta reunión de la Conferencia de las partes en el convenio sobre la diversidad biológica que actúa como reunión de las partes en el protocolo de Cartagena sobre seguridad de la biotecnología) en el evento paralelo organizado por ENSSER realizado en octubre del 2012 en Hyderabad, India; además de seminarios en el departamento de Ecología de la Escuela Nacional de Ciencias Biológicas del IPN, Facultad de Ciencias de la UNAM y talleres organizados por la CONABIO, CIBIOGEM, INIFAP y otras instancias.

Hasta la fecha el trabajo se ha citado en publicaciones científicas (Cerritos *et al.*, 2012; Weiss 2012; Zhang 2013; Pan 2013), documentos de agencias gubernamentales (SEMARNAT 2012; Pérez *et al.*, 2012) y de organizaciones no gubernamentales.

A raíz de su publicación, la prensa retomó algunas ideas en reportajes realizados por *Inter Press Service*, el sitio SciDeb.Net, entre otros, las cuales fueron posteriormente difundidas en múltiples medios electrónicos nacionales e internacionales. También se realizaron entrevistas en radio (pueden consultarse por internet) en el programa *La esencia de la ciencia* con Javier Mena (<http://radiomexicointernacional.imer.gob.mx/>) del Instituto Mexicano de la radio, en el programa *Aqua* de Radio Educación (<http://www.e-radio.edu.mx/>) y el podcast del grupo editorial Reforma (www.reforma.com).

4.2 Síntesis del artículo “*Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin*” (omitiendo las figuras, metodología, resultados y materiales suplementarios)

Los complejos entre las plantas cultivadas y sus parientes silvestres en sus centros de origen de cultivos y/o diversidad son buenos modelos biológicos para comprender diversas preguntas sobre las dinámicas ecológicas y evolutivas que los afectan, en concreto la dinámica del flujo de genes (por ejemplo, del maíz al teocintle, Baltazar *et al.*,

2005; Ellstrand *et al.*, 2007; en la familia del betabel, Bartsch *et al.*, 1999; Viard *et al.*, 2004; Fénart *et al.*, 2007; Arnaud *et al.*, 2009; o las especies del género *Brassica*, Jørgensen y Andersen, 1994; Snow *et al.*, 1999).

El flujo genético puede tener un impacto dramático en la estructura genética de las poblaciones (Ehrlich y Raven 1969) ya sea histórico o reciente, con transgenes u otros elementos genéticos, puede limitar la diferenciación de la población por homogeneización de la reserva genética (Slatkin 1987).

Las estimación del flujo de genes se han basado históricamente en Nm (número de migrantes por generación), así como en F_{st} (índice de fijación, una medida de la diferenciación de la población), pero ambos parámetros se han desarrollado sobre modelos de poblaciones poco realistas (Whitlock y McCauley 1999; Paetkau *et al.*, 2004) suponiendo, por ejemplo, las poblaciones se encuentran en equilibrio (Broquet y Petit 2009). En contraste, el uso de redes de haplotipos y la covarianza genética, así como el análisis de redes de poblaciones (*Popgraph*), pueden proporcionar información sobre la relación histórica y espacial entre los genotipos (Dyer 2009). Por ejemplo, los patrones históricos de flujo de genes puede deducirse de redes que conectan los haplotipos a través de pasos mutacionales, correlacionando su origen con su distribución actual y contribuir a la diferenciación entre polimorfismos ancestrales y migración. Esta distinción es particularmente útil en el análisis de especies que han diversificado o divergido muy recientemente (Londo *et al.*, 2006). Por otro lado, *Popgraph* aplica herramientas generadas en la genética del paisaje (ver capítulo 1) que permiten la diferenciación entre fenómenos que subyacen la diferenciación genética entre las poblaciones (Dyer y Nason 2004), el aislamiento por distancia (cuando las poblaciones más cercanas genéticamente también lo son geográficamente y la diferenciación va creciendo al aumentar la distancia) y las migraciones a larga distancia (cuando la dispersión está limitada por barreras físicas o biológicas, no por la distancia misma, y por lo tanto, no hay una correlación entre la distancia genética y la geográfica). Estos enfoques incorporan información geográfica para evaluar la contribución del espacio físico en la estructuración de la diversidad genética (Dyer 2009; Manel *et al.*, 2003).

Si bien la estimación del flujo de genes es fundamental para analizar la estructura genética de las poblaciones, esto debe ser complementado con una evaluación de las dinámicas ecológicas que afectan a los patrones naturales de flujo

de genes y, por tanto, la evolución de una especie determinada (por ejemplo, Paetkau 1999; Palstra *et al.*, 2007).

Las metapoblaciones son ensamblajes de las poblaciones existentes en un equilibrio entre extinción, colonización y recolonización (Levins 1969; Hanski y Gaggiotti 2004 y sus referencias). El concepto de metapoblación en plantas es restrictivo a los siguientes criterios (Hanski 1998; Freckleton y Watkinson 2002): 1) los hábitats adecuados de la metapoblación son parches separados espacialmente; 2) todos los parches pueden extinguirse, pero no al mismo tiempo; y 3) la recolonización de cada parche después de la extinción local es posible (Honnay *et al.*, 2009). En este trabajo se ha documentado la estructura metapoblacional del algodón (*Gossypium hirsutum*) en México, lo que tiene amplias repercusiones en su manejo y conservación.

Dada la distribución de las poblaciones silvestres y cultivadas de *Gossypium hirsutum* en su centro de origen y diversidad (México), así como la reciente liberación de variedades de algodón transgénico, nos encontramos con un sistema modelo apropiado para abordar el papel de la dinámica metapoblacional sobre la estructura genética con flujo genético histórico y reciente.

El germoplasma de algodón cultivado actual se originó en Mesoamérica, donde fue semi-domesticado en tiempos prehispánicos (Valle de Tehuacán, México, fechado en torno a 5500-4300 AP, Smith y Stephens 1971). En varios estudios se utilizaron datos aloenzimas y RFLP para identificar posibles lugares de domesticación del algodón y para evaluar la diversidad genética (Wendel y Albert 1992; Brubaker y Wendel, 1994), pero necesitamos más información en la que se analice la variación en toda la distribución de algodón silvestre y conocer la contribución del flujo de genes sobre la estructura genética. Estas cuestiones son de especial importancia si tenemos en cuenta la importancia económica de las actuales variedades de algodón cultivadas, que son la fuente más importante de fibra natural y la tercera fuente de aceite vegetal en el mundo (FAOSTAT, 2009). Por otra parte, los estudios disponibles que abordan la fisiología y la adaptabilidad de algodón a las cambiantes condiciones ambientales se han llevado a cabo con un solo genotipo silvestre de algodón (*G. hirsutum* var. *Yucatanense*, Llamado TX2094), o utilizando el cultivar comercial obsoleto *Deltapine 14* (Delta and Pine Land Co; Applequist *et al.*, 2001). La resolución de la diversidad genética en las poblaciones silvestres de *G. hirsutum* también podría aumentar el éxito

de las estrategias de mejoramiento enfocadas en la generación de variedades adaptadas a diferentes ambientes.

El algodón es una especie que puede auto-polinizarse (McGregor 1976), por lo tanto, el flujo de genes se produce principalmente a través de la dispersión de semillas, que puede tener lugar en distancias largas, dado que sus semillas tienen la capacidad de resistir el agua salina (Stephens 1966) y también puede ser dispersadas por las aves, el agua dulce y el viento (observaciones personales). Sin embargo, la polinización cruzada puede ser más común de lo que se pensaba y depende de las condiciones ambientales (Stephens & Finkner 1953, Simpson 1954, McGregor 1976). Estos fenómenos deben tratarse con mayor atención, en particular, la dispersión de semillas a través del viento y de los medios antropogénicos, ya que pueden representar un medio importante para el del flujo de genes, sobre todo en los países que albergan poblaciones silvestres, variedades no GM y GM. En México, el algodón transgénico ha sido sembrado en un creciente número de hectáreas desde su aprobación (1996), y para 2009, sumaban ya 172.000 (SAGARPA, 2010). A pesar de la gran extensión del cultivo de algodón transgénico, la dispersión de transgenes no había sido evaluada hasta esta investigación.

Dada la compleja historia del género *Gossypium* y su capacidad para la migración a larga distancia, en este trabajo se estudia si las barreras geográficas han afectado el flujo de genes entre poblaciones silvestres de *G. hirsutum*, lo que resulta en una estructura genética que no se ajusta al modelo de aislamiento por distancia. Como un medio para estudiar la cuestión anterior, se evalúa la estructuración geográfica de las poblaciones de *G. hirsutum* en México, se presenta el área de distribución potencial y se propone una estructura de metapoblación. Con el fin de complementar el análisis anterior, documentamos el flujo de genes a partir de datos históricos con microsatélites de cloroplasto en una red de haplotipos y en otra de poblaciones. Por último, para evaluar si los últimos modelos de flujo de genes se ajustan a las inferencias históricas, utilizamos transgenes como marcadores recientes.

4.3 Discusión

En este estudio hemos demostrado que el flujo de genes presenta un modelo de migración a larga distancia entre las poblaciones silvestres *G. hirsutum*, tanto históricamente como recientemente. Los análisis sugieren que las barreras geográficas

pueden ayudar a mantener la estructura genética, pero no lo suficiente para impedir la migración entre las poblaciones. Por otra parte, esta estructura podría ser exacerbada por la dinámica ecológica metapoblacional que se propone.

4.4 Metapoblaciones de algodón: distribución actual, la dinámica ecológica y el cambio de uso del suelo

Basándonos en los estándares cualitativos y coherentes con los criterios de delimitación presentadas por otros estudios (Hanski 1998; Freckleton y Watkinson 2003; Honnay *et al.*, 2009), en este trabajo se propone la existencia de ocho metapoblaciones distintas de algodón silvestre en México: 1) Metapoblación Baja California Sur (Sur de Baja California Sur; MBC); 2) Metapoblación Pacífico Norte (Centro y Sur de Sinaloa y Norte de Nayarit; MPN); 3) Metapoblación Golfo Norte (Norte de Veracruz, Este de San Luis Potosí y Sur de Tamaulipas; MGN); 4) Metapoblación Bahía de Banderas (Suroeste de Nayarit y Noroeste de Jalisco; MBB); 5) Metapoblación Golfo Sur (Centro y Sureste de Veracruz; MGS); 6) Metapoblación Pacífico Sur (Sureste de Guerrero, línea costera de Oaxaca, Centro Oeste, Centro y Sur de Chiapas; MPS); 7) Metapoblación Pacífico Centro (línea costera del Centro y Sur de Jalisco, Colima, Michoacán, Noroeste y Centro de Guerrero; MPC) y 8) Metapoblación Península de Yucatán (Quintana Roo, Yucatán, Campeche, Noreste y Este de Tabasco; MPY). Esta dinámica podría verse favorecida por el hecho de que el 55% de las poblaciones viven en áreas perturbadas, lo que sugiere, de acuerdo a Fryxell y colaboradores (1979), en "evaluaciones previas de las poblaciones de algodón silvestre", que un proceso de alteración del hábitat debido a las perturbaciones humanas y otros procesos abióticos (cambios de uso del suelo, así como los huracanes y las tormentas tropicales) han ocurrido y dado forma al hábitat de la especie a lo largo de las costas del Pacífico y del Golfo de México. La supervivencia de estas poblaciones en áreas perturbadas probablemente está relacionado con la capacidad de la especie a crecer en sitios con baja cobertura vegetal y alta exposición solar, hábito perenne, madurez sexual durante el primer año de vida, poblaciones compuestas de plantas en diferentes etapas de vida, así como la dispersión a larga distancia de las semillas. Sin embargo, aunque las perturbaciones favorecen en algunos casos el crecimiento de las poblaciones, también podrían conducir a la extinción de las mismas sin posibilidad de re-colonización,

especialmente en escenarios de cambio extremo en el uso del suelo, como la pérdida del hábitat por la construcción de desarrollos turísticos que abarcan grandes extensiones de territorio. Este podría ser el caso de la región costera del Golfo de México (Metapoblaciones Golfo Norte y Golfo Sur (en el artículo GNM y GSM, por sus siglas en Inglés), que ha sido objeto de cambios en el uso del suelo debido a la agricultura, áreas de pastoreo y desarrollos urbanos. Lo que puede explicar las pocas poblaciones e individuos encontradas en esta región.

Mientras que las dinámicas ecológicas actualmente son afectadas por las actividades humanas, la estructura genética encontrada sólo se puede explicar en el tiempo evolutivo. El modelado del nicho ecológico de poblaciones de *G. hirsutum* en México se basó en datos reales obtenidos a partir de poblaciones silvestres de algodón, lo que confiere una entrada más precisa para los algoritmos de inferencia de distribución como GARP (ver capítulo 3), a diferencia de los estudios anteriores en los que se infiere la distribución de esta especie pero utilizando registros de algodón cultivado (Rogers *et al.* 2007).

4.5 Flujo de genes entre metapoblaciones y cultivares de algodón

Para evaluar el flujo de genes histórico, se utilizaron microsátélites del cloroplasto (alelos heredados vía materna) para detectar el flujo de genes a través de la migración histórica de las semillas. Nuestros datos sugieren migración a larga distancia y son consistentes con las observaciones anteriores (Stephens 1958, Wendel 1989, Wendel y Albert de 1992, Andersson y de Vicente 2010). Curiosamente, al evaluar el flujo reciente utilizando transgenes como marcadores en poblaciones silvestres *G. hirsutum*, nos encontramos con altas tasas de migración ($m = 66/270 = 0.24$), pero esto no parece ser debido a la migración de semillas, ya que sólo 15.9 % de las plantas que fueron positivas para la presencia de las proteínas recombinantes tiene el haplotipo del algodón domesticado (haplotipo 2). Esta observación podría implicar una primera migración a larga distancia de las semillas, su posterior crecimiento y floración, y necesariamente el entrecruzamiento o flujo secundario de estas plantas con las poblaciones silvestres. Posteriormente, los transgenes se insertados en el genoma nuclear y pueden dispersarse tanto a través del polen o semillas (explicando que el 84.1 % de las plantas positivas presente haplotipos exclusivamente silvestres).

Como el algodón fue domesticado hace siglos, el flujo de genes entre antiguos cultivares domesticados y sus parientes silvestres podría haber ocurrido históricamente, probablemente a través de la dispersión de semillas, favorecido por las actividades humanas y los fenómenos ambientales. Así, algunos de los patrones genéticos observados podrían ser el producto de tales eventos ancestrales que conducen el flujo de genes. No obstante, en el supuesto de que la estructura genética observada se ve afectada por los acontecimientos históricos del flujo de genes entre el algodón cultivado y silvestre, se repitió el análisis de haplotipos eliminando al haplotipo 2 (el único haplotipo en las muestras de algodón cultivado) y no se encontraron cambios significativos con respecto a la estructura reportada aquí (datos no mostrados).

La red de haplotipos que hemos presentado nos ha ayudado a distinguir polimorfismos ancestrales de los recientes acontecimientos de flujo génico. Además, todas las aproximaciones se han complementado con la estimación del flujo reciente de genes utilizando como marcadores a los transgenes en poblaciones existentes de algodón silvestre (Fig. 4 del artículo).

4.6 Los transgenes en metapoblaciones de algodón silvestre

Los rasgos que se han introducido mediante ingeniería genética a las variedades de algodón incluyen la resistencia lepidópteros (*CryIAb/Ac*, *Cry2Ac*, *CryIF* y *vip3A*), tolerancia a los herbicidas (CP4-EPSPS), y resistencia a los antibióticos (*PAT / Bar*, *nptII* y *aph4*). Usando estos rasgos, solos o en diferentes combinaciones, se han modificado variedades de algodón que han sido liberadas al medio ambiente desde 1996 (Traxler y Godoy 2004).

Quince años después de la introducción de variedades de algodón transgénico en México, se ha documentado la presencia de proteínas recombinantes en las poblaciones silvestres de algodón en su centro de origen y diversidad (ver figura 4a). Analizamos la presencia de las proteína recombinantes con los kits ELISA disponibles en México. Estos nos han permitido detectar 18 de 21 eventos aprobados (CERA, 2010) entre los individuos de las poblaciones silvestres de algodón. Los eventos restantes indetectables (3) han sido poco sembrados en el país. Es importante subrayar que las combinaciones de las proteínas recombinantes detectadas difieren entre metapoblaciones. Esto sugiere que las combinaciones han sido el resultado de

múltiples eventos de flujo de forma independiente. En el caso de las semillas positivas para transgenes que albergan el haplotipo presente en el algodón cultivado (2), podríamos estar detectando las plantas de algodón GM escapadas que se han dispersado en hábitats adecuados para la sobrevivencia pero que no permiten el desarrollo morfológico “normal” de planta cultivada. Este último caso podría ser frecuente en las metapoblaciones Pacífico Norte y Golfo Norte, porque el algodón GM se cultiva muy cerca o incluso dentro de su área (véase PNM y GNM por sus siglas en Inglés en la Figura 4 y Tabla 2).

En el artículo se analizan las proteínas recombinantes encontradas en las poblaciones y los eventos comerciales aprobados para su cultivo en México, así como las combinaciones encontradas que no son de venta comercial y sus posibles causas. Estas combinaciones transgénicas no se pueden explicar como principales eventos de flujo génico, ya que no están presentes con esa configuración en las líneas de algodón GM actualmente disponibles. Este es el caso de una semilla de la metapoblación Golfo Norte que expresa las cuatro proteínas recombinantes analizadas. Este hallazgo sugiere que la recurrencia de eventos de flujo de genes y apilamiento de genes podría haber ocurrido en esta metapoblación. En contraste, algunas semillas de las metapoblaciones Pacífico Norte y Pacífico Sur sólo expresan la proteína Cry2Ac, la cual no se comercializa de forma individual. Este fenómeno podría implicar la segregación independiente de los transgenes o, alternativamente, silenciamiento transcripcional o post-transcripcional de los genes con los que está disponible comercialmente (*Cry1Ab/Ac* y CP4-EPSPS). A fin de distinguir entre estas hipótesis, es necesario realizar análisis con ADN de las semillas y de las plantas madre.

Nuestros datos confirman que el flujo génico a larga distancia es posible en su centro de origen y diversidad. Sin embargo, el flujo de genes se favorece no sólo por factores biológicos como los mencionados anteriormente, sino también, por la intervención humana y la dinámica agrícola (Dyer *et al.*, 2009). Algunas otras prácticas también pudieran ocasionarlo, como la dispersión accidental de semillas de algodón destinadas a la alimentación animal (hemos observado su transporte en diversas partes de México y desde los EE.UU. hacia el centro-sur de México). Tampoco puede descartarse la dispersión de semillas transgénicas de sitios no autorizados para su cultivo, ya que no hay un marcador visual que permita distinguir la segregación de las semillas GM o la diferenciación de los productores de algodón. Estos fenómenos se producen porque las semillas que están separadas de su fibra se

venden como alimento para animales sin ser previamente trituradas en una "torta". Este escenario sucede porque se le presta poca atención a la semilla viable una vez separada de la fibra.

Dados los patrones documentados, los estudios futuros deben abordar los posibles escenarios que se pueden esperar en términos del flujo transgénico y la acumulación de transgenes en las poblaciones silvestres, así como las consecuencias que su presencia pudiera tener en el centro de origen y diversidad del algodón, como se ha documentado en el caso del maíz en México (Piñeyro-Nelson *et al.*, 2009; Dyer y Taylor, 2008).

El flujo de transgenes debe haber ocurrido principalmente a través de una migración primaria de las semillas y un evento secundario de polinización cruzada. Sólo en los casos en donde la distancia es muy corta se espera la polinización cruzada entre cultivares y plantas silvestres. En este estudio no encontramos una correlación entre la presencia del transgén y la pérdida de diversidad genética. Sin embargo, para explorar si la presencia de transgenes podría tener consecuencias en las poblaciones silvestres de algodón, se requieren estudios a largo plazo.

En general, este estudio demuestra que los análisis basados en ELISA son útiles para evaluar la presencia de transgenes en metapoblaciones de algodón silvestre. Sin embargo, los estudios futuros también deben considerar métodos basados detección de ADN para corroborar los resultados, así como la determinación de los acontecimientos específicos implicados. Este enfoque de múltiples técnicas ha sido sugerido en otros estudios que se ocupan de la detección de transgenes en centros de origen y diversidad (Serratos-Hernández *et al.*, 2007; Piñeyro-Nelson *et al.*, 2009).

La evidencia de investigaciones anteriores sugiere que *G. tomentosum* (en Hawai), *G. mustelium* (en Brasil) y *G. darwinii* (en Galápagos), están en peligro de extinción por el resultado de la hibridación con los algodones domesticados (Ellstrand 2003, Andersson & de Vicente 2010). En algunos casos, los híbridos interespecíficos (*G. hirsutum* x *G. barbadense*) pueden actuar como puentes genéticos para la transferencia de genes de algodón domesticado a otros parientes silvestres (*G. darwinii*; Ellstrand 2003, Andersson & de Vicente 2010).

4.7 Conclusiones

La interacción del flujo génico a larga distancia y las barreras geográficas de México ha formado la estructura genética de las poblaciones actuales de *G. hirsutum*, favorecida por la dinámica ecológica que obstaculiza la homogeneización genética dentro y entre poblaciones.

En lo que respecta al flujo de genes reciente, además de la importancia del estudio del flujo de transgenes y la acumulación en las poblaciones silvestres de algodón en su centro de origen y diversidad, se requieren esfuerzos de conservación y monitoreo permanente de las poblaciones silvestres. Estos esfuerzos se basan en la preservación del hábitat, tanto el ocupado actualmente por el algodón silvestre, como del disponible para su colonización, teniendo en cuenta una perspectiva de metapoblación (Meirmans *et al.*, 2003) donde la "persistencia de la metapoblación depende de la existencia de una cierta cantidad de hábitat adecuado pero actualmente desocupado" (Freckleton y Watkinson 2002, 2003). Se requieren además estudios demográficos cuantitativos de las poblaciones silvestres de algodón, documentar los patrones espacio-temporales de la dispersión de semillas y polen, así como las tasas de polinización cruzada entre individuos silvestres y cultivados en México. En términos generales, la conservación de zonas costeras puede ser útil para ayudar a la conservación de las poblaciones silvestres de algodón.

Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin

A. WEGIER,*† A. PIÑEYRO-NELSON,*‡ J. ALARCÓN,§ A. GÁLVEZ-MARISCAL,
¶ E. R. ÁLVAREZ-BUYLLA*‡ and D. PIÑERO*

*Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado postal 70-725, CP 04510, México DF, México,

†Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Progreso 5, Coyoacán, 04010, México DF, México,

‡Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Apartado postal 70-725, CP 04510, México DF, México, §Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Liga Periférico-Insurgentes Sur 4903, Parques del Pedregal, Tlalpan 14010, México DF, México, ¶Departamento de Alimentos y Biotecnología, Facultad de Química,

Universidad Nacional Autónoma de México, CP 04510, México DF, México

Abstract

Over 95% of the currently cultivated cotton was domesticated from *Gossypium hirsutum*, which originated and diversified in Mexico. Demographic and genetic studies of this species at its centre of origin and diversification are lacking, although they are critical for cotton conservation and breeding. We investigated the actual and potential distribution of wild cotton populations, as well as the contribution of historical and recent gene flow in shaping cotton genetic diversity and structure. We evaluated historical gene flow using chloroplast microsatellites and recent gene flow through the assessment of transgene presence in wild cotton populations, exploiting the fact that genetically modified cotton has been planted in the North of Mexico since 1996. Assessment of geographic structure through Bayesian spatial analysis, BAPS and Genetic Algorithm for Rule-set Production (GARP), suggests that *G. hirsutum* seems to conform to a metapopulation scheme, with eight distinct metapopulations. Despite evidence for long-distance gene flow, genetic variation among the metapopulations of *G. hirsutum* is high ($H_e = 0.894 \pm 0.01$). We identified 46 different haplotypes, 78% of which are unique to a particular metapopulation, in contrast to a single haplotype detected in cotton cultivars. Recent gene flow was also detected ($m = 66/270 = 0.24$), with four out of eight metapopulations having transgenes. We discuss the implications of the data presented here with respect to the conservation and future breeding of cotton populations and genetic diversity at its centre of crop origin.

Keywords: *Gossypium hirsutum*, long distance gene flow, metapopulations, Mexico, transgene flow

Received 1 October 2010; revision received 6 July 2011; accepted 15 July 2011

Introduction

The complexes of wild and cultivated varieties of crop plants at their centres of crop origin and/or diversity (hereafter, CCO) provide useful systems for addressing fundamental questions on population structure, genetics,

Correspondence: Ana Wegier, Fax: (+52 55) 36 26 87 00 Ext. 104
E-mail: awegier@gmail.com

Present address: CENID-COMEF, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Progreso 5, Coyoacán, 04010, México DF, México.

and specifically, gene flow dynamics (e.g. maize to teosinte; Baltazar *et al.* 2005; Ellstrand *et al.* 2007; the beet family; Bartsch *et al.* 1999; Viard *et al.* 2004; Féniart *et al.* 2007; Arnaud *et al.* 2009; or *Brassica* spp. Jørgensen & Andersen 1994; Snow *et al.* 1999). In cases where genetically modified varieties have been released at the CCO, transgenes become useful markers for addressing ongoing patterns, dynamics, and pervasiveness of gene flow (maize, van Heerwaarden *et al.* 2009; *Cucurbita*, Sasu *et al.* 2009; *Sorghum*, Sahoo *et al.* 2010). At the same time, these cases become particularly relevant for assessing the

general viability of GMO cultivation when there is potential for transgene flow into wild relatives at the CCO.

In spite of the effects of recent gene flow involving transgenes or other genetic elements, historical gene flow may still have a dramatic impact on population genetic structure (Ehrlich & Raven 1969). It may counteract the effects on effective population size of drift and inbreeding (Ebert *et al.* 2002), but may also constrain population differentiation by homogenizing the gene pool (Slatkin 1987). Gene flow estimation has historically relied on estimates of Nm (number of migrants per generation) and F_{st} (Fixation index; a measure of population differentiation). However, both of these parameters have been developed based on simplified and typically unrealistic population models (Whitlock & McCauley 1999; Paetkau *et al.* 2004) that assume, for example, that populations are at equilibrium (Broquet & Petit 2009).

In contrast, the use of haplotype networks and genetic covariance estimates, such as those used in Popgraph analyses, can provide information regarding the historical and spatial relationships among genotypes (Dyer 2009). For instance, historical gene flow patterns can be inferred from haplotype networks that connect each particular haplotype through mutational steps. This enables assignment of extant haplotypes to an ancestral population, while differentiating between ancestral polymorphisms and migration. This distinction is particularly useful when analysing species that have diversified or diverged quite recently, as is the case for the majority of cultivars (Londo *et al.* 2006). On the other hand, Popgraph draws from tools generated by landscape genetics that allow for the differentiation between isolation by distance and long distance migrations, which are phenomena that can underlie genetic differentiation among populations (Dyer & Nason 2004). These approaches explicitly incorporate geographical information to assess the contribution of physical space in structuring genetic diversity (Manel *et al.* 2003; Dyer 2009).

In the present study, we complement these types of historical gene flow analyses with estimates of ongoing gene flow using transgenes. While gene flow estimation is instrumental in the analysis of the genetic structure of populations, it should be complemented with a direct assessment of pollen and seed dispersal rates that impact on the natural patterns of gene flow. Otherwise, the consequences of dispersal-related life history variation among populations—and, hence, gene flow itself—will remain poorly quantified (e.g. Palstra *et al.* 2007). Therefore, we have also pursued the analysis of landscape features that can impact the genetic structure of populations by documenting the metapopulation structure of cotton in Mexico.

Metapopulations are assemblages of populations that exist in a balance between extinction and colonization (Levins 1969; Hanski & Gaggiotti 2004 and references therein). For plants, several criteria have been proposed that further constrain this metapopulation concept (Hanski 1998; Freckleton & Watkinson 2002), including: (i) that suitable metapopulation habitats are in spatially separated patches; (ii) that all patches can become extinct but they cannot do so at the same time; and (iii) that recolonization of each patch after local extinction is possible (Honnay *et al.* 2009).

The complex of wild and cultivated cotton populations in Mexico is an ideal system with which to address the role of metapopulation dynamics on recent and historical gene flow patterns, and on the genetic structure of populations. These studies are also instrumental for breeding and conservation programs for crops at their CCO. The germplasm of current cultivated cotton originated in Mesoamerica, where it was semi-domesticated in pre-Hispanic times (Tehuacán Valley, Mexico, dated around 5500–4300 BP; Smith & Stephens 1971). Previous studies used allozymes and RFLP data to identify possible venues of cotton domestication and to assess broad range genetic diversity (Wendel & Albert 1992; Brubaker & Wendel 1994). However, although cultivated cotton varieties are the most important source of natural fibre and the third source of oil in the world (FAOSTAT 2009), only two varieties (*G. hirsutum* var. *yucatanense*; called TX2094 and Deltapine 14; Delta and Pine Land Co; Applequist *et al.* 2001) have been used as reference for wild germplasm. Thus, broadening the genetic studies of wild populations of *G. hirsutum* will increase the success of breeding strategies focused on generating varieties adapted to extreme environments.

The *Gossypium* genus originated from African relatives between 12.5 (Seelanan *et al.* 1997) and 25 (Wendel & Albert 1992; Wendel *et al.* 2010) million years ago, and its salt-tolerant seeds enabled its spread around the world (Stephens 1966; Seelanan *et al.* 1997). Only four out of more than fifty *Gossypium* species have been domesticated (Wendel *et al.* 2009): two diploids in Asia and Africa (*G. herbaceum* and *G. arboreum*) and two tetraploids in America (*G. hirsutum* and *G. barbadense*). Current diploid and allopolyploid *Gossypium* species on the American continent cannot hybridize amongst themselves (Beasley 1940, 1942). Cotton is mainly self-pollinated, although cross-pollination may rarely occur (Stephens & Finkner 1953; Simpson 1954; McGregor 1976), and gene flow occurs via seed dispersal by water (Stephens 1966), and probably by wind and birds. In Mexico, GM cotton has been cultivated since 1996 and 172 000 ha were approved for sowing in 2009 (SAGARPA 2010). Despite the extent of GM cotton cultivation,

the dispersal of transgenes into non-GM and wild cotton has not yet been evaluated.

Given the complex history of the *Gossypium* genus and its capability for long distance migration, we first evaluated the geographical structuring of *G. hirsutum* populations in Mexico by generating a potential distribution based on climatic data. We hypothesized that geographic barriers have affected long distance gene flow among wild *G. hirsutum* populations, rendering a genetic structure that does not conform to an isolation-by-distance pattern across the area as a whole. We then documented historical gene flow using chloroplast microsatellite data to construct a haplotype network. Lastly, we used transgenes as markers to assess whether recent gene flow has taken place and if its patterns and dynamics conform to our historical inferences.

Materials and methods

Assessment of wild cotton populations and modelling of a potential distribution map

We selected populations of wild *Gossypium hirsutum* to be collected for this work by first performing an analysis of hundreds of historical specimens at the MEXU National Herbarium and XAL Herbarium. Twenty accessions were used that were clearly referenced as wild specimens and whose geographical reference fell within the formerly established natural habitats of this species. Concomitantly, we used the collections made by Paul A. Fryxell between 1968 and 1975 to guide our field search for wild populations. The specimens collected by Fryxell had clear features of wild cotton, as well as a precise description of both the habitat and location. Based on previous reports (Fryxell 1979; Wendel & Albert 1992; Applequist *et al.* 2001), we used the following objective criteria to classify a cotton plant as wild: (i) it is present in the expected habitat and distribution for the species' wild populations; (ii) it is a perennial shrub or tree, and (iii) its fruits have less than 22% lint content. We also delimited our unit of study, considering a population as comprised by a set of individuals that may potentially cross-pollinate among themselves and that are set at a distance of a maximum of 14 km among them. This distance criterion was set as a conservative limit, because this is the maximum pollinator (honeybee) movement range reported to date (Beekman & Ratnieks 2000).

We characterized the ecological niche for this species, based on 185 collection points of wild cotton plants surveyed between 2002 and 2007. We used the niche model proposed by Wiley (Wiley *et al.* 2003) and analysed our data using the GARP program (Genetic Algo-

rithm for Rule-set Production; Scachetti-Pereira 2001), which incorporated 23 bioclimatic covers from Worldclim, with a convergence limit of 0.01%, a 5% of omission, and a 10% commission threshold. Models were selected using the methodology proposed by Anderson *et al.* (2003).

The potential distribution map of *G. hirsutum* wild populations in Mexico was delimited through comparison with cartography from the Biogeographic Regions of CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad 1997). The predicted areas of distribution of wild cotton were validated through field inspections of places diagnosed to contain *G. hirsutum* wild populations according to the potential distribution maps, but where no former collections had been undertaken or where no entries were available at any database consulted. With these data, we analysed the population structure using a metapopulation scheme (see 'Results' section).

Cotton seed collection

Cotton seeds were collected between 2002 and 2008 in the identified wild populations of this species. The size of surveyed populations varied in number from 4 to 24 plants, with 1 up to <14 km separating individuals within a population. In total, 336 individual plants distributed in 36 populations were collected (Fig. 1).

Additionally, seeds from commercial cotton cultivars from Sonora, Mexicali, Chihuahua (Mexico), Texas, Virginia (USA), Argentina, Brazil, India and Egypt, were used to assess genetic diversity. We also collected seed from populations present outside the potential distribution area: Cuautla (18.89 N, -99.97 W), Tepoztlán (18.97 N, -99.09 W), Cuernavaca (18.87 N, -99.205 W; in the state of Morelos), Durango (23.18 N, -104.52 W) and Sonora (28.80 N, -110.57 W). All of these were considered feral populations because they have more than 25% of lint and are far away from potential distribution areas of wild cotton.

Laboratory procedures

DNA and chloroplast microsatellite analyses. Collected cotton seeds were germinated in growth chambers with a 12 h/30 °C light and 12 h/20 °C darkness regime, in 80–90% humidity. Genomic DNA was isolated from young seedling leaves using a modified CTAB method from Sul & Korban (1996; see Table S1).

DNA sequences were amplified through PCR using two specific primer sets for *G. hirsutum* chloroplast microsatellites (AF351292 (GAA)₉ and AF351313 (CA)₁₂, from Reddy *et al.* 2001). Additionally, ten PCR primer sets were used for the analysis of simple sequence

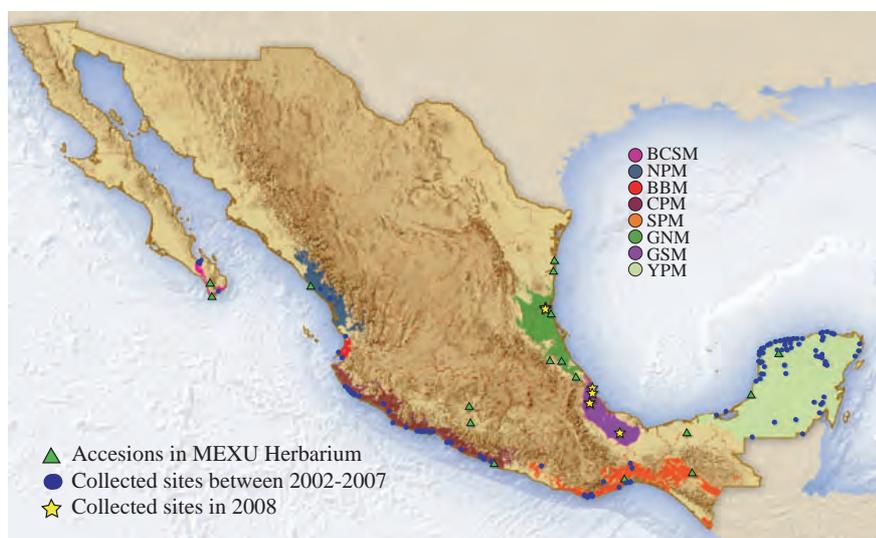


Fig. 1 Map showing collection sites, potential distribution area and metapopulations of *G. hirsutum* in Mexico. Symbols: green triangles: cotton collections identified by Fryxell at the MEXU herbarium; blue circles: 2002–2007 cotton collections; yellow stars: 2008 collections discovered with the use of the potential distribution map. Metapopulations are coded as follows: Baja California Sur (BCSM): fuchsia; North Pacific (NPM): grey; Banderas Bay (BBM): red; Central Pacific (CPM): burgundy; South Pacific (SPM): orange; Gulf North (GNM): dark green; Gulf South (GSM): purple; and Yucatán Peninsula (YPM): lime.

repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms (CCMP1–CCMP10 from Weising & Gardner 1999). The PCR procedure and individual conditions are shown in Table S1. DNA fragments were sequenced on an ABI Prism 3730xl Analyzer at the High-Throughput Sequencing and Genotyping Unit in the University of Illinois.

Immunoassays to detect the presence of transgenes in wild cotton populations. A total of 270 individual cotton seeds from 36 populations ($N \geq 20$ seeds per population; see Table 1) were individually analysed for transgene presence via immunoassays for the most common recombinant proteins present in cultivated cotton for which ELISA kits were available (Cry1Ab/Ac, Cry2A, CP4-EP-

Table 1 Presence of recombinant proteins in *G. hirsutum* metapopulations

Metapopulation	Populations	N	Positive seeds	Positive 1 protein	Positive 1 + proteins
BCS (<i>S of BCS</i>)	2	17	0	0	0
North Pacific (<i>Center and S of Sinaloa and N of Nayarit</i>)	3	37	25	19	6
Banderas Bay (<i>SW Nayarit and NW of Jalisco</i>)	2	15	0	0	0
Center Pacific (<i>Coastal line of C and S of Jalisco, Colima, Michoacán and NW and C of Guerrero</i>)	6	24	0	0	0
South Pacific (<i>SE of Guerrero, Coastal line of Oaxaca, CW, C and South tip of Chiapas</i>)	8	44	13	13	0
Yucatán Peninsula (<i>Quintana Roo, Yucatán, Campeche and NE and E of Tabasco</i>)	11	88	0	0	0
Gulf South (<i>C and SE of Veracruz</i>)	3	21	14	12	2
Gulf North (<i>N of Veracruz, E of San Luis Potosí and S of Tamaulipas</i>)	1	24	14	0	14
Total	36	270	66	44	22

The region comprised within each metapopulation is described in parentheses; the number of wild cotton populations in each metapopulation is presented in column two. Symbols: *N*: total number of seeds analysed per metapopulation; positive: total number of seeds positive for recombinant protein presence; positive 1 protein: number of seeds positive for only one recombinant protein; positive 1 + protein: number of seeds positive for more than 1 and up to 4 different recombinant proteins (see text for a complete description).

SPS and PAT/Bar). The embryo of each seed was separated from its seed coat and divided into four pieces with a surgical knife. Each piece was placed in a 2-mL microcentrifuge tube for separate homogenization with an appropriate volume of extraction buffer. Each sample was analysed using duplicate assays in each ELISA plate. Immunoassays were conducted according to the manufacturer's instructions. The ELISA plates were read in a spectrophotometer at 450 nm for proteins PAT/bar, Cry2A and CP4-EPSPS-event NK603 (Envirologix™ plates) and at 650 nm for proteins CP4-EPSPS and Cry1Ab/Ac (Agdia® plates).

We considered a sample to be positive only when its absorbance was equal to or above a reading three standard deviations above the average intensity of all negative controls and blank samples. At least one duplicate of a blank (extraction buffer), one negative control, and one positive control were included in each ELISA plate.

Data analyses

Molecular diversity. We determined the number and frequency of all unique chloroplast DNA haplotypes and estimated molecular diversity using Arlequin v3.5 (Excoffier & Lischer 2010). We used the rarefaction approach (using ADZE; Szpiech *et al.* 2008) to see if heterogeneous population sizes could affect the estimation of genetic diversity among populations and also to generate estimates that would be comparable among different populations (Petit *et al.* 1998; Kalinowsky 2004).

Genetic structure and gene flow analyses. We examined population structure by performing a Bayesian spatial analysis using the program BAPS 5.1 (Corander *et al.* 2008), which uses stochastic optimization to find the optimal partition. Simulations were run from $K = 2$ to $K = 10$ with 100 replicates for each K .

We sought evidence for isolation by distance and/or long-distance dispersal events using Population Graph (GeneticStudio software; Dyer 2009). This is a graph-theoretic approach that analyses how genetic variation is distributed across the investigated landscape, by plotting migration and enabling the assessment of the dependence or independence of evolutionary trajectories among populations. Within a graph, populations are represented as nodes and the genetic covariation among populations determines the topology. The pattern of connections between populations is estimated conditional on the entire data set. The pattern can be used to test for isolation-by-graph-distance, where in an extreme case, if covariance between two populations equals zero, no connection is drawn (IBGD; Dyer & Nason 2004). Plotting the Population Graph onto a

map also allows the inferred population pairs to have 'extended edges', 'normal edges', or 'compressed edges', which imply that genetic distance is either higher, equal to, or lower, respectively, than the one expected by geographic data (Dyer 2009).

We investigated the evolutionary history and relationships among the haplotypes found in this study and differentiation of the ancestral polymorphism and gene flow by constructing a minimum-spanning network of haplotypes using tcs 1.21 (Clement *et al.* 2000). We used the methods described by Templeton & Sing (1993) to break loops (ambiguous connections) within our network, while using predictions derived from coalescence theory (reviewed in Rosenberg & Nordborg 2002).

Distances between GM cotton release sites and wild cotton populations. Mexico was divided into over 80 000 hexagons, 25 km² each, to compare areas against experimental release centres. Centroids of these hexagons were used to calculate the distance between the release sites and the potential distribution model, with an error of 25 km. The sites where permits to release genetically modified cotton in Mexico have been granted (from 1996 to 2008) were plotted on a map of Mexico, under the assumption that all plots approved were actually planted (Fig. 4a). The minimum distance separating a granted GM cotton release site from all populations of wild cotton was determined (Table 2).

Results

Wild populations of G. hirsutum in Mexico: potential distribution and actual metapopulation structure

A potential distribution map was generated using computational and geographic tools (GARP). This map was based on a comprehensive survey of existing wild *G. hirsutum* populations comprising 185 collection points (recorded between 2002 and 2007) distributed in 36 populations (blue points in Fig. 1). The potential distribution is plotted in Fig. 1 and represents those areas that had over 75% of confidence of translating into the actual wild cotton distribution, according to our survey data. Thus, the actual distribution area for this species may possibly be even broader than that considered here. Nevertheless, the fact that all predicted populations were either corroborated or led to the finding of new populations helped us to validate the precision of the ecological niche prediction model used here. The potential distribution map identified seven new populations along the Gulf of Mexico in 2008 (yellow stars in Fig. 1). In previous years, without the guidance of this model, efforts to find wild populations in this area had proved unsuccessful.

During the 7-year fieldwork period (2002–2008), we observed that 85% of wild cotton populations were in coastal ecosystems and some were in low dry forests. Cotton plants were in populations of 4 to 20 individuals (5 was the mode).

The spatial distribution and the ecological setting of the populations investigated here suggest the existence of eight discrete bioclimatic areas. These are separated by intermediate zones that lack adequate climatic and ecological conditions for *G. hirsutum* to grow, and that effectively form geographical barriers to seed and pollen flow. Each discrete area described here is considered to be a metapopulation because cotton plant populations were discontinuous due to the discrete occurrence of favourable habitats. Furthermore, each metapopulation was separated from one another by at least 150 km or was isolated by evident geographical barriers.

The eight metapopulations proposed here are: Baja California Sur (BCSM), North Pacific (NPM), Banderas Bay (BBM), Center Pacific (CPM), South Pacific (SPM), Yucatán Peninsula (YPM), Gulf South (GSM) and Gulf North (GNM; Fig. 1). Although we lack quantitative dynamic data for all of the populations surveyed, the number of plants per population, as well as the number of populations that form a metapopulation, varied substantially (Table 1). In the northern part of the country, the maximum number of populations per metapopulation is three (BCSM, NPM, BBM and GSM). In the south of Mexico, three metapopulations

have six, eight, and eleven populations (CPM, SPM and YPM, respectively). With regard to suitable habitats for cotton growth within metapopulations, the YPM has the largest contiguous range of suitable habitats and bears the largest populations. It is also the most genetically diverse.

Genetic variation, historical gene flow, and population structure of *G. hirsutum* in Mexico

Overall, genetic variation among wild metapopulations of *G. hirsutum* is high ($H_e = 0.894 \pm 0.01$). We found a total of 46 haplotypes, 78% of which are unique to a particular metapopulation (Fig. 2). The highest haplotype diversity was found in BBM (0.94) and YPM (0.93; for haplotype diversity between metapopulations, see Table S2). The remaining metapopulation diversity ranges between values of 0.6 and 0.8, except for GSM, which is exceptionally low (0.34). In contrast, the analysed commercial cotton seeds and inferred feral populations have only one haplotype (number 2; Fig. 2a). The only exception to this trend is the feral population in Cuernavaca, Morelos, where two haplotypes were found (2 and 25; Fig. 2a).

The 46 haplotypes found in this study group into six distinct lineages, of which those of BBM and YPM are well differentiated (Fig. 2a). This haplotype network has a complex topology, where some populations with unique lineages have haplotypes that are not sampled,

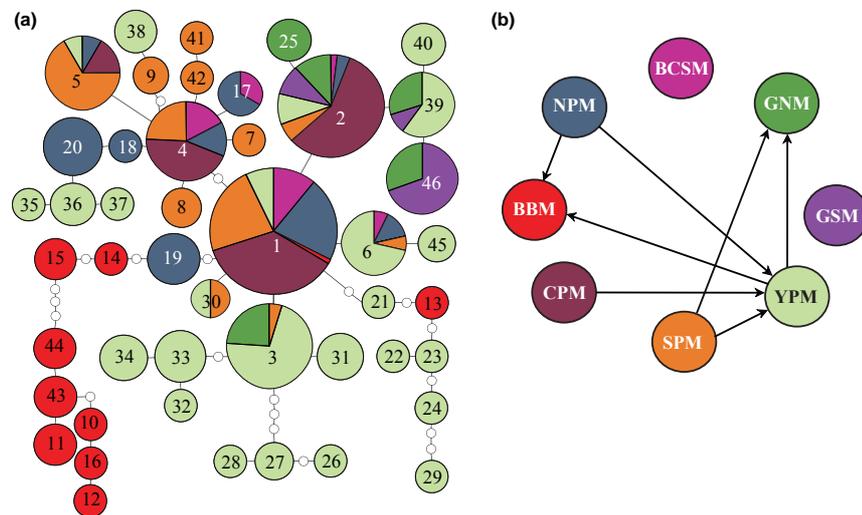


Fig. 2 Haplotype network and historical gene flow in wild *G. hirsutum* metapopulations. (a) Haplotype network. Haplotypes documented in this work are depicted in circles; sizes of nodes show the frequency of a particular haplotype while colours represent the presence of a particular haplotype within each metapopulation. (b) Historical gene flow patterns among metapopulations, as inferred from the haplotype network (metapopulation colour-codes and labels are as in Fig. 1).

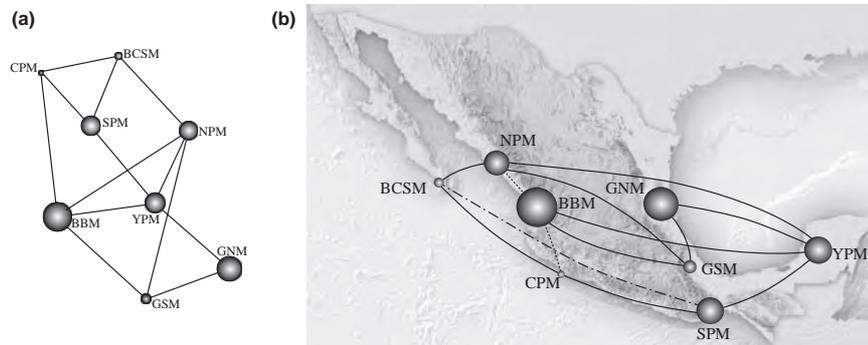


Fig. 3 *Popgraph* analysis showing significant connections among wild cotton metapopulations. (a) Three-dimensional *Popgraph* representing the genetic covariance among metapopulations of *G. hirsutum*, based on chloroplast markers. The length of a line between any two metapopulations is proportional to their covariance; within-metapopulation genetic variance is proportional to sphere size. (b) Geographic distances among metapopulations and their relation with the *Popgraph* analysis shown in (a). Edges that are significantly longer (---), shorter (· · ·) or congruent (—) with the predicted genetic covariance with respect to geographic distance are plotted. Names are as in Fig. 1.

while shared haplotypes among several populations appear to be ancestral. Derived haplotypes are generally unique. The unique haplotype of cultivated cotton (2) is shared and frequent in almost all metapopulations. Haplotypes 35, 36, and 37 show ancient gene flow, while 38 and 5 seem to have recently migrated from the YPM; lastly, haplotype 13 shows evidence of ancestral gene flow (Fig. 2b).

We estimated the genetic structure among the surveyed wild cotton populations by modelling our data using the BAPS approach. We found that the optimal number of clustered groups was eight; thus, the description of each cluster by this algorithm is consistent with the metapopulation scheme derived from the potential distribution analysis used above. Furthermore, the rarefaction approach was consistent with the Population Graph tool, as the diameter representing the genetic variation is not correlated with the sample size per population (Fig. 3).

Given the inferences of genetic variation and metapopulation genetic diversity presented above, which are the result of historical events, we addressed the question of whether contemporary long distance gene flow has taken place, by evaluating transgene flow.

Recent long distance gene flow: presence of transgenes in wild cotton metapopulations

The potential for long and short-distance ongoing transgene flow that could be occurring from GM cotton plants to native wild cotton populations was evaluated through plotting the frequency distribution of the distance between the GM cotton parcels and the nearest wild cotton population (Table 2). In this analysis, we

found that 1.4% of 5985 permits to sow GM cotton issued by the pertinent Mexican authority between 1996 and the beginning of 2008, fall within the area of distribution of two metapopulations of wild cotton (NPM and GNM), while 4.2% are within a 300-km radius from three metapopulations (NPM, GNM and GSM). The remaining 94.4% of GM field releases approved are over 300 km apart from all wild cotton metapopulations (Table 2).

We identified actual transgene flow by assessing the presence of recombinant proteins in wild cotton populations through ELISA tests. The immunoassays yielded 66 positive seeds out of 270 seeds tested (24.4%) for at least one recombinant protein (Table 1). These positive cases were distributed among four metapopulations (Fig. 4): NPM (25/37; 67.6%), GNM (14/24; 58.3%), GSM (14/21; 66.7%) and surprisingly, SPM (13/44; 29.5%). The latter is at a lineal distance of 755 km from the southernmost and nearest approved GM cotton plot. Furthermore, 3 out of 3 populations comprising the NPM had positive testing plants for transgene presence; 1 out of 1 in GNM; 2 out of 3 in GSM and 3 out of 7 in the SPM. Interestingly, two-thirds of all positive samples yielded positive results for a single recombinant protein, while one-third did so for two and up to four different transgene-codified proteins (Table 1).

Of the 66 positive seeds, 15.9% had the haplotype common to the domesticated cultivars (haplotype 2). In the GNM, 6 out of 14 positive seeds for Cry1Ab/Ac have haplotype 2. In the GSM, three out of five individuals positive for CP4-EPSPS had this haplotype. In the NPM, two seeds positive for Cry1Ab/1Ac and Cry2A shared this haplotype. In the SPM, none of the positive seeds for recombinant proteins had this haplotype.

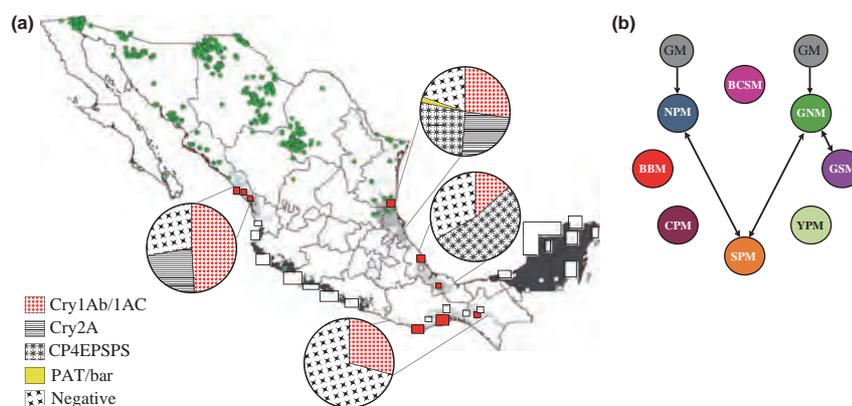


Fig. 4 Contemporary gene flow among cotton metapopulations as inferred by transgene presence. (a) Map of Mexico showing the regions where GM cotton cultivars have been approved for planting, as well as wild cotton metapopulations and populations positive for recombinant protein presence. GM cotton cultivation sites are plotted as green circles; metapopulations without recombinant proteins (BCSM, BBM, CPM, and YPM) are coloured in dark grey; metapopulations with recombinant proteins (NPM, SPM, GNM and GSM) are in pale grey; wild cotton populations with transgene presence are plotted as red squares while populations without transgenes are depicted as white squares. Pie charts with the frequency of particular recombinant proteins are set aside each transgene-harboursing metapopulation. (b) Diagram showing possible venues of present gene flow between GM cultivars and some wild cotton metapopulations. Arrows show the probable trajectories of transgene flow.

Discussion

In this study, we have shown that long distance gene flow has taken place among *G. hirsutum* wild populations, both historically and recently. Evidence from the population genetic analyses and the metapopulation scheme suggests that geographical barriers can hinder population structuring, but not sufficiently to suppress migration among metapopulations.

Cotton metapopulations: current distribution, metapopulation dynamics, and changes in land use

In this work, we propose the existence of eight distinct wild cotton metapopulations in Mexico. While the standards utilized in this work to define metapopulations are qualitative, they are consistent with the delimitation criteria put forward by other scholars (Hanski 1998; Freckleton & Watkinson 2003; Honnay *et al.* 2009). In line with metapopulation theory, we found that 34% of the historically characterized *G. hirsutum* populations continue to dwell in their original geographic zones. Local extinction and recolonization was also observed in 68 of 171 collection points surveyed for which at least two visits were performed during this work.

This dynamic turnover could be favoured by the fact that 55% of surveyed wild populations live in disturbed areas. This suggests, based on Fryxell's and others' previous assessments of wild cotton populations, that a

process of habitat alteration due to human and abiotic perturbations (changes in land use, as well as hurricanes and tropical storms) has taken place. These phenomena have shaped the species' habitat along the Pacific and Gulf of Mexico coastal lines. The survival of these populations in disturbed areas is probably related to the ability of this species to grow well in places with low plant cover and high solar exposure, as well as having a perennial habit, being sexually mature during the first year of life, having populations composed of plants at different life stages, and presenting long distance seed dispersal. Nevertheless, while habitat perturbations have not affected all cotton populations, they could drive a significant number of them to extinction, especially in a scenario where extreme changes in land use would hinder recolonization. This could be the case for the coastal region of the Gulf of Mexico (GNM and GSM), which has been subjected to land use changes due to promotion of agriculture and cattle grazing areas (GNM) or to the establishment of hotel resorts that deplete coastal dunes (GSM). This probably accounts for the smaller number of wild cotton populations documented in this study for that part of the country.

While the dynamics currently affecting metapopulation structuring are probably significantly influenced by human activity, the structure unveiled in this work can only be explained in evolutionary time. The modelling of the ecological niche for *G. hirsutum* populations in Mexico was also based on data obtained from actual wild cotton populations. This confers a more precise

Table 2 Linear distances between GM cotton release-sites and wild cotton metapopulations

Cotton growing region	Minimum distance between GM cotton plot and a wild cotton metapopulation (km)	Number of granted permits (1996–2008)
Tamaulipas and Sinaloa	0	85
Tamaulipas, Sinaloa and South Sonora	1–100	152
	101–200	42
	201–300	56
Comarca Lagunera	301–400	919
	401–500	1200
South Chihuahua	501–600	378
	601–700	274
North Chihuahua	701–800	1375
North Sonora	801–900	210
	901–1000	210
Mexicali, Baja California Norte	1001–1100	1084

input for distribution inference algorithms such as GARP, than was obtained by previous studies where this distribution was inferred using data from cultivated cotton (Rogers *et al.* 2007).

Ancestral gene flow among cotton metapopulations and cotton cultivars

We evaluated historical gene flow using chloroplast microsatellites (maternally inherited alleles) to detect historical gene flow through seed migration. Our data indeed suggest long distance seed migration that is consistent with previous suggestions of the potential for seed dispersal through marine currents, given the viability of seeds subjected to prolonged incubation periods in salt water (Stephens 1958). This finding is consistent with the signature of molecular markers during chromosomal speciation (Wendel 1989; Wendel & Albert 1992; Andersson & de Vicente 2010). Interestingly, when assessing recent long distance gene flow through transgene presence in wild *G. hirsutum* populations, we find high migration rates ($m = 66/270 = 0.24$), but this does not seem to be due to seed migration, since only 15.9% of the plants that were positive for transgene presence have the haplotype common to domesticated cotton used to generate transgenic lines (haplotype 2). This observation could imply low seed migration out of GM fields. Nevertheless, once a single or a few transgenic individuals are dispersed into particular wild populations, they produce pollen that may fertilize local wild plants. Since transgenes are inserted within the nuclear genome, they can be dispersed both via pollen or seed.

As cotton was domesticated centuries ago, ancient gene flow between domesticated cultivars and its wild relatives could probably have occurred historically via seed dispersal, favoured by human activities and environmental phenomena. Thus, some of the genetic patterns observed could be the product of these types of ancestral events. Nevertheless, we assumed that the observed genetic structure is affected by historic gene flow events among cultivated and wild cotton and we repeated the haplotype analysis, this time eliminating haplotype 2 (the only haplotype in cultivated specimens). We did not find significant changes with respect to the structure reported here (data not shown).

The haplotype network that we have put forward has helped us to distinguish ancient polymorphisms from recent gene flow events. Furthermore, these approximations have been complemented by the estimation of recent gene flow using transgenes as markers in extant wild cotton populations (Fig. 4).

Transgenes in wild cotton metapopulations

Fifteen years after the introduction of GM cotton cultivars into Mexico, we have documented the presence of recombinant proteins in wild cotton populations at its CCO (see Fig. 4a). We assayed recombinant protein activity using ELISA kits available in Mexico. These enabled us to detect 18 out of 21 approved events (CERA 2010) among individuals of wild cotton populations. The remaining undetectable events (3) have been scarcely sown. The traits that have been introduced, alone or in different combinations, into currently sown cotton varieties through genetic engineering include Lepidoptera resistance (*Cry1Ab/Ac*, *Cry2Ac*, *Cry1F* and *vip3A*), herbicide tolerance (CP4-EPSPS), and antibiotic resistance (*PAT/Bar*, *nptII* and *aph4*; Traxler & Godoy-Avila 2004).

The combinations of recombinant proteins detected in this study differ among metapopulations, which suggests that each combination could have been the result of independent and multiple transgene flow events into the Mexican wild cotton populations. This observation is additionally supported by the fact that 84.1% of seeds that tested positive for transgene expression had a haplotype other than the one present in the cultivars (2). Since cotton is assumed to be self pollinated, transgene flow must also have occurred mostly via seed and secondary cross-pollination events (Dyer *et al.* 2009).

The combinations of transgenes found within metapopulations and the possible transgenic events from which they could have originated are as follows: in PNM, plants expressing *Cry1Ab/Ac*, could have originated from event MON531; in GSM, CP4-EPSPS protein could involve either MON88913 or MON1445/1698; for

the GNM and GSM metapopulations, the recombinant protein combinations found -Cry1Ab/Ac and CP4-EPSPS- suggest that the most likely transgenic event could be MON531 × MON1445. For GNM, the Cry1Ab/Ac, Cry2Ac, and CP4-EPSPS proteins could originate from MON15985 × MON1445 or MON88913. These events were approved for planting in Mexico between 1996 and 2003. In the case of seeds positive for transgenes that harbour haplotype 2 and have transgene combinations consistent with a commercial transgenic variety (15.9%), we could be detecting feral GM cotton plants that have dispersed into suitable habitats, but, given the environmental conditions, do not grow to resemble their cultivated counterparts.

In contrast, we found some transgene combinations that cannot be explained as primary gene flow events, given the transgene combinations present in the currently available GM cotton lines. This is the case of a seed from GNM that expresses all four recombinant proteins assayed. This finding suggests that recurrent gene flow events and gene stacking could already have occurred in this metapopulation. In contrast, some seeds from NPM and SPM only expressed the Cry2Ac protein, which is not contained individually in any commercial event. This phenomenon could involve independent segregation of transgenes from some lines and a later introgression into wild cotton. Alternatively, it could represent transcriptional or post-transcriptional gene silencing of either the CP4-EPSPS or Cry1Ab gene/transcripts that are present in all commercially available lines expressing Cry2Ac. In order to distinguish between these hypotheses, DNA-based analyses should be undertaken using transgene specific primers, both from the DNA of the seed and that of the mother plant.

Detected transgenes were aggregated in space ($p = 0.001$). This type of distribution could be favoured by the dynamics of plant metapopulations. In the particular case of wild cotton, populations where recolonization has taken place have few plants, and thus can be subject to a genetic bottleneck and to genetic drift. Transgene frequencies and spatial patterns documented here also suggest that transgene introduction is relatively recent and has not been fixed in all metapopulations. These findings could imply that these new alleles do not confer a high selective advantage. In the particular metapopulations where not all populations are positive for transgene presence (GSM and PSM), but are in close proximity to some of the positive populations, three hypotheses would explain the intermixing of positive and negative populations. Firstly, negative populations are more recent than positive ones and are the product of colonization/recolonization from wild (non-transgenic) seed. Alternatively, these populations did have transgenes, but the transgenes were eliminated by

genetic drift or selection. Finally, transgenes in these populations may exist but are silenced or have not reached some populations simply due to random events.

Our transgene data confirm that long-distance gene flow is preeminent in wild cotton at its CCO. Given present day management practices, some means of seed movement at long distances include the accidental dispersal of cotton seed intended for animal feed. We observed this happening in trucks from the USA to the centre-south of Mexico. This phenomenon takes place because seeds that are separated from their fibre are later sold as animal feed without being previously mashed into a 'cake'. This venue for GM seed dispersal could very well be occurring for GM seed processed in Mexico, because very little attention is paid to the disposal of this seed once the fibre has been removed. Given the documented patterns, future studies should address the possible scenarios to be expected in terms of transgene flow and accumulation, as well as the consequences these may have for wild cotton conservation at its CCO, as has been documented for maize in Mexico (Dyer & Taylor 2008; Piñeyro-Nelson *et al.* 2009).

In this study, we found no correlation between transgene presence and loss of genetic diversity. Nevertheless, in order to explore whether the presence of transgenes could have consequences in wild cotton populations, sustained and long-term analyses should be pursued. In particular, biomonitoring studies that assess the consequences of both transgene and foreign germplasm introduction into wild metapopulations of *G. hirsutum*, should be undertaken (see, for example, Meirmans *et al.* 2009).

This study confirms that ELISA-based analyses are useful when assessing the presence of transgenes in wild cotton metapopulations. Nevertheless, future studies should also consider DNA-based detection methods to corroborate our findings, as well as to determine the specific events involved. This multiple-technique approach has been suggested in other studies dealing with transgene detection at CCO (Serratos-Hernández *et al.* 2007; Piñeyro-Nelson *et al.* 2009).

Lastly, gene flow from cultivated cotton can put the wild germplasm of several *Gossypium* species at risk. Evidence from previous investigations suggests that *G. tomentosum* (in Hawaii), *G. mustelium* (in Brazil) and *G. darwinii* (in Galapagos) are at risk of extinction as a result of hybridization with domesticated tetraploid cotton (Ellstrand 2003; Andersson & de Vicente 2010). In some cases, interspecific hybrids (*G. hirsutum* × *G. barbadense*) may act as genetic bridges for gene transfer from domesticated cotton to other wild relatives (*G. darwinii*; Ellstrand 2003; Andersson & de Vicente 2010). As a con-

sequence, conservation programs should include all *Gossypium* tetraploid species.

Conclusions

The interplay of historical long distance gene flow and geographic barriers in Mexico has shaped the genetic structure of extant populations of *G. hirsutum*. Extinction and recolonization events in particular populations have hindered genetic homogenization among metapopulations.

Potential distribution analyses and molecular markers independently show the existence of eight metapopulations. We were able to record intense dynamics of recent local extinctions and colonizations that go back to the collections made by Paul Fryxell. In spite of their integrity, these metapopulations are connected through long distance migration events. In particular, through the assessment of transgene presence, we were able to detect recent gene flow, which supports the connectivity of these metapopulations. This scenario of long distance colonization, the existence of metapopulations, and the presence of transgenes at its CCO calls for conservation efforts both *in situ* and *ex situ*. These types of endeavours rely upon preservation of the habitat currently occupied by wild cotton plants or on opening up of new habitats for wild cotton colonization. The metapopulation perspective must be kept in mind (Meirmans *et al.* 2003), as 'metapopulation persistence relies on the existence of a certain amount of suitable but currently unoccupied habitat' (Freckleton & Watkinson 2002, 2003). Coastal dunes appear to be particularly important areas in this respect. In addition, demographic studies of wild cotton populations, documentation of spatio-temporal patterns of seed and pollen dispersal, and rates of cross-pollination among wild individuals should also be the basis for guiding these conservation efforts.

Acknowledgements

The authors thank Victoria Sork, René Cerritos, José Casar, Lluvia Flores, Valeria Alavez and three anonymous reviewers for useful comments on earlier versions of this manuscript; Rafael Lira, Alejandro Casas, Judith Márquez-Guzmán, Luis Felipe Jiménez, and Francisca Acevedo gave useful advice for this work. AW and APN would like to thank Alejandra Vázquez-Lobo, Rodolfo Salas, Argelia Cuenca, Lev Jardón, Diego Ortega, Adriana Caballero and Luz María Martínez for their help in field-work and in various laboratory tasks. PhD scholarships from Consejo Nacional de Ciencia y Tecnología and Universidad Nacional Autónoma de México were granted to AW and APN. This work is part of the PhD dissertation of AW in Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México. Research was done with support from CONABIO, through grants

GEF-CIBIOGEM 00013572, PNUD-CIBIOGEM 00051868 to AW and DP. ERAB and APN research was financed by grants UC-Mexus ECO-IE359 and 'Red Temática de Investigación CONACYT: Complejidad, Ciencia y Sociedad' (124909).

References

- Anderson RP, Lew D, Peterson AT (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, **162**, 211–232.
- Andersson MS, de Vicente CM (2010) *Gene Flow Between Crops and Their Wild Relatives*, 564 pp. Johns Hopkins University, Baltimore.
- Appelquist WL, Cronn R, Wendel JF (2001) Comparative development of fiber in wild and cultivated cotton. *Evolution & Development*, **3**, 3–17.
- Arnaud J-F, Fénart S, Godé C, Deledicque S, Touzet P, Cuguen J (2009) Fine-scale geographical structure of genetic diversity in inland wild beet populations. *Molecular Ecology*, **18**, 3201–3215.
- Baltazar BM, Sanchez-Gonzalez JD, Cruz-Larios L, Schoper JB (2005) Pollination between maize and teosinte: an important determinant of gene flow in Mexico. *Theoretical and Applied Genetics*, **110**, 519–526.
- Bartsch D, Lehen M, Clegg J, Pohl-Orf M, Schuphan I, Ellstrand NC (1999) Impact of gene flow from cultivated beet on genetic diversity of wild sea beet populations. *Molecular Ecology*, **8**, 1733–1741.
- Beasley JO (1940) Hybridization of American 26 chromosome and Asiatic 13 chromosome species of *Gossypium*. *Journal of Agricultural Research*, **69**, 175–181.
- Beasley JO (1942) Meiotic chromosome behavior in species hybrids, haploids, and induced polyploids of *Gossypium*. *Genetics*, **27**, 25–54.
- Beekman M, Ratnieks LF (2000) Long-range foraging by the honey-bee, *Apis mellifera* L. *Functional Ecology*, **14**, 490–496.
- Broquet T, Petit EJ (2009) Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 193–216.
- Brubaker CL, Wendel JF (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *American Journal of Botany*, **81**, 1309–1326.
- CERA (2010) *GM Crop Database*. Center for Environmental Risk Assessment (CERA), ILSI Research Foundation, Washington DC. http://cera-gmc.org/index.php?action=gm_crop_database (Accessed on February 1, 2011).
- Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.
- Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (2003) *México: Imagen desde el espacio*. CONABIO, México, DF.
- Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9**, 539.
- Dyer RJ (2009) GeneticStudio: a suite of programs for the spatial analysis of genetic marker data. *Molecular Ecology Resources*, **9**, 110–113.

- Dyer RJ, Nason JD (2004) Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1727.
- Dyer JA, Taylor JE (2008) A crop population perspective on maize seed systems in Mexico. *Proceedings of the National Academy of Sciences, USA*, **105**, 470–475.
- Dyer G, Serratos-Hernández A, Perales H *et al.* (2009) Dispersal of transgenes through maize seed systems in Mexico. *PLoS ONE*, **4**(5), e5734.
- Ebert D, Haag D, Kirkpatrick M, Riek M, Hottinger JW, Pajunen VI (2002) A selective advantage to immigrant genes in a *Daphnia* metapopulation. *Science*, **295**, 485–487.
- Ehrlich R, Raven PH (1969) Differentiation of populations. *Science*, **165**, 1228–1232.
- Ellstrand NC (2003) *Dangerous liaisons? When Cultivated Plants Mate With Their Wild Relatives*, 244 pp. Johns Hopkins University, Baltimore.
- Ellstrand NC, Garner LC, Hegde S, Guardagnuolo R, Blancas L (2007) Spontaneous hybridization between maize and teosinte. *Journal of Heredity*, **98**, 183–187.
- Excoffier L, Lischer HEL (2010) Arlequin suite v3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- FAOSTAT (2009) *Statistical Databases*. Food and Agriculture Organization of the United Nations, Rome, Italy. <http://faostat.fao.org> (accessed on February 1, 2011).
- Fénart S, Austerlitz F, Cuguen J, Arnaud J-F (2007) Long distance pollen-mediated gene flow at a landscape level: the weed beet as a case study. *Molecular Ecology*, **16**, 3801–3813.
- Freckleton RP, Watkinson A (2002) Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. *Ecology*, **90**, 419–434.
- Freckleton RP, Watkinson AR (2003) Are all plant populations metapopulations? *Journal of Ecology*, **91**, 321–324.
- Fryxell PA (1979) *The Natural History of the Cotton Tribe*, 245 pp. Texas A & M University Press, College Station/London.
- Hanski I (1998) Metapopulation dynamics. *Nature*, **396**, 41–49.
- Hanski I, Gaggiotti O (2004) *Ecology, Genetics and Evolution of Metapopulations*. Elsevier Academic Press, London, UK.
- van Heerwaarden J, Van Eeuwijk FA, Ross-Ibarra J (2009) Genetic diversity in a crop metapopulation. *Heredity*, **104**, 28–39.
- Honnay O, Jacquemyn H, Van Looy K, Vandepitte K, Breyne P (2009) Temporal and spatial genetic variation in a metapopulation of the annual *Erysimum cheiranthoides* on stony river banks. *Journal of Ecology*, **97**, 131–141.
- Jørgensen RB, Andersen B (1994) Spontaneous hybridization between oilseed rape (*Brassica napus*) and weedy *B. campestris* (Brassicaceae). *American Journal of Botany*, **81**, 1620–1626.
- Kalinowsky ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics*, **5**, 539–543.
- Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, **15**, 237–240.
- Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences, USA*, **103**(25), 9578–9583.
- Manel S, Schwartz K, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.
- McGregor SE (1976) *Insect Pollination of Cultivated Crop Plants*. US Department of Agriculture, *Agriculture Handbook*, 496, 411.
- Meirmans PG, Bousquet J, Isabel N (2009) A metapopulation model for the introgression from genetically modified plants into their wild relatives. *Evolutionary Applications*, **2**, 160–171. doi:10.1111/j.1752-4571.2008.00050.x.
- Paetkau D, Slade R, Burdens M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.
- Palstra FP, O'Connell MF, Ruzzante DE (2007) Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. *Molecular Ecology*, **16**, 4504–4522.
- Petit RJ, Mousadirk A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.
- Piñeyro-Nelson A, Van Heerwaarden J, Perales HR *et al.* (2009) Transgenes in Mexican maize: molecular evidence and methodological considerations for GMO detection in landrace populations. *Molecular Ecology*, **18**, 750–761.
- Reddy OU, Pepper AE, Abdurakmonov I *et al.* (2001) New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. *Journal Cotton Science*, **5**, 103–113.
- Rogers DJ, Reid RE, Rogers JJ, Addison SJ (2007) Prediction of the naturalization potential and weediness risk of transgenic cotton in Australia. *Agriculture Ecosystems and Environment*, **119**, 177–189.
- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory, and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**, 380–390.
- SAGARPA (2010) <http://www.sagarpa.gob.mx> (Accessed on February 1, 2010).
- Sahoo L, Schmidt JJ, Pedersen JF, Lee DJ, Lindquist JL (2010) Growth and fitness components of wild x cultivated *Sorghum bicolor* (Poaceae) hybrids in Nebraska. *American Journal Botany*, **97**, 1610–1617.
- Sasu MA, Ferrari MJ, Du D, Winsor JA, Stephenson AG (2009) Indirect costs of a nontarget pathogen mitigate the direct benefits of a virus-resistant transgene in wild *Cucurbita*. *Proceedings of the National Academy of Sciences, USA*, **106**, 19067–19071.
- Scachetti-Pereira R (2001) *Desktop GARP*. <http://www.nhm.ku.edu/desktopgarp/index.html> (Accessed on July 2010).
- Seelanan T, Schnabel A, Wendel JF (1997) Congruence and consensus in the cotton tribe (Malvaceae). *Systematic Botany*, **22**, 259–290.
- Serratos-Hernández JA, Gómez-Olivares JL, Salinas-Arreortua N, Buendía-Rodríguez E, Islas-Gutiérrez F, de-Ita A (2007) Transgenic proteins in maize in the soil conservation area of Federal District, México. *Frontiers in Ecology and the Environment*, **5**, 247–252.
- Simpson D (1954) Natural cross-pollination in cotton. *US Department of Agriculture Technical Bulletin*, 1094.

- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.
- Smith C, Stephens S (1971) Critical identification of Mexican archaeological cotton remains. *Economic Botany*, **25**, 160–168.
- Snow AA, Andersen B, Jorgensen RB (1999) Costs of transgenic herbicide resistance introgressed from *Brassica napus* into weedy *B. rapa*. *Molecular Ecology*, **8**, 605–615.
- Stephens SG (1958) Factors affecting seed dispersal in *Gossypium* and their possible evolutionary significance. *North Carolina Agricultural Experiment Station Technical Bulletin*, No. 131. 32 pp.
- Stephens SG (1966) The potentiality for long range oceanic dispersal of cotton seeds. *The American Naturalist*, **100**, 199–210.
- Stephens SG, Finkner MD (1953) Natural crossing in cotton. *Economic Botany*, **7**, 257–269.
- Sul IW, Korban SS (1996) A highly efficient method for isolating genomic DNA from plant tissues. *Plant Tissue Culture Biotechnology*, **2**, 113–116.
- Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.
- Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659–669.
- Traxler G, Godoy-Avila S (2004) Transgenic cotton in Mexico. *AgBioForum*, **7**, 57–62.
- Viard F, Arnaud J-F, Delescluse M, Cuguen J (2004) Tracing back seed and pollen flow within the crop-wild *Beta vulgaris* complex: genetic distinctiveness Vs. hot spots of hybridization over a regional scale. *Molecular Ecology*, **13**, 1357–1364.
- Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledenous angiosperms. *Genome*, **42**, 9–19.
- Wendel JF (1989) New World tetraploid cottons contain Old World cytoplasm. *Proceedings of the National Academy of Science of USA*, **86**, 4132–4136.
- Wendel JF, Albert VA (1992) Phylogenetics of the cotton genus (*Gossypium* L.): character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany*, **17**, 115–143.
- Wendel JF, Brubaker CL, Álvarez I, Cronn RC, Stewart J McD (2009) Evolution and natural history of the cotton genus. In: *Genetics and Genomics of cotton, Plant Genetics and Genomics: Crops and Models 3* (ed. Paterson AH), pp. 1–20. Springer, New York.
- Wendel JF, Brubaker CL, Seelanan T (2010) The origin and evolution of *Gossypium*, Chapter 1. In: *Physiology of Cotton* (eds Stewart JM, Oosterhuis D, Heitholt JJ, Mauney JR), pp. 3–22. Springer, Netherlands.
- Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F_{ST} doesn't equal $1/(4Nm+1)$. *Heredity*, **82**, 117–125.
- Wiley EO, McNyset KM, Peterson AT, Robins CR, Stewart AM (2003) Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography*, **16**, 120–127.

A.W. is interested in the evolution, applied population genetics of plants, as well as the study of the centers of origin and diversification of cultivated plants. A.P.-N. is interested in molecular genetics and plant evolutionary development. J.A. is interested in modeling the distribution of species and in the analysis of spatial information. A.G. is interested in molecular detection and quantification of GM sequences as well as heterologous proteins. E.R.Á.-B. is interested in genetics, evolutionary development and plant conservation as well as biomathematics. D.P. is interested in the fields of population genetics and phylogeography. He is currently involved in the study of Mexican pines.

Data accessibility

Microsatellite data available in DRYAD: doi:10.5061/dryad.rd8fn

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Description of DNA extraction procedure and PCR conditions

Table S2 Pairwise comparison matrix of genetic distance and genetic diversity among metapopulations

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supporting information

S1 Description of DNA extraction procedure and PCR conditions

Description of DNA extraction procedure and PCR conditions DNA extraction (modified from Sul and Korban, 1996). Two grams of leaf tissue from seedlings were frozen in liquid nitrogen, ground with a mortar and pestle and transferred to a microtube; 600 µL of preheated 2x CTAB/PVP solution was added, vortexed 15 seconds and incubated at 65°C for 15 minutes; 600 µL of chloroform/isoamyl-alcohol (24:1) was added, shaken vigorously and centrifuged for 10 minutes at 3,000 rpm. The supernatant was transferred to a new microtube and 600 µL of fresh chloroform/isoamyl-alcohol (24:1) was added and the centrifugation step was repeated. The supernatant was transferred to a new microtube with 1 mL of 100% ethanol; the sample was mixed by inversion, incubated at 4°C for 30 minutes and centrifugated 2 minutes at 13000 rpm. The supernatant was discarded and the DNA pellet was rinsed twice with 70% (v/v) ethanol and dissolved in 400 µL 0.2 TE bufferDNA concentration was estimated through visual inspection in an agarose gel stained with ethidium bromide.

PCR amplification conditions

CODE	PCR program	Final volume (25µL)
CCMP2 Weising et al. 1999	94°C x 3'; 30 cycles of 94°C x 45'' + 60°C x 15'' + 72°C x 30''; 72°C x 5'	20ng DNA, 10mM Tris-HCl, 1.5mM MgCl ₂ , 50mM KCl, pH 8.3), 200µM dNTPs, forward and reverse primer 5 pmol, 0.5U Taq polymerase (Promega GoTaq).
CCPM3 Weising et al. 1999	94°C x 3'; 30 cycles of 94°C x 45'' + 55°C x 45'' + 72°C x 30''; 72°C x 5'	
CCMP5 Weising et al. 1999	94°C x 3'; 30 cycles of 94°C x 15'' + 58°C x 45'' + 72°C x 1'; 72°C x 7'	
CCMP10 Weising et al. 1999	94°C x 3'; 30 cycles of 94°C x 15'' + 52°C x 15'' + 72°C x 30''; 72°C x 7'	
AF351313 Reddy et al. 2002	94°C x 3'; 30 cycles of 94°C x 1' + 55°C x 1' + 72°C x 2'; 72°C x 7'	40ng DNA, 10mM Tris-HCl, 1.5mM MgCl ₂ , 50mM KCl, pH 8.3), 400µM dNTPs, forward and reverse primer 10pmol, 1U Taq polymerase (Promega GoTaq).

S2 Pairwise comparison matrix of genetic distance and genetic diversity among metapopulations

D_{est}/F_{ST}	BCSM	NPM	BBM	CPM	SPM	YPM	GSM	GNM	Genetic diversity
BCSM		0.20	0.84	0.19	0.10	0.83	0.97	0.95	0.7±0.1
NPM	0.08		0.84	0.36	0.22	0.82	0.98	0.95	0.78±0.04
BBM	0.56	0.57		0.88	0.82	0.96	1.00	1.00	0.94±0.05
CPM	0.13	0.24	0.66		0.24	0.79	0.80	0.65	0.902±0.01
SPM	0.03	0.12	0.54	0.11		0.78	0.96	0.91	0.805±0.04
YPM	0.14	0.17	0.36	0.16	0.14		0.95	0.56	0.93±0.01
GSM	0.69	0.61	0.60	0.62	0.53	0.23		0.32	0.34±0.17
GNM	0.45	0.46	0.51	0.39	0.36	0.10	0.12		0.77±0.06

F_{st} values are plotted under the diagonal; D_{est} values are plotted above. The last column corresponds to the estimation of haplotypic diversity according to $F_{ST} = 1 / (2N_e m + 1)$. All F_{ST} values were significant ($p=0.01$).

5. Determinación de los centros de diversidad genética de *Gossypium hirsutum* en México

Resumen

El establecimiento y el monitoreo de los centros de diversidad genética es importante para establecer políticas que permitan la subsistencia de una especie (Pistorius 1997). La metodología para su establecimiento y diseño de medidas de protección no es claro (Gepts 2004). Por ello, se propone por un lado utilizar la teoría de la genética de la conservación para establecer los centros de diversidad genética, equiparándolo al concepto de “unidad de conservación” y por el otro, la metodología para medir y conservar la diversidad genética. Las tres medidas que consideradas para este análisis son: la riqueza alélica, la heterocigosis y el coeficiente de consanguinidad, F .

Se utilizaron datos de cinco microsatélites ligados del cloroplasto como un solo locus (Wegier *et al.*, 2011) de poblaciones silvestres de *Gossypium hirsutum* para hacer un *agrupamiento* espacial, permitiendo la inclusión de coordenadas geográficas en el análisis, para obtener mejores inferencias sobre la estructura genética (Corander *et al.*, 2008) y se encontró una partición óptima de las poblaciones en 6 grupos. Las metapoblaciones que mostraban el menor valor de heterocigosis, y por tanto, que potencialmente podrían tener una peor respuesta inmediata a la selección, fueron Golfo Sur y Pacífico Centro, mientras que Bahía de Banderas y Península de Yucatán mostraron los valores más elevados y la mayor riqueza alélica.

De acuerdo a Palsbøll *et al.* (2007), una tasa de dispersión (proporción de individuos que van de un grupo a otro) menor al 10% indica que pertenecen a diferentes unidades de manejo. La alta tasa de migración encontrada (mayor a 10%) en *G. hirsutum*, indica que los sucesos en una población tienen alta probabilidad de repercutir en el resto, esto justifica una estrategia de conservación que involucre a las seis unidades de manejo identificadas, ya que, no son independientes. Las poblaciones de *G. hirsutum* tienen una alta capacidad de sobrevivir ante cambios adversos según sus tamaños efectivos, sin embargo, el número encontrado de plantas adultas es reducido, por lo que se requiere una estrategia para la conservación de la especie.

5.1 Introducción

La determinación de los centros de diversidad genética es importante para establecer políticas que permitan la subsistencia de una especie (Pistorius 1997), sin embargo, la metodología para su delimitación y diseño de las medidas para su protección no están claras (Gepts 2004). Por ello se propone que se utilice la teoría de la Genética de la conservación porque es posible establecer los centros de diversidad genética de una especie utilizando como referencia el término de “unidad de conservación”, la cual consiste en poblaciones o grupos que por sus características genéticas y ecológicas sean importantes de conservar (Geist y Kuehn, 2005).

Las unidades de conservación pueden ser unidades evolutivamente significativas, unidades de manejo o alguna área geográfica que los investigadores consideren importante conservar (Manel *et al.*, 2003). Las unidades evolutivamente significativas representan grupos de poblaciones que han divergido hace mucho tiempo, mientras que las unidades de manejo representan grupos de poblaciones que han divergido recientemente en donde podemos reconocer a los grupos de poblaciones como independientes debido al bajo flujo génico entre dichos grupos (Moritz, 1994). Después de encontrar cuáles son los centros de diversidad genética se pueden proponer estrategias para proteger por separado cada uno de éstos, utilizando la información disponible para ello (Palsbøll *et al.*, 2007).

Después de que se definen los centros de diversidad genética, la diversidad genética puede ser monitoreada dentro de cada centro utilizando tres medidas: 1) La heterocigosis, que afecta a la respuesta inmediata a la selección, ya sea artificial o natural (Allendorf, 1986 en Petit *et al.*, 1998); 2) La riqueza alélica, que muestra el límite de la respuesta a la selección (James, 1971 en Petit *et al.*, 1998) y 3) El coeficiente de consanguinidad, ya que ésta puede causar un descenso en la adecuación promedio de una población (Hedrick, 2001).

5.2 Centros de diversidad genética

La identificación y el mantenimiento de los centros de diversidad genética son importantes porque están relacionados con la adecuación de una población (Reed y Frankham, 2003) y ligados al potencial evolutivo de una especie (England *et al.*, 2003). De acuerdo a la ley de bioseguridad de organismos genéticamente modificados (LBOGM), un centro de diversidad genética es: “Aquella área geográfica del

territorio nacional donde existe diversidad morfológica, genética o ambas, de determinadas especies, que se caracteriza por albergar poblaciones de los parientes silvestres y que constituye una reserva genética” (Nueva Ley DOF 18-03-2005; www.diputados.gob.mx/LeyesBiblio/pdf/LBOGM.pdf).

En este capítulo de la tesis se propone primero, equiparar los centros de diversidad genética con un término ampliamente utilizado en la literatura de la genética de la conservación, el de las unidades de conservación. y segundo, proponer una estrategia metodológica para su determinación, con el objetivo final de preservar la diversidad de una especie, que en este caso será *Gossypium hirsutum*.

Según la LBOGM, también debe tomarse en cuenta que la diversidad morfológica de una población debe protegerse. Esta diversidad depende de los fenotipos encontrados en la misma y a su vez, los fenotipos son el resultado de la interacción entre el genotipo de cada individuo con el ambiente, por lo tanto, la diversidad morfológica depende de tres factores: 1) La información genética, 2) el ambiente (Gillespie, 2004) y 3) el fenotipo de los padres (Santure y Spencer, 2006; Wade, 1998; Bayer *et al.*, 2009). Por ello, la protección conjunta del ambiente y de la diversidad genética permitirá la conservación de la diversidad morfológica de una especie.

5.3 Unidades de conservación

El establecimiento de las unidades de conservación puede ayudarnos a tomar mejores decisiones para la conservación de la diversidad genética de una especie. Las unidades de conservación son: “Unidades evolutivamente significativas, unidades de manejo o cualquier unidad geográfica que los investigadores consideren importante conservar” (Manel *et al.*, 2003). En este capítulo estableceremos metodologías para definir tanto “unidades evolutivamente significativas” (ESU, *evolutionarily significant unit*) como “unidades de manejo” (MU, *management units*). Las ESU sirven para establecer grupos de poblaciones que han divergido hace bastante tiempo mientras que en las MU se establecen grupos de poblaciones que divergieron hace menos tiempo y donde hay un flujo génico tan bajo entre los grupos de poblaciones, que podemos reconocer a dichos grupos de poblaciones como independientes (Moritz, 1994).

5.3.1 Unidades evolutivamente significativas (ESU)

Las ESU han sido propuestas para ayudar a la conservación de los procesos ecológicos y evolutivos de una especie (Moritz, 1999), no obstante, existen varias definiciones. El término ESU fue propuesto por Ryder (1986) como: “subconjuntos del término más inclusivo de especies, que poseen atributos genéticos significativos para las generaciones presentes y futuras de la especie en cuestión” (Ryder, 1986 en Fraser y Bernatchez, 2001). La definición más usada sobre ESU fue propuesta por Moritz (1994), quien define ESU como grupos “recíprocamente monofiléticos para los alelos de mtADN y que muestran una divergencia significativa de las frecuencias alélicas de *loci* nucleares”. El concepto de ESU propuesto por Moritz ha sido debatido y cuestionado (Crandall *et al.*, 2000; Fraser y Bernatchez, 2001; Hey *et al.* 2003). Una de las mayores críticas es que cuando se tiene un proceso de subdivisión poblacional y especiación, se producen relaciones polifiléticas, que con el tiempo se vuelven parafiléticas y posteriormente monofiléticas. Por lo tanto, si utilizamos un criterio de monofilia recíproca en secuencias de mtADN para definir ESU, este criterio puede no satisfacerse a pesar de que sí se tengan ESU en la población, ya que pudo no haber transcurrido el tiempo suficiente para que se formen relaciones monofiléticas. Por ello, el criterio de monofilia recíproca puede ser demasiado estricto (Crandall *et al.*, 2000). Sin embargo, Moritz (2002) ha aclarado que el concepto de ESU sirve como un complemento al concepto de especie. Otra crítica que se ha hecho a la propuesta de Moritz es que este criterio es útil en animales, pero no en los grupos con bajas tasas de mutación en el mtADN como las plantas (Palmer, 1992; Fraser y Bernatchez, 2001).

Otra propuesta es definir una ESU como un conjunto de individuos, o poblaciones, que tienen caracteres hereditarios que pueden ser unidos en un *linaje*, utilizando el concepto de especie filogenética (PSC, *Phylogenetic species concept*; Vogler y DeSalle, 1994), pero se requiere de una elección correcta de caracteres homólogos con suficiente variación intrapoblacional.

Crandall *et al.*, (2000) propusieron otra forma para determinar la conservación de las poblaciones basándose en el intercambio genético y ecológico, tanto reciente como antiguo (Mastretta *et al.*, 2011). En la siguiente sección se verán algunas metodologías para calcular el intercambio genético. El intercambio ecológico se refiere a que los individuos puedan moverse entre poblaciones para poblar el mismo nicho ecológico. Para que haya intercambio ecológico, las poblaciones deben

compartir las mismas adaptaciones fundamentales, como su morfología y características similares de su ciclo de vida y de su demografía (Crandall *et al.*, 2000).

Analizando las posibilidades y la falta de consenso, Fraser y Bernatchez (2001) proponen que la metodología es dependiente del sistema de estudio. Ellos proponen el uso de las monofilias recíprocas para animales (como en Moritz 1994), mientras que en linajes que han experimentado una radiación adaptativa rápida, se propone delimitar las ESU mediante la investigación del efecto del intercambio ecológico (como Crandall *et al.*, 2000) o del uso del PSC (como Vogler y DeSalle, 1994).

5.3.2 Unidades de manejo (MU)

De acuerdo a la definición original de Moritz (1994), las MU son “poblaciones con una divergencia significativa de sus frecuencias alélicas en los *loci* nucleares o mitocondriales, sin importar la distinción filogenética de los alelos”. Después se ha precisado que lo más importante cuando se definen las MU, es encontrar poblaciones que estén lo suficientemente aisladas para que merezcan ser delimitadas, monitoreadas y manejadas por separado (Palsbøll *et al.*, 2007). De acuerdo a Palsbøll *et al.* (2007), se ha encontrado que cuando existen tasas de dispersión mayores al 10% (donde la tasa de dispersión representa la proporción de individuos que van de una población a otra), las poblaciones deben estar demográficamente correlacionadas (Hastings, 1993), por ello cuando las tasas de dispersión son menores al 10% entre dos grupos de poblaciones, podemos definir dichos grupos de poblaciones como dos MU.

Cabe aclarar que la tasa de dispersión en algunas ocasiones no está relacionada con el flujo génico, que está determinado por el parámetro $N_e m$, donde N_e es el tamaño efectivo de la población y m es la probabilidad de que un individuo muestreado sea un migrante. Se han documentado ocasiones donde existe una baja tasa de dispersión y un alto flujo génico, ya que los machos se aparean con hembras de otras poblaciones y después regresan a su población nativa (Amos y Balmford, 2001). Teóricamente, también sería posible encontrar casos donde existe una alta tasa de dispersión y un bajo flujo génico, en caso de que los individuos migrantes no se reprodujeran con los individuos de una cierta población. Por ello, en el futuro se requiere el desarrollo de la estimación de parámetros poblacionales de interés bajo modelos demográficos más complejos (Palsbøll *et al.*, 2007).

Para inferir las MU, se ha propuesto el uso de pruebas de asignación (Pearse y Crandall, 2004), como STRUCTURE (Pritchard *et al.*, 2000) o BAPS (Corander y Tang, 2007). Esto nos permite establecer *clusters* donde tenemos grupos de poblaciones. Después de la formación de *clusters*, se debe establecer la tasa de dispersión entre las poblaciones (Palsbøll *et al.*, 2007). La tasa de dispersión puede ser estimada de varias formas. Pueden utilizarse métodos directos de estimación del flujo génico, que a través de la identificación genética de padres e hijos, permiten una estimación directa. También pueden realizarse estimados del flujo génico utilizando medidas basadas en el estadístico F_{ST} (Wright, 1931). Sin embargo, ambos enfoques poseen problemas. Los métodos directos pueden hacer una mala estimación de la tasa de migración cuando se tienen poblaciones grandes y un número pequeño de migrantes, además están sujetos a las condiciones espacio-temporales del experimento, que pueden cambiar a través del tiempo. Mientras que los métodos indirectos parten de modelos neutrales que asumen modelos simplificados de estructura poblacional, como un tamaño poblacional constante y una tasa de migración simétrica y constante (Wilson y Rannala, 2003), aunque poseen la ventaja de que promedian el valor del flujo génico a través de la historia de las poblaciones. Algunos programas que pueden ayudarnos a estimar las tasas de dispersión, a través de las tasas de migración, son: 1) LAMARC (Kuhner, 2006), que puede calcular el flujo génico entre varias poblaciones pero ignora el tiempo de divergencia entre las poblaciones. 2) IM (Hey y Nielsen, 2007), que sirve para calcular el flujo génico entre grupos de poblaciones y 3) BayesAss+ (Wilson y Rannala, 2003), estima tasas de migración recientes con información genética de *loci* que no están ligados.

5.4 Monitoreo de la diversidad genética

Después de que se establecen las unidades de conservación, es necesario hacer un monitoreo de la diversidad genética por población para decidir las medidas que deben utilizarse para la preservación de la misma. La diversidad genética se define como la variedad de alelos y genotipos presentes en un grupo bajo estudio, donde éste puede ser una población, una especie o un grupo de especies (Frankham, 2002). Las tres medidas que consideramos más importantes para el análisis de la diversidad genética son: la riqueza alélica, la heterocigosis y el coeficiente de consanguinidad, F .

La riqueza alélica es el número de alelos que encontramos en una población (Kalinowski, 2004). Teóricamente, se ha mostrado que la riqueza alélica determina el límite de la respuesta a la selección (James, 1971 en Petit *et al.*, 1998) y determina el potencial evolutivo a largo plazo (James, 1971 en England *et al.*, 2003). Existen programas que utilizan algoritmos que son capaces de calcular la riqueza alélica de una población, la riqueza de los alelos únicos de cada población (Kalinowski, 2005) y la riqueza alélica de alelos que no se encuentran compartidos entre todas las poblaciones (Szpiech *et al.*, 2008). Adicionalmente, aportan información si se debe decidir cuáles son las poblaciones prioritarias para la conservación de la diversidad genética, tomando en cuenta tanto a la riqueza alélica como a la heterocigosis de las poblaciones (Castellanos-Morales, 2006).

La heterocigosis es una medida de la diversidad genética (Nei, 1978) que está relacionada con la respuesta inmediata a la selección, ya sea artificial o natural (Allendorf, 1986 en Petit *et al.*, 1998). La heterocigosis (H) se calcula como:

$$H = 1 - \sum_{i=1}^n \widehat{p}_i^2$$

Un estimador sin sesgo de la heterocigosis (H_E) en un locus, cuando se tienen tamaños de muestra pequeños, puede calcularse como (Hedrick, 2004):

$$H_E = \frac{N}{2N-1} \left(1 - \sum_{i=1}^n \widehat{p}_i^2 \right)$$

El parámetro θ (que es igual a $4Nu$ en marcadores provenientes de autosomas diploides y a $2Nu$ en marcadores provenientes de organelos haploides, como la mitocondria o el cloroplasto; donde N es el tamaño efectivo de la población y u es la tasa de mutación del marcador) también puede relacionarse con la heterocigosis y por tanto, con la respuesta inmediata a la selección. Si el marcador molecular que estamos utilizando se comporta de acuerdo a un modelo de evolución neutral, ambos parámetros pueden relacionarse según la siguiente ecuación (Gillespie, 2004):

$$H = \frac{\theta}{\theta + 1}$$

En secuencias, el parámetro θ se puede estimar utilizando el estimador de Watterson (1975) que relaciona el número de sitios segregantes con el valor real de θ . También se puede estimar utilizando el valor de θ utilizando el parámetro π , que nos dice la proporción de diferencias nucleotídicas entre pares de secuencias (Hedrick,

2004). Para microsatélites existen otras formas de estimar el valor de θ usando la homocigosis muestral y suponiendo un modelo de mutación paso a paso (Xu y Fu, 2004).

Adicionalmente, se ha propuesto una medida de la contribución de la heterocigosis y de la riqueza alélica de cada población a la heterocigosis y la riqueza alélica total de toda la especie (Petit *et al.*, 1998). Esta medida puede ayudarnos a cuantificar cuáles poblaciones tienen mayor diversidad genética dentro de cada unidad de conservación.

También es importante medir el valor la consanguineidad dentro de una población mediante el parámetro F_{IS} , una medida de la desviación de la heterocigosis observada respecto a la heterocigosis esperada bajo equilibrio Hardy-Weinberg (Keller y Waller, 2002). El valor de F_{IS} puede ser calculado mediante el programa Arlequín (Excoffier *et al.*, 2005). Un alto número de apareamientos consanguíneos en una población pueden ocasionar una depresión por endogamia, y por tanto, un descenso en la adecuación promedio de una población (Hedrick, 2001). Existen varias técnicas para relacionar el efecto del coeficiente de consanguinidad con el descenso en la adecuación promedio, por ejemplo, el uso de una recta de regresión lineal entre el coeficiente de consanguinidad y la adecuación funciona si suponemos que las mutaciones son independientes (Keller y Waller, 2002), aunque en ocasiones no se encuentra una relación lineal (Wang *et al.*, 2002). También se ha propuesto: estimar el efecto de la depresión por endogamia a partir de la heterocigosis, en los casos donde la heterocigosis se relaciona con la adecuación de un individuo (Mitton, 1997 en Edmands, 2007); relacionar los cambios en el coeficiente de consanguinidad para inferir selección contra los individuos que se aparean de forma consanguínea (Ritland, 1990); y por otra parte, usar el patrón de bandas compartidas entre dos individuos y la adecuación de su progenie para estimar la depresión por endogamia (Bensch *et al.*, 1994).

5.5 Métodos

5.5.1 Información genética utilizada

Se utilizaron datos de cinco microsatélites de cloroplasto de *Gossypium hirsutum* silvestre tomados de Wegier *et al.*, 2011. que fueron amplificados en 8 metapoblaciones: 1) Metapoblación Baja California Sur (Sur de Baja California Sur;

MBC); 2) Metapoblación Pacífico Norte (Centro y Sur de Sinaloa y Norte de Nayarit; MPN); 3) Metapoblación Golfo Norte (Norte de Veracruz, Este de San Luis Potosí y Sur de Tamaulipas; MGN); 4) Metapoblación Bahía de Banderas (Suroeste de Nayarit y Noroeste de Jalisco; MBB); 5) Metapoblación Golfo Sur (Centro y Sureste de Veracruz; MGS); 6) Metapoblación Pacífico Sur (Sureste de Guerrero, línea costera de Oaxaca, Centro Oeste, Centro y Sur de Chiapas; MPS); 7) Metapoblación Pacífico Centro (línea costera del Centro y Sur de Jalisco, Colima, Michoacán, Noroeste y Centro de Guerrero; MPC) y 8) Metapoblación Península de Yucatán (Quintana Roo, Yucatán, Campeche, Noreste y Este de Tabasco; MPY). Se amplificaron los cinco microsatélites en 10, 47, 13, 75, 50, 77, 11 y 16 individuos en cada población, respectivamente.

5.5.2 Establecimiento de los centros de diversidad genética de Gossypium hirsutum

Para establecer las unidades de manejo, primero utilizamos el programa *Bayesian Analysis of Population Structure* (BAPS v5.2) desarrollado por Corander *et al.* (2008) para establecer la estructura poblacional de *G. hirsutum* en México. Realizamos un agrupamiento espacial de individuos mediante la codificación de la información de los cinco microsatélites ligados del cloroplasto como un solo locus. Este análisis permite la inclusión de las coordenadas geográficas de donde fueron colectados los individuos, proporcionando mejores inferencias sobre la estructura genética, particularmente en evaluaciones donde la información genética es poco abundante (Corander *et al.*, 2008). Para establecer el número de grupos K que conforma la estructura poblacional de *G. hirsutum* se analizaron valores de K que iban de 2 a 8, realizando 10 iteraciones para cada valor de K . Después se realizaron 100 repeticiones para los valores de K que poseían un valor de verosimilitud más grande y se seleccionó el resultado donde se encontraba la estructura poblacional con el valor del logaritmo de máxima verosimilitud más alto.

5.5.3 Determinación de la tasa de dispersión entre grupos de poblaciones

Después de que se determinó la estructura poblacional, se estimó el valor de θ (que puede interpretarse como el número esperado de mutaciones esperadas entre un par de linajes de acuerdo a Hein *et al.*, 2005) para cada uno de los grupos definidos por BAPS suponiendo un modelo de alelos infinitos (Kimura y Crow, 1964).

Alavez (2008) ha demostrado que cuando se tienen 2 o más *loci* de microsatélites ligados, el valor de θ calculado sumando el valor estimado de θ de cada locus bajo un modelo de alelos infinitos (Kimura y Crow, 1964) es más exacto que cuando se supone un modelo de mutación de mutación *stepwise* para los microsatélites ligados (Ohta y Kimura, 1964; Xu y Fu, 2004). Por ello se calculó el valor de θ usando un modelo de alelos infinitos (Kimura y Crow, 1964). Podemos encontrar el valor de N_e a través de la siguiente fórmula:

$$\theta = 2N_e * l * u$$

Donde N_e es el tamaño efectivo de la población, l es el número de microsatélites y u es la tasa de mutación por generación. u se igualó a $8 * 10^{-6}$, que es la tasa de mutación por año estimada para los microsatélites de 11 pares de bases en *Arabidopsis thaliana* multiplicada por el tiempo generacional en *G. hirsutum*, que es igual a la edad donde la especie alcanza la mayor reproducción. El cálculo de θ nos permitirá estimar el valor del N_e , necesario para calcular el valor de la tasa de migración entre poblaciones.

De acuerdo a Palsbøll *et al.* (2007), una tasa de dispersión (que representa la proporción del número de individuos que van de una población a otra) menor al 10% entre dos grupos de poblaciones indica que dichos grupos de poblaciones deben ser asignadas a diferentes unidades de manejo. Podemos relacionar la tasa de dispersión con el flujo génico, aunque en ciertas ocasiones esto puede no ser apropiado (Palsbøll *et al.*, 2007). El flujo génico se mide mediante el parámetro $N_e m$, donde m representa la probabilidad de que un individuo sea un migrante. Si entre dos poblaciones m es menor al 10%, podemos definir a los dos grupos de poblaciones como unidades de manejo independientes. En marcadores genéticos provenientes de organelos haploides, este parámetro puede relacionarse con el parámetro F_{ST} , que representa la proporción de la diversidad genética que se debe a diferencias en la frecuencia alélica entre poblaciones, a partir de la siguiente relación:

$$F_{ST} = \frac{1}{(2 N_e m + 1)}$$

Se obtuvo el valor del flujo génico entre cada grupo de poblaciones establecidos con BAPS mediante el cálculo del valor de F_{ST} (Excoffier *et al.*, 1984) entre cada par de grupos de poblaciones utilizando el programa Arlequín (Excoffier *et al.*, 2005) y se

probó la significancia de cada valor de F_{ST} mediante 10,000 permutaciones de haplotipos entre poblaciones. El valor de N_e tomando cada par de grupos de poblaciones se obtuvo mediante la media armónica del N_e de dichos grupos de poblaciones (Gillespie, 2004):

$$\frac{1}{N_e^{dos\ grupos}} = \frac{1}{2} \left(\frac{1}{N_e^{grupo\ 1}} + \frac{1}{N_e^{grupo\ 2}} \right)$$

A partir del valor de N_e promedio entre cada par de grupos de poblaciones se calculó el valor de m entre cada grupo de poblaciones.

5.5.4 Monitoreo de la diversidad genética

Se midió la diversidad genética dentro de cada población utilizando: i) La heterocigosis esperada, o diversidad genética (Nei, 1987 en <http://cmpg.unibe.ch/software/arlequin3/arlequin31.pdf>), mientras que ii) La riqueza alélica (el número de alelos) y iii) la riqueza alélica privada (el número de alelos exclusivos de una población) fueron calculadas con el programa ADZE (Szpiech *et al.*, 2008) mediante el enrarecimiento, que permite establecer el número de alelos que esperaríamos en una muestra de tamaño g de cada población. El valor de g debe ser menor o igual a la población con el tamaño más pequeño (en este caso la metapoblación Baja California Sur: $g = 10$). Este es el valor más grande de g que podemos utilizar, aunque valores mayores de g dan una mejor estimación de la riqueza alélica privada (Szpiech *et al.*, 2008).

5.6 Resultados del análisis bioinformático: Identificación de los centros de diversidad genética en *Gossypium hirsutum*

5.6.1 Estructura genética

Se encontró una partición óptima de las poblaciones en 6 grupos con BAPS (Tabla 1) con un valor del logaritmo de máxima verosimilitud igual a -780.5852.

Tabla 1.- Subestructura poblacional inferida con BAPS junto con el valor de N_e y el valor de θ , para cada grupo definido por BAPS, suponiendo un modelo de alelos infinitos.

Linaje	Metapoblaciones	θ	N_e
1	MBC y MPN	2.191865066	27398.3133
2	MBB	4.698412698	58730.1587
3	MPC	1.463559685	18294.4961
4	MPS	2.395488077	29943.601
5	MPY	4.731835368	59147.9421
6	MGS y MGN	1.862857143	23285.7143

Donde Metapoblación Baja California Sur (MBC); Metapoblación Pacífico Norte (MPN); Metapoblación Golfo Norte (MGN); Metapoblación Bahía de Banderas (MBB); Metapoblación Golfo Sur (MGS); Metapoblación Pacífico Sur (MPS); Metapoblación Pacífico Centro (MPC) y Metapoblación Península de Yucatán (MPY).

5.6.2 Determinación de las unidades de manejo

Se estimó el valor de θ , suponiendo un modelo de alelos infinitos, y el valor de N_e para cada grupo definido por BAPS (Tabla 1). El valor del N_e dentro de cada grupo es mayor a 10,000. Se estimó el valor del parámetro F_{ST} entre todos los grupos definidos por BAPS (Tabla 2). En todas las comparaciones pareadas se rechazó la hipótesis de que no existen diferencias entre los grupos de poblaciones (todos los resultados fueron significativos), lo que refuerza la estructura poblacional encontrada por BAPS.

Tabla 2. Valores de F_{ST} (p -value < 0.001) en comparaciones pareadas entre grupos de poblaciones (bajo la diagonal). Ver abreviaturas de poblaciones en Tabla 1.

Linaje	MBC y MPN	MBB	MPC	MPS	MPY
MBC y MPN	0				
MBB	0.513468	0			
MPC	0.204102	0.603209	0		
MPS	0.093164	0.4734938	0.108256	0	
MPY	0.201378	0.33051077	0.183885	0.165162	0
MGS y MGN	0.52456	0.529885	0.490002	0.443297	0.149363

5.6.3 Monitoreo de la diversidad genética dentro de cada unidad de manejo

Se calculó el valor de la heterocigosis para cada población (Tabla 3), las poblaciones que mostraron el menor valor de heterocigosis, y por tanto que potencialmente podrían tener una peor respuesta inmediata a la selección, fueron las Metapoblaciones Golfo Sur y Pacífico Centro. Las metapoblaciones de Bahía de Banderas y Península de Yucatán mostraron el valor de heterocigosis más elevado.

En cuanto a la riqueza alélica (Tabla 3), las poblaciones que mostraban un mayor número esperado de haplotipos fueron las metapoblaciones de Bahía de Banderas y Península de Yucatán. Las metapoblaciones de Pacífico Norte y Pacífico Sur también poseen una cantidad importante de haplotipos únicos (en promedio se pueden encontrar 2 haplotipos únicos esperados por cada 10 haplotipos muestreados), y en cuanto a contribución a la diversidad genética, un poco más abajo se encuentran las metapoblaciones de Baja California Sur y Golfo Norte (con un promedio de 1 haplotipo único esperado por cada 10 haplotipos muestreados), mientras que las metapoblaciones de Golfo Sur y Pacífico Centro no tienen una aportación importante de haplotipos únicos.

Tabla 3.- Heterocigosis esperada en cada población, además de la riqueza alélica y riqueza alélica privada (tomando en cuenta el haplotipo completo) usando el método del enrarecimiento en una muestra al azar de tamaño de 10 de cada población.

Metapoblación	Heterocigosis	Riqueza alélica	Riqueza alélica privada
Baja California	0.73333333	4	0.625061
Pacífico Norte	0.78075856	4.81917	2.08928
Bahía de Banderas	0.96153846	8.26923	7.5
Pacífico Centro	0.65801802	3.11305	0.0245746
Pacífico Sur	0.81306122	5.54742	2.42994
Península de Yucatán	0.92173616	7.34237	4.33044
Golfo Sur	0.34545455	2.81818	0.0479558
Golfo Norte	0.775	4.3735	0.72709

5.7 Discusión

Para tomar decisiones sobre la conservación de la diversidad genética de una población, primero debemos establecer unidades de conservación. En cada una de las unidades de conservación está representada una fracción de la diversidad genética de una especie y cada una de estas unidades de conservación representa un centro de diversidad genética. Las unidades de conservación pueden dividirse en: unidades

evolutivamente significativas y unidades de manejo. Las unidades evolutivamente significativas tratan de dividir linajes que divergieron y se han mantenido aisladas desde hace mucho tiempo, mientras que las unidades de manejo dividen a las poblaciones en linajes (grupos o *clusters*), que contienen grupos de poblaciones, donde el flujo genético se encuentra restringido.

Por otro lado, se debe tratar de conservar la variabilidad genética de cada una de estas unidades de conservación. Para ello, debe mantenerse un monitoreo de la diversidad genética de cada población dentro de cada unidad de conservación. Este monitoreo debe realizarse calculando la riqueza alélica, la heterocigosis y el valor del coeficiente de consanguinidad. En las decisiones sobre la conservación de la diversidad genética que se tomen (por ejemplo, la elección de la conservación de ciertas poblaciones), se debe procurar que la riqueza alélica y la heterocigosis tengan valores altos, pues esto permitirá que las poblaciones mantengan un mayor potencial evolutivo que permita la subsistencia de la especie. Adicionalmente, se debe procurar la conservación del ambiente, ya que la conservación del ambiente junto con el mantenimiento de la diversidad genética permitirán la conservación de la diversidad morfológica de la especie bajo estudio.

Para garantizar la conservación de una especie es indispensable que se establezcan las medidas de protección necesarias. Esto se puede realizar integrando la información disponible, por ejemplo, análisis del modelos de flujo génico, datos de las dinámicas ecológicas e incluso, cuando sea el caso, se debe considerar el análisis social y económico que pudiera afectar, reducir o tolerarse para que se conserve la diversidad genética de las áreas.

5.7.1 Estructura genética encontrada en las metapoblaciones de Gossypium hirsutum

Se encontró una partición óptima en 6 grupos con BAPS (Tabla 1). Las poblaciones del algodón *G. hirsutum* en México por lo tanto están estructuradas, lo cual significa que distintas regiones de México muestran diferencias importantes en su composición genética, ya sea porque cada grupo de poblaciones tiene diferentes alelos, o porque los grupos de poblaciones poseen los mismos alelos en frecuencias diferentes entre grupos o una combinación de ambas.

El valor del N_e dentro de cada grupo es mayor a 10,000. Esto representa un valor alto ya que se ha sugerido que se necesitan valores de N_e de al menos 500 para que las poblaciones mantengan su potencial evolutivo (Frankham *et al.*, 2002).

Los análisis publicados en Wegier *et al.*, 2011 prueban la elevada probabilidad de flujo génico entre las poblaciones de *G. hirtutum* bajo el modelo de migración a larga distancia. Aún con una alta tasa de migración permanece la diferenciación genética entre los grupos y por lo tanto la estructura, lo cual puede deberse a la dinámica metapoblacional, por lo que estos resultados deben incorporarse al análisis integral de las medidas de protección de las áreas. En el mismo trabajo se demuestra el flujo génico de plantas domesticadas a plantas silvestres (capítulo 4). Las plantas domesticadas son genéticamente homogéneas por lo que un constante flujo de ellas a las poblaciones silvestres puede ocasionar un decremento en la diversidad genética. Además, actualmente la diversidad genética parece mantenerse en las poblaciones que presentan transgenes, sin embargo, hay incertidumbre de las consecuencias que pudieran presentarse con el tiempo en las interacciones del ecosistema y en las plantas mismas después de varias generaciones.

Las poblaciones que mostraron un mayor número esperado de haplotipos fueron las metapoblaciones de Bahía de Banderas y Península de Yucatán. Por ello y por su alta diversidad genética, su monitoreo y conservación es prioritaria para el mantenimiento de la diversidad genética en *G. hirsutum*. Sin embargo, la conservación de dichas poblaciones no garantizaría la conservación de la diversidad genética en México como refleja la riqueza alélica privada. Las metapoblaciones de Pacífico Norte y Pacífico Sur también poseen una cantidad importante de haplotipos únicos (en promedio 2 por cada 10 muestreados), y Baja California Sur y Golfo Norte (1 por cada 10 muestreados), mientras que las metapoblaciones Golfo Sur y Pacífico Centro no tienen una aportación importante de haplotipos únicos.

No se cuantificó el valor de la tasa de consanguinidad (F_{IS}), porque se necesita información de marcadores genotípicos (Keller y Waller, 2002), con la cual no se cuenta en este momento. F_{IS} es importante porque se relaciona con la diversidad genética (Frankham *et al.*, 2002) y la adecuación promedio de la población (Hedrick, 2001).

5.7.2 Identificación de los centros de diversidad genética en Gossypium hirsutum y medidas necesarias para su conservación

Las unidades de conservación de *G. hirsutum* identificadas en este análisis fueron seis, debido a la diferenciación entre las metapoblaciones ubicadas en Baja California Sur y Pacífico Norte así como las de Golfo Norte y Sur, que se comportan como unidades evolutivas.

Por otro lado el modelo de migración a larga distancia y los estudios publicados sobre flujo en el artículo de Wegier *et al.*, 2011 (capítulo 4), indican que aunque se mantiene la estructura, los niveles de flujo son altos, por lo tanto la distancia no ejerce una barrera ante el flujo y son otros factores los que limitan la dispersión entre las poblaciones. Es relevante también para el establecimiento de las medidas de protección, saber que la principal migración es vía semilla con un posterior contacto vía polen entre las poblaciones, y entre plantas cultivadas con sus parientes silvestres. Esto sumado a que el algodón es una planta importante económicamente, y por lo tanto, las dinámicas sociales también deben ser consideradas para generar medidas de protección para la especie.

La alta tasa de migración encontrada, indica que lo que suceda en una población tiene alta probabilidad de repercutir en el resto de las mismas, esto justifica que la estrategia de conservación involucre a las seis unidades de manejo identificadas, ya que, en otras palabras, no se encuentran aisladas entre ellas.

Las metapoblaciones tienen una alta capacidad actualmente de sobrevivir ante cambios adversos según lo reflejado por sus tamaños efectivos, sin embargo el número encontrado de plantas adultas es muy reducido ya que las mismas se ven afectadas por prácticas como cambios de uso de suelo y otras perturbaciones, por lo que es necesario iniciar una estrategia activa para la conservación de la especie.

Se ha demostrado que la capacidad de dispersión de las semillas se debe a múltiples factores. Los naturales, podemos dividirlos en bióticos y abióticos, y los artificiales o factores antropogénicos. Los bióticos son las aves que transportan las fibras para hacer sus nidos y animales que las consumen por su alto contenido lipídico. Los abióticos pueden mover las semillas mayores distancias que los bióticos, ya que son las corrientes marinas y de agua dulce, el viento en huracanes y tormentas, y la combinación de ambos. Debido a que las semillas son quiescentes (germinan cuando las condiciones ambientales son adecuadas) no pierden su viabilidad durante la migración (esto se encuentra fuertemente sustentado por la historia evolutiva de la

especie; Wendel *et al.*, 2011). Los factores antropogénicos se encuentran principalmente ligados a los eslabones de la cadena productiva: en el cultivo, transporte cultivo-despepitadora, de la despepitadoras mismas, las semillas sin fibras transportadas desde las despepitadora a puntos de concentración o a su punto final y del punto final mismo), que pueden ser ranchos vacunos, ya que la semilla se utiliza como alimento para ganado; pero también, pueden dispersarse semillas de forma intencional para ornato y otros fines, sin poder identificar si las semillas tienen transgenes, ya que no se puede identificar su diferencia con las plantas que no los tienen sin realizar análisis específicos para ello.

La estrategia para la conservación de las seis unidades de manejo debe ser integral con acciones diversas, ya que las vías de migración y los factores que pudiesen disminuir la diversidad son complejos y diversos.

- 1) Para disminuir el riesgo de migración entre silvestres y cultivados se deben utilizar barreras físicas, químicas y biológicas en los cultivares.
- 2) Para disminuir el riesgo de migración de semillas cultivadas por medios antropogénicos:
 - a. Todas estas fuentes se pueden evitar con el transporte adecuado y extrayendo los aceites de la semilla en el mismo sitio donde se localiza la despepitadora y antes de ser comercializada y trasladada para alimentar ganado, como sucede en otros países del mundo.
 - b. Los puntos 1 y 2 podrían dejar de ser un riesgo con los instrumentos propuestos por la LBOGM.
 - c. Las semillas genéticamente modificadas importadas de los Estados Unidos para alimento para ganado deberían entrar al país no viables o etiquetadas y bien empaquetadas para evitar su dispersión. Esto se puede regular con los instrumentos del Protocolo de Cartagena y el Protocolo de Nagoya-Kuala Lumpur sobre responsabilidad y compensación de la biotecnología, si se argumenta la probabilidad suficiente de que puede haber daños resultantes de organismos vivos modificados (en el centro de origen y diversidad del algodón a largo plazo) que tienen su origen transfronterizo.
 - d. Realizar campañas de información efectivas que permitan a la población tomar decisiones informadas sobre el movimiento de semillas GM y sus consecuencias.

- 3) Para lograr la conservación de las plantas dentro de su ecosistema se debe considerar la dinámica ecológica metapoblacional, esto es que el nicho debe ser conservado cuando las plantas están presentes y también cuando no lo están. En otras palabras, las estrategias de conservación deben estar orientadas a la conservación del hábitat más que a cada una de las plantas que actualmente podemos encontrar.
- 4) Se debe monitorear la frecuencia de los transgenes en las metapoblaciones e investigar sus consecuencias sobre la diversidad de las plantas pero también en los ecosistemas (lo anterior ligado al punto tres, es decir, garantizar la conservación del hábitat).
- 5) El diseño de la estrategia de conservación *in situ* de las poblaciones debe integrar las dinámicas sociales, culturales y económicas que se presentan en las localidades donde habitan las plantas.
- 6) Las medidas de bioseguridad que se han aplicado durante los 16 años en los que se ha cultivado algodón GM en México no han evitado el flujo génico por lo que se deben reconsiderar y revisar, además de verificar su adecuada y correcta aplicación. Debe también considerarse si las medidas de bioseguridad propuestas son viables en nuestro país dadas las dinámicas sociales, culturales y económicas que se presentan.

5.8 Conclusiones

- 1) El primer paso para la conservación de la diversidad genética es determinar los centros de diversidad genética (unidades de conservación) y utilizar una estrategia integral para generar las medidas de protección.
- 2) Se debe monitorear la diversidad genética dentro de cada *Unidad de conservación* mediante la estimación de la heterocigosis y la riqueza alélica, además del monitoreo periódico de los transgenes (temporalidad que debe establecerse después de un periodo de monitoreo piloto).
- 3) A lo largo de su extensión, México cuenta con 6 centros de diversidad genética (o unidades de manejo) que deben ser monitoreados por separado, ya que entre cada centro existe una distinta composición genética que es de interés para la conservación de la diversidad genética de la especie.

- 4) El flujo y la migración encontrado entre ellas, aunado al fenómeno de flujo a larga distancia, indican que rápidamente lo que afecte a una de las metapoblaciones puede afectar al resto, por lo que no se puede priorizar la conservación de alguna de las áreas y por lo tanto, se deben conservar todas.
- 5) Las poblaciones dentro de los centros de diversidad genética de *G. hirsutum* en México tienen un aporte diferente a dicha diversidad, y existe variación genética que es exclusiva de ciertas poblaciones. Por ello, al momento de realizar estrategias para la conservación de la diversidad genética de *G. hirsutum* se debe tomar en cuenta que existen poblaciones que tienen una composición genética exclusiva que, en caso de que desaparezca la población, puede ser irrecuperable.
- 6) Es urgente interrumpir el flujo génico constante entre cultivados y silvestres en México, ya que la diversidad genética encontrada es alta pero el número de individuos adultos es pequeño.

Agradecimientos

A la CARB-CONABIO por su apoyo y coordinación en el Proyecto “Análisis para la determinación de los centros de origen y diversidad de las especies mexicanas del género *Gossypium*” dentro del marco del proyecto financiado por la DGSPRNR “Conocer el estado de las especies prioritarias de las que México es centro de origen y diversidad genética, en el marco del artículo 86 de la Ley de Bioseguridad de los Organismos Genéticamente Modificados”.

A Diego Ortega Del Vecchyo, Valeria Alavez, Lev Jardón, Leticia Moyers y Daniel Piñero, por su colaboración en el proyecto. Al comité sinodal de esta tesis, Valeria Alavez y Adriana Uscanga, por sus observaciones al manuscrito.

Agradecemos el apoyo del Herbario Nacional (MEXU) y al Instituto de Biología, UNAM, por las facilidades otorgadas para la realización de esta investigación, a los herbarios ARIZ y NYBG, por permitir la consulta de sus imágenes.

6. Diversidad genética y conservación de *Gossypium hirsutum* silvestre y cultivado en México: Conclusiones y perspectivas

Resumen

Este capítulo final retoma las ideas y conclusiones de la investigación para realizar un análisis integral, desde el marco conceptual hasta las aplicaciones prácticas que se pueden desprender de los resultados, para la conservación de la diversidad genética de la especie a largo plazo.

La aplicación de conceptos de la genética de poblaciones, de la conservación y del paisaje, permiten generar resultados tangibles para comprender los procesos evolutivos y simplificar la gestión de la diversidad. Sin embargo, las diferentes actividades humanas deben considerarse con sus contextos sociales, económicos y jurídicos, en programas a largo plazo. Aquí se plantean estrategias de manejo de las poblaciones silvestres, se identifican algunas causas del flujo génico entre plantas cultivadas y silvestres y se proponen medidas de bioseguridad que pueden contribuir a limitarlo. Finalmente, aún es necesario evaluar las medidas propuestas en campo y monitorear el efecto de los transgenes en los ecosistemas.

6.1 Conclusiones y perspectivas

6.1.1 Biología evolutiva

La revisión de la biología evolutiva del género fue trascendental para la formulación de las hipótesis y ahora sabemos que los patrones de migración a larga distancia y la tolerancia al estrés ambiental, como la salinidad, han caracterizado la evolución de la especie, el género y la tribu *Gossypiae* (Capítulo 2). Es importante, por lo tanto, señalar que la revisión de la historia evolutiva de especies y géneros cercanos puede aportar información (a veces de manera rápida y oportuna) para la toma de decisiones sobre el uso y manejo de las especies, además de ser una de las principales herramientas para el diseño de estrategias para la conservación de la diversidad genética.

6.1.1.1 Perspectivas biología evolutiva. Muchos de los esfuerzos en biología evolutiva están enfocados a la reconstrucción histórica del origen, función y efecto de los genes. Ahora tenemos la posibilidad de estudiar poblaciones silvestres en las que conocemos el origen del transgén, su función y secuencia, por lo que podemos monitorear su frecuencia. Sin embargo, representa un reto teórico y práctico, el análisis de su efecto sobre la diversidad genética, específica y funcional del ecosistema.

6.1.2 Flujo génico

La diferencia entre el flujo génico ancestral y reciente es a veces imperceptible en especies que divergieron hace poco tiempo, sin embargo, es muy importante para la toma de decisiones. La construcción de una red de haplotipos que conecta a los mismos por pasos mutacionales, permitió reconocer algunos eventos de migración más antiguos que otros, sin embargo, es imposible fecharlos con este método. Los algodones genéticamente modificados tienen construcciones genéticas nuevas para la especie, y por lo tanto, pueden ser usados como un marcador molecular del que conocemos su secuencia, función y periodo exacto de aparición, permitiendo saber con mayor exactitud cuándo ocurre el flujo génico. En esta investigación se infirió el flujo génico ancestral (en la especie; capítulo 4) e histórico (del género; capítulo 2) y se comprobó que es un fenómeno continuo utilizando a los transgenes como marcadores moleculares (capítulo 4).

6.1.3. Genética del paisaje

La aplicación de conceptos de la genética de poblaciones y la ecología del paisaje pueden generar resultados tangibles que permiten comprender los procesos evolutivos que determinan la evolución colectiva de las poblaciones pero también simplificar la gestión de la diversidad (Rieseberg y Burke 2001). En la genética del paisaje se han generado herramientas para el análisis gráfico de las poblaciones, lo cual es fundamental, para responder preguntas de casos específicos, donde el flujo entre las poblaciones está limitado por factores bióticos y abióticos del paisaje que las separa, más que por su capacidad de movimiento.

Las redes de poblaciones son herramientas generadas en la genética del paisaje (ver capítulos 1 y 4) que permiten la diferenciación entre fenómenos que subyacen la diferenciación genética entre las poblaciones porque incorporan información

geográfica para evaluar la contribución del espacio físico en la estructuración de la diversidad genética (Manel *et al.*, 2003; Dyer y Nason 2004, Dyer 2009), por ejemplo, entre los modelos de flujo génico: el aislamiento por distancia (cuando las poblaciones más cercanas genéticamente también lo son geográficamente y la diferenciación va creciendo al aumentar la distancia) y las migraciones a larga distancia (cuando la dispersión está limitada por barreras físicas o biológicas, no por la distancia misma, y por lo tanto, no hay una correlación entre la distancia genética y la geográfica).

En esta investigación el análisis de la red de poblaciones y su extrapolación espacial, fue crucial para interpretar los resultados y probar las hipótesis (Capítulo 4). La separación geográfica entre cultivos genéticamente modificados de áreas donde no se desea su presencia es una práctica común en muchos países, sin embargo, como hemos visto en esta tesis (Capítulos 2, 4 y 5), pretender la separación (o aislamiento) por distancia no es una medida funcional para las especies del género *Gossypium*. Es entonces necesario que se evalúe la eficiencia de esta medida en otros cultivos GM antes de ser utilizada.

6.1.3.1. Perspectivas genética del paisaje. Las herramientas de genética del paisaje pueden apoyar en el futuro a comprender el aporte de la migración del polen y diferenciarlo de la migración de las semillas. Además, surgen preguntas sobre la influencia e importancia de los efectos de borde en una dinámica ecológica metapoblacional como la encontrada, que pueden ser utilizados para generar estrategias de conservación.

Las perturbación en las dunas costeras y selvas bajas en donde encontramos al algodón es constante (principalmente por influencia humana y climática, como huracanes y tormentas tropicales), además las condiciones del suelo son poco favorables para el crecimiento (suelos salinos y arenosos, bajos en nutrientes y alta insolación), por lo que el análisis genético y su correlación con estos factores de estrés puede contribuir al fitomejoramiento de cultivos poco tolerantes a estas condiciones e incluso variedades de algodón.

6.1.4. Genética de la conservación

Utilizar la teoría y herramientas de la genética de la conservación para delimitar las áreas de diversidad genética que pueden ser consideradas como centros de diversidad

genética así como sugerir las estrategias para su conservación es viable (Capítulo 5), además, de ser una de las formas relativamente rápidas para arrojar gran cantidad de información, sin estar sujeta a las condiciones espacio temporales en las que se llevó a cabo el estudio; sin embargo, se debe continuar monitoreando la eficacia de las medidas sugeridas. Una vez más nos enfrentamos al reto de integrar la información genética con la ecológica; es por ello que hemos propuesto el término “unidad de conservación” (grupos que por sus características genéticas y ecológicas sean importantes de conservar; Manel *et al.*, 2003), las cuales pueden ser “unidades evolutivamente significativas, unidades de manejo o cualquier unidad geográfica que los investigadores consideren importante conservar”. La delimitación de los centros de diversidad involucra información más allá de lo biológico, como la diversidad de usos y lingüística, de la misma forma que las unidades de manejo, en las que se pueden incluir rutas de transporte y comercio. Se pueden integrar muchos métodos e información para su delimitación, sin embargo incluir la historia evolutiva de la especie y sus parientes es crucial para lograr un resultado efectivo a largo plazo.

6.2 Estrategias de bioseguridad y medidas de protección para la conservación de la diversidad genética de *Gossypium hirsutum* en México

6.2.1 Estrategias de bioseguridad

Es indudable que las diferentes actividades humanas tienen gran influencia sobre las plantas útiles. Es importante entonces pensar en ellas dentro de los contextos sociales, económicos y jurídicos en los que están inmersas, además de las dinámicas ecológicas y evolutivas mencionadas en este trabajo. Las medidas de bioseguridad deben ser consideradas entonces, dentro de las posibilidades del país, además de monitorear su cumplimiento y contemplar la mitigación de consecuencias en caso de presentarse resultados inesperados. Si lo anterior no puede cumplirse, entonces desde el análisis de riesgo deberá considerarse que se carece de un programa estratégico de monitoreo y de documentación de resultados inesperados; sobre todo de aquellos cultivares de los que somos centro de origen y diversidad, como el maíz, calabaza, chile, frijol, entre muchos más, donde los riesgos pueden tener consecuencias mayores.

Los centros de origen y diversidad albergan a los parientes de la misma especie o a los parientes cercanos, pero también son las áreas donde habitan los organismos con los que estas especies han mantenido interacciones por millones de años (en la mayoría de los casos), las cuales a su vez mantienen complejas redes de interacción; la posibilidad de daño ambiental de un organismos GM en un ambiente en donde se tienen estas posibilidades es mayor, difícil de medir y de mitigar, en cambio en un área donde el cultivo GM es una especie introducida su impacto puede ser similar al de la misma especie sin En el caso del algodón hemos identificado algunas de las **posibles causas por cuales se pueden escapar las semillas GM**. La primera es el resultado de esta investigación y la podemos denominar como **la causa biológica**: la capacidad de dispersión de la especie. Las características biológicas de la especie indican que la distancia es una medida que no previene el flujo a larga distancia como se ha demostrado en este trabajo (Capítulos 2, 4 y 5). Por lo tanto, generalizar las medidas de bioseguridad para todos los cultivos, áreas o países, es incorrecto, es necesario que estas se analicen caso por caso, ya que su efectividad depende tanto de la especie como del ambiente receptor, además del contexto social y cultural en el que se pretende que las medidas sean efectivas.

La segunda causa es la responsabilidad legal: en la cadena productiva, las semillas viables de algodón GM dejan de estar custodiadas por los responsables legales en diferentes etapas. La primera en el campo de cultivo por eventos asociados al clima (viento, tormentas, huracanes), canales de riego y vías de comunicación cercanas (al separarse del fruto, fácilmente alcanzan canales de riego y caminos, de donde continúan migrando por el agua o por el movimiento de vehículos), animales (las aves recogen las fibras y semillas para hacer nidos; los animales que comen las semillas y algunas de ellas continúan viables) y humanos (que recogen semillas de carreteras y cultivos con diversos fines); durante la transportación, debido a que los transportes son inadecuados (abiertos, tapados con lonas o tienen cajas con espacios por donde las semillas se pueden caer), sin registro de sus rutas y limpieza que permita el control de las semillas. Así, este fenómeno ocurre del cultivo al acopio temporal, de éste a la despepitadora y desde ésta a cualquier sitio donde alguien desee comprar la semilla (en México la mayoría se distribuye ranchos vacunos). Finalmente, la responsabilidad legal deja de existir cuando la semilla es separada de su fibra, y por razones comerciales pero no biológicas (la semilla es igual de viable con y sin fibra), se le comienza a llamar “grano”, debido a que su destino comercial no debiese ser

germinar en campos de cultivo, pero se carece de justificación biológica o legal para ello.

La tercera causa es similar a la anterior, pero tiene un origen distinto, se trata de la “**migración trasfronteriza**”. México importa semilla de algodón GM viable para alimentar ganado. Esta semilla (mal nombrada “grano”) carece de etiquetas e información para el usuario y está fuera de los planes de contingencia de las instituciones reguladoras de GM de México. La información que permitiría su uso responsable como rutas por las que se transporta, destinos y eventos que contiene, es inexistente.

La cuarta causa son los “actores desinformados” que manejan semillas “huérfanas” (semillas GM sin responsable legal); estos actores no utilizan la semillas GM por su modificación genética, carecen de la información adecuada para su manejo porque no se les proporciona. La historia de las semillas “huérfanas” es variable, principalmente son trasladadas a los ranchos vacunos (a veces por medio de varios intermediarios) donde se utilizarán como alimento para ganado. Para mejorar el rendimiento de este alimento las semillas son almacenadas en lugares frescos, aireados con baja probabilidad de incendio (las semillas son inflamables debido al alto contenido de lípidos). En países como Argentina donde el “picudo” del algodón (*Anthonomus grandis*) es un gran problema, se han cambiado las prácticas del transporte y confinamiento de las semillas, por lugares cerrados para evitar la dispersión de la plaga y con ello se ha demostrado que puede realizarse un manejo confinado y monitoreado en toda la cadena productiva (ver decretos de ley en: <http://www.programapicudo.com.ar/>: Res. Ministerio de Economía, Producción y Empleo, Provincia del Chaco N° 340/2009). En México se procura mantener las semillas en enramados y sitios con techo pero abiertos, donde el viento y animales las dispersan. Además, el 3% de las semillas consumidas por las vacas puede continuar viable al salir del tracto digestivo por lo que se generan plantas voluntarias en las zonas de pastoreo (los dueños del ganado y terrenos también son actores desinformados).

La quinta causa es la vía “ilegal” que también existe y consiste en cultivo y uso de la “semilla peluda” sin autorización, llamada así porque se le ha separado de la fibra en la despepitadora y se utiliza para cultivo sin los permisos requeridos, registros, programas de para la contención de plagas, eliminación de plantas voluntarias y sin pagar el precio de las patentes y regalías a las empresas.

Una vez descritas las formas en las que se dispersan las semillas, se proponen estrategias que podrían resolver algunas de las problemáticas:

1) **Fomentar la industria del aceite de algodón y alimentos procesados de alta calidad para el ganado.** Ambos se obtienen en otras partes del mundo y los importamos porque no los producimos. Este proceso destruye la viabilidad de la semilla, de la cual en cambio se obtiene el aceite (la semilla de algodón es la segunda fuente de aceite vegetal del mundo (FAOSTAT 2010), y en México se desaprovecha). Una vez extraído el aceite se obtiene “el bagazo”, el cual se puede procesar como hojuelas para ganado y enriquecer con calcio y vitaminas para generar un producto con mayor valor. Esto aumentaría ganancias y empleos en la cadena productiva y productos derivados, los cuales a su vez generarían valor agregado y solventarían la inversión que se necesita para implementar esta medida de bioseguridad.

2) **Las semillas en todos los trayectos deben estar completamente contenidas.** Como se explicó en párrafos anteriores es una medida utilizada para la prevención de la expansión de las plagas como el picudo, por lo cual tendría doble ventaja. La limpieza y desinfección de los transportes es también una medida con doble beneficio. Es contradictorio que durante los trayectos previos a la siembra, la semilla sea cuidada en empaques herméticos y etiquetados, se realicen avisos a las autoridades sobre los puertos de entrada, rutas y sitios de establecimiento, mientras que después de la cosecha las semillas se pueden trasladar sin empaques, en los vehículos abiertos (construidos de reja o sin lonas que las tapen) y sin registro o aviso alguno a las autoridades, que pudieran entre otras cosas, actuar en caso de accidente, monitorear la presencia de plantas voluntarias en los caminos, etcétera. En países como Argentina existe un registro de los vehículos así como normas para su obligatoria inspección y desinfección entre fronteras, lo cual previene plagas y enfermedades, pero también que las semillas queden atrapadas en huecos de los vehículos o se dispersen, porque la presencia de plantas voluntarias disminuye la efectividad de algunas de las estrategias regionales para evitar la evolución de plagas (organismos blanco (objetivo de la modificación genética) u otros que son controlados con agroquímicos como el picudo, que actualmente es el principal problema y para el cual el transgénico no ofrece resistencia diferente a las plantas son modificación).

3) Sin duda una buena e indispensable medida es **respetar la ley principalmente en dos de sus mandatos**. Primero, **la determinación de las áreas que son centro de origen y diversidad genética** a las que se hace referencia en los artículos 86, 87 y 88 de la LBOGM (DOF 18-03-2005). Segundo, **la custodia legal de las semillas desde el inicio hasta su destrucción o transformación es responsabilidad de quien tramita el permiso para su uso**.

4) Sin embargo, el problema más complejo es adoptar sin **cuestionar la información generada en países con dinámicas biológicas, sociales, económicas y políticas diferentes al nuestro**. México es megadiverso biológica y culturalmente, alberga los procesos ecológicos y evolutivos que han dado origen a más de ciento cincuenta cultivos importantes para la humanidad y conserva costumbres de intercambio de semillas, uso tradicional de uso y mejoramiento de plantas en sistemas de agricultura de conservación. Los estudios realizados en este escenario difícilmente son extrapolables para comprender las dinámicas de otros países y viceversa.

Las observaciones ecológicas que realizamos de las especies son producto de la interacción de su genotipo con el ambiente, si el ambiente varía, entonces la observación puede variar, tal incertidumbre justifica la necesidad de que los estudios utilizados para los análisis de riesgo de especies para las cuales México sea centro de origen, domesticación o de diversidad genética sea realizados en el país, en las condiciones más cercanas posibles al sitio evaluado.

5) Es necesaria **una visión integral en dos diferentes niveles**. El **primero el contexto donde la nueva tecnología impactará**, comprendiendo lo biológico (con ecología y genética de poblaciones, del paisaje y de la conservación) integrada a disciplinas sociales y económicas, además de investigar los posibles efectos que pudieran tener los desarrollos biotecnológicos sobre la salud (ya que nuestras costumbres no han sido analizadas), por ejemplo, los cambios en la dieta en las comunidades. **El segundo nivel es entender y analizar el impacto de la tecnología completa** ya que la liberación implica al OGM y a los agroquímicos del paquete tecnológico. Legalmente son regulados por separado en casi todas las legislaciones del mundo, pero biológicamente es necesario integrar el impacto. Es una cuestión histórica que ha permeado a las investigaciones, sin embargo, no es posible separarlos. Incluso se piensa que las plantas resistentes a herbicidas tendrán menor impacto en los ecosistemas si se escapan, sin embargo, se debe analizar si el herbicida para el cual es resistente es ampliamente distribuido en el país y los usos que se le

dan, para comprender la probabilidad de selección positiva en poblaciones silvestres, y de evolución a plantas en los cultivos resistentes a todos los herbicidas disponibles (súper malezas). De igual forma, la liberación de algunos OGM implica el uso de herbicidas se ha generado información suficiente para poner en duda los efectos en la salud, y por lo tanto, es importante investigar y considerarlo. Por ejemplo, en sitios como la Península de Yucatán (donde el sistema de comunicación de aguas subterráneas es único en el mundo) y donde se libera soya GM resistente a glifosato desde 2005, se ponen en riesgo la apicultura (por el OGM *per se*), el turismo y salud de la población local por el herbicida utilizado (principalmente compuesto por glifosato).

6) En conclusión, **las investigaciones requieren tiempo y recursos**, sin embargo, tomar decisiones sin información científica suficiente puede iniciar procesos irreversibles y costosos a largo plazo, que requerirán más tiempo y recursos. **Utilizar el principio precautorio cuando la incertidumbre es grande implica responsabilidad en la toma de decisiones y no debe confundirse con temor al desarrollo.**

6.2.2 Medidas de protección

Para lograr conservar las poblaciones silvestres de algodón y su diversidad genética, es necesario implementar una estrategia con *medidas de protección*, además de las ya mencionadas medidas de bioseguridad. Actualmente las poblaciones habitan áreas fuertemente perturbadas y aunque parecen sobrevivir en condiciones de estrés extremos, el número observado de plantas adultas es bajo (336 entre 2002-2007 (ver Capítulo 4); casi 800 aumentando el esfuerzo de colecta entre 2008- 2012). En los estados más afectados observamos cambio de uso de suelo principalmente por áreas ganaderas (Veracruz y Tamaulipas) y crecimiento turístico en las zonas costeras (Jalisco, Nayarit y Península de Yucatán). En los capítulos 4 y 5 se describen los resultados sobre la dinámica ecológica y la diversidad genética encontrada en las poblaciones silvestres de *Gossypium hirsutum* utilizadas para proponer las siguientes medidas de manejo:

1. Evitar el flujo génico entre *Gossypium hirsutum* cultivado y sus parientes silvestres. Las plantas domesticadas son homogéneas genéticamente (como se demuestra en el Capítulo 4) y el intenso flujo

génico con las poblaciones silvestres puede disminuir rápidamente la diversidad genética de las últimas.

2. Determinar legalmente los centros de origen y de diversidad genética, junto con las estrategias y medidas necesarias para su protección.
3. Identificar y procurar áreas para la conservación de la especie y el nicho ecológico en cada una de las metapoblaciones. Los centros de enseñanza, instituciones gubernamentales, jardines botánicos entre otros, pueden servir como refugio de la diversidad genética de cada una de las metapoblaciones.
4. Divulgar el conocimiento adquirido entre las comunidades cercanas a las poblaciones para que sean participes de las actividades y conozcan la importancia de la conservación de sus recursos.
5. Monitorear organismos blanco del algodón GM y especies indicadoras de la salud ecosistémica en las poblaciones silvestres de *Gossypium hirsutum* para identificar las posibles consecuencias de la presencia de transgenes en las poblaciones.
6. Realizar un seguimiento demográfico de las especies del género *Gossypium* que habitan en México para monitorear el tamaño de las poblaciones, su expansión o disminución.
7. Aumentar la información que permita diferenciar las plantas silvestres y las domesticadas, incluyendo cuando se encuentran en condiciones de estrés ambiental.
8. Conocer a fondo la biología reproductiva de las poblaciones silvestres de *Gossypium hirsutum* ya que la información actual fue obtenida de plantas domesticadas y el proceso de domesticación implica en muchos de los casos modificaciones en el sistema reproductivo.
9. Monitorear las frecuencias de los transgenes presentes en las poblaciones silvestres de *Gossypium hirsutum* periódicamente.
10. Analizar integralmente la presencia de los transgenes en los ecosistemas de selvas bajas y dunas costeras, midiendo las posibles consecuencias en poblaciones y comunidades.
11. Enseñar biología evolutiva a los tomadores de decisiones. Comprender los procesos evolutivos por los que se genera y se mantiene la

diversidad, podría sensibilizar sobre la importancia de estrategias de conservación largo plazo.

6.2.3 Perspectivas generales

Las nuevas preguntas que se abrieron con esta investigación son diversas e interdisciplinarias. Es necesario ampliar la información ya publicada; realizar análisis con marcadores del núcleo para conocer el flujo de polen y caracterizarlo; profundizar en la detección de los transgenes, usar estrategias evento-específicas (inicialmente, PCR-punto final, PCR tiempo real para cuantificar el número de copias); hacer un modelo sobre la dispersión de transgenes en una dinámica metapoblacional mediante monitoreos periódicos y analizar si hay un patrón de selección.

Otras preguntas están relacionadas con la determinación de los centros de origen y diversidad genética, así como las estrategias de bioseguridad y conservación de las especies. Pero en general, la “receta maestra” que podría resultar acertada para su estudio y conservación, inicia con la revisión de la biología evolutiva, las dinámicas ecológicas, características de historia de vida de las especies cercanas y la biología reproductiva; después puede continuar con la validación de la distribución mediante revisiones de herbarios y sistemas de información geográfica, para finalmente localizar las poblaciones y probar hipótesis sobre los patrones históricos que modelaron la diversidad genética de la especie e integrar la evolución y la ecología, en el diseño de las estrategias de manejo y conservación (identificando las Unidades de manejo).

Las especies con centros de origen mesoamericanos comparten algunos aspectos de su historia, tales como los procesos continuos de domesticación, sin tiempo definido e identificable de principio a fin, e incluso continúan en la actualidad. Los caracteres distinguibles a los domesticados están integrados en la variación de los silvestres en un continuo de formas difíciles de identificar. Otra característica que comparten es que los procesos de selección se pudieron llevar a cabo en muchos sitios al mismo tiempo por lo que podríamos decir que, además de ser continuos en el tiempo, son múltiples en el espacio. Muchas preguntas en el algodón quedan abiertas en este sentido, ya que falta caracterizar la variación morfológica, anatómica y funcional de los algodones silvestres y compararlas con los cultivados para contribuir

a la determinación de los centros de diversidad pero también para reconocer a las plantas escapadas que crecen con pocos recursos y las consecuencias del efecto del flujo entre éstas y sus parientes silvestres, en las plantas y el ecosistema.

Los retos en Biología evolutiva son amplios en los sitios donde las especies han evolucionado (centro de origen de la especie o taxón) ya que habitan otras especies con las que se mantienen interacciones que forman parte de las dinámicas ecológicas entre poblaciones, comunidades y ecosistemas. Por lo tanto, para conocer el efecto de los transgenes en el ecosistema se deben realizar investigaciones a corto, mediano y largo plazo con este objetivo.

Finalmente, uno de los retos urgentes e importantes en bioseguridad es el desarrollo y validación de una metodología que permita la evaluación de los daños ambientales ocasionados por OGM, la cual debería permitir la comparación de los mismos en tiempo y espacio, ponderar entre efectos reversibles e irreversibles y estar construida sobre la teoría ecológica-evolutiva y no sobre la económica o social.

7. Glosario

Adecuación.- “El éxito de una entidad para reproducirse; por tanto, es la contribución promedio de un alelo o genotipo para las siguientes generaciones.” (Futuyma, 2009).

Centro de diversidad genética.- (ver unidad de conservación)

Deriva génica.- “Cambios al azar en las frecuencias de dos o más alelos o genotipos dentro de una población” (Futuyma,2009).

Fenotipo.- “Las propiedades bioquímicas, fisiológicas, morfológicas y de comportamiento de un organismo manifestadas a través de su vida; o una fracción de dichas propiedades, especialmente aquellas afectadas por un alelo particular o una porción del genotipo” (Futuyma, 2009).

Flujo génico.- Migración de genes a una cierta población. “El número de migrantes por generación se mide por el parámetro Nm , donde N es el tamaño efectivo de la población y m es la probabilidad de que un individuo muestreado sea un migrante” (Palsbøll, 2007).

Heterocigosis.- Es una medida de la diversidad genética (Nei, 1978). Puede interpretarse como la probabilidad de que dos alelos muestrados sean diferentes.

Consanguineidad.- “Apareamiento entre parientes que sucede más frecuentemente de lo que esperaríamos si las cruzas se realizaran de forma aleatoria en la población” (Futuyma, 2009).

Linaje o Cluster.- Grupos de poblaciones donde existe flujo génico dentro de las poblaciones contenidas en el *cluster*, pero existe flujo genético restringido entre *clusters*

(<http://web.abo.fi/fak/mnf//mate/jc/software/IntroductionToBAPSMETHODS.pdf>).

Monofilia recíproca.- La monofilia es “un taxón, árbol filogenético o árbol génico donde los miembros derivaron de un taxón ancestral común” (Futuyma, 2009). La monofilia recíproca es cuando dos o más grupos de taxones son monofiléticos con un diferente ancestro común más frecuente.

Potencial evolutivo.- Capacidad de cambio de una población o especie.

Radiación adaptativa.- “Divergencia evolutiva de miembros de un linaje filogenético hacia una variedad de diferentes formas adaptativas; usualmente los taxa difieren en el uso de recursos o hábitats y han divergido a lo largo de un intervalo corto de tiempo geológico”. (Futuyma, 2009)

Riqueza alélica.- “El número de alelos que encontramos en una población” (Kalinowski, 2004).

Tamaño efectivo de la población.- “El tamaño de una población ideal panmíctica donde observamos la misma pérdida de variación genética debido a la deriva génica que en la población observada” (Palsbøll, 2007).

Unidades de conservación.- “Unidades evolutivamente significativas, unidades de manejo o cualquier unidad geográfica que los investigadores consideren importante conservar” (Manel *et al.*, 2003).

Unidades de manejo (MU).- “Grupos de poblaciones donde el grado de conectividad es tan bajo entre dichos grupos que cada grupo debe ser monitoreado y manejado por separado” (Taylor y Dizon, 1999 en Palsbøll, 2007).

Unidades evolutivamente significativas (ESU).- “Una colección de individuos o poblaciones de una especie con una historia evolutiva y genética única donde se necesita una protección por separado para cada ESU” (Manel *et al.*, 2003).

Tasa de dispersión.- La proporción de individuos que van de una población a otra (Palsbøll, 2007).

8. Literatura citada

- Alavez, V. (2008) Evolución de microsatélites de cloroplasto en *Pinus pseudostrabus* y *Pinus montezumae*: efecto del ligamiento en la estimación de parámetros poblacionales y aplicaciones en genética de poblaciones. Tesis de licenciatura, Facultad de Ciencias, UNAM.
- Allendorf, F.W. (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*. **5**: 181-190.
- Altieri, M. (2004) Linking ecologists and traditional farmers in the search for sustainable agriculture. *Frontiers in Ecology and the Environment*. **2**: 35-42.
- Álvarez, I., Cronn, R. y Wendel, J.F. (2005) Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. *Plant Systematics and Evolution*. **252** (3): 199-214
- Amos, B. y Balmford, A. (2001) When does conservation genetics matter? *Heredity*. **87**: 257-265.
- Avise, J.C., Arnold, C.J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. y Saunders, N.C. (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*. **18**: 489-522.
- Baer, C.F. (1998) Species-wide population structure in a southeastern US freshwater fish, *Heterandria formosa*: gene flow and biogeography. *Evolution*. **52**: 183-193.
- Basra, A.S. y Malik, C.P. (1984) Development of the cotton fiber. *International Review of Cytology*. **89**: 65-113.
- Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinel, T. y Lukowitz, W. (2009) Paternal control of embryonic patterning in *Arabidopsis thaliana*. *Science*. **232**: 1485-1488.
- Bensch, S., Hasselquist, D. y von Schantz, T. (1994) Genetic similarity between parents predicts hatching failure: Nonincestuous inbreeding in the great reed warbler? *Evolution*. **48**: 317-326.
- Booth, J.E. (1968) *Principles of Textile Testing*. 3rd ed. Butterworths, London. 583.

- Brown, M.S. y Menzel, M.Y. (1950) New trispecies hybrids in cotton. *Journal of Heredity*. **41**: 291-295.
- Brown, M.S. y Menzel, M.Y. (1952) The cytology and crossing behavior of *Gossypium gossypoides*. *Bulletin of the Torrey Botanical Club*. **79**: 110-125.
- Brubaker, C.L. y Wendel, J.F. (1994) Reevaluation the origin of domesticated cotton (*Gossypium hirsutum*: Malvaceae) using nuclear restriction fragment length polymorfisms (RFLPs). *American Journal of Botany*. **81**: 1309-1326.
- Brush, S.B. (2001) Genetically Modified Organisms in Peasant Farming: Social Impact and Equity. *Indiana Journal of Global Legal Studies*. **9**: 1,8.
- Castellanos-Morales, G. (2006) El efecto de la fragmentación en la variabilidad genética de los perros llaneros (*Cynomys ludovicans*) en Janos-Nuevo Casas Grandes, Chihuahua, México. Tesis de maestría, UNAM, Instituto de Ecología.
- CERA 2011, <http://www.cera-gmc.org>.
- Cerritos, R., Wegier, A. y Alavez, V. (2012) *Toward the Development of Novel Long-Term Pest Control Strategies Based on Insect Ecological and Evolutionary Dynamics*. En: *Integrated Pest Management and Pest Control*. ISBN 978-953-307-926-4. InTech, Open Access publisher.
- Chaudhary, B., Flagel, L., Stupar, R. M., Udall, J. A. y Verma, N. (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics*. **182**: 503–517.
- Corander, J. y Tang, J. (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences*. **205**: 19-31.
- Corander, J., Marttinen, P., Sirén, J. y Tang, J. (2008) Enhanced bayesian modelling in BAPS software for learning genetic structures of populations, *BMC Bioinformatics*. **9**: 539.
- Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M. y Wayne R.K. (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*. **15**: 290-295.
- Cronn, R.C., Small, R.L., Wendel, J.F. (1999) Duplicated genes evolve independently after polyploidy formation in cotton. *Proceedings of the National Academy of Science USA*. **96**: 14406–14411.

- Cronn, R.C., Small, R.L., Haselkorn, T. y Wendel, J.F. (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany*. **89**: 707–72
- Cronn, R.C., Zhao, X., Paterson, A.H. y Wendel, J.F. (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *Journal of Molecular Evolution*. **42**: 685-705.
- DeJooode, D.R. y Wendel, J.F. (1992) Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum*. *American Journal of Botany*. **79**: 1311-1319.
- DeLanghe, E.A.L. (1986) *Lint development*. En: J.R. Mauney and J.M. Stewart (eds.). Cotton Physiology. The Cotton Foundation, Memphis, Tenn.
- Dyer, R.J., Nason, J.D. y Garrick, R.C. (2010) Landscape modelling of gene flow: Improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*. **19**: 3746–3759.
- Edmands, S. (2007) Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*. **16**: 463-475.
- Ellstrand, N.C. (2003) *Dangerous liaisons? When cultivated plants mate with their wild relatives*. Baltimore: Johns Hopkins University.
- Ellstrand, N.C., Garner, L.C., Hegde, S., Guardagnuolo, R. y Blancas, L. (2007) Spontaneous hybridization between maize and teosinte. *Journal of Heredity*. **98**: 183–187.
- Ellstrand, N.C., Prentice, H.C., Hancock, J.F. (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology & Systematics*. **30**: 539-563
- Endrizzi, J.E., Turcotte, E.L. y Kohel, R.J. (1985) Genetics, cytology, and evolution of *Gossypium*. *Adv. Genet.* **23**: 271-375.
- Engels, J.M.M., Ebert, A.W., Thormann, I. y Vicente, M. (2006) Centres of Crop Diversity and/or Origin, Genetically Modified Crops and Implications for Plant Genetic Resources Conservation. *Genetic Resources and Crop Evolution*. **53**(8): 1675-1688

- England, P.R, Osler G.H.R., Woodworth, L.M., Montgomery, M.E., Briscoe, D.A. y Frankham R. (2003) Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conservation Genetics*. **4**: 595-604.
- Esquinas-Alcázar, J. (2005) Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nature Reviews Genetics*. **6**: 946-953.
- Excoffier, L., Laval, G. y Schneider (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*. **1**: 47-50.
- Excoffier, L., Smouse, P.E. y Quatro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. **131**: 479-491.
- FAOSTAT 2010 at: <http://faostat.fao.org/>
- Frankham R. (1995) Conservation Genetics. *Annual Review of Genetics* Vol. **29**: 305-327
- Frankham, R. (2002) *Introduction to conservation genetics*. Cambridge University Press. Reino Unido.
- Frankham, R., Ballou, J.D. y Briscoe, D.A. (2002) *Introduction to conservation genetics*. Cambridge University Press. Reino Unido.
- Fraser, D.J. y Bernatchez, L. (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*. **10**: 2741-2752.
- Fryxell, P.A. (1965) Stages in the evolution of *Gossypium*. *Adv. Frontiers Pl. Sci*. **10**: 31-56.
- Fryxell, P.A. (1968) A redefinition of the tribe Gossypieae. *Bot. Gaz*. **129**: 296-308.
- Fryxell, P.A. (1971) Phenetic analysis and the phylogeny of the diploid species of *Gossypium* L. (Malvaceae). *Evolution*. **25**: 554-562.
- Fryxell, P.A. (1979) *The natural history of the cotton tribe*. Texas A&M University Press, College Station, Texas.
- Fryxell, P.A. (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheede*. **2**: 108-165.

- Fryxell, P.A., Craven, L.A. y Stewart, J.McD. (1992) A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. *Syst. Bot.* **17**: 91-114.
- Futuyma, D. (2009) *Evolution*. Sinauer Associates, Inc. Estados Unidos de América.
- Geist, J. y Kuehn, R. (2005) Genetic diversity and differentiation of central European freshwater pearl mussel (*Margaritifera margaritifera* L.) populations: implications for conservation and management. *Molecular Ecology*. **14**: 425-439.
- Gepts, P. (2004) Crop domestication as a long-term selection experiment. *Plant Breed Rev.* **24** (Part 2): 1-4.
- Gerstel, D.U. (1953) Chromosomal translocations in interspecific hybrids of the genus *Gossypium*. *Evolution*. **7**: 234-244.
- Gillespie, J.H. (2004) *Population genetics: a concise guide*. John Hopkins University. Estados Unidos de América.
- Gossett, D.R., Millhollon, E.P. y Lucas, M.C. (1994) Antioxidant response to NaCl stress in salt-tolerant and salt-sensitive cultivars of cotton. *Plant Cell Rep.* **13**: 498-503
- Gossett, D.R., Lucas, M.C., Millhollon, E.P., Caldwell, W.D. y Barclay, A. (1992) Antioxidant status in salt stressed cotton. *Proc. Beltwide Cotton Physiol. Conf. Memphis*. **3**: 1036-1039.
- Haldane, J.B.S. (1957) The cost of natural selection. *Journal of Genetics*. **55**: 511-524.
- Harlan, J.R. (1971) Agricultural origins: centers and noncenters. *Science*. **174**: 468-474.
- Hastings, A. (1993) Complex interactions between dispersal and dynamics: Lessons from coupled logistic equations. *Ecology*. **74**: 1362-1372.
- Hedrick, P.W. (2001) Conservation genetics: where are we now? *Trends in Ecology and Evolution*. **16**: 629-636.
- Hedrick, P.W. (2004) *Genetics of populations*. Jones & Bartlett Publishers, Estados Unidos de América.
- Hein, J., Schierup, M.H. y Wiuf, C. (2005) *Gene genealogies, variation and evolution: a primer in coalescent theory*, New York, Oxford University Press.

- Hey, J. y Nielsen, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*. **104**: 2785–2790.
- Hey, J., Waples, R.S., Arnold, M.L., Butlin, R.K. y Harrison, R.G. (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology and Evolution*. **18**: 597-603.
- Hovav, R., Chaudhary, B., Udall, J.A., Fligel, L. y Wendel, J.F. (2008) Parallel domestication, convergent evolution and duplicated gene recruitment in allopolyploid cotton. *Genetics*. **179**(3):1725 – 1733
- <http://cmpg.unibe.ch/software/arlequin3/arlequin31.pdf> (Revisado el 28 de agosto).
- <http://web.abo.fi/fak/mnf//mate/jc/software/IntroductionToBAPSMETHODS.pdf> (Revisado el 30 de julio a las 11:00 a.m.)
- <http://www.salud.gob.mx/unidades/cdi/nom/compi/ley180305.html> (Revisado el 16 de julio)
- Hutchinson, J.B., Silow, R.A. y Stephens, S.G. (1947) *The Evolution of Gossypium*. Oxford Univ. Press. London.
- James, J.W. (1971) The founder effect and response to artificial selection. *Genetical Research*. **12**: 249-266.
- Kalinowski, S.T. (2004) Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics*. **5**: 539-543.
- Kalinowski, S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*. **5**: 187-189.
- Keller, L.F. y Waller, D.M. (2002) Inbreeding effects in wild populations. *Trends in Ecology and Evolution*. **17**: 230-241
- Kimura, M. y Crow, J. (1964) The number of alleles that can be maintained in a finite population. *Genetics*. **49**: 725-738.
- Kosmidou-Dimitropoulou, K. (1986) *Hormonal influences in fiber development*. En: J.R. Mauney and J.M. Stewart (eds.) Cotton Physiology. The Cotton Foundation, Memphis. Estados Unidos de América.
- Kuhner, M. (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*. **22**: 768-770.

- LaDuke, J.C. y Doebley, J. (1995) A chloroplast DNA based phylogeny of the Malvaceae. *Syst. Bot.* **20**: 259-271.
- Liu, Q., C.L. Brubaker, A.G. Green, D.R. Marshall, P.J. Sharp y S.P. Singh. (2001). Evolution of the *fad2-1* fatty acid desaturase 5' UTR intron and the molecular systematic of *Gossypium* (Malvaceae). *American Journal of Botany.* **88**: 92–102.
- Lu, B.R. y Snow, A.A. (2005) Gene flow from genetically modified rice and its environmental consequences. *BioScience.* **55**: 669-678.
- Manel, S., Schwartz, M.K., Luikart, G. y Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution.* **18**: 189-197.
- Mastretta-Yanes, A., Wegier, A., Vázquez-Lobo, A. y Piñero, D. (2012) Distinctiveness and rarity in a subtropical conifer and its conservation. *Conservation genetics.* Doi: 10.1007/s10592-011-0277-y
- Menzel, M.Y. y Brown, M.S. (1954) The significance of multivalent formation in three-species *Gossypium* hybrids. *Genetics.* **39**: 546-557.
- Meyer, H. (2011) Systemic risks of genetically modified crops: the need for new approaches to risk assessment. *Environmental Sciences Europe.* **23**:7.
- Mitton, J.B. (1997) *Selection in Natural Populations.* Oxford University Press. Reino Unido.
- Moritz, C. (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology and Evolution.* **9**: 373-375.
- Moritz, C. (1999) Conservation units and translocations: strategies for conserving evolutionary processes. *Heredity.* **130**: 217-228.
- Moritz, C. (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology.* **51**: 238-254.
- Nei, M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics.* **89**: 583-590.
- Nei, M. (1987) *Molecular Evolutionary Genetics.* New York, Columbia University. Estados Unidos de América.

- Ohta, T. y Kimura, M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*. **22**: 201-204.
- Palmer, J.D. (1992) *Mitochondrial DNA in plant systematics: applications and limitations*. En: Soltis, P.S, Soltis D.E. y Doyle J.J. Molecular systematic of plants. Chapman y Hall. Estados Unidos de América.
- Palsbøll, P.J., Bérubé, M. y Allendorf, F.W. (2007) Identification of management units using population genetic data. *Trends in Ecology and Evolution*. **22**: 11-16.
- Pan, X. (2013) Determining Gene Flow in Transgenic Cotton. En: Zhang, B (ed.) Transgenic Cotton, Methods and Protocols, Serie: Methods in Molecular Biology. Volumen 958.
- Pearse, D.E. y Crandall, K.A. (2004) Beyond F_{st} : Analysis of population genetic data for conservation. *Conservation genetics*. **5**: 582-602.
- Petit, R.J., el Mousadik, A. y Pons O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation biology*. **12**: 844-855.
- Phillips, L.L y Strickland, M.A. (1966) The cytology of a hybrid between *Gossypium hirsutum* and *G. longicalyx*. *Can. J. Genet. Cytol.* **8**:91-95.
- Pistorius, R.J. (1997), *Scientists, Plants and Politics: A History of the Plant Genetic Resources Movement*. Rome: International Plant Genetic Resources Institute, IPGRI.
- Pritchard, J.K., Stephens, M. y Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*. **155**: 945-959.
- Rapp, R.A., Haigler, C.H., Flagel, L., Hovav, R.H., Udall, J.A. y Wendel, J.F. (2010) Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. *BMC Biol.* **8**: 139–153.
- Reddy, O.U.K., Pepper, A.E., Abdurakhmonov, I., Saha, S., Jenkins, J.N., Brooks, T., Bolek, Y. y El-Zik, K.M. (2001) New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. *The journal of cotton science*. **5**: 103-113.
- Reed, D.H. y Frankham, R. (2003) Correlation between fitness and genetic diversity. *Conservation Biology*. **17**: 230-237.

- Rieseberg, L.H. y Burke, J.M. (2008) Molecular evidence and the origin of the domesticated sunflower. *Proc Natl Acad Sci USA*. 105:E46.
- Ritland K. (1990) Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution*. **44**: 1230-1241.
- Rocha, M. y Gasca, J. (2007) *Ecología molecular de la conservación*. En: Eguiarte, L., Souza, V. y Aguirre, X. (eds.). *Ecología Molecular*. Instituto Nacional de Ecología. México, D. F.
- Ryder, O.A. (1986) Species conservation and systematic: the dilemma of subspecies. *Trends in Ecology and Evolution*. **1**: 9-10.
- Ryser, U. (1985) Cell wall biosynthesis in differentiating cotton fibers. *Eur. J. Cell Biol.* **39**: 236-256.
- Santure, A.W. y Spencer, H.G. (2006) Influence of mom and dad: quantitative genetic models for maternal effects and genomic imprinting. *Genetics*. **173**: 2297-2316.
- Saunders, J.H. (1961) *The Wild Species of Gossypium and Their Evolutionary History*. Oxford University Press, London. Reino Unido.
- Seelanan, T., Schnabel, A. y Wendel, J.F. (1997) Congruence and consensus in the cotton tribe. *Syst. Bot.* **22**: 259-290.
- SEMARNAT (2012) *Recursos fitogenéticos. Compendio de estadísticas ambientales 2012*. http://app1.semarnat.gob.mx/dgeia/Compendio_2012/dgeiawf-semarnat.gob.mx_8080/ibi_apps/WFServletb14e.html (Consultado 15 enero 2013).
- Senchina, D.S., Álvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A. y Wendel, J.F. (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution*. **20**: 633-643.
- Slatkin, M. E. (1987) Gene flow and the geographic structure of natural populations. *Science*. **236**: 787-792.
- Slatkin, M. y Hudson, R.R. (1991) Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics*. **129**: 555-562.
- Small, R. L. y Wendel, J. F. (2000) Phylogeny, duplication, and intraspecific variation of Adh sequences in New World diploid cottons (*Gossypium* L.,

- Malvaceae). *Molecular Phylogenetics and Evolution*. **16**: 73-84.
- Sokal, R.R., Oden, N.L. y Wilson, C. (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature*. **351**: 143–145.
 - Sork, V.L. y Waits, L. (2010) Contributions of landscape genetics – approaches, insights, and future potential. *Molecular Ecology* **19**: 3489–3495.
 - Spielman, R.S. y Smouse, P.E. (1976) Multivariate classification of human populations. I. Allocation of Yanomama Indians to villages. *American Journal of Human Genetics*. **184**: 637–644.
 - Storfer, A., Murphy, M.A., Evans, J.S. *et al.* (2007) Putting the “landscape” in landscape genetics. *Heredity*. **98**: 128–142.
 - Szpiech, Z.A., Jakobsson, M. y Rosenberg, N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*. **24**: 2498-2504.
 - Taliercio, E. y Haigler, C.H. (2011) The effect of calcium on early fiber elongation in cotton ovule culture. *Journal of Cotton Science*. **15**: 154-161.
 - Taylor, P.D., Fahrig, L., Henein, K. y Merriam, G. (1993) Connectivity is a vital element of landscape structure. *Oikos*. **68**: 571–573.
 - Taylor, B.L. y Dizon, A.E. (1999) First policy then science: why a management unit based solely on genetic criteria cannot work. *Molecular Ecology*. **8**: S11-S16.
 - Ulloa, M., Stewart, J.M., Garcia-C, E.A., Goday, A. S., Gaytan-M, A. y Acosta, N.S. (2006). Cotton Genetic Resources in the Western States of Mexico: *in situ* Conservation Status and Germplasm Collection for *ex situ* Preservation. *Genetic Resources and Crop Evolution*. **53**: 4: 653-668.
 - Valiček, P. (1978) Wild and cultivated cottons. *Cot. Fib.Trop.* **33**: 363-387.
 - Vavilov, N.I. (1992) *Origin and geography of cultivated plants*. Cambridge University Press, Estados Unidos de América.
 - Vogler, A.P. y Desalle R. (1994) Diagnosing units of conservation management. *Conservation Biology*. **8**: 354-363.
 - Vollesen, K. (1987) *The native species of Gossypium (Malvaceae) in Africa, Arabia and Pakistan*. Kew. Reino Unido.

- Wade, M.J. (1998) *The evolutionary genetics of maternal effects*. En: Mousseau, T.A. y Fox C.W. (eds.) *Maternal effects as adaptations*. Oxford University Press. Reino Unido.
- Walter, R. y Epperson, B.K. (2001) Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. *Molecular Ecology*. **10**: 103-111.
- Wang, S., Hard, J.J. y Utter, F. (2002) Salmonid inbreeding: a review. *Reviews in Fish Biology and Fisheries*. **11**: 301-319.
- Watt, G. (1907) *The Wild and Cultivated Cotton Plants of the World*. Longmans, Green and Co., London. Reino Unido.
- Watterson, G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*. **7**: 256-276.
- Wegier, A. , Piñeyro-Nelson, A., Alarcón, J., Gálvez-Mariscal, A., Alvarez-Buylla, E. y Piñero, D. (2011) Recent long-distance transgene flow conforms to historical patterns of gene flow in wild cotton (*Gossypium hirsutum*) at its center of origin. *Molecular Ecology*. **20** (19):4182-94 Doi: 10.1111/j.1365-294X.2011.05258.x
- Weir y Cockerham (1984) Estimating F-statistics for the analysis of population structure. *Evolution*. **38**: 1358.
- Weising, K. y Gardner R.C. (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*. **42**: 9-19.
- Weiss, A. 2012. Germination and dormancy of crop-wild sunflower hybrid cross types. Tesis de Licenciatura. The Ohio State University. School of Environment and Natural Resources.
- Wendel, J.F. y Cronn, R.C. (2003) Polyploidy and the evolutionary history of cotton. *Adv Agron*. **78**: 139–186.
- Wendel, J. F. y Percival, A. E. (1990) Molecular divergence in the Galapagos Island-Baja California species pair, *Gossypium klotzschianum* and *G. davidsonii* (Malvaceae). *Plant Systematics and Evolution*. **171**: 99-115.
- Wendel, J. F. y Cronn, R. C. (2003) Polyploidy and the evolutionary history of cotton. *Advances in Agronomy*. **78**: 139-186.
- Wendel, J.F. y Percy, R.G. (1990) Allozyme diversity and introgression in the

Galapagos endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochemical Systematics and Ecology*. **18**: 517-528.

- Wendel, J.F. y Albert, V.A. (1992). Phylogenetics of the cotton genus (*Gossypium* L.): Character state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany*. **17**: 115-143.
- Wendel, J.F., Schnabel, A. y Seelanan, T. (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences USA*. **92**: 280-284.
- Wendel, J. F., Brubaker, C. L. y Seelanan, T. (2010) *The origin and evolution of Gossypium*. En: Stewart, J.M., Oosterhuis, D., Heitholt, J.J. y Mauney, J.R. (Eds.). Physiology of cotton. Springer, Netherlands.
- Wendel, J. F., Brubaker, C. L., Alvarez, I., Cronn, R. C. y Stewart, J. McD. (2009) Evolution and natural history of the cotton genus. En: Paterson, A. (ed.) Genomics of cotton. Plant Genetics and Genomics. Crops and Models 3. Springer, New York. Estados Unidos de América.
- Wendel, J.F., Rowley, R. y Stewart, J. (1994) Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Systematics and Evolution*. **192**: 49-59.
- Wendel, J.F. (1989) New World tetraploid cottons contain Old World cytoplasm. *Proceedings of the National Academy of Science USA*. **86**: 4132-4136.
- Wendel, J.F. y Cronn, R.C. (2003) Polyploidy and the evolutionary history of cotton. *Advances in agronomy*. **78**: 139-186.
- Wilson, G.A. y Rannala, B. (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*. **163**: 1177-1191.
- Wright, S. (1931) Evolution in mendelian populations. *Genetics*. **16**: 97-159.
- Xu, H. y Fu, Y.X. (2004) Estimating effective population size or mutation rate with microsatellites. *Genetics*. **166**: 555-563.
- Zhang B (2013) Transgenic Cotton: From Biotransformation Methods to Agricultural Application. En: Zhang, B. (ed.) Methods and Protocols. Serie Methods in Molecular Biology. Vol. 958.