

Informe final* del Proyecto KE004
Diversidad genética de las especies de cucurbita en México e hibridación entre plantas genéticamente modificadas y especies silvestres de cucurbita

Responsable: Dr. Rafael Lira Saade
Institución: Universidad Nacional Autónoma de México
Facultad de Estudios Superiores Iztacala
División de Investigación y Posgrado
Unidad de Biotecnología y Prototipos
Dirección: Av. de los Barrios # 1, Los Reyes Iztacala, Tlalnepantla, Mex, 54090, México.
Correo electrónico: rlira@servidor.unam.mx
Teléfono/Fax: Tel: 5623 1333 ext. 39780 Fax: 5623 1193
Fecha de inicio: Agosto 15, 2013.
Fecha de término: Febrero 7, 2017.
Principales resultados: Base de datos, informe final.
Forma de citar el informe final y otros resultados:** Lira, S. R., Eguarte, F. L., y Montes, H. S., 2018. Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de *Cucurbita*. Universidad Nacional Autónoma de México. Facultad de Estudios Superiores Iztacala. **Informe Final SNIB-CONABIO. Proyecto No. KE004.** Ciudad de México.

Resumen:

México es el centro de origen y diversidad del género *Cucurbita*, ya que dentro de sus límites geográficos prosperan 15 de los 20 taxa que lo conforman. Los taxa mexicanos de *Cucurbita* incluyen plantas domesticadas (*C. argyrosperma* ssp. *argyrosperma*, *C. pepo* ssp. *pepo*, *C. ficifolia* y *C. moschata*), así como dos de sus ancestros silvestres (*C. pepo* ssp. *fraterna* y *C. argyrosperma* ssp. *sororia*). La revisión de todos los estudios realizados hasta la fecha con marcadores moleculares para las especies del género que crecen en México ha revelado que falta incluir a varias de las especies silvestres nativas de México, además de que no existe un verdadero registro de la diversidad genética de todas las especies. Otro aspecto que también se puede derivar de todos esos trabajos es que el genoma de las diferentes especies de *Cucurbita* es relativamente pequeño, de unos 337.41 Mpb/C a 422.5 Mpb/C, apenas alrededor del doble que en *Arabidopsis thaliana* (115 a 211 Mbp/C) y mucho menor que el maíz (2, 500 Mbp/C), y hay estudios detallados de su transcrito y mapas con marcadores moleculares relativamente densos (Gong et al., 2008; Gong y Lelley, 2008; Blanca et al. 2011) por lo que su estudio no debe ser tan complejo. Por otra parte, otros estudios muestran que las características reproductivas de los taxa de *Cucurbita* y las posibilidades de hibridar entre ellos, implican problemas en el contexto de la liberación de Organismos Genéticamente Modificados (OGMs), por lo que en México la introducción de éstos últimos debe restringirse desde el nivel experimental. En este contexto, se deben realizar experimentos que sustenten lo anterior, donde se evalúan los posibles efectos de híbridos silvestre-OGMs en diferentes componentes de la adecuación en plantas silvestres de *Cucurbita*. En estos experimentos, además, será importante estudiar los componentes masculinos y femeninos de la fecundidad de dichos híbridos y especialmente determinar el movimiento de su polen. Evidentemente, se necesitarán hacer estudios detallados y análisis de paternidad, utilizando la técnica de microsatélites. Considerando lo anterior, los objetivos de este proyecto son 1) Analizar la diversidad genética de las especies de *Cucurbita* que crecen en México y 2) Evaluar las posibles consecuencias en diversos aspectos de la biología reproductiva y adecuación, por la hibridación de tres taxa silvestres de *Cucurbita* (*C. pepo* ssp. *fraterna*, *C. argyrosperma* ssp. *sororia* y *C. okeechobeensis* ssp. *martinezii*) y dos cultivares genéticamente modificados (CGMs) de la planta domesticada *Cucurbita pepo* ssp. *pepo*.

-
- * El presente documento no necesariamente contiene los principales resultados del proyecto correspondiente o la descripción de los mismos. Los proyectos apoyados por la CONABIO así como información adicional sobre ellos, pueden consultarse en www.conabio.gob.mx
 - ** El usuario tiene la obligación, de conformidad con el artículo 57 de la LFDA, de citar a los autores de obras individuales, así como a los compiladores. De manera que deberán citarse todos los responsables de los proyectos, que proveyeron datos, así como a la CONABIO como depositaria, compiladora y proveedora de la información. En su caso, el usuario deberá obtener del proveedor la información complementaria sobre la autoría específica de los datos.

Ciudad de México a 22 de noviembre de 2016

**INFORME FINAL
PRIMERA ETAPA**

Proyecto CONABIO KE004

DIVERSIDAD GENÉTICA DE LAS ESPECIES DE *Cucurbita* EN MÉXICO E HIBRIDACIÓN ENTRE PLANTAS GENÉTICAMENTE MODIFICADAS Y ESPECIES SILVESTRES DE *Cucurbita*

Responsables:

Rafael Lira Saade

Profesor Titular C TC, Facultad de Estudios Superiores Iztacala, UNAM. Avenida de Los Barrios, #1 Colonia Los Reyes Iztacala, 54090, Tlalnepantla, Edo. de México.

Correo electrónico: rlira@unam.mx

Luis E. Eguiarte Fruns

Investigador Titular C TC, Laboratorio de Evolución Molecular, Instituto de Ecología, UNAM. Cd. Universitaria, Anexo al Jardín Botánico. CP 04510. México, D.F.

Correo electrónico: fruns@unam.mx

Salvador Montes Hernández

Investigador del Campo Experimental Bajío, INIFAP. Km. 6.5 Carr. Celaya-San Miguel de Allende. 38110, Celaya, Gto.

Participantes

Dra. Gabriela Castellanos Morales
Postdoctoral, UBIPRO, FES-Iztacala, UNAM.

Dra. Alejandra Vázquez Lobo
Investigador asociado al proyecto,
Profesor Investigador Asociado "C" de TC, Universidad Autónoma del Estado de Morelos
Genómica de *C. argyrosperma* spp. *argyrosperma* y análisis genéticos de *C. ficifolia*.

M. en C. Guillermo Sánchez de la Vega
Estudiante de doctorado, Instituto de Ecología, UNAM
Análisis de la diversidad genética de *Cucurbita argyrosperma* Huber (Cucurbitaceae) en México.

Biól. Helena Hernández Rosales
Estudiante de doctorado, Instituto de Ecología, UNAM

Genética de poblaciones y filogeografía de *Cucurbita moschata* Duchesne ex Lam. Duchesne ex Poir. (Cucurbitaceae) en México.

Biól. Josué Barrera Redondo
Estudiante de Doctorado, Instituto de Ecología, UNAM
Genómica comparada de la domesticación en *Cucurbita argyrosperma*.

Biól. Leslie Mariel Paredes Torres
Estudiante de Licenciatura, UBIPRO, FES-Iztacala
Filogenia con marcadores de cloroplasto de *Cucurbita*. Concluido.

P. de B. Gustavo García Jaramillo
Estudiante de Licenciatura, UBIPRO, FES-Iztacala
Etnobotánica de las especies de *Cucurbita* en el Valle de Tehuacán-Cuicatlán. En revisión.

P. de B. Paulina Hernández Vargas
Estudiante de Licenciatura, UBIPRO, FES-Iztacala.
Documentación sobre la existencia de *Cucurbita maxima* Duchesne ex Lam. en México. En revisión.

P de B. Karen Ruíz Mondragón
Estudiante de Licenciatura, Facultad de Ciencias, UNAM.
Filogeografía de *Cucurbita pepo* en México.

Biól. Thalita Estrella Legaspi-Serrano
Estudiante de Maestría, UBIPRO, FES-Iztacala.
Morfología de *Cucurbita*.

Dra. Erika Aguirre Planter
Técnico Académico, Instituto de Ecología, UNAM.

Dra. Laura Espinosa Asuar
Técnico Académico, Instituto de Ecología, UNAM.

RESUMEN DEL PROYECTO ORIGINAL

México es el centro de origen y diversidad del género *Cucurbita*, ya que dentro de sus límites geográficos prosperan 15 de los 20 taxa que lo conforman. Los taxa mexicanos de *Cucurbita* incluyen plantas domesticadas (*C. argyrosperma* ssp. *argyrosperma*, *C. pepo* ssp. *pepo*, *C. ficifolia* y *C. moschata*), así como dos de sus ancestros silvestres (*C. pepo* ssp. *fraterna* y *C. argyrosperma* ssp. *sororia*). La revisión de todos los estudios realizados hasta la fecha con marcadores moleculares para las especies del género que crecen en México ha revelado que falta incluir a varias de las especies silvestres nativas de México, además de que no existe un verdadero registro de la diversidad genética de todas las especies. Otro aspecto que también se

puede derivar de todos esos trabajos es que el genoma de las diferentes especies de *Cucurbita* es relativamente pequeño, de unos 337.41 Mpb/C a 422.5 Mpb/C, apenas alrededor del doble que en *Arabidopsis thaliana* (115 a 211 Mbp/C) y mucho menor que el maíz (2, 500 Mbp/C), y hay estudios detallados de su transcriptoma y mapas con marcadores moleculares relativamente densos (Gong et al., 2008; Gong y Lelley, 2008; Blanca et al. 2011) por lo que su estudio no debe ser tan complejo. Por otra parte, otros estudios muestran que las características reproductivas de los taxa de *Cucurbita* y las posibilidades de hibridar entre ellos, implican problemas en el contexto de la liberación de Organismos Genéticamente Modificados (OGMs), por lo que en México la introducción de éstos últimos debe restringirse desde el nivel experimental. En este contexto, se deben realizar experimentos que sustenten lo anterior, donde se evalúen los posibles efectos de híbridos silvestre-OGMs en diferentes componentes de la adecuación en plantas silvestres de *Cucurbita*. En estos experimentos, además, será importante estudiar los componentes masculinos y femeninos de la fecundidad de dichos híbridos y especialmente determinar el movimiento de su polen. Evidentemente, se necesitarán hacer estudios detallados y análisis de paternidad, utilizando la técnica de microsatélites. Considerando lo anterior, los objetivos de este proyecto son 1) Analizar la diversidad genética de las especies de *Cucurbita* que crecen en México y 2) Evaluar las posibles consecuencias en diversos aspectos de la biología reproductiva y adecuación, por la hibridación de tres taxa silvestres de *Cucurbita* (*C. pepo* ssp. *fraterna*, *C. argyrosperma* ssp. *sororia* y *C. okechobeensis* ssp. *martinezii*) y dos cultivares genéticamente modificados (CGMs) de la planta domesticada *Cucurbita pepo* ssp. *pepo*.

Introducción del proyecto original.

Cucurbita se reconoce actualmente como un género estrictamente americano, pues todos sus miembros crecen espontáneamente o fueron domesticados en América. Incluye 15 especies o agrupaciones taxonómicas que en total comprenden a 20 taxa, de los cuales 15 crecen espontáneamente o se cultivan en México (Lira et al., 1995; Sanjur et al., 2002). En cinco de estas agrupaciones se ubican los taxa domesticados (*C. argyrosperma* Huber ssp. *argyrosperma*, *C. ficifolia* Bouché, *C. maxima* Duch. ex Lam. ssp. *máxima*, *C. moschata* (Duch. ex Lam.) Duch. ex Poir. y *C. pepo* L. ssp. *pepo*), los cuales en muchos casos incluyen numerosas razas o variedades locales, cultivares comerciales comestibles u ornamentales que en otros trabajos han sido reconocidos en diferentes categorías taxonómicas como subespecies, variedades o formas (Lira et al., 1995; Sanjur et al., 2002). En el Nuevo Mundo, los únicos géneros que pueden ser comparables con *Cucurbita* por el número de especies comestibles domesticadas son *Capsicum* (con 3-5 especies) y *Solanum* (con 7 especies) (Nee, 1990).

En todos los estudios con marcadores moleculares de *Cucurbita* publicados hasta la fecha ha hecho falta incluir a varias especies silvestres nativas de México, por lo que consideramos que es urgente retomarlos incluyendo a todas las especies silvestres y domesticadas del género. Las investigaciones hasta ahora realizadas para entender el genoma de las Cucurbitaceae y en particular los de *Cucurbita* son realmente escasas, pero indican que se trata de genomas relativamente pequeños, por lo que estudiarlos más ampliamente es una tarea relativamente sencilla que debe abordarse.

Las características reproductivas de los taxa de *Cucurbita* (polinización cruzada mediante abejas diurnas de los géneros *Peponapis* y *Xenoglossa*) y las posibilidades de formar híbridos de distinta fertilidad entre plantas domesticadas de *C. pepo* ssp. *pepo*, *C. argyrosperma* ssp. *argyrosperma* y *C. moschata* y entre éstas con al menos cuatro de los taxa silvestres anuales (*C. pepo* ssp. *fraterna*, *C. argyrosperma* ssp. *sororia*, *C. okechobeensis* ssp. *martinezii* y *C. lundeliana*), implicarían problemas a futuro en el contexto de la liberación de Organismos Genéticamente Modificados, por lo que se necesitan estudios detallados antes de considerar la introducción de éstos al ambiente. Mientras tanto, no debe permitirse de manera tan abierta su liberación ni siquiera a nivel de experimentación en nuestro país.

OBJETIVOS GENERALES

- 1) Analizar la diversidad genética de las especies silvestres y domesticadas de *Cucurbita* que crecen en México.
- 2) Evaluar las posibles consecuencias en diversos aspectos de la biología reproductiva y adecuación, por la hibridación de tres taxa silvestres de *Cucurbita* (*C. pepo* ssp. *fraterna*, *C. argyrosperma* ssp. *sororia* y *C. okechobeensis* ssp. *martinezii*) y dos cultivares genéticamente modificados (CGMs) de la planta domesticada *Cucurbita pepo* ssp. *pepo*.

OBJETIVOS ESPECÍFICOS ORIGINALES DEL PROYECTO

Estudios de diversidad genética

- 1) Organizar la base de datos del banco nacional de germoplasma del INIFAP-Celaya y la base de datos de CONABIO, contemplando la correcta determinación y georeferenciación de las accesiones.

- 2) Utilizar la información del punto anterior para mapear la distribución potencial de las colecciones y detectar sitios en los que se requiera hacer trabajo de recolección.
- 3) Llevar a cabo trabajo de recolección en los sitios detectados por el punto anterior.
- 4) Determinar, mediante el uso de marcadores moleculares y utilizando los materiales disponibles (colecciones de germoplasma y ejemplares de herbario) y en su caso los recolectados, la distribución de la diversidad genética de las especies del género que crecen en México.
- 5) Mapear la distribución de la diversidad genética documentada.

Estudios de hibridación

- 1) Conocer el aporte de la paternidad de los padres CGMs y el nivel o grado de hibridación entre las poblaciones silvestres y cultivadas. Estudiar la biología floral, reproductiva, polinización y éxito reproductivo de todos los taxa de *Cucurbita* utilizados y los híbridos resultantes.
- 2) Examinar la selectividad de los diferentes tipos de polen usado, en las flores hembras de los progenitores silvestres.
- 3) Evaluar la distancia de movimiento del polen de plantas de los CGMs a las plantas silvestres de *Cucurbita*.
- 4) Conocer los efectos de la introgresión entre los taxa cultivados y silvestres al darle seguimiento a los patrones de permanencia de los marcadores presentes en los CGMs en las generaciones siguientes de las poblaciones con diferente grado o nivel de hibridación, así como también la adecuación del (o los) híbrido(s). Inferir sobre los riesgos reales y potenciales relacionados con la posible liberación de CGMs del género *Cucurbita*, considerando los modos de reproducción de este género.

METAS ESPECÍFICAS DE LA PRIMERA ETAPA ORIGINAL DEL PROYECTO

- 1) Construcción de Invernadero y gestión para la importación de OGM's.
- 2) Recopilación de información, armado de bases de datos y elaboración de mapas de distribución real y potencial.
- 3) Colectas a partir de mapas de distribución, extracción de ADN, montaje de marcadores moleculares (microsatélites nucleares y secuencias de regiones de cloroplasto y mitocondria) y obtención de resultados.
- 4) Realizar análisis estadísticos y de genética de poblaciones de los microsatélites.
- 5) Iniciar secuencias de cloroplasto.
- 6) Terminar de secuenciar cloroplasto y armar las bases de datos de cloroplasto.
- 7) Iniciar análisis del cloroplasto. Concluir análisis de los microsatélites.
- 8) Iniciar experimentos de flujo génico y biología reproductiva.

AVANCES DE LA PRIMERA ETAPA

1) Construcción de invernadero y gestión para la importación de OGM's

En cuanto a los avances en la construcción del invernadero, la construcción del mismo ya se ha concluido. Actualmente se encuentra desocupado en espera a que se liberen los permisos para la importación de *OGM's* o de los experimentos de cruza con cultivares que se decidan realizar.

En cuanto a la gestión para la importación de *OGM's*, el convenio de colaboración revisado por los abogados de la UNAM se encuentra en revisión por parte de los abogados del INIFAP. Considerando que el Campo Experimental Bajío, INIFAP ya cuenta con un aviso para la utilización de *OGM's*, se procedió a solicitar un alcance específico para la importación de *CGM's* y un alcance de primera utilización para el invernadero. Ambos documentos se encuentran en la Dirección General del INIFAP. Sin embargo, seguimos esperando respuesta (ver Anexo 1), por lo que no se ha logrado iniciar esta parte del proyecto.

Durante la segunda etapa del proyecto se continuará gestionando la obtención del permiso para importación y utilización de *OGM's*. Sin embargo, este proceso ha resultado más complicado de lo previsto. De tal manera que si no se logra un avance significativo en la obtención de dichos permisos se comenzarán los experimentos de hibridación entre especies silvestres y poblaciones no-transgénicas usando cultivares comerciales de *Cucurbita* que sirvan como análogo para los experimentos y así evaluar el potencial de flujo génico entre taxa y tipos de poblaciones. Si este fuera el caso, se tendrá que evaluar la pertinencia de continuar con la gestión de los permisos para realizar los experimentos con *OGM's*.

Por otro lado, también se trabajó con la Comisión de Bioseguridad del Instituto de Ecología de la UNAM para avisar del posible uso de tejido (hojas congeladas) de plantas de *Cucurbita* genéticamente modificadas. Dicho aviso se encuentra en proceso.

Porcentaje de avance: 90%

2) Recopilación de información, armado de bases de datos y elaboración de mapas de distribución real y potencial.

Las bases de datos correspondientes a los ejemplares de herbario incluidos en el Sistema Nacional de Información sobre Biodiversidad (SNIB) y a las accesiones del Banco de Germoplasma de *Cucurbita*, INIFAP, Celaya, a cargo del Dr. Salvador Montes, fueron revisadas y

depuradas por parte de la CONABIO. Esta revisión implicó eliminar aquellos datos con coordenadas que caen fuera del territorio continental y aquellos que probablemente correspondan a introducciones o identificaciones incorrectas. La base de datos final se encuentra en el Anexo 2.

La base de datos incluye 786 números de registro del Banco de Germoplasma de *Cucurbita* de Celaya; fue revisada, se incorporaron datos de georreferencia por parte de la CONABIO y se encuentra en proceso de correcciones. El proceso de correcciones se está realizando actualmente y continuará durante la Segunda Etapa del proyecto, de manera que será entregada en diciembre de 2017. Asimismo, se incorporará la información de las colectas realizadas durante el desarrollo del proyecto a la base de datos, en un archivo aparte proporcionado por CONABIO (Anexo 3).

A partir de la base de datos actual se obtuvieron mapas de ocurrencia de taxones en México en colaboración con la CONABIO (Anexo 4) y mapas de distribución potencial de especies (Anexo 5, Figura 1) y de grupos de taxones (Anexo 5, Figura 2) con la colaboración del Biól. Jonás Aguirre Liguori (Instituto de Ecología, UNAM). Los mapas de distribución potencial se obtuvieron en el programa MaxEnt con 19 capas bioclimáticas de World Clim V.1.4 con una resolución de 30 Arco-segundos para generar los modelos de distribución de especies y de grupos de especies monofiléticos. Se cortaron las capas bioclimáticas de manera que incluyeran a México y sólo una parte de las países fronterizos. Para obtener el modelo, se usaron 20,000 puntos de *background* y 40% de datos para probar el modelo y 10 réplicas de *bootstrap*. Con base en los modelos resultantes, las áreas de distribución se convirtieron a valores binarios (presencia o ausencia de la especie), que fueron graficados con el paquete de R-raster.

A partir de la base de datos del SNIB, se puede concluir que la región con mayor diversidad de taxones para el género es la región Huasteca en los estados de Hidalgo y San Luis Potosí. También que existen estados pobremente muestreados que, de acuerdo con los mapas de distribución potencial, probablemente contengan una mayor diversidad que la representada en colecciones. De éstos destacan Tlaxcala, Aguascalientes y Colima (Anexo 5, Figura 1). En cuanto al número de registros por taxón, el que cuenta con un mayor número es *C. moschata*, seguido por *C. a. argyrosperma* y *C. a. sororia* (Anexo 5, Figura 2).

El análisis de las bases de datos y mapas ha permitido determinar qué taxones están ausentes de la colección y para qué regiones geográficas del país no existen accesiones en el Banco de Germoplasma de *Cucurbita*. Esto es relevante para la planificación del muestreo que se realizará. Particularmente para las colectas de *C. argyrosperma* ssp. *argyrosperma* y *C. moschata*.

Porcentaje de avance: 90%

3) Colectas a partir de mapas de distribución, extracción de ADN, montaje de marcadores moleculares (microsatélites nucleares y secuencias de regiones de cloroplasto)

Se ha avanzado de manera considerable en las colectas a partir de mapas de distribución. Sin embargo, es necesario realizar nuevas colectas durante la segunda etapa del proyecto que cubran los estados de Veracruz y Querétaro para tener una mejor representación de la variación genética de las especies de *Cucurbita* presente en todo el país. Se continuará con la germinación y extracción de ADN del nuevo material colectado, así como la amplificación de los marcadores moleculares.

El montaje de marcadores moleculares se ha realizado de manera exitosa. Hasta el momento se han montado 12 microsatélites nucleares, varias regiones del ADN del cloroplasto (*rpl20-rps12*, *psbD-trnT*, *psbJ-petA* y *atpI-atpH*) con poca variación intraespecífica y una región del ADN de la mitocondria (*trnL-trnF*) con variación intraespecífica en algunas de las especies.

Se avanzó en la obtención de los primeros datos genómicos que originalmente estaban considerados dentro de las metas para el segundo año del proyecto. Se enviaron a servicio de secuenciación *next-gen* muestras poblacionales de *C. argyrosperma* y se obtuvieron las primeras lecturas para el ensamble del transcriptoma y el genoma de esta especie. Además, los genomas de *C. argyrosperma* spp. *sororia* y *C. moschata* se encuentran en proceso de secuenciación. Durante la segunda etapa del proyecto se elegirán muestras poblacionales adicionales de *C. argyrosperma* spp. *argyrosperma*, así como individuos de *C. argyrosperma* spp. *sororia*, *C. moschata*, *C. pepo* y *C. ficifolia* para realizar tGBS que permitan identificar genes candidatos a adaptación y genes relacionados con la domesticación.

Una vez terminadas las colectas, se incluirá soporte documental de los registros de las colectas, tanto notas de campo como fotografías en apoyo a la captura y/o revisión de la información. Esta actividad se realizará durante la segunda etapa del proyecto.

Además se participará y proporcionará apoyo en la construcción de una página de difusión de información sobre las calabazas y chayotes de México.

Los avances realizados hasta el momento se detallan a continuación.

Colecta:

Entre los meses de diciembre 2013 y diciembre 2015 se realizaron cuatro salidas para coleccionar frutos de *C. argyrosperma* spp. *argyrosperma*, *C. argyrosperma* spp. *sororia* y *C. moschata*, en los estados de Jalisco, Guerrero, Oaxaca, Chiapas, la península de Yucatán, Nayarit, Sinaloa y Sonora. Las muestras provienen principalmente de terrenos de cultivo, potreros y huertos de traspatio. Se eligieron frutos que no presentaran daño mecánico o por patógenos. Todos los frutos colectados se depositaron en el Instituto de Ecología, UNAM para continuar su maduración y extraer las semillas. Además, se obtuvieron frutos de taxones para la inferencia filogenética, la documentación de *C. maxima* en México, la caracterización etnobotánica de las especies de *Cucurbita* en el valle de Tehuacán-Cuicatlán, y la caracterización morfológica de los frutos y semillas colectados.

En el caso de *C. argyrosperma* se colectaron 598 frutos de 53 poblaciones, distribuidas en 12 estados del país, para las dos subespecies. Se han colectado 236 frutos de 35 poblaciones de *C. argyrosperma* spp. *argyrosperma*, y 362 frutos de 18 poblaciones de *C. argyrosperma* spp. *sororia*.

Para *C. moschata* se han colectado 165 frutos de 40 municipios de 14 estados del país. Finalmente, se han obtenido 65 muestras de 6 accesiones provenientes de Guanajuato, Oaxaca y Brasil de la especie *C. maxima*, así como semillas provenientes de Chile y EUA.

Adicionalmente, se obtuvieron tejidos foliares de 40 individuos de 12 especies diferentes a partir de ejemplares del herbario MEXU del Instituto de Biología, UNAM. También se colectaron muestras de algunas especies silvestres como *C. lundelliana*, *C. palmata* y *C.*

radicans. Esto último con la finalidad de obtener una filogenia completa para el género *Cucurbita*.

Las semillas no utilizadas en el análisis serán almacenadas a 4°C en bolsas selladas y etiquetadas en el Laboratorio de Evolución Molecular y Experimental del Instituto de Ecología, UNAM.

Germinación:

En los meses de julio 2014, junio 2015 y junio 2016 se colocaron semillas para su germinación en condiciones controladas de invernadero. Las semillas se germinaron en sustrato orgánico con arena y grava bajo condiciones controladas de invernadero con temperatura de 30 a 35°C y un riego diurno por día.

Durante este periodo se germinaron semillas provenientes de distintas poblaciones obtenidas del BG-Celaya y de las colectas realizadas en 2013, 2014 y 2015. Después de 30 días de desarrollo, se cosecharon hojas de cada individuo para realizar extracción de ADN. Adicionalmente se cosecharon al menos 2 hojas por individuo para su preservación a -70°C en los ultracongeladores del Instituto de Ecología, UNAM. Hasta el momento se han germinado más de 1,000 individuos representantes de cinco especies de *Cucurbita* provenientes de 21 estados del país.

La mayoría de las accesiones facilitadas por el BG-Celaya no germinaron y solo se obtuvieron individuos de 6 accesiones para *C. argyrosperma*, 10 individuos de *C. maxima* y unas cuantas para *C. moschata* y 9 accesiones para *C. pepo*.

Hasta el momento se han germinado casi todas las muestras obtenidas en los viajes de colecta realizados en 2013 y 2014. Algunos individuos colectados durante la salida de 2015 se encuentran en proceso de crecimiento para realizar extracción de ADN. Aún es necesario germinar semillas adicionales provenientes de los individuos colectados en la salida realizada en octubre 2015 para realizar extracción de ADN y se está comenzando la germinación de *C. ficifolia* para realizar análisis de variación y estructura genética en esta especie. Además, durante la segunda etapa del proyecto se continuara con la germinación conforme se colecten nuevos individuos de las distintas especies consideradas en el proyecto.

Extracción de ADN:

Se realizó extracción de ADN de tejido fresco de los individuos germinados en el invernadero. Se utilizó un protocolo estándar de CTAB (Vázquez-Lobo, 1996; Anexo 6), así como el kit DNeasy Plant Mini Kit para algunas muestras de herbario. Hasta el momento se ha realizado extracción de ADN de alrededor de 1350 individuos, incluyendo las especies *C. argyrosperma*, *C. moschata*, *C. pepo* y *C. maxima*.

Para *C. argyrosperma* se ha realizado extracción de ADN de 534 individuos de 34 poblaciones de 15 estados del país. Además, se ha extraído ADN de 450 individuos de *C. moschata* provenientes de 18 localidades de 14 Estados del país. Para *C. pepo* se ha extraído ADN de 366 individuos provenientes de 24 poblaciones de 18 Estados del país. Finalmente, se extrajo ADN de 47 individuos de *C. maxima* de Guanajuato, Oaxaca y Brasil. Adicionalmente, se ha realizado extracción de 36 muestras que representan 19 taxa e incluyen cultivares comerciales de *C. pepo* para la filogenia del género *Cucurbita*.

De igual manera, durante la segunda fase del proyecto se continuará con la extracción de ADN de las muestras faltantes, incluyendo a *C. ficifolia*, así como de las muestras de los individuos obtenidos durante las nuevas colectas. Además, se incorporará a la filogenia una muestra de *C. radicans* que fue colectada en el mes de noviembre 2016.

Amplificación de Marcadores Moleculares. Cloroplasto:

La secuenciación de genes del cloroplasto se está llevando a cabo para los análisis filogeográficos de *C. argyrosperma*, *C. moschata* y *C. pepo*, así como para la inferencia de la filogenia del género. Las condiciones utilizadas para la amplificación de los marcadores moleculares se encuentra en el Anexo 7. Las regiones amplificadas del cloroplasto fueron enviadas inicialmente al servicio de secuenciación Sanger de la Universidad de Washington. Actualmente se está utilizando el servicio de secuenciación de MacroGen (www.macrogenusa.com).

Inicialmente se amplificaron las regiones *rpl20-rps12*, previamente utilizada por Zheng et al. (2013), y la región *trnL-trnF* (Taberlet et al. 1991). En relación a la baja variación encontrada en estas regiones del cloroplasto, se realizó una búsqueda de varias regiones de cloroplasto

reportadas como altamente variables en otras plantas (Shaw et al. 2005, 2007) y que no se encuentren duplicadas en la mitocondria (Alverson et al. 2010). Se mandaron a hacer cinco pares de oligos (*ndhF-rpl32*, *psbD-trnT^(GGU)*, *psbJ-petA*, *atpI-atpH* y *petL-psbE*). Se han estandarizado y amplificado con éxito las regiones *psbD-trnT^(GGU)*, *psbJ-petA* y *atpI-atpH* en todas las especies de *Cucurbita*. Se detectó que las regiones *psbD-trnT^(GGU)* y *psbJ-petA* son las que presentan mayor variación intraespecífica. Actualmente se está trabajando en la amplificación de estas dos regiones para individuos provenientes de distintas poblaciones de *C. argyrosperma*, *C. moschata* y *C. pepo*. Además se amplificaron las cinco regiones para los análisis de inferencia filogenética.

Durante la segunda etapa del proyecto se continuará con la amplificación de estas dos regiones del genoma del cloroplasto para todas las especies, incluyendo *C. ficifolia*, para obtener datos comparativos de variación genética a nivel interespecífico.

Amplificación de Marcadores Moleculares. Mitocondria:

En el proyecto original no se había considerado la amplificación de regiones de la mitocondria. Sin embargo, dada la baja variación hallada en el cloroplasto hasta el momento, se decidió incluir esta región en el análisis por que presenta variación genética importante en algunas de las especies para realizar los análisis filogeográficos. Por lo tanto, se continuará trabajando con este marcador durante la segunda fase del proyecto.

La secuenciación de genes de la mitocondria se está llevando a cabo para los análisis filogeográficos de *C. argyrosperma*, *C. moschata* y *C. pepo*. Actualmente se esta trabajando con la región mitocondrial *trnL-trnF*. Ésta es una región del cloroplasto que se encuentra duplicada e invertida en el genoma mitocondrial de las especies del género *Cucurbita* (Alverson et al. 2010). Esta región ha resultado de medianamente a altamente variable en las especies en las que ha sido amplificada exitosamente (*C. argyrosperma*, *C. maxima*, *C. moschata* y *C. pepo*). Las regiones amplificadas de la mitocondria están siendo enviadas al servicio de secuenciación MacroGen. Las condiciones utilizadas para la amplificación de este marcador se encuentran en el Anexo 7.

Finalmente, se intento estandarizar la región *NAD1* de la mitocondria para incorporarla en la filogenia y para determinar si presenta variación para los análisis filogeográficos. Sin embargo, no se tuvo éxito en la amplificación de esta región.

Amplificación de Marcadores Moleculares. Microsatélites:

Se eligieron 20 pares de oligos (= *primers*) de microsatélites de los más de 300 disponibles para el género, usando como criterio la composición del motivo de repetición (dinucleótido) y el largo estimado del amplificado. Hasta la fecha se han estandarizado condiciones para la amplificación de 13 pares de oligos. De estos, se eligieron 12 microsatélites previamente utilizados (Gong et al. 2008) que fueron montados en reacción multiplex (reacciones de amplificación simultánea de tres loci) y se ha iniciado el proceso de genotipado en *C. argyrosperma*, *C. moschata* y *C. pepo*, el cual continuará durante la segunda etapa del proyecto. Además, se comenzará el montaje y amplificación de estos marcadores en *C. ficifolia*. Las condiciones utilizadas para la amplificación de los marcadores moleculares se encuentra en el Anexo 7.

Amplificación de Marcadores Moleculares. GBS:

Dentro del marco de actividades a realizar durante el segundo año del proyecto original se encuentra contemplado la obtención de datos genómicos para un estudio de genómica comparada así como para análisis de genómica de poblaciones de *C. argyrosperma* spp. *argyrosperma*. Se avanzó considerablemente en esta parte del proyecto ya que se enviaron para genotipación mediante un método de secuenciación de siguiente generación llamado tGBS (*Tunable Genotyping-by-sequencing*) 96 ADNs de individuos de diferentes poblaciones de *C. a. argyrosperma* y *C. a. sororia* de diversas poblaciones de dichos taxones en México. El servicio de genotipado se está realizando con la compañía Data2Bio (<http://www.data2bio.com/services/tGBS/>) para obtener polimorfismos en hasta 10,000 loci nucleares. Para garantizar la efectividad de este método la concentración y la calidad del DNA se evaluaron mediante electroforesis en geles de agarosa y se estimó el cociente 260/280 nm en un equipo Nanodrop, que debe estar en valores de entre 1.7 a 2.0.

Actualmente el proceso de genotipado ha sido concluido y se recibieron los datos crudos. Se están realizando análisis de estos datos y son muy promisorios y polimórficos. Los resultados preliminares indican la presencia de flujo génico entre *C. argyrosperma* spp. *argyrosperma* y su ancestro silvestre *C. argyrosperma* spp. *sororia* entre poblaciones de diferentes estados. Con base en los resultados obtenidos a partir de los microsatélites nucleares se seleccionarán individuos adicionales de *C. argyrosperma* spp. *argyrosperma*, *C. argyrosperma* spp. *sororia*, así como de las especies *C. moschata*, *C. pepo* y *C. ficifolia* que serán enviados, durante la segunda etapa del proyecto, al servicio de tGBS para realizar análisis de diversidad y estructura genética y para la detección de regiones importantes para el proceso de domesticación.

Con el objetivo de complementar los análisis de tGBS y para futuros estudios de genómica de la conservación y la domesticación, adicionalmente estamos secuenciando y armando el genoma nuclear completo a alta resolución (más de 100x) de *C. argyrosperma* spp. *argyrosperma* y los transcriptomas de diferentes tejidos del mismo individuo, además de genomas a menor resolución de *C. argyrosperma* spp. *sororia* y *C. moschata*.

Para la obtención del genoma se germinaron 5 semillas y se seleccionó la planta más robusta de *C. argyrosperma* spp. *argyrosperma* para cosechar tejidos foliares para la extracción de ADN total y diferentes tejidos (tallo, raíz, botón floral y zarcillo) para la extracción de ARN (Anexo 8). Actualmente, ya se cuenta con un ensamble de alta calidad de *C. argyrosperma* spp. *argyrosperma* generado con secuencias de Illumina y PacBio, además de las secuencias consenso de los genomas de *C. argyrosperma* spp. *sororia* y *C. moschata* (detallado en Anexo 8). La CONABIO y el INECOL están colaborando con los investigadores responsables del proyecto para que parte de los análisis bioinformáticos de alta demanda de memoria RAM, como el ensamble de genomas/transcriptomas, el mapeo de lecturas, la corrección de bases y la búsqueda de modelos génicos, se realicen en los *clusters* de dichas instituciones.

Se realizó extracción de ADN total a partir de tejido foliar de *C. a. argyrosperma*, el cual se secuenció en una línea de Illumina HiSeq2000 PE100 con una librería de ~350 pb, una línea de Illumina MiSeq PE300 con una librería de ~1000 bp y 8 celdas de PacBio P6/C4 con MagBeads y librerías de tamaños mayores a 8kb.

También se realizaron extracciones de ARN a partir de tejido de hojas jóvenes, tallo, raíz, botones florales y zarcillos para obtener el transcriptoma de *C. a. argyrosperma* utilizando el kit RNeasy Plant Mini Kit (QIAGEN) con el protocolo propuesto por el fabricante. El transcriptoma permitirá reconocer genes codificantes y ARNs no codificantes dentro del genoma de referencia para completar el proceso de ensamblado y anotación de genes. La extracción fue exitosa y las muestras de ARN fueron precipitadas en etanol y cloruro de litio para su envío al centro QB3 de la Universidad de Berkeley, donde se realizó la construcción de librerías *ribo-depleted* indexadas de 350 pb para cada tejido y se corrieron en una sola línea HiSeq2000 PE100.

Se obtuvo un primer ensamblaje del genoma de *C. argyrosperma* spp. *argyrosperma* con una cobertura de ~140x utilizando las secuencias de Illumina. Se han realizado los mapeos correspondientes al genoma del cloroplasto (15 contigs; 151,740 bp) y la mitocondria (151 contigs; 932,802 bp) que se encuentran en proceso de anotación. Para el genoma nuclear se obtuvieron 187,409 scaffolds con un tamaño de 343.5 Mbp (Anexo 8). Además, se ensambló *de novo* el transcriptoma, obteniendo 289,255 contigs, de los cuales se encontraron 106,420 marcos abiertos de lectura y 82.1% de estos fueron anotados (Anexo 8). Se lograron predecir 26,997 modelos génicos soportados por evidencia transcripcional dentro del genoma, de los cuales se lograron anotar el 96.8%. También se identificaron 541,734 elementos repetitivos, 2,196 proteínas pertenecientes a elementos móviles y 10,656 RNAs no codificantes.

Se han realizado mapeos de lecturas de *C. argyrosperma* ssp. *sororia* y *C. moschata* contra el genoma ensamblado de *C. argyrosperma* ssp. *argyrosperma*. Se encontraron 2,169,909 SNPs y 232,620 indels entre el genoma de *C. moschata* y el genoma de *C. argyrosperma* ssp. *argyrosperma*, además de 1,379,450 SNPs y 194,014 indels entre el genoma de *C. argyrosperma* ssp. *sororia* y el genoma de *C. argyrosperma* ssp. *argyrosperma*.

Recientemente se ensambló una segunda versión del genoma utilizando las lecturas obtenidas con PacBio. Dichas lecturas son muy largas y permitieron obtener un genoma mejor reconstruido de *C. argyrosperma* ssp. *argyrosperma*. Se obtuvieron 1018 scaffolds con un tamaño total de 241 Mbp, una N50 de 638,754 pb y una L50 de 107 scaffolds.

Porcentaje de avance: 80%

4) Realizar análisis estadísticos y de genética de poblaciones de los microsatélites

Se han obtenido datos para microsatélites para *C. argyrosperma* spp. *argyrosperma*, *C. argyrosperma* spp. *sororia*, *C. moschata*, *C. pepo* spp. *pepo* y *C. pepo* spp. *fraterna*. Estos marcadores han resultado polimórficos en todas las especies. Aún se está trabajando en la amplificación de estos marcadores para todas las especies y se comenzará la amplificación en *C. ficifolia*. Durante la segunda etapa del proyecto se continuará con la amplificación de los 6 a 12 microsatélites nucleares y una vez realizada la genotipación de todos los individuos se llevarán a cabo los análisis finales de variación y estructura genética por especie y comparando entre especies.

A continuación se presentan análisis preliminares realizados con los datos obtenidos hasta el momento.

En el caso de *C. argyrosperma*, ya se tienen los primeros datos de 9 microsatélites para 339 individuos de 14 estados para ambas subespecies (298 individuos *C. argyrosperma* spp. *argyrosperma*, 41 individuos de *C. argyrosperma* spp. *sororia*). Actualmente se están obteniendo los tamaños alélicos para construir la base de datos, misma que será analizada una vez se hayan amplificado todos los individuos. Los resultados preliminares de heterocigosis esperada (H_E) sugieren que las poblaciones de Autlán en el estado de Jalisco, Tanquián en S.L.P. y Ometepec, Guerrero de la ssp. *argyrosperma* presentan valores mayores de variación genética que las demás poblaciones. Las poblaciones de la península de Yucatán y Tehuantepec presentan los valores más bajos. Los patrones de estructura muestran que existen tres pozas génicas (vertiente del pacífico, centro-norte y península de Yucatán) (Anexo 9). Quedan por analizar las poblaciones de la ssp. *sororia* que serán de suma importancia para hacer inferencias de flujo génico entre ambas subespecies.

Por otra parte, para *C. moschata* se han amplificado 11 microsatélites nucleares en 295 individuos de 18 localidades de 14 estados de la República Mexicana. Los resultados preliminares sugieren que la mayoría de las poblaciones mantienen de medianos a altos valores de diversidad genética medida como Heterocigosis esperada (H_E) donde San Luis Potosí y Puebla poseen la variación genética más alta, mientras que Buctzotz (Yucatán) y Coyuca (Guerrero) son

las localidades con los niveles más bajos. De manera general los valores de endogamia no son significativos lo que refleja que esta especie se autopoliniza en baja medida. Los patrones actuales de estructura genética aun sugieren la presencia de tres pozas génicas diferenciadas (Norte, Centro y Sureste) (Anexo 10) aunque estos patrones podrían cambiar al analizar los datos considerando cuestiones importantes de manejo y practicas de cultivo.

Para *C. pepo* se han amplificado 10 microsatélites nucleares para 339 individuos de 18 poblaciones provenientes de 17 estados del país, incluyendo una población de *C. pepo* spp. *fraterna* e individuos representantes de algunas variedades comerciales. Los resultados preliminares sugieren que en general *C. pepo* presenta alta variación genética, siendo Tlaxcala, el Estado de México y Zacatecas los estados con mayor variación (Anexo 11). Actualmente se están estandarizando 2 microsatélites nucleares adicionales para estos individuos y se están amplificando los 10 microsatélites en muestras adicionales, incluyendo algunas muestras de *C. pepo* spp. *ovifera*.

En *C. ficifolia* los avances han sido más lentos, en particular debido a que en el invernadero del Instituto de Ecología de la UNAM no germinaban las semillas y si lo hacían, se morían las plántulas, debido a que es un invernadero muy caliente (solo crecen bien especies del desierto y de lugares tropicales calientes). Actualmente se están creciendo las plantas en un invernadero donde crecen bien, en la UAEM de la parte alta de la Ciudad de Cuernavaca. En la segunda etapa del proyecto se obtendrán y analizarán los datos de esta especie.

Porcentaje de avance: 75%

5) Secuencias de cloroplasto

Se avanzó con la secuenciación de distintas regiones del cloroplasto. Se han obtenido secuencias para poblaciones de *C. argyrosperma* spp. *argyrosperma*, *C. argyrosperma* spp. *sororia*, *C. moschata*, *C. maxima*, *C. pepo* spp. *pepo* y *C. pepo* spp. *fraterna*, además de secuencias para uno o dos individuos de los taxa restantes del género para realizar análisis filogenéticos. Estas regiones han resultado poco variables a nivel intraespecífico, por lo que se

ha optado por complementar los datos con secuencias de una región del ADN de la mitocondria que presenta variación intraespecífica.

Durante la segunda etapa del proyecto se continuara con la amplificación y el análisis evolutivo de estas regiones para realizar estimar la variación entre especies. Los avances realizados en la obtención de secuencias se detallan a continuación.

En el caso de *C. argyrosperma* se amplificaron 51 muestras para la región *rpl20- rps12* del cloroplasto. Sin embargo, no se encontró variación. Adicionalmente se amplificaron secuencias de las regiones *psbD-trnT*, *psbJ-petA* y *atpI-atpH* para individuos provenientes de distintas poblaciones del país. Se encontró poca variación en el fragmento *psbD-trnT*, por lo que no se considera que sean adecuadas para estudios de diferenciación poblacional. De igual manera, se amplificaron estas regiones del cloroplasto en distintos individuos de distintas poblaciones de *C. moschata* donde los marcadores de cloroplasto han resultado monomórficos.

En el caso de *C. maxima* se amplificó la región *trnL-trnF* del cloroplasto para 25 muestras. Adicionalmente se amplificaron las regiones *psbD-trnT*, *psbJ-petA* y *atpI-atpH*, encontrando variación en la región *psbJ-petA* (Anexo 12).

Finalmente, para *C. pepo* se han amplificado las regiones *psbD-trnT* y *psbJ-petA* de todos los individuos. Esta especie sí presenta variación para realizar análisis de filogeografía de esta especie en México (Anexo 11).

En cuanto a la amplificación de los marcadores de cloroplasto para la inferencia filogenética, se amplificó exitosamente la región *rpl20-rps12* de la mayoría de las muestras con que se cuenta. Estas muestras representan a 19 taxa, incluyendo cultivares comerciales de *C. pepo*. Adicionalmente, se amplificó la región *trnL-trnF* para varias especies silvestres. Varias de las especies de las que aún no se cuenta con muestra se encuentran reportadas en el GenBank, por lo que las secuencias fueron descargadas. Se encontró poca variación en esta región, de manera que se amplificaron las regiones *psbD-trnT*, *psbJ-petA* y *atpI-atpH* y se recuperaron regiones adicionales de los genomas de cloroplasto recientemente publicados por Kistler et al. (2016) para completar a las especies representantes del género, así como para tener los grupos externos necesarios para enraizar la filogenia. Además, se obtuvieron secuencias reportadas por Zheng et al. (2013). En total se consideraron 6 regiones del ADN del cloroplasto, que abarcan

6,514 pbs, de los cuales 6,127 fueron utilizadas para realizar la reconstrucción filogenética (Anexo 13). Recientemente se obtuvieron frutos de *C. radicans* por lo que es necesario volver a realizar los análisis filogenéticos. De esta manera se obtendrá la primera filogenia reportada que incluye a esta especie.

Amplificación de regiones de la Mitocondria:

Como se menciona con anterioridad, en el proyecto original no se había contemplado incluir secuencias de ADN mitocondrial. Se decidió incluir la región mitocondrial *trnL-trnF*, dada la baja resolución que han resultado tener los marcadores de cloroplasto. Ambos marcadores son de herencia materna y funcionan muy bien para llevar a cabo análisis filogeográficos. La amplificación de este marcador ha sido exitosa en *C. argyrosperma*, *C. moschata*, *C. maxima* y *C. pepo* por lo que se continuará con la amplificación de todas las muestras obtenidas hasta el momento durante la segunda etapa del proyecto.

En el caso de *C. argyrosperma* se amplificó la región mitocondrial *trnL-trnF*. Esta especie ha demostrado poca variación genética en la región amplificada hasta el momento (un solo sitio variable), por lo que es recomendable realizar la amplificación de individuos provenientes de otras poblaciones, así como de otras regiones con la finalidad de encontrar más variación que permita realizar análisis de inferencia filogeográfica más finos.

En cuanto a *C. moschata* también se amplificó la región mitocondrial *trnL-trnF*. Sin bien inicialmente, se analizaron 7 individuos de 6 poblaciones localizadas en distintos estados del país actualmente se han integrado datos del resto de localidades incluidas en los análisis anteriores sin embargo el avances en los análisis formales aun no se ha completado. Se ha encontrado suficiente variación genética en esta región de la mitocondria, por lo que se ha incrementado el tamaño de muestra para este marcador.

Para *C. maxima* se amplificó la región *trnL-trnF* mitocondrial para 22 muestras, y se están analizando estos primeros datos. Cabe señalar que algunos de los haplotipos presentes en *C. maxima* se comparten con *C. moschata* lo que podría indicar hibridación entre estas especies, o que algunas muestras fueron clasificadas erróneamente. Finalmente, *C. pepo* presento baja

variación genética, con únicamente 2 haplotipos, uno de ellos compartido con *C. pepo* spp. *fraterna* (Anexo 11).

Porcentaje de avance: 95%

6) Terminar secuencias de cloroplasto y armar bases de datos de cloroplasto

Actualmente se está realizando el trabajo necesario para concluir la germinación, extracción de ADN y amplificación de marcadores moleculares para *C. argyrosperma*, *C. moschata* y *C. pepo*, que son las especies con las que se ha tenido mayor avance. Además, se está iniciando el trabajo con *C. ficifolia*. El genoma de cloroplasto de estas especies ha demostrado baja variación genética por lo que se tuvieron que probar distintas regiones para encontrar variación. El polimorfismo hallado en las tres regiones del cloroplasto que se están amplificando en este momento es muy variable, siendo prácticamente monomórficos en *C. argyrosperma* y *C. moschata* y con variación moderada en *C. pepo*. Sin embargo, es importante continuar con la amplificación de estas regiones durante la segunda etapa del proyecto para obtener datos que sean comparables entre especies. Estos datos, aunados a las secuencias del ADN mitocondrial y de los microsatélites nucleares permitirán realizar hipótesis relacionadas con la historia evolutiva de estas especies y el proceso de domesticación que podrán ser puestas a prueba por medio de métodos ABC (*Approximate Bayesian Computation*).

Las secuencias y genotipos obtenidos serán almacenados en un repositorio de datos (NCBI y/o DataDryad) hasta que se concluya la secuenciación de todas las muestras por especie, ya que es preferible tener números de acceso consecutivos y que todas las secuencias estén asociadas al mismo archivo. Estos constituirán un paquete organizado de archivos de datos o ruta en donde estén depositados, asociados con los análisis realizados por el proyecto más los metadatos de los archivos y variables. Se incluirá un archivo "LEEME" donde se describa cada archivo (o conjunto de archivos) y cómo está asociado a los análisis realizados.

El repositorio será también publicado por los autores junto con las publicaciones que deriven del proyecto en repositorios de datos curados (e.g. DataDryad). El repositorio seguirá

las recomendaciones de manejo y publicación de datos de White et al. (2013) y de <http://datadryad.org/pages/faq>.

Porcentaje de avance: 90%

7) Análisis de cloroplasto y mitocondria

En vista de las dificultades iniciales para la germinación de muestras provenientes del banco de germoplasma, así como para la estandarización de marcadores polimórficos, aun nos encontramos realizando trabajo de laboratorio para la obtención de datos. Los análisis finales de cloroplasto, mitocondria y microsatélites nucleares se realizarán durante la segunda etapa del proyecto, una vez concluido el trabajo de laboratorio, como se indica en el cronograma de actividades. Con los datos obtenidos hasta el momento es posible realizar análisis preliminares. Los resultados preliminares generales para los microsatélites nucleares se describen en el punto 4 (realizar análisis estadísticos y de genética de poblaciones de los microsatélites) del presente documento, mientras que los resultados preliminares para las secuencias del cloroplasto y la mitocondria se presentan a continuación. Los resultados detallados por especie se encuentran en los anexos respectivos a cada especie.

Los análisis iniciales indican que tanto *C. argyrosperma* (Anexo 9) como *C. moschata* (Anexo 10) presentan baja o nula variación genética en las regiones del cloroplasto amplificadas hasta el momento, por lo que estas regiones no son útiles para realizar análisis filogeográficos de estas especies y es necesario complementar la información con datos de otras regiones (ADN mitocondrial) que si presenten variación o datos de tGBS. Estos datos sugieren un fuerte cuello de botella asociado al proceso de domesticación.

Por otra parte, para *C. moschata* la región mitocondrial que se ha amplificado hasta el momento presenta variación adecuada para realizar la reconstrucción filogeográfica de la especie. En total se amplificaron 182 individuos de 18 poblaciones para llevar a cabo análisis preliminares de diversidad genética. Los resultados preliminares del ADN mitocondrial de *C. moschata* sugieren alta diversidad de haplotipos en Veracruz y baja variación en Campeche, Yucatán y Sinaloa (Anexo 10). Sin embargo aun falta analizar las ultimas secuencias y re analizar

los datos para tener resultados más formales. Estos resultados sugieren la presencia de tres grupos genéticos que coinciden con los detectados con los microsatélites nucleares (Anexo 10).

Para *C. pepo* los análisis preliminares sugieren la presencia de dos linajes de cloroplasto, y coincide con lo reportado previamente para la especie (Zheng et al. 2013) y con los resultados obtenidos hasta el momento para los microsatélites nucleares (Anexo 11). Las poblaciones con mayor variación genética son Zacatecas, Hidalgo y San Luis Potosí. Además la presencia de un solo haplotipo en cada población sugiere fuerte estructura genética en esta especie (Anexo 11). El haplotipo encontrado con mayor frecuencia en los cultivares de *C. pepo* se comparte con la especie silvestre *C. pepo* spp. *fraterna*. Además, las variedades Spaguetti, Delicata honeyboat y Straight neck están más cercanamente relacionadas con *C. pepo* spp. *ovifera* variedades *texana* y *ozarkana* (Anexo 11).

En el caso de *C. maxima* se amplificó la región *trnL-trnF* del cloroplasto para 25 muestras. Adicionalmente se amplificó la región *psbJ-petA* para 65 muestras provenientes de distintas partes de la República Mexicana, así como para variedades comerciales de Brasil, Chile y Estados Unidos de Norteamérica. Las secuencias ya han sido ensambladas y analizadas. Se comprobó la *C. maxima* en México está presente en los estados de Guanajuato, Jalisco, Oaxaca, Puebla, Distrito Federal y Sinaloa. Se identificaron dos haplotipos distribuidos en distintas partes del país, y mayor variación en la accesión del Distrito Federal a diferencia de las accesiones de Sonora, Guanajuato y Chile (Anexo 12). La tesis de licenciatura, de Paulina Hernández, resultante de estos análisis se encuentra en proceso de revisión por parte del jurado de examen y la versión final de la misma será enviada a la CONABIO una vez la alumna haya presentado su examen. Cabe señalar que algunos de los haplotipos presentes en esta especie se agrupan con los haplotipos amplificados para *C. moschata*. Es necesario contar con datos nucleares para determinar si se trata de eventos de hibridación entre estas especies. Los resultados de este análisis fueron presentados en el XX Congreso Mexicano de Botánica realizado en la Ciudad de México del 4 al 9 de septiembre 2016 (Anexo 14).

En cuanto a los análisis para la inferencia filogenética, se trabajó con 6 regiones del ADN del cloroplasto, que abarcan aproximadamente 6,519 pbs de 20 taxa representantes del género (Anexo 13). Las secuencias fueron alineadas y se finalizó la fase de análisis de los datos. Los

análisis indican que estos marcadores resuelven de manera adecuada las relaciones filogenéticas de las especies integrantes del género con altos valores de soporte de las ramas (Anexo 13). Además, se realizó un análisis filogenético con las secuencias obtenidas para la región *trnL-trnF* de la mitocondria en la que se observa que *C. pepo* y *C. argyrosperma* forman un clado monofilético cada uno, mientras que las relaciones entre los haplotipos presentes en *C. moschata* y *C. maxima* no se resuelven adecuadamente (Anexo 13). Además, se calcularon las fechas de divergencia, las tasas de sustitución molecular de los distintos linajes del género, así como las tasas de diversificación del género (Anexo 13).

La tasa de diversificación calculada para el género *Cucurbita* fue de 0.320 taxa/millón de años considerando los 22 taxa del género. Para el grupo Digitata fue de 0.627 taxa/millón de años, para el grupo Máxima de 0.858 taxa/millón de años, para el grupo Pepo de 1.183 taxa/millón de años, para el grupo Argyrosperma de 0.923 taxa/millón de años y para el grupo Okeechobeensis de 3.923 taxa/millón de años (Anexo 13).

Esta información se utilizó para culminar la tesis de licenciatura titulada “**FILOGENIA MOLECULAR DEL GÉNERO *Cucurbita* L. (CUCURBITACEAE) USANDO SECUENCIAS DE CLOROPLASTO**” de Leslie Paredes, quien presentó su examen de defensa el 2 de septiembre de 2016 (se integra la tesis en el Anexo 13). También se presentaron los resultados de este trabajo en el XX Congreso Mexicano de Botánica (Anexo 14) y se está trabajando en un manuscrito que será sometido en una revista internacional indizada.

Las conclusiones principales de este trabajo son:

- Se generó una filogenia molecular de cloroplasto mejor resuelta que las previamente publicadas, en la que se observó una clara definición de los grupos pertenecientes al género *Cucurbita* con altos valores de soporte de ramas.
- Dentro del grupo Pepo, se observó la diferenciación de las subespecies mexicanas de las subespecies y variedades de E. U. A., lo cual soporta los dos eventos de domesticación en este grupo.
- *C. moschata* y *C. lundelliana* se observaron dentro de los grupos Argyrosperma y Okeechobeensis, respectivamente, por lo que se propone su inclusión formal a estos grupos y su reevaluación como posibles subespecies.

- Se obtuvo una fecha de divergencia estimada de aproximadamente 9.64 Ma para el género *Cucurbita*. Similar al de otros taxa endémicos de América y al de sus polinizadores, lo que sugiere un proceso de coevolución polinizador-planta.
- Los grupos actuales del género *Cucurbita*, divergieron durante el Pleistoceno, entre 1.75 Ma y 280 mil años atrás.
- El grupo Okeechobeensis es el grupo más reciente del género *Cucurbita* de acuerdo con las fechas estimadas de divergencia con 280 mil años.
- *C. moschata* presentó una tasa de sustitución molecular más alta seguida de *C. cordata*. Aunque no se observó una correlación con su ciclo de vida o distribución, es posible que esté relacionado a su diversidad morfológica o su manejo intensivo.
- La tasa de diversificación del género *Cucurbita* fue de 0.320 spp/Ma, similar a otros géneros de la flora mexicana.
- La región *trnL-trnF* originaria del cloroplasto se encontró duplicada en la mitocondria de los taxa que pertenecen a los grupos mesofíticos: Pepo, Argyrosperma y Okeechobeensis. Encontrándose la misma topología que con el cloroplasto.

Porcentaje de avance: 70%

8) Experimentos de flujo génico y biología reproductiva

El inicio de los experimentos de flujo génico y biología reproductiva estaba contemplado para el segundo año del proyecto original. La construcción del invernadero destinado a la realización de estos experimentos fue llevada a cabo y el invernadero ya se encuentra listo para el inicio de la fase experimental. Sin embargo, no se ha logrado avanzar en la obtención del permiso para importación y utilización de OGM's (ver punto 1). Por lo tanto, se ha tomado la decisión de iniciar los experimentos dentro de la segunda etapa del proyecto, considerando el flujo génico e hibridación entre dos líneas de variedades comerciales y una población de cada taxa silvestre como experimento piloto, como se detallan en el protocolo anexo para la segunda etapa del proyecto.

La temporada ideal para el inicio de los experimentos es después del mes de marzo, preferentemente durante el mes de junio. Se utilizará una población de cada taxa, considerando a cuatro taxa silvestres: *C. argyrosperma* spp. *sororia*, *C. lundelliana*, *C. pepo* spp. *fraterna* y *C. okeechobensis* spp. *martinezii*. Las cruza serán llevadas a cabo entre estos taxa silvestres y dos líneas comerciales modernas pero no transgénicas de *C. pepo* que sirvan como análogo de OGM's. Se realizaran experimentos de carga de polen, ecología de la polinización y evaluación de los efectos de la paternidad que se detallan en la propuesta para la segunda etapa del proyecto.

Porcentaje de avance: 30%

Actividades adicionales no contempladas en la primera etapa

Se enviaron muestras de ADN para la obtención de SNPs con métodos *next-gen* (tGBS) de poblaciones de *C. argyrosperma* para análisis genómicos de variación genética. Actualmente nos encontramos en el proceso de envío de muestras adicionales para análisis de SNPs de *C. argyrosperma* spp. *argyrosperma* y *C. argyrosperma* spp. *sororia* así como muestras de *C. moschata* y *C. pepo* (spp. *pepo* y spp. *fraterna*). Durante la segunda fase del proyecto se analizarán con este mismo método muestras de *C. ficifolia*. Los análisis serán realizados una vez genotipadas todas las especies consideradas en este análisis.

Se obtuvieron los primeros ensamblajes del transcriptoma y el genoma de *C. argyrosperma* spp. *argyrosperma*. La CONABIO está colaborando con los investigadores responsables del proyecto para que parte de los análisis bioinformáticos de alta demanda de memoria RAM se realicen en el *cluster* de la CONABIO, con la finalidad de acortar los tiempos de procesamiento de los datos. Ya se tiene acceso al servidor para llevar a cabo los análisis genómicos.

Ya se ha avanzado en el ensamblaje de una segunda versión del genoma de *C. argyrosperma* ssp. *argyrosperma*, utilizando lecturas obtenidas con la plataforma PacBio. Dichas lecturas permitirán generar un genoma de referencia de mayor calidad al reducir el número de contigs/scaffolds. Durante la segunda fase del proyecto se volverán a anotar los genes dentro de

este nuevo ensamble, esperando que se consiga un número cercano a 20,000 genes, de acuerdo a lo observado en taxa cercanamente emparentados como *Cucumis sativus* y *Citrullus lanatus* (Guo *et al.*, 2012; Huang *et al.*, 2009). La segunda versión del genoma de *C. argyrosperma* ssp. *argyrosperma* se utilizará como referencia para ensamblar los genomas de *C. argyrosperma* spp. *sororia* y *C. moschata*, permitiendonos generar genomas de buena calidad a un costo menor. Además, se pretende incluir genomas de una especie del grupo Digitata, una especie del grupo Foetidissima, *C. ficifolia*, *C. pepo* spp. *pepo* y *C. pepo* spp. *fraterna* que nos permitan identificar paralelismos evolutivos asociados al proceso de domesticación.

Durante la segunda fase del proyecto, y una vez obtenidos los genomas de referencia para cada especie, se compararán los genomas de estas calabazas con respecto a los genomas reportados de las demás cucurbitáceas para detectar cambios estructurales, eventos de duplicación génica, cambios en el cariotipo y expansión/contracción de familias génicas.

Por otra parte, se estandarizó la amplificación de una región de la mitocondria que resultó polimórfica. Durante la segunda etapa del proyecto se continuará con la amplificación de esta región para llevar a cabo análisis de filogeografía de *C. argyrosperma*, *C. moschata* y *C. pepo*.

Adicionalmente se está realizando un estudio de maestría sobre la variación morfológica del género *Cucurbita* en México llevado a cabo por Thalita E. Legaspi (Anexo 15). Además, se realizó un estudio etnobotánico de las diferentes especies de *Cucurbita* en el Valle de Tehuacán. La tesis de Gustavo García Jaramillo, resultado de este análisis está en fase de tramites de titulación y se incluye en el Anexo 16. En el Anexo 16 también se incluyen los avances de tesis de Paulina Hernández Vargas sobre la documentación de *C. maxima* en México, quien se encuentra en fase de revisión por parte del jurado de examen, y de Karen Yazmín Ruíz Mondragón sobre la filogeografía de *C. pepo* en México utilizando marcadores de cloroplasto, quien se encuentra en fase de registro de tesis.

ENTREGABLES INDICADORES DE ÉXITO DE LA PRIMERA ETAPA DEL PROYECTO QUE SE CONCLUIRAN DURANTE LA SEGUNDA ÉTAPA DEL PROYECTO:

| Entregables | |
|---|--------|
| Base de datos BG Celaya | dic-17 |
| Mapas de distribución potencial actualizados | jun-17 |
| 4 tesis de licenciatura (Gustavo, Paulina, Leslie y Karen) | jun-17 |
| Un artículo científico en revista internacional, arbitrada e indexada | dic-17 |
| Un artículo de difusión | dic-17 |

En cuanto a los entregables indicadores de éxito de la primera etapa del proyecto. La base de datos BG Celaya se encuentra en fase de correcciones y la versión final se entregará en diciembre de 2017. Los mapas de distribución potencial actualizados se podrán procesar una vez se hayan finalizado las nuevas colectas y se cuente con todos los registros actualizados, por lo que se plantea podrán ser entregados a finales de junio de 2017.

En cuanto a las tesis de licenciatura planteadas como entregables, se ha finalizado una de las tesis (Leslie Paredes; Anexo 13), dos tesis se encuentran en fase final de revisión por parte del jurado de examen (Gustavo García y Paulina Hernández; Anexo 16) y los exámenes profesionales se llevarán a cabo dentro del primer semestre de 2017. Una de las tesis (Karen Ruíz; Anexo 16) está en proceso de redacción, registro y revisión por el tutor. Los tramites de solicitud de jurado se comenzarán a realizar durante el primer semestre de 2017.

Finalmente, se están trabajando los borradores de dos artículos científicos que serán sometidos a revistas internacionales arbitradas e indexadas (Anexo 16). Uno sobre la variación genética en microsatélites nucleares de *C. argyrosperma* (se está considerando la revista *Frontiers in Ecology and the Environment*) y otro sobre la biogeografía histórica del género *Cucurbita* (se están considerando las revistas *Systematic Biology* o *New Phytologist*). Los artículos serán sometidos a revisión en el transcurso del primer semestre de 2017.

Literatura Citada

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

- Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., ... Xu, Y. (2012). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics*, 45(1), 51–58. <http://doi.org/10.1038/ng.2470>
- Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., ... Li, S. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics*, 41(12), 1275–1281. <http://doi.org/10.1038/ng.475>
- Kistler, L., et al. 2016. Gourd and squashes (*Cucurbita* spp.) adapted to megafaunal extinction and ecological anachronism through domestication. *Proceedings of the National Academy of Science*, DOI: 10.1073/pnas.1516109112.
- Shaw J., et al. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92:142-166.
- Shaw J., Lickey E. B., Schilling E. E., Small R. L. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94:275-288.
- White E. P., et al. 2013. Nine simple ways to make it easier to (re)use your data. *Ideas Eco Evo* 6.
- Zheng, Y., et al. 2013. Chloroplast phylogeny of *Cucurbita*: evolution of the domesticated and wild species. *Journal of Systematics and Evolution* 5:326-334.

INDICE DE ANEXOS

- Anexo 1. Alcance específico para la importación de CGM's y alcance de primera utilización para el invernadero.
- Anexo 2. Base de datos final de los ejemplares de herbario incluidos en el SNIB y accesiones del Banco de Germoplasma de *Cucurbita*, INIFAP, Celaya.
- Anexo 3. Base de datos de las colectas realizadas durante el desarrollo del presente proyecto.
- Anexo 4. Mapas de ocurrencia de especies de *Cucurbita* en México.
- Anexo 5. Mapas de distribución potencial de especie y de grupos de taxones de *Cucurbita* en México.
- Anexo 6. Protocolo de extracción de ADN en *Tissue-lyser*.
- Anexo 7. Lista de marcadores moleculares utilizados y protocolos de amplificación para cada uno.
- Anexo 8. Resultados preliminares de extracción de ADN y ARN, secuenciación, ensamble de genoma y transcriptoma de *Cucurbita argyrosperma* spp. *argyrosperma*, y genomas de *C. argyrosperma* spp. *sororia* y *C. moschata*.
- Anexo 9. Variación genética de *C. argyrosperma*. Resultados preliminares obtenidos a partir del análisis de marcadores moleculares (microsatélites nucleares y tGBS) amplificados hasta el momento.
- Anexo 10. Variación genética de *C. moschata*. Resultados preliminares obtenidos a partir del análisis de marcadores moleculares (microsatélites nucleares, secuencias de mitocondria y cloroplasto) amplificados hasta el momento.

- Anexo 11. Variación genética de *C. pepo*. Resultados preliminares obtenidos a partir del análisis de marcadores moleculares (microsatélites nucleares y secuencias de cloroplasto) amplificados hasta el momento.
- Anexo 12. Variación genética de *C. maxima*. Resultados obtenidos a partir del análisis de secuencias de la región *psbJ-petA* del cloroplasto.
- Anexo 13. Análisis filogenéticos del género *Cucurbita*. Tesis de licenciatura presentada por Leslie Paredes el 2 de septiembre de 2016. Resultados obtenidos a partir del análisis de marcadores moleculares (secuencias de cloroplasto y mitocondria) para 20 taxa del género *Cucurbita*.
- Anexo 14. Carpeta comprimida que contiene los trabajos presentados en el XX Congreso Mexicano de Botánica realizado en la Ciudad de México del 4 al 9 de septiembre 2016. Se anexan: posters, presentaciones orales y constancias de participación.
- Anexo 15. Resultados preliminares de análisis de variación morfológica en *C. moschata*, *C. argyrosperma argyrosperma*, *C. argyrosperma sororia*, *C. pepo* y *C. digitata*.
- Anexo 16. Carpeta comprimida que contiene los avances los entregables indicadores de éxito de la primera etapa del proyecto. Tesis de los alumnos Gustavo García Jaramillo, Paulina Hernández Vargas y Karen Yazmín Ruíz Mondragon. Borradores de dos manuscritos de artículos científicos a someter en revistas internacionales arbitradas e indizadas.