



Centro de Investigación y de Estudios Avanzados del  
Instituto Politécnico Nacional  
Unidad de Genómica Avanzada

Genética de poblaciones del Género de Cicadas  
*Dioon* (Zamiaceae)

Tesis que presenta:  
Francisco Gabriel Pérez Zavala

Para obtener el grado de:  
Maestría en Ciencias  
en la especialidad de Biotecnología de plantas

Directora de Tesis:  
Angélica Cibrián Jaramillo

Nothing in biology makes sense except in the light of evolution.  
-Theodosius Dobzhansky (1973)

Nothing in evolution makes sense except in light of population genetics.  
-Michael Lynch (2007)



# Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología por apoyarme con la beca numero 396097, asi como al proyecto de investigación científica básica titulado: population genomics of cycads and their cyanobacteria: unraveling the evolutionary Basis of adaptation to novel and changing environments, Con numero de proyecto 169701 sin estos financiamientos este trabajo no habría sido posible.

Esta tesis va dedicada a mi papá y mamá, sin todo su apoyo, todos sus consejos, toda su ayuda y educación no habría llegado a este punto en mi vida. También al resto de mi familia por que ellos tambien me han apoyado mucho en todo.

A la Dra. Angelica por su dirección en este proyecto y por todo lo que aprendí durante la estancia en su laboratorio.

Al Dr. S. M. Rovito y el Dr. M. A. Gomez Lim por sus comentarios sobre mi proyecto.

A Flor por su paciencia, apoyo y dedicación a mi proyecto.

A mis compañeros de laboratorio (por orden alfabético) Erick, Eunice, Josean, Pablo Moo, Pablo Peláez, Rafa. También a las que ya no están Fer y Sarita.

A mis amigos fuera del laboratorio y de generación.

A los que pagan impuestos.

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. El género <i>Dioon</i> (Cycadales, Zamiaceae)	1
1.2. Taxonomía y Morfología	2
1.3. Reproducción	2
1.4. Dispersión	3
1.5. Usos humanos: Importancia económica y cultural	4
1.6. Antecedentes	4
<b>2. Hipótesis y objetivos</b>	<b>7</b>
<b>3. Metodología</b>	<b>9</b>
3.1. Colecta de Material y Extracción de ADN	9
3.2. SSRs	10
3.2.1. Diseño de Primers	10
3.2.2. PCRs	10
3.2.3. Genotipificación	11
3.2.4. Errores genotipificación y Alelos nulos	11
3.2.5. Identificación de loci bajo selección	12
3.3. Diversidad Genética	13
3.3.1. Equilibrio de Hardy-Weinberg	13
3.4. Descripción de la Estructura Poblacional	13
3.4.1. $F_{ST}$ y $G_{ST}$	13
3.4.2. $R_{ST}$	14
3.4.3. Asignamiento de individuos a grupos genéticos	15
3.4.4. Prueba de Mantel	16
3.5. Conectividad genética	16
3.5.1. Graphos	16
3.6. Detección de la reducción del tamaño poblacional	17
<b>4. Resultados</b>	<b>19</b>
4.1. Selección	19
4.2. Diversidad genética	19
4.2.1. Número de alelos, Heterocigosidad y Equilibrio de Hardy-Weinberg	19
4.2.2. Frecuencias alélicas	21
4.3. Estructura Genética	23
4.3.1. $F_{ST}$ y $R_{ST}$	23
4.3.2. STRUCTURE	24
4.3.3. STRUCTURE con submuestreo	26

4.4. Análisis de componentes principales . . . . .	27
4.5. Prueba de Mantel . . . . .	28
4.6. Grafos de poblaciones . . . . .	29
4.7. Índice $M$ para detectar cuellos de botella . . . . .	29
<b>5. Discusión</b>	<b>31</b>
5.1. Estructura genética a nivel de genero: Un gradiente . . . . .	31
5.2. ¿Reciente divergencia? . . . . .	31
5.3. Sección Noroeste-Oeste . . . . .	32
5.3.1. Evidencia de un cuello de botella en las poblaciones del norte de México . . . . .	32
5.3.2. Un complejo de especies . . . . .	33
5.3.3. Poblaciones fragmentadas y con poco flujo genético entre ellas . . . . .	33
5.4. Sección Noreste-Este-Suroeste . . . . .	34
5.4.1. Deficiencia de heterocigocidad y alta estructura dentro de cada población se explica con tres posibles escenarios . . . . .	34
5.4.2. Alto flujo genético. . . . .	35
5.5. Sección Valle de Tehuacan-Sierra Madre del Sur . . . . .	36
5.5.1. Cuello de botella y exogamia . . . . .	36
5.5.2. Aislamiento del resto de las especies . . . . .	36
<b>6. Conclusiones</b>	<b>39</b>
<b>Bibliografía</b>	<b>41</b>
<b>A. Anexos</b>	<b>47</b>
A.1. Protocolo de Extracción de ADN total con CTAB . . . . .	47
A.2. PCA . . . . .	49

# Introducción

## 1.1. El género *Dioon* (Cycadales, Zamiaceae)

Las cícadas son un grupo monofilético de plantas cuyo origen se remonta probablemente a hace 250-350Ma siendo de los grupos actuales de plantas con semilla con la presencia más antigua en el registro fósil. Se encuentran dentro del grupo de las gimnospermas, pero hay que aclarar que las cícadas forman un linaje muy diferenciado de las otras gimnospermas y que ha pasado por radiaciones adaptativas recientes (11Ma) [Nagalingum et al., 2011]. México es uno de los centros de diversidad del orden Cycadales, con 54 especies pertenecientes a tres de los géneros de la familia Zamiaceae: *Zamia*, *Ceratozamia* y *Dioon*. Dentro de las cícadas, el género *Dioon* es el que cuenta con la mayor distribución geográfica en México (figura 1.1), además de ser un género hermano al resto de los géneros de cícadas excepto el género *Cycas* [Nicolalde-Morejón et al., 2014].

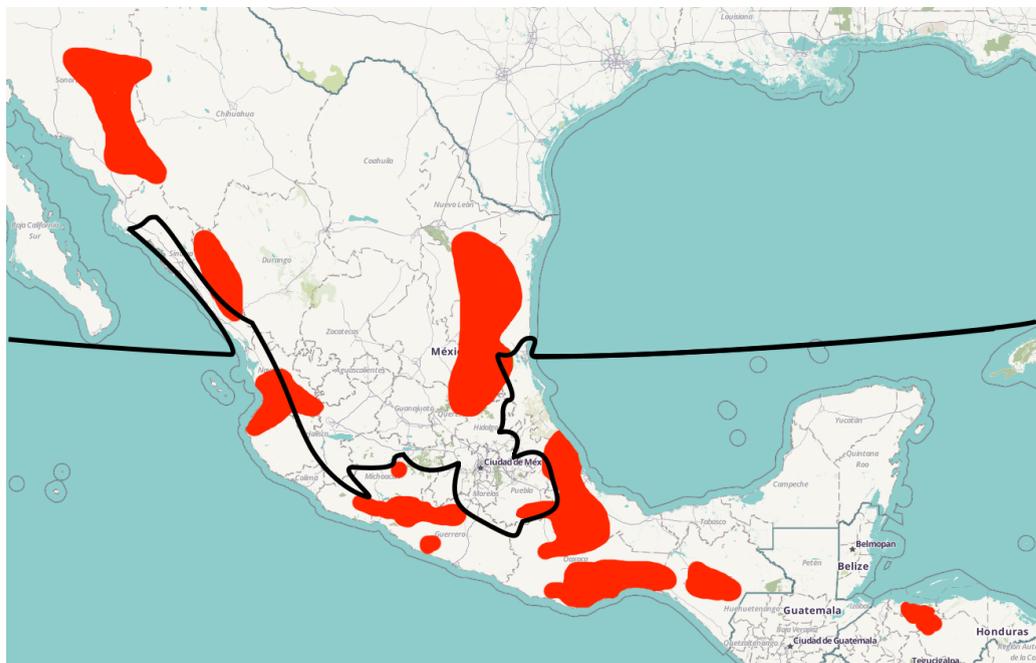


Figura 1.1: **Distribución geográfica de las 15 especies del género *Dioon*** incluye la zona de transición de las regiones Neártica y Neotropical (línea negra). *D. mejiae* es la única especie fuera de México y se localiza en la zona norte de Honduras. Figura basada en [Nicolalde-Morejón et al., 2014] y [Bonta pers. comm.].

El género *Dioon* cuenta con 15 especies descritas de las cuales 14 son endémicas de México y una de Honduras; y está distribuido a lo largo de todo México, desde el estado de Sonora hasta Chia-

pas, y en la parte norte de Honduras. En México *Dioon* se localiza principalmente en los sistemas montañosos de la Sierra Madre Oriental, Sierra Madre Occidental y Sierra Madre del Sur [González et al., 2008], aunque hay especies como *D. edule* que se puede encontrar también en playas de Veracruz.

El rango de distribución de *Dioon* incluye 12 de 19 regiones biogeográficas según Arriaga [1997] y el rango de altitud va desde el nivel del mar hasta hasta más de 3000 msnm: *Dioon* se puede encontrar desde playas de Veracruz hasta montañas, desde las zonas más xéricas hasta bosques de pino-encino [Nicolalde-Morejón et al., 2014]. Para adaptarse a ambientes tan contrastantes en los que se encuentran estas plantas, las cícadas presentan adaptaciones únicas; resistencia extrema al fuego, salinidad y sequías, además de poder germinar sobre casi cualquier sustrato incluso sobre roca desnuda, otras adaptaciones incluyen: raíces coraloides capaces de fijar nitrógeno y otros nutrientes, regeneración de la planta completa a partir de tallo o raíz, resistencia a ceniza volcánica que bloquea los estomas que casi cualquier planta, toxinas para evitar la herbívora.[Norstog et al., 1997].

## 1.2. Taxonomía y Morfología

El género *Dioon* ha experimentado cambios sustanciales desde los trabajos clásicos que solo reconocían tres especies. Fue en los últimos 30 años cuando el género *Dioon* comenzó a ser más estudiado, lo que resultó la descripción de 13 especies más, dando un total de 15 especies descritas [Nicolalde-Morejón et al., 2009] [Salas-Morales et al., 2016], sin embargo, existen autores que afirman que hay más de 15 especies y autores que sostienen no hay tantas especies como las que se reconocen. El segundo argumento se basa en que algunas características utilizadas para separar las especies como el ancho de los folíolos de la hoja o espinas en el margen de los folíolos son muy variables dentro de cada población, por ejemplo *D. edule* [González-Astorga et al., 2003]. Otra crítica es el uso de barreras políticas entre estados, como barrera para definir especies (ver: [Nicolalde-Morejón et al., 2009]).

Los caracteres morfológicos utilizados para dividir especies son: largo y ancho de folíolos, folíolos imbricados (sobrelapados), el ángulo en que se insertan los folíolos del raquis, si los folíolos tienen espinas o no, color de la pubescencia de las hojas emergentes. Estas características resultan crípticas además de que la variación intraespecie es demasiado alta para estas características [González-Astorga et al., 2003], [Perez-Zavala FG. obs. pers.].

## 1.3. Reproducción

Todas las cícadas incluyendo las del género *Dioon* son dioicas, y además del cono reproductivo, no poseen ningún otro tipo de dimorfismo sexual. A nivel cromosómico se cree que los cuerpos satélites al cromosoma tres pueden ser responsables de la determinación de sexos aunque nadie lo ha corroborado [Abraham y Mathew, 1962]. Una de las más grandes desventajas de las cícadas como modelo de estudio es su ciclo de vida tan largo, por ejemplo en el género *Dioon* se ha encontrado que la edad aproximada de reproducción es de  $\approx 500$  años, aunque hay especies como *Dioon merolae* que se reproducen a los  $\approx 100$  años [Lázaro-Zermeño et al., 2012].

Las cícadas presentan características y métodos únicos en el reino de las plantas para realizar la polinización. Hoy se sabe que los principales vectores de polinización son insectos de los órdenes Coleoptera y Thysanoptera y no el viento como es el caso para muchas otras gimnospermas [Terry et al., 2005]. Estos polinizadores se cree que pueden recorrer distancias considerables y llegar de una población de cícadas a otra. En cícadas parece no existir barreras reproductivas postcigóticas, ya que se sabe que especies del género *Dioon* se pueden cruzar con cualquier otra especie y tener descendencia,

el caso más extremo de hibridación en cícadas es el de *Zamia pumila* (Cuba) con *Encephalartos villosu* (Sudáfrica) donde la cruce de estas dos plantas generó híbridos viables y fértiles [Chamberlain, 1926].

#### 1.4. Dispersión

La dispersión de semillas del género *Dioon* es llevada a cabo principalmente por ardillas (*Sciurus spp.*), urracas (*Calocitta formosa*) [García et al., 2016], guacamaya verde (*Ara militaris*) [Chemnick, 2013] y ratón silvestre (*Peromyscus mexicanus*) [Vovides, 1990] (Figura 1.2). Se ha observado que no llevan las semillas a más de 40m de distancia de la planta madre. Hay que mencionar que todas las observaciones excepto para el ratón silvestre son derivadas de trabajos cuyo fin no era el de medir la dispersión de las cícadas y caen dentro de la categoría de pruebas anecdóticas. Hoy en día no hay ningún trabajo dedicado a determinar cual es la capacidad real de los dispersores en especies de cícadas mexicanas, el único trabajo de dispersión para la familia Zamiaceae es el de W. Tang donde encuentra que la dispersión de *Zamia pumila*, planta de Florida, Estados Unidos, es en realidad muy baja (un máximo de 4m) [Tang, 1989].

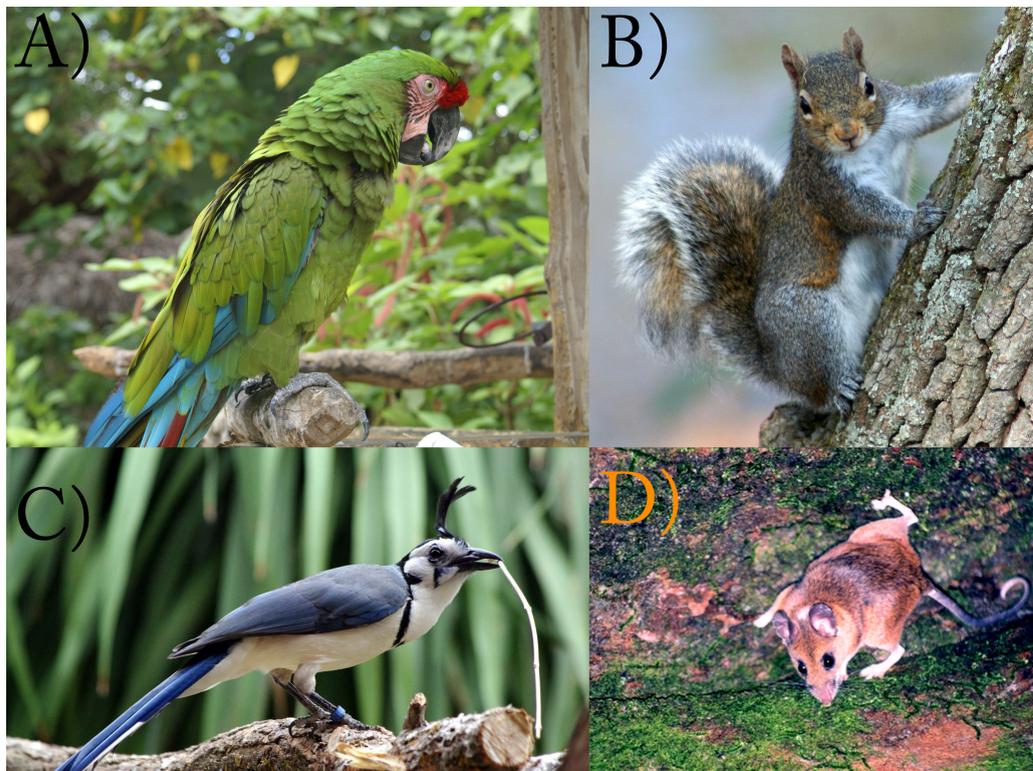


Figura 1.2: **Dispersores de semillas del género *Dioon***. Todos ellos de amplia distribución: A) Guacamaya verde (*Ara militaris*), B) Ardillas (*Sciurus spp.*), C) Urraca copetona (*Calocitta formosa*) y D) Ratón silvestre (*Peromyscus mexicanus*). Creditos de las fotos: A) Charles Pluta, B) Chiswick Chap, C) Xuan Mai y D) Juan Cruzado-Cortez.

Para el caso de *Peromyscus mexicanus* se tiene reportado que puede consumir no solo la sarcotesta, que es la cubierta carnosa de la semilla, sino también la semilla sin fallecer siempre y cuando tenga una dieta mixta de varias semillas [González-Christen, 1990]. En el mismo estudio Gonzalez-Christen [1990] reporta que en los comederos de *Peromyscus mexicanus* encuentra un 50% de semillas de *D.*

*edule* y la mitad de los comederos tienen plantulas de *D. edule*. No se menciona en este estudio nada sobre la distancia de dispersión del ratón silvestre.

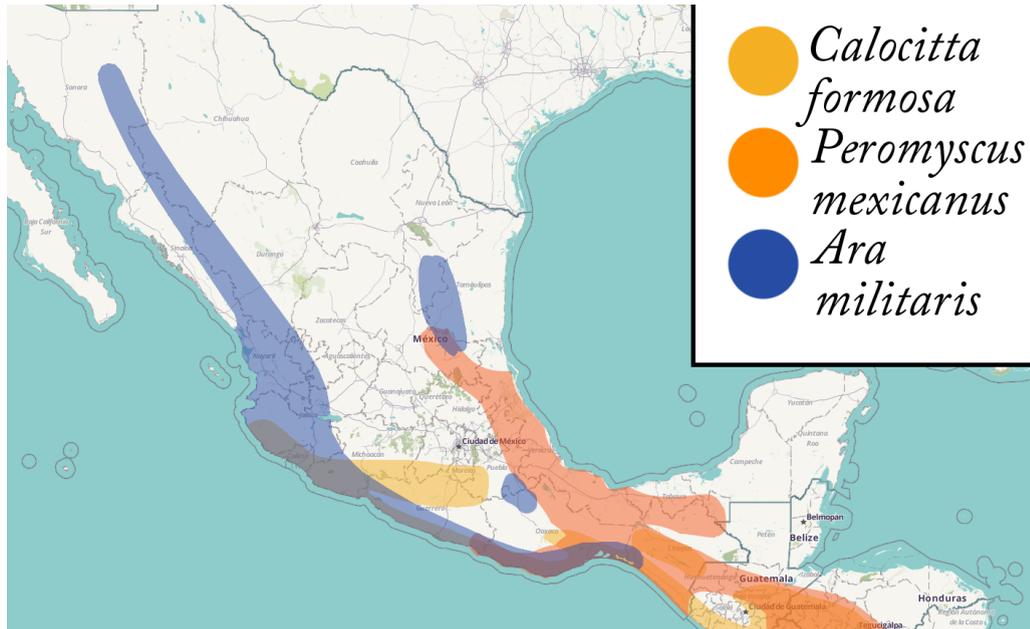


Figura 1.3: **Distribución de los dispersores de semillas del género *Dioon*.** La distribución de las ardillas del género *Sciurus* no se muestra debido a que son ubicuas en México y Centro America. Estas tres especies cubren la distribución del género *Dioon* excepto en Honduras. Ardillas del género *Sciurus* cubren toda la distribución del género. Figura basada en información obtenida de [Marateo et al., 2015, Ellis, 2010, NatureServe y Patterson, 2007a, NatureServe y Patterson, 2007b, Domínguez y Ruz, 2008].

## 1.5. Usos humanos: Importancia económica y cultural

El principal uso de las cícadas en México es la alimentación, se usan las semillas aunque también se come la sarcotesta, el tallo y la raíz. Para evitar la intoxicación, que puede llegar a ser letal, las personas que consumen las semillas de cícadas deben nixtamalizar las semillas para eliminar las toxinas que contienen [Bonta et al., 2006]. Otro uso muy importante del género *Dioon* es en ceremonias católicas sincreticas como el día de los muertos o la fiesta de la Santa Cruz en Chiapas [Pérez-Farrera y Vovides, 2006, Bonta y Osborne, 2005] donde se usan las hojas para decorar altares.

Un rasgo de la importancia cultural de cícadas de la familia Zamiaceae en México, son las grandes analogías con el maíz. En muchas de las cícadas mexicanas incluyendo las del genero *Dioon* se pueden encontrar referencias al maíz en sus nombre comunes, por ejemplo, “viejo maíz”, “amo del maíz”, “pastor del maíz”, “teocinte” y términos similares. En México y Honduras las comidas preparadas con semillas de *Dioon* suelen recibir nombres idénticos a los de la comida preparada con maíz, por ejemplo “tamal”, “tortilla”, “atole”, etc. [Bonta y Osborne, 2005].

## 1.6. Antecedentes

Trabajos previos a éste estudio con enfoque de genética de poblaciones para *Dioon* hay siete que cubren ocho de las 15 especies; *D. edule* [Gonzalez-Astorga et al., 2003, Rubio-Mendez, 2013], *D.*

*angustifolium* [Gonzalez-Astorga et al., 2005], *D. sonorence*, *D. tomasellii* (algunas poblaciones hoy *D. stevensonii*) y *D. holmgrenii* [González-Astorga et al., 2008, Gonzalez-Astorga et al., 2009] y para *D. caputoi* y *D. merolae* [Cabrera-Toledo et al., 2008, Cabrera-Toledo et al., 2010]. De forma general todos los trabajos reportan patrones de diversidad genética complejos, que se explicarían por migración entre especies *D. sonorence*, *D. tomaselli*, *D. holmgrenii* y *D. edule*. Así como cuellos de botella en especies geográficamente aisladas como, *D. caputoi* y *D. sonorence*. Rubio-Mendez [2013] encuentra alelos nuevos en plantulas de *D. edule* que no se encuentran en plantas adultas, lo que se podría explicarse por alta migración y/o dispersión de polen.

En las especies de la Sierra Madre Occidental, *D. sonorence*, *D. tomasellii* y *D. stevensonii*, encontramos que las poblaciones están fragmentadas y perturbadas, siendo *D. sonorence* la más aislada y perturbada pero también la más diversa. Para las especies del centro de México y la Sierra Madre Oriental, *D. holmgrenii*, *D. edule* y *D. angustifolium*, encontramos que estas especies tienen exceso de heterocigotos y presentan conectividad genética entre poblaciones complejas. Para las especies *D. caputoi* y *D. merolae* los antecedentes mencionan que hay poca diferenciación entre ellas.

Tabla 1.1: Resumen de los antecedentes de este trabajo.

Especie	Resultado	Estudio
<i>D. sonorence</i>	Alta diversidad relativa a <i>D. tomasellii</i> , <i>D. stevensonii</i> y <i>D. holmgrenii</i> .	González-Astorga, 2008
	Poblaciones fragmentadas y perturbadas. Gradientes de frecuencias alélicas de norte a sur.	González-Astorga, 2009
<i>D. tomasellii</i>	Poblaciones genéticamente divididas por montañas, diversidad intermedia entre <i>D. sonorence</i> y <i>D. holmgrenii</i> .	González-Astorga, 2008
<i>D. stevensonii</i>	(Antes <i>D. tomaselli</i> ). Poblaciones al sur del Eje Volcánico Transversal altamente diversas.	González-Astorga, 2008
<i>D. holmgrenii</i>	Exceso de heterocigotos, se discute el posible origen híbrido de esta especie.	González-Astorga, 2008
<i>D. edule</i>	Exceso de heterocigotos conectividad genética entre poblaciones compleja: posible flujo genético o migración.	González-Astorga, 2003
	Alelos nuevos (no presentes en adultos) en juveniles y plantulas.	Rubio-Mendez, 2013
<i>D. angustifolium</i>	(Antes <i>D. edule</i> ). Conectividad genética entre poblaciones compleja: posible flujo genético o migración.	González-Astorga, 2003
		González-Astorga, 2005
<i>D. caputoi</i>	Evidencia de cuello de botella y exogamia en plantulas. Poca diferenciación entre poblaciones relativa a otras especies de la familia Zamiaceae.	Cabrera-Toledo, 2008
	Poca diferenciación con <i>D. merolae</i> .	Cabrera-Toledo, 2010
<i>D. merolae</i>	La poca diferenciación con <i>D. caputoi</i> tal vez es debido a flujo genético o historias compartidas recientes.	Cabrera-Toledo, 2010

Como se mencionó antes el género *Dioon* es un modelo atractivo por su amplia distribución, adaptaciones y presencia en la cultura humana sin embargo este es el primer estudio con el enfoque de genética de poblaciones hecho a nivel de toda la distribución el género para *Dioon*, el cual servirá para entender la ecología y la dinámica de las poblaciones de este género, entender esto nos permitirá

## 6 FRANCISCO GABRIEL PÉREZ ZAVALA

plantear hipótesis evolutivas y filogeográficas, que nos permitirán entender la historia natural de las cícadas en México.

# Hipótesis y objetivos

## Hipótesis

Las especies del género *Dioon* presentan una alta conectividad genética en sus poblaciones, probablemente por dispersión tanto de polen como semillas o reciente diversificación.

## Objetivo general

Describir la diversidad, estructura y conectividad genética del género *Dioon*.

## Objetivos específicos

1. Generar una colección de muestras género *Dioon*, cubriendo lo más posible las 15 especies y su distribución.
2. Diseñar marcadores (microsatellites) transferibles a todas las especies del género *Dioon* a partir del transcriptoma de *D. edule*.
3. Genotipificar la colección de muestras generadas así como corregir posibles errores de genotipificación.
4. Realizar pruebas de neutralidad en los marcadores, calcular la diversidad genética a partir de los marcadores seleccionados y calcular si los marcadores seleccionados están en equilibrio de Hardy-Weinberg
5. Calcular los índices  $F_{ST}$  y  $R_{ST}$  así como utilizar métodos de agrupamiento para medir la estructura genética, así como generar redes de conexiones genéticas.
6. Calcular el índice  $M$  de Garza-Williamsom para detectar cuellos de botella en las poblaciones.



# Metodología

## 3.1. Colecta de Material y Extracción de ADN

Se colectó tejido foliar de un total de 483 individuos de 30 poblaciones en México y una población en Honduras cubriendo 14 de las 15 especies reconocidas para el género *Dioon* a lo largo de toda su distribución. Los puntos de colecta se muestran en la figura 3.1. No se pudieron obtener muestras de la especie *D. planifolium*.

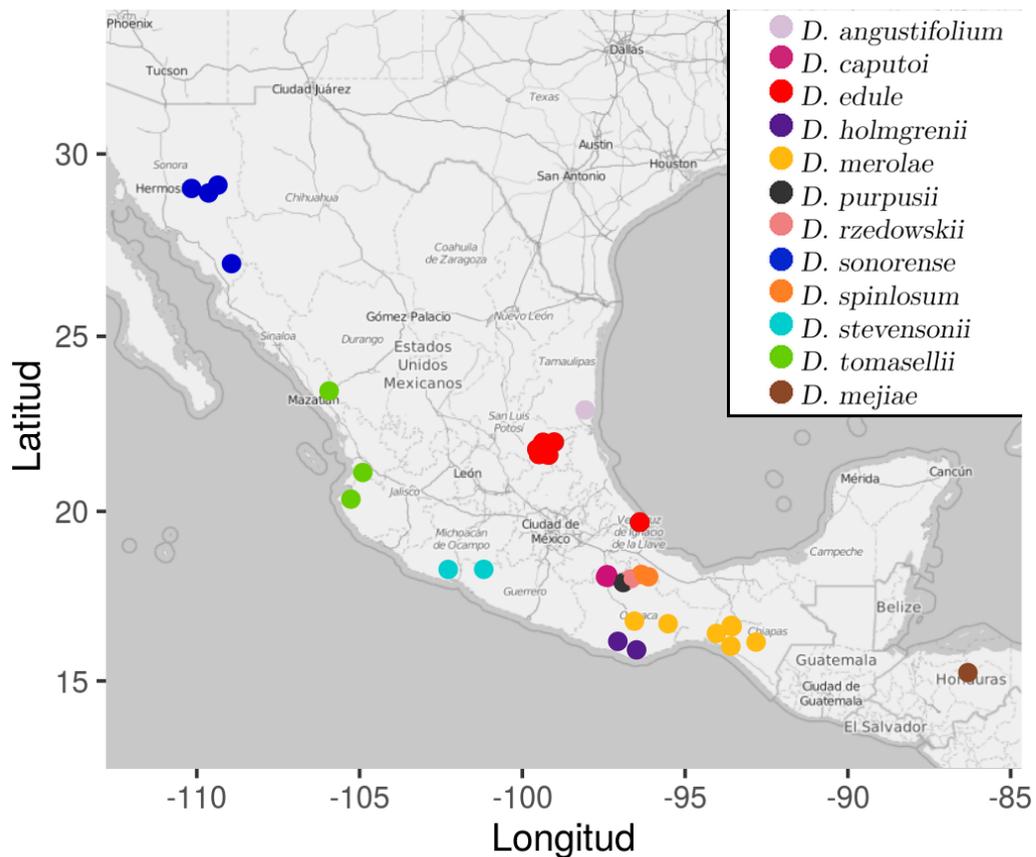


Figura 3.1: **Puntos de Muestreo**, las poblaciones muestreadas cubren toda la distribución del género *Dioon* y todas 14 de las 15 especies.

Para la colecta de tejido foliar se dio preferencia a las hojas más jóvenes y en mejor estado, evitando siempre que fuera posible las que presentaban daños visibles. Las muestras de cada individuo fueron

transportadas en sobres de papel con silica en gel hasta su almacenamiento final a  $-80^{\circ}\text{C}$ . La extracción de ADN se realizó de acuerdo al método de extracción con CTAB de Doyle [1991] con modificaciones descritas en los Anexos (A.1).

### 3.2. SSRs

Los microsatélites, llamados STRs (**S**hort **T**andem **R**epeats) o SSRs (**S**imple **S**equences **R**epeats), son secuencias repetidas en tándem, los motivos repetidos son cortos de entre 2-6 pb, y con una región de repetidos  $> 100$  pb. Los SSRs son altamente polimórficos, abundantes y distribuidos uniformemente en la eucromatina [Schlötterer, 2004], no solo son secuencias satélites a los genes como se creía. La tasa de mutación de los microsatélites varía entre  $10^{-2}$  y  $10^{-6}$  por locus por generación, pero se asume generalmente una tasa de mutación de  $10^{-3}$  por locus por generación [Weber y Wong, 1993, Jarne y Lagoda, 1996]. La tasa de mutación de los SSRs no solo varía entre loci sino también entre distintos motivos y tipos de SSRs [Bachtrog et al., 2000], tamaño del tándem, con los tándems más largos mutando más rápido que los cortos [Jin et al., 1996, Wierdl et al., 1997, Schlötterer et al., 1998], incluso varía entre distintas especies, sexo y edad [Ellegren, 2000]. Una tasa de mutación tan cambiante se debe a un mecanismo de mutación complejo dado por deslizamiento o “slippage”, mecanismo que no está completamente entendido hasta el día de hoy [Ellegren, 2000, Ellegren, 2004].

El alto polimorfismo de los SSRs y su alta tasa de mutación, hace a los SSRs excelentes candidatos para detectar la estructura a tiempos recientes de divergencia, necesitando muy pocos para identificar el número de poblaciones reales [Haasl y Payseur, 2011].

#### 3.2.1. Diseño de Primers

Para diseñar los primers se utilizó el transcriptoma de *D. edule* obtenido a través del OneKP project ([www.onekp.com](http://www.onekp.com)), un total de 121,771 contigs fueron analizados buscando repetidos en tándem usando el algoritmo mreps v2.5 [Kolpakov y Kucherov, 1999]. Se buscaron tandems de cinco o más repetidos priorizando repetidos di- tri- o tetranucleótidos con flancos 3' y 5' de 15-30 nucleótidos para el diseño de primers. Los primers fueron diseñados con el programa Primer3Plus [Untergasser et al., 2007] después se probó la formación de heterodímeros y auto alineamiento con el programa OligoAnalyser v3.1 ([www.idtdna.com/calc/analyser](http://www.idtdna.com/calc/analyser)) (Integrated DNA Technologies, Coralville, Iowa, USA). Ochenta de 150 SSRs identificados en el transcriptoma de *D. edule* tenían sitios de unión de primers candidatos. Se diseñaron primers para los 80 loci.

#### 3.2.2. PCRs

Las reacciones se planearon de acuerdo al método de Schuelke [2000] para marcar fragmentos de PCR. Los PCRs se hicieron en un volumen total de  $15\mu\text{L}$  los cuales contenían  $1.5\mu\text{L}$  de buffer de PCR 10x,  $0.3\mu\text{L}$  de dNTPs 10mM,  $0.9\mu\text{L}$  de  $\text{MgCl}_2$ ,  $0.09\mu\text{L}$  del primer forward 10mM,  $0.36\mu\text{L}$  del primer Reverse 10mM,  $0.36\mu\text{L}$  del primer M13(-21) marcado 10mM, 0.25U de Taq,  $1\mu\text{L}$  de ADN templado de aproximadamente  $100\text{ng}/\mu\text{L}$  y  $10.24\mu\text{L}$  de  $\text{H}_2\text{O}$ .

Las condiciones de PCR fueron:  $95^{\circ}\text{C}$  por 2 min; seguidos de 30 ciclos de:  $95^{\circ}\text{C}$  por 30 s,  $55-58^{\circ}\text{C}$  por 30 s (temperatura de acuerdo a la  $T_m$  de cada par de primers), y  $72^{\circ}\text{C}$  por 1 min; luego ocho ciclos a:  $95^{\circ}\text{C}$  por 30 s,  $53^{\circ}\text{C}$  por 30 s, y  $72^{\circ}\text{C}$  por 1 min; terminando con  $70^{\circ}\text{C}$  por 10 min. Se probaron 50 de los candidatos en muestras de *D. edule*, *D. holmgrenii*, *D. spinulosum*, *D. sonorensis* y *D. angustifolium* y se eligieron 21 loci los cuales produjeron bandas de manera consistente en las especies probadas.

### 3.2.3. Genotipificación

Se enviaron a genotipar un total de 483 individuos, de los 21 loci previamente seleccionados solo se eligieron 15 que amplificaron con todas las especies, y dieron resultados consistentes en después del análisis de los fragmentos. La genotipificación se llevo a cabo en Macrogen Inc., Korea. Los resultados de las corridas fueron analizados en el programa Geneious v8.1.7 (<http://www.geneious.com>, [Kearse et al., 2012]). Los primers y loci usados en este trabajo se encuentran en la tabla 3.1.

Tabla 3.1: Lista de loci utilizados para este trabajo

Locus	ID	Secuencia	Patron	Tandems	TM	Tamaño esperado
2014311	P10	F: TTAGGGGGCCGAGGAG R: CCGACCGCCAAGAGAA	CCA	8	58	243pb
2015232	P12	F: TGGAGATCAACAACGACCAA R: CGAGACCCAGAGARCCTGAC	CAT	8	58	234pb
2018276	P17	F: CACCCTGCCAAAGGTCAT R: GCACCCATTGTTGGACA	TG	14	58	250pb
2019766	P18	F: CGTGCGACCAGCAAGA R: TCGCCGGTGAGAACAAG	TCG	8	58	234pb
2015331	P21	F: CCATCCCTCCATCTTCACA R: CACCCCTTCGACACG	CCT	8	55	207pb
2018116	P22	F: GGCAGATTAGCTCCAGCAG R: CAGAGCTTCCAATTCCTTGC	GAC	8	55	188pb
2018893	P23	F: GCGGTAGCTGGAGAGGTTC R: AGTCTGGGGCCTCATCAAC	GAA	10	55	245pb
2019765	P25	F: CAAATTCCTGTGGGAGATGG R: GCAGGCAGTTTGAAAGAAC	AGC	9	54.5	159pb
2001304	De1	F: CTGGGCCTCGACATAACATT R: CTGGGCCTCGACATAACATT	ATAA	5	56	173pb
2001955	De3	F: CTGCCGAGGAGGGACA R: CGCAGGGTTGGAGAGC	TGC	8	56	195pb
2002349	De5	F: AGAGCTGCTTCCACGTTTCAT R: GCGGAAACTTCTTCAACAGC	TCT	7	56	169pb
2002757	De6	F: TGGGAAATGCACACCTAAAA R: ACCTGGGCCACTTGAGG	CA	20	56	236pb
2003643	De7	F: CGAACTTGAAGACGATGACG R: CGGGTAGCACCAAAGATTGT	GCA	7	56	227pb
2011833	P7	F: ATCGATCCGCTGTTCCACC R: CATCTGCGGTTGCATGAC	GCA	6	55	229pb
2013566	P9	F: TGCAAACATGGTCCTTGG R: TGGTGGCATTTCAGAAGCA	TCT	6	55	246pb

### 3.2.4. Errores genotipificación y Alelos nulos

Errores de genotipificación pueden ser causados por concentraciones bajas de ADN templado durante el PCR [Wandeler et al., 2003], ADN degradado o por mutaciones en el sitio de unión de los primers [Shaw et al., 1999], incluso por la dominancia de un alelos sobre otro o amplificación preferencial [Wattier et al., 1998].

Tales errores en los genotipos pueden causar desviaciones del Equilibrio de Hardy-Weinberg, en particular una deficiencia de heterocigotos [Shaw et al., 1999], potencialmente sesgando los análisis de genética de poblaciones. Estas desviaciones son a menudo muy similares a las causadas por la endogamia, desviaciones de panmixia o el efecto Wahlund. Sin embargo, el dominio un alelo sobre otro, los alelos nulos, y el “stuttering” o tartamudeo, también tienen sus propias “firmas alélicas”. Por lo tanto, es posible discriminar entre las desviaciones debidas a la no panmixia y las causadas por los diversos errores de genotipado [Van Oosterhout et al., 2004].

Para identificar y corregir errores en los genotipos en los datos de los SSRs, se utilizó el programa Micro-Checker v2.2.3 [Van Oosterhout et al., 2004] con los parámetros default. Se corrigieron las frecuencias de alélicas usando el algoritmo Brookfield 1 [Brookfield, 1996] implementado en el programa el programa Micro-Checker v2.2.3 siempre que se detectaron nulos. El programa se corrió usando los genotipos totales y tratando todos los individuos como una sola población, usando locus por locus por población, y usando los genotipos totales por población, no se obtuvieron resultados diferentes en ninguna de las corridas.

### 3.2.5. Identificación de loci bajo selección

Una tarea previa a la mayoría de los análisis es separar los loci bajo selección de los loci neutrales. Ignorar el efecto de la selección en estudios poblacionales y evolutivos puede llevar a predicciones y estimaciones erróneas, además que distinguir loci bajo selección y neutrales puede llevar a entender el proceso de especiación por el que atraviesa una especie, para este fin se utilizó el programa BayeScan v2.1 [Foll y Gaggiotti, 2008].

El programa es capaz de identificar loci candidatos de estar bajo selección, usando las diferencias de frecuencias alélicas entre poblaciones. La selección es medida por la descomposición de  $F_{ST}$  en un componente población-específico ( $\beta$ ) compartida por todos los loci, y un componente locus-específico ( $\alpha$ ) compartido por todas las poblaciones utilizando una regresión logística. Desviaciones de la neutralidad en un locus dado se asumen cuando es necesario el componente específico del locus ( $\alpha$ ) para explicar el patrón observado de la diversidad ( $\alpha$  significativamente diferente de 0). Un valor positivo de alfa sugiere la selección diversificadora, mientras que los valores negativos indican selección estabilizadora [Foll y Gaggiotti, 2008].

El programa se corrió con parámetros default (periodos de “burn-in” más largos no alteraban los resultados). Una primera corrida del programa detectó 7 de los 15 loci con evidencia decisiva (según la escala de Jeffrey [Rouder et al., 2009]), pero según Foll y Gaggiotti [2008] esto puede ser por una violación al modelo demográfico asumido por el método de BayeScan, o aislamiento de una o varias poblaciones. Foll y Gaggiotti recomiendan en estos casos, hacer varios análisis entre poblaciones semejantes en vez de un solo análisis global.

El análisis se repitió tomando como conjunto de datos poblaciones especies hermanas según las filogenias de Gonzalez [2008] y Gutierrez-Ortega [Datos sin publicar].

### 3.3. Diversidad Genética

#### 3.3.1. Equilibrio de Hardy-Weinberg

El equilibrio de Hardy-Weinberg es un principio que indica que la variación genética (o mejor dicho frecuencias alélicas) en una población permanecerá constante de una generación a la siguiente en ausencia de factores perturbadores (como selección o mutación). Cuando el apareamiento es al azar en una población grande, la ley prevé que ambas frecuencias alélicas y genotípicas se mantienen constantes de una generación a otra porque están en equilibrio [Hardy, 2003], [Weinberg, 1908]. El equilibrio de Hardy-Weinberg sirve para idealizar una población que nos servirá como hipótesis nula de que no hay fuerzas perturbando una población.

El método para comprobar si una población está en equilibrio de Hardy-Weinberg es simplemente comparar las frecuencias de una población contra las frecuencias esperadas según éste. Para éste fin se hizo una  $\chi^2$  utilizando los paquetes de R *gstudio* [Dyer, 2014a] y *pegas* [Paradis, 2010].

### 3.4. Descripción de la Estructura Poblacional

En esencia, la genética de poblaciones busca patrones en las frecuencias alélicas, patrones de discontinuidad que pudieran indicar barreras a la dispersión, patrones asociados a una única subpoblación que podrían representar adaptación local, patrones con huellas de selección, etc. Estos patrones son la estructura poblacional, o dicho de otra manera, es la diferencia sistemática en las frecuencias alélicas entre subpoblaciones en una población.

Poblaciones de la mayoría de las especies (si no es que todas) muestran algún nivel de estructura poblacional, por lo que es crucial para la conservación y manejo de especies. También resulta útil para hacer inferencias sobre la historia demográfica de las poblaciones (como eventos de migración, reducciones recientes de los tamaños poblacionales, colonización de nuevos ambientes). Sin embargo, aunque la definición resulta simple, aún no hay consenso de como se debe medir la estructura poblacional.

#### 3.4.1. $F_{ST}$ y $G_{ST}$

Los índices  $F_{ST}$  y  $G_{ST}$  son probablemente las medidas de estructura poblacional más usadas. El  $F_{ST}$  fue desarrollado por Sewall Wright [Wright, 1921, Wright, 1943, Wright, 1949] para describir patrones observados en loci con dos alelos de dos subpoblaciones, bajo estas condiciones:  $F_{ST} = 0$  cuando las dos poblaciones son idénticas genéticamente (no hay estructura poblacional) y  $F_{ST} = 1$  cuando completamente diferentes (no hay alelos compartidos entre las subpoblaciones).

$$F_{ST} = \frac{\sigma^2[p]}{\bar{p}(1 - \bar{p})}$$

Donde  $\sigma^2[p]$  es la varianza de las frecuencias alélicas entre subpoblaciones, y  $\bar{p}$  es el promedio de las frecuencias alélicas de todas las subpoblaciones.

El  $F_{ST}$  tiene una gran ventaja en términos de familiaridad: **1)** Ha existido el tiempo suficiente como

para que tengamos idea de cuales son sus propiedades: Aumenta con la deriva genética pero se reduce con la migración. La selección sobre la metapoblación reduce el  $F_{ST}$ , pero si la selección solo actúa solo sobre una subpoblación tiende a aumentarlo. **2)** Hay tantos trabajos que han usado el  $F_{ST}$  (y muchos más que lo seguirán haciendo), que siempre se tendrá una referencia para comparar [Whitlock, 2011].

El  $G_{ST}$  es un índice similar al  $F_{ST}$ , derivado por Masatoshi Nei [Nei, 1973] para tratar con múltiples alelos en un locus. Excepto por detalles en el proceso de estimación  $G_{ST} = F_{ST}$ .

$$G_{ST} = \frac{(H_T - H_S)}{H_T}$$

El índice  $G_{ST}$  es definido en términos de  $H_S$  (Heterocigosidad de la subpoblación) y  $H_T$  (Heterocigosidad de la metapoblación). El problema con el índice  $G_{ST}$  (recuerda que bajo ciertas condiciones  $G_{ST} = F_{ST}$ , incluso en mucha literatura se usa  $G_{ST}$  y  $F_{ST}$  de manera intercambiable), es que numerosos trabajos recientes [Hedrick, 1999, Hedrick, 2005, Jost, 2008, Jost, 2009, Meirmans y Hedrick, 2011] han señalado una dificultad en la interpretación de  $G_{ST}$ . Con dos poblaciones y dos alelos,  $G_{ST}$  oscila de 0 a 1, como se espera. Sin embargo, si se utilizan marcadores muy polimorficos como los SSRs,  $G_{ST}$  no puede llegar a 1, incluso cuando no hay alelos son compartidos entre las dos poblaciones, ya que siempre habrá algo de heterocigosidad en las poblaciones. Esto no es una falla del  $G_{ST}$  pero hay que tener cuidado al interpretarlo o al compararlo entre diferentes tipos de marcadores. A pesar de esto se recomienda siempre presentar el  $F_{ST}$  o  $G_{ST}$  por la ventajas en cuanto a familiaridad [Meirmans y Hedrick, 2011].

Para Calcular el  $F_{ST}$  se utilizó el paquete R *gstudio* [Dyer, 2014a]. El  $F_{ST}$  se calculo de manera de manera pareada (una subpoblación vs otra subpoblación, no población vs metapoblación).

### 3.4.2. $R_{ST}$

El índice  $R_{ST}$  fue creado por M. Slatkin [1995] como un analogo del  $F_{ST}$  para SSRs debido a que algunas asunciones del  $F_{ST}$  no aplican para los SSRs ya que tienen otro modelo de mutacion.

$$R_{ST} = \frac{S - S_w}{S}$$

Donde  $S$  es la diferencia promedio al cuadrado de tamaños de alelos entre todos los pares de alelos y  $S_w$  la suma promedio de los cuadrados de las diferencias en tamaños de alelos dentro de la subpoblación. Teóricamente  $R_{ST}$  es más adecuado que  $F_{ST}$  cuando se usan SSRs pero los trabajos revisados por Balloux y Lugon-Moulin [2002] encuentran que el  $R_{ST}$  es mejor para medir estructura y divergencia interespecifica y el  $F_{ST}$  es mejor para diferenciación intraespecifica.

El  $R_{ST}$  se calculo en el programa Arlequin v3.5.2.2 [Excoffier y Lischer, 2010] de manera pareada (una subpoblación vs otra subpoblación, no población vs metapoblación).

### 3.4.3. Asignamiento de individuos a grupos genéticos

El programa *STRUCTURE* implementa un método de agrupamiento basado en modelos que infieren la estructura poblacional utilizando la información de los genotipos. El método es descrito en el artículo [Pritchard et al., 2000] y actualizado en [Falush et al., 2003] y [Falush et al., 2007]. Las aplicaciones de éste método incluyen la detección de estructura poblacional, la identificación de poblaciones genéticamente distintas, la asignación de los individuos a poblaciones genéticas, e identificación de migrantes y de individuos con genotipos mezclados de dos o más poblaciones. El método de *STRUCTURE* asume un modelo donde existen  $K$  poblaciones ( $K$  generalmente es desconocida, es dada por el usuario y luego evaluada estadísticamente) y cada población es caracterizada por un conjunto de frecuencias alélicas para cada locus. Luego cada individuo es asignado probabilísticamente a una población, dos o más poblaciones de acuerdo al genotipo del individuo. Éste tipo de aproximaciones asume que los marcadores utilizados no están ligados o que están débilmente ligados. El modelo no asume ningún proceso de mutación, por lo que puede ser utilizado con todo tipo de marcadores.

Los parámetros con que se corrió *STRUCTURE* fueron: un periodo de “burn-in” de 100,000, un periodo de 200,000 repeticiones después del periodo de “burn-in”, con el modelo Admixture, sin ningún tipo de prior además de los genotipos. El análisis se corrió con valores de  $K$  de 1 hasta 16, con 20 iteraciones para cada valor de  $K$ .

Para estimar el valor de  $K$  real se usó el método descrito en [Evanno et al., 2005] e implementado en Structure Harvester [Earl et al., 2012]. El método consta de 4 pasos: 1) se calcula el promedio de la probabilidad de  $L(K)$  de 20 corridas de *STRUCTURE*. 2) se calcula la diferencia promedio entre valores sucesivos de  $K$ ,  $L'(K) = L(K) - L(K - 1)$ . Esta diferencia es el ritmo de cambio de la probabilidad con respecto a  $K$ . 3) se calcula la diferencia absoluta entre valores sucesivos de  $L'(K)$ . Esta diferencia corresponde a la velocidad de cambio en segundo orden de  $L(K)$ ,  $L''(K) = L'(K) - L'(K - 1)$ . Por último se calcula  $\Delta K$  como el promedio de los valores absolutos de  $L''(K)$  entre la desviación estándar de  $L(K)$ ,  $\Delta K = m(-L''(K)) / sd[L(K)]$ . Evanno encontró que el valor modal de la distribución de  $\Delta K$  casi siempre detecta el valor real de  $K$  o el  $K$  que representa el nivel más alto de estructura poblacional.

Algunos problemas conocidos de éste tipo de aproximaciones son: **1)** éste tipo de métodos están destinados a encontrar particiones bien definidas en los datos, sin embargo no siempre existen grupos bien definidos en poblaciones naturales y se sabe que programas como *STRUCTURE* tienen un desempeño relativamente pobre cuando hay conectividad entre poblaciones a lo largo de un gradiente [Pritchard et al., 2007]. **2)** La robustez del programa se ve comprometida gravemente cuando el muestreo no es equilibrado [Puechmaille, 2016]. Por ejemplo, si se muestrean tres poblaciones genéticamente distintas, pero con tamaños de muestra desiguales ( $N=30, 12$  y  $12$ ). Es probable que las poblaciones con tamaño de muestra pequeño se agrupen juntas (aunque sean genéticamente muy diferentes) y que la población con un tamaño de muestra grande se divida en múltiples grupos (aunque esté formada por individuos genéticamente muy similares).

Ya que nuestros datos presentan una varianza muy elevada en el número de muestras por especie y por población y un primer análisis de *STRUCTURE* mostró diferencias en tamaños en los clusters muy grandes (ver sección 4.3.2), se decidió submuestrear los grupos genéticos con número de muestras muy elevado tratando de reducir lo más posible la varianza entre muestreos.

Usando solo 335 de las 483 muestras disponibles se volvieron a correr los análisis de *STRUCTURE*.

RE. Puechmaille también recomienda usar otro tipo de estimadores además de  $\Delta K$  sin embargo los estimadores que él recomienda solo son aplicables cuando se corre el programa STRUCTURE usando la información de que localidad es cada muestra como prior.

#### 3.4.4. Prueba de Mantel

En esencia la prueba de Mantel es una prueba estadística que dice si dos matrices se encuentran correlacionadas. En el caso de genética de poblaciones una prueba de Mantel entre una matriz de diferenciación genética y una matriz de distancia genética puede revelar la presencia de aislamiento de poblaciones debido a la distancia. A pesar de su utilidad la prueba de mantel no puede ser utilizada para cuantificar la variación explicada por distancia [Manel et al., 2003].

Se realizaron cuatro pruebas de Mantel: una con todas las poblaciones, y tres dividiendo los grupos según las filogenias de Gonzalez [2008] y Gutierrez-Ortega [Datos sin publicar]. Las pruebas se realizaron con los paquetes de R Analyses of Phylogenetics and Evolution (ape) v3.5 [Paradis et al., 2004] y Analysis of Ecological Data : Exploratory and Euclidean Methods in Environmental Sciences (ade4) v1.7-4 [Chessel et al., 2009].

### 3.5. Conectividad genetica

#### 3.5.1. Graphos

Típicamente, un grafo se representa gráficamente como un conjunto de puntos (vértices o nodos) unidos por líneas (aristas). Desde un punto de vista práctico, los grafos permiten estudiar las interrelaciones entre unidades que interactúan unas con otras. Por ejemplo, una red de computadoras puede representarse y estudiarse mediante un grafo, en el cual los vértices representan computadoras y las aristas representan conexiones (las cuales, a su vez, pueden ser cables o conexiones inalámbricas) [Trudeau, 2013]. En el contexto de genética de poblaciones cada vértice es una población y las aristas pueden ser interpretadas como migración/flujo genético.

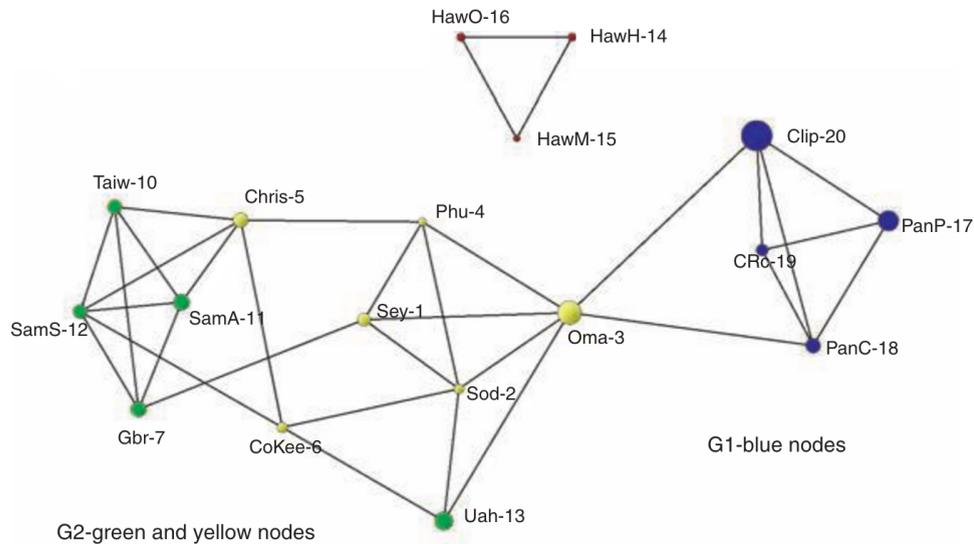


Figura 3.2: **Ejemplo de un grafo poblacional del pez de amplia distribución *Scarus rubroviolaceus***: rojo = Islas Hawaianas, azul = Pacífico del este, verde = Pacífico centro/occidental, amarillo = Océano Índico. Como se ve en la figura los nodos de las poblaciones hawaianas (rojo) se encuentran aisladas del resto (muy alta estructura poblacional), las poblaciones del Pacífico del este (azul) presentan alta estructura poblacional con las poblaciones del Pacífico Centro/occidental y Océano Índico y viceversa. Estas últimas poblaciones están diferenciadas entre sí aunque no tanto como con las demás. Figura tomada de [Fitzpatrick et al., 2011].

Los grafos de poblaciones son un método desarrollado por Dyer y Nason [2004], desde un punto de vista general muestran las conexiones entre poblaciones pero hay que tener en cuenta varias cosas para analizar un grafo. El tamaño de cada nodo o población, es proporcional al componente de varianza genética dentro de la población y el tamaño o distancia de las aristas es proporcional a su varianza interpoblacional. En otras palabras: cuando un nodo es muy grande quiere decir que su variación genética se explica por individuos de la misma población y viceversa, cuando una arista es muy pequeña entre dos poblaciones quiere decir que la varianza de esas dos poblaciones se explica por individuos de la población con la que está conectada, (esto se podría interpretar como migración/genotipos muy similares, pero hay que tener en cuenta que un grafo de poblaciones no mide dirección de la migración).

Los grafos se realizaron con los paquetes de R *genetic studio* (*gstudio*) y *popgraph* [Dyer, 2014b] con parámetros default excepto por el nivel de alfa (que es el nivel de significancia con que se mantienen las aristas), se hicieron grafos con un alfa de 0.5 (default) y de 0.1.

### 3.6. Detección de la reducción del tamaño poblacional

Después de una reducción en una población, en los loci con SSRs, se espera que la velocidad con que se pierden alelos sea mayor a la velocidad con que el rango de alelos disminuye. Garza y Williamson usaron esta propiedad única de los SSRs para detectar cambios en los tamaños poblacionales [Garza y Williamson, 2001]. El índice  $M$  que ellos propusieron se define como:

$$M = \frac{k}{r}$$

Donde  $k$  es el número de alelos y  $r$  es el rango de alelos de SSRs. Garza y Williamson señalan que un valor de  $M < 6.8$  es una fuerte evidencia de que la población sufrió un cuello de botella reciente. aunque por la manera en que se comporta el índice  $M$  valores un poco más grandes podrían indicar tres cosas: el cuello de botella es extremadamente reciente y el índice  $M$  aún no llega a su punto mínimo (aún no se pierden algunos alelos raros por deriva génica) o lo contrario el cuello de botella es muy antiguo o no fue suficientemente fuerte y la población esta en proceso de recuperación (figura 3.3).

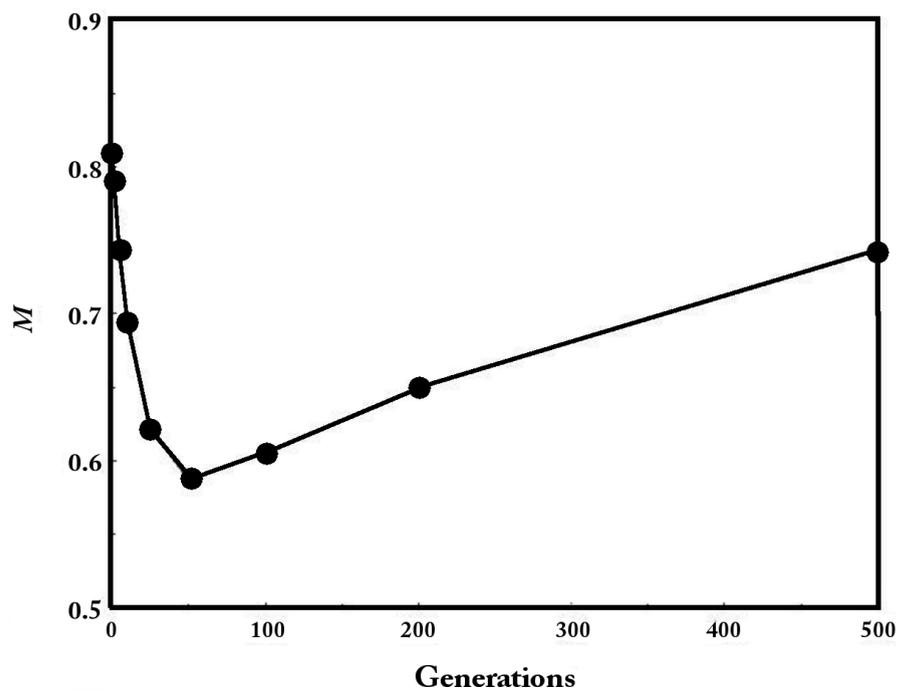


Figura 3.3: **Recuperación del índice  $M$  después de un cuello de botella severo:** incluso después de 500 generaciones el índice  $M$  no alcanza su valor inicial, por otro lado el índice  $M$  no llega a su punto mínimo inmediatamente después del cuello de botella. Figura tomada de [Garza y Williamson, 2001].

El índice  $M$  fue calculado para cada locus por población en el programa Arlequin v3.5.2.2. Después se realizó una prueba de T de una muestra para determinar si el valor promedio era menor a 0.68, valor que Garza y Williamson recomiendan como punto crítico, valores menores a 0.68 se consideran como evidencia de cuello de botella reciente.

# Resultados

## 4.1. Selección

Se busco selección en nuestros loci usados con el programa BayeScan usando como poblaciones el clado tomasellii (*D. sonorensis*, *D. tomasellii* y *D. stvensoni*), los clados purpusii y spinulosum (*D. holmgrenii*, *D. purpusii*, *D. argenteum*, *D. rzedowskii*, *D. spinulosum*, *D. mejiae*, *D. edule*, *D. angustifolium* y *D. califanoi*) y *D. caputoi* junto con *D. merolae* estos últimos se consideraron a parte ya que hay literatura que dice que están aislados del resto y han sufrido cuellos de botella recientes.

El análisis de selección dio negativo para la mayoría de los loci (tabla 4.1). Solo el locus P12 en el clado tomasellii muestra una señal de selección direccional y el locus De5 en los clados purpusii y edule muestra también señal de selección direccional. Se corrieron los análisis sin estos dos loci, no hubo grandes diferencias en los resultados ni siquiera en el  $F_{ST}$  o  $R_{ST}$  por lo que se muestran los resultados de las pruebas con los 15 loci.

Tabla 4.1: **Resultados de BayeScan.** un valor de  $\log_{10}$  (BF) de 1 a 1.5 es considerado como evidencia de selección, valores de 0.5 a 1 son considerados como señal débil de selección. Valores positivos de alpha indican selección direccional mientras que valores negativos señalan selección estabilizadora.

ID	Clado tomasellii				Clados purpusii y spinulosum				D. caputoi y D. merolae			
	Prob	$\log_{10}$ (BF)	alpha	Fst	Prob	$\log_{10}$ (BF)	alpha	Fst	Prob	$\log_{10}$ (BF)	alpha	Fst
P10	0.061	-1.189	-0.024	0.429	0.037	-1.418	0.007	0.387	0.025	-1.598	-0.001	0.426
P12	0.893	0.920	1.423	0.727	0.030	-1.510	0.003	0.386	0.683	0.333	-0.912	0.259
P17	0.092	-0.993	0.028	0.441	0.142	-0.782	-0.083	0.368	0.045	-1.327	0.014	0.430
P18	0.338	-0.292	0.354	0.513	0.044	-1.335	-0.010	0.383	0.434	-0.116	0.402	0.518
P21	0.074	-1.100	0.024	0.440	0.058	-1.212	-0.020	0.381	0.135	-0.806	0.103	0.450
P22	0.075	-1.094	0.009	0.437	0.715	0.399	0.534	0.508	0.090	-1.006	0.003	0.428
P23	0.061	-1.186	0.027	0.441	0.785	0.562	0.763	0.559	0.033	-1.470	0.000	0.427
P25	0.062	-1.181	0.026	0.441	0.036	-1.423	0.000	0.385	0.084	-1.039	0.002	0.428
De1	0.056	-1.227	-0.020	0.431	0.739	0.452	0.611	0.525	0.098	-0.965	0.003	0.428
De3	0.056	-1.223	-0.015	0.431	0.491	-0.015	0.326	0.460	0.089	-1.010	0.050	0.438
De5	0.071	-1.115	0.024	0.440	0.938	1.178	0.969	0.605	0.090	-1.005	-0.001	0.427
De6	0.102	-0.943	-0.056	0.422	0.082	-1.051	-0.033	0.378	0.087	-1.019	-0.005	0.426
De7	0.189	-0.634	0.134	0.466	0.039	-1.389	0.009	0.387	0.094	-0.983	0.003	0.428
P7	0.078	-1.070	-0.036	0.428	0.043	-1.347	0.007	0.387	0.094	-0.986	0.005	0.428
P9	0.083	-1.046	0.043	0.444	0.125	-0.845	0.067	0.401	0.095	-0.978	-0.003	0.427

## 4.2. Diversidad genetica

### 4.2.1. Número de alelos, Heteozigosidad y Equilibrio de Hardy-Weinberg

Analizados juntos los 15 loci, el 100 % de ellos son polimorficos pero a nivel poblacional se cuenta con un promedio de 8.16 loci polimorficos por población. El número de alelos por población tiene como promedio de 2.56 (tabla 4.2). Hay que notar que los loci P10, P12, P17, P23 y De3 son los que están fuera del equilibrio de Hardy-weinberg en más poblaciones, las poblaciones con más loci fuera del

equilibrio son las del clado spinulosum excepto por *D. mejiae*, todas poblaciones de Oaxaca, también una población de *D. holmgrenii* y la población de Querétaro de *D. edule* tienen muchos loci fuera del equilibrio, las poblaciones de *D. caputoi* también tienen muchos loci fuera del equilibrio, relativo a sus loci polimórficos.

Tabla 4.2: **Resumen de la variación genética de las 31 poblaciones cada una analizada de manera independiente.** A = número de alelos, R= rango de alelos, He = heterocigosidad esperada, Ho = heterocigosidad observada,  $F_{IS}$ = índice de endogamia, valores alrededor de 0 indican cruza de individuos al azar, valores de  $F_{IS}$  alrededor de 1 indican fuerte cruza selectiva dentro de la población, valores de  $F_{IS}$  negativos indican cruza con individuos de otras poblaciones. Los ID en azul significan que ese locus tiene una deficiencia de heterocigotos y los valores en rojo significan exceso. Los valores de A, R, He y Ho son valores promedio de todos los loci.

Población	A	R	Ho	He	$F_{IS}$	Loci poli- mórficos	Loci con $p < 0.05$ ( $He \neq Ho$ )	ID de loci con $p < 0.05$
Dson novi Son	2.6	3.6	0.456	0.441	0.101	10	0	-
Dson Maza Son	2.5	3.5	0.397	0.433	0.614	4	0	-
Dson Nuri Son	2.222	3	0.303	0.401	0.549	9	2	P10 De1
Dtom Panu Sin	2.3	3.2	0.357	0.414	0.503	10	0	-
Dtom Comp Nay	2.875	3.875	0.432	0.474	0.598	8	0	-
Dtom Tuit Jal	3	4.111	0.374	0.491	0.754	9	1	De7
Dste Alt Gue	3.222	4	0.473	0.561	0.647	9	2	P23 De3
Dste Higue Mich	3.333	5	0.298	0.576	0.812	3	1	P25
Dpur Coyu Oax	3.714	5.714	0.564	0.514	0.412	7	0	-
Dcal Oax	2.556	3.667	0.322	0.384	0.171	9	1	De6
Darg Oax	2.571	3	0.497	0.505	0.308	7	0	-
Dhol Limon Oax	3.2	5.7	0.608	0.569	0.559	10	2	P18 P25
Dhol Loxi Oax	3.286	5.786	0.446	0.465	0.3	14	5	P22 P23 De1 De3 De7
Dspi Soya Oax	3.5	5.786	0.375	0.461	0.466	14	3	P17 P7 P9
Dspi CerroB Oax	3.462	5.538	0.414	0.518	0.501	13	5	P12 P21 De1 De3 P9
Drze SnBar Oax	3	4	0.353	0.581	0.779	11	7	P10 P12 P22 P23 P25 De3 De7
Dmej Honduras	3	3.818	0.454	0.533	0.277	11	2	P12 P22
Dedu VVerde Qro	3.364	4	0.183	0.615	0.749	11	7	P10 P12 P17 P18 P21 De7 P7
Dedu Fara Ver	2.9	3.9	0.361	0.47	0.683	10	2	P12 De6
Dedu Cd. V S.L.P.	3.917	6.583	0.512	0.688	0.67	12	3	P12 P21 De1
Dang Ald Tam	3.333	3.667	0.5	0.527	0.339	6	0	-
Dcap Gra Pue	3.143	4.857	0.495	0.381	-0.075	7	4	P10 P17 P23 De3
Dcap Gua Pue	2.333	3.333	0.528	0.395	-0.265	6	2	P23 De3
Dcap LomaP Pue	3	4.4	0.548	0.485	0.072	5	3	P10 P23 De3
Dmer Mina Chi	2.4	4.4	0.315	0.317	0.005	5	0	-
Dmer AP Chi	3.6	4.8	0.241	0.305	0.364	5	2	P17 P21
Dmer Jiq1 Chi	2.75	4.5	0.539	0.422	-0.23	4	3	P10 P17 P23
Dmer Jiq2 Chi	2.833	4.333	0.298	0.317	0.248	6	4	P12 P17 P23 De3
Dmer Flor Chi	3	4.2	0.414	0.418	0.041	5	1	P10
Dmer Santi Oax	3	4.571	0.364	0.332	0.025	7	4	P10 P17 P23 De3
Dmer LomaC Oax	3.333	4.667	0.666	0.537	-0.234	6	3	P17 P23 De3
Promedio	3.0072	4.371	0.422	0.468	0.346	8.161	2.225	

4.2.2. Frecuencias alélicas

En la figura 4.1 se representan las frecuencias y distribución de cada alelo, la mayoría de los alelos en realidad son muy poco frecuentes y para cada locus la mayor parte de las frecuencias se distribuye en pocos alelos, como por ejemplo el locus P12 con 10 alelos, la mayoría de los alelos se encuentran en una frecuencia muy baja y en realidad solo dominan dos alelos en todos las poblaciones.

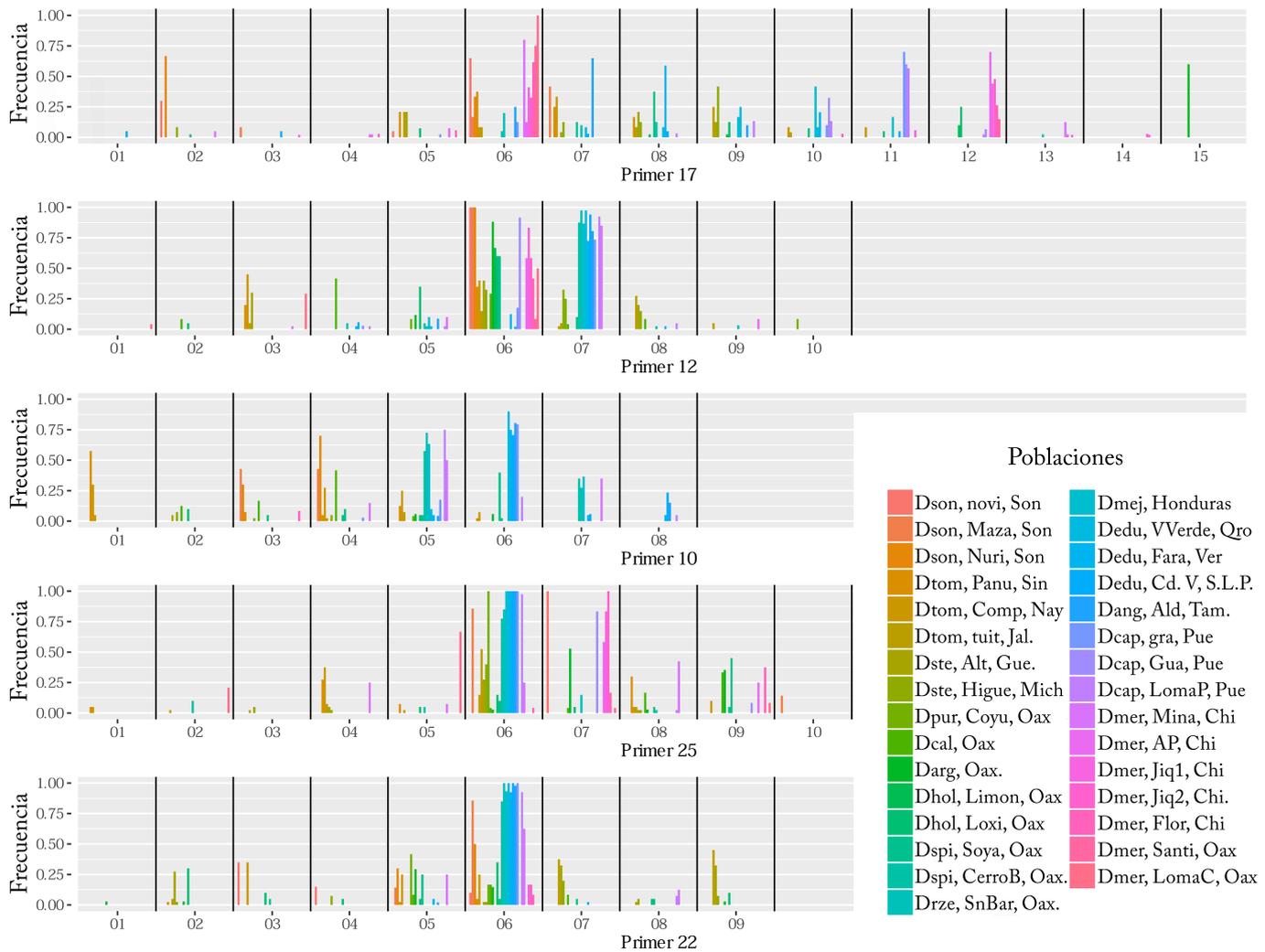


Figura 4.1: Frecuencias de alelos primeros 5 primers

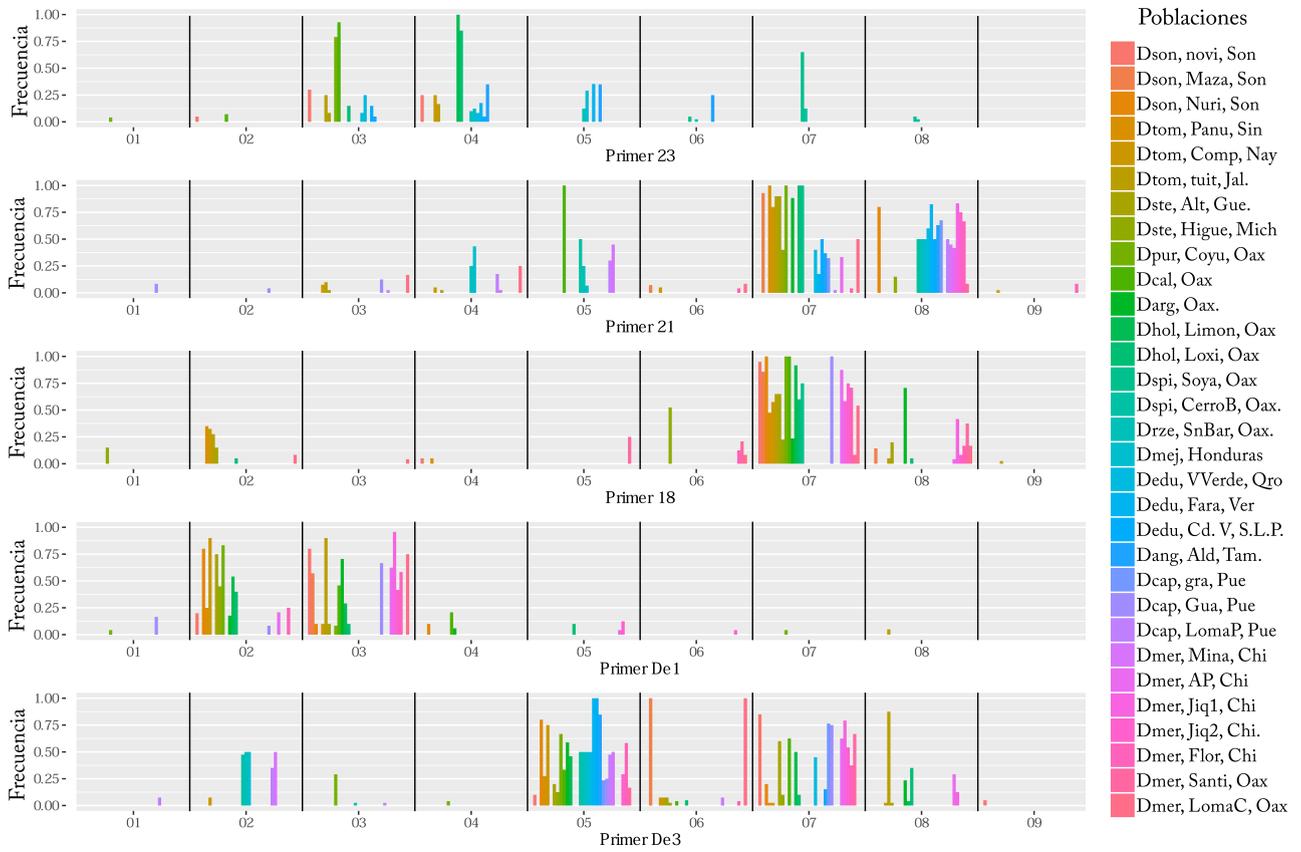


Figura 4.2: Frecuencias de alelos

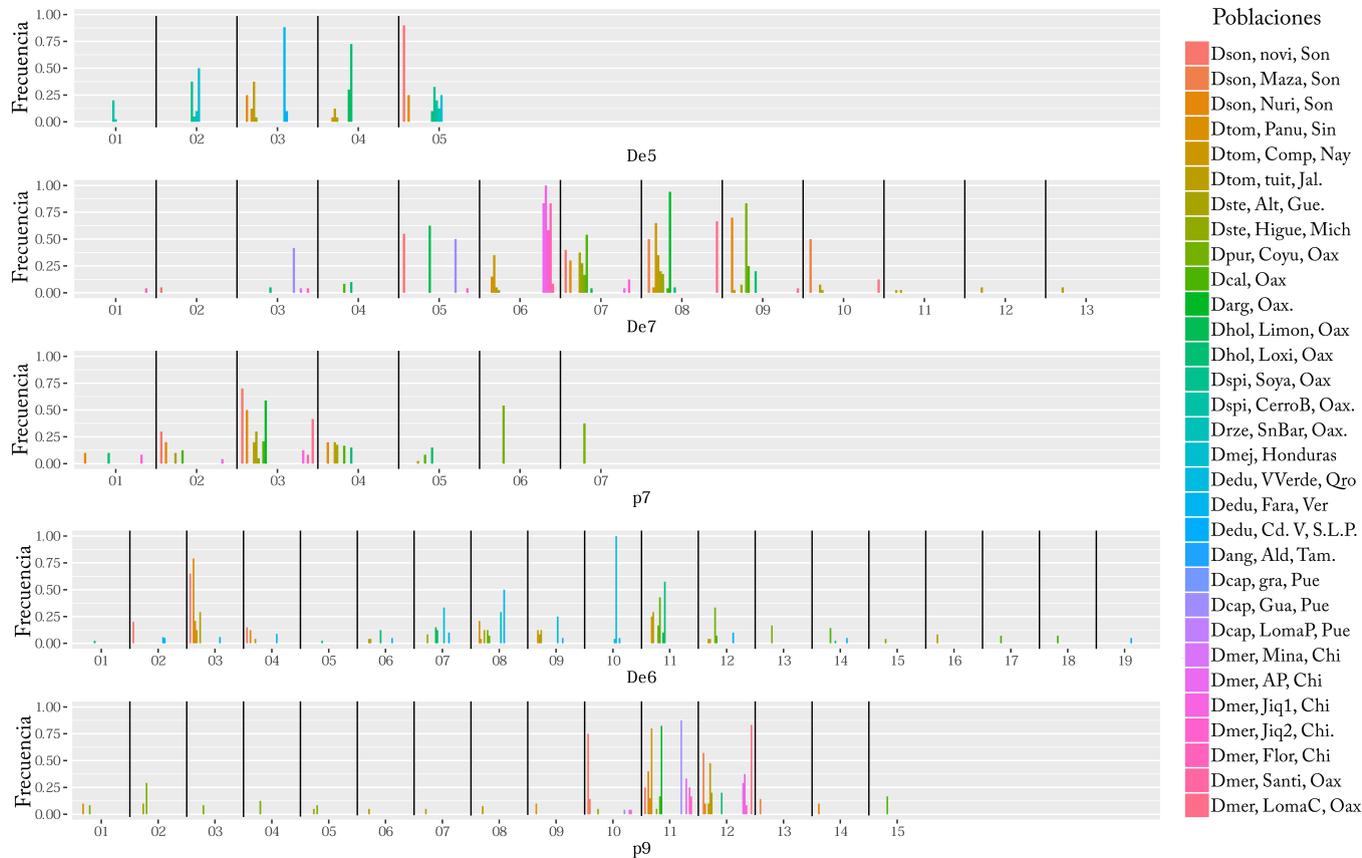


Figura 4.3: Frecuencias de alelos

### 4.3. Estructura Genetica

#### 4.3.1. $F_{ST}$ y $R_{ST}$

El índice  $F_{ST}$  muestra que las especies *D. caputoi* y *D. merolae* tienen una alta estructuración con el resto de las especies excepto entre ellas, tenemos el mismo caso para las especies *D. sonorensis*, *D. tomaselli* y una población de *D. stevensonii* así como para las especies de Oaxaca, *D. mejiae*, *D. edule* y *D. angustifolium* formando 3 grandes grupos. Las especies *D. purpusii*, *D. califanoi* y *D. argenteum* junto con la población más sureña de *D. stevensonii* en están bien diferenciadas del resto de las poblaciones (figura 4.4). Por otro lado el índice  $R_{ST}$  encuentra más estructura a nivel de género manteniendo solo a las especies de *D. caputoi* y *D. merolae* como grupo bien diferenciado del resto, también vemos que a todas las poblaciones de *D. tomaselli* forman un grupo bien definido y que son similares a las poblaciones de *D. edule*, por su parte cada población de *D. edule* es similar a distintas poblaciones de Oaxaca.

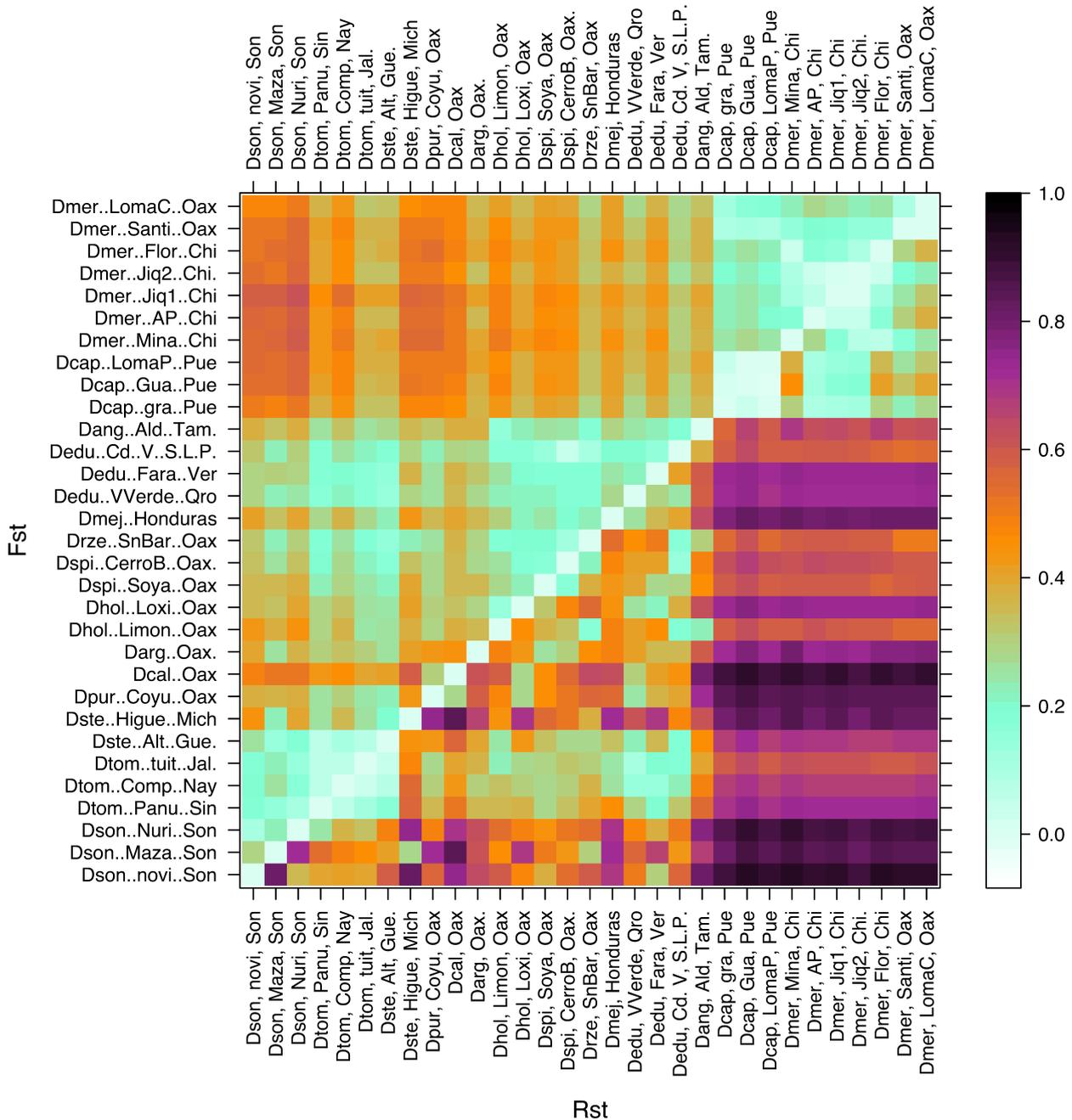


Figura 4.4: **Estructura poblacional pareada** entre las 31 poblaciones sobre del eje comparación usando  $F_{ST}$  y debajo de eje  $R_{ST}$ . Los valores de ambos índices van de 0 a 1, donde 0 indica no hay estructura poblacional y 1 indica total diferenciación.

### 4.3.2. STRUCTURE

La prueba de Evanno (figura: 4.6) revela que el número de cluster es igual a dos siendo las especies *D. caputoi*, *D. merolae*, *D. mejiae* y una población de *D. edule* miembros de un cluster y el resto de las especie miembros del otro cluster. A niveles más bajos de estructura genética ( $K=12$ ) vemos que

las poblaciones de *D. merolae* se separan en tres clusters y poblaciones como la de *D. rzedowskii*, la más sureña de *D. stevensoni* y dos poblaciones de *D. edule* tienen muchos migrantes. Especies como *D. califanoi* y *D. agenteum* no se separan de otras especies.

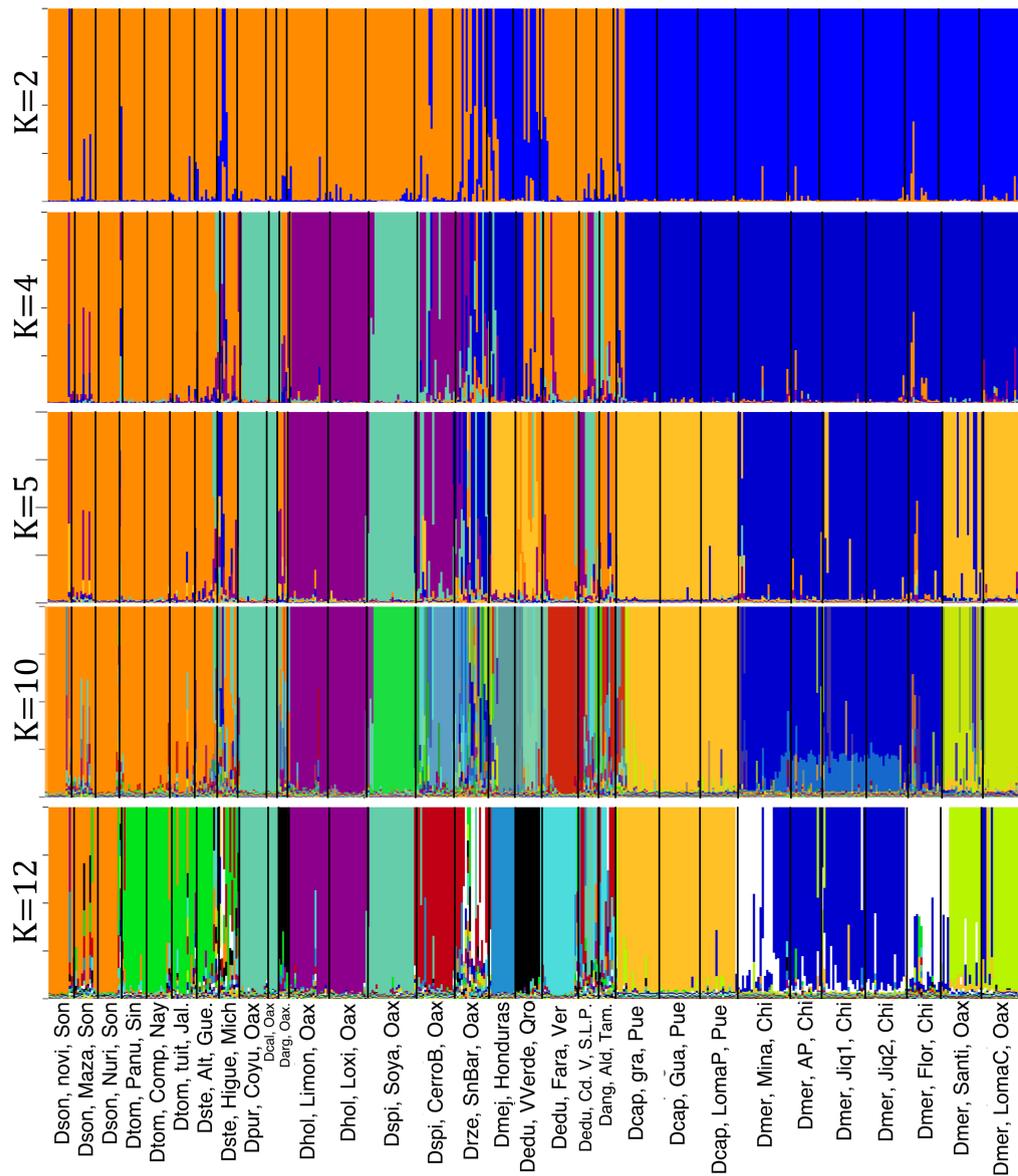


Figura 4.5: Graficos de estructura poblacional para cinco valores de K: K=2 es valor dado por la prueba de Evanno y K=12 es el valor más grande donde no se presentan clusters “fantasmas”.

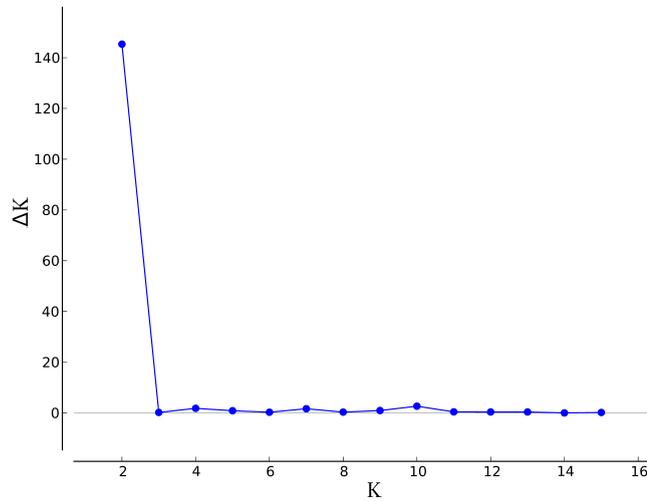


Figura 4.6:  $\Delta K$  para cada  $K$  dada a STRUCTURE. El valor más alto de  $\Delta K$  es igual al nivel más alto de estructura genética jerárquica.

### 4.3.3. STRUCTURE con submuestreo

Los resultados del STRUCTURE con submuestreo dan resultados diferentes a los resultados obtenidos usando todos los individuos mientras que la prueba de Evanno da como resultado  $K=2$  los miembros de los clusters cambian (figura 4.8) de la misma forma a niveles más bajos de estructura genética ( $K=12$ ) los miembros de cada cluster cambian, por ejemplo *D. merolae* ya no se divide en dos clusters si no que solo dos, *D. sonorensis* y *D. tomaselli* forman parte de un solo cluster y los migrantes de la población de *D. edule* de Cd. Valles SLP. ahora sabemos que son de las dos poblaciones de *D. spinulosum*.

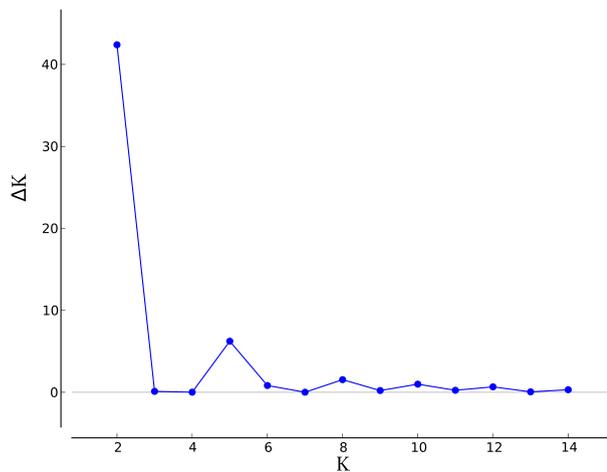


Figura 4.7:  $\Delta K$  tomando los resultados del submuestreo

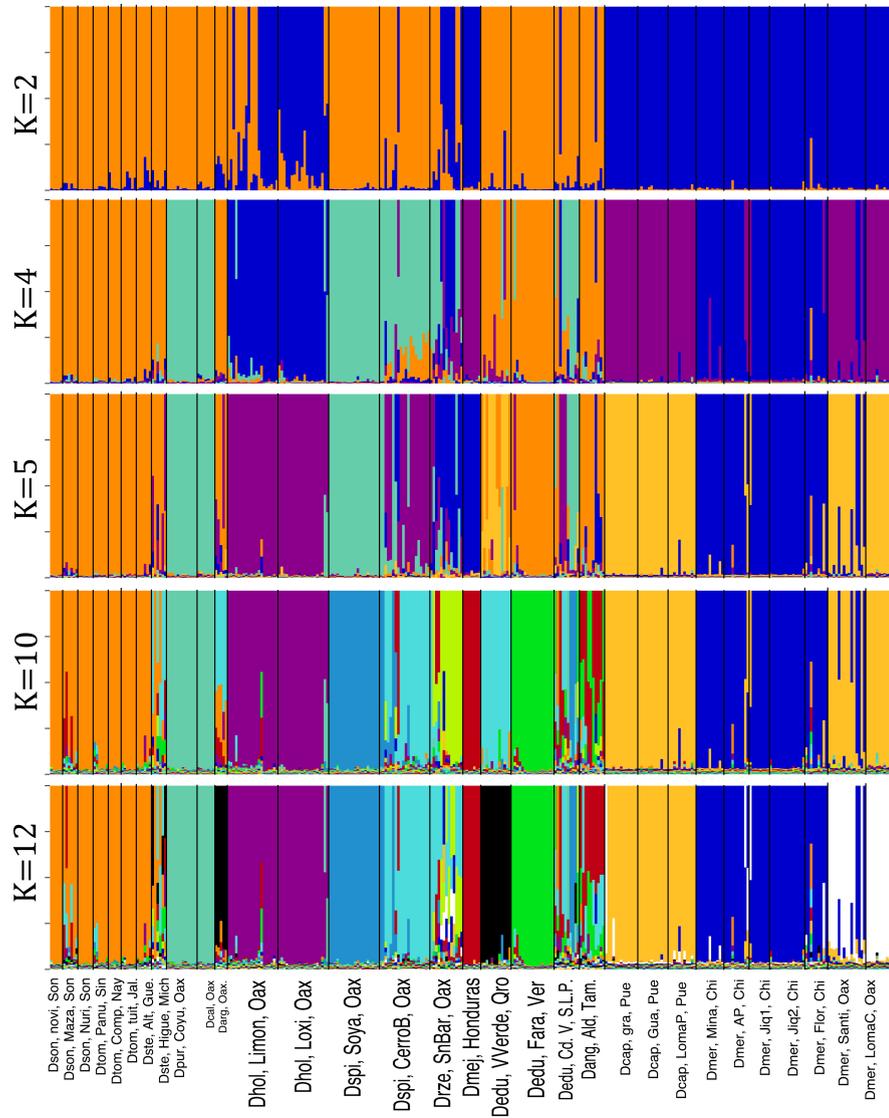


Figura 4.8: Graficos de estructura poblacional para cinco valores de  $K$  haciendo un submuestreo de los datos,  $K=2$  es valor dado por la prueba de Evanno y  $K=12$  es el valor más grande donde no se presentan clusters "fantasmas"

#### 4.4. Análisis de componentes principales

El análisis de componentes principales (PCA) nos revela algo bastante similar al  $F_{ST}$ : Graficando el PC1 (8.87%) contra el PC2 (3.77%) se forman dos grupos, en la figura 4.9 se observan dos grandes clusters, uno de ellos formado por las poblaciones del norte sobrelapadas con las poblaciones del centro y sur de México. Las poblaciones de *D. caputoi* y *D. merolae* se agrupan en un cluster bien definido y separado de los demás. Graficando el PC1 contra el PC3 (3.3%) se observa una tendencia muy similar, excepto por que las poblaciones de *D. holmgrenii* y una poblacion de *D. edule* se agrupan con las poblaciones del norte (figura A.1). En la representación del PC1 contra el PC4 (3.17%) vemos dos grandes grupos de los cuales se separan tres especies: *D. purpusii*, *D. califanoi* y *D. mejiae* (figura A.2).



Figura 4.9: **Análisis de componentes principales:** el PC1 explica el 8.87% de la variación y el PC2 el 3.77%. Se necesitan 181 PCs para explicar el 100% de la variación.

### 4.5. Prueba de Mantel

La prueba de mantel indica que hay una correlación positiva de distancia geográfica con la distancia genética tanto a nivel global como hecha en los tres grandes clados encontrados por Gonzalez [2008] y Gutierrez-Ortega [Datos sin publicar], esto apunta a que existe aislamiento por distancia. Los detalles se muestran en la tabla 4.3.

Tabla 4.3: **Resultados de las pruebas de Mantel**, tanto la prueba de Z como el método de Monte Carlo encuentran una relación positiva entre las matrices de distancias genéticas y geográficas.

	Prueba Z		Prueba MC	
	Z-score	P	Correlación	P
Clado tomasellii	42.13	0.05	0.34	0.029
Clados purpusii y spinulosum	48.88	0.085	0.37	0.04
D. caputoi y D. merolae	22.43	0.001	0.83	0.0001
Todos los individuos	1240	0.001	0.34	0.0001

### 4.6. Grafos de poblaciones

Los grafos poblacionales concuerdan en parte con los índices  $F_{ST}$  y  $R_{ST}$  separando a las especies *D. caputoi* y *D. merolae* del resto de las especies. En la figura 4.10 podemos observar que hay tres grafos: uno compuesto por poblaciones de Chiapas de *D. merolae*, otro compuesto por *D. caputoi* y poblaciones de Oaxaca de *D. merolae* y el tercero formado por el resto de las poblaciones, vemos que las poblaciones con más conexiones son la población de *D. rzedowskii* y dos poblaciones de *D. edule*, la población de cd. Valles S.L.P. y la de Valle Verde, Querétaro, con un valor de alfa de 0.01 vemos que solo las conexiones la población de *D. edule* de Valle Verde, Querétaro se mantienen pero en general se mantiene la topología del grafo (figura 4.10).

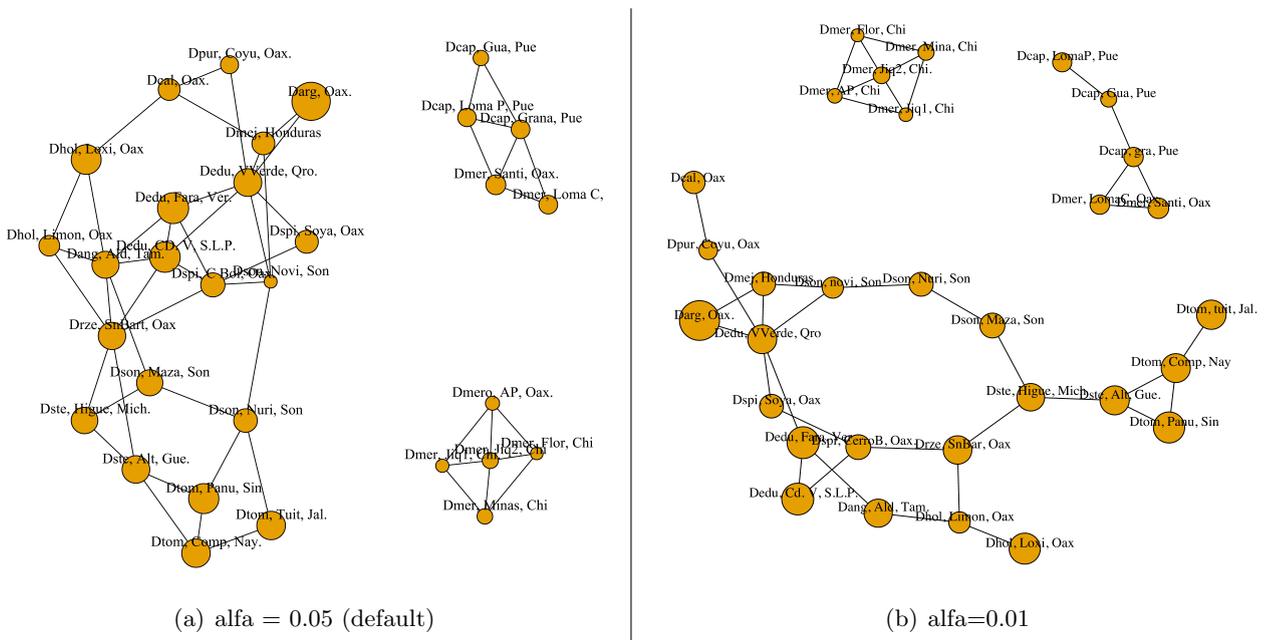


Figura 4.10: **Conexiones genéticas entre poblaciones.** El tamaño del círculo representa cuanto variación genética se explica por la diversidad de la misma población, la distancia de la conexión es proporcional a la variación que se explica por la diversidad de la población con que se conecta. Como se ve en la figura se forman tres grupos, uno formado por la mayoría de poblaciones excepto *D. caputoi* y *D. merolae* otro formado por poblaciones de *D. caputoi* y *D. merolae* y el ultimo solo conformado por poblaciones de *D. merolae*.

### 4.7. Índice $M$ para detectar cuellos de botella

El índice  $M$  nos indica que la mayoría de las poblaciones han sufrido un cuello de botella reciente la que concuerda con que la mayoría de los alelos de casi todos los loci son raros esto si se descartan los loci que son monomorficos para cada población (tabla 4.4), Se eliminaron los loci con  $A = 1$  y  $A = 2$  ya que estos siempre darán un valor de  $M$  igual a 1.

Tabla 4.4: **Indice  $M$** : muestra que casi todas las poblaciones pasaron por un cuello de botella. Se eliminaron los loci con  $A = 1$  y  $A=2$  ya que estos siempre darán un valor de  $M$  igual a 1, aumentando el valor promedio y ocultando si la población sufrió un cuello de botella. \* valores significativamente  $< 0.68$ .

Poblacion	Todos los loci		Sin loci con A=1 y 2	
	Promedio	s.d.	Promedio	s.d.
Dson novi Son	0.842	0.207	0.59*	0.092
Dson Maza Son	0.836	0.17	0.672	0.061
Dson Nuri Son	0.824	0.206	0.613*	0.107
Dtom Panu Sin	0.81	0.246	0.525*	0.126
Dtom Comp Nay	0.905	0.163	0.667	0.118
Dtom tuit Jal	0.908	0.163	0.678	0.138
Dste Alt Gue	0.912	0.152	0.691	0.114
Dste Higue Mich	0.861	0.202	0.583*	0.083
Dpur Coyu Oax	0.801	0.199	0.634*	0.109
Dcal Oax	0.864	0.207	0.591*	0.132
Darg Oax	0.917	0.138	0.694	0.039
Dhol Limon Oax	0.707	0.213	0.59*	0.124
Dhol Loxi Oax	0.658	0.229	0.564*	0.161
Dspi Soya Oax	0.644	0.206	0.584*	0.158
Dspi CerroB Oax	0.712	0.245	0.532*	0.114
Drze SnBar Oax	0.88	0.196	0.58*	0.091
Dmej Honduras	0.874	0.202	0.56*	0.068
Dedu VVerde Qro	0.913	0.179	0.594*	0.138
Dedu Fara Ver	0.892	0.186	0.594*	0.094
Dedu Cd. V. S.L.P.	0.746	0.224	0.618*	0.164
Dang Ald Tam.	0.938	0.108	0.75	0
Dcap gra Pue	0.67	0.173	0.615*	0.117
Dcap Gua Pue	0.776	0.201	0.608*	0.068
Dcap LomaP Pue	0.767	0.207	0.592*	0.06
Dmer Mina Chi	0.81	0.243	0.556*	0.157
Dmer AP Chi	0.9	0.177	0.65	0.15
Dmer Jiq1 Chi	0.849	0.239	0.472*	0.028
Dmer Jiq2 Chi	0.833	0.214	0.611*	0.142
Dmer Flor Chi	0.853	0.177	0.657	0.08
Dmer Santi Oax	0.679	0.171	0.679	0.171
Dmer LomaC Oax	0.752	0.14	0.711	0.104

# Discusión

## 5.1. Estructura genética a nivel de genero: Un gradiente

La estructura genética observada en el genero *Dioon* es baja a nivel general. Dentro de cada clado los valores son ligeramente más altos a los reportados en trabajos previos [Cabrera-Toledo et al., 2008, González-Astorga et al., 2008]. La prueba de Mantel nos dice que hay aislamiento por distancia, lo que explicaría los valores tan altos de diferenciación genética entre poblaciones del Norte y del Sur de Mexico. Aunque hay que mencionar que hay conexiones en el grafo poblacional en las que parece que no hay aislamiento por distancia. En el PCA claramente se observa que todas las poblaciones forman un gradiente de diferenciación: Las poblaciones más norteñas están separadas de poblaciones sureñas, pero ambas se sobrelapan con poblaciones intermedias. Esto se puede explicar por que hay flujo genético entre poblaciones cercanas, aunque estructura dada por la distancia, o que son especies de reciente muy divergencia.

El valor más alto de  $\Delta K$  fue 2, lo que revela grupos muy similares entre el programa STRUCTURE y el PCA. Un número de grupos igual a dos es muy bajo para un genero con 15 especies. Cavender-Bares *et al* [2015] en un trabajo en el genero *Quercus* ( $\approx 140$  especies) encuentra que el valor más alto de  $\Delta K$  fue 2, ellos mencionan que para especies con tan amplia distribución un  $K=2$  representa más bien los grandes clados antes que especies individuales ya que se encontró diferenciación muy baja entre algunas especies. En sus resultados también encuentran grandes zonas híbridas. De la misma manera en este trabajo también se encontraron poblaciones con muchos híbridos o con ancestría mezclada sobre todo en los clados *purpusii* y *spinulosum*, estas poblaciones son también las que más conexiones tienen en el grafo poblacional, las conexiones que se mantienen en nuestros datos.

Algo que hay que notar es que algunas de las conexiones más estables (que no se pierden aun si hacemos más estricto el punto para mantener las conexiones), como por ejemplo *D. sonorensis* de Sonora con *D. edule* de Querétaro y *D. edule* de Querétaro con *D. mejiae* de Honduras, no se explican por los dispersores antes mencionados. Estas conexiones se pueden explicar por otro tipo de dispersión, como la dispersión humana que es independiente a barrera geográficas como montañas y ríos, y que ya ha sido observada para el genero *Dioon* y otras cícadas mexicanas. A pesar de su toxicidad extrema, las cícadas siguen siendo usadas como alimento en algunos lugares como en la Sierra Gorda y en Honduras, o como narcóticos en Sonora y también en Sierra Gorda [Bonta M, *in preparation*]. Estas “conexiones” culturales podrían estar relacionadas a las conexiones genéticas que observamos en los mismos lugares (Figura 4.10).

## 5.2. ¿Reciente divergencia?

De manera especulativa, otra explicación a la diferenciación en gradiente, seria que la diversificación del género *Dioon* es relativamente reciente, menor a los 9-12Ma que proponen Nagalingum

*et al.* [2011] y Salas-Leiva *et al.* [2013] para este género. Cavender-Bares *et al.* [2015] encuentran de manera similar para el género *Quercus* subsección *Virentes*, que la edad del nodo corona (Antepasado común más reciente de las especies muestreadas) fue hace aproximadamente 11Ma. Once Ma es la edad del ancestro común de todas las especies vivientes lo que no quiere decir que hace 11Ma ya existían ni comenzaron a dividirse las especies de *Quercus*. Cavender-Bares *et al.* [2015] explican que fueron varios los eventos que llevaron a la diversificación de *Quercus* entre ellos; cambios en los niveles de los mares, actividad volcánica, sequías, enfriamiento global y dispersión a larga distancia. Estos procesos ambientales iniciaban procesos de especiación independientes y no afectan a una sola especie ya que sus resultados son congruentes con las dinámicas demográficas de otras especies de plantas. Estos eventos pudieron haber afectado al género *Dioon* teniendo varios procesos de especiación entre 0.99Ma y 8.1Ma según los datos Cavender-Bares *et al.* [2015].

Ya que en el género *Dioon* hay tendencias distintas entre regiones geográficas, se dividirá la siguiente parte de la discusión en tres secciones. La sección Noroeste-Oeste, formado por las especies *D. sonorense*, *D. tomaselli* y *D. stevensonii*. La sección Noreste-Este-Suroeste formado por las especies, *D. holmgrenii*, *D. spinulosum*, *D. rzedowskii*, *D. mejiae*, *D. purpusii*, *D. argenteum*, *D. califanoi* *D. edule* y *D. angustifolium*. Y finalmente la sección Valle de Tehuacan-Sierra Madre del Sur formado por las especies *D. caputoi* y *D. merolae*. En términos del equilibrio de Hardy-Weinberg existen pocos loci fuera de equilibrio para el grupo Noroeste-Oeste, la mayoría de loci fuera del equilibrio con deficiencia de heterocigotos para el grupo Noreste-Este-Suroeste y la mayoría de loci con exceso de heterocigotos para el grupo Valle de Tehuacan-Sierra Madre del Sur.

La sección Noroeste-Oeste incluye las mismas especies que el clado *tomaselli*, la sección Noreste-Este-Suroeste es formado por los clados *spinulosum* y *purpusii* y finalmente se decidió discutir a las especies *D. merolae* y *caputoi* en una sección aparte.

### 5.3. Sección Noroeste-Oeste

#### 5.3.1. Evidencia de un cuello de botella en las poblaciones del norte de México

Aunque las poblaciones de la sección Noroeste-Oeste están en equilibrio de Hardy-Weinberg tienen los promedios de alelos y rango de alelos más bajos. Esto sugiere que el cuello de botella detectado con el índice  $M$  no es reciente: Durante un cuello de botella reciente se detectaría un exceso de heterocigotos y una pérdida de los alelos raros, seguido al final una estabilización de la heterocigosidad dentro del equilibrio de Hardy-Weinberg. Ya que las poblaciones ya presentan una estabilización de la heterocigosidad podemos decir que el cuello de botella no es contemporáneo.

Aunque no se hicieron pruebas para datar el cuello de botella (o efecto fundador), podemos hipotetizar que éste ocurrió hace algunos milenios, como discuten Cabrera-Toledo *et al.* [2008] para la especie *D. caputoi* y Gonzalez-Astorga y Castillo-Campos [2004], Lowe *et al.* [2005], Feyissa *et al.* [2007] para distintas especies de arboles. En teoría, se necesitaría un periodo de 15 generaciones para alcanzar el equilibrio si se tuviera una heterocigosidad del 100% debido al sobrelape de generaciones. Ya que la cícadas son plantas con un ciclo de vida muy largo ( $\approx 500$  años para alcanzar edad reproductiva [Vovides, 1990, Octavio-Aguilar *et al.*, 2008, Cabrera-Toledo *et al.*, 2008]), se necesitaría un periodo de 7,500 años para alcanzar el equilibrio de Hardy-Weinberg. Pudo ser más reciente dependiendo de la heterocigosidad inicial después del cuello de botella.

Un cuello de botella en realidad no es sorprendente ni inconsistente para estas poblaciones. Toda

la literatura que hay de estas población mencionan que actualmente estas poblaciones están muy deterioradas y sufren riesgo de desaparecer. El principal factor del deterioro es el cambio de uso de suelo por actividades ganaderas, pero se agrava por la explotación humana para preparación de bebidas alcohólicas, tráfico de plantas y semillas, así como sequías extremas y el cambio climático. Lo anterior es congruente con un cuello de botella reciente, sin embargo ya que la evidencia apunta a que el cuello de botella no es tan reciente podemos hipotetizar que hubo otros eventos afectando estas poblaciones, como efecto fundador por dispersión a larga distancia o eventos históricos que causaron la reducción de la población.

### 5.3.2. Un complejo de especies

En cuanto a la estructura genética poblacional del grupo Noroeste-Oeste, los valores de  $F_{ST}$  que encontramos son considerados altos en la literatura en general pero al parecer son valores normales entre poblaciones de cícadras siendo similares a los valores encontrados en el género *Dioon* para poblaciones de las mismas especies [González-Astorga et al., 2008] y menor a otros géneros de cícadras [Gonzalez-Astorga et al., 2009]. Gonzales-Astorga *et al.* [2009, 2008] encontraron valores similares de  $F_{ST}$  y un número de alelos bajos similar a nuestros resultados (excepto por la población más sureña de *D. stevensonii* en Michoacán, en la que tenemos resultados contradictorios para esta población). Las gráficas de estructura poblacional de STRUCTURE nos dividen a las especies de la sección Noroeste-Oeste en uno o dos clusters, dependiendo si se usa submuestreo o no, considerando que son tres especies mínimo esperaríamos tres clusters. El grafo poblacional las ubica como un grupo con conexiones entre si mismo y con la Sección Noreste-Este-Suroeste. Excepto por la población de *D. stevensonii* en Michoacán y la población de *D. sonorese* en la Presa Novillo, Sonora éste grupo solo muestra conexiones dentro de si mismo. La prueba de Mantel nos dice que si hay aislamiento por distancia esto quiere decir que poblaciones cercanas se mezclan entre ellas y no con las mas distantes. El PCA apoya lo anterior pues el grupo Noroeste-Oeste conforma un gradiente entre estas especies. Toda esta evidencia encaja con la definición de complejo de especies: Un grupo de especies estrechamente relacionado muy similares en características (genéticas en este caso) al punto donde los limites en ellas no es claro [Brown et al., 1995, De Queiroz, 2007]. Por todo lo anterior, podemos considerar a las especies de la sección Noroeste-Oeste como un complejo de especies.

### 5.3.3. Poblaciones fragmentadas y con poco flujo genético entre ellas

Valores tan altos de  $F_{IS}$  pueden indicar cruza selectiva en cada población de la sección Noroeste-Oeste, sin embargo no hay reportes de endogamia (cruza con individuos estrechamente emparentados) en cícadras. Otra explicación puede ser que las poblaciones de este grupo estén genéticamente diferenciadas (considerar una sola población cuando en realidad se tienen dos poblaciones diferentes.) esto concuerda con los reportes de Gonzalez-Astorga *et al.* [2008, 2009] donde mencionan que estas poblaciones están altamente fragmentadas, aunque para mantenerse así debe de haber poco flujo genético dentro de cada población así como entre poblaciones.

Álvarez-Yépez J. *et al.* apoyan la hipótesis de que hay poco flujo genético entre poblaciones, en su estudio encuentran que la fecundación de *D. sonorese* es muy baja, al igual que el reclutamiento [Álvarez-Yépez et al., 2011]. La baja fecundidad nos dice, primero que la polinización es poco eficiente en estas poblaciones, de lo que podemos inferir que si los polinizadores no pueden fecundar una planta por completo dentro de una población, será mucho más difícil hacerlo entre dos poblaciones distintas y aunque ocurra polinización entre dos poblaciones. El bajo reclutamiento disminuye la probabilidad de que una de estas semillas llegue a la etapa adulta y se reproduzca. En cuanto a las semillas, aunque la distribución de algunos de los dispersores es la misma que estas poblaciones, no hay reporte de ningún

dispersor para la sección Noroeste-Oeste [Gutiérrez-Ortega et al., 2014]. En el grafo poblacional se observa que la sección Noroeste-Oeste es una de las que menos conexiones tiene lo que reafirma el poco flujo entre poblaciones.

Un resultado contrastante al trabajo de Gonzales-Astorga *et al.* [2008, 2009] para el género *Dioon* es el  $F_{IS}$  ellos encontraron valores cercanos a cero mientras que nosotros encontramos un  $F_{IS} \approx 0.5$ . Los valores distintos de  $F_{IS}$  pueden ser debido a el tipo de marcadores usados, aloenzimas contra SSRs; siendo los SSRs capaces de detectar estructura más fina. Freville H. *et al.* [2001] reportan un caso similar; Ellos compararon aloenzimas contra SSRs y encontraron que los SSRs son capaces de detectar el flujo genético cuando las aloenzimas no pueden. Contreras-Negrete *et al.* [2015] encontraron algo similar en cactáceas reportando que cuando se usan SSR y aloenzimas, las aloenzimas no pueden detectar loci fuera del equilibrio de Hardy-Weinberg, mientras que los SSRs si encuentran desviaciones significativas al equilibrio. Lemaire *et al.* [2000] también reportan discrepancias entre resultados de aloenzimas y SSRs posiblemente debido a que las aloenzimas podrían estar sujetas a selección mientras que los SSRs se pueden filtrar para evitar locus bajo selección, dando resultados más confiables en genética de poblaciones.

## 5.4. Sección Noreste-Este-Suroeste

### 5.4.1. Deficiencia de heterocigocidad y alta estructura dentro de cada población se explica con tres posibles escenarios

En el Sección Noreste-Este-Suroeste vemos un mayor número de loci fuera del equilibrio de Hardy-Weinberg con deficiencia de heterocigotos (excepto para una población de *D. holmgrenii*). Este grupo también es el que tiene las poblaciones con los números de alelos y rango de alelos más grandes (tabla 4.2.), lo que hace a la sección Noreste-Este-Suroeste, la más diversa de las tres secciones. La deficiencia de heterocigotos y un  $F_{IS}$  alto se puede explicar de tres maneras: **1**) Cruza selectiva o endogamia, **2**) Efecto Wahlund por generaciones intercaladas y **3**) Efecto Wahlund por migración (semillas o polen).

El efecto Wahlund es la desviación del equilibrio de Hardy-Weinberg hacia una deficiencia de heterocigotos causado por la estructura poblacional, por ejemplo cuando se muestrea una localidad, pero en realidad esta población esta dividida por alguna barrera ya sea física o ecológica, aun cuando ambas poblaciones estén en equilibrio al juntarlas (o muestrearlas como una) provocaremos una deficiencia de heterocigotos.

Como se mencionó antes, la cruza selectiva en cíadas no se ha reportado, por el contrario hay trabajos que reportan que la polinización es al azar o sin preferencia de los polinizadores por plantas en particular, además de que el viento es un factor que si bien no es que dominante para la polinización, es un factor importante [Norstog et al., 1986, Terry et al., 2005]. El Efecto Wahlund por generaciones intercaladas sería congruente con las observaciones de Vovides [1990]. En el caso del efecto Wahlund por solapamiento de generaciones, la barrera es el tiempo; Por ejemplo, como menciona Vovides algunas especies del género *Dioon* se reproducen cada cuatro años, si una parte de la población se reproduce en 1990 y otra en 1992 estas nunca se cruzarían entre si, formando cada una una población independiente ya que, aunque estén en el mismo espacio el tiempo las separaría. El problema de esta hipótesis es que como menciona Ornduff [1987, 1991] solo haría falta una temporada de sequía (esto retrasa la reproducción) o algún estrés mecánico como fuego o herbívora (esto apresuraría la reproducción) para sincronizar la reproducción de las plantas, en plantas tan longevas como las cíadas esto podría ocurrir

varias veces a lo largo de su vida.

#### 5.4.2. Alto flujo genético.

El efecto Wahlund por migración es la explicación más congruente con nuestro grafo poblacional, PCA y  $F_{ST}$  pareado y también es consistente con la biología del género *Dioon* ya que varios trabajos mencionan varios dispersores para el género *Dioon* tales como: ardillas (*Sciurus spp.*), urracas (*Calocitta formosa*) [García et al., 2016], guacamaya verde (*Ara militaris*) [Chemnick, 2013] y ratón silvestre (*Peromyscus mexicanus*) [Vovides, 1990] todos estos animales de amplia distribución en México. Para otros géneros de cícadas mexicanas se han reportado dispersores como pecaríes (*Pecari tajacu*) [Pérez Farrera y Vovides, 2004] y cenizotes (*Mimus polyglottos*) [Eckenwalder, 1980] y aunque no se han documentado Chemnick [2013] menciona que los cenizotes, la chara papán (*Psilorhinus morio*), así como murciélagos, podrían ser posibles dispersores del género *Dioon*. A pesar de la gran cantidad de posibles dispersores varios estudios independientes encuentran que para el género *Dioon* y otras cícadas mexicanas la distribución de plántulas es mayormente alrededor de las plantas madre por gravedad [Yáñez-Espinosa, 2007, Octavio-Aguilar et al., 2008, Salomé-Castañeda, 2009, Chávez-Acuña, 2010, García et al., 2016]. Nuestros resultados sugieren alta dispersión tal vez por aves y roedores podrían jugar un papel muy importante en la dispersión de semillas [Chemnick, 2013], lo que concuerda con los lugares donde más semillas germinan [Vovides, 1990, Yáñez-Espinosa, 2007, Salomé-Castañeda, 2009, Álvarez-Yépez et al., 2011], cerca de las plantas madre o cerca de otras plantas nodrizas, como árboles, y entre rocas, siempre lejos de la luz para protegerse del sol para evitar desecación, que es donde aves (árboles) y roedores (entre rocas) llevan las semillas para comer la sarcotesta.

El hecho de que no encontremos individuos con genotipos de otras poblaciones dentro de una en el STRUCTURE va en contra de la propuesta anterior. Lo que puede indicar que los eventos de migración no son actuales: Zhivotovsky [2015] propone un modelo en donde el efecto Wahlund puede tener un efecto importante sobre la frecuencia de los heterocigotos, baja diversidad de la población aceptora y alta en la población donadora. En este caso la población aceptora será una población cualquiera y la donadora serían todas las demás poblaciones. Y como discute Christiansen [1988] en caso que efecto Wahlund cuando hay solapamiento de generaciones se necesitarían muchas generaciones para llegar a las frecuencias de heterocigotos esperadas según el equilibrio de Hardy-Weinberg. Christiansen [1998] proporciona una serie de fórmulas para calcular la edad (en generaciones) de cuando ocurrió la migración, pero se requieren saber a que generación pertenece cada individuo para poder aplicarlas. Aún considerando el caso de 100% de homocigotos el tiempo sería en este caso menor a 7,500 años ya que aun no se ha alcanzado el equilibrio de Hardy-Weinberg.

Murawski et al. [2016] mencionan que otra manera de explicar el efecto Wahlund sería tener pocos donadores de polen por ser femenino. De ser esto cierto se apoyaría la hipótesis de reciente divergencia ya que la poca diferenciación que observamos no es dada por la migración. Sabemos que no fue flujo genético por polen ya que esto aumentaría los heterocigotos.

La sección Noreste-Este-Suroeste, al igual que la sección Noroeste-Oeste, tiene un promedio de  $F_{ST}$  que son normales para poblaciones similares en cícadas [González-Astorga et al., 2008]. La sección Noreste-Este-Suroeste también es el grupo con más conexiones en el grafo poblacional con un promedio de 2.8 conexiones por población. En el PCA las poblaciones de la sección Noreste-Este-Suroeste no solo se solapan más que la sección Noroeste-Oeste sino que hay individuos que genéticamente son más parecidos a individuos de otras poblaciones que a las que pertenecen. Ambos gráficos de

estructura poblacional confirman lo anterior mostrando que en el Sección Noreste-Este-Suroeste hay poblaciones con muchos migrantes así como individuos híbridos. Esto apunta a que las cícadas de este grupo son plantas con una alta dispersión tanto de semillas como polen, por lo que al igual que la sección Noroeste-Oeste podemos considerar la sección Noreste-Este-Suroeste como otro complejo de especies.

## 5.5. Sección Valle de Tehuacan-Sierra Madre del Sur

### 5.5.1. Cuello de botella y exogamia

El Sección Valle de Tehuacan-Sierra Madre del Sur muestra tendencias muy diferentes a los otros dos grupos. Tiene un exceso de heterocigotos en casi todas las poblaciones para la mayoría de los loci (relativo al número de loci polimórficos); un  $F_{IS}$  cercano a cero o negativo (excepto en dos poblaciones); y el número y rango de alelos es intermedio a los grupos Norte y Centro. Esto es congruente con los resultados obtenidos por Cabrera-Toledo D. *et al.* [2008], como mencionan Cabrera-Toledo D. *et al.* un exceso heterocigotos se explica de dos maneras: 1) Un cuello de botella reciente en el que aún no se pierden los alelos raros o 2) Exogamia (cruza con individuos muy poco emparentados, por ejemplo individuos de otra población); los valores negativos de  $F_{IS}$  apoyan la exogamia al igual que nuestro grafo poblacional y el índice  $F_{ST}$  ya que hay poca diferenciación entre poblaciones del Sección Valle de Tehuacan-Sierra Madre del Sur.

Un cuello de botella reciente es apoyado por el índice  $M$ , además de que poblaciones de *D. caputoi* presentan una proporción 3:1 de machos:hembras [Cabrera-Toledo et al., 2008], esto es evidencia de una población perturbada ya que la proporción de sexos para poblaciones no perturbadas es 1:1 [Yáñez-Espinosa, 2007] (no hay información de proporción de sexos para *D. merolae* pero se ha reportado que las poblaciones son explotadas gravemente año con año [Pérez-Farrera y Vovides, 2006]). Es importante decir que estas dos hipótesis (cuello de botella vs exogamia) no son contrarias y hay evidencia de ambas en este trabajo.

Una proporción de sexos de 3:1 tendría un impacto importante en la diversidad genética reduciendo el tamaño efectivo poblacional y aumentando la fuerza de la deriva genética (lo que causaría pérdida de diversidad). Aunque también es posible que no afecte de manera dramática al tamaño efectivo poblacional ya que un solo cono hembra puede ser fecundado por polen de tres o más plantas macho.

### 5.5.2. Aislamiento del resto de las especies

La sección Valle de Tehuacan-Sierra Madre del Sur presenta valores de  $F_{ST}$  pareados similares a los de la sección Noroeste-Oeste, sin embargo el  $R_{ST}$  nos dice que las poblaciones del Sección Valle de Tehuacan-Sierra Madre del Sur se parecen más entre ellas que las poblaciones de los otros grupos entre ellos mismos. En el PCA podemos observar que el Sección Valle de Tehuacan-Sierra Madre del Sur está completamente separado de los otros dos grupos lo mismo podemos decir en el grafo poblacional donde poblaciones de estos grupo no forman ninguna conexión con poblaciones de los otros dos grupos. Todo esto es evidencia de aislamiento, lo que concuerda con el paisaje del Valle de Tehuacan-Cuicatlan, que es donde se localizan las poblaciones de *D. caputoi*. El valle se encuentra aislado geográficamente, lo que evitaría flujo con otras especies [Beraldi-Campesi et al., 2006, Cornejo-Romero, 2013]. La especie

*D. merolae* es la más sureña de México y su distribución es en montañas de Oaxaca y Chiapas por lo que también se encontraría aislada geográficamente.



# Conclusiones

- Las 14 especies analizadas del genero *Dioon* se agrupan solo en dos grandes grupos genéticos, los que presentan patrones diferentes de acuerdo a su posición geográfica.
- La dispersión de semilla y/o polen para el genero *Dioon* ha sido subestimada al menos para las especies de la sección Noreste-Este-Suroeste, aquí se demuestra que el flujo genético entre poblaciones es a larga distancia posiblemente por aves y pequeños roedores para semillas, coleópteros para polen.
- También se observa aislamiento por distancia y que las diferencias genéticas entre poblaciones no son marcadas, más bien conforman un gradiente.
- Todas o la mayoría de las poblaciones han sufrido cuellos de botella (o efecto fundador), siendo el de la sección Valle de Tehuacan-Sierra Madre del Sur el más reciente.
- Algunas de las conexiones genéticas entre poblaciones, no se explican por dispersores comunes como aves o mamíferos, ya que no concuerda con la capacidad de dispersión ni distribución de los vectores animales mencionados en este trabajo. Es posible que al menos para algunas especies, haya un componente significativo de dispersión humana.

## Perspectivas

- Validar la dispersión a larga distancia con observaciones detalladas en campo.
- Estudiar si las conexiones que no se explican por vectores animales se explican por dispersión humana mediante la integración de información geográfica, etnográfica, arqueológica y genética de poblaciones humana.
- Estimar modelos demográficos explícitos de migración para ahondar en la historia de *Dioon*, y poder descartar que patrones encontrados en este trabajo se deban a una reciente diversificación de *Dioon* en México.
- Genotipar plantas hembras y las semillas de uno de sus conos para ver el número de donadores de polen por cono: Un número reducido de donadores explicaría la deficiencia de heterocigotos en poblaciones del Noreste, Este y Suroeste, Un número elevado explicaría el exceso de heterocigotos en poblaciones de *D. caputoi* y *D. merolae*.
- Realizar simulaciones *forward* usando los modelos de Christiansen FB y Zhivotvsky LA y simulaciones *backward* (coalecencia) para comenzar a datar cuando ocurrieron cambios en el tamaño efectivo poblacional y un aproximado de cuantas generaciones se necesitan para observar los mismos patrones que estamos viendo en nuestros datos.



# Bibliografía

- [Abraham y Mathew, 1962] Abraham, A. y Mathew, P. (1962). Cytological studies in the cycads: sex chromosomes in cycas. *Annals of Botany*, 26(2):261–266.
- [Álvarez-Yépez et al., 2011] Álvarez-Yépez, J. C., Dovčiak, M., y Búrquez, A. (2011). Persistence of a rare ancient cycad: effects of environment and demography. *Biological Conservation*, 144(1):122–130.
- [Bachtrog et al., 2000] Bachtrog, D., Agis, M., Imhof, M., y Schlötterer, C. (2000). Microsatellite variability differs between dinucleotide repeat motifs?evidence from drosophila melanogaster. *Molecular Biology and Evolution*, 17(9):1277–1285.
- [Beraldi-Campesi et al., 2006] Beraldi-Campesi, H., Cevallos-Ferriz, S. R., Centeno-García, E., Arenas-Abad, C., y Fernández, L. P. (2006). Sedimentology and paleoecology of an eocene–oligocene alluvial–lacustrine arid system, southern mexico. *Sedimentary Geology*, 191(3):227–254.
- [Bonta et al., 2006] Bonta, M., FLORES PINOT, O., Graham, D., Haynes, J., y Sandoval, G. (2006). Ethnobotany and conservation of tiusinte (*dioon mejiae* standl. & lo williams, zamiaceae) in northeastern honduras. *Journal of Ethnobiology*, 26(2):228–257.
- [Bonta y Osborne, 2005] Bonta, M. y Osborne, R. (2005). Cycads in the vernacular: A compendium of local names (nombres vernaculares de cycadalesâun compendio). *Proceedings of CYCAD*, pages 147–175.
- [Brookfield, 1996] Brookfield, J. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, 5(3):453–455.
- [Brown et al., 1995] Brown, J. K., Frohlich, D., y Rosell, R. (1995). The sweetpotato or silverleaf whiteflies: biotypes of *bemisia tabaci* or a species complex? *Annual review of entomology*, 40(1):511–534.
- [Cabrera-Toledo et al., 2010] Cabrera-Toledo, D., González-Astorga, J., Nicolalde-Morejón, F., Vergara-Silva, F., y Vovides, A. P. (2010). Allozyme diversity levels in two congeneric *dioon* spp.(zamiaceae, cycadales) with contrasting rarities. *Plant Systematics and Evolution*, 290(1-4):115–125.
- [Cabrera-Toledo et al., 2008] Cabrera-Toledo, D., González-Astorga, J., y Vovides, A. P. (2008). Heterozygote excess in ancient populations of the critically endangered *dioon caputoi* (zamiaceae, cycadales) from central mexico. *Botanical Journal of the Linnean Society*, 158(3):436–447.
- [Chamberlain, 1926] Chamberlain, C. J. (1926). Hybrids in cycads. *Botanical Gazette*, pages 401–418.
- [Chávez-Acuña, 2010] Chávez-Acuña, I. J. (2010). Análisis estructural de dos poblaciones de *dioon edule* lindl. (zamiaceae) en comunidades indígenas y mestizas de la región xi'iuy de la

- palma, s.l.p. [conjunto] : implicaciones ecológicas y culturales. *Universidad Autónoma de San Luis Potosí, tesis de maestría*.
- [Chemnick, 2013] Chemnick, J. (2013). Seed dispersal agents for selected mexican cycads in: <http://www.aloesinwonderland.com/article/3762-seed-dispersal-agents-for-selected-mexican-cycads>. *Santa Barbara Aloe Cycad Nursery*.
- [Chessel et al., 2009] Chessel, D., Dufour, A.-B., y Dray, S. (2009). Analysis of ecological data: Exploratory and euclidean methods in environmental sciences. *Version 1.4-14. 2 October 2010; 1: 4*, 11.
- [Cornejo-Romero, 2013] Cornejo-Romero, A. (2013). Estructura genética y filogeografía de dos cactus endémicos del valle de tehuacán-cuicatlán. *Universidad Autónoma Metropolitana*.
- [De Queiroz, 2007] De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology*, 56(6):879–886.
- [Domínguez y Ruz, 2008] Domínguez, R. M. y Ruz, C. B. (2008). Hábitos alimenticios de *ara militaris* en la reserva de la biosfera tehuacán-cuicatlán, México. *Mesoamericana*, 20(4).
- [Dyer, 2014a] Dyer, R. (2014a). *gstudio*: Analyses and functions related to the spatial analysis of genetic marker data. URL <http://CRAN.R-project.org/package=gstudio>, r package version, 1.
- [Dyer, 2014b] Dyer, R. (2014b). *popgraph*: R package that constructs and manipulates population graphs. r package version 1.4.
- [Earl et al., 2012] Earl, D. A. et al. (2012). Structure harvester: a website and program for visualizing structure output and implementing the evanno method. *Conservation genetics resources*, 4(2):359–361.
- [Eckenwalder, 1980] Eckenwalder, J. E. (1980). Dispersal of the west indian cycad, *zamia pumila* l. *Biotropica*.
- [Ellegren, 2000] Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in genetics*, 16(12):551–558.
- [Ellegren, 2004] Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature reviews genetics*, 5(6):435–445.
- [Ellis, 2010] Ellis, J. M. S. (2010). White-throated magpie-jay (*calocitta formosa*), neotropical birds online. [http://neotropical.birds.cornell.edu/portal/species/overview?p\\_p\\_spp=514796](http://neotropical.birds.cornell.edu/portal/species/overview?p_p_spp=514796). Online; accessed[2016-08-25].
- [Evanno et al., 2005] Evanno, G., Regnaut, S., y Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular ecology*, 14(8):2611–2620.
- [Excoffier y Lischer, 2010] Excoffier, L. y Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. *Molecular ecology resources*, 10(3):564–567.
- [Falush et al., 2003] Falush, D., Stephens, M., y Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4):1567–1587.
- [Falush et al., 2007] Falush, D., Stephens, M., y Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular ecology notes*, 7(4):574–578.

- [Fitzpatrick et al., 2011] Fitzpatrick, J., Carlon, D., Lippe, C., y Robertson, D. (2011). The west pacific diversity hotspot as a source or sink for new species? population genetic insights from the indo-pacific parrotfish *scarus rubroviolaceus*. *Molecular Ecology*, 20(2):219–234.
- [Foll y Gaggiotti, 2008] Foll, M. y Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics*, 180(2):977–993.
- [García et al., 2016] García, Mario Valerio Velasco y Valdez-Hernández, J. I., Ramírez-Herrera, C., Hernández-Hernández, M. L., López-Upton, J., López-Mata, L., y López-Sánchez, H. (2016). Estructura, heterogeneidad de estadios y patrón de dispersión espacial de *dioon holmgrenii* (zamiaceae). *Botanical Sciences*, 94(1):75–87.
- [Garza y Williamson, 2001] Garza, J. y Williamson, E. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular ecology*, 10(2):305–318.
- [González et al., 2008] González, D., Vovides, A. P., y Bárcenas, C. (2008). Phylogenetic relationships of the neotropical genus *dioon* (cycadales, zamiaceae) based on nuclear and chloroplast dna sequence data. *Systematic Botany*, 33(2):229–236.
- [González-Astorga et al., 2008] González-Astorga, J., Vergara-Silva, F., Vovides, A. P., Nicolalde-Morejón, F., Carera-Toledo, D., y Pérez-Farrera, M. A. (2008). Diversity and genetic structure of three species of *dioon* lindl.(zamiaceae, cycadales) from the pacific seaboard of mexico. *Biological Journal of the Linnean Society*, 94(4):765–776.
- [Gonzalez-Astorga et al., 2009] Gonzalez-Astorga, J., Vovides, A. P., Cabrera-Toledo, D., y Nicolalde-Morejón, F. (2009). Diversity and genetic structure of the endangered cycad *dioon sonorensis* (zamiaceae) from sonora, mexico: Evolutionary and conservation implications. *Biochemical Systematics and Ecology*, 36(12):891–899.
- [Gonzalez-Astorga et al., 2005] Gonzalez-Astorga, J., Vovides, A. P., Cruz-Angon, A., Octavio-Aguilar, P., y Iglesias, C. (2005). Allozyme variation in the three extant populations of the narrowly endemic cycad *dioon angustifolium* miq.(zamiaceae) from north-eastern mexico. *Annals of Botany*, 95(6):999–1007.
- [Gonzalez-Astorga et al., 2003] Gonzalez-Astorga, J., Vovides, A. P., Ferrer, M. M., y Iglesias, C. (2003). Population genetics of *dioon edule* lindl.(zamiaceae, cycadales): biogeographical and evolutionary implications. *Biological Journal of the Linnean Society*, 80(3):457–467.
- [González-Astorga et al., 2003] González-Astorga, J., VOVIDES, A. P., y IGLESIAS, C. (2003). Morphological and geographic variation of the cycad *dioon edule* lindl.(zamiaceae): ecological and evolutionary implications. *Botanical Journal of the Linnean Society*, 141(4):465–470.
- [González-Christen, 1990] González-Christen, A. (1990). Algunas interacciones entre *dioon edule* (zamiaceae) y *peromyscus mexicanus* (rodentia: Cricetidae).
- [Gutiérrez-Ortega et al., 2014] Gutiérrez-Ortega, J. S., Kajita, T., y Molina-Freaner, F. E. (2014). Conservation genetics of an endangered cycad, *dioon sonorensis* (zamiaceae): implications from variation of chloroplast dna. *Botanical Sciences*, 92(3):441–451.
- [Haasl y Payseur, 2011] Haasl, R. J. y Payseur, B. A. (2011). Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity*, 106(1):158–171.
- [Hardy, 2003] Hardy, G. (2003). Mendelian proportions in a mixed population. 1908. *The Yale journal of biology and medicine*, 76(2):79.
- [Hedrick, 1999] Hedrick, P. W. (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, pages 313–318.

- [Hedrick, 2005] Hedrick, P. W. (2005). A standardized genetic differentiation measure. *Evolution*, 59(8):1633–1638.
- [Jarne y Lagoda, 1996] Jarne, P. y Lagoda, P. J. (1996). Microsatellites, from molecules to populations and back. *Trends in ecology & evolution*, 11(10):424–429.
- [Jin et al., 1996] Jin, L., Macaubas, C., Hallmayer, J., Kimura, A., y Mignot, E. (1996). Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proceedings of the National Academy of Sciences*, 93(26):15285–15288.
- [Jost, 2008] Jost, L. (2008).  $G_{ST}$  and its relatives do not measure differentiation. *Molecular Ecology*, 17(18):4015–4026.
- [Jost, 2009] Jost, L. (2009).  $D$  vs.  $G_{ST}$ : Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). *Molecular Ecology*, 18(10):2088–2091.
- [Kearse et al., 2012] Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647–1649.
- [Kolpakov y Kucherov, 1999] Kolpakov, R. y Kucherov, G. (1999). Finding maximal repetitions in a word in linear time. In *Foundations of Computer Science, 1999. 40th Annual Symposium on*, pages 596–604. IEEE.
- [Lázaro-Zermeño et al., 2012] Lázaro-Zermeño, J. M., González-Espinosa, M., Mendoza, A., y Martínez-Ramos, M. (2012). Historia natural de *dioon merolae* (zamiaceae) en chiapas, méxico. *Botanical Sciences*, 90(1):73–87.
- [Manel et al., 2003] Manel, S., Schwartz, M. K., Luikart, G., y Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology & evolution*, 18(4):189–197.
- [Marateo et al., 2015] Marateo, G., Juárez, M., Grilli, P., y Pagano, L. (2015). Military macaw (*ara militaris*), neotropical birds online. [http://neotropical.birds.cornell.edu/portal/species/distribution?p\\_spp=23174](http://neotropical.birds.cornell.edu/portal/species/distribution?p_spp=23174). Online; accessed[2016-08-25].
- [Meirmans y Hedrick, 2011] Meirmans, P. G. y Hedrick, P. W. (2011). Assessing population structure:  $F_{ST}$  and related measures. *Molecular Ecology Resources*, 11(1):5–18.
- [Nagalingum et al., 2011] Nagalingum, N., Marshall, C., Quental, T., Rai, H., Little, D., y Mathews, S. (2011). Recent synchronous radiation of a living fossil. *Science*, 334(6057):796–799.
- [NatureServe y Patterson, 2007a] NatureServe y Patterson, B. (2007a). *Peromyscus mexicanus* (ratón mexicano). [http://naturalhistory.si.edu/mna/full\\_image.cfm?image\\_id=2142](http://naturalhistory.si.edu/mna/full_image.cfm?image_id=2142). Online; accessed[2016-08-25].
- [NatureServe y Patterson, 2007b] NatureServe y Patterson, B. (2007b). *Sciurus mamíferos de america del norte*. [http://naturalhistory.si.edu/mna/image\\_menu.cfm?genus=122](http://naturalhistory.si.edu/mna/image_menu.cfm?genus=122). Online; accessed[2016-08-25].
- [Nei, 1973] Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12):3321–3323.
- [Nicolalde-Morejón et al., 2014] Nicolalde-Morejón, F., González-Astorga, J., Vergara-Silva, F., Stevenson, D. W., Rojas-Soto, O., y Medina-Villarreal, A. (2014). Biodiversidad de zamiaceae en México. *Revista mexicana de biodiversidad*, 85:114–125.

- [Nicolalde-Morejón et al., 2009] Nicolalde-Morejón, F., Vergara-Silva, F., González-Astorga, J., Vovides, A. P., y de los Monteros, A. E. (2009). Reciprocal illumination of morphological characters upon a molecular hypothesis supports the proposal of a new species of cycad from Mexico. *Systematics and Biodiversity*, 7(01):73–79.
- [Norstog et al., 1997] Norstog, K., Nicholls, T. J. K. J., y Trevor, J. N. (1997). *The biology of the cycads*.
- [Norstog et al., 1986] Norstog, K. J., Stevenson, D. W., y Niklas, K. J. (1986). The role of beetles in the pollination of *Zamia furfuracea* L. fil. (Zamiaceae). *Biotropica*, pages 300–306.
- [Octavio-Aguilar et al., 2008] Octavio-Aguilar, P., González-Astorga, J., y Vovides, A. P. (2008). Population dynamics of the Mexican cycad *Dioon edule* Lindl. (Zamiaceae): life history stages and management impact. *Botanical Journal of the Linnean Society*, 157(3):381–391.
- [Paradis, 2010] Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3):419–420.
- [Paradis et al., 2004] Paradis, E., Claude, J., y Strimmer, K. (2004). Ape: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2):289–290.
- [Pérez Farrera y Vovides, 2004] Pérez Farrera, M. A. y Vovides, A. P. (2004). Spatial distribution, population structure, and fecundity of *Ceratozamia matudai* Lundell (Zamiaceae) in the Triunfo Biosphere Reserve, Chiapas, Mexico. *The Botanical Review*, 70(2):299–311.
- [Pérez-Farrera y Vovides, 2006] Pérez-Farrera, M. A. y Vovides, A. P. (2006). The ceremonial use of the threatened *âspadañaâ* cycad (*Dioon merolae*, Zamiaceae) by a community of the central depression of Chiapas, Mexico. *Boletín de la Sociedad Botánica de México*, 78:107–113.
- [Pritchard et al., 2000] Pritchard, J. K., Stephens, M., y Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.
- [Pritchard et al., 2007] Pritchard, J. K., Wen, X., y Falush, D. (2007). Documentation for structure software: Version 2.3.
- [Puechmaille, 2016] Puechmaille, S. J. (2016). The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimators alleviate the problem. *Molecular ecology resources*.
- [Rouder et al., 2009] Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., y Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, 16(2):225–237.
- [Rubio-Mendez, 2013] Rubio-Mendez, G. (2013). Estudio para la conservación a largo plazo de *Dioon edule* en San Luis Potosí. *Universidad Autónoma de San Luis Potosí, tesis de maestría*.
- [Salas-Morales et al., 2016] Salas-Morales, S. H., Chemnick, J., y Gregory, T. J. (2016). A new cycad species in the genus *Dioon* (Zamiaceae) from the Mixteca region of Oaxaca, Mexico. *Cactus and Succulent Journal*, 88(1):35–42.
- [Salomé-Castañeda, 2009] Salomé-Castañeda, E. (2009). Parámetros demográficos de *Dioon spinulosum* (Zamiaceae), en San Miguel Soyaltepec, Oaxaca. *Instituto de Ecología, Tesis de Maestría*.
- [Schlötterer, 2004] Schlötterer, C. (2004). The evolution of molecular markers: just a matter of fashion? *Nature Reviews Genetics*, 5(1):63–69.
- [Schlötterer et al., 1998] Schlötterer, C., Ritter, R., Harr, B., y Brem, G. (1998). High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution*, 15(10):1269–1274.

- [Shaw et al., 1999] Shaw, P., Pierce, G., y Boyle, P. (1999). Subtle population structuring within a highly vagile marine invertebrate, the veined squid *loligo forbesi*, demonstrated with microsatellite dna markers. *Molecular Ecology*, 8(3):407–417.
- [Tang, 1989] Tang, W. (1989). Seed dispersal in the cycad *zamia pumila* in florida. *Canadian Journal of Botany*, 67(7):2066–2070.
- [Terry et al., 2005] Terry, L. I., Walter, G. H., Donaldson, J. S., Snow, E., Forster, P. I., y Machin, P. J. A. (2005). Pollination of australian *macrozamia* cycads (zamiaceae): effectiveness and behavior of specialist vectors in a dependent mutualism. *American Journal of Botany*, 92(6):931–940.
- [Trudeau, 2013] Trudeau, R. J. (2013). *Introduction to graph theory*. Courier Corporation.
- [Untergasser et al., 2007] Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., y Leunissen, J. A. (2007). Primer3plus, an enhanced web interface to primer3. *Nucleic acids research*, 35(suppl 2):W71–W74.
- [Van Oosterhout et al., 2004] Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., y Shipley, P. (2004). Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3):535–538.
- [Vovides, 1990] Vovides, A. P. (1990). Spatial distribution, survival, and fecundity of *dioon edule* (zamiaceae) in a tropical deciduous forest in veracruz, mexico, with notes on its habitat. *American Journal of Botany*, pages 1532–1543.
- [Wandeler et al., 2003] Wandeler, P., Smith, S., Morin, P. A., Pettifor, R., y Funk, S. (2003). Patterns of nuclear dna degeneration over time a case study in historic teeth samples. *Molecular Ecology*, 12(4):1087–1093.
- [Wattier et al., 1998] Wattier, R., Engel, C., Saumitou-Laprade, P., y Valero, M. (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus *gv1ct* in *gracilaria gracilis* (rhodophyta). *Molecular Ecology*, 7(11):1569–1573.
- [Weber y Wong, 1993] Weber, J. L. y Wong, C. (1993). Mutation of human short tandem repeats. *Human molecular genetics*, 2(8):1123–1128.
- [Weinberg, 1908] Weinberg, W. (1908). *Über den nachweis der vererbung beim menschen*. publisher not identified.
- [Whitlock, 2011] Whitlock, M. C. (2011).  $G'_{ST}$  and  $D$  do not replace  $F_{ST}$ . *Molecular Ecology*, 20(6):1083–1091.
- [Wierdl et al., 1997] Wierdl, M., Dominska, M., y Petes, T. D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*, 146(3):769–779.
- [Wright, 1921] Wright, S. (1921). Systems of mating, I-V. *Genetics*, 6(2):111–178.
- [Wright, 1943] Wright, S. (1943). Isolation by distance. *Genetics*, 28(2):114.
- [Wright, 1949] Wright, S. (1949). The genetical structure of populations. *Annals of eugenics*, 15(1):323–354.
- [Yáñez-Espinosa, 2007] Yáñez-Espinosa, Laura y Sosa-Sosa, F. (2007). Population structure of *dioon purpusii* rose in oaxaca, mexico. *Neotropical Biology and Conservation*, 2(1):46–54.

# Anexos

## A.1. Protocolo de Extracción de ADN total con CTAB

### Soluciones:

- a) Buffer de extracción (cantidades para 500 mL):
  - 10 g de CTAB
  - 10 g de PVP
  - 50 mL de Tris-HCl pH 8.0 1 M
  - 25 mL de EDTA 0.5 M
  - 200 mL NaCl 0.5 M
  - 0.25g de Espermidina\* opcional
  - Aforar a 500 mL con H<sub>2</sub>O
- b) Buffer P3 (cantidades para 1 L):
  - 294.5 g de CH<sub>3</sub>CO<sub>2</sub>K
  - ≈ 500 mL de H<sub>2</sub>O (o suficiente para disolver el CH<sub>3</sub>CO<sub>2</sub>K)
  - ≈ 200 mL de CH<sub>3</sub>COOH (o suficiente para ajustar el pH a 5.5)
  - Ajustar a 1 L con H<sub>2</sub>O
- c) Cloroformo / alcohol isoamilico (24:1)
- d) C<sub>2</sub>H<sub>3</sub>NaO<sub>2</sub> 3 M
- e) Etanol al 100 %
- f) Agua ultrapura o agua tratada con DEPC o agua libre de nucleasas

### Antes de empezar:

- I Precalienta el thermomixer a 65C°.
- II Coloca el etanol a -20C° y el acetato de Na a 4C°.
- III Agrega 2% (v/v) de β-mercaptoetanol al buffer de extracción que se va a utilizar, deseche el sobrante.
- IV No extraiga más de 24 muestras seguidas.

### Procedimiento:

1. En un tubo de 2 mL agrega ≈100g de tejido finamente molido y agrega 500L del buffer de extracción.
2. Incuba de 20 minutos a una hora a 55C°.

3. Centrifuga a 12,000 rcf por 2 minutos y colecta el sobrenadante en un tubo de 1.5mL
4. Agrega 150  $\mu\text{L}$  del buffer P3, agita el tubo e incuba en hielo por 5 minutos.
5. Centrifuga a 12,000 rcf por 2 minutos y colecta el sobrenadante en un tubo de 1.5mL
6. Agrega 500  $\mu\text{L}$  de Cloroformo / alcohol isoamilico y agita en vortex a maxima velocidad por 10 segundos.
7. Centrifuga por 20 minutos a 12,500 rcf.
8. Colecta la fase acuosa en un nuevo tubo de 1.5 mL
9. \*(opcional) repite los pasos 7 y 8
10. Agrega 125  $\mu\text{L}$  de  $\text{C}_2\text{H}_3\text{NaO}_2$  3 M y 1 mL de etanol 100 %, precipita a  $-20\text{C}^\circ$  por 2 horas o a  $-80\text{C}^\circ$  por 30minutos. se puede dejar precipitando toda la noche.
11. Centrifuga a 20,000 rcf por 10 minutos.
12. Desecha el sobrenadante y agrega 150  $\mu\text{L}$  de agua libre de nucleasas, resuspende y agrega 350  $\mu\text{L}$  de etanol 100 % (puedes doblar las cantidades si el ADN no se resuspende en 150  $\mu\text{L}$  de agua)
13. Centrifuga a 20,000 rcf por 10 minutos
14. Desecha el sobrenadante y deja que el etanol se evapore por completo.
15. Resuspende en 50, 100 o 200  $\mu\text{L}$  de agua libre de nucleasas.

## A.2. PCA



Figura A.1: Análisis de componentes principales: el PC1 explica el 8.87% de la variación y el PC3 el 3.3%. Se necesitan 181 PCs para explicar el 100% de la variación.

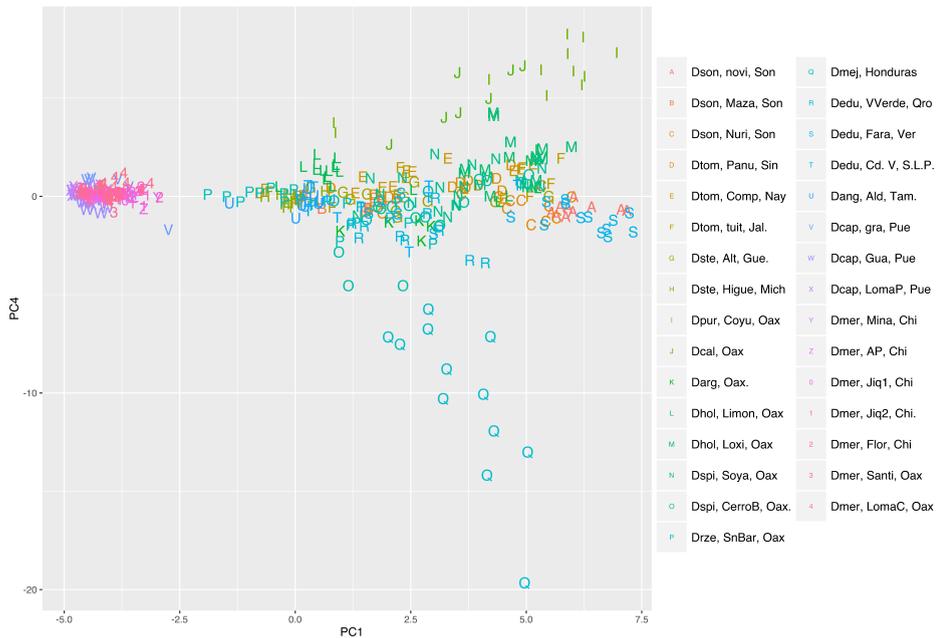


Figura A.2: Análisis de componentes principales: el PC1 explica el 8.87% de la variación y el PC4 el 3.17%. Se necesitan 181 PCs para explicar el 100% de la variación.