



CENTRO DE INVESTIGACIONES BIOLÓGICAS  
DEL NOROESTE, S.C.

---

---

Programa de Estudios de Posgrado

**ANÁLISIS DE LA ESTRUCTURA GENÉTICA  
POBLACIONAL DE LA PITAYA AGRIA (*Stenocereus  
gummosus*) EN EL DESIERTO SONORENSE**

**TESIS**

Que para obtener el grado de

**Maestro en Ciencias**

Uso, Manejo y Preservación de los Recursos Naturales  
(Orientación |Ecología de Zonas Áridas|)

P r e s e n t a

|**Oscar Adrián Lozano Garza**|

La Paz, Baja California Sur, |Marzo| de |2013|

## ACTA DE LIBERACION DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 12:00 horas del día 15 del Mes de Febrero del 2013, se procedió por los abajo firmantes, miembros de la Comisión Revisora de Tesis avalada por la Dirección de Estudios de Posgrado del Centro de Investigaciones Biológicas del Noroeste, S. C., a liberar la Tesis de Grado titulada:

### "ANÁLISIS DE LA ESTRUCTURA GENÉTICA POBLACIONAL DE LA PITAYA AGRIA (*Stenocereus gummosus*) EN EL DESIERTO SONORENSE"

Presentada por el alumno:

#### Oscar Adrián Lozano Garza

Aspirante al Grado de MAESTRO EN CIENCIAS EN EL USO, MANEJO Y PRESERVACION DE LOS RECURSOS NATURALES CON ORIENTACION EN Ecología de Zonas Áridas

Después de intercambiar opiniones los miembros de la Comisión manifestaron su **APROBACION DE LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

#### LA COMISION REVISORA



Francisco Javier García de León  
DIRECTOR DE TESIS



José Luis León de la Luz  
CO-TUTOR



Susana Favela Lara  
CO-TUTOR



DRA. ELISA SERVIERE ZARAGOZA,  
DIRECTORA DE ESTUDIOS DE POSGRADO

## **CONFORMACIÓN DE COMITÉ**

### **COMITÉ TUTORIAL**

**Dr. Francisco Javier García de León**

Director de Tesis

Centro de Investigaciones Biológicas del Noroeste, S. C., La Paz, B.C.S.

**Dr. José Luis León de la Luz**

Co-Tutor

Centro de Investigaciones Biológicas del Noroeste, S. C., La Paz, B.C.S.

**Dra. Susana Favela Lara**

Co-Tutor

Universidad Autónoma de Nuevo León, San Nicolás de los Garza, N.L.

### **COMITÉ REVISOR DE TESIS**

**Dr. Francisco Javier García de León**

**Dr. José Luis León de la Luz**

**Dra. Susana Favela Lara**

### **JURADO EN EXAMEN DE GRADO**

**Dr. Francisco Javier García de León**

**Dr. José Luis León de la Luz**

**Dra. Susana Favela Lara**

**Suplente**

**Dr. Pedro Peña Garcillán**

## RESUMEN

*Stenocereus gummosus* (Englem.) Gibson y Horak, coloquialmente llamada pitaya agria, es una cactácea columnar endémica de México con distribución en la región oeste del Desierto Sonorense, se encuentra presente en casi todo lo largo de la Península de Baja California. En general, las plantas de esta región han sido poco estudiadas en términos de estructuración genética, donde los fenómenos de especiación y la filogeografía han sido explicadas principalmente por eventos vicariantes, al grado de reconocer algunos de estos eventos incluso sin contar con evidencias geológicas, por lo que el debate al respecto está vigente.

La falta de precisión sobre el uso de la estrategia de reproducción vegetativa por parte de *S. gummosus*, ha hecho pensar que su diversidad genética es baja debido a la dispersión clonal, además se ha supuesto que el aumento de la temperatura postglacial pudo ocasionar una expansión hacia las localidades norteñas de su actual distribución, como se ha registrado para otras comunidades vegetales en el hemisferio norte.

Empleando marcadores microsatélite para analizar la genética poblacional de la pitaya agria, se encontró que la diversidad genética describe un patrón de estructuración de aislamiento por distancia coherente con una expansión poblacional con el flujo genético limitado por la dispersión; no se encontraron señales de estructuración fuertes atribuibles a barreras geográficas y los resultados además sugieren eventos de dispersión a larga distancia durante la colonización. Aunque la diversidad genética en la Península fue mayor en la región sur, se discute la posibilidad de la existencia de más de un refugio durante el último máximo glacial.

Palabras clave: *Stenocereus gummosus*; microsatélites; estructura poblacional.

## ABSTRACT

*Stenocereus gummosus* (Englem.) Gibson and Horak is a Mexican endemic columnar cactus with the vernacular name of “sour pitahaya” and a natural distribution in the western half of the Sonoran Desert, mostly found all along the Baja California Peninsula. The plants of this region have been scarcely investigated in terms of genetic structure, while speciation and phylogeography for this area have been explained mainly by vicariant events. Although some of those events have been recognized by phylogeographical breaks, there is no geological evidence for them, which is why the debate about this matter is opened.

The absence of certainty about the reproduction strategy usage for *S. gummosus* has lead to suspect that its genetic diversity may be compromised because of its clonal reproduction ability. In addition, it has been supposed that the postglacial warming process could have facilitated a recent northward expansion range to complete its current distribution, as it has been found happens to other plant communities on the north hemisphere.

Through the analysis of microsatellite molecular markers to infer about the population structure of the sour pitahaya, it has been found that its genetic diversity follows a pattern of isolation by distance, coherent with a population expansion with restricted genetic flow by organism dispersal. No strong structuration due to geographic barriers has been detected, and results also suggest the participation of long distance dispersal events during the colonization. Even though genetic diversity was higher at the southern peninsular region, a potential model of multiple refuges during the last glacial maximum is discussed.

Key words: *Stenocereus gummosus*; microsatellites; population structure.

**DEDICATORIA**

*... a la vida y a México que alberga mucha de la que me inspira y me da vida.*

## AGRADECIMIENTOS

Agradezco primeramente a las instituciones y en especial a la gente que las hace funcionar, al CONACyT y al CIBNOR, por cumplir su responsabilidad de impulsar y producir conocimiento, financiando mi manutención con becas y otros apoyos económicos, y por facilitarme espacio físico y condiciones para desarrollar este trabajo. Al proyecto de Filogeografía Comparada en el Noroeste de México y su responsable (Francisco J. García de León), por financiar la parte experimental y parte del taller ReGeneC2012 y a José Luis León de la Luz por su apoyo económico y logístico en los muestreos, en el mismo taller y en el Congreso 32 de la Sociedad Internacional para el Estudio de Cactáceas y Suculentas.

Al equipo del Herbario CIBNOR (José Luis León, Raymundo Domínguez y Alfonso Medel, más Carina Gutiérrez) por su enorme apoyo para los muestreos de campo; incluidos a todos los demás colaboradores de recolectas: Emiliano, Tania, Adal, Arturo y Oscar. Al personal de Posgrado CIBNOR, en particular a Horacio por su servicio siempre tan presto y atento y a Diana Dorantes por su revisión y edición del resumen en inglés.

A Paco mi director, por su constante atención y facilitamiento de oportunidades para mi desarrollo profesional y por creer en mi esfuerzo a pesar de las dificultades técnicas. A todo el comité tutorial (incluido Pedro, el suplente), y a los profesores de las asignaturas que cursé, por sus enseñanzas, consejos, observaciones y recomendaciones.

¡A toda la familia GenCon! A quienes les debería de dedicar como dos hojas de descripción por todo el apoyo, pero ya saben que valoro muchísimo todas sus enormes participaciones (y también las pequeñas): Miguel, Sarai, Caro, Fausto, Cary, David, Emiliano, Tania, Cristina, Paulina y las estancieras Tzitziki y Laura. De colado al Quina que también echa la mano. No está de más agradecerles también todos los buenos y reconfortantes ratos fuera y lejos del laboratorio.

A mis compañeros de generación y de vida por promover a la salud mental: desde los LBG que me echaron la mano con revisiones y críticas a mi trabajo (Babis, Arcadio y Ernesto), los de maestría por todo el tiempo compartido, el apoyo moral, la buena vibra y la buena fiesta (Arturo, Dany, Dome, Ely, Emiliano, Jorge, Jose, Magy, Matus, Rubén y Tona) y a los del cubículo extendido por lo correspondiente y por generar un excelente espacio compartido (Anita (tu llevas agradecimiento extra), Diana, Fercho, Laura, Quinatzin, Yarelis, Yarelys).

A la Manada Cazadora Kiliwa y al grupo Scout 11 Kiliwa B.C.S., que me regaló el espacio para agarrar fuerzas e inspiración. Otra enorme lista de gente.

A mi familia del árbol y adjuntos que no necesito nombrar ni explicar mi agradecimiento, los llevo en la piel, en el corazón y un lado en el camino. A mi familia otorgada, por el grandísimo soporte, empuje, aceptación y amor que no paran de invitarme a seguir caminando.

**¡A todos gracias y buenas vibras!**

## CONTENIDO

<b>RESUMEN</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>ii</b>
<b>DEDICATORIA</b> .....	<b>iii</b>
<b>AGRADECIMIENTOS</b> .....	<b>iv</b>
<b>CONTENIDO</b> .....	<b>v</b>
<b>LISTA DE FIGURAS</b> .....	<b>vii</b>
<b>LISTA DE TABLAS</b> .....	<b>viii</b>
<b>1. INTRODUCCIÓN</b> .....	<b>1</b>
<b>2. ANTECEDENTES</b> .....	<b>4</b>
2.1. Especie de estudio.....	4
2.2. Estudios genéticos.....	6
2.3. Otras plantas, otras <i>Stenocereus</i> y otras regiones .....	7
2.4. Microsatélites en otras cactáceas .....	9
<b>3. JUSTIFICACIÓN</b> .....	<b>11</b>
<b>4. OBJETIVOS</b> .....	<b>12</b>
4.1. General.....	12
4.2. Particulares.....	12
<b>5. HIPÓTESIS</b> .....	<b>13</b>
<b>6. MATERIALES Y MÉTODOS</b> .....	<b>14</b>
6.1. Área de estudio .....	14
6.2. Trabajo de campo.....	14
6.3. Trabajo de laboratorio. Estandarizaciones .....	16
6.3.1. Extracción de ácidos nucleicos. ....	16
6.3.2. Secuenciación y diseño de iniciadores de PCR para microsatélites.....	16
6.3.3. Condiciones de amplificación por PCR. ....	20
6.3.4. Electroforesis en gel de acrilamida con tinción de plata. ....	20
6.3.5. Pruebas de polimorfismo por PCR.....	21
6.3.6. PCR con iniciadores marcados por fluoróforos. ....	21
6.3.7. Electroforesis en capilar automatizada.....	21
6.4. Trabajo de laboratorio. Genotipificación.....	22
6.4.1. Amplificación y lectura electroforética .....	22
6.4.2. Determinación de rangos alélicos y reglas de lectura automatizada. ....	23
6.5. Análisis de datos .....	23
6.5.1. Calidad y utilidad de los marcadores. ....	23
a. Detección de Alelos nulos.....	23
b. Pruebas de desequilibrio de ligamiento.....	24
6.5.2. Estimación de parámetros poblaciones .....	24
a. Diversidad genética por localidad .....	24



b. Equilibrio de Hardy-Weinberg.....	24
6.5.3. Análisis de estructuración poblacional.....	25
a. Análisis Factorial de Correspondencia (AFC) y Análisis Discriminante de Componentes Principales (ADCP).....	25
b. Análisis de Varianza Molecular (AMOVA).....	25
c. Aislamiento por distancia.....	27
d. Dendograma de distancia genética.....	27
e. Índices de Fijación de Wright.....	28
f. Forma teórica de la estructura genética.....	28
g. Análisis de clusters.....	29
<b>7. RESULTADOS.....</b>	<b>30</b>
7.1. Estandarización.....	30
7.1.1. Extracción de ADN.....	30
7.1.2. Secuenciación y diseño de iniciadores.....	30
7.1.3. Pruebas de amplificación y polimorfismo.....	31
7.2. Genotipificación.....	32
7.3. Calidad y utilidad de los marcadores genotipificados.....	33
7.3.1. Detección de alelos nulos.....	33
7.3.2. Pruebas de desequilibrio de ligamiento.....	35
7.4. Diversidad por localidad.....	35
7.5. Análisis de estructura poblacional.....	41
7.5.1. Análisis Factorial de Correspondencia (AFC) y Análisis Discriminante de Componentes Principales (ADCP).....	41
7.5.2. Análisis de varianza molecular (AMOVA).....	43
7.5.3. Aislamiento por distancia.....	45
7.5.4. Dendograma de distancia genética.....	45
7.5.5. Índices de Fijación de Wright ( $F_{ST}$ ).....	46
7.5.6. Forma teórica de la estructura genética.....	47
7.5.7. Análisis de clusters.....	48
<b>8. DISCUSIÓN.....</b>	<b>50</b>
8.1. Desarrollo y calidad de los marcadores.....	50
8.2. Estructuración genética, expansión poblacional y colonización sur-norte.....	52
<b>9. CONCLUSIONES.....</b>	<b>59</b>
<b>10. LITERATURA CITADA.....</b>	<b>61</b>
<b>ANEXO I. EXTRACCIÓN DE ADN PARA <i>Stenocereus gummosus</i>.....</b>	<b>68</b>
<b>ANEXO II. EJEMPLOS DE ELECTROFEROGRAMAS CON TAMAÑOS ALÉLICOS ASIGNADOS.....</b>	<b>70</b>
<b>ANEXO III. FORMA TEÓRICA DE LA ESTRUCTURA GENÉTICA PARA ALFA EN INCREMENTO.....</b>	<b>71</b>

## LISTA DE FIGURAS

<b>Figura 1.</b> a) Mapa de distribución de <i>Stenocereus gummosus</i> y eventos vicariantes propuestos para otros taxa. b), c) y d) Fotografías de <i>S. gummosus</i> .....	2
<b>Figura 2.</b> Mapa de sitios de recolecta.....	15
<b>Figura 3.</b> Eventos vicariantes considerados para modelos de AMOVA. ....	26
<b>Figura 4.</b> Distribución de frecuencias por longitud de lecturas de pirosecuenciación .....	30
<b>Figura 5.</b> Éxito de amplificación acumulativo .....	32
<b>Figura 6.</b> Comparación de heterocigocidad con 7 loci y con 6 loci, excluyendo Sgum05 .....	34
<b>Figura 7.</b> a) Rarefacción de la riqueza alélica. b) Gradiente latitudinal de $R   S = 18$ .....	40
<b>Figura 8.</b> Dispersión de muestras a) por localidad (AFC) y b) por grupo regional (ADCP).....	42
<b>Figura 9.</b> Resultados de AMOVA.....	44
<b>Figura 10.</b> Correlación entre distancia genética y geográfica por pares de localidades.....	44
<b>Figura 11.</b> Cladograma NJ de distancias genéticas .....	45
<b>Figura 12.</b> <i>Population Graph</i> .....	48
<b>Figura 13.</b> Resultados de Structure Harvester de probabilidad para $K$ : a) $Ln(P K)$ y b) $\Delta K$ .....	49
<b>Figura 14.</b> Gráficos de barras de Structure para a) $K = 2$ y b) $K = 5$ .....	49

## LISTA DE TABLAS

<b>Tabla I.</b> Loci microsatélite de cactáceas.....	10
<b>Tabla II.</b> Localidades de muestreo. ....	14
<b>Tabla III.</b> Loci adquiridos para pruebas de amplificación y polimorfismo .....	17
<b>Tabla IV.</b> Iniciadores de PCR con fluoróforo acoplado y juegos de lectura electroforética.....	22
<b>Tabla V.</b> Modelos de agrupación jerárquica comparados por medio del AMOVA.....	27
<b>Tabla VI.</b> Condiciones de amplificación para loci con éxito mayor a 50%.....	31
<b>Tabla VII.</b> Éxito de amplificación de muestras genotipificadas por ensayo.....	32
<b>Tabla VIII.</b> Muestras por población incorporadas al análisis .....	33
<b>Tabla IX.</b> Frecuencia de alelos nulos por loci, por localidad .....	34
<b>Tabla X.</b> Prueba de desequilibrio de ligamiento .....	35
<b>Tabla XI.</b> Diversidad genética por loci, por localidad .....	36
<b>Tabla XII.</b> Diversidad genética por localidad (promedios 6 loci).....	39
<b>Tabla XIII.</b> Resultados de AMOVAS.....	43
<b>Tabla XIV.</b> Índices de fijación de Wright $F_{ST}$ .....	47

## 1. INTRODUCCIÓN

*Stenocereus gummosus* (Englem.) Gibson y Horak, coloquialmente llamada pitaya agria, es una cactácea columnar endémica de México con distribución silvestre en la región oeste del Desierto Sonorense (Figura 1a): se encuentra a lo largo de la Península de Baja California (PBC), desde la región de Los Cabos en el extremo sur hasta los 32° de latitud norte, además de ubicarse en un área relativamente pequeña en la zona costera del estado de Sonora y en algunas de las islas del Golfo de California (Turner et al., 1995).

La pitahaya agria, ampliamente presente en 8 de las 14 ecorregiones de la PBC propuestas por González-Abraham y colaboradores (2010), comparte su patrón de distribución interrumpido por el Golfo de California con otras plantas características de la PBC, por lo que se ha discutido sobre los eventos que conformaron la distribución actual de las poblaciones de estas especies y sobre la direccionalidad de la dispersión (Cody et al., 1983; Gibson, 1989), hecho que puede representar un patrón biogeográfico que ayudaría a explicar la migración de vegetales y la conformación de las actuales regiones desérticas. En particular para *S. gummosus*, se ha propuesto que las islas de la parte central del golfo han funcionado como un puente para la migración desde la península hacia la costa de la placa continental. Particularmente en la PBC, por análisis de loci aloenzimáticos polimórficos, se encontró una mayor relación entre las localidades del centro de la PBC con las continentales e insulares y se detectó una incipiente señal de disminución de la heterocigocidad esperada con el aumento de la latitud (Clark-Tapia y Molina-Freaner, 2003) sin significancia estadística al analizar exclusivamente las localidades peninsulares.

Por otra parte, la estructura poblacional y la filogenia de algunos vertebrados del Desierto Sonorense han sido explicadas en lo general por eventos vicariantes (Riddle et al., 2000; Murphy y Aguirre-León, 2002). En particular para la PBC, además de la del Golfo de California, se han propuesto los tres siguientes eventos (Figura 1a): 1) la presencia de un canal marino a mitad de la Península sobre el actual Desierto Vizcaíno, durante el Pleistoceno medio; 2) una transgresión del mar sobre el delta del Río Colorado, hacia el Plioceno tardío y 3) durante el Plioceno reiteradas transgresiones marinas sobre el Istmo de La Paz, separando el extremo sur del de la Península (Riddle et al., 2000). Aún se encuentra en debate la existencia de estos eventos debido a falta de evidencia geológica (Jacobs et al., 2004; Lindell et al., 2006). La apertura del Golfo de California y los propuestos canales transpeninsulares, se pudieron haber completado hace 5 M y 1 M años atrás respectivamente (Carreño y Helenes, 2002). Pero posteriormente a estos eventos vicariantes

atribuidos a cambios geológicos, sucedieron reiterados cambios climáticos durante el Pleistoceno (último millón de años), caracterizado por la alternancia de condiciones glaciales y sus respectivos períodos interglaciares, que definitivamente también debieron influenciar la estructuración genética en la región peninsular (Hewitt, 1996; Petit et al., 1999).

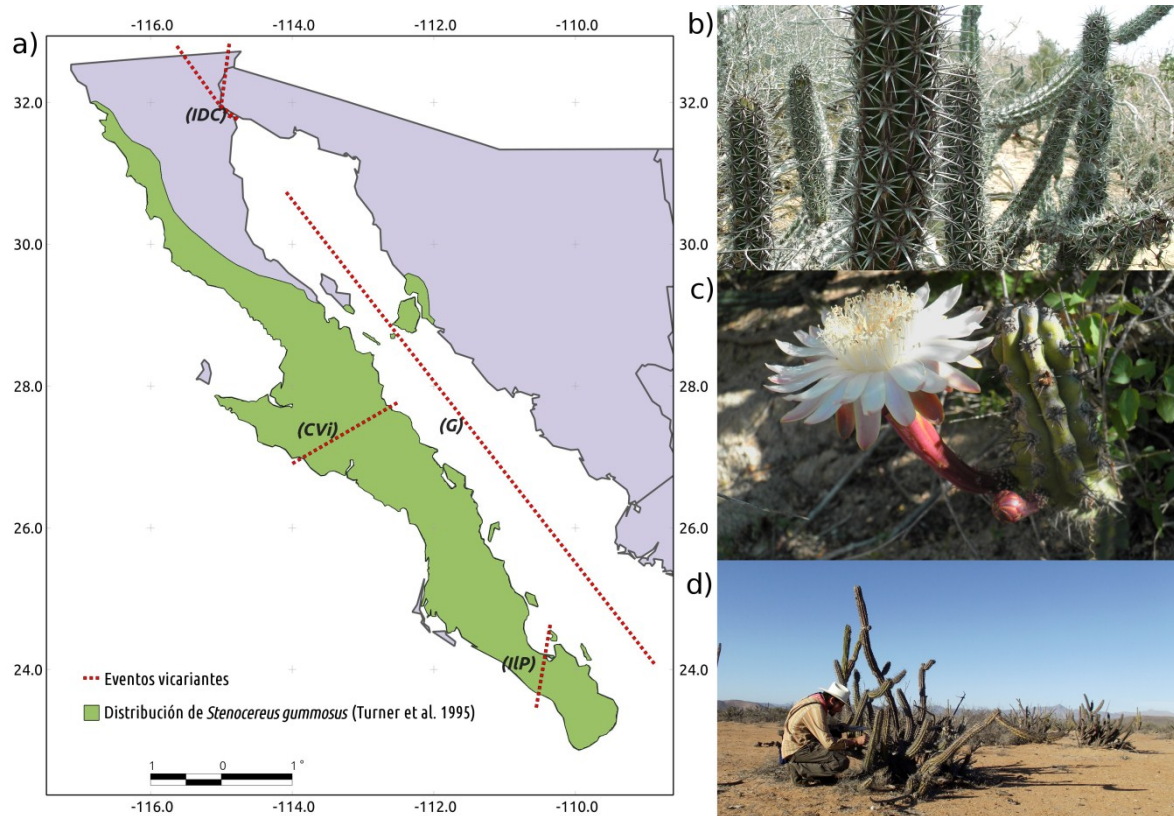


Figura 1. a) Mapa de distribución de *Stenocereus gummosus* y representación de eventos vicariantes propuestos para otros taxa: Golfo de California (G), Canal del Vizcaíno (CVi), inundación marina del delta del Río Colorado (IDC) e inundación del Istmo de la Paz (IIP). b) Fotografía de matorrales de *S. gummosus*. c) Fotografía de flor de *S. gummosus*. d) Fotografía del muestreo de *S. gummosus*.

Sin embargo, el grupo de las plantas vasculares, con alrededor de 30% de endemismos en la región peninsular del Desierto Sonorense (Rebman y Roberts, 2012), ha recibido menos atención con respecto a estudios de filogeografía y genética poblacional. Una importante proporción de estos vegetales son de afinidad neo-tropical, condición que los hace particularmente susceptibles a las temperaturas de congelamiento (Van Devender, 2002); por tal razón, es más probable que la distribución actual haya sido alcanzada por colonizaciones recientes, y que la diversidad genética entre las poblaciones o regiones detectada hoy en día haya sido determinada por la expansión poblacional y no por barreras más antiguas que limitaran el flujo entre poblaciones (Petit et al.,

2001a). Por ejemplo, la información palinológica y restos de macrofósiles en cuevas de la rata coleccionista o cambista *Neotoma sp.*, evidencian expansiones y retracciones de la comunidad vegetal en la zona de los desiertos cálidos de Norte América como efecto del cambio climático postglacial del Holoceno (Van Devender et al., 1994; Thompson y Anderson, 2000). En cuanto a información molecular, dos plantas características de zonas áridas, *Lophocereus schottii* (Cactaceae) y *Euphorbia lomelii* (Euphorbaceae), ambas encontradas en el Desierto Sonorense, también presentan patrones de estructuración genética mayormente concordante con causas climáticas recientes que con los eventos vicariantes antes mencionados (Nason et al., 2002; Garrick et al., 2009).

Por otro lado, los datos palinológicos además sugieren la existencia de pequeños semidesiertos a bajas altitudes en la actual extensión de los desiertos de Norteamérica, durante el último máximo glacial del Holoceno (Van Devender et al., 1994; Thompson y Anderson, 2000), lo que significaría la existencia de pequeños refugios locales capaces de albergar la diversidad genética de *Stenocereus gummosus* a distintas latitudes de la Península, sitios a partir de donde surgiría la distribución actual.

La presente investigación analiza la estructura genética poblacional de *Stenocereus gummosus* por medio de marcadores microsatélite, para verificar si su diversidad genética corresponde a una expansión poblacional reciente, significando una pérdida de diversidad genética en el sentido de la colonización.

## 2. ANTECEDENTES

### 2.1. Especie de estudio

*Stenocereus gummosus*, aunque considerada cactácea columnar, presenta una forma de crecimiento de ramificación bastiónica o decumbente (Vázquez Sánchez et al., 2012), forma matorrales espinosos de hasta 10 m de diámetro y 1 a 3 m de alto variando con la altura de la vegetación circundante (Figura 1b, 1d). La corteza del tallo es mucilaginoso debido al contenido de polisacáridos y polifenoles (Robichaux, 1999). Sus flores son autoincompatibles, tubulares de color blanco o rosado pálido, de hasta 6 cm de ancho y 11 de largo, y presentan antesis nocturna de una única noche y la mañana posterior a la apertura para polinización (Figura 1c); la floración tiene lugar especialmente entre julio y septiembre y los frutos se desarrollan completamente de uno a dos meses después de la fecundación (Wiggins, 1980; Clark-Tapia y Molina-Freaner, 2003). Su polinización parece ser generalista, comúnmente mediada por lepidópteros de la familia Sphingidae y por coleópteros Nitidulidae y Cantharidae (Gibson y Horak, 1978; León de La Luz y Domínguez-Cadena, 1991), pero dado que comparten características florales con otras plantas quiropterófilas, podrían también ser polinizadas por murciélagos (Del Coro Arizmendi et al., 2002), aunque no hay datos que lo soporten, ya que la recompensa de néctar y polen es de bajo nivel para un organismo de esas dimensiones.

Los individuos adultos cuentan con capacidad de reproducción vegetativa. Sus clones o rametos forman raíces adventicias en segmentos añosos del tallo, siguiendo por un desprendimiento paulatino del organismo del que se originan (Gibson y Nobel, 1986). A pesar de su producción abundante de flores, frutos y de semillas viables, la baja proporción de observaciones de desarrollo de plántulas en el campo no podría explicar el éxito poblacional de esta especie en su actual área de distribución sin invocar la propagación clonal (Gibson, 1989; León de La Luz y Domínguez-Cadena, 1991), sin embargo la diversidad genotípica indica que la especie combina ambas estrategias de reproducción, sexual y clonal, para la regeneración de individuos (Molina Freaner y Clark Tapia, 2005). Una evaluación de la resistencia a tratamientos de deshidratación-hidratación de plántulas de *S. gummosus*, encontró una notable capacidad de ese estadio para resistir las condiciones del medio natural que habita (León de La Luz y Domínguez-Cadena, 1991) descartando las condiciones ambientales de aridez como causa del bajo registro de plántulas en campo, por lo que el consumo de semillas y plántulas pudiera ser una causa que limita su propagación.

Los frutos, llamados pitayas, son un recurso alimenticio abundante en las comunidades vegetales peninsulares (alrededor del 30% de los botones florales producidos), lo que los hace accesibles a consumidores variables, algunos dispersores y otros propiamente herbívoros. La cantidad de semilla varía en función al peso y al volumen del fruto con un promedio estimado de 674 por fruto, con una viabilidad natural estimada en 87.5%. Entre los consumidores del fruto y las semillas, se ha reportado a los roedores de los géneros *Perognathus*, *Ammospermophilus* y *Dipodomys*, quienes abren las semillas para aprovechar su contenido; algunas especies de hormigas almacenan las semillas en sus colonias para su consumo posterior; en cualquiera de estos dos casos, solo un bajo porcentaje de semillas (las que escapan al consumo) tienen la oportunidad de al menos germinar y eventualmente establecerse si se desarrollan en algún sitio pertinente. Las aves en cambio, principalmente paserínidos e ictéridos, consumen la pulpa del fruto (los restos arillados de las semillas), excretándolas ya limpias y escarificadas con el embrión intacto. A partir de excretas del carpintero *Melanerpes uropygialis* (Picidae) se obtuvo un índice de 85% de viabilidad (León de La Luz y Domínguez-Cadena, 1991) lo cual sugiere y soporta que las aves consumidoras de pitayas son los mejores dispersores de sus semillas.

Dos hipótesis han sido propuestas para explicar el origen de *S. gummosus*. La primera plantea que una población vicariante de *S. stellatus* especiara posterior a la separación de la Península; la otra hipótesis plantea una diferenciación previa a la apertura del Golfo (Gibson, 1989). La propuesta de origen de las poblaciones en la costa de Sonora (Cody et al., 1983) y los datos de frecuencias alélicas de aloenzimas (Clark-Tapia y Molina-Freaner, 2003) tienen mayor coherencia con la primera hipótesis del origen de la especie. Por otra parte, se ha encontrado que *S. eruca*, ha especiado recientemente a partir *S. gummosus*, respondiendo selectivamente a las condiciones edafológicas y microclimáticas muy particulares de las paleodunas en los llanos de Bahía Magdalena (Gibson, 1989).

## 2.2. Estudios genéticos

Clark-Tapia y Molina-Freaner (2003), detectaron un déficit de heterocigotos generalizado en las localidades analizadas con 11 loci polimórficos de aloenzimas ( $H_O$  entre 0.067–0.175,  $H_E$  entre 0.207–0.351) y un promedio alto de coeficiente de endogamia ( $F_{IS} = 0.608$ ,  $SD_{F_{IS}} = 0.05$ ) que no puede ser atribuido a autofertilización, pues las flores son auto-incompatibles. Las potenciales explicaciones para estos valores podrían encontrarse en: el entrecruzamiento entre plantas estrechamente emparentadas, en ventaja selectiva de homocigotos, en que la estructura clonal sea



suficientemente fuerte para afectar la escala poblacional o que ésta no haya sido tomada en cuenta para el esquema de muestreo, en que existan variaciones geográficas en el sistema de reproducción o inclusive en un patrón de aislamiento por distancia hacia adentro de las poblaciones con estructuras familiares o vecindarios locales muy amplios, efecto de la polinización mediada por insectos de baja dispersión.

Los niveles de diferenciación genética entre las localidades muestreadas fueron significativos pero en general bajos ( $G_{ST} = 0.102$ ), esos datos sugieren altos niveles de flujo genético actual o una expansión reciente de las poblaciones (Clark-Tapia y Molina-Freaner, 2003), además de ser cercanos a los observados en otras cactáceas columnares con marcadores similares, en particular más parecidos a especies de polinización generalista por murciélagos, aves y abejas (Parker y Hamrick, 1992; Nassar et al., 1997, 2001; Holland y Fleming, 1998; Nason et al., 2002).

El análisis de la identidad clonal para *S. gummosus* y su especie hermana *S. eruca*, identificada por los mismos marcadores aloenzimáticos anteriormente mencionados, (Molina Freaner y Clark Tapia, 2005) encontró para *S. gummosus* una prevalencia baja de clones en el espacio muestreado (100 x 50 m) con sólo dos genets repetidos una vez, dos rametos con genotipo multilocus idéntico separados por 6.6 m y otro par por 9.3 m; en cambio *S. eruca* en un área mayor (100 m<sup>2</sup>) y un muestreo menormente saturado, contó con 8 genets iterados, el más abundante repetido 21 veces con una distancia máxima de 92 m entre rametos, describiendo una distribución agregada pero con cierto grado de entremezcla, según la clasificación de Carrillo-Ángeles y Del Carmen Mandujano (2011). No obstante ha sido recomendado el tomar con cautela la información de marcadores de bajo polimorfismo y, en particular de aloenzimas, al hacer inferencias sobre la proporción de reproducción sexual debido a la capacidad limitada que tienen para distinguir efectivamente entre genotipos verdaderamente idénticos sobreestimando la cuenta de rametos idénticos (Arnaud-Haond et al., 2005) y por el contrario, las diferencias debidas a mutaciones somáticas entre rametos del mismo geneto no pueden ser identificadas por aloenzimas, subestimando la cuenta de clones (Arnaud-Haond et al., 2007).

### **2.3. Otras plantas, otras *Stenocereus* y otras regiones**

Para la Península de Baja California, la historia biogeográfica en función a la genética poblacional ha sido evaluada para otras dos plantas: *Lophocereus schottii*, cactácea distribuida en rangos de latitud similares a *S. gummosus* tanto en la PBC como con la costa de Sonora, de

polinización mediada específicamente por polillas, con capacidad de reproducción clonal de arquitectura agregada con esporádicas dispersiones mayores de rametos atribuidas a corrientes y arroyos intermitentes (Parker y Hamrick, 1992; Holland y Fleming, 1998); y *Euphorbia lomelii*, euforbiácea polinizada por colibríes, con distribución natural en la PBC desde el extremo sur hasta los 29.5° de latitud y con pequeñas poblaciones dispersas a lo largo de la costa del estado de Sonora (Dressler, 1957). A diferencia de las discontinuidades biogeográficas en animales explicados por los supuestos eventos vicariantes (Riddle et al., 2000), estas dos plantas exhibieron patrones de estructuración genética mucho más relacionados con flujo genético actual y/o con expansiones poblacionales recientes, debido al clima postglacial, aunque la direccionalidad de la dispersión fue diferente; *L. schottii* debió migrar en dirección sur-norte y *E. lomelii* en dirección opuesta, encontrándose las poblaciones más recientes hacia la región de Los Cabos. En el caso de *E. lomeli* también se detectó una ligera pero significativamente señal de estructura poblacional con RFLP de intrones nucleares, atribuida a la inundación del Istmo de La Paz, al sur de la PBC (Nason et al., 2002; Dyer y Nason, 2004; Garrick et al., 2009).

El valle de Tehuacán-Cuicatlán al centro de México aloja muchas especies de cactáceas, entre ellas algunas columnares, para las cuales se han documentado diversos usos humanos desde hace 10,000 años (Casas et al., 1999a). Por ejemplo, *Stenocereus stellatus* y *S. pruinosus* subsisten en medio silvestre además de contar con una historia de domesticación con selección artificial y manejo de especímenes de vida libre (Casas et al., 1999b; Parra et al., 2008). Estudios genéticos comparativos de poblaciones silvestres, manejadas y cultivadas han encontrado que el manejo y cultivo tradicionales de *S. pruinosus* ha permitido mantener la diversidad genética de esta especie a pesar de la selección morfológica, y los niveles de flujo genético entre las diferentes poblaciones estudiadas fue notablemente alto ( $Nm_{Fst} > 2$ ) (Parra et al., 2010). En el caso de *S. stellatus*, tanto en el Valle de Tehuacán como en La Mixteca (cerca de la costa Pacífico), para un sistema de manejo tradicional similar en jardines caseros y manejo *in situ* en terrenos de silvicultura contrastados con poblaciones silvestres, se encontró que el manejo humano resulta en diversidad significativamente mayor en los jardines caseros ( $H_E = 0.289$ ) y en las zonas de cultivo ( $H_E = 0.270$ ) que en las poblaciones naturales ( $H_E = 0.253$ ) (Casas et al., 2005). En el caso de *S. gummosus*, de la PBC, los frutos son usualmente colectados por campesinos para su consumo, pero no se cuenta con registros de cultivos (León de la Luz et al., 1995), mas es importante mencionar que la cultura Seri valora mucho el fruto de esta planta y se ha documentado la movilización de ramas de especímenes adultos a jardines caseros y el uso tradicional de trozos de la corteza para técnicas de pesca (Felger y Moser, 1985; Anderson et al., 2001).

En términos generales, en el hemisferio norte, la alternancia de los cambios climáticos entre períodos glaciares e interglaciares, propios del Pleistoceno, resultaron en múltiples contracciones y expansiones de la distribución de muchos taxa que encontraron resguardo en *zonas refugio* o zonas con menor impacto en la intensidad de los cambios globales, hacia el frío o el calor, según sea el caso (Hewitt, 1996, 2000). Esto ha implicado que muchas de las actuales comunidades vegetales hayan alcanzado su distribución en muy poco tiempo. En general la ruta de colonización es acompañada por una pérdida de la riqueza alélica en función a la distancia de migración de los fundadores y del número de eventos fundacionales, por efecto de cuellos de botella (Petit et al., 2001a); la heterocigocidad, dependiendo completamente del flujo intrapoblacional de polen, se espera que disminuya mucho menos drásticamente durante una colonización (Nei et al., 1975; Austerlitz et al., 1997). En una planta que depende de los polinizadores para el transporte de sus gametos, es probable encontrar que los genotipos homocigotos aumenten con un proceso de colonización rápido (Hewitt, 1996).

En contraste de lo esperado, se han detectado zonas de reciente colonización con niveles de diversidad genética considerablemente altos en Europa y en Norteamérica (Petit et al., 2004). Esta señal ha sido explicada por la confluencia de dos o más rutas de expansión que aumentan la heterocigocidad al mezclarse (Konnert y Bergmann, 1995; Petit et al., 2001b) y también por elevada frecuencia por generación de dispersión a larga distancia a través de semillas ( $\geq 0.01$  individuos por generación), llamado *efecto de reajuste*, muy por el contrario de lo que sucede con el *efecto de embolismo*, donde la dispersiones de larga distancia es muy esporádica y el efecto fundador deriva en una pérdida de la diversidad muy abrupta en la dirección de la colonización (Bialozyt et al., 2006).

#### **2.4. Microsatélites en otras cactáceas**

En los últimos años, los estudios de genética poblacional han sido mayormente basados en marcadores de ADN, tanto con secuencias como con microsatélites (Schlötterer, 2004), estos últimos, son marcadores nucleares codominantes y su polimorfismo depende del número de repeticiones para un motivo de 2, 3 o hasta 6 nucleótidos, con tasas de mutación estimadas entre  $1 \times 10^{-6}$ – $1 \times 10^{-2}$  por locus por generación, que derivan en altos niveles de polimorfismo, suficientes para estudios genéticos que busquen respuestas a procesos ecológicos actuales (Schlötterer, 2000). Dados los procesos implicados en generar y corregir variantes alélicas de los microsatélites, la tasa

de mutación difiere no sólo entre taxa, sino entre loci según su ubicación en el genoma y entre tamaños alélicos, siendo más susceptibles a mutación los loci con mayor número de repeticiones debido a desfase o resbalón de la ADN polimerasa (Xu et al., 2000; Li et al., 2002; Lai y Sun, 2003; Harr y Schlötterer, 2004). Estos niveles de mutación generan homoplasia, es decir, cuando se llevan a cabo mutaciones convergentes a partir de distintos estados o tamaños alélicos con el modelo de mutación por pasos, por lo que se vuelve imposible detectar diferencias en cuestión de linajes conforme se acumulan las mutaciones, limitando este tipo de marcadores a inferencias históricas recientes conforme mayor sea la tasa de mutación (Rousset, 1996; Estoup et al., 2002).

En la Tabla I se muestran los resultados de polimorfismo y diversidad ( $A$ ,  $H_E$  y  $H_O$ ) obtenidos por marcadores microsatélite para 6 especies de cactáceas de la subfamilia Cactoideae; los primeros dos fueron aislados a partir de librerías genómicas (a y b), los siguientes tres por medio de librerías genómicas enriquecidas para algún motivo repetitivo (c, d y e) y los de *Stenocereus pruinosus* fueron de amplificación cruzada a partir de *P. chichipe*, siendo la única especie de las presentadas sin marcadores específicos. *E. grussoni* y *S. pruinosus* exhibieron menor número promedio de alelos por locus (3.33 y 4.42 respectivamente) y corresponden a las dos especies con más y menos loci analizados (12 y 4) por lo que la comparación debe tomarse con reservas; *M. crucigera* presentó el más alto (8.88). La proporción de loci polimórficos encontrados es muy variable entre los estudios revisados (15–91%).

Tabla I. Número y porcentaje de loci polimórficos (Polim.) y diversidad genética medida en número medio de alelos por locus ( $A$ ), heterocigocidad observada ( $H_O$ ) y esperada ( $H_E$ ) para 4 estudios de caracterización de loci microsatélite (<sup>a</sup>, <sup>b</sup>, <sup>c</sup> y <sup>d</sup>) y 2 estudios poblacionales con loci microsatélite (<sup>e</sup> y <sup>f</sup>). Se especifica la Tribu a la que pertenece cada especie, la región geográfica donde se llevó a cabo el estudio y el número de poblaciones analizadas (n.p.).

Especie	Tribu	Región geográfica	n.p.	Polim.	$A$	$H_O$	$H_E$
<i>Ariocarpus bravoanus</i> <sup>a</sup>	Thelocactineae	S. L. Potosi, México	1	8 (15%)	5.60	0.574	0.512
<i>Echinocactus grusonii</i> <sup>b</sup>	Cactaceae	S. L. Potosi, México	1	12 (21%)	3.33	0.550	0.508
<i>Mammillaria crucigera</i> <sup>c</sup>	Cactaceae	Tehuacán, México	2	8 (53%)	8.88	0.320	0.757
<i>Pilosocereus machrisii</i> <sup>d</sup>	Cactaceae	Caatinga, Brasil	2	10 (91%)	4.75	0.443	0.487
<i>Polaskia chichipe</i> <sup>e</sup>	Pachycereeae	Tehuacán, México	3	7 (70%)	5.93	0.631	0.683
<i>Stenocereus pruinosus</i> <sup>f</sup>	Pachycereeae	Tehuacán, México	3	4 (57%)	4.42	0.670	0.720

<sup>a</sup> (Hughes et al., 2008); <sup>b</sup> (Hardesty et al., 2008); <sup>c</sup> (Solórzano et al., 2009); <sup>d</sup> (Pérez et al., 2011); <sup>e</sup> (Otero-Arnaiz et al., 2005); <sup>f</sup> (Parra et al., 2010).

Para la identificación de clones los niveles de polimorfismo en los marcadores son importantes. Parra y colaboradores determinaron como rametos del mismo geneto (clones), a las

muestras que compartieron todos los alelos en 4 loci utilizados en *S. pruinosus* (Parra et al., 2010), sin embargo, los resultados de Arnaud-Haond y colaboradores (2005) demuestran que tanto un número pequeño de loci como niveles bajos de polimorfismo, resultan en asignaciones erróneas de diversidad clonal, sobreestimando el número de rametos del mismo geneto y subestimando el número de individuos genéticamente distintos.

### 3. JUSTIFICACIÓN

La historia de la biota del Desierto Sonorense ha sido mayormente abordada por hipótesis imperantes sobre vicarianzas geológicas, que implican millones de años, explicando así la distribución espacial de la diversidad genética de los organismos distribuidos en esa región geográfica. Sin embargo, para los vegetales se han presentado recientemente explicaciones que dan más importancia a eventos climáticos ambientales del último millón de años (Pleistoceno), y particularmente de los últimos 10,000 años (Holoceno), aspectos que poco habían sido tomados en cuenta con anterioridad, pero no hay un consenso al respecto. El hecho de producir información para explicar la genética poblacional de más especies con características particulares, como la pitaya con su sistema dual de reproducción (sexual-asexual) y su vasta presencia a lo largo de heterogeneidad ambiental de la Península de Baja California, aspira a ayudar a discernir entre unas hipótesis y otras o a precisar las hipótesis imperantes.

Siendo que las cactáceas se han diversificado en tierras mexicanas, podemos contarlas como recursos del país, y entre mejor conozcamos los recursos con que contamos, mejor se puede planear su uso, lograr su aprovechamiento y prever su preservación. Es pertinente puntualizar en este sentido, que las investigaciones de genética poblacional para otras plantas de la Península, han sido llevadas a cabo por equipos extranjeros.

Entre las plantas con capacidad de reproducción vegetativa, esta es la primera cactácea para la que se desarrollan marcadores microsatélite especie-específicos con una técnica de secuenciación masiva (de segunda generación), lo que permitirá, posterior a este estudio, continuar con investigaciones genéticas y genómicas con otros alcances, a partir de información ya obtenida que ha sido objeto de este trabajo.

## 4. OBJETIVOS

### 4.1. General

Identificar si la estructura genética poblacional de *Stenocereus gummosus* es coherente con los procesos de contracción-expansión poblacional impulsados por efecto de los reiterados cambios climáticos (glaciar-interglaciar) del Pleistoceno, en donde la porción sur peninsular pudo actuar como una zona refugio.

### 4.2. Particulares

- Obtener genotipos de marcadores microsátélites específicos que permitan analizar los procesos microevolutivos de *Stenocereus gummosus* a nivel población.
- Evaluar la distribución geográfica de la diversidad genética intra e interpoblacional de *S. gummosus* en la Península de Baja California.
- Interpretar las evidencias genéticas para discernir si son coherentes con las hipótesis de cambios climáticos glaciar-interglaciar y con un refugio en la zona sur de la Península de Baja California.

## 5. HIPÓTESIS

La estructura genética poblacional de *Stenocereus gummosus* en la Península de Baja California, es resultado de una colonización reciente en sentido sur-norte, a partir de la región de Los Cabos.

**Escenario I.** Una colonización por difusión simple a partir de un refugio con el flujo genético limitado por la distancia, implicaría un modelo de aislamiento por distancia con un coeficiente de correlación parcial  $r \approx 1$ , la riqueza alélica a lo largo de la distribución disminuiría en función de la ruta de colonización y la heterocigocidad sería determinada por el sistema de reproducción, en particular por el flujo local de polen.

**Escenario II.** Una colonización con alta frecuencia de eventos de dispersión a larga distancia (EDLD) con el flujo genético limitado por la distancia, implicaría un modelo de aislamiento por distancia con ajuste inversamente proporcional a la frecuencia de los EDLD, con la riqueza alélica no relacionada a la ruta de la colonización y la heterocigocidad sujeta a la frecuencia de los EDLD y a la selección que pudiera presentarse ante cuellos de botella por efecto fundador.

**Escenario III.** La ausencia de colonización, es decir, una presencia constante de *S. gummosus* durante los procesos de cambio climático glacial-interglacial a lo largo de su distribución actual, con el flujo genético limitado por la distancia, implicaría un modelo de aislamiento por distancia con  $r \approx 1$ , la riqueza alélica dependería de la talla efectiva poblacional y la heterocigocidad sería determinada mayormente por el sistema de reproducción. Si además se presentaran barreras ecológicas o geológicas al flujo genético, en lugar de un modelo de aislamiento por distancia, se encontraría estructuración en función a estas barreras con índices de fijación altamente significativos entre las poblaciones.



## 6. MATERIALES Y MÉTODOS

### 6.1. Área de estudio

El área de estudio del presente trabajo incluye localidades ubicadas por todo el rango de distribución geográfica de *S. gummosus* reportado por Turner y colaboradores (1995), las localidades fueron seleccionadas buscando obtener muestras de sitios distintos según la experiencia de campo del equipo técnico del Herbario del CIBNOR, S.C. Se visitaron 16 sitios para la toma de muestras cuyos detalles y ubicación se encuentran en la Tabla II y en la Figura 2.

Tabla II. Localización geográfica de los sitios de muestreo en coordenadas decimales (Longitud y Latitud), fechas de recolecta, número de muestras recolectadas (# de muestras) y clave del nombre de cada localidad. Los registros se encuentran en orden latitudinal (sur-norte) para las localidades de la PBC, con la Isla Tiburón y la localidad de Sonora al final.

Clave	Localidad	# de muestras	Longitud	Latitud	Fecha de recolecta
L05	San Lucas	40	-109.994675	22.948053	17/04/2011
L10	Caduaño. Santa Teresita	38	-109.720482	23.214793	17/04/2011
L13	Todos Santos. La Paz Km 37	34	-110.205400	23.559110	25/03/2011
L01	El Comitán	35	-110.431716	24.133240	14/07/2011
L03	El Cien	26	-111.007639	24.346583	15/06/2010
L03	El Cien 2	10	-111.106000	24.347760	24/04/2011
L04	Puente Querétaro 2	12	-111.677040	25.307140	24/04/2011
L04	Puente Querétaro	30	-111.609461	25.350833	15/06/2010
L26	El 40. Mesa Salina	37	-112.932360	26.657390	07/04/2011
L20	Camino a Santa Águeda	41	-112.258920	27.271769	05/04/2011
L24	Rancho Piedra Blanca	42	-113.159389	28.229661	06/04/2011
L22	Punta Prieta	33	-114.155100	29.063345	04/04/2011
L29	San Fernando Velicata	44	-115.189722	29.970389	27/04/2011
L30	San Pedro	36	-116.030778	31.286556	28/04/2011
L36	Isla Tiburón	39	-112.257236	28.771077	28/04/2011
L41	Bahía Sargento	45	-112.202189	29.212369	01/05/2011

### 6.2. Trabajo de campo

Las recolectas de muestras se realizaron entre los meses de marzo y julio del año 2011, con excepción de dos muestreos preliminares realizados en junio del 2010, para las localidades El cien y Puente Querétaro (Tabla II). Una etiqueta única de seis caracteres fue asignada a cada muestra, componiéndose primeramente de la clave de localidad seguida por una "P" de pitaya y un número secuencial de dos dígitos (por ejemplo "L01P15").

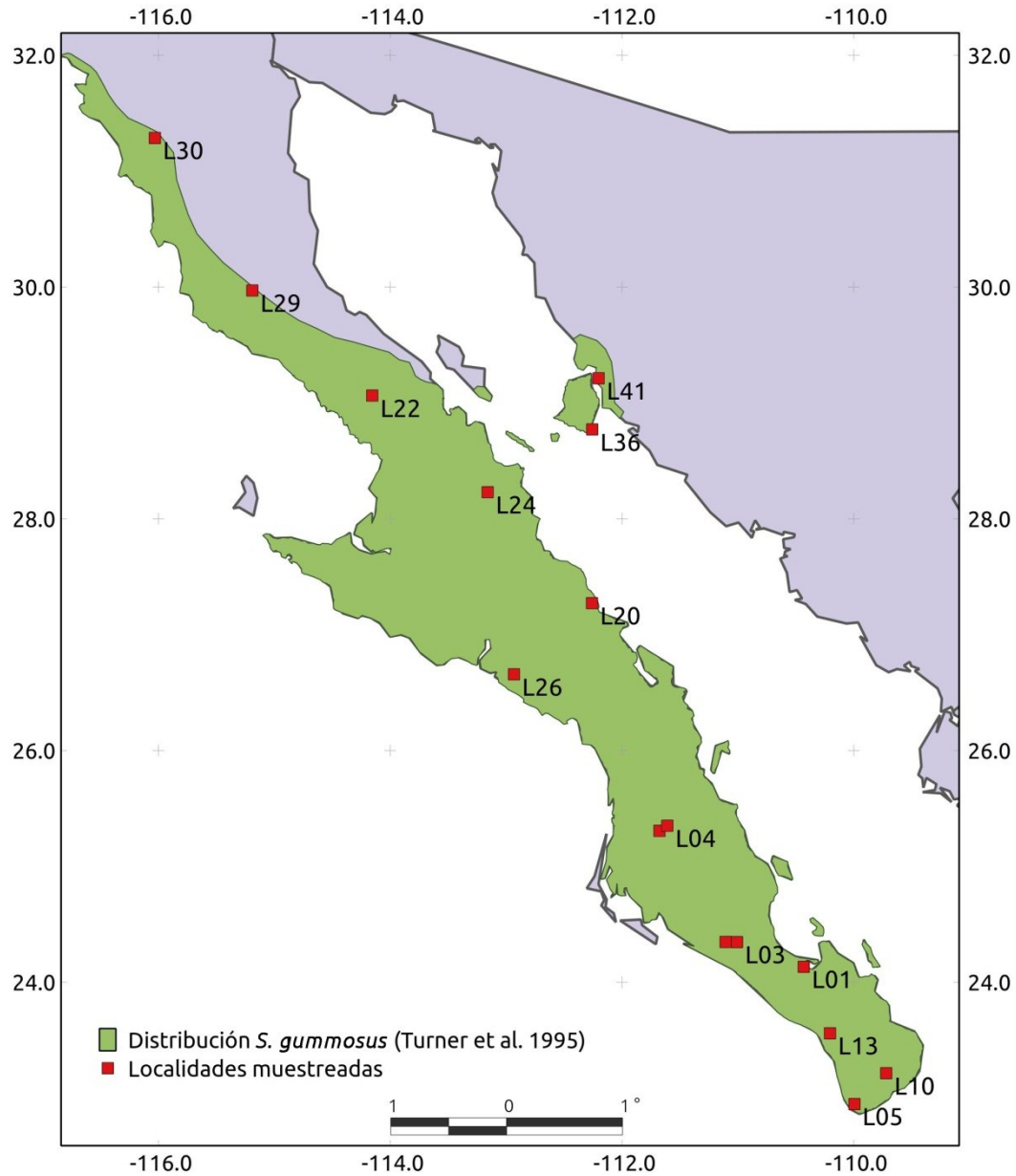


Figura 2. Mapa de localización de sitios de recolecta de muestras y rango de distribución geográfica de *Stenocereus gummosus* (Turner et al., 1995).

Los individuos fueron muestreados caminando hacia dos direcciones contrarias siguiendo un semicírculo imaginario. Para evitar la recolecta de dos muestras correspondientes a rametos de un mismo individuo genético, se eligió cada individuo una vez recorridos al menos 30 m (40-50 pasos) según la densidad lo permitiera, considerando como criterio los resultados de Molina Freaner y Clark Tapia (2005), quienes determinaron una distancia máxima entre rametos menor a 10 m.

La toma de muestra constó del corte de un segmento de costilla de la rama más turgente, utilizando una navaja y pinzas para incidir hasta la base troncal de la costilla y tomar el parénquima fresco adyacente a la corteza exterior, omitiendo el centro leñoso con gran cantidad de fibras. Cada muestra fue preservada en un tubo de 2 mL con etanol 96°.

### **6.3. Trabajo de laboratorio. Estandarizaciones**

#### **6.3.1. *Extracción de ácidos nucleicos***

Para el total de muestras recolectadas (Tabla II), se utilizó el método reportado por de la Cruz y colaboradores (1997) con algunas modificaciones (ver ANEXO I). La extracción de ADN y su calidad fue verificada para las primeras reacciones de extracción, a través de electroforesis en geles de agarosa al 1% (buffer TAE 1X, 10 mg/L de bromuro de etidio) por 50 minutos a 100 voltios. La totalidad de las muestras fueron cuantificadas en un NanoDrop 2000 (Thermo Fisher Scientific, Waltham MA). La solución de ADN concentrado se almacenó en tubos eppendorf de 2 mL a -20° C, con sus etiquetas correspondiente. Posteriormente, se prepararon soluciones de trabajo diluidas con agua milliQ® (Millipore, Billerica, MA., E.U.A.) a una concentración de 30 ng/μL.

#### **6.3.2. *Secuenciación y diseño de iniciadores de PCR para microsatélites***

Se utilizaron 50ng de ADN libre de ARN para una muestra de *Stenocereus gummosus* correspondiente a la L01 (Tabla II), para determinar secuencias de regiones aleatorias del genoma total por pirosecuenciación-454 con la química GS-FLX Titanium en el Centro de Secuenciación de la Universidad de California en Santa Cruz. La corrida fue realizada bajo las condiciones recomendadas por el prestador de servicios, en una placa de reacciones donde estuvieron presentes 8 especies (*S. gummosus* incluida) identificadas por etiquetas genéticas discriminantes entre sí por su secuencia.

Todas las lecturas 454 contenidas en un archivo formato “sff”, fueron convertidas de flujogramas a un solo archivo formato fasta (secuencias) y otro fasta-qual (calidad de secuenciación por sitio) donde las lecturas no contuvieran los adaptadores para la reacción de secuenciación, con el software sff\_extract ([http://bioinf.comav.upv.es/sff\\_extract/index.html](http://bioinf.comav.upv.es/sff_extract/index.html)).

Para identificar lecturas que contuvieran motivos microsatélites di-, tri-, tetra-, penta- y hexa-nucleótidos que contaran con al menos cinco motivos repetidos para dinucleótidos y cuatro para el resto, fue utilizado el software MSATCOMMANDER (Faircloth, 2008), cuyos resultados

fueron el primer filtro para generar un nuevo archivo fasta y su respectivo fasta-qual donde todas las lecturas presentes contuvieran loci potenciales para el análisis poblacional, es decir que contuvieran microsatélites en la secuencia.

Posteriormente se utilizaron los archivos de loci potenciales para diseñar iniciadores de PCR con el software PRIMER3 (Rozen y Skaletsky, 2000) implementado en el software QDD (Megléczy et al., 2010), con el que se descartaron posibles repeticiones entre las lecturas para evitar redundancia de datos y disminuir los errores de determinación de alelos; los criterios utilizados para dirigir el diseño de iniciadores fueron: 1) temperatura de alineamiento entre 50-70° C con un máximo de 2° C de diferencia entre cada par de iniciadores, 2) productos de PCR entre 90 y 320 pares de bases de longitud, 3) contenido de GC > 40%, 4) longitud de iniciadores entre 17 y 27 nucleótidos, y 5) los parámetros por defecto de PRIMER3 para auto-complementariedad de los iniciadores y complementariedad entre cada par de iniciadores.

Finalmente, se mandaron sintetizar a Sigma-Aldrich (St. Louis, Missouri, E.U.A.) 40 pares de iniciadores (Tabla III) que fueron elegidos favoreciendo aquellos diseñados para amplificar los motivos microsatélite con mayor número de repeticiones y que tuvieran la menor diferencia de temperatura de alineamiento entre los dos oligonucleótidos iniciadores, excluyendo los trinucleótidos y exanucleótidos, dado que se han asociados con elementos expresables y mayor ocurrencia entre sitios codificantes, por lo tanto sujetos a selección (Metzgar et al., 2000), condición indeseada para marcadores neutrales. Los nombres fueron asignados con un acrónimo del nombre específico, “Sgum”, seguido por un número consecutivo de dos dígitos, 01–40.

Tabla III. Loci obtenidos para pruebas de amplificación y polimorfismo. Los iniciadores fueron nombrados con el nombre del locus correspondiente más una F (*Forward*) para los de la lectura secuenciada o más una R (*Reverse*) para los de la cadena inversa complementaria. Entre corchetes se muestra la talla del motivo microsatélite o SSR (Simple Sequence Repeat, de sus siglas en inglés).

Nombre	Iniciadores (5' - 3')	Talla [SSR]	Motivo	PRIMER3 <i>T<sub>m</sub></i> (°C)
Sgum01	F: CCACTCTTCCGTCCAATCAT R: CACAAAACCCTCAAGGCAAT	147 [10]	(AG) <sub>5</sub>	59
Sgum02	F: TGAATCTACCCCGATCTCCA R: CATGTCATTTCTGATGCATTT	186 [80]	(TTTA) C (ATTT) CA (TTTA) <sub>2</sub> TC (TTTC) TT (TTTG) N <sub>9</sub> (TTTA) <sub>6</sub> (TTTG) <sub>3</sub> (TTTC)	60
Sgum03	F: TGCATTGATTCCATCCAAC R: TAGGTGCGATAACGTGTTGC	140 [24]	(AT) <sub>8</sub> G (AT) <sub>2</sub> G (AT) <sub>1</sub>	58
Sgum04	F: CCACTCTTCCGTCCAATCAT R: CAAGGCAATAGCAGCAGATG	136 [15]	(CT) <sub>5</sub> C (CT) <sub>2</sub>	60

Continuación Tabla III.

<b>Nombre</b>	<b>Iniciadores (5' - 3')</b>	<b>Talla [SSR]</b>	<b>Motivo</b>	<b>PRIMER3 T<sub>m</sub> (°C)</b>
Sgum05	F: TCGTAAAACGAATTGAGGGC R: GCTTTTAGCGAGGATTCCCT	122 [34]	(GGGA) <sub>1</sub> (TGGA) <sub>5</sub> (CGGA) <sub>1</sub> TG (GGGA) <sub>1</sub>	59
Sgum06	F: ATGAGATCGTGTGGGTCCAA R: ACAGAATCTCATGTGGCAGC	90 [24]	(AG) <sub>12</sub>	60
Sgum07	F: CTGACCTGCCTGACACCAC R: GCTATCATGGCCACCACC	90 [14]	(AG) <sub>7</sub>	60
Sgum08	F: GTCGAAACAAGCCTAGTCGC R: CCGACCTACACTTCCACTGC	116 [12]	(AG) <sub>6</sub>	63
Sgum09	F: TTTTCATATAATCGATAAGCCACG R: ATCCCGCAACACTATTGGTC	260 [10]	(AT) <sub>5</sub>	61
Sgum10	F: CCACAGGCAGAAGAGAGGAA R: GAGACAGAGTTAGGGTTGGCTC	91 [10]	(AT) <sub>5</sub>	65
Sgum11	F: TAAGAGACAGATGCGCGAAG R: AATCGGACCCTAATTCGACC	139 [10]	(AG) <sub>5</sub>	60
Sgum12	F: CAGGTTCAAGGGTGGAGTAA R: GGATGCGGCTATAGGAGTGT	110 [10]	(AC) <sub>5</sub>	62
Sgum13	F: AGCTCAGGTTGTGGACCAAG R: GCGTATTTGACCCTCACAA	105 [10]	(AG) <sub>5</sub>	61
Sgum14	F: ATGCATTAGTTTGCTTGGGC R: TAGAAGCAACACCAGGAGCA	126 [10]	(AG) <sub>5</sub>	59
Sgum15	F: CATGACTCAGGTGGTGGATG R: CAGCCTAGGAAGCCAATTCA	91 [10]	(AG) <sub>5</sub>	61
Sgum16	F: AACAGGGACACCAAAGTGG R: TGCACCCCAATTGTAGTTAGC	187 [10]	(AT) <sub>5</sub>	61
Sgum17	F: CAATGAGGGAGGGAGAATCA R: CAAAACCCTATTGGCCTCAA	153 [10]	(AG) <sub>5</sub>	59
Sgum18	F: TCTAGCAGGCCGAACCTAAA R: TCATTTTGAGGTCACTCTATTGATTC	150 [10]	(AT) <sub>5</sub>	65
Sgum19	F: AGTCTTTGTGCTGGGGAAGA R: GCCTCACCTGCTACTTACCG	317 [10]	(AG) <sub>5</sub>	62
Sgum20	F: GATCAATTAATGTTTCGTAACGCAT R: CAGCATCAGTGCCATACTACCT	124 [10]	(AG) <sub>5</sub>	65
Sgum21	F: AATCACGCCCGATGCTTCTT R: AGTCGAGCGATGCAACCAAC	116 [16]	(AT) <sub>1</sub> (GT) <sub>8</sub> (AT) <sub>7</sub> (GT) <sub>1</sub>	61
Sgum22	F: AGACCCGTGACAAGGTGCAA R: TGCTCCTGTGGTAACCTCCTC	106 [16]	(AACCA) <sub>5</sub> (AACT)	67
Sgum23	F: GAACAGAGGCTGACATGGTACA R: TTGGTGGACCTTGGATGCAG	91 [18]	(TG) <sub>12</sub>	64

Continuación Tabla III

Nombre	Iniciadores (5' - 3')	Talla [SSR]	Motivo	PRIMER3 Tm (°C)
Sgum24	F: CCCCAAAGCGGACAATACC R: TGACATACATGTTGGGCACTT	130 [60]	(AC) <sub>1</sub> (AT) <sub>5</sub> (AC) <sub>13</sub> [(AT) <sub>2</sub> GT] <sub>3</sub> (AT) <sub>2</sub>	60
Sgum25	F: ATTTTGTTTTATATCTGTGTTTCATTC R: AGACGCTTCAATTGTAACAAGG	101 [20]	(GT) <sub>10</sub>	63
Sgum26	F: ACGGTCGGGGCTGGTTG R: AAAGCTCACGCACGACCAAG	151 [24]	(CT) <sub>2</sub> N <sub>4</sub> (CT) <sub>1</sub> N <sub>4</sub> (CT) <sub>9</sub>	60
Sgum27	F: GAATAAATGTGGCCATATGAATGC R: AACGACGCGCTGCAACC	92 [38]	[CT (AT)] <sub>3</sub> (AT) <sub>5</sub> GC AC (AT) <sub>2</sub> [GT (AT)] <sub>2</sub>	61
Sgum28	F: AGGAGGTTCCCTTCCCTGCT R: TGGAAAACGAGAATGCTGGAA	124 [24]	(CT) <sub>12</sub>	62
Sgum29	F: CGTCTCATTTGCTGTGCTCAGAT R: CTGTGTGGGGACGAGAGCAG	96 [25]	(ACCAG) <sub>5</sub>	67
Sgum30	F: TCAGCTTAATCAACGCTCGTTTTTC R: AGCAAAGCGCTGCAATCAAC	114 [38]	(AT) AC (AT) <sub>9</sub> GC AC (AT) <sub>2</sub> [GT (AT)] <sub>2</sub>	64
Sgum31	F: TAACGCGCCGTCTTCAATTC R: TCCTTGCTTTAGCAGGATTCCC	142 [34]	(GGAT) <sub>5</sub> (GGAC) <sub>1</sub> (GGAT) <sub>1</sub> GG (GGAC) <sub>1</sub>	63
Sgum32	F: GGCAGTGCAAATTACACCCA R: TTGAAGAACTCAAATTTGTGCAAG	138 [20]	(CT) <sub>10</sub>	62
Sgum33	F: CCCTCTTCTCTCCAGTTTG R: ACCAGCGACCGTTTCTAGCC	99 [23]	(CT) <sub>9</sub> G (CT) <sub>2</sub>	61
Sgum34	F: TTTGGGAGCCATTATCGCAA R: GCGTGCGTCTGTCTTCCCTA	118 [25]	(AG) <sub>2</sub> G (AG) <sub>10</sub>	61
Sgum35	F: TCACTGCCTTTGGGCTGC R: TTCTCAGGGGAAATCACGCA	206 [115]	(CT) <sub>7</sub> N <sub>85</sub> (CT) <sub>8</sub>	59
Sgum36	F: AGCACATAACATATGGAACACGACA R: AAGTTAGTCCAATAACCCATCTCA	142 [18]	(AC) <sub>9</sub>	68
Sgum37	F: CTGGTGCTGCTTACCGGATG R: TCGGCTAGTCGGGTCCAGTT	148 [27]	(TG) <sub>2</sub> (AG) <sub>6</sub> C (AG) <sub>5</sub>	64
Sgum38	F: GCATCCAATCATTTCATTCTCCA R: CTTGAACCGTCGTGCAAACC	129 [24]	(AG) <sub>8</sub> AA (AG) <sub>1</sub> TG (AG) <sub>1</sub>	63
Sgum39	F: CCTGCTGGCCACTTCATGTC R: ATCGCTCAAATCGACCACCA	111 [18]	(GA) <sub>9</sub>	62
Sgum40	F: TGATCAATTTGACAAGATCCAACCTCA R: TGGTTCAATTCATCCCTTTTCAA	152 [43]	(AACG) <sub>1</sub> (AACA) <sub>1</sub> N <sub>7</sub> (AACA) <sub>1</sub> (AACG) <sub>6</sub>	65
Ppri13	F: TTCAAATCATTCAATAAAACAAGAGC R: TCTTGATCCTCGACCTCCAT	101 [13]	(TC) <sub>13</sub>	60

### 6.3.3. *Condiciones de amplificación por PCR*

Para determinar las condiciones óptimas de amplificación de los productos blanco de los 40 pares de iniciadores adquiridos, se realizaron pruebas de amplificación por gradientes de 4 temperaturas de alineamiento ( $\pm 2^{\circ}\text{C}$  y  $\pm 5^{\circ}\text{C}$  de la temperatura de alineamiento determinada por PRIMER3, Tabla III) y de 3 concentración de cloruro de magnesio (1.5 mM, 2.0 mM y 2.5mM [ $\text{MgCl}_2$ ]), variando la concentración de polimerasa añadida a las reacciones a largo de las pruebas a partir de 0.2 unidades por reacción.

Las reacciones de amplificación por PCR se llevaron a cabo en un volumen total de 10  $\mu\text{L}$  en un termociclador DNA Engine marca BioRad de 96 plazas, conteniendo 50 ng aproximadamente de ADN de dos muestras distintas (L24P04 y L26P06), 1.0x de Buffer para PCR Invitrogen [20 mM Tris-HCl pH 8.4, 50 mM KCl], 0.2 mM de cada dNTP, 0.4  $\mu\text{M}$  de cada iniciador, 0.1 M de BSA (New England Biolabs) y una concentración locus-dependiente de TaqPolimerasa (Invitrogen) y de  $\text{MgCl}_2$  (Tabla VI). Se añadió adicionalmente glicerol al 5% para el locus Sgum24 para aumentar el rendimiento de amplificación. El programa de amplificación utilizado en el termociclador constó de los siguientes pasos:  $95^{\circ}\text{C}$  por 6 min como desnaturalización inicial; 35 ciclos de  $94^{\circ}\text{C}$  por 60 s, 35 s en gradiente de temperatura ( $T_m \pm 2^{\circ}\text{C}$  y  $\pm 5^{\circ}\text{C}$ ) y 15 s de extensión a  $72^{\circ}\text{C}$ ; finalmente 15 minutos a  $72^{\circ}\text{C}$  como extensión final posterior a los 35 ciclos (exceptuando Sgum05 y Sgum36 de los 15 s de extensión dentro de los ciclos).

### 6.3.4. *Electroforesis en gel de acrilamida con tinción de plata*

Los productos amplificados fueron sometidos a electroforesis desnaturalizante en geles de acrilamida al 6% y Urea 7.5 M, a 1700 V, 40 mA y 45 W por 100-120 min para ser visualizados por tinción de plata (Bassam y Caetano-Anollés, 1993). Las tallas alélicas fueron determinadas en comparación con el estándar de talla molecular 10 bp ladder (Invitrogen, Carlsbad, California, E.U.A.) con el programa Gel Analyzer (<http://www.sequentix.de>), una vez que los geles fueron digitalizados. Según la facilidad de determinación de las bandas correspondientes al producto blanco de cada reacción, se eligieron las condiciones óptimas para continuar subsecuentemente con las pruebas de polimorfismo.

### 6.3.5. *Pruebas de polimorfismo por PCR*

Para identificar el polimorfismo alélico de los loci microsatélite que se logró determinar las condiciones óptimas de amplificación (Tabla VI), se seleccionaron 10 muestras para amplificar por PCR, pertenecientes a localidades dispersas a lo largo del área de distribución total: L01, L04, L05, L13, L24, L26, L29, L36 y L41 (ver Figura 2). Los productos de PCR fueron analizados por el mismo método electroforético especificado anteriormente (ver 6.3.4).

### 6.3.6. *PCR con iniciadores marcados por fluoróforos*

Una vez determinados los loci polimórficos y la temperatura óptima de alineamiento de sus iniciadores de PCR ( $T_m$ ), se ajustó el programa de amplificación para los iniciadores  $F$  (Tabla IV) marcados con fluoróforos de distinto color (matriz DS-33 de ABI PRISM®, Life Technologies, Carlsbad CA), quedando una desnaturalización inicial a 95° C por 6 minutos, 35 ciclos compuestos por 60 s a 94° C, 40 s a la  $T_m$  específica (Tabla VI) y 20-40 s a 72° C, con un tiempo de extensión final a 72° C por 25 minutos para mejorar la calidad de los picos y disminuir los errores de lectura (Guichoux et al., 2011). Las condiciones de amplificación fueron las mismas que durante la estandarización de las  $T_m$ , a excepción de la concentración de cada iniciador que en este caso fue de 0.3  $\mu$ M en cada reacción.

Se añadió adicionalmente glicerol al 5% para el locus Sgum24 para aumentar el rendimiento de amplificación. También se analizó un par de iniciadores de *Pachycereus pringlei* por amplificación cruzada, (locus Ppri13 , Tabla III, Gutiérrez-Flores et al., en preparación).

### 6.3.7. *Electroforesis en capilar automatizada*

La lectura de los tamaños alélicos fue realizada por electroforesis en capilar automatizada, en equipos ABI 310–3730 PRISM® (Life Technologies, Carlsbad, California, E.U.A.), determinando los tamaños en comparación con el estándar de talla LIZ600 de Applied Biosystems (Carlsbad, California, E.U.A.).

Los productos fueron mezclados para las corridas electroforéticas, considerando los rangos alélicos y los colores de detección de los fluoróforos. Para determinar la combinación de loci se tomó lectura de 12 muestras de distintas localidades (L01, L04, L05, L10, L13, L20, L22, L26, L29, L30, L36 y L41). Finalmente se lograron combinar MD1 y MD2 (Tabla IV).



Las mezclas de productos de PCR para electroforesis se compusieron de 8  $\mu\text{L}$  de HiDi-formamida de alta pureza (Applied Biosystems), 0.55  $\mu\text{L}$  de LIZ600 (Applied Biosystems), 0.5  $\mu\text{L}$  de agua MilliQ (Millipore) y 1.0  $\mu\text{L}$  de mezcla de los productos de PCR correspondientes a cada juego (Tabla IV) en mismas proporciones, siempre buscando un volumen final entre 0.25–0.5  $\mu\text{L}$ , según el número de marcadores analizados en la misma corrida.

Tabla IV. Lista de combinaciones de iniciadores *F* de PCR con fluoróforo acoplado. MD1, MD2 y MDr corresponde al juego de análisis en electroforesis. Se indica el nombre con la clave (No. Acceso) con que cada locus fue registrado en Gene Bank. Colores: amarillo (NED), azul (6FAM), verde (VIC) y rojo (PET). \*No. Acceso en trámite.

MD1			MD2			MDr	
Gene Bank No. Acceso	Locus	Color	Gene Bank No. Acceso	Locus	Color	Locus (colores como en MD1 y MD2)	
JX512846	Sgum06	NED	JX512845	Sgum05	NED	Sgum05	Sgum29
JX512847	Sgum12	VIC	JX567729	Sgum24	PET	Sgum06	Sgum31
JX512851	Sgum36	NED	JX512848	Sgum25	FAM	Sgum12	Sgum36
JX512852	Sgum39	FAM	JX512849	Sgum29	VIC	Sgum25	Sgum39
*	Ppri13	PET	JX512850	Sgum31	FAM		

#### 6.4. Trabajo de laboratorio. Genotipificación

##### 6.4.1. Amplificación y lectura electroforética

Cada reacción de cada locus para cada muestra fue realizada por separado. Se llevaron a cabo dos ensayos de amplificación para la genotipificación. El primero se realizó para la totalidad de las muestras especificadas en la Tabla II y con el procedimiento de combinación de loci MD1 y MD2 (Tabla IV). En el segundo ensayo sólo se emplearon los loci con mayor éxito de amplificación (Sgum24 y Ppri13 fueron excluidos), con el fin de completar datos faltantes del primer ensayo y validar la lectura alélica con duplicados; este segundo ensayo fue analizado con los productos de todos los loci combinados en un solo juego de lectura MDr (Tabla IV).

El programa de amplificación para el segundo ensayo fue utilizando 40 s de extensión en todos los ciclos y aumentando las concentraciones de los iniciadores a 0.4  $\mu\text{M}$ , de dNTPs a 0.25 mM y de *Taq*Polimerasa a 0.4 U/Rx.

Las placas de mezcla de productos MD1 y MD2 fueron enviadas al Carver Center, University of Illinois (Urbana, Illinois, E.U.A.) para análisis de fragmentos en secuenciador ABI PRISM 3730x bajo las condiciones recomendadas por el proveedor. Las placas correspondientes al

juego de marcadores MDr fueron analizadas en el Laboratorio de Genética para la Conservación del CIBNOR, S.C. con un equipo ABI PRISM 310.

#### 6.4.2. *Determinación de rangos alélicos y reglas de lectura automatizada*

Para reducir errores de registro que deriven en errores de interpretación (Guichoux et al., 2011), la determinación de los tamaños alélicos fue realizada empleando el software GeneMapper versión 4.1 (Applied Biosystems) (ANEXO II). Para esto se deben establecer reglas que permitan homogenizar la lectura y registro de datos:

La señal mínima para lectura de picos homocigotos en este estudio, fue de 140 unidades de fluorescencia relativa (RFU, por sus siglas en inglés) y de 70 RFU para picos heterocigotos. Las lecturas menores fueron omitidas de la base de datos de genotipos.

Se establecieron los rangos de tamaño alélico (bins) para cada marcador (barras en gris, ANEXO II), que son los límites para asignación de alelos con los que el software compara los perfiles de lectura de cada muestra analizada, asignando así la talla alélica correspondiente a cada marcador. Los alelos que exhibieron una talla intermedia entre dos bins, fueron corregidos manualmente revisando la presencia de artefactos de amplificación o picos dobles por falta de adición de una base (stutters, por tartamudeo en inglés) (Guichoux et al., 2011).

Debido a la longitud de los fragmentos, en los genotipos heterocigotos se espera que el pico correspondiente al alelo más pequeño o corto sea más grande en concentración y por lo tanto en RFU, en comparación con el más grande o largo. Por lo tanto, se evaluó la proporción de la altura del primer pico en RFU con respecto a la del segundo ( $RFU_1/RFU_2$ ) para todos los genotipos heterocigotos; cuando la razón fue menor a 0.7, no se consideró al primer pico como un alelo.

### 6.5. Análisis de datos

#### 6.5.1. *Calidad y utilidad de los marcadores*

*a. Detección de alelos nulos.* Un alelo nulo es un fragmento de ADN que no logró amplificarse, ya sea por problemas en la PCR o por mutaciones sitio-específicas en la región de alineamiento para los iniciadores (Shaw et al., 1999), lo que provoca una sobreestimación de los genotipos homocigoto por la ausencia de amplificación; si esta condición es muy persistente la heterocigocidad será subestimada (Carlsson, 2008). Se evaluó la presencia de alelos nulos mediante el algoritmo de Maximización de la expectativa (Dempster et al., 1977) a través del software

FreeNA (Chapuis y Estoup, 2007) calculando la frecuencia de ocurrencia de alelos nulos por locus y por localidad.

*b. Pruebas de desequilibrio de ligamiento.* Para garantizar la distribución independiente entre los alelos de distintos loci, se estimó la probabilidad de cada par de loci se encuentren ligados, por medio de pruebas exactas de Fisher usando Cadenas de Markov de 10,000 permutaciones efectuadas con el software GENEPOP 4 (Rousset, 2008).

### 6.5.2. Estimación de parámetros poblaciones

*a. Diversidad genética por localidad.* Para evaluar la diversidad genética dentro de las localidades muestreadas se estimó para cada locus la heterocigocidad esperada insesgada ( $H_E$ ) y la heterocigocidad observada ( $H_O$ ) por medio del software Arlequin 3.5.1.2 (Excoffier y Lischer, 2010), además del número de alelos por locus ( $A$ ), número de alelos efectivos ( $A_e$ ), el índice de información ( $I$ ) como en Brown y Weir (1983) y el índice de fijación o coeficiente de endogamia ( $F$ ) calculados con el paquete GenAlEx 6.5 (Peakall y Smouse, 2012). Se analizó la diversidad genética intralocalidad, esperando encontrar variación sobre el eje de direccionalidad de una potencial expansión demográfica (Hewitt, 1996). Para comparar los niveles de diversidad genética entre localidades se calculó la riqueza alélica ( $R$ ), utilizando el método de rarefacción para corregir el efecto por tamaños de muestra desiguales, por medio del software HP-Rare (Kalinowski, 2005). La significancia estadística del cambio clinal en la diversidad intralocalidad, medida por la riqueza alélica, fue determinada por una regresión lineal con la latitud de cada localidad efectuada con el software R (R Development Core Team, 2008) probando el nivel de significancia para 3 muestras balanceadas ( $S$ ) de distinto tamaño ( $n$ ), además de probar utilizando por separado la heterocigocidad esperada insesgada ( $H_E$ ).

*b. Equilibrio de Hardy-Weinberg.* El equilibrio de Hardy-Weinberg (E-HW) establece que existe un equilibrio entre las frecuencias alélicas y las genotípicas en las poblaciones panmícticas cuando no existen fuerzas evolutivas tales como la deriva genética, selección, mutación y la migración (Hartl y Clark, 2007). Para evaluar si las localidades de *S. gummosus* presentan desviación del equilibrio HW, se empleó la prueba exacta de Fisher a partir del coeficiente de endogamia ( $F_{IS}$ ) con el software Arlequin 3.5 (Excoffier y Lischer, 2010) con el algoritmo de Cadenas de Markov por 1,000,000 cadenas y 100,000 pasos de dememorización. Los valores  $p$  de significancia estadística fueron evaluados con un  $\alpha$  de 0.05 con corrección de Bonferroni (Rice,

1989) donde valores de  $p \leq 0.000658$  correspondieron a rechazo a la hipótesis nula de panmixia; subsecuentemente, se empleó el método de tasa de falsos positivos (FDR, por sus siglas en inglés) aplicando una corrección a los valores de  $p$  con el método propuesto por Benjamini y Hotchberg (1995) manteniendo  $\alpha$  en 0.05, con la función `p.adjust()` del paquete estadístico R (R Development Core Team, 2008).

### 6.5.3. *Análisis de estructuración poblacional*

**a. Análisis Factorial de Correspondencia (AFC) y Análisis Discriminante de Componentes Principales (ADCP).** Empleando el software GENETIX (Belkhir et al., 2004), se realizó un análisis canónico multivariado AFC, el cual permite la visualización de cada uno de los individuos genotificados en un espacio definido por los valores de una tabla de contingencia construida en factores a partir de las frecuencias alélicas de cada grupo (localidades), de tal forma que, al combinar los factores que mejor explican la varianza entre grupos, se puede representar a la totalidad de muestras utilizadas en una nube de puntos dispuestos según su conjunto de alelos. Se analizó la distribución de los puntos por localidad y posteriormente por región con el ADCP (Jombart et al., 2010), detectando centroides y su varianza espacial para cada región identificada (Sur de la Península, Centro-Norte de la Península y Sonora-Isla Tiburón), a partir de un análisis discriminante de componentes principales con el paquete `adeget` de R (Jombart, 2008).

**b. Análisis de Varianza Molecular (AMOVA).** El AMOVA consta en determinar la cantidad de variabilidad genética contenida dentro de subgrupos, dentro grupos y entre ellos, por medio de un análisis jerarquizado con una matriz de distancias (Excoffier et al., 1992) a partir de los estadísticos  $F$  de Wright (Weir y Cockerham, 1984) y fueron llevados a cabo en el programa ARLEQUIN 3.5 (Excoffier y Lischer, 2010). Los modelos de organización espacial probados, fueron considerando regiones geográficas a lo largo de la distribución tomando en cuenta los eventos vicariantes invocados para explicar la diversidad genética de los vertebrados en el noroeste de México (Lindell et al., 2006; Riddle y Hafner, 2006), se consideró uno y dos canales transpeninsulares y el Golfo de California (Figura 3). Primero se incluyó sólo el conjunto de localidades de la Península de Baja California, incorporando posteriormente la totalidad de localidades, para conformar 5 modelos distintos considerando algún evento vicariante, 2 modelos sin considerar ningún nivel de agrupación más que las propias localidades de muestreo y finalmente, además de estos modelos *a priori*, se evaluó una agrupación *post-hoc* inferida a partir

de los resultados de análisis de distancias genéticas (ver Figura 11), resultando en 8 modelos distintos (Tabla V).

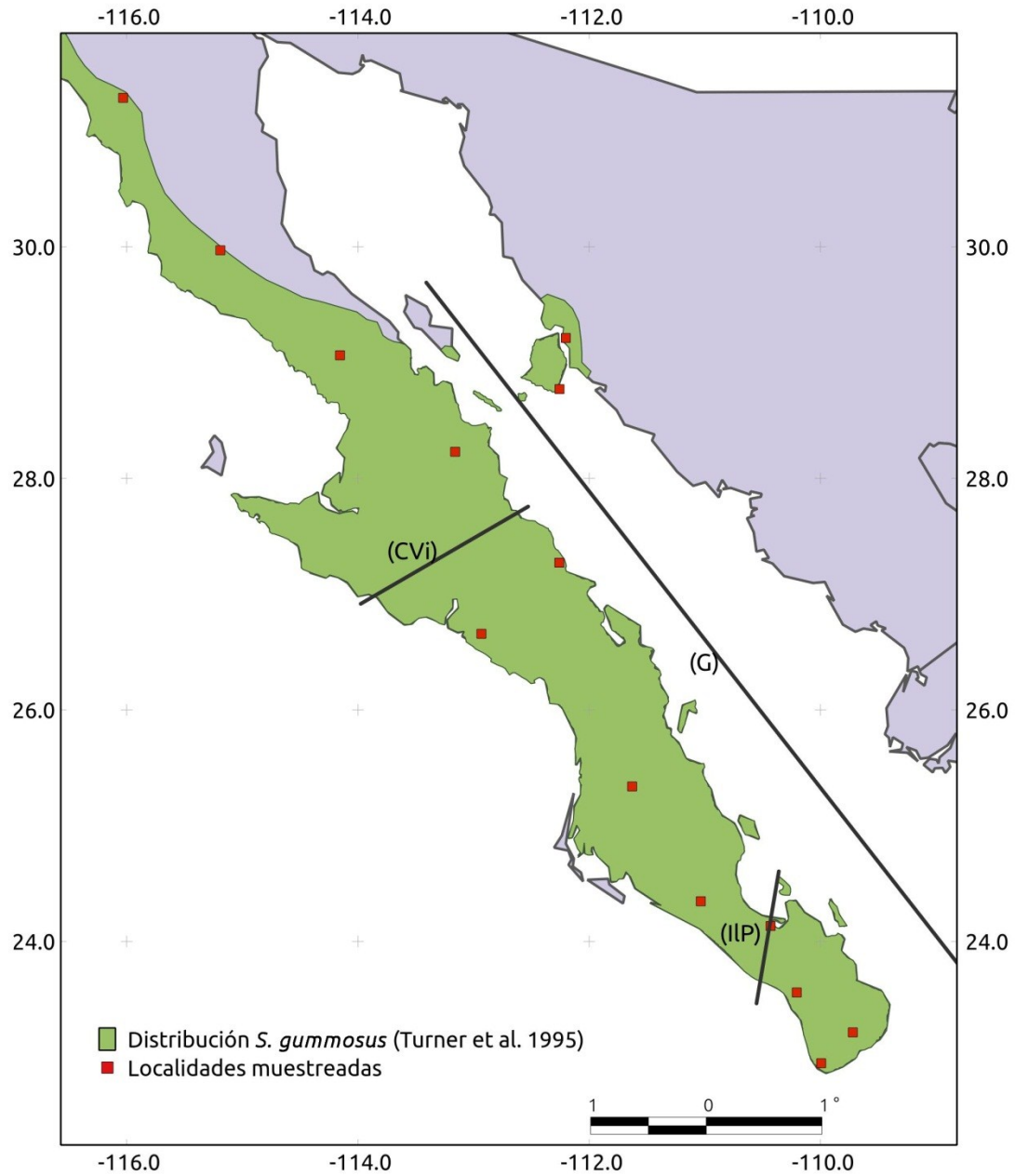


Figura 3. Eventos vicariantes geológicos teorizados a partir de información biológica (ver texto), tomados en cuenta para la conformación de grupos para el AMOVA.

Tabla V. Modelos de agrupación espacial de la diversidad genética entre grupos, tomando en cuenta las hipótesis de vicarianza (ver texto) propuestas para explicar la distribución genética de especies codistribuidas con *S. gummosus*, comparados por medio del AMOVA. Se especifica el tamaño de muestra ( $n$ ) tomado en cuenta.

Modelo	Localidades combinadas	Nivel	No. grupos	$n$
Localidades	Cada localidad independiente	Península	14	228
CVi	Canal Vizcaíno	Península	2	228
CTP	2 Canales transpeninsulares	Península	3	228
IIP	Istmo de La Paz	Península	2	228
Localidades	Cada localidad independiente	Total	16	277
ViG	Canal Vizcaíno + Golfo de California	Total	3	277
CsG	2 Canales + Golfo de California	Total	4	277
NJ	Grupos a partir de árbol $D_C$ NJ (Figura 11)	Total	3	277

*c. Aislamiento por distancia.* Tomando en cuenta que la dispersión movilizadora de la diversidad genética de una planta se reduce básicamente a la fase gamética (polen) y la de semilla, ante una expansión poblacional, se espera que la probabilidad de reproducción sea mayor entre individuos más cercanos, tendiendo a generar un patrón de aislamiento por distancia que, a lo largo de una expansión poblacional, resultaría en un aumento de la varianza genética interpoblacional proporcional a la distancia (Rousset, 1997). Buscando esta relación de menor distancia genética a menor distancia geográfica, se realizaron pruebas de Mantel (1967), que constan en evaluar por permutaciones si la correlación entre dos matrices pudiera ser debida al azar. Las pruebas se llevaron a cabo con el servicio en línea Isolation By Distance de la Universidad de San Diego (Jensen et al., 2005), por 5000 permutaciones. La distancia geográfica por pares de localidades fue obtenida con la aplicación Geographic Distance Matrix Generator (Ersts, 2006) y la matriz correspondiente de distancias genéticas fue computada con el software Genetix 4.05 (Belkhir et al., 2004), empleando la distancia de cuerda  $D_C$  (Cavalli-Sforza y Edwards, 1967), pues entre las utilizadas para microsatélites, esta es la que resulta menos afectada por eventos demográficos acompañados de disminución abrupta de diversidad genética (Takezaki y Nei, 1996). Las estimaciones de distancia fueron permutadas con sus valores originales como también transformados a logaritmos.

*d. Dendograma de distancia genética.* Se examinaron las relaciones de similitud genética entre las localidades utilizando la distancia genéticas  $D_C$  (Cavalli-Sforza y Edwards, 1967) entre cada par de localidades para reconstruir un dendograma no enraizado; esta distancia no hace ningún supuesto respecto a una talla de población constante o tasas de mutación de los loci y las

magnitudes resultantes no son proporcionales al tiempo evolutivo, pero es generalmente usada para obtener la topología correcta de árboles entre poblaciones estrechamente relacionadas con datos de marcadores microsatélites (Takezaki y Nei, 1996). El dendograma consenso fue reconstruido por el método del vecino más cercano Neighbor-Joining (NJ) (Saitou y Nei, 1987) y la robustez de las ramas fue probada por un bootstrap de 50,000 permutaciones, empleando el software PHYLIP (Felsenstein, 2005). Finalmente, el árbol consenso fue proyectado y editado con el software Dendroscope 1.4 (Huson et al., 2007).

*e.* **Índices de Fijación de Wright.** De entre los tres niveles jerárquicos de los índices  $F$  de Wright, se utilizó el que toma en cuenta las subdivisiones poblacionales,  $F_{ST}$  según Weir y Cockerham (1984) para cada par de localidades, más un valor de significancia  $p$  calculado por 1000 permutaciones de las frecuencias originales ante la hipótesis nula de no diferencia entre localidades, ambos calculados por el software Genetix 4.05 (Belkhir et al., 2004). Para ajustar el valor crítico de significancia, dado que cada comparación pareada de  $F_{ST}$  implicó una prueba de hipótesis independiente, se ajustaron los valores de  $p$  por el método de Benjamini y Hotchberg (1995) con la función `p.adjust()` del software R (R Development Core Team, 2008).

*f.* **Forma teórica de la estructura genética.** A partir de la teoría de gráficos, es posible conjuntar información de marcadores genéticos para generar un gráfico cuya topología es resultante de las interacciones históricas y contemporáneas entre las entidades en cuestión (Dyer y Nason, 2004), llamado *Population Graph*. Estos gráficos poblacionales se componen de nodos y vértices entrelazados, donde cada nodo representa a una entidad de análisis (localidades, en este caso) con un tamaño proporcional a la varianza genética intrapoblacional, cada vértice indica conectividad o cercanía genética y su longitud es proporcional a la varianza genética interpoblacional entre los nodos conectados; así el conjunto de interconexiones permite analizar cómo se distribuye la variación genética entre las entidades analizadas, proyectando el gráfico en un dibujo de espacio tridimensional. A partir de información genotípica, se calcula un estadístico  $D$ , equivalente al  $\Phi$  utilizado en AMOVA (Excoffier et al., 1992), que describe una distancia estadística entre las poblaciones en un espacio genético de  $m$  dimensiones como el número de variables genéticas independientes (alelos). Los valores  $D$  permiten generar un gráfico saturado de conexiones que posteriormente es transformado en un gráfico con topología informativa obtenido de la matriz de la incidencia mínima, que es el número mínimo de conexiones con distancia  $D$ , que describan suficientemente la covarianza genética entre los nodos, por análisis de independencia condicional (Magwene, 2001). El análisis de la información y la construcción del gráfico fueron llevados a cabo

con el paquete GeneticStudio (Dyer, 2009) implementado en el software R (R Development Core Team, 2008).

**g. Análisis de clusters.** Para determinar el número más probable de poblaciones, se utilizó el software STRUCTURE 2.2 (Pritchard et al., 2000), con el cual se estima un valor de probabilidad de pertenencia para cada individuo a cada grupo genéticamente homogéneo de entre  $K$  grupos, esto a través de simulación por el método de Montecarlo y Cadenas de Markov (MCMC, por sus siglas en inglés) calculando la probabilidad de cada valor de  $K$  en exploración, dada la información de los genotipos multilocus, en este caso, 6 loci microsatélite. Se empleó el modelo de mezcla de poblaciones con 10,000 réplicas o cadenas de pre-simulación, para posteriormente calcular valores de probabilidad hasta las 100,000 réplicas, efectuando 10 repeticiones para cada valor  $K$ , que fue de 1 hasta 15. Finalmente se evaluó la probabilidad acumulada y la probabilidad posterior para cada valor de  $K$ , introduciendo los resultados de la simulación al servicio en línea Structure Harvester (Earl y vonHoldt, 2011), el cual permite detectar el número de grupos reales para datos simulados bajo modelos de intercambio genético, empleando el estadístico  $\Delta K$  de Evanno (Evanno et al., 2005). Se tomaron los gráficos generados por el mismo software STRUCTURE.



## 7. RESULTADOS

### 7.1. Estandarización

#### 7.1.1. *Extracción de ADN*

En función de la cantidad de ADN total cuantificado en cada muestra, se eligieron las 542 con mayor cantidad disponible para proceder con la amplificación de los loci polimórficos.

#### 7.1.2. *Secuenciación y diseño de iniciadores*

La pirosecuenciación arrojó un total de 42,554 lecturas con un promedio de longitud de 311.11 nucleótidos y desviación estándar de 128.27 (29 y 839 nucleótidos como longitudes mínima y máxima, Figura 4), entre las cuales en 1,539 (3.62%) se detectaron repeticiones microsatélite. Una vez filtradas las lecturas con loci potencialmente útiles, se logró diseñar 179 pares de iniciadores de PCR que cumplieran los criterios establecidos anteriormente (ver 6.3.2), correspondiente a un 11.63% de los loci microsatélite encontrados originalmente.

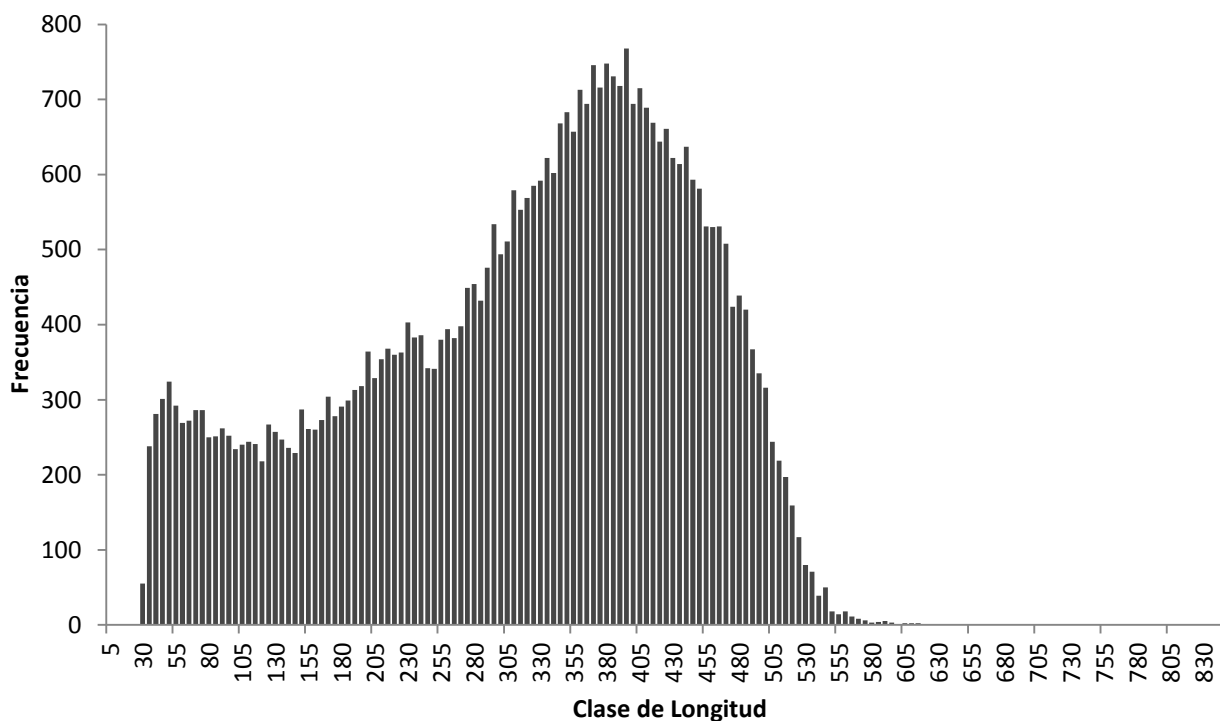


Figura 4. Distribución de frecuencias de lecturas de pirosecuenciación Titanium GS-FLX, por intervalos de longitud de 5 nucleótidos.

### 7.1.3. Pruebas de amplificación y polimorfismo

De los 40 loci trabajados, 6 tuvieron un éxito de amplificación menor al 50%, 2 mostraron polimorfismo pero con dificultad de amplificación (alrededor de 50%) e inconsistencia en los tamaños alélicos. De los 32 restantes (Tabla VI), 23 presentaron sólo una talla alélica con las muestras amplificadas y 9 exhibieron al menos 2 alelos distintos.

Tabla VI. Condiciones de amplificación para loci microsatélite con éxito mayor a 50%. Se enlistan por dos criterios de orden: número de alelos encontrados (*A*) y por nombre. Se detallan las condiciones determinadas como óptimas para amplificación: concentración de cloruro de magnesio (MgCl<sub>2</sub>), temperatura de alineamiento de los iniciadores (*T<sub>m</sub>*) y unidades de enzima Taq-Polimerasa (TaqPol). \* Loci polimórficos.

Nombre	MgCl <sub>2</sub> [mM]	<i>T<sub>m</sub></i> (°C)	TaqPol (U/Rx)	Tipo de Motivo	<i>A</i>
Sgum01	2.5	65	0.2	Dinucleótido	1
Sgum04	2.0	61	0.3	Tetranucleótido	1
Sgum07	1.5	65	0.2	Dinucleótido	1
Sgum08	2.0	68	0.2	Dinucleótido	1
Sgum09	2.0	56	0.2	Dinucleótido	1
Sgum10	2.5	64	0.2	Dinucleótido	1
Sgum11	2.0	63	0.2	Dinucleótido	1
Sgum15	2.0	64	0.2	Dinucleótido	1
Sgum17	2.0	65	0.2	Dinucleótido	1
Sgum19	2.0	65	0.2	Dinucleótido	1
Sgum21	1.5	58	0.2	Dinucleótido	1
Sgum22	2.0	68	0.2	Tetranucleótido	1
Sgum26	2.5	60	0.4	Dinucleótido	1
Sgum27	2.0	57	0.4	Dinucleótido	1
Sgum28	2.5	58	0.2	Dinucleótido	1
Sgum30	1.5	68	0.2	Dinucleótido	1
Sgum32	2.5	64	0.2	Dinucleótido	1
Sgum33	2.0	58	0.4	Dinucleótido	1
Sgum34	2.0	58	0.4	Dinucleótido	1
Sgum35	2.0	61	0.4	Dinucleótido	1
Sgum37	1.5	62	0.2	Dinucleótido	1
Sgum38	2.0	58	0.4	Dinucleótido	1
Sgum40	2.0	63	0.2	Tetranucleótido	1
Sgum05*	1.5	58	0.2	Tetranucleótido	3
Sgum06*	2.0	61	0.3	Dinucleótido	4
Sgum12*	2.0	58	0.2	Dinucleótido	2
Sgum24*	2.0	55	0.3	Dinucleótido	6
Sgum25*	2.5	55	0.3	Dinucleótido	4
Sgum29*	2.5	64	0.3	Pentanucleótido	4
Sgum31*	2.0	64	0.4	Tetranucleótido	2
Sgum36*	2.0	58	0.2	Dinucleótido	7
Sgum39*	2.5	55	0.4	Dinucleótido	3
Ppri13*	2.0	58	0.4	Dinucleótido	10

## 7.2. Genotipificación

De los 10 loci polimórficos seleccionados para amplificación, Sgum24 y Ppri13 tuvieron un éxito de amplificación menor al 15% de las muestras incorporadas al primer ensayo, por lo que fueron excluidos para el segundo ensayo (ver 6.4.1) y para el resto del análisis. Sgum31 tuvo un incremento en muestras genotipificadas con el segundo ensayo similar al del resto de loci (Tabla VII, Figura 5), pero no alcanzó el 50% de datos acumulados, por lo tanto, también fue excluido de los análisis posteriores.

Tabla VII. Éxito de amplificación en porcentaje del total de muestras genotipificadas por ensayo.

Ensayo	Sgum05	Sgum06	Sgum12	Sgum24	Sgum25	Sgum29	Sgum31	Sgum36	Sgum39	Ppri13
<b>1 (%)</b>	53.9	69.6	46.1	14.6	52.4	45.0	20.3	39.3	62.2	14.8
<b>2 (%)</b>	13.5	5.5	12.7	-	13.8	17.5	13.1	14.4	7.9	-
<b>% Total</b>	67.3	75.1	58.9	14.6	66.2	62.5	33.4	53.7	70.1	14.8

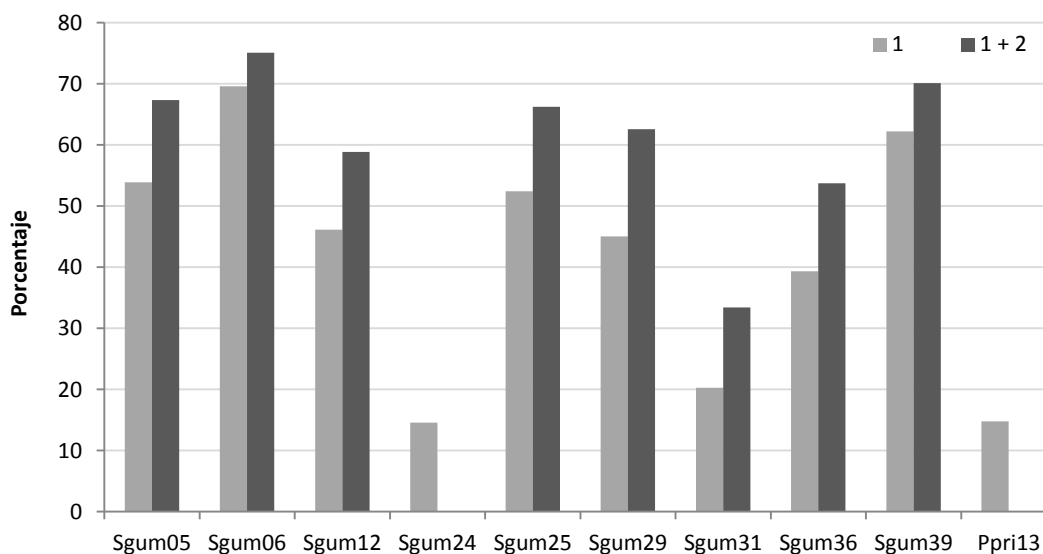


Figura 5. Éxito de amplificación acumulado entre los dos ensayos de amplificación (Tabla VII).

Para validar el trabajo de genotipificación se amplificaron por duplicado 341 genotipos (20% de los datos totales), entre los cuales el 7% tuvo alguna discordancia de tamaños alélicos. 30 casos difirieron en un alelo, poniendo en duda la condición de heterocigocidad (4.4% de los datos duplicados), 9 genotipos más tuvieron una asignación discordante en ambos alelos (2.6% del total de duplicados). Dado que fueron menos muestras manipuladas durante el segundo ensayo de

amplificación, se optó por conservar las lecturas alélicas correspondientes a este, disminuyendo así la probabilidad de errores por contaminación cruzada.

Finalmente se identificaron las muestras con un máximo de 1 genotipo faltante para los 7 loci con mayor éxito de amplificación, conjuntando 277 muestras (51% del total inicial) para los análisis posteriores (Tabla VIII).

Tabla VIII. Total de muestras por población incorporadas al análisis con genotipos de los loci 7 loci microsatélite (Sgum05, Sgum06, Sgum12, Sgum25, Sgum29, Sgum36 y Sgum39).

<b>Clave</b>	<b>Localidad</b>	<b>≤ 1 genotipos faltantes (% del total)</b>
L05	Cabo San Lucas	22 (58%)
L10	Caduaño	14 (35%)
L13	Todos Santos	16 (47%)
L01	El Comitán	16 (46%)
L03	El Cien	17 (47%)
L04	Puente Querétaro	21 (50%)
L26	El Cuarenta	17 (46%)
L20	Sta. Águeda	17 (41%)
L24	Piedra Blanca	31 (74%)
L22	Punta Prieta	16 (48%)
L29	San Fernando Velicata	27 (61%)
L30	San Pedro	14 (39%)
L36	Isla Tiburón	20 (51%)
L41	Bahía Sargento	29 (64%)
<b>Total</b>		<b>277 (51%)</b>

### 7.3. Calidad y utilidad de los marcadores genotipificados

#### 7.3.1. *Detección de alelos nulos*

La estimación de frecuencia de alelos nulos se muestra en el Tabla IX. Los resultados se muestran catalogados por clases, en color gris oscuro para frecuencias excesivas, en gris intermedio para altas, gris claro para moderadas y sin color las frecuencias bajas o iguales a cero. Los límites fueron tomados del estudio de Chapuis y Stoup (2007) donde valores menores a 0.2 fueron determinados como tolerables para identificar estructura genética a partir de  $F_{ST}$  y  $D_C$ .

Tabla IX. Frecuencia de alelos nulos por loci, por localidad, calculada por el algoritmo de Dempster (1977). Los valores entre 0.00-0.049 se muestran con fondo blanco, entre 0.05-0.19 en gris claro, entre 0.20-0.249 en gris medio y los  $\geq 0.25$  en gris oscuro.

Localidad	Sgum05	Sgum06	Sgum12	Sgum25	Sgum29	Sgum36	Sgum39
L05	0.290	0.000	0.000	0.105	0.091	0.256	0.001
L10	0.200	0.001	0.000	0.083	0.112	0.000	0.000
L13	0.224	0.000	0.073	0.000	0.238	0.176	0.120
L01	0.310	0.000	0.201	0.000	0.134	0.086	0.001
L03	0.327	0.000	0.100	0.293	0.138	0.240	0.000
L04	0.252	0.000	0.105	0.000	0.181	0.087	0.000
L26	0.344	0.005	0.124	0.099	0.134	0.185	0.001
L20	0.254	0.000	0.134	0.000	0.190	0.085	0.001
L24	0.341	0.000	0.221	0.001	0.025	0.196	0.000
L22	0.275	0.000	0.000	0.000	0.165	0.225	0.000
L29	0.341	0.000	0.174	0.001	0.109	0.254	0.000
L30	0.291	0.000	0.123	0.001	0.055	0.245	0.001
L36	0.364	0.039	0.168	0.000	0.132	0.179	0.000
L41	0.258	0.000	0.053	0.000	0.011	0.164	0.101

Debido a la alta frecuencia de alelos nulos estimada para el locus Sgum05 y a la prevalencia de estos en todas las localidades, se realizó una comparación de su efecto en la heterocigocidad. En la Figura 6 se observa como en todas las localidades la  $H_o$  aumenta al excluir el locus Sgum05. En consecuencia, se decidió proceder en los análisis sin dicho locus.

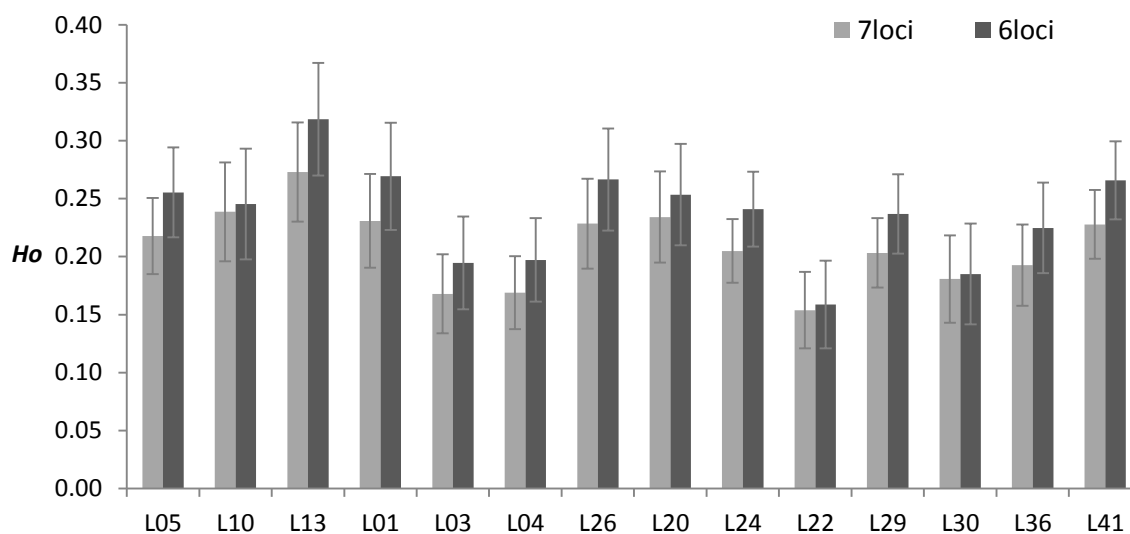


Figura 6. Comparación de Heterocigocidad observada ( $H_o$ ) calculada con 7 loci (Sgum05, Sgum06, Sgum12, Sgum25, Sgum29, Sgum36 y Sgum39) en gris claro y con 6 loci excluyendo el locus Sgum05 en gris oscuro. Las líneas grises al centro de las columnas indican la desviación estándar debida al cálculo de  $H_o$  de cada locus.

### 7.3.2. Pruebas de desequilibrio de ligamiento

Para esta prueba por pares de loci, ninguna de las comparaciones exhibió valores significativos (Tabla X), es decir que los loci empleados en este trabajo, presuntamente, se segregan de manera independiente y, por lo tanto esto no sería un agente causante del desajuste al EHW.

Tabla X. Valores  $p$  para la segregación independiente por par de loci calculados en GenePop (Rousset, 2008).

Par de loci	valor $p$	Par de loci	valor $p$
Sgum06 y Sgum12	0.96696	Sgum25 y Sgum36	0.99636
Sgum06 y Sgum25	0.82271	Sgum29 y Sgum36	0.71581
Sgum12 y Sgum25	0.99954	Sgum06 y Sgum39	0.52422
Sgum06 y Sgum29	0.82274	Sgum12 y Sgum39	0.47597
Sgum12 y Sgum29	0.09477	Sgum25 y Sgum39	0.92079
Sgum25 y Sgum29	0.99913	Sgum29 y Sgum39	1.00000
Sgum06 y Sgum36	0.95357	Sgum36 y Sgum39	0.97589
Sgum12 y Sgum36	0.46887		

### 7.4. Diversidad por localidad

En la Tabla XI se muestran los valores de diversidad calculados para cada localidad por locus, así como los resultados de la prueba de desviación de desviación del EHW. El número de alelos por locus ( $A$ ) varió de 1 hasta 7, con un promedio de 2.76; el número efectivo de alelos, es decir, la cantidad equivalentemente frecuente, tomó valores desde 1 hasta 2.86, con una media de 1.61. El índice de información, equivalente al índice ecológico de Shannon-Weaver, tomó valores de 0.0 en los loci que dentro de una localidad fueron monomórficos y alcanzó hasta 1.23, reportando los valores más altos en los loci Sgum06 y Sgum36, con una media global de 0.56. La heterocigocidad observada ( $H_O$ ) registro valores entre 0 y 0.75 con un promedio de 0.237, la heterocigocidad esperada no sesgada ( $H_E$ ) se encontró entre 0 y 0.67 con un promedio de 0.329 y los valores del índice de fijación ( $F$ ) variaron entre -0.432 y 1.0 con un promedio global de 0.242.

Se encontró desviación significativa del Equilibrio Hardy-Weinberg ( $H_E - H_O \neq 0.0$ ) en 2 de las comparaciones usando la corrección de los dos métodos (Bonferroni y FDR): las localidades de L05, L24 y L29 en el locus Sgum36 y la localidad L03 en el locus Sgum25; este último tuvo un  $F$  de 1.0, que podría deberse a endogamia, alelos nulos o a desventaja selectiva de heterocigotos. Con el método de FDR, se encontraron 11 comparaciones de loci dentro de localidades con desviación significativa del EHW ( $p_{BH} \leq 0.05$ , Tabla XI), siendo 8 comparaciones correspondientes al locus

Sgum36, y una para los loci Sgum12, Sgum25 y Sgum29 respectivamente. Las localidades L03 y L24 mostraron dos loci con desviación significativa del EHW cada una.

Tabla XI. Índices de diversidad genética por locus, por localidad. Se especifica el tamaño de muestra ( $n$ ), el número de alelos totales encontrados ( $A$ ), el número efectivo de alelos ( $A_e$ ), el índice informativo ( $I$ ), la heterocigocidad observada ( $H_O$ ) y esperada insesgada ( $H_E$ ), los valores de  $p$  para desviación del EHW ( $p_{HWE}$ ) y con el ajuste FDR ( $p_{BH}$ ), en negritas  $p_{HWE} \leq 6.58 \times 10^{-5}$  y  $p_{BH} \leq 0.05$ ; además del índice de Fijación ( $F$ ).

Localidad	Estimador	Sgum06	Sgum12	Sgum25	Sgum29	Sgum36	Sgum39
	$n$	22	22	21	22	19	20
L05	$A$	4	2	3	2	4	1
San Lucas	$A_e$	2.225	1.252	1.561	1.482	2.063	1.000
	$I$	1.005	0.354	0.667	0.507	0.950	0.000
	$H_O$	0.6818	0.2273	0.2381	0.2273	0.1579	0.0000
	$H_E$	0.5634	0.2061	0.3682	0.3330	0.5292	-
	$p_{HWE}$	0.5802	1.0000	0.0843	0.1794	<b>0.0000</b>	-
	$p_{BH}$	0.8819	1.0000	0.2613	0.4011	<b>0.0000</b>	-
	$F$	-0.2383	-0.1282	0.3375	0.3016	0.6935	-
	$n$	14	14	13	13	13	14
L10	$A$	4	2	3	2	2	2
Cadauño	$A_e$	2.579	1.153	1.617	1.352	1.166	1.153
	$I$	1.091	0.257	0.700	0.429	0.271	0.257
	$H_O$	0.5714	0.1429	0.3077	0.1539	0.1539	0.1429
	$H_E$	0.6349	0.1376	0.3969	0.2708	0.1477	0.1376
	$p_{HWE}$	0.5345	1.0000	0.2538	0.2340	1.0000	1.0000
	$p_{BH}$	0.8291	1.0000	0.4705	0.4605	1.0000	1.0000
	$F$	0.0667	-0.0769	0.1938	0.4091	-0.0833	-0.0769
	$n$	16	16	16	14	14	16
L13	$A$	4	2	3	2	5	3
Todos Santos	$A_e$	2.681	1.679	1.796	1.600	1.719	1.290
	$I$	1.118	0.594	0.787	0.562	0.882	0.447
	$H_O$	0.7500	0.3125	0.4375	0.0714	0.2143	0.1250
	$H_E$	0.6472	0.4173	0.4577	0.3889	0.4339	0.2319
	$p_{HWE}$	0.7202	0.5306	0.3984	0.0068	0.0325	0.1902
	$p_{BH}$	1.0000	0.8291	0.6442	<b>0.0468</b>	0.1193	0.4130
	$F$	-0.1963	0.2271	0.0132	0.8095	0.4878	0.4435
	$n$	16	14	16	16	14	16
L01	$A$	3	2	3	2	5	1
El comitán	$A_e$	2.179	1.690	1.471	1.519	1.832	1.000
	$I$	0.895	0.598	0.602	0.525	0.905	0.000
	$H_O$	0.6250	0.1429	0.3750	0.1875	0.2857	0.0000
	$H_E$	0.5585	0.4233	0.3307	0.3528	0.4709	-
	$p_{HWE}$	1.0000	0.0255	1.0000	0.1086	0.0123	-
	$p_{BH}$	1.0000	0.1078	1.0000	0.2751	0.0640	-
	$F$	-0.1552	0.6500	-0.1707	0.4514	0.3708	-

Continuación Tabla XI.

Localidad	Estimador	Sgum06	Sgum12	Sgum25	Sgum29	Sgum36	Sgumr39
L03 El cien	<i>n</i>	17	15	17	16	16	17
	<i>A</i>	5	2	3	2	5	2
	<i>A<sub>e</sub></i>	1.895	1.642	1.615	1.205	2.599	1.125
	<i>I</i>	0.957	0.580	0.678	0.311	1.198	0.224
	<i>H<sub>O</sub></i>	0.4706	0.2667	0.0000	0.0625	0.2500	0.1177
	<i>H<sub>E</sub></i>	0.4866	0.4046	0.3922	0.1754	0.6351	0.1141
	<i>p<sub>HWE</sub></i>	0.3918	0.2260	<b>0.0001</b>	0.0963	0.0013	1.0000
	<i>p<sub>BH</sub></i>	0.6442	0.4605	<b>0.0013</b>	0.2711	<b>0.0160</b>	1.0000
	<i>F</i>	0.0037	0.3182	1.0000	0.6322	0.5937	-0.0625
L04 Puente Querétaro	<i>n</i>	21	21	21	20	19	20
	<i>A</i>	5	2	2	2	4	2
	<i>A<sub>e</sub></i>	1.793	1.208	1.049	1.280	1.847	1.051
	<i>I</i>	0.837	0.314	0.113	0.377	0.867	0.117
	<i>H<sub>O</sub></i>	0.5714	0.0952	0.0476	0.0500	0.3684	0.0500
	<i>H<sub>E</sub></i>	0.4530	0.1765	0.0476	0.2244	0.4708	0.0500
	<i>p<sub>HWE</sub></i>	0.8003	0.1447	1.0000	0.0104	0.3147	1.0000
	<i>p<sub>BH</sub></i>	1.0000	0.3333	1.0000	0.0606	0.5694	1.0000
	<i>F</i>	-0.2923	0.4474	-0.0244	0.7714	0.1964	-0.0256
L26 El cuarenta	<i>n</i>	17	17	17	17	14	17
	<i>A</i>	3	2	2	2	3	1
	<i>A<sub>e</sub></i>	1.973	1.486	1.335	1.335	1.562	1.000
	<i>I</i>	0.769	0.508	0.418	0.418	0.668	0.000
	<i>H<sub>O</sub></i>	0.5294	0.2941	0.2353	0.1875	0.3529	0.0000
	<i>H<sub>E</sub></i>	0.5865	0.4866	0.3191	0.3528	0.6702	-
	<i>p<sub>HWE</sub></i>	0.8986	0.1386	0.0678	0.1077	0.0068	-
	<i>p<sub>BH</sub></i>	1.0000	0.3291	0.2239	0.2751	<b>0.0468</b>	-
	<i>F</i>	-0.4316	0.4603	-0.1724	0.7655	0.2057	-
L20 Santa Águeda	<i>n</i>	17	17	17	16	17	17
	<i>A</i>	4	2	4	2	5	1
	<i>A<sub>e</sub></i>	2.321	1.895	1.449	1.519	2.861	1.000
	<i>I</i>	0.975	0.665	0.644	0.525	1.220	0.000
	<i>H<sub>O</sub></i>	0.7059	0.1765	0.2941	0.0588	0.2857	0.0000
	<i>H<sub>E</sub></i>	0.5080	0.3369	0.2585	0.2585	0.3730	-
	<i>p<sub>HWE</sub></i>	0.0584	0.0960	1.0000	0.0148	0.2355	-
	<i>p<sub>BH</sub></i>	0.2017	0.2711	1.0000	0.0663	0.4605	-
	<i>F</i>	0.0699	0.3773	0.2402	0.4514	0.4574	-
L24 Piedra Blanca	<i>n</i>	31	26	29	29	30	31
	<i>A</i>	5	2	1	3	5	2
	<i>A<sub>e</sub></i>	2.397	1.899	1.000	1.736	2.707	1.033
	<i>I</i>	1.023	0.666	0.000	0.670	1.232	0.083
	<i>H<sub>O</sub></i>	0.5807	0.1539	0.0000	0.3793	0.3000	0.0323
	<i>H<sub>E</sub></i>	0.5923	0.4827	-	0.4313	0.6412	0.0323
	<i>p<sub>HWE</sub></i>	0.9026	0.0007	-	0.7419	<b>0.0000</b>	1.0000
	<i>p<sub>BH</sub></i>	1.0000	<b>0.0109</b>	-	1.0000	<b>0.0000</b>	1.0000
	<i>F</i>	0.0036	0.6750	-	0.1052	0.5242	-0.0164



Continuación Tabla XI.

Localidad	Estimador	Sgum06	Sgum12	Sgum25	Sgum29	Sgum36	Sgumr39
L22 Punta Prieta	<i>n</i>	16	16	16	16	13	16
	<i>A</i>	4	2	2	2	3	2
	<i>A<sub>e</sub></i>	2.276	1.205	1.064	1.133	1.489	1.133
	<i>I</i>	0.972	0.311	0.139	0.234	0.619	0.234
	<i>H<sub>O</sub></i>	0.5000	0.1875	0.0625	0.0000	0.0769	0.1250
	<i>H<sub>E</sub></i>	0.5786	0.1754	0.0625	0.1210	0.3415	0.1210
	<i>p<sub>HWE</sub></i>	0.6372	1.0000	1.0000	0.0330	0.0052	1.0000
	<i>p<sub>BH</sub></i>	0.9495	1.0000	1.0000	0.1193	<b>0.0441</b>	1.0000
<i>F</i>	0.1080	-0.1034	-0.0323	1.0000	0.7658	-0.0667	
L29 San Fernando Velicata	<i>n</i>	27	26	26	27	22	27
	<i>A</i>	5	2	1	3	5	2
	<i>A<sub>e</sub></i>	2.202	1.954	1.000	1.668	2.588	1.038
	<i>I</i>	0.982	0.681	0.000	0.679	1.197	0.092
	<i>H<sub>O</sub></i>	0.6296	0.2308	0.0000	0.2963	0.2273	0.0370
	<i>H<sub>E</sub></i>	0.5563	0.4977	-	0.4081	0.6279	0.0370
	<i>p<sub>HWE</sub></i>	0.3585	0.0131	-	0.0135	<b>0.0002</b>	1.0000
	<i>p<sub>BH</sub></i>	0.6055	0.0640	-	0.0640	<b>0.0038</b>	1.0000
<i>F</i>	-0.1533	0.5273	-	0.2603	0.6296	-0.0189	
L30 San Pedro	<i>n</i>	14	14	14	14	11	13
	<i>A</i>	2	2	1	2	4	1
	<i>A<sub>e</sub></i>	1.849	1.849	1.000	1.415	2.200	1.000
	<i>I</i>	0.652	0.652	0.000	0.469	0.956	0.000
	<i>H<sub>O</sub></i>	0.4286	0.2857	0.0000	0.2143	0.1818	0.0000
	<i>H<sub>E</sub></i>	0.4762	0.4762	-	0.3042	0.5714	-
	<i>p<sub>HWE</sub></i>	1.0000	0.2382	-	0.3466	0.0028	-
	<i>p<sub>BH</sub></i>	1.0000	0.4605	-	0.6000	<b>0.0269</b>	-
<i>F</i>	0.0667	0.3778	-	0.2696	0.6667	-	
L36 Isla Tiburón	<i>n</i>	20	19	20	18	18	19
	<i>A</i>	4	2	2	2	3	2
	<i>A<sub>e</sub></i>	2.694	1.566	1.051	1.180	1.889	1.296
	<i>I</i>	1.136	0.547	0.117	0.287	0.826	0.389
	<i>H<sub>O</sub></i>	0.6000	0.1579	0.0500	0.0556	0.2222	0.2632
	<i>H<sub>E</sub></i>	0.6449	0.3713	0.0500	0.1571	0.4841	0.2347
	<i>p<sub>HWE</sub></i>	0.1202	0.0275	1.0000	0.0860	0.0094	1.0000
	<i>p<sub>BH</sub></i>	0.2948	0.1100	1.0000	0.2613	0.0597	1.0000
<i>F</i>	0.0457	0.5632	-0.0256	0.6364	0.5279	-0.1515	
L41 Bahía Sargento	<i>n</i>	29	28	28	29	29	29
	<i>A</i>	5	2	2	2	7	3
	<i>A<sub>e</sub></i>	2.240	1.280	1.074	1.399	2.182	1.364
	<i>I</i>	1.031	0.377	0.154	0.460	1.185	0.486
	<i>H<sub>O</sub></i>	0.5517	0.1786	0.0714	0.2759	0.3448	0.1724
	<i>H<sub>E</sub></i>	0.5632	0.2227	0.0701	0.2904	0.5511	0.2716
	<i>p<sub>HWE</sub></i>	0.2424	0.3474	1.0000	1.0000	0.0020	0.1046
	<i>p<sub>BH</sub></i>	0.4605	0.6000	1.0000	1.0000	<b>0.0212</b>	0.2751
<i>F</i>	0.0032	0.1837	-0.0370	0.0333	0.3633	0.3541	

Continuación Tabla XI

Localidad	Estimador	Sgum06	Sgum12	Sgum25	Sgum29	Sgum36	Sgumr39
	<i>n</i>	277	265	271	267	249	272
Todas las localidades	<i>A</i>	9	2	5	3	10	3
	<i>I</i>	0.960	0.508	0.358	0.461	0.927	0.166
	<i>F</i>	-0.079	0.321	0.120	0.493	0.457	0.042

Aunque no se encontró desviación del EHW en la mayoría de loci hacia adentro de cada localidad, el efecto del promedio de loci para cada localidad resultó en déficit de heterocigotos significativo para 10 localidades (Tabla XII). Los  $F_{IS}$  resultaron mayores a cero en todos los casos, alcanzando un valor máximo de 0.449 para L03 y un mínimo de 0.026 para L26, con un valor global para todos los loci de 0.284 ( $p = 0.0000$ ).

Tabla XII. Promedio para todos los loci por localidad de los índices de diversidad genética; tamaño de muestra ( $n$ ), riqueza alélica ( $R$ ), número efectivo de alelos ( $A_e$ ), riqueza alélica  $S = 18$  ( $R_{18}$ ), la heterocigocidad esperada insesgada ( $H_E$ ), heterocigocidad observada ( $H_O$ ) y el coeficiente de endogamia ( $F_{IS}$ ). El tamaño de las barras en la última columna es proporcional a los valores de  $F_{IS}$  entre 0 - 1.

Localidad	$nx$	$R$	$A_e$	$R_{18}$	$H_E$	$H_O$	$F_{IS}$
L05	21.0	2.67	1.597	2.46	0.333	0.255	0.211
L10	13.5	2.50	1.503	2.40	0.288	0.245	0.091
L13*	15.3	3.17	1.794	2.89	0.429	0.318	0.233
L01	15.3	2.67	1.615	2.44	0.356	0.269	0.213
L03*	16.3	3.17	1.680	2.80	0.368	0.195	0.449
L04*	20.3	2.83	1.371	2.24	0.237	0.197	0.138
L26*	16.5	3.00	1.841	2.60	0.403	0.267	0.347
L20	16.8	2.17	1.448	2.07	0.289	0.254	0.064
L24*	29.3	3.00	1.795	2.36	0.363	0.241	0.317
L22*	15.5	2.50	1.383	2.23	0.233	0.159	0.315
L29*	25.8	3.00	1.742	2.42	0.355	0.237	0.311
L30*	13.3	2.00	1.552	1.94	0.305	0.185	0.383
L36*	19.0	2.50	1.613	2.34	0.324	0.225	0.286
L41*	28.7	3.50	1.590	2.68	0.328	0.266	0.172

\* Déficit de heterocigotos significativo, después de corrección Bonferroni ( $\alpha = 0.05$ ).

En cuanto a la rarefacción de la riqueza alélica ( $R$ ), en la Figura 7a se observa cómo se alcanza en muchas ocasiones el límite que el tamaño de muestra permite, describiendo curvas que alcanzan su valor máximo de  $R$  antes de alcanzar la forma asintótica, por lo tanto se tomaron los valores para la muestra balanceada ( $S$ ) de 18. Los resultados de la regresión lineal para estos valores de  $R$  de las localidades de la Península, confirmaron una tendencia decreciente de la diversidad en

función de la latitud ( $R^2 = 0.383$ ,  $p = 0.032$ ) (Figura 10). Cuando se utilizaron valores de  $R|S > 22$ , se perdió la significancia estadística de la regresión ( $\alpha = 0.05$ ) (Datos no mostrados). Para la prueba con  $H_E$  no se encontró ninguna tendencia ni significancia con la regresión lineal ( $R^2 = 0.055$ ,  $p = 0.464$ ).

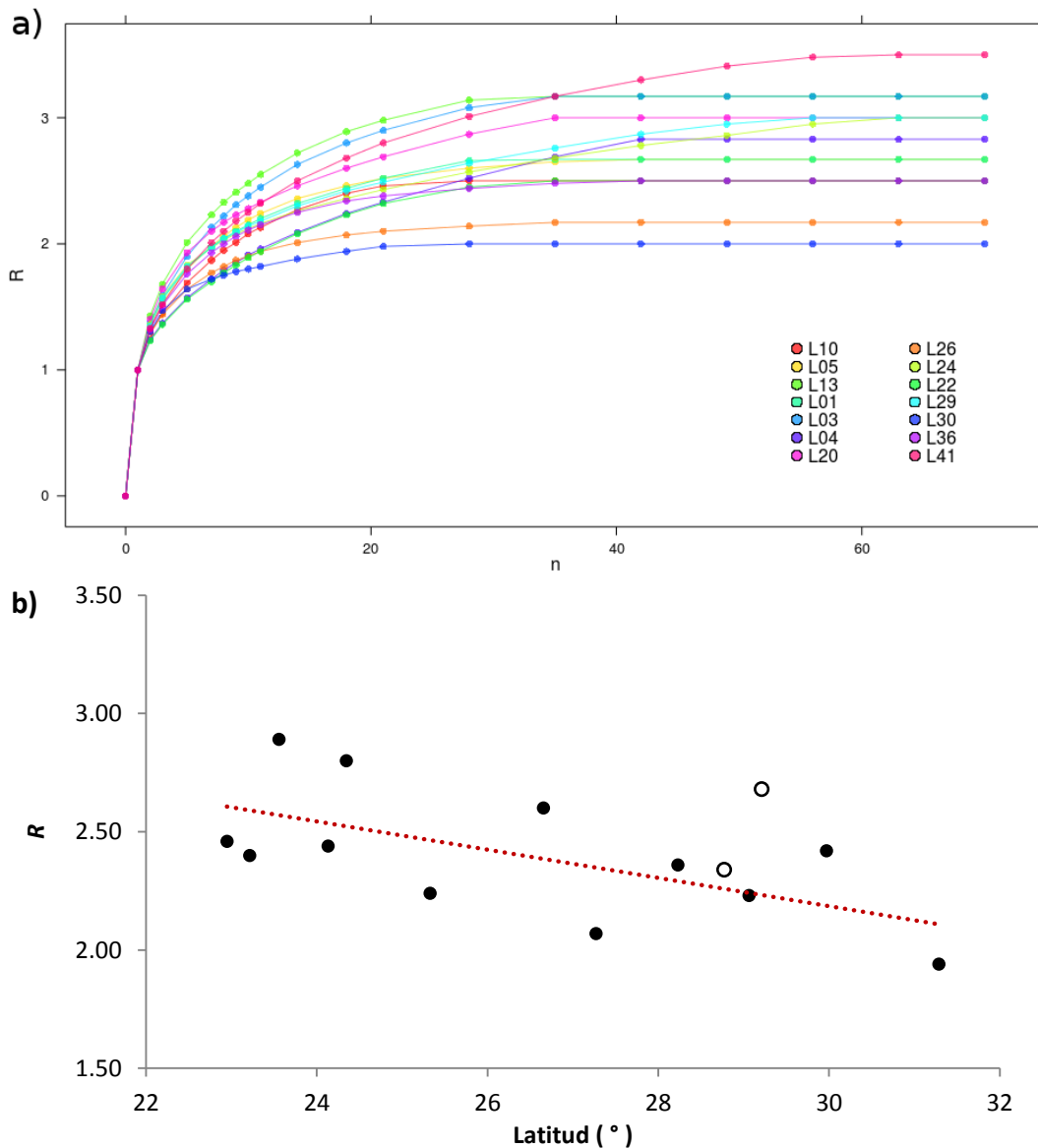


Figura 7. a) Rarefacción de la riqueza alélica ( $R$ ) por localidad obtenido tomando en cuenta los loci Sgum06, Sgum12, Sgum25, Sgum29, Sgum36 y Sgum39. El aumento en  $n$  corresponde al número de alelos. b) Gradiente latitudinal en  $R | S = 18$  a lo largo de la Península de Baja California (círculos rellenos), y dos muestras correspondientes a la Isla Tiburón y a Sonora (círculos vacíos); el modelo de la regresión lineal se muestra en color rojo con la línea punteada.

## 7.5. Análisis de estructura poblacional

### 7.5.1. *Análisis Factorial de Correspondencia (AFC) y Análisis Discriminante de Componentes Principales (ADCP)*

El AFC efectuado para la totalidad de las muestras, logró explicar el 58.55% de la varianza genética con los primeros 3 factores o ejes (24.67% 19.24% 14.63%, ejes 1, 2 y 3, respectivamente), pero al graficar la ubicación tridimensional de cada muestra con los valores de las cargas correspondientes a esos primeros 3 factores (Figura 8a), no se logró detectar una separación que refleje a las localidades como grupos diferenciados genéticamente. Sin embargo, se aprecia cierta orientación o tendencia de agrupación, la dispersión de puntos sobre el eje 1 permite detectar que las muestras de las localidades más norteñas se ubican hacia los valores más negativos (L30, L29, L22, L24), mezcladas con las localidades de la Isla Tiburón y de Sonora (L36 y L41) que se encuentran en un rango de valores más amplio; las de la región central de la Península (L26, L20, L04 y L03), quedan mezcladas con todas las demás localidades, L03 ocupa un rango más disperso que el resto; finalmente las localidades más sureñas (LL01, L13, L10 y L05) quedaron orientadas hacia los valores más altos del eje 1 y notablemente con menos cohesión espacial como grupo. Esta tendencia de organización regional, se logra identificar con mayor facilidad al analizar las localidades en grupos (Sur, Centro, Norte de la Península y Sonora-Isla Tiburón) posterior al ADCP (Figura 8b), aunque los elipses de varianza de las cargas a partir de los centroides de cada grupo quedan ampliamente traslapados, siendo muy cercana la probabilidad de pertenencia a cualquiera de los grupos para todas la muestras (datos de probabilidad no mostrados).

Posteriormente se repitió el AFC incorporando solamente las muestras de la Península de Baja California. Aunque el porcentaje de la varianza explicada con los primeros 3 factores fue ligeramente mayor (60.48%), al graficar las cargas por muestra, no se observó mejor nivel de separación entre las localidades como grupos genéticos (no se incluye figura), en comparación con la Figura 8a.

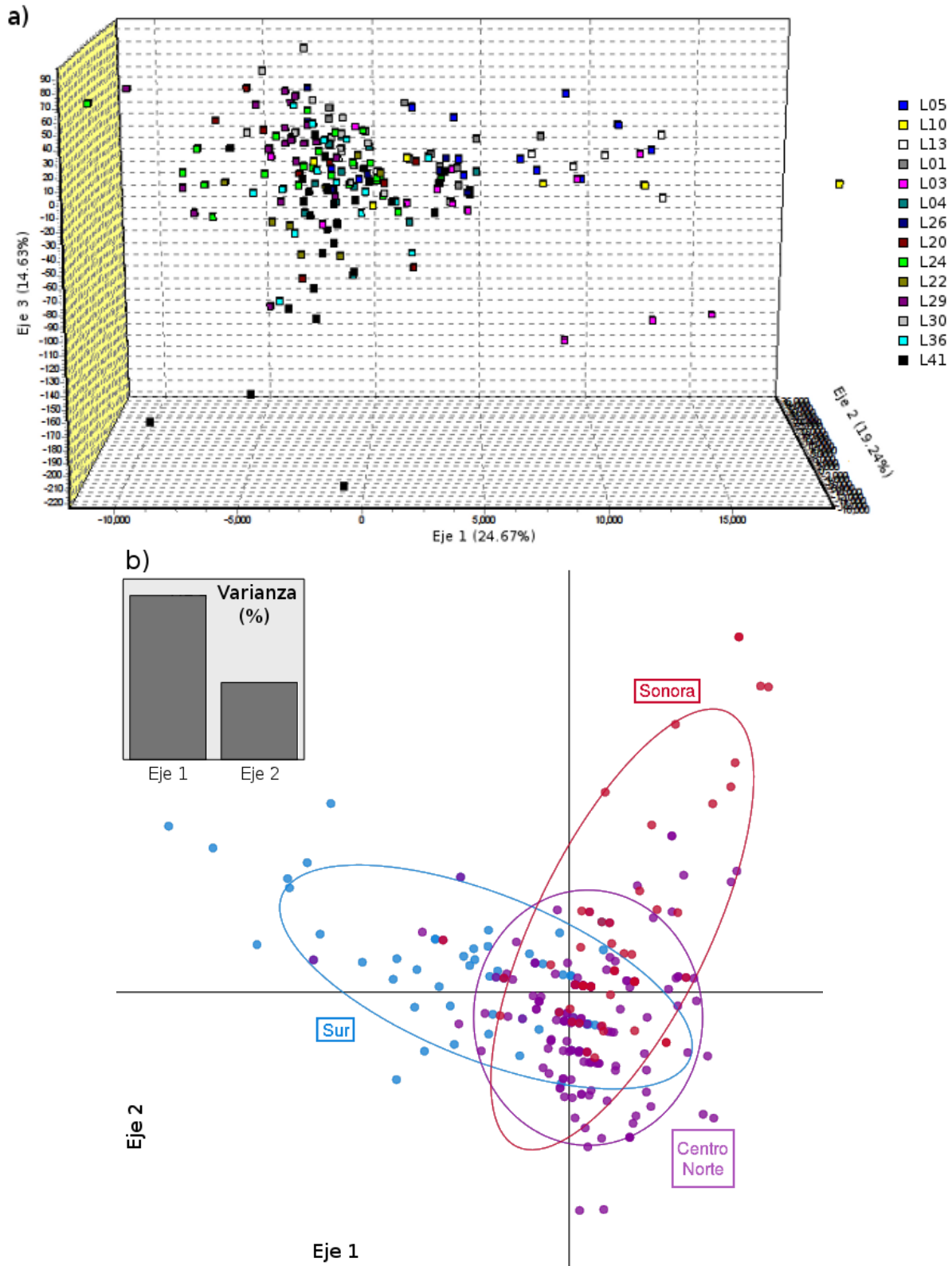


Figura 8. a) Gráfico de dispersión de los valores de las cargas de cada una de las muestras diferenciadas por el color de la localidad de origen, resultado de los primeros 3 ejes o factores calculados por el AFC. b) Valores de las cargas de cada muestra después del ADCP, con las regiones como agrupación, desplegando elipses proyectados a partir de los centroides con la varianza de las cargas de cada grupo.

### 7.5.2. Análisis de varianza molecular (AMOVA)

Al efectuar el AMOVA con el modelo donde cada localidad es un grupo, el 95.38% de la varianza genética se encontró repartida entre los individuos ( $F_{ST} = 0.0462$ ,  $p = 0.000$ ) y el restante 4.62% entre las localidades. Para el modelo equivalente con sólo las localidades de la Península fue prácticamente el mismo resultado con 95.15% de la varianza entre los individuos ( $F_{ST} = 0.0485$ ,  $p = 0.000$ ). En la Tabla XIII se resumen los resultados de los distintos modelos sometidos a AMOVA.

Tabla XIII. Resultados de AMOVAS para distintas formas de agrupación, tomando en cuenta eventos históricos (ver 6.5.3.b y Figura 3). Los porcentajes en las columnas 3, 4 y 5 corresponden al total de la varianza explicada por el nivel jerárquico especificado; los valores  $F_{CT}$  corresponden al nivel Entre grupos,  $F_{SC}$  a Entre subpoblaciones dentro de grupos y  $F_{ST}$  a Dentro de subpoblaciones.

Modelo	Nivel	Entre grupos (%)	Entre subpobs. dentro de grupos (%)	Dentro subpobs. (%)	$F_{CT}$	$p_{Fct}$	$F_{SC}$	$p_{Fsc}$	$F_{ST}$	$p_{Fst}$
Localidades	Península	4.85	--	95.15	--	--	--	--	0.0485	0.000
CVi	Península	3.03	3.21	93.76	0.0303	0.000	0.0331	0.002	0.0624	0.000
CTP	Península	2.44	3.05	94.51	0.0244	0.008	0.0312	0.003	0.0549	0.000
IIP	Península	1.67	4.04	94.28	0.0167	0.046	0.0411	0.000	0.0572	0.000
Localidades	Total	4.62	--	95.38	--	--	--	--	0.0462	0.000
ViG	Total	2.08	3.21	94.71	0.0208	0.008	0.0328	0.000	0.0529	0.000
CsG	Total	1.94	3.05	95.01	0.0194	0.009	0.0311	0.002	0.0499	0.000
NJ	Total	3.82	1.82	94.36	0.0382	0.000	0.0190	0.044	0.0564	0.000

Al realizar la comparación de los resultados para los 8 modelos probados (Figura 9) podemos ver como en todos los casos la mayor parte de la varianza se encontró dentro de subpoblaciones (localidades) entre un 93.76% y hasta un 95.01%. Al probar una estructuración concordante con la agrupación de los canales propuestos en la PBC, el modelo CVi fue el que detectó menor variación dentro de las subpoblaciones (93.76%) y mayor variación entre grupos (3.03%), con estadísticos  $F$  muy bajos pero en todos los casos significativos (Tabla XIII). Al incorporar las agrupaciones que toman en cuenta la separación del Golfo de California (G) el porcentaje de la varianza detectado al interior de las subpoblaciones fue de 3.21% y 3.05%, pero en ambos casos la varianza entre grupos se vio reducida, hasta 1.94% tomando en cuenta los dos canales y el Golfo (CsG). Finalmente, cuando se evaluó el modelo de agrupación inferido con el resultado del cladograma NJ, la varianza genética entre las subpoblaciones dentro de los grupos fue de 1.82%, el valor menor encontrado en este nivel de jerarquía, y entre grupos de 3.82%, pero la varianza hacia adentro de las subpoblaciones se mantuvo sin variación.

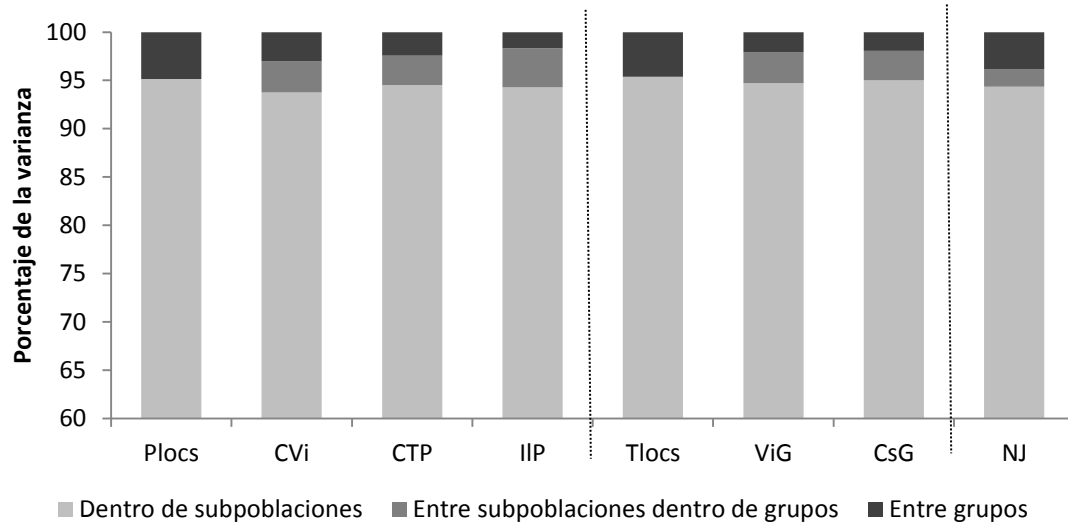


Figura 9. Porcentaje de la varianza por jerarquía, explicada por el AMOVA para cada modelo planteado (ver Tabla V): los primeros cuatro son sólo para las localidades de la Península, los siguientes para la totalidad de las muestras, donde los primeros modelos corresponden a las localidades como jerarquía de agrupación (Plocs y Tlocs). El resto de modelos contemplan las separaciones que hubieran generado los canales transpeninsulares atravesando el Desierto del Vizcaíno (CVi), el Istmo de la Paz (IIP) o ambos (CTP) (Figura 3), considerando la separación del CVi y la del Golfo de California (ViG) y considerando CTP y el Golfo (CsG). Finalmente, un modelo de agrupación conformado a partir de los resultados del cladograma de distancias genéticas (Figura 11). El eje x muestra el porcentaje cortando el gráfico a partir de 60, para ilustrar los cambios a la varianza entre y dentro de grupos.

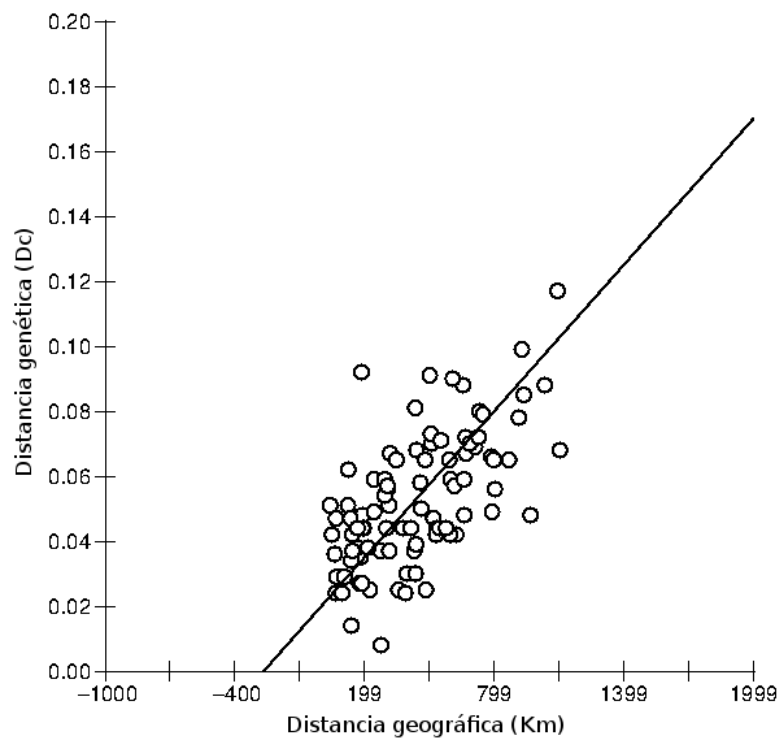


Figura 10. Correlación entre distancia genética y distancia geográfica por pares de localidades ( $r = 0.62$ ;  $p = 0.0002$ ) evaluada por test de Mantel.

### 7.5.1. Aislamiento por distancia

La prueba de Mantel para las 14 localidades, resultó positiva ( $p = 0.0002$ ) con un valor de  $r = 0.62$  y una  $Z = 2418.45$ , confirmando la existencia del patrón de aislamiento por distancia. Cuando se realizó el mismo análisis sólo para las muestras de la península, los parámetros del test de Mantel fueron muy similares ( $r = 0.60$ ,  $Z = 1889.66$ ,  $p = 0.0002$ ). En la Figura 10 se observa gráficamente la relación entre la distancia geográfica y la distancia genética.

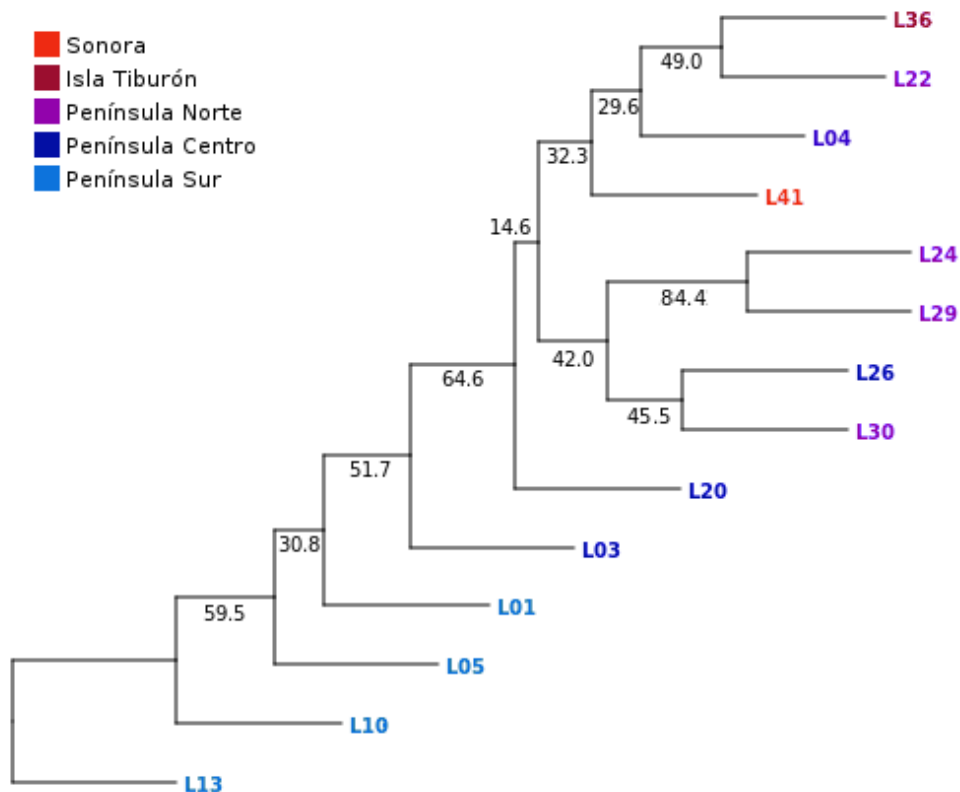


Figura 11. Cladograma no enraizado reconstruido por el método del vecino más cercano. Los números en las ramas representan el soporte de remuestreo (bootstrapping, 50,000 permutaciones); la longitud de las ramas no representa la distancia genética. Para el significado de los acrónimos ver Tabla II.

### 7.5.2. Dendograma de distancia genética



En la Figura 11 se muestra el dendograma consenso reconstruido por el método del vecino más cercano, aunque con bajo soporte de bootstrap (menor a 50%), se observa un patrón de clados anidados a nivel regional, del sur hacia el centro de la Península, pero estas localidades no se ordenan ascendentemente con la latitud (L13, L10, L05, L01, L03, L20). Las localidades del norte (sin contar a L22) se separan en dos clados recíprocos, en uno se incluye a L26 del centro de la PBC con L30, localidad del extremo norte de la distribución. La localidad L04 quedó agrupada en la misma rama que L22, L36 y L41, siendo las últimas dos de la Isla Tiburón y Sonora. El valor más alto de bootstrap fue para el clado L24-L29 con 84.4%, seguido por la rama que integra a las localidades del sur de la Península (incluyendo a L03) con un 64.6%.

### 7.5.3. *Índices de Fijación de Wright ( $F_{ST}$ )*

Los resultados de la prueba de diferenciación genética mediante el cálculo de  $F_{ST}$  pareados por sitios de muestro, detectaron diferencias genéticas significativas en un tercio de las comparaciones (Tabla XIV), exhibiendo valores desde 0.04 hasta 0.192, diferenciación baja y moderada (Freeland, 2005).

No se logró detectar diferencia significativa de las localidades L01 ni L26 con ninguna de las demás. Las señales de diferenciación más notables son ambas correspondientes a la localidad más norteña (L30) diferenciándose con un  $F_{ST} = 0.192$  ( $p = 0.009$ ) de L10, la segunda localidad más sureña, y  $F_{ST} = 0.16$  ( $p = 0.000$ ) en comparación con L22, que se encuentra en la misma región norte que L30. Entre las localidades del sur no se detectó diferencia significativa al compararlas entre sí. Sin embargo algunas de las localidades del sur si mostraron diferencia con respecto a otras localidades, por ejemplo L05 presentó valores desde  $F_{ST} = 0.052$  hasta  $F_{ST} = 0.091$  ( $p \leq 0.049$ ) con respecto a las localidades más al norte de la Península y con la Isla Tiburón (L36); L10, con valores ligeramente más elevados, se diferenció con un patrón similar de las mismas que L05 y L13, sólo se diferenció significativamente de L30; L10 y L13, se distinguieron moderadamente de L03 y L04. La localidad L20 del centro de la Península, se diferenció moderadamente de L04 en la misma región con un  $F_{ST} = 0.055$  ( $p = 0.047$ ). Entre las localidades del norte de la Península, L22 se diferenció del resto con índices de fijación moderados de 0.069 y 0.06 ( $p \leq 0.05$ ) en comparación con L24 y L29 respectivamente, y de L30 con  $F_{ST} = 0.16$  ( $p = 0.00$ ), el segundo valor más alto encontrado, por el contrario no se diferenció significativamente de las localidades no peninsulares (L36 y L41). L36 a mitad del Golfo de California, tiene mayor similitud con las localidades del norte de la PBC, y L41 en Sonora no mostró ninguna diferencia significativa con localidades del sur de la Península.

Tabla XIV. Bajo la diagonal central se encuentran los valores de  $F_{ST}$  (Weir & Cockerham, 1984) para cada par de localidades y en la parte superior se muestran sus correspondientes valores de  $p$  posterior a la corrección de FDR (Benjamini y Hochberg, 1995); en color negro se muestran los valores de  $F_{ST}$  significativos ( $p \leq 0.05$ , en negritas), clasificados por su magnitud (Freeland, 2005), con fondo gris claro los valores bajos ( $F_{ST} \leq 0.05$ ), con gris intermedio los valores moderados menores a 0.15 y con gris oscuro los moderados mayores de 0.15.

	L05	L10	L13	L01	L03	L04	L26	L20	L24	L22	L29	L30	L36	L41
L05		0.521	0.559	0.800	0.411	0.249	0.559	0.118	<b>0.026</b>	<b>0.047</b>	<b>0.000</b>	<b>0.050</b>	<b>0.040</b>	0.113
L10	0.01		0.327	0.312	<b>0.047</b>	0.079	0.482	<b>0.030</b>	<b>0.009</b>	0.205	<b>0.000</b>	<b>0.009</b>	0.086	0.103
L13	0.00	0.00		0.933	0.238	<b>0.047</b>	0.566	0.439	0.211	0.108	0.108	<b>0.047</b>	0.227	0.126
L01	-0.02	0.01	-0.03		0.559	0.258	0.816	0.643	0.414	0.105	0.126	0.149	0.226	0.172
L03	0.01	<b>0.062</b>	0.02	0.00		0.113	0.249	0.266	<b>0.047</b>	<b>0.009</b>	<b>0.009</b>	0.115	<b>0.047</b>	0.064
L04	0.01	0.04	<b>0.051</b>	0.02	0.03		0.474	<b>0.047</b>	<b>0.034</b>	<b>0.047</b>	<b>0.000</b>	<b>0.017</b>	<b>0.026</b>	0.566
L26	0.00	0.00	-0.01	-0.02	0.02	0.00		0.329	0.179	0.184	0.063	0.051	0.226	0.406
L20	0.03	<b>0.074</b>	0.00	-0.01	0.02	<b>0.055</b>	0.01		0.829	<b>0.047</b>	0.385	0.376	0.220	<b>0.047</b>
L24	<b>0.052</b>	<b>0.088</b>	0.02	0.00	<b>0.050</b>	<b>0.059</b>	0.03	-0.02		<b>0.026</b>	0.578	0.168	0.172	<b>0.037</b>
L22	0.055	0.02	0.04	0.05	0.101	0.062	0.03	<b>0.070</b>	<b>0.069</b>		<b>0.050</b>	<b>0.000</b>	0.679	0.079
L29	0.091	<b>0.111</b>	0.03	0.03	<b>0.094</b>	0.118	0.06	0.01	0.00	<b>0.060</b>		0.073	0.230	<b>0.009</b>
L30	<b>0.083</b>	<b>0.192</b>	<b>0.087</b>	0.05	0.05	0.123	0.10	0.01	0.04	<b>0.160</b>	0.06		<b>0.047</b>	<b>0.009</b>
L36	<b>0.054</b>	0.04	0.02	0.02	<b>0.066</b>	<b>0.057</b>	0.02	0.02	0.02	-0.01	0.02	<b>0.097</b>		0.107
L41	0.02	0.03	0.03	0.02	0.03	0.00	0.00	<b>0.041</b>	<b>0.040</b>	0.04	<b>0.079</b>	<b>0.109</b>	0.02	

#### 7.5.4. Forma teórica de la estructura genética

En la Figura 12 se muestra el gráfico poblacional resultante de la varianza genética encontrada. La topología está construida por los 14 nodos correspondientes cada uno a las localidades en análisis y 17 vértices. La localidad L10 queda en un extremo conectada solamente con L13, seguida por el resto de localidades pertenecientes a la región del sur. En cuanto a las localidades no peninsulares que quedaron conectadas entre sí, L36 se encontró también conectada con L22 en la Península a una latitud cercana, y L41 comparte un vértice con L04 que se ubica en la región centro de la PBC, relación que implicaría conectividad a larga distancia a través del Golfo de California. Los vértices que conectan a L30 con L03 y L26 respectivamente, representan las conexiones de mayor distancia geográfica a lo largo de la PBC sugiriendo una relación de las localidades encontradas a lo largo de la costa Pacífico. La topografía del gráfico en general representa una organización similar a la encontrada con el árbol de distancia genéticas (Figura 11),

con la ventaja de tener resueltos los valores bajos de bootstrap al detectar conectividad o relación con un alfa bajo ( $\alpha = 0.01$ ) entre los nodos ubicados en clados poco soportados en el árbol.

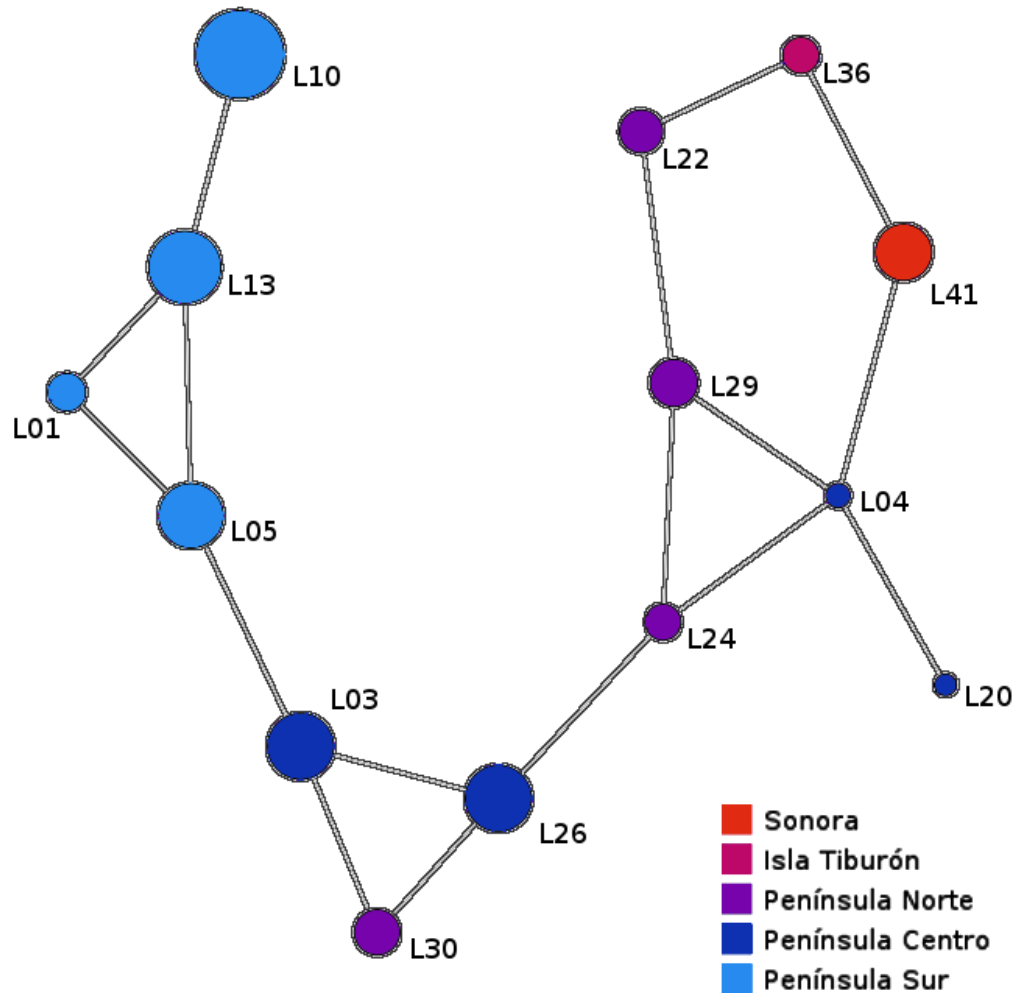


Figura 12. *Population Graph* ( $\alpha = 0.01$ ). Se muestran las leyendas con coloración clasificada *a priori* por latitud con cortes de agrupación correspondientes a los sitios de los canales transpeninsulares y al Golfo de California (Figura 3), el tamaño de los nodos es proporcional a la varianza intra-localidad.

#### 7.5.5. *Análisis de clusters*

En las pruebas de estructura por probabilidad bayesiana, STRUCTURE no identificó separación de grupos genéticamente homogéneos. A pesar de encontrar 2 grupos genéticos como el

valor con la mayor probabilidad posterior (Figura 13b), los coeficientes de identidad (Figura 14a) muestran un alto porcentaje de mezcla genética entre ellos.

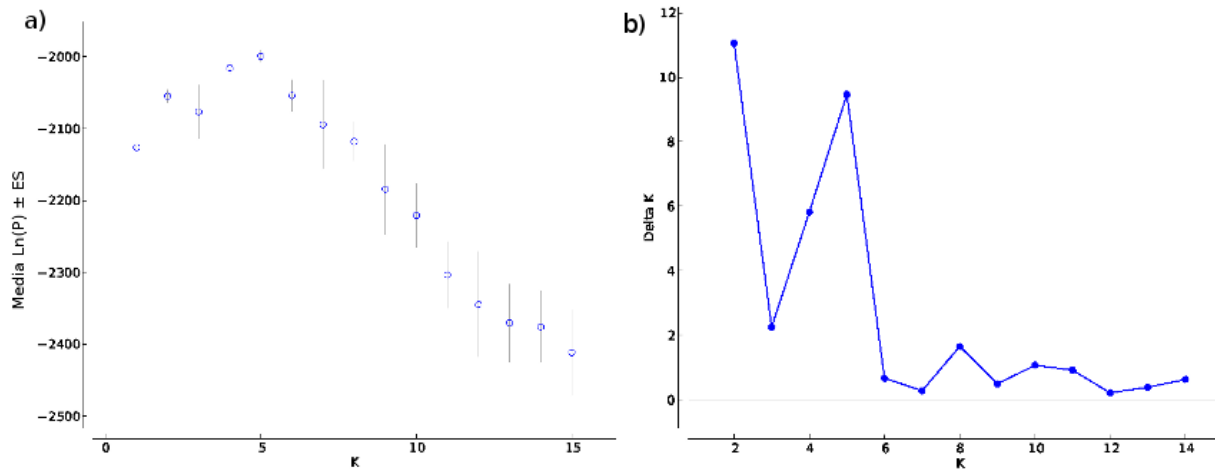


Figura 13. a) Representación del logaritmo natural de la probabilidad posterior promedio para 10 repeticiones probando distintos números de poblaciones ( $K$ ) de *Stenocereus gummosus*. b) Número de grupos genéticamente homogéneos ( $K$ ) calculado a partir de la tasa de cambio de la función de verosimilitud o  $\Delta K$  (Evanno et al., 2005).

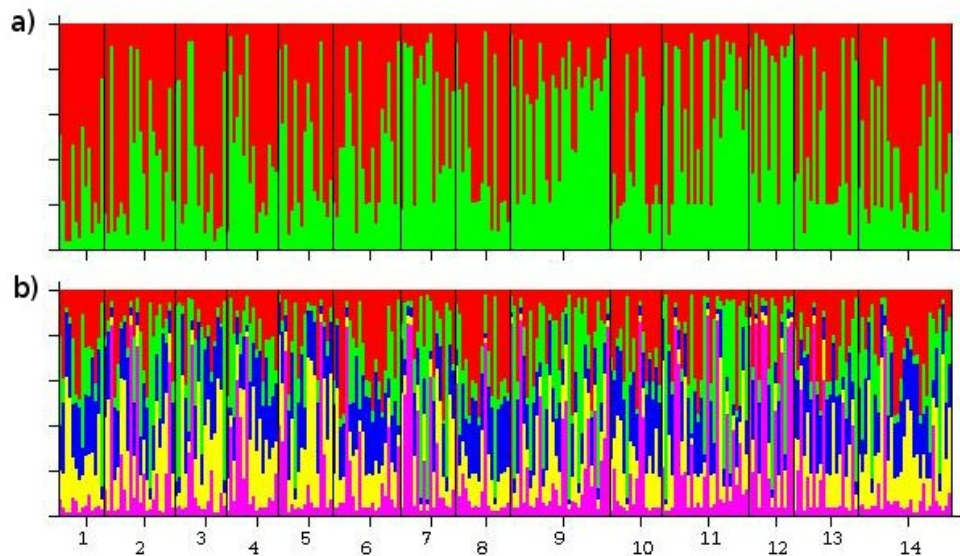


Figura 14. Coeficientes de identidad para a)  $K = 2$  y b)  $K = 5$  como número de grupos genéticos homogéneos, donde cada barra vertical corresponde a un individuo y cada color corresponde a la proporción de probabilidad de pertenecer a cada grupo. Los números en la parte inferior corresponden a las localidades de origen, para su correspondencia ver Tabla VIII.

STRUCTURE también indicó una posible subestructura, observándose asimismo una fuerte señal de que existan 5 grupos (Figura 13a y b), sin embargo la gráfica de los coeficientes de identidad (Figura 14b) arroja los mismos resultados de mezcla entre los grupos genéticos; estos

mismos análisis se realizaron excluyendo las localidades no peninsulares tanto incluyendo como ignorando el locus Sgum05, sin encontrar diferencias en los resultado (datos no mostrados).

## 8. DISCUSIÓN

### 8.1. Desarrollo y calidad de los marcadores

Las librerías enriquecidas para obtener microsatélites encontraron loci polimórficos en porcentajes de 53%, 70% y 91%, por encima de librerías genómicas (15% y 21%), en ambos casos con especies de la subfamilia Cactoideae (Tabla I). En este estudio, la pirosecuenciación con la química de titanium GS-FLX (Life Technologies) produjo un porcentaje de loci polimórficos de 30%; esta técnica permite elegir un mayor número de loci identificando el tipo de repetición y el número de repeticiones (Abdelkrim et al., 2009), por lo que se esperaba obtener mayor proporción de loci polimórficos. Esta discordancia podría deberse a una calidad baja del ADN utilizado para secuenciar o a que probablemente el sistema de reproducción que emplea *S. gummosus* afecte la riqueza alélica. Sin embargo, el número medio de alelos por locus en *S. gummosus* (5.33) estuvo muy cercano al promedio de los antecedentes revisados (5.48) (Tabla I).

Lamentablemente no hubo datos suficientes a lo largo de todas las localidades como para precisar la identidad clonal a través del análisis de los genotipos multilocus de las muestras, no obstante los resultados reportados por Molina Freaner y Clark Tapia (2005) fueron tomados en cuenta para diseñar un muestreo que evitara recolectar individuos clonales. Con esta certeza se estudió la estructura genética de la pitahaya agria con 6 loci microsatélite polimórficos para 12 localidades peninsulares, una insular y una continental, utilizando 7 metodologías distintas. Los coeficientes de endogamia fueron menores (global  $F_{IS} = 0.284$ ) a los determinados con loci aloenzimáticos (global  $F_{IS} = 0.608$ ), y la heterocigocidad resultó mayor ( $0.159 \leq H_O \leq 0.318$ ,  $0.233 \leq H_E \leq 0.429$ ) que anteriormente ( $0.063 \leq H_O \leq 0.175$ ,  $0.207 \leq H_E \leq 0.351$ ) (Clark-Tapia y Molina-Freaner, 2003), permitiendo una mejor resolución para estudios de genética poblacional.

El éxito de amplificación obtenido por locus fue muy bajo para los loci Ppri13, Sgum24 y Sgum31; particularmente para Ppri13 se registraron 16 tamaños alélicos que, junto con el bajo éxito de amplificación, sugieren inespecificidad de los iniciadores de PCR. Sgum31, aunque también problemático por falta de amplificación, exhibió solamente 4 alelos y la mayoría de genotipos homocigotos. La falta de rendimiento en general se puede atribuir a una variabilidad muy elevada en los sitios de alineamiento de los iniciadores de PCR (Chapuis y Estoup, 2007), pues además se observó que los genotipos que se lograron amplificar para estos tres loci fueron siempre en algunas pocas localidades, sugiriendo un patrón de alelos fijados por mutaciones en los sitios de alineamiento de los iniciadores. Para determinar esta condición se requiere secuenciar los

fragmentos amplificadas (Shaw et al., 1999), no obstante esto no fue realizado, por lo que los loci no fueron usados para los análisis posteriores.

Un locus que presenta alta frecuencia de alelos nulos puede causar subestimación en la heterocigocidad esperada y sobrestimación de  $F_{ST}$  y  $D_C$  (Chapuis y Estoup, 2007; Carlsson, 2008). Sgum05 presentó una alta prevalencia entre localidades y una alta frecuencia de alelos nulos (20%–36%), como los análisis preliminares fueron muy similares al tomarlo y al no tomarlo en cuenta, se optó por excluirlo de los análisis subsecuentes. Sgum36 si fue considerado en el resto de análisis a pesar de presentar 5 localidades con estimaciones de alelos nulos con proporciones cercanas a 25%, pues las 9 localidades restantes mostraron valores moderados.

Varios de los modelos de análisis poblacional utilizados en este estudio asumen que las poblaciones en cuestión se encuentran en Equilibrio Hardy-Weinberg (EHW) (a excepción del AFC, el ADCP, el dendograma de distancia genéticas y el gráfico poblacional). Encontrar desviación del EHW se vuelve trascendentes no sólo por estos modelos que lo asumen, sino porque es posible identificar algunas de las fuerzas evolutivas que lo producen, siempre y cuando se consideren los mismos datos genéticos y otros datos biológicos. De las estimaciones con desviación significativa del EHW por locus para cada localidad (Tabla XI), ninguno podría ser atribuido a la segregación, pues todos los pares de loci tuvieron valores de  $p$  mayores a 0.09 para la prueba de desequilibrio de ligamiento (Tabla X). De los 11 casos en desviación, 8 correspondieron a las estimaciones de alelos nulos más altas (22%–29%), 5 del locus Sgum36 y una de Sgum12, Sgum25 y Sgum29 respectivamente, por lo tanto su cuenta de homocigotos debe estar sobreestimada alejando artificialmente la estimación de  $H_O$  y  $H_E$ . La localidad de El Cien (L03) contó con el coeficiente de endogamia más alto ( $F_{IS} = 0.4488$ ) y dos de sus loci (Sgum25 y Sgum36) con 29% y 24% de estimación de alelos nulos, a lo que se podría atribuir el desequilibrio.

Sin embargo, al analizar cada localidad utilizando la información de los 6 loci (Tabla XII) podemos encontrar que 10 localidades muestran un déficit de heterocigotos significativo. Una hipótesis alternativa para explicar este déficit, es considerar la reproducción vegetativa de esta especie; en efecto los coeficientes de endogamia ( $F_{IS}$ ) siempre fueron mayores a 0. Entre más agregada sea la arquitectura clonal en una planta de polinización mediada por insectos, más aumenta la proporción de geitonogamia (polinización entre flores del mismo individuo genético), incluso en plantas autoincompatibles como *S. gummosus* (Charpentier, 2002). Con la intención de discernir si la clonalidad pudo haber influido en los valores  $H_O$  y  $F_{IS}$ , durante la estandarización de

los marcadores se realizaron estimaciones de genotipos multilocus idénticos y su probabilidad de ser clones, para la localidad de Bahía Sargento, Sonora (L41), para la cual se lograron obtener los genotipos completos de 9 loci polimórficos; tras comparar los genotipos de 38 individuos separados por más de 30 metros, se encontraron 33 individuos genéticos distintos, indicando que en esa localidad la dispersión de rametos puede llegar al menos hasta 30-35 m. Estos resultados son congruentes con la explicación de Parker y Hamrick (1992) sobre la arquitectura clonal de *Lophocereus schottii*, una cactácea columnar codistribuida con *S. gummosus*, que aunque es agregada cuenta con cierto grado de dispersión de los rametos atribuida a arroyos intermitentes. De tal forma que para las poblaciones de *S. gummosus*, el impacto de la endogamia por geintonogamia asociada a la arquitectura clonal no habría de ser tan elevado.

Otra explicación alternativa sobre los niveles de  $F_{IS}$  encontrados y la desviación del EHW en 10 localidades podría ser la selección a favor de los homocigotos. Esto podría ser una señal resultante de un proceso de colonización reciente con eventos fundaciones seguidos por rápido crecimiento poblacional (Nei et al., 1975), como el que tendría que haber sucedido a partir de un único refugio pleistocénico, o a partir de múltiples refugios locales a lo largo de la Península donde la reproducción vegetativa tendría mayor éxito ante el estrés ambiental, particularmente de la temperatura y la humedad (Carrillo-Angeles y Del Carmen Mandujano, 2011), que implicarían menos eventos fundacionales necesarios para alcanzar la actual distribución, permitiendo la persistencia de la diversidad genética pero aumentando el coeficiente de endogamia.

## **8.2. Estructuración genética, expansión poblacional y colonización sur-norte**

En el Análisis Factorial de Correspondencia (AFC), todas las localidades quedaron traslapadas en el espacio tridimensional de las variables artificiales (Figura 8a) explicando el 60% de la varianza, pero también se pudo detectar una aparente organización regional, la cual fue identificable con el Análisis Discriminante de Componentes Principales (ADCP), que a pesar de asignar probabilidades de pertenencia similares a cada grupo para todas las muestras, reveló que las localidades del centro y del norte de la PBC quedaron completamente mezcladas al graficar las cargas de las variables artificiales, por lo que estas dos regiones se tomaron como un solo grupo, mientras que las localidades del grupo sur y las no peninsulares convergen con el centroide tienden a organizarse hacia distintas zonas en el área conformada por los ejes 1 y 2 (Figura 8b), permitiendo así detectar una tendencia regional en la diferenciación entre las muestras analizadas.



Si la estructura genética poblacional de *S. gummosus* es resultado de una expansión poblacional, es decir donde las limitaciones al flujo genético serían principalmente determinadas por la capacidad de dispersión de sus propágulos sexuales, la tendencia de organización regional observada no sería el resultado de estructuración abrupta debido a un evento vicariante o una barrera ecológica y, para el caso de procesos selectivos locales generalizados es útil aprovechar medidas de distancia genética como  $D_C$  (Cavalli-Sforza y Edwards, 1967) que resulta menos sensible a procesos demográficos como cuellos de botella (Takezaki y Nei, 1996). Los resultados de AMOVA evaluando los 8 modelos, donde 5 de los cuales fueron explícitamente planteados a partir de las hipótesis de eventos viacariantes que fueron previamente sugeridos para explicar la arquitectura genética de animales codistribuidos en la Baja California (Riddle et al., 2000), mostraron en todos los casos la mayoría de los componentes de la varianza atribuida a la diferencia hacia adentro de las subpoblaciones ( $\geq 93.7\%$ ) y, el índice de fijación más bajo ( $F_{ST} = 0.0485$ ) se encontró cuando se compararon todas las localidades como grupos independientes, contando además con el mayor porcentaje de varianza hacia adentro de las subpoblaciones (95.38%). Esto podría indicar que la diversidad genética se encuentra heterogéneamente distribuida a lo largo de toda la distribución. Aun así, todos los modelos presentados tuvieron índices significativos, pero explicaron muy pocos componentes de la varianza al descomponerla en los grupos en cuestión. El modelo sobre El Canal transpeninsular del Vizcaíno como evento vicariante, fue el que logró la mejor combinación de porcentaje entre poblaciones dentro de grupos (3.21%) y entre grupos (3.03%), reduciendo el explicado por la varianza hacia adentro de las subpoblaciones al nivel más bajo detectado de 93.76%. Las diferencias genéticas entre los grupos analizados existen, pero son muy pequeñas y podrían deberse a limitaciones muy recientes o incluso contemporáneas al flujo genético, tanto de índole fisiográfico como ambiental o ecológico, por lo que deberían de analizarse por separado. El uso de marcadores mitocondriales y/o cloroplásticos podrían ayudar a dilucidar estas señales de estructuración genética con mejor resolución, pensando en factores geológicos más antiguos, como fue encontrado para *E. lomelii* precisamente a mitad de la PBC, concordante con el canal hace 1 M años en el Vizcaíno (Garrick et al., 2009).

En última instancia, los modelos sometidos a AMOVA, son muy pocos para el número de combinaciones diferentes de grupos que podrían realizarse a partir de la misma información genética, pero hacerlo sería un trabajo muy exhaustivo, complejo de sintetizar y que correría con el riesgo de no contar con fundamentación biológica para muchos de los posibles modelos de agrupación (Dyer y Nason, 2004). Por lo que se emplearon otros métodos que exploran las

relaciones entre las muestras incorporadas a partir de los datos genotípicos, alélicos y las relaciones geográficas para detectar modelos de estructuración o diferenciación.

A diferencia de los datos aloenzimáticos (Clark-Tapia y Molina-Freaner, 2003), los 6 loci microsatélite analizados permitieron detectar un patrón de aislamiento por distancia al evaluar la  $D_C$  de Cavalli-Sforza y Edwards (1967) con la distancia geográfica (Figura 10). A reserva de detectar otros señales de estructuración entre poblaciones atribuibles a barreras limitantes del flujo genético y no a la limitación de la dispersión (Meirmans, 2012), este patrón es concordante con la estructura genética de otras especies sujetas a historias ambientales similares (Hewitt, 1996) en la misma región del Desierto Sonorense, como *Lophocereus schottii*, cactácea columnar, y la euforbiácea *Euphorbia lomelii* (Nason et al., 2002; Garrick et al., 2009), que al igual que muchas plantas han sufrido cambios recientes de su distribución extendiendo sus límites en respuesta a los efectos de los cambios climáticos ocurridos en las épocas glacial e interglacial (Hewitt, 2000; Petit et al., 2001a).

Considerando las condiciones tropicales de la región de Los Cabos, al sur de la PBC (González-Abraham et al., 2010) podemos explicar que la riqueza alélica ( $R$ ) sea mayor en las localidades del sur (Figura 7b), lo que es una evidencia concordante con la tendencia de expansión y desplazamiento de las comunidades vegetales como respuesta al calentamiento postglacial descrita por Hewitt (1996). Sin embargo, considerar que la actual distribución se extendió únicamente desde esta región, implicaría necesariamente un proceso de colonización mucho más veloz que si hubieran existido múltiples refugios locales a lo largo de la PBC. En consecuencia, entre más rápido se haya logrado la colonización, mayor sería el número de eventos de dispersión a larga distancia y, conforme estos aumenten, la diversidad genética resulta homogenizada por *efecto de reajuste* (Bialozyt et al., 2006). Esto no fue lo observado con la  $R$ , por el contrario, se encontró una disminución significativa con el aumento de la latitud a lo largo de la Península (Figura 7b). Estos resultados apoyan la noción de la direccionalidad sur-norte de la expansión, concordante con lo encontrado por Nason y colaboradores en *L. schottii* (2002). Pero la señal de diferenciación medida en  $F_{ST}$  pareados de Punta Prieta (L22) con respecto al resto de localidades norteñas (Tabla XI), la agrupación de Puente Querétaro (L04) con localidades distantes y separadas por el Golfo en el cladograma de distancias genéticas y en el gráfico poblacional (L22, L36 y L41 en Figura 11 y Figura 12) y la noción de selección de homocigotos en El Cien (L03) con el más alto índice de endogamia ( $F_{IS} = 0.449$ , Tabla XII), refuerzan la hipótesis sobre la existencia de múltiples refugios a lo largo de la PBC con tamaño efectivo pequeño. A diferencia de *L. schottii* (Nason et al., 2002) S.

*gummosus* mostró valores de  $H_E$  relativamente constantes a lo largo de la Península, lo cual puede explicarse si durante el proceso de expansión poblacional se encontraron distintas rutas de colonización (Konnert y Bergmann, 1995; Petit et al., 2001b), lo cual también concuerda con un proceso de colonización a partir de múltiples refugios. La evidencia palinológica indica la existencia de semidesiertos locales durante la última glaciación, a bajas latitudes en los actuales desiertos de Norteamérica (Van Devender et al., 1994; Thompson y Anderson, 2000); estas áreas pudieron funcionar como refugios que permitirían la expansión poblacional de *S. gummosus* a partir de poblaciones no necesariamente sureñas y con tamaños de población efectiva ( $N_e$ ) reducidos por las condiciones ambientales. La señal genética que debieran dejar estos eventos múltiples de expansión local –por efecto de la deriva génica previa a la expansión poblacional–, podrían atenuarse con los eventos de dispersión a larga distancia provenientes de otros refugios, donde las tallas de población hayan sido más grandes, a saber de aquellos que se encontraban en latitudes más sureñas con condiciones más óptimas (Hewitt, 1996).

Un modelo así, con diversos refugios locales dispersos por la PBC, soporta la hipótesis de diversificación de *Stenocereus eruca* a partir de *S. gummosus* (Gibson, 1989) como respuesta selectiva local de fenotipos exitosos con un aumento en la tasa de reproducción clonal, que le permitiera diferenciarse reciente y rápidamente; la distribución de *S. eruca* se encuentra cercana a las localidades de El Cien (L03) que reportó el coeficiente de endogamia más alto ( $F_{IS} = 0.449$ ) y a Puente Querétaro (L04) que tuvo el segundo valor más bajo de heterocigocidad esperada y en el gráfico poblacional es representado por el nodo más pequeño junto con Santa Águeda (L20).

El cladograma de distancias genéticas mostró las localidades analizadas en tres grupos principales (Figura 11) que no concuerda con ninguno de los escenarios de vicarianza analizados. Por el contrario, conjunta localidades que habrían sido separadas por los canales en la Península y por la formación del Golfo de California, rechazando la idea de diferenciación genética debido a esos factores de barreras físicas en el pasado lejano, al menos con los loci microsatélite estudiados. Por otra parte, que la topología del árbol no haya exhibido un patrón de clados anidados en serie desde el sur hasta el norte de la Península, puede sugerir que una colonización postglacial (Holocénica) debió llevarse a cabo en varias direcciones. Al igual que con los análisis multivariados (AFC y DACP, Figura 8) pero con mejor resolución, se observó una agrupación regional, en la que las localidades del sur quedan en una rama junto con el extremo sur, después de la mitad de la PBC surgen dos ramas, una agrupando a la localidad de El Cuarenta (L26, centro-oeste de la PBC) con el grupo peninsular norte y otra que incluye a Puente Querétaro (L04, centro de la PBC) con Punta

Prieta (L22, norte de la PBC) y las dos localidades no peninsulares, estando ubicadas las tres últimas en latitudes muy cercanas. La ubicación de Puente Querétaro (L04) en un clado distinto a sus las localidades más próximas, podría indicar un proceso selectivo particular en respuesta a características ambientales propias del sitio como el sustrato rocoso (observación de campo). La ubicación de la localidad insular y la continental, podría estar estrechamente relacionado con su origen, reforzando la hipótesis de Cody (1983) sobre el origen peninsular de las poblaciones sonorenses. Debido a que la medición de distancia genética empleada no asume ningún modelo mutacional subyacente, ni hace inferencias evolutivas entre los alelos y sus tallas, no es posible establecer el tiempo transcurrido desde cada divergencia descrita por la topología: estas relaciones son sólo descriptivas del estado, no de la historia. El bajo valor de bootstrap para la ubicación de la Santa Águeda (L20) en el árbol de distancias (Figura 11), queda resuelto en el gráfico poblacional (Figura 12) quedando sólo conectada con Puente Querétaro (L04, centro peninsular), concordante con el patrón de aislamiento por distancia explicando las relaciones a lo largo del eje latitudinal de la PBC.

El análisis bayesiano, buscando el número más probable de poblaciones, no detecto una estructura genética de grupos homogéneos estructurados entre las localidades, más bien se resolvió con amplia cantidad de mezcla genética independientemente de si los datos se agrupan en 2 o 5 grupos (Figura 14a y b). Cuando los valores de  $F_{ST}$  son menores a 0.2, STRUCTURE muestra dificultades para la identificación de estructura, además de que el modelo asumido por el software no es adecuado para datos con patrones de aislamiento por distancia (Pritchard et al., 2007).

En general, las diferencias genéticas significativas encontradas de  $F_{st}$  por pares de localidades (Tabla XIV,  $\alpha = 0.05$ ) describen la agrupación en las 3 ramas del árbol de distancias genéticas (Figura 11) y la topología obtenida para el gráfico poblacional (Figura 12), que en análisis conjunto coinciden también con un patrón de relaciones sur–centro–norte en la Península y Península centro–Sonora. Entre las 4 localidades nortenas de la PBC, Punta Prieta (L22) se diferencia de todo este grupo ( $F_{st} = 0.06, 0.063$  y  $0.16; p \leq 0.049$ ), siempre mostró menor diferencia a otras localidades (en particular Sonora e Isla Tiburón). Esta diferencia local dentro de una región particular, tiene coherencia con la existencia de refugios locales de tamaño efectivo pequeño y aumento poblacional con heterocigocidad disminuida por cuello de botella (mostró la menor  $H_O = 0.159$ ) (Nei et al., 1975; Petit et al., 2001a).

Las débil señal de diferenciación genética observada para la separación de grupos debida a la presencia del Golfo de California (Figura 9) y que el gráfico poblacional no desplegara una sub-estructuración de redes desconectadas entre las poblaciones de la PBC y las no peninsulares en ninguno de los casos (ANEXO III), hace pensar que el Golfo no ha representado una barrera al flujo genético; esto es concordante con las conclusiones de Garrick y colaboradores (2009) que atribuyen la conectividad genética entre estas dos regiones a una introducción no necesariamente reiterativa, argumentando la posibilidad de dispersión por aves a través del Golfo. De cualquier forma, es importante considerar que sólo se incorporó una localidad insular y una de continental contra 12 peninsulares, por lo que sería recomendable incluir más información de esos sitios.

Por otra parte, es importante considerar la posibilidad de que las semillas o incluso ramas de la pitaya agria hayan sido transportadas por los Seris entre la Península, las islas del Golfo y Sonora, pues se ha documentado el cultivo de pitaya agria a partir ramas en jardines caseros de esta etnia y particularmente una alta movilidad entre Sonora y la Península de este grupo, en donde estos frutos formarían parte de la subsistencia durante los viajes (Felger y Moser, 1985), similar a lo que sucede con otras especies de *Stenocereus* en el centro de México (Casas et al., 1999a), además del uso que le dan a la corteza en técnicas de pesca (Anderson et al., 2001). Las localidades conocidas con presencia de *S. gummosus* en vida silvestre en Sonora, se ubican alrededor y en medio de Punta Chueca y Desemboque, las dos comunidades más grandes habitadas por los actuales Seris. Las Islas Tiburón y San Esteban, estuvieron ocupadas por agrupaciones de los Seris hasta el primer tercio del siglo XX y se ha inferido además la presencia de un subgrupo en la Península (Felger y Moser, 1985). Sin embargo, se reconoce que los Seris no se han caracterizado por desarrollo de técnicas agrícolas.

## 9. CONCLUSIONES

a. El polimorfismo de los loci microsatélite reportados en este estudio fue bajo (número de alelos por locus), sin embargo fue comparable a lo reportado en otros organismos relacionados con *S. gummosus*; estos niveles son suficientes para indagar sobre procesos recientes en la historia evolutiva de esta cactácea.

b. Aunque es probable que la capacidad de reproducción clonal aporta importantes elementos para la estructuración poblacional, sobre todo a nivel local o intrapoblacional (coeficiente de endogamia), la presente investigación no tiene los elementos suficientes para determinar los efectos de esta propiedad biológica. Será necesario un esquema de muestreo diferente para analizar los efectos de la capacidad de reproducción clonal a la distribución de la diversidad genética intra e interpoblacional, inclusive aumentar el número de muestras genotipificadas por localidad y el número de loci registrados por muestra.

c. La información biológica conocida sobre genética poblacional de otras plantas de la PBC, la capacidad de dispersión de semillas en otras especies del mismo género y en la propia especie y el patrón generalizado de colonizaciones postglaciales de plantas, explican satisfactoriamente la detección de baja diferenciación y el aislamiento por distancia.

d. La disminución de la riqueza alélica con el aumento de la latitud, el aislamiento por distancia y la distancia genética descrita por clados anidados, confirman la importancia del proceso de expansión poblacional post-glacial, como modelador de la estructura genética de *Stenocereus gummosus*.

e. Para reforzar la hipótesis de múltiples refugios pleistocénicos, será de utilidad incluir a *Stenocereus eruca* en el análisis de distancias genéticas y en el resto de análisis exploratorios de la

estructura poblacional, además de analizar el mismo escenario en más organismos codistribuidos, especialmente polinizadores, dispersores y en otras plantas.

*f.* La conectividad de las localidades no peninsulares con algunos sitios en particular de la Península de Baja California, no permiten desechar la hipótesis sobre que existan causas antropogénicas implicadas en la dispersión de la pitaya agria hacia algunas de las islas del Golfo y en Sonora.

## 10. LITERATURA CITADA

- Abdelkrim, J., B. Robertson, J. A. Stanton, y N. Gemmell . 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*. 46(3): 185–92.
- Anderson, E.F., W. Barthlott, y R. Brown . 2001. *The Cactus Family*. Timber Press, 776p.
- Arnaud-Haond, S., F. Alberto, S. Teixeira, G. Procaccini, E.A. Serrão, y C.M. Duarte . 2005. Assessing genetic diversity in clonal organisms: low diversity or low resolution? Combining power and cost efficiency in selecting markers. *The Journal of Heredity*. 96(4): 434–40.
- Arnaud-Haond, S., C.M. Duarte, F. Alberto, y E. a Serrão . 2007. Standardizing methods to address clonality in population studies. *Molecular Ecology*. 16(24): 5115–5139.
- Austerlitz, F., B. Jung-Muller, B. Godelle, y P. H. Gouyon . 1997. Evolution of coalescence times , genetic diversity and structure during colonization. *Theoretical Population Biology*. 51(2): 148–164.
- Bassam, B.J. y G. Caetano-Anollés . 1993. Silver staining of DNA in polyacrylamide gels. *Applied Biochemistry and Biotechnology*. 42(2-3): 181–188.
- Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, y F. Bonhomme . 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions. Université Montpellier II.
- Benjamini, Y. y Y. Hochberg . 1995. Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. 57(1): 289–300.
- Bialozyt, R., B. Ziegenhagen, y R.J. Petit . 2006. Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*. 19(1): 12–20.
- Brown, A. y B. Weir . 1983. Measuring genetic variability in plant populations. En: Tanksley S. y Orton T. (eds.) *Isozymes in Plant Genetics and Breeding*. Elsevier Science Publications, Amsterdam. 219–239p.
- Carlsson, J. 2008. Effects of microsatellite null alleles on assignment testing. *The Journal of Heredity*. 99(6): 616–623.
- Carreño, A.L. y J. Helenes . 2002. Geology and ages of the Islands. En: Case T.J. et al. (eds.) *A new Island Biogeography of the Sea of Cortés*. Oxford University Press, New York. 14–40p.
- Carrillo-Angeles, G.I. y M. del Carmen Mandujano . 2011. Patrones de distribución espacial en plantas clonales. *Boletín de la Sociedad Botánica de México*. 89: 1–18.
- Carson, H.L. 1990. Increased genetic variance after a population bottleneck. *Trends in Ecology & Evolution*. 5(7): 228–230.
- Casas, A., N. Alitgioma, J. Caballero, A. Posta, y A.V. Banuet . 1999a. Use, management and domestication of columnar cacti in south-central Mexico: a historical perspective. *Journal of Ethnobiology*. 19(1): 71–95.



- Casas, A., J. Cruse-Sanders, E. Morales, A. Otero-Arnaiz, y A. Valiente-Banuet . 2005. Maintenance of phenotypic and genotypic diversity in managed populations of *Stenocereus Stellatus* (Cactaceae) by indigenous peoples in central Mexico. *Biodiversity and Conservation*. 15(3): 879–898.
- Casas, A., A. Valiente-Banuet, A. Rojas-Martínez, y P. Dávila . 1999b. Reproduction biology and the process of domestication of the columnar cactus *Stenocereus stellatus* in Central Mexico. *American Journal of Botany*. 86(4): 534–542.
- Cavalli-Sforza, L.L. y S. V. Edwards . 1967. Phylogenetic analysis: models and estimation procedures. *Evolution*. 21(3): 550–570.
- Chapuis, M.-P. y A. Estoup . 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*. 24(3): 621–31.
- Charpentier, A. 2002. Consequences of clonal growth for plant mating. *Evolutionary Ecology*. 15: 521–530.
- Clark-Tapia, R. y F. Molina-Freaner . 2003. The genetic structure of a columnar cactus with a disjunct distribution: *Stenocereus gummosus* in the Sonoran desert. *Heredity*. 90(6): 443–450.
- Cody, M.L., R. Moran, y H. Thompson . 1983. The plants. En: Case T.J. y Cody M.L. (eds.) *Island Biogeography in the Sea of Cortez*. University of California Press, Berkeley, CA. 49–97p.
- Del Coro Arizmendi, M., A. Valiente-Banuet, A. Rojas-Martínez, y P. Dávila-Aranda . 2002. Columnar cacti and the diets of nectar-feeding bats. En: Fleming T.H. y Valiente-Banuet A. (eds.) *Columnar cacti and their mutualists: evolution, ecology, and conservation*. University Of Arizona Press., Tucson, Arizona. 264–281p.
- David, P. 1998. Heterozygosity–fitness correlations: new perspectives on old problems. *Heredity*. 80(5): 531–537.
- Dempster, A.P., N.M. Laird, y D.B. Rubin . 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*. 39(1): 1–38.
- Van Devender, T.R. 2002. Environmental history of Sonoran Desert. En: Fleming T.H. y Valiente-Banuet A. (eds.) *Columnar cacti and their mutualists: evolution, ecology, and conservation*. University of Arizona Press., Tucson, Arizona. 3–24p.
- Van Devender, T.R., T.L. Burgess, J.C. Piper, y R.M. Turner . 1994. Paleoclimatic Implications of Holocene Plant Remains from the Sierra Bacha, Sonora, Mexico. *Quaternary Research*. 41(1): 99–108.
- Dressler, R.L. 1957. The genus *Pedilanthus*. *Contributions of the Gray Herbarium*. 182: 1–188.
- Dyer, R.J. 2009. GeneticStudio: a suite of programs for spatial analysis of genetic-marker data. *Molecular Ecology Resources*. 9(1): 110–103.
- Dyer, R.J. y J.D. Nason . 2004. Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*. 13(7): 1713–1727.

- Earl, D.A. y B.M. vonHoldt . 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. 4(2): 359–361.
- Ersts, P.J. 2006. Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation.
- Estoup, A., P. Jarne, y J.-M. Cornuet . 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*. 11(9): 1591–1604.
- Evanno, G., S. Regnaut, y J. Goudet . 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 14(8): 2611–2620.
- Excoffier, L. y H.E.L. Lischer . 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.
- Excoffier, L., P.E. Smouse, y J.M. Quattro . 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. 131(2): 479–491.
- Faircloth, B.C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*. 8(1): 92–94.
- Felger, R.S. y M.B. Moser . 1985. *People of the desert and sea: ethnobotany of the Seri Indians*. University Of Arizona Press., Tucson, Arizona. 435p.
- Felsenstein, J. 2005. PHYLIP (Phylogeny Infering Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle.
- Freeland, J.R. 2006. *Molecular Ecology*. John Wiley & Sons Ltd, West Sussex, England. 400p.
- Garrick, R.C., J.D. Nason, C.A. Meadows, y R.J. Dyer . 2009. Not just vicariance: phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja Peninsula. *Molecular Ecology*. 18(9):1916–1931.
- Gibson, A.C. 1989. The systematics and evolution of the subtribe *Stenocereinae*. The macharocerei of *Stenocereus*. *Cactus and Succulent Journal*. 61: 104–112.
- Gibson, A.C. y K.E. Horak . 1978. Systematic anatomy and phylogeny of mexican columnar cacti. *Annals of the Missouri Botanical Garden*. 65: 999–1057.
- Gibson, A.C. y P.S. Nobel . 1986. *The cactus primer*. Harvard University Press, Cambridge, Massachusetts. 286p.
- González-Abraham, C.E., P. Peña-Garcillán, y E. Ezcurra . 2010. Ecorregiones de la Península de Baja California: una síntesis. *Boletín de la Sociedad Botánica de México*. 87: 69–82.
- Guichoux, E., L. Lagache, S. Wagner, P. Chaumeil, P. Léger, O. Lepais, C. Lepoittevin, T. Malausa, E. Revardel, F. Salin, y R.J. Petit . 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources*. 11(4): 591–611.

- Hardesty, B.D., S.L. Hughes, V.M. Rodriguez, y J.A. Hawkins . 2008. Characterization of microsatellite loci for the endangered cactus *Echinocactus grusonii*, and their cross-species utilization. *Molecular Ecology Resources*. 8(1): 164–167.
- Harr, B. y C. Schlötterer . 2004. Patterns of microsatellite variability in the *Drosophila melanogaster* complex. *Genetica*. 120(1-3): 71–77.
- Hartl, D.L. y A.G. Clark . 2007. Principles of population genetics. Sinauer, Sunderland, MA. 652p.
- Hewitt, G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*. 58(3): 247–276.
- Hewitt, G.M. 2000. The genetic legacy of the Quaternary ice ages. *Nature*. 405(6789): 907–913.
- Holland, J.N. y T.H. Fleming . 1998. The evolution of obligate pollination mutualisms: senita cactus and senita moth. *Oecologia*. 114: 368–375.
- Hughes, S.L., V.M. Rodriguez, B.D. Hardesty, R.T. Bárcenas Luna, H.M. Hernández, R.M. Robson, y J.A. Hawkins . 2008. Characterization of microsatellite loci for the critically endangered cactus *Ariocarpus bravoanus*. *Molecular Ecology Resources*. 8(5): 1068–1070.
- Huson, D.H., D.C. Richter, C. Rausch, T. DeZulian, M. Franz, y R. Rupp . 2007. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics*. 8(1): 460.
- Jacobs, D.K., T. A. Haney, y K.D. Louie . 2004. Genes, diversity, and geologic process on the Pacific Coast. *Annual Review of Earth and Planetary Sciences*. 32(1): 601–652.
- Jensen, J.L., A.J. Bohonak, y S.T. Kelley . 2005. Isolation by distance, web service. *BMC Genetics*. 6(1): 13.
- Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*. 24(11): 1403–1405.
- Jombart, T., S. Devillard, y F. Balloux . 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*. 11(1): 94.
- Kalinowski, S.T. 2005. hp-rare 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*. 5(1): 187–189.
- Konnert, M. y F. Bergmann . 1995. The geographical distribution of genetic variation of silver fir (*Abies alba*, Pinaceae) in relation to its migration history. *Plant Systematics and Evolution*. 196(1-2): 19–30.
- De la Cruz, M., F. Ramírez, y H. Hernández . 1997. DNA Isolation and amplification from cacti. *Plant Molecular Biology Reporter*. 15(4): 319–325.
- Lai, Y. y F. Sun . 2003. The relationship between microsatellite slippage mutation rate and the number of repeat units. *Molecular Biology and Evolution*. 20(12): 2123–2131.
- León de La Luz, J.L. y R. Domínguez-Cadena . 1991. Evaluación de la reproducción por semilla de la pitahaya agria (*Stenocereus gummosus*) en Baja California Sur, México. *Acta Botánica Mexicana*. 14: 75–87.

- León de la Luz, J.L., R. Domínguez-Cadena, M. Cruz-Estrada, y R. Rodríguez-Estrella . 1995. Reproductive phenology of *Stenocereus gummosus* (Engelm.) Gibson & Horak. Implications for its cultivation. *Genetic Resources and Crop Evolution*. 42(1): 61–67.
- Li, Y.-C., A.B. Korol, T. Fahima, A. Beiles, y E. Nevo . 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*. 11(12): 2453–2465.
- Lindell, J., A. Ngo, y R.W. Murphy . 2006. Deep genealogies and the mid-peninsular seaway of Baja California. *Journal of Biogeography*. 33: 1327–1331.
- Magwene, P.M. . 2001. New tools for studying integration and modularity. *Evolution*. 55(9): 1734–1745.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*. 27: 209–220.
- Megléczy, E., C. Costedoat, V. Dubut, A. Gilles, T. Malausa, N. Pech, y J.-F. Martin . 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*. 26(3): 403–404.
- Meirmans, P.G. 2012. The trouble with isolation by distance. *Molecular Ecology*. 21(12): 2839–2846.
- Metzgar, D., J. Bytof, y C. Wills . 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research* 10(1): 72–80.
- Molina Freaner, F. y R. Clark Tapia . 2005. Clonal diversity and allelic relationships between two closely related species of columnar cacti from the Sonoran Desert: *Stenocereus eruca* and *S. gummosus*. *International Journal of Plant Science* 166(2): 257–264.
- Murphy, R.W. y G. Aguirre-León . 2002. Nonavian reptiles; origins and evolution. En: Case, T.J., Cody, M.L., Ezcurra E. (eds.) *A new island biogeography of the Sea of Cortés*. Oxford University Press, New York, NY. 181–220p.
- Nason, J.D., J.L. Hamrick, y T.H. Fleming . 2002. Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. *Evolution*. 56(11): 2214–26.
- Nassar, J.M., J.L. Hamrick, y T.H. Fleming . 2001. Genetic variation and population structure of the mixed-mating cactus, *Melocactus curvispinus* (Cactaceae). *Heredity*. 87(Pt 1): 69–79.
- Nassar, J.M., N. Ramirez, y O. Linares . 1997. Comparative pollination biology of Venezuelan columnar cacti and the role of nectar-feeding bats in their sexual reproduction. *American Journal of Botany*. 84(7): 918–927.
- Nei, M., T. Maruyama, y R. Chakraborty . 1975. The bottleneck effect and the genetic variability in populations. *Evolution*. 29(1): 1–10.
- Otero-Arnaiz, A., A. Casas, J.L. Hamrick, y J. Cruse-Sanders . 2005. Genetic variation and evolution of *Polaskia chichipe* (Cactaceae) under domestication in the Tehuacán Valley, central Mexico. *Molecular Ecology*. 14(6):1603–11.

- Parker, K.C. y J.L. Hamrick . 1992. Genetic diversity and clonal structure in a columnar cactus, *Lophocereus schottii*. *American Journal of Botany*. 79(1): 86–96.
- Parra, F., A. Casas, J.M. Peñaloza-Ramírez, A.C. Cortés-Palomec, V. Rocha-Ramírez, y A. González-Rodríguez . 2010. Evolution under domestication: ongoing artificial selection and divergence of wild and managed *Stenocereus pruinosus* (Cactaceae) populations in the Tehuacan Valley, Mexico. *Annals of Botany*. 106(3): 483–96.
- Parra, F., N. Pérez-Nasser, R. Lira, D. Pérez-Salicrup, y A. Casas . 2008. Population genetics and process of domestication of *Stenocereus pruinosus* (Cactaceae) in the Tehuacán Valley, México. *Journal of Arid Environments*. 72(11): 1997–2010.
- Peakall, R. y P.E. Smouse . 2012. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*. 28(19): 2537–2539.
- Pérez, M.F., M.F. Téó, D.C. Zappi, N.P. Taylor, y E.M. Moraes . 2011. Isolation, characterization, and cross-species amplification of polymorphic microsatellite markers for *Pilosocereus machrisii* (Cactaceae). *American Journal of Botany*. 98(8): e204–e206.
- Petit, R. J., J. Jouzel, D. Raynaud, N. I. Barkov, J.-M. Barnola, I. Basile, M. Bender, J. happellaz, M. Davis, G. Delaygue, M. Delmotte, V. M. Kotlyakov, M. Legrand, V. Y. Lipenkov, C. Lorius, L. Pépin, C. Ritz, E. Saltzman y M. Stievenard . 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*. 399(6735): 429–436.
- Petit, R.J., R. Bialozyt, S. Brewer, R. Cheddadi, y B. Comps . 2001a. From spatial patterns of genetic diversity to postglacial migration processes in forest trees. En: Silvertown J. y Antonovics J. (eds.) *Integrating ecology and evolution in a spatial context*. Blackwell Science Ltd., Oxford, 295–318p.
- Petit, R.J., R. Bialozyt, P. Garnier-Gere, y A. Hampe . 2004. Ecology and genetics of tree invasions: from recent introductions to Quaternary migrations. *Forest Ecology and Management*. 197: 117–137.
- Petit, R.J., U.M. Csaikl, J. Cottrell, B. Van Dam, J.D. Deans, S. Dumolin-lape, S. Fineschi, R. Finkeldey, A. Gillies, I. Glaz, A.J. Lowe, P.G. Goicoechea, J.S. Jensen, A.O. Ko, R.C. Munro, M. Olalde, S.F. Madsen, y G. Ma . 2001b. Chloroplast DNA variation in european white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*. 156: 5–26.
- Pritchard, J.K., M. Stephens, y P. Donnelly . 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155(2): 945–959.
- Pritchard, J.K., X. Wen, y D. Falush . 2007. Documentation for STRUCTURE software: Version 2.2.
- R Development Core Team . 2009. R: A language and environment for statistical computation. R Foundation for Statistical Computing, Vienna, Austria.
- Rebman, J.P. y N.C. Roberts . 2012. Baja California Plant Field Guide. San Diego Natural History Museum, San Diego, California, USA. 480p.
- Rice, W.R. 1989. Analysing tables of statistical tests. *Evolution*. 43(1): 223–225.

- Riddle, B.R. y D.J. Hafner . 2006. A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. *Journal of Arid Environments*. 66(3): 435–461.
- Riddle, B.R., D.J. Hafner, L.F. Alexander, y J.R. Jaeger . 2000. Cryptic vicariance in the historical assembly of a Baja California peninsular desert biota. *Proceedings of the National Academy of Sciences*. 97(26): 14438–14443.
- Robichaux, R. 1999. *Ecology of Sonoran Desert plants and plant communities*. University of Arizona Press., Tucson, Arizona. 303p.
- Rousset, F. 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*. 142(4): 1357–1362.
- Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*. 145(4): 1219–1228.
- Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*. 8(1): 103–106.
- Rozen, S. y H. Skaletsky . 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*. 132: 365–386.
- Saitou, N. y M. Nei . 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4): 406–425.
- Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma*. 109(6): 365–371.
- Schlötterer, C. 2004. The evolution of molecular markers - just a matter of fashion?. *Nature Reviews Genetics*. 5(1): 63–69.
- Shaw, P.W., G.J. Pierce, y P.R. Boyle . 1999. Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. *Molecular Ecology*. 8: 407–417.
- Solórzano, S., A.C. Cortés-Palomec, A. Ibarra, P. Dávila, y K. Oyama . 2009. Isolation, characterization and cross-amplification of polymorphic microsatellite loci in the threatened endemic *Mammillaria crucigera* (Cactaceae). *Molecular Ecology Resources*. 9(1): 156–158.
- Takezaki, N. y M. Nei . 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*. 144: 389–399.
- Thompson, R.S. y K.H. Anderson . 2000. Biomes of western North America at 18,000, 6000 and 0 14C yr BP reconstructed from pollen and packrat midden data. *Journal of Biogeography*. 27(3): 555–584.
- Turner, R.M., J.E. Bowers, y T.L. Burgess . 1995. *Sonoran Desert plants: an ecological atlas*. University Of Arizona Press., Tucson, Arizona. 504p.
- Vázquez Sánchez, M., T. Terrazas, y S. Arias . 2012. Habit and growth form in the tribe cactaceae (Cactaceae, Cactoideae). *Botanical Sciences*. 90(2): 97–108.

- Weir, B.S. y C.C. Cockerham . 1984. Estimating F-statistics for the analysis of population structure. *Evolution*. 38(6): 1358–1370.
- Wiggins, I.L. 1980. *Flora of Baja California*. Stanford University Press, Stanford. 1025p.
- Xu, X., M. Peng, y Z. Fang . 2000. The direction of microsatellite mutations is dependent upon allele length. *Nature Genetics*. 24(4): 396–399.

## ANEXO I. EXTRACCIÓN DE ADN PARA *Stenocereus gummosus*

Basado en de la Cruz *et al.*, (1997)

### Tejido

- Corteza interna almacenada en etanol 96°

### Reactivos

- Buffer CTAB de maceración [100 mM tris-HCl pH 8, 20 mM EDTA pH 8, 4% CTAB, 1.5 M NaCl, 4% PVP-40, 2.0% 2-Mercaptoetanol, 5.0% *Ácido ascórbico* y 5.0% *DIECA*]. Los últimos 4 reactivos (en *itálicas*) se agregan en el momento de uso del buffer con las muestras.
- Buffer STE de extracción [100 mM Tris-HCl pH 8, 50 mM EDTA pH 8, 100 mM NaCl, 2.0% 2-Mercaptoetanol]. último reactivo (en *itálicas*) se agrega al momento del uso del buffer con las muestras.
- Solución SDS 20%
- Cloruro de sodio 5 M
- Acetato de sodio 3 M
- Isopropanol frío
- Solución Etanol 80%

### Material

- Tubos Eppendorf 1.5 mL
- Puntas de 1000  $\mu$ L y 200  $\mu$ L
- Pipetas de 200-1000  $\mu$ L, 50-200  $\mu$ L, 20-100  $\mu$ L y 2-20  $\mu$ L
- Guantes

### Equipo

- Incubadora (Termoblock)
- Microcentrífuga
- Campana extracción
- Balanza analítica
- Congelador -4° y -20° C
- Vortex

### Protocolo

\*\* Todos los pasos que impliquen manipulación de soluciones con 2-Mercaptoetanol han de realizarse en la campana de flujo laminar.



**Lisis**

1. Después de presionarlo sobre papel kimwipe, secar ~15mg de tejido de corteza (~2 mm<sup>3</sup>) cortado en 5–6 trozos, en un tubo eppendorf de 1.5 mL al termoblock a 55° C por 1 hora.
2. Una vez seco, agregar 180 µL de buffer CTAB y vortexear (cuidando que no se vierta el contenido, el 2-Mercaptoetanol puede contaminar el área).
3. Agregar 675 µL de buffer STE.
4. Agregar 60 µL de solución SDS 20%.
5. Vortexear de nuevo por 30 segundos e Incubar a 65°C toda la noche (~8 horas).

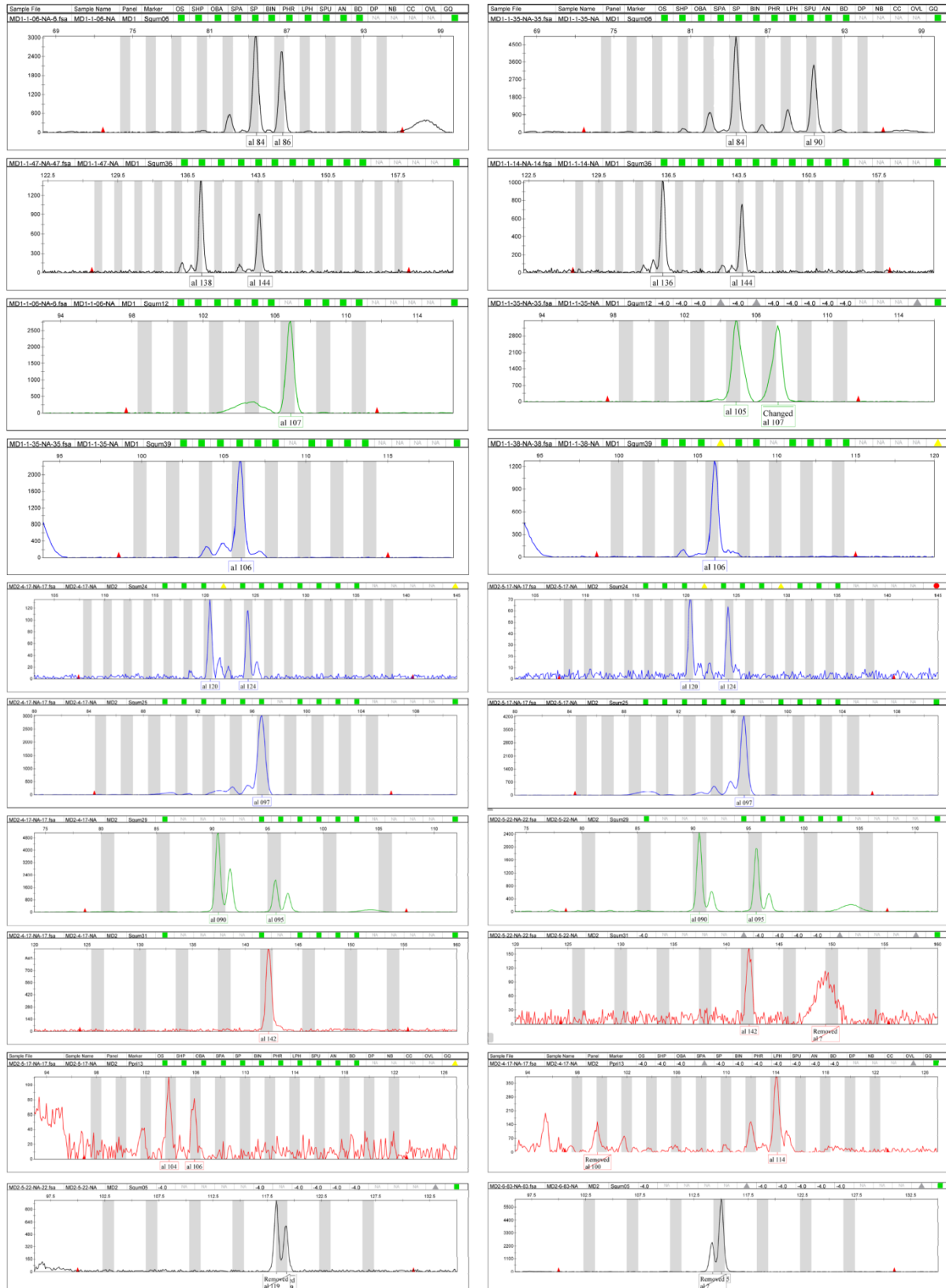
**Precipitación**

6. Agregar 255 µL de NaCl 5M e incubar por 15 minutos a -4° C.
7. Centrifugar a 14 KRPM por 10 minutos.
8. Colectar con una pipeta la fase intermedia en un nuevo tubo de 1.5 mL, dejando los restos del fondo y la capa sobrenadante para posterior eliminación. Colectar al menos 250 µL y a lo mucho 600 µL; esto dependerá de la calidad del tejido inicial y la facilidad para tomar sólo la fase intermedia.
9. Agregar 60 µL de Acetato de sodio 3M en el tubo nuevo que contiene la fracción recién colectada.
10. Agregar 600 µL de isopropanol frío.
11. Agitar suavemente e incubar a -20° C por 3–4 horas para precipitar el ADN genómico. (de preferencia toda la noche).

**Lavado y Resuspensión**

12. Centrifugar a 14 KRPM por 10 minutos.
13. Eliminar el sobrenadante con cuidado de conservar el pellet al fondo del tubo.
14. Agregar 200 µL de etanol 80% y mezclar manualmente por 1 minuto. Dejar a -4°C por 10 min.
15. Centrifugar a 14 KRPM por 5 minutos.
16. Eliminar el etanol del tubo vertiéndolo con cuidado de no perder el pellet.
17. Dependiendo la calidad del pellet se pueden repetir pasos 14 al 16 para eliminar viscosidad.
18. Dejar secar por al menos 60 minutos a temperatura ambiente o a no más de 40° C.
19. Resuspender el pellet en 100 o 150 µL de agua miliQ.

## ANEXO II. EJEMPLOS DE ELECTROFEROGRAMAS CON TAMAÑOS ALÉLICOS ASIGNADOS



### ANEXO III. FORMA TEÓRICA DE LA ESTRUCTURA GENÉTICA (ALFA EN INCREMENTO)

En este grupo de gráficos poblacionales se indica el valor de  $\alpha$  utilizado para la prueba de los vértices. Este cambio sistemático del valor  $\alpha$ , significa ir eliminando vértices que explican la varianza en todo el sistema analizado en comparación con el análisis de  $F_{ST}$  con un  $\alpha$  fijo en 0.05. Los gráficos a), c) y d) fueron reconstruidos con 277 muestras (Tabla VIII) y los 6 loci utilizados a partir de los análisis poblacionales (ver 7.3.2). La diferencia entre a) y b) se debe a la incorporación de las muestras para las que faltaron hasta 2 genotipos, ( $n = 288$ ). Es importante denotar que al incorporar esta información para el gráfico b), desaparece el vértice entre L03 y L30, par de localidades con una de las distancias geográficas más largas entre todos los pares posibles.

