



CENTRO DE INVESTIGACIONES BIOLÓGICAS  
DEL NOROESTE, S.C.

---

---

Programa de Estudios de Posgrado

ESTRUCTURA GENÓMICA NEUTRAL DEL ABULÓN AZUL  
(*Haliotis fulgens*) EN LA COSTA OCCIDENTAL DE LA  
PENÍNSULA DE BAJA CALIFORNIA, MÉXICO.

T E S I S

Que para obtener el grado de

Doctora en Ciencias

Uso, Manejo y Preservación de los Recursos Naturales  
(Orientación en Biología Marina)

P r e s e n t a

**Paulina Mejía Ruiz**

La Paz, Baja California Sur, septiembre de 2022.

## ACTA DE LIBERACIÓN DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 15:00 horas del día 18 del Mes de agosto del 2022, se procedió por los abajo firmantes, miembros de la Comisión Revisora de Tesis avalada por la Dirección de Estudios de Posgrado y Formación de Recursos Humanos del Centro de Investigaciones Biológicas del Noroeste, S. C., a liberar la Tesis de Grado titulada:

"ESTRUCTURA GENÓMICA NEUTRAL DEL ABULÓN AZUL (*Haliotis fulgens*) EN LA COSTA OCCIDENTAL DE LA PENÍNSULA DE BAJA CALIFORNIA, MÉXICO."

Presentada por el alumno:

**Paulina Mejía Ruiz**

Aspirante al Grado de DOCTOR EN CIENCIAS EN EL USO, MANEJO Y PRESERVACIÓN DE LOS RECURSOS NATURALES CON ORIENTACIÓN EN Biología Marina

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

### LA COMISIÓN REVISORA



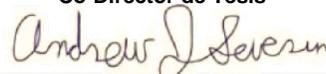
Dr. Fausto Valenzuela Quiñonez  
Co-Director de Tesis



Dr. Ricardo Pérez Enríquez  
Co-Director de Tesis



Dr. Raúl Octavio Martínez Rincón  
Co-Tutor de Tesis



Dr. Andrew J. Severin  
Co-Tutor de Tesis



Dr. John Carlos Garza  
Co-Tutor de Tesis



Dra. Gracia Alicia Gómez Anduro,  
Directora de Estudios de Posgrado y  
Formación de Recursos Humanos

La Paz, Baja California Sur, a 29 de junio de 2022.

Los miembros del comité de tesis de la estudiante Paulina Mejía Ruíz del Programa de Doctorado en Uso Manejo y Preservación de los Recursos Naturales, revisamos el contenido de la tesis y otorgamos el Vo.Bo. dado que la tesis no representa un plagio de otro documento como lo muestra el reporte de similitud realizado:

- Herramienta antiplagio: iThenticate®
- Filtros utilizados:
  - Citas y referencias bibliográficas excluidas
  - Fuentes que sean menores a 15 palabras
  - Coincidencias inferiores a 15 palabras
- Porcentajes de similitud: Se muestra captura de pantalla

tesis doctorado		Índice de similitud																											
Por: Paulina Mejía Ruíz		4%																											
A partir de: 29 Jun 2022 11:06:54																													
36,398 words - 46 matches - 21 sources																													
<a href="#">Incluir citas</a> <a href="#">Incluir bibliografía</a> <a href="#">Excluyendo coincidencias &lt; 15 palabras</a> <a href="#">Limitar el tamaño de las coincidencias a 15 palabras</a>																													
<p>il a la que se distribuyen,</p> <p>as especies sobreexplotadas</p> <p>dades biológicas y las</p> <p>miento y caracterización de</p> <p>un elevado poder para</p> <p>en varios recursos</p> <p>población de origen</p> <p>ulón azul (<i>Haliotis fulgens</i>)</p> <p>mia (PBC). A pesar de las</p> <p>debajo del nivel óptimo de</p> <p>bien establecidas. Por lo</p> <p>enómica neutra de <i>H. fulgens</i></p> <p>o y las unidades biológicas.</p> <p>os de novo mediante un</p> <p>des a lo largo de la PBC</p> <p>ades costeras de la PBC</p> <p>s Norte, Tortugas Sur,</p> <p>por las características</p> <p>tinante hacia el sur desde la</p> <p>fue desarrollar un panel</p> <p>os genéticos detectados</p> <p>de abulón azul con lo cual</p> <p>s: (a) global que incluye los</p>	<table border="1"> <tbody> <tr> <td>1</td> <td>481 words / 2% - Internet de 21-dic-2021 12:00a. m.</td> <td><a href="#">peeri.com</a></td> </tr> <tr> <td>2</td> <td>53 words / &lt; 1% coincidencia - Internet de 14-jun-2022 12:00a. m.</td> <td><a href="#">cibnor.repositorioinstitucional.mx</a></td> </tr> <tr> <td>3</td> <td>30 words / &lt; 1% coincidencia - Internet de 13-jun-2022 12:00a. m.</td> <td><a href="#">cibnor.repositorioinstitucional.mx</a></td> </tr> <tr> <td>4</td> <td>18 words / &lt; 1% coincidencia - Internet de 02-jun-2022 12:00a. m.</td> <td><a href="#">cibnor.repositorioinstitucional.mx</a></td> </tr> <tr> <td>5</td> <td>64 words / &lt; 1% coincidencia - Internet de 01-nov-2021 12:00a. m.</td> <td><a href="#">dspace.cibnor.mx</a></td> </tr> <tr> <td>6</td> <td>58 words / &lt; 1% coincidencia - Internet de 24-sept-2018 12:00a. m.</td> <td><a href="#">tel.archives-ouvertes.fr</a></td> </tr> <tr> <td>7</td> <td>45 words / &lt; 1% coincidencia - Internet de 27-jun-2019 12:00a. m.</td> <td><a href="#">viz.be</a></td> </tr> <tr> <td>8</td> <td>36 words / &lt; 1% coincidencia - Internet de 12-ago-2021 12:00a. m.</td> <td><a href="#">imeccocal.cicese.mx</a></td> </tr> <tr> <td>9</td> <td>27 words / &lt; 1% coincidencia - Internet de 14-jun-2016 12:00a. m.</td> <td><a href="#">scien.net</a></td> </tr> </tbody> </table>	1	481 words / 2% - Internet de 21-dic-2021 12:00a. m.	<a href="#">peeri.com</a>	2	53 words / < 1% coincidencia - Internet de 14-jun-2022 12:00a. m.	<a href="#">cibnor.repositorioinstitucional.mx</a>	3	30 words / < 1% coincidencia - Internet de 13-jun-2022 12:00a. m.	<a href="#">cibnor.repositorioinstitucional.mx</a>	4	18 words / < 1% coincidencia - Internet de 02-jun-2022 12:00a. m.	<a href="#">cibnor.repositorioinstitucional.mx</a>	5	64 words / < 1% coincidencia - Internet de 01-nov-2021 12:00a. m.	<a href="#">dspace.cibnor.mx</a>	6	58 words / < 1% coincidencia - Internet de 24-sept-2018 12:00a. m.	<a href="#">tel.archives-ouvertes.fr</a>	7	45 words / < 1% coincidencia - Internet de 27-jun-2019 12:00a. m.	<a href="#">viz.be</a>	8	36 words / < 1% coincidencia - Internet de 12-ago-2021 12:00a. m.	<a href="#">imeccocal.cicese.mx</a>	9	27 words / < 1% coincidencia - Internet de 14-jun-2016 12:00a. m.	<a href="#">scien.net</a>	
1	481 words / 2% - Internet de 21-dic-2021 12:00a. m.	<a href="#">peeri.com</a>																											
2	53 words / < 1% coincidencia - Internet de 14-jun-2022 12:00a. m.	<a href="#">cibnor.repositorioinstitucional.mx</a>																											
3	30 words / < 1% coincidencia - Internet de 13-jun-2022 12:00a. m.	<a href="#">cibnor.repositorioinstitucional.mx</a>																											
4	18 words / < 1% coincidencia - Internet de 02-jun-2022 12:00a. m.	<a href="#">cibnor.repositorioinstitucional.mx</a>																											
5	64 words / < 1% coincidencia - Internet de 01-nov-2021 12:00a. m.	<a href="#">dspace.cibnor.mx</a>																											
6	58 words / < 1% coincidencia - Internet de 24-sept-2018 12:00a. m.	<a href="#">tel.archives-ouvertes.fr</a>																											
7	45 words / < 1% coincidencia - Internet de 27-jun-2019 12:00a. m.	<a href="#">viz.be</a>																											
8	36 words / < 1% coincidencia - Internet de 12-ago-2021 12:00a. m.	<a href="#">imeccocal.cicese.mx</a>																											
9	27 words / < 1% coincidencia - Internet de 14-jun-2016 12:00a. m.	<a href="#">scien.net</a>																											

### Firmas del comité



Dr. Fausto Valenzuela Quiñonez



Dr. Ricardo Pérez Enríquez



Dr. Raúl O. Martínez Rincón



Dr. Andrew J. Severin



Dr. John Carlos Garza

## **Conformación de Comité**

### **Comité Tutorial**

Dr. Fausto Valenzuela Quiñonez  
Centro de Investigaciones Biológicas del Noroeste, S.C.  
Co-Director de Tesis

Dr. Ricardo Pérez Enríquez  
Centro de Investigaciones Biológicas del Noroeste, S.C.  
Co-Director de Tesis

Dr. Raúl Octavio Martínez Rincón  
Centro de Investigaciones Biológicas del Noroeste, S.C.  
Co-Tutor de Tesis

Dr. Andrew J. Severin  
Universidad del Estado de Iowa  
Co-Tutor de Tesis

Dr. John Carlos Garza  
Universidad de California Santa Cruz  
Co-Tutor de Tesis

### **Comité Revisor de Tesis**

Dr. Fausto Valenzuela Quiñonez  
Dr. Ricardo Pérez Enríquez  
Dr. Raúl Octavio Martínez Rincón  
Dr. Andrew J. Severin  
Dr. John Carlos Garza

### **Jurado de Examen**

Dr. Fausto Valenzuela Quiñonez  
Dr. Ricardo Pérez Enríquez  
Dr. Raúl Octavio Martínez Rincón  
Dr. Andrew J. Severin  
Dr. John Carlos Garza

### **Suplentes**

Dr. Pedro Cruz Hernández  
Dr. Juan Antonio de Anda Montañez

## Resumen

La identificación de las poblaciones pesqueras y la delimitación de la escala espacial a la que se distribuyen, deberían ser de alta prioridad para el diseño de estrategias de manejo y recuperación, tanto de las especies sobreexplotadas y amenazadas, como de las mantenidas en niveles sostenibles. Sin embargo, a menudo las unidades biológicas y las unidades de manejo no concuerdan entre sí. El enfoque genómico poblacional posibilita el aislamiento y caracterización de polimorfismos de nucleótido simple (SNPs por sus siglas en inglés) altamente informativos con un elevado poder para diferenciar entre poblaciones y mejorando la resolución espacial de los límites de la población en varios recursos pesqueros. Además, posibilita el desarrollo de paneles de SNPs para análisis de asignación a la población de origen (trazabilidad), que ya se ha probado en el manejo pesquero y la industria de la acuicultura. El abulón azul (*Haliotis fulgens*) representa una importante pesquería artesanal en la costa oeste de la Península de Baja California (PBC). A pesar de las regulaciones existentes, *H. fulgens* está catalogado como deteriorado con sus poblaciones por debajo del nivel óptimo de recuperación. Además, la identificación y delimitación de sus poblaciones aún no se encuentran bien establecidas.

Por lo tanto, el primer objetivo de este estudio fue caracterizar los patrones espaciales de variación genómica neutra de *H. fulgens* a lo largo de la PBC y evaluar la concordancia entre las unidades de manejo actuales del recurso y las unidades biológicas. Para probar esta hipótesis, se utilizó un conjunto de 2,170 SNPs neutrales putativos descubiertos *de novo* mediante un enfoque de ADN asociado al sitio de restricción de doble digestión (ddRAD-seq) en 10 localidades a lo largo de la PBC (n=175). Los resultados revelaron tres grupos genéticos regionales: Isla Guadalupe y las localidades costeras de la PBC dividida en norte (Isla San Jerónimo y Faro San José) y sur (Isla Cedros, Punta Eugenia, Tortugas Norte, Tortugas Sur, Asunción Norte, Asunción Sur y La Bocana). El flujo de genes contemporáneo podría explicarse por las características oceanográficas locales, donde es bidireccional dentro de la región sur, pero con un flujo predominante hacia el sur desde la región norte con un modelo de diferenciación de Aislamiento por Distancia.

El segundo objetivo fue desarrollar un panel reducido de SNPs para la asignación de poblaciones que describiera con precisión los tres grupos genéticos detectados con los análisis previos. Para ello se mapearon las lecturas crudas con el genoma de referencia de abulón azul, con lo cual se obtuvo un mayor número de SNPs (10,523). Se analizaron los datos a dos escalas jerárquicas: (a) global que incluye los tres grupos genéticos (n= 176) 1) Isla Guadalupe (GI), 2) región norte de la PBC (GCN) y 3) región sur de la PBC (GCS); y (b) la escala costera que incluye solo los dos grupos costeros (n=156). Se clasificaron los loci de mayor a menor en función de su valor de  $F_{ST}$  y se seleccionaron seis paneles SNP (de 50 a 500), comenzando con el  $F_{ST}$  más alto, para maximizar el poder de detectar la diferenciación genética entre los grupos genéticos en ambas escalas. El modelo predictivo a escala global mostró un alto poder de asignación para casi todos los conjuntos de datos de SNPs para dos de los tres grupos (GI y SCG), con un rango de 0.81 (d.s.= 0.25) a 1. A escala costera, las precisiones de asignación aumentaron para ambos grupos, sin embargo, NCG mostró los valores más bajos, que van desde 0.03 (s.d.= 0.16) con los 50 SNP más informativos, hasta 0.91 (d.s.= 0.23) con 500 SNP.

Los resultados obtenidos en este estudio indicaron que las áreas administrativas no coincidían con las unidades biológicas de la pesquería de *H. fulgens*; por lo tanto, las áreas de evaluación y manejo del stock deben revisarse. Los paneles SNP desarrollados son relevantes para los administradores de abulón azul, ya que podrían contribuir al plan de gestión pesquera en desarrollo.

**Palabras clave:** genómica poblacional, manejo pesquero, trazabilidad, ddRAD-seq, panel SNPs, asignación origen poblacional.

**ORCID:** <https://orcid.org/0000-0003-4176-8887>

Vo. Bo Co-Directores de Tesis



Dr. Fausto Valenzuela Quiñonez



Dr. Ricardo Pérez Enríquez

## Summary

Stock identification of fishery populations and their spatial delimitation should be of high priority for the design of management and recovery strategies of overexploited and endangered wildlife, as well as those in sustainability levels. However, a mismatch between biological and management units of fishery resources often occurs. Genomic approaches simplify the ability to isolate informative Single Nucleotide Polymorphisms (SNPs) with high population differentiation power, improving the spatial resolution of population boundaries in several fisheries resources. The use of SNPs panels in assignment analyses has also been proven in fishery management and the aquaculture industry. The green abalone (*Haliotis fulgens*) represents an important artisanal fishery along the western coast of the Baja California Peninsula (BCP). Despite the existing regulations, the green abalone is listed as deteriorated with its populations below the optimum recovery level. In addition, their populations' identity and delimitation are not yet well established.

Thus, the first aim of this study was to characterize the spatial patterns of neutral genomic variation of *H. fulgens* along the BCP to test whether the genomic structure patterns support the current management areas of this fishery resource. To test this hypothesis, a set of 2,170 putative neutral SNPs discovered *de novo* by a double digest restriction-site associated DNA approach was used on 10 locations along the BCP (n=175). The results revealed a population structure with three putative groups: Guadalupe Island and northern (Isla San Jerónimo y Faro San José) and southern BCP locations (Isla Cedros, Punta Eugenia, Tortugas Norte, Tortugas Sur, Asunción Norte, Asunción Sur y La Bocana). The contemporary gene flow might be explained by local oceanographic features, where it is bidirectional within the southern region but with a predominant southward flow from the northern region, following an Isolation-by-Distance model.

The second aim was to develop a reduced SNPs panel for population assignment that accurately describe the three genetic groups detected with previous analysis. For that purpose, the raw reads were aligned to the green abalone reference genome, from which a larger number of SNPs was obtained (10,523). The data were analyzed at two hierarchical scales: (a) the global scale including the three genetic groups (176 individuals) 1) Guadalupe Island (GI), 2) northern region of the BCP (NCG) and 3) southern region of the BCP (SCG); and (b) the coastal scale including only the two coastal groups (n=156). Six SNP panels (from 50 to 500), starting with the highest  $F_{ST}$  were selected to maximize the power to detect genetic differentiation between the genetic groups at both scales. The predictive model at global scale displayed a high discriminatory power for almost all the SNPs datasets for two of the three groups (GI and SCG), ranging from 0.81 (s.d.= 0.25) to 1. At the coastal scale, assignment accuracies increased for both groups, nonetheless, NCG showed the lowest values, ranging from 0.03 (d.s.= 0.16) with the 50 most informative SNPs, to 0.91 (s.d= 0.23) using 500 SNPs.

The results obtained in this study indicated that the administrative areas did not match the biological units of *H. fulgens* fishery; hence, the stock assessment and management areas should be revised. The developed SNP panels are relevant to managers of green abalone as they might contribute to the fishery management plan under development.

**Keywords:** Population genomics, fisheries management, *traceability*, ddRAD-seq, SNPs panel, assignment population origin.

**ORCID:** <https://orcid.org/0000-0003-4176-8887>

Vo. Bo Co-Directores de Tesis



Dr. Fausto Valenzuela Quiñonez



Dr. Ricardo Pérez Enríquez

**Dedicatoria**

**A mis Padres Rosa Ruíz y Fernando Mejía**

**A Paulo Mora Martínez y a mis Hijos León y Máximo**

## **Agradecimientos**

Al CIBNOR por el apoyo y las facilidades para realizar mis estudios de Posgrado.

A todo el personal administrativo y académico del CIBNOR que lo hacen funcionar y que siempre están dispuestos a ayudar y hacer más fácil la vida de los estudiantes, en especial a Tania Verónica Núñez Valdez, Lic. Osvelia Ibarra Morales, Dra. Elisa Serviere Zaragoza, Ing. Horacio Sandoval Gómez, Lic. Leticia González Rubio, Lic. Claudia Olachea León, Lic. Ana María Talamantes Cota, Lic. Adriana Franzoni García.

Al personal de Posgrado y en especial a la Dra. Gracia Gómez Anduro por su empatía y comprensión frente a los imprevistos y contratiempos presentados durante mi Doctorado.

Al CONACyT por la beca otorgada con número de registro CVU 492144.

A mis Co-Directores de Tesis Dr. Fausto Valenzuela Quiñonez y Dr. Ricardo Pérez Enríquez por su apoyo incondicional, porque me mostraron la manera en que se guía a un estudiante, por retarme y orientarme durante este camino, por toda su confianza y sobre todo por la paciencia que siempre me tuvieron, muchas gracias.

A mis Co-Tutores de Tesis Dr. Raúl O. Martínez Rincón, Dr. John Carlos Garza y Dr. Andrew Severin Thank you so much for your comments and patience during this thesis.

A mi Comité revisor durante el predoctoral, gracias por sus comentarios en ese momento Dr. Pedro Cruz Hernández, Dra. Elisa Serviere Zaragoza, Dr. Juan Antonio de Anda Montañez, Dr. Sergio Flores Ramírez, Dr. José Luis Gutiérrez González y Dr. Noé Díaz Viloría.

Al Laboratorio de Genética Molecular del CIBNOR, en especial a la Dra. Norma Yolanda Hernández Saavedra, Dra. Crisalejandra Rivera Pérez y a la M en C. Delia Rojas Posadas por todas las facilidades, asesoría y apoyo brindados.

Al Laboratorio de Genómica y Bioinformática del CIBNOR, en especial al Dr. Ricardo Vázquez Juárez por permitirme utilizar su Laboratorio y equipo especializado para la realización de mi tesis y a la M en C. Gabriela Mendoza Carreón por el apoyo técnico.

A los técnicos y demás personas de las cooperativas de la FEDECOOP por su apoyo para la toma de muestras, y por su hospitalidad durante las salidas de campo. Mario Ramade Villanueva, Espinoza-Montes, A., Salinas, C., Miranda-Saucedo, C., Bonilla-Gutiérrez, J.C., Villavicencio, G., Aguilar-Osuna, D., Aguilar-Osuna, J.D., Franco, R., Castro, A., Cortés-Estrada, C., Villegas, M.A., Bracamontes, M., Raúl Concepción Urías Arellano†, Javier Alberto Gonzáles Gallegos, Manuel

Antonio Villegas Urías, Cesar Cortez Estrada, Alfonso Urías Candelario y a las familias de los técnicos y pescadores que fueron tan atentos con nosotros durante las salidas de campo.

A Arturo Hernández Velasco de COBI por brindarnos apoyo y alojamiento durante la colecta de Isla Natividad.

A la CONANP-Reserva de la Biósfera Isla Guadalupe (Marisol Torres Aguilar y Donaxí Borjes Flores) por su apoyo para las colectas en Isla Guadalupe.

A la SEMAR 2ª región Naval de Ensenada, B.C por su invaluable apoyo para el traslado a isla Guadalupe e isla Cedros.

Al RAD-team Fausto, Paulo, Jorge, Cami, disfruté mucho aprender juntos sobre este bello y complicado mundo de la genómica.

A mi hermanito abulonero Jorge Mares Mayagoitia porque reímos, sufrimos y nos apoyamos en todo momento, por hacer los tiempos felices más felices y los difíciles más llevaderos.

A mis compañeros de campo (Paulo, Jorge, Luis Villavicencio, Luis Custodio, Horacio Cabrera, Salwa, Cami) por su apoyo en las colectas de las muestras de abulón y por hacer de las salidas de campo algo inolvidable.

A Fausto y Cris porque además de su apoyo académico nos brindaron una linda amistad y su cariño durante este tiempo.

A mis compañeros y amigos de cubo Gina, Betza, Jorge, ese cubo no hubiera sido lo mismo sin ustedes.

A los ex gen-conianos porque siempre estaré agradecida por su ayuda y cariño durante mi formación.

A las pilas (Eli, Dona, Aby, Magda, Adri, Mariana) por su cariño y apoyo siempre, aunque estemos regadas en la república, las quiero...

A mi familia paceña, Susi, Fede, Kass, las pilas, Paty, Dany, Marquito, Leo, Sara, Dellis, Tania, Mitzi, Jorge, Javi. Gracias por el inmenso apoyo y cariño, aunque no siempre nos vemos el cariño no disminuye.

A mi amiga y hermana Patito Rosas por tus consejos, jalones de orejas, por todo el cariño que le tienes a mis pequeños y la lista es infinita, simplemente gracias.

A la Dra. Martha Patricia por su apoyo en edición y R.

A Annita y a Gina por sus correcciones de inglés.

A mi amado esposo Paulo Mora por todo el apoyo y cariño, gracias por no dejarme rendirme nunca y por la bella familia que tenemos.

A mis hermanos y a mis padres por ser mi pilar y por todo el apoyo y cariño que siempre me brindan. Y en especial a mis padres, aunque no alcanzan las palabras para agradecer todo lo que hacen por mí. Los amo.

A mis hijos, aunque son tan pequeños son lo más importante y una gran motivación en mi vida.

## Contenido

<b>Resumen</b> .....	i
<b>Summary</b> .....	iii
<b>Dedicatoria</b> .....	v
<b>Agradecimientos</b> .....	vi
<b>Contenido</b> .....	ix
<b>Lista de figuras</b> .....	xii
<b>Lista de tablas</b> .....	xv
<b>1. INTRODUCCIÓN</b> .....	1
<b>2. ANTECEDENTES</b> .....	5
2.1 Biología de la especie .....	5
2.2 Pesquería, medidas de regulación, estado del recurso y problemáticas .....	6
2.3 Características del área de estudio .....	8
2.4 Genómica y manejo de recursos pesqueros naturales .....	10
2.4.1 Polimorfismos de una sola base (SNPs) en estudios de genómica poblacional .....	10
2.4.2 Estudios de asignación poblacional usando marcadores SNPs.....	11
2.4.3 Estudios de genética y trazabilidad de poblaciones en haliótidos en el Pacífico nororiental .....	13
<b>3. JUSTIFICACIÓN</b> .....	15
<b>4. HIPÓTESIS</b> .....	17
<b>5. OBJETIVOS</b> .....	18
5.1 Objetivos general .....	18
5.2 Objetivos particulares .....	18
<b>6. MATERIAL Y MÉTODOS</b> .....	19
6.1 Área de estudio y toma de muestras .....	19
6.2 Preparación de librerías genómicas y secuenciación.....	21
6.2.1 Extracción de ADN genómico.....	21
6.2.2 Preparación de librerías ddRAD .....	21
6.3 Análisis bioinformático.....	22
6.3.1 Limpieza, filtrado y control de calidad de secuencias crudas.....	22
6.3.2 Optimización de parámetros, identificación de SNPs y obtención de genotipos de novo..	22
6.3.3 Estimación del error de genotipificación .....	24
6.4 Detección de loci candidatos a selección ( <i>loci</i> adaptativos).....	24
6.5 Equilibrio de Hardy-Weinberg.....	25
6.6 Desequilibrio de ligamiento .....	25
6.7 Estructura genómica poblacional neutral .....	25
6.7.1 Diferenciación genética poblacional ( $F_{ST}$ ) .....	25
6.7.2 Análisis Discriminante de Componentes Principales (DAPC).....	26
6.7.3 Análisis de clusters .....	26
6.7.4 Análisis de Varianza Molecular (AMOVA) .....	27
6.8 Análisis de diversidad genética .....	27
6.9 Tasas de migración relativa .....	28

6.10 Relación entre estructura genómica y distancia geográfica .....	28
6.11 Identificación de SNPs y obtención de genotipos utilizando genoma de referencia .....	29
6.11.1 Limpieza, filtrado y control de calidad de secuencias crudas .....	29
6.11.2 Alineamiento contra genoma de referencia .....	29
6.11.3 Identificación de SNPs y obtención de genotipos con genoma de referencia .....	30
6.12 Asignación poblacional.....	30
6.12.1 Selección de SNPs altamente informativos.....	30
6.12.2 Análisis de asignación y prueba de poder in silico .....	31
6.12.3 Análisis de estructura genética poblacional .....	32
<b>7. RESULTADOS</b> .....	<b>33</b>
7.1 Preparación de librerías genómicas y secuenciación.....	33
7.1.1 Extracción de ADN genómico .....	33
7.1.2 Preparación de librerías ddRAD .....	33
7.2 Análisis bioinformático .....	35
7.2.1 Limpieza, filtrado y control de calidad de secuencias crudas.....	35
7.2.2 Optimización de parámetros, Identificación de SNPs y obtención de genotipos de novo..	36
7.2.3 Estimación de error de genotipificación .....	37
7.3 Detección de loci candidatos a selección ( <i>loci</i> adaptativos).....	37
7.4 Equilibrio de Hardy-Weinberg (E-HW) .....	37
7.5 Desequilibrio de ligamiento .....	37
7.6 Estructura genómica poblacional neutral .....	38
7.6.1 Diferenciación genética poblacional ( $F_{ST}$ ) .....	38
7.6.2 Análisis Discriminante de Componentes Principales (DAPC).....	39
7.6.3 Análisis de clusters .....	40
7.6.4 Análisis de Varianza Molecular (AMOVA) .....	41
7.7 Análisis de diversidad genética .....	43
7.8 Tasas de migración relativa.....	44
7.9 Relación entre la estructura genómica neutral y la distancia geográfica .....	46
7.10 Identificación de SNPs y obtención de genotipos usando genoma de referencia .....	48
7.10.1 Limpieza, filtrado y control de calidad de secuencias crudas.....	48
7.10.2 Alineamiento contra genoma de referencia .....	49
7.10.3 Identificación de SNPs y obtención de genotipos con genoma de referencia .....	49
7.11 Asignación poblacional.....	50
7.11.1 Selección de SNPs altamente informativos.....	50
7.11.2 Análisis de asignación y prueba de poder .....	50
7.11.3 Análisis de estructura genética poblacional .....	53
<b>8. DISCUSIÓN</b> .....	<b>57</b>
8.1 Preparación de librerías .....	57
8.2 Análisis bioinformático.....	58
8.2.1 Limpieza, filtrado y control de calidad de secuencias crudas.....	58
8.2.2 Optimización de parámetros, Identificación de SNPs y Obtención de genotipos .....	58
8.3 Estructura genómica poblacional.....	61

8.4 Diversidad genética .....	66
8.5 Implicaciones de los patrones de estructuración poblacional en el manejo del recurso.....	68
8.6 Asignación poblacional.....	68
8.6.1 Selección de SNPs altamente informativos, análisis de asignación y pruebas de poder ....	68
8.6.2 Análisis de estructura genética poblacional .....	73
8.6.3 Potenciales usos y alcances de los paneles de SNPs para el manejo del recurso .....	73
<b>9. CONCLUSIONES</b> .....	76
<b>10. LITERATURA CITADA</b> .....	78
<b>11. ANEXOS</b> .....	92
Anexo A. Preparación de librerías ddRAD.....	92
Anexo B. Descripción del pipeline STACKS y los principales parámetros .....	93
Anexo C. Optimización de parámetros, Identificación de SNPs y obtención de genotipos .....	95
Anexo D. Pruebas no paramétricas de Gao para comparaciones múltiples de valores de Heterocigosidad observada ( <i>H<sub>o</sub></i> ) en 10 localidades de <i>H. fulgens</i> .....	100
Anexo E. Pruebas no paramétricas de Gao para comparaciones múltiples de valores de <i>Riqueza alélica</i> ( <i>R<sub>a</sub></i> ) en 10 localidades de <i>H. fulgens</i> .....	100
Anexo F. Matriz de asignación poblacional con los seis paneles de SNPs a escala global.....	101
Anexo G. Matriz de asignación poblacional con los seis paneles de SNPs a escala costera.....	101
Anexo H. Matriz de asignación poblacional utilizando todos los marcadores a escala global y costera.....	102
Anexo I. Resultados de Análisis discriminante de componentes principales (DAPC) en <i>H. fulgens</i> con los seis paneles de SNPs a escala global.....	103
Anexo J. Resultados de Análisis discriminante de componentes principales (DAPC) en <i>H. fulgens</i> con los seis paneles de SNPs a escala costera. ....	109
Anexo K. Artículo publicado.-Population genomics reveals a mismatch between management and biological units in green abalone ( <i>Haliotis fulgens</i> ).....	115
Anexo L. Artículo sometido.- A novel SNP panel enables population assignment of green abalone ( <i>Haliotis fulgens</i> ): implications and potential uses for fishery management. ....	138

## Lista de figuras

- Figura 1.** A) Sistema de corrientes en la Península de Baja California. Se muestra en punteado la Corriente de California y en flechas más pequeñas la Contracorriente Costera. **B)** Concentración de pigmento promedio fotosintético anual ( $\text{mg}/\text{m}^3$ ) mostrando los principales Centros de Actividad Biológica (BACs). Tomado y modificado de Lluch-Belda et al. (2000). ..... 9
- Figura 2.** Mapa de los sitios de muestreo a lo largo de la Península de Baja California, México. GI, Isla Guadalupe; SJI, Isla San Jerónimo; FSJ, Faro San José; IC, Isla Cedros; PE, Punta Eugenia; TN, Tortugas Norte; TS, Tortugas Sur; AN, Asunción Norte; AS, Asunción Sur; Bo, La Bocana. Las líneas negras y los números romanos indican las cuatro zonas administrativas de la pesquería de abulón: Zona I (Islotes Coronado, B.C. a Punta Malarrimo, B.C.S.), zona II (Punta Malarrimo a Bahía Asunción, B.C.S), zona III (Bahía asunción a Punta Holcomb) y zona IV (Pulta Holcomb a Arroyo El Conejo, B.C.S). ..... 20
- Figura 3.** Pruebas de parametrización. Combinación de parámetros utilizados para la prueba de parametrización en el módulo `denovo_map.pl` (-m, -M y -n) y `populations` (-p) del pipeline STACKS..... 23
- Figura 4.** Muestras de ADN de individuos de *H. fulgens*. Gel de agarosa-Synergel al 1% de las diluciones de trabajo, carril derecho muestra la escalera de 1 kb (Thermo Fisher®). ADN de *H. fulgens* a 250 ng/ $\mu\text{L}$  para la elaboración de librerías genómicas..... 33
- Figura 5.** Productos de ligación de librería genómica. Gel agarosa-Synergel al 2%. Se muestran los productos de ligación después de realizar el PCR1 de la librería genómica, al centro se muestra la escalera 1kb plus (Thermo Fisher®). ..... 34
- Figura 6.** Pools para formar librerías genómicas. Gel agarosa-Synergel al 2%, a la izquierda se muestra la escalera 1kb plus (Thermo Fisher®). Muestra los dos pools antes de formar la librería, cada uno con un index distinto; se observa que su peso molecular está por debajo de las 500 pb. .... 35
- Figura 7.** Análisis Discriminante de Componentes Principales (DAPC) utilizando un set de 2,170 SNPs neutrales en 175 individuos de *H. fulgens* pertenecientes a 10 localidades. **A)** Valores del Criterio de Información Bayesiano (BIC) versus el número de grupos (K). **B)** Gráfica de la distribución poblacional basada en la primera función discriminante. **C)** Gráfica de DAPC, los puntos representan los individuos con colores que indican una localidad de muestreo y la inclusión de la elipse del 95 % de probabilidad de pertenencia al clúster. Las leyendas están ordenadas de norte a sur GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana. .... 40
- Figura 8.** Numero de grupo más probables (K) y probabilidades de pertenencia. **A)** Número de grupos (K) más probable encontrados con el método de Evanno et al. (2005) a partir del valor de  $\Delta K$  utilizando un set de 2,170 SNPs neutrales en 175 individuos de *H. fulgens* pertenecientes a 10 localidades de muestreo. **B)** Probabilidades de pertenencia de cada individuo a los 3 grupos. .... 41
- Figura 9.** Coeficientes de asignación para **A)** K=2 y **B)** K=3 como grupo genéticos distintos más probables. Cada línea vertical representa el genotipo determinado por las frecuencias alélicas para cada individuo y cada color corresponde a la proporción de probabilidad de asignación de cada grupo. El eje horizontal superior e inferior corresponde a las abreviaciones de las 10 localidades de muestreo ordenadas de norte a sur GI= Isla Guadalupe, SJI= Isla San Jerónimo,

FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana. .... 41

**Figura 10.** Redes de migración relativa para el abulón azul (*Haliotis fulgens*) basadas en 2,170 SNPs neutrales entre 10 localidades de muestreo usando la estimación basada en el  $G_{ST}$  de Nei en el paquete *divMigrate*, Sólo se presentan las tasas de migración  $> 0.6$ , los círculos representan las localidades, mientras que las flechas indican la dirección y magnitud de los niveles relativos de migración. Las flechas más oscuras indican una relación de migración más fuerte en comparación con las flechas más claras. **A)** todas las localidades y **B)** región costera sur. GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana. .... 45

**Figura 11.** Gráficas de aislamiento por distancia de los valores de  $F_{ST}$  linealizados contra la distancia geográfica del abulón azul (*Haliotis fulgens*) basado en 2,170 SNPs neutrales y 10 localidades de muestreo. Línea negra: aislamiento por distancia para todas las localidades, línea azul: aislamiento por distancia sin Isla Guadalupe y línea roja: aislamiento por distancia para las localidades costeras centro-sur. .... 47

**Figura 12.** Prueba de Mantel con estimador de densidad de kernel bidimensional. Gráficas de aislamiento por distancia de valores de  $F_{ST}$  linealizados contra la distancia genética por pares de poblaciones para las 10 localidades de *H. fulgens* basada en 2,170 SNPs neutrales A) usando todas las localidades, B) sin Isla Guadalupe y C) usando sólo las localidades costeras de la región centro-sur. .... 48

**Figura 13.** Resultados del modelo predictivo para la escala global. Precisiones de asignación estimadas vía Monte-Carlo y la función de clasificación SVM (Support Vector Machine), con dos proporciones de individuos para entrenar al modelo (0.75 y 0.8) de la base de datos de entrenamiento ( $n= 52$ ) y seis paneles de SNPs (50 a 500 SNPs más diferenciados con base en su valor de  $F_{ST}$ ) representados por diferentes colores. Detalles de la gráfica de cajas: la línea dentro de las cajas es la mediana; los bordes superior e inferior de las cajas son los percentiles 25 y 75; los extremos de los bigotes son los valores mínimos y máximos de los datos no atípicos; los puntos negros son los valores atípicos. .... 51

**Figura 14.** Resultados del modelo predictivo para la escala costera. Precisiones de asignación estimadas vía Monte-Carlo y la función de clasificación SVM (Support Vector Machine), con dos proporciones de individuos para entrenar al modelo (0.75 y 0.8) de la base de datos de entrenamiento ( $n= 42$ ) y seis paneles de SNPs (50 a 500 SNPs más diferenciados con base en su valor de  $F_{ST}$ ) representados por diferentes colores. Detalles de la gráfica de cajas: la línea dentro de las cajas es la mediana; los bordes superior e inferior de las cajas son los percentiles 25 y 75; los extremos de los bigotes son los valores mínimos y máximos de los datos no atípicos; los puntos negros son los valores atípicos. .... 52

**Figura 15.** Análisis discriminante de componentes principales (DAPC) en *H. fulgens*. Figuras **A, B** y **C** muestran los resultados utilizando los 2,216 SNPs derivados de novo a **escala global**. Figuras **D, E** y **F** muestran los resultados usando los 9,954 SNPs derivados del alineamiento contra el genoma de referencia del abulón azul a **escala global**. **A, D)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos (K). **B, E)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C, F)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 28 y 35 componentes principales y dos funciones discriminantes (dimensiones) para los datos de novo y con genoma de referencia

respectivamente durante los análisis. Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana..... 54

**Figura 16.** Análisis discriminante de componentes principales (DAPC) en *H. fulgens*. Figuras **A, B** y **C** muestran los resultados utilizando los 2,216 SNPs derivados de novo a **escala costera**. Figuras **D, E** y **F** muestran los resultados usando los 9,869 SNPs derivados del alineamiento contra el genoma de referencia del abulón azul a **escala costera**. **A, D)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos (K). **B, E)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C, F)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 27 y 30 componentes principales y dos funciones discriminantes (dimensiones) para los datos de novo y con genoma de referencia respectivamente durante los análisis. Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana..... 55

**Figura 17.** Patrones de circulación oceánica superficial (<200m) en la porción sur del Sistema de Corrientes de California (SCC) a lo largo de las cuatro estaciones del año, las flechas indican la dirección de la corriente. **a)** invierno, **b)** primavera, **c)** verano y **d)** otoño. Tomado y modificado de Gaxiola-Castro y Durazo (2010)..... 66

## Lista de tablas

<b>Tabla 1.</b> Localidades de muestreo de abulón azul ( <i>Haliotis fulgens</i> ) durante las temporadas de pesca del 2017 y 2018 y temporada de evaluación del recurso 2018. Se muestran el ID de las localidades, la localidad y nombre de las Sociedades Cooperativas Pesqueras (SCPP), así como el tamaño de muestra total y por localidad n.....	19
<b>Tabla 2.</b> Abreviaciones de las localidades de muestreo (ID), número de individuos de cada localidad de muestreo utilizados para elaborar las librerías (n), muestras duplicadas para estimar el error de genotipificación (eg).....	34
<b>Tabla 3.</b> Se muestra el número de individuos por localidad después de eliminar aquellos con <1,000,000 de lecturas, número de lecturas promedio retenidas por localidad obtenidas de novo y número de individuos por localidad después de eliminar aquellos con > 10% de genotipos faltantes.....	36
<b>Tabla 4.</b> Resumen del número de SNPs obtenidos de novo retenidos después de cada filtro para <i>Haliotis fulgens</i> . .....	38
<b>Tabla 5.</b> Valores de $F_{ST}$ (Weir y Cockerman, 1984) para comparaciones pareadas por localidades de muestreo utilizando el set de 2,170 SNPs neutrales en 175 individuos de <i>H. fulgens</i> pertenecientes a 10 localidades (triángulo inferior). Valores de P para valores pareados de $F_{ST}$ (triángulo superior). En negrita se muestran los valores significativos ( $P \leq 0.001$ después de la corrección de BF).....	39
<b>Tabla 6.</b> Análisis de Varianza Molecular (AMOVA) con 2,170 loci neutrales en 175 individuos de <i>H. fulgens</i> . AMOVA para examinar la partición de la variación genética en cuatro niveles jerárquicos entre grupos regionales de poblaciones identificadas por los análisis previos y con diferentes estructuras hipotéticas. Se muestra la fuente de variación, los grados de libertad (gl), Índices de fijación ( $F_{CT}$ : variación debido a diferencias entre los grupos; $F_{SC}$ : variación debido a diferencias entre poblaciones dentro de los grupos; $F_{IS}$ : variación debido a diferencias entre individuos dentro de las poblaciones y $F_{IT}$ : variación debido a diferencias entre individuos) con su valor de probabilidad ( $P \leq 0.05$ ) y el porcentaje de variación.....	42
<b>Tabla 7.</b> Tabla resumen de la diversidad genética para el set de marcadores neutrales (2,170 SNPs) en 175 individuos de <i>H. fulgens</i> pertenecientes a 10 localidades. Se muestran los valores de heterocigosidad observada ( $H_O$ ) y esperada ( $H_E$ ) para cada localidad y global. Valores del coeficiente de endogamia ( $F_{IS}$ ) para cada localidad y el promedio. ....	44
<b>Tabla 8.</b> Se muestra el número de individuos por localidad después de eliminar aquellos con <1,000,000 de lecturas, así como el número de lecturas promedio retenidas por localidad obtenidas con genoma de referencia. ....	49
<b>Tabla 9.</b> Valores de diversidad genética reportados en especies marinas obtenidos con marcadores moleculares tipo SNPs. Se muestra la cita del trabajo, la especie, el tipo de RAD-seq utilizado, tamaño de muestra (n), # de SNPs utilizados para la estimación de los índices heterocigosidad observada ( $H_O$ ) y Heterocigosidad esperada ( $H_E$ ), valores de diferenciación genética ( $F_{ST}$ ) con su valor de probabilidad asociado (P), coeficiente de endogamia ( $F_{IS}$ ) o su análogo ( $G_{IS}$ ) y resultados de pruebas de asilamiento por distancia (IBD) con su respectivo valor de probabilidad asociado (P).....	67

## 1. INTRODUCCIÓN

Las estrategias efectivas de manejo pesquero incluyen la delimitación de stocks (unidades biológicas fundamentales sobre la cual se realiza un manejo) y requieren del conocimiento de la escala espacial de las unidades biológicas (Selkoe y Toonen, 2006; Hedgecock *et al.*, 2007; Funk *et al.*, 2012). Evaluar el estado de la población o desarrollar una estrategia de manejo podría no ser sostenible si no se conoce primero cómo están delimitadas las poblaciones (Funk *et al.*, 2012). Sin embargo, las unidades de manejo se basan predominantemente en unidades administrativas (Bernatchez *et al.*, 2017) por lo que con frecuencia no hay concordancia entre las unidades de manejo y las unidades biológicas (e.g. *Gadus morhua*, *Melanogrammus aeglefinus*, *Merlangius merlangus*, *Micromesistius poutassou*, *Merluccius merluccius*, *Clupea harengus*, *Thunnus albacares* Reiss *et al.*, 2009; Mullins *et al.*, 2018). Esta discrepancia puede llevar a una reducción en la productividad o la sobre explotación de las subpoblaciones discretas (Frank y Brickman, 2000; Kenchington, 2003; Hilborn *et al.*, 2003).

En el ambiente marino, determinar la estructura genética y definir las unidades de manejo es desafiante debido a que no existen barreras físicas aparentes y los organismos muestran grandes tamaños efectivos poblacionales y un alto flujo genético. Esta situación resulta en una baja diferenciación poblacional en donde las herramientas genéticas tradicionales con un limitado número de loci suelen tener poca potencia para definir las unidades biológicas (DeWoody y Avise, 2000; Hauser y Carvalho, 2008; Allendorf *et al.*, 2010).

Los avances en tecnologías de secuenciación masiva y en los métodos de genotipificación de ADN asociado a sitios de restricción (RAD-seq) han permitido descubrir y genotipificar una gran cantidad de loci en especies no modelo, lo que permite hacer frente a las limitaciones tecnológicas anteriores (Hammer Hansen, Therkildsen y Pujolar, 2014).

El uso de marcadores tipo Polimorfismos de Nucleótido Sencillo o SNP (Single Nucleotide Polymorphism, por sus siglas en inglés) ha aumentado significativamente en las últimas décadas, dado que permiten realizar inferencias más precisas acerca de la estructura genómica

poblacional neutral (Luikart *et al.*, 2003; Baird *et al.*, 2008; Grewe *et al.*, 2015), además de mejorar la resolución espacial de los límites de población en varios recursos pesqueros (e.g., Thunnus alalunga, Laconcha *et al.*, 2015; Salmo salar L., Gilbey *et al.*, 2016; Thunnus albacares, Pecoraro *et al.*, 2018) a un menor costo y tiempo de inversión en comparación con las técnicas tradicionales (Jenkins *et al.*, 2019).

Recientemente, algunos estudios, se han enfocado en la identificación de SNPs altamente informativos que muestran frecuencias alélicas fuera de las esperadas bajo neutralidad (*outliers*) y mayores niveles de divergencia entre las poblaciones, permitiendo el desarrollo de paneles reducidos de SNPs para capturar la diferenciación en la escala espacial de interés a un menor costo (Jenkins *et al.*, 2019).

Específicamente, los métodos de asignación han sido utilizados en la conservación y manejo de especies ya que permiten la discriminación de individuos residentes e inmigrantes en poblaciones locales y la identificación de poblaciones de origen de individuos inmigrantes (Manel *et al.*, 2005). El uso de paneles reducidos de SNPs en los análisis de asignación se ha probado en la gestión pesquera y en la industria acuícola para ayudar a abordar la pesca ilegal, la trazabilidad, detección del fraude de especies (Shackell y Dodds, 2008), para promover la eco certificación (Nielsen *et al.*, 2012; Jenkins *et al.*, 2019) y en estudios de dispersión (Berry *et al.*, 2004; Reynes *et al.*, 2021).

La pesquería de abulón en México representa una actividad de gran importancia económica y social para las comunidades pesqueras establecidas a lo largo de la Península de Baja California (PBC) (León y Muciño, 1996). A lo largo de la PBC se explotan de manera comercial 5 especies de abulón (abulón azul: *Haliotis fulgens*, abulón amarillo: *H. corrugata*, abulón negro: *H. cracherodii*, abulón chino: *H. sorenseni* y abulón rojo: *H. rufescens*) (DOF, 2018) sin embargo la pesca está sustentada casi en su totalidad por el abulón amarillo y el abulón azul. Se estimó que en el periodo del 2000 al 2018 su producción fue de aproximadamente 274 toneladas anuales con un valor de 67.3 millones de pesos (INAPESCA, 2021).

El abulón azul tiene un rango de distribución que va desde Bahía Concepción, California, EUA hasta Bahía Magdalena, BCS, México (Guzmán-del Proo, 1992) el cual está influenciado por las Corrientes de California y Ecuatorial del Norte, generando condiciones oceanográficas y de circulación oceánica heterogéneas (Chelton, 1982; Lluch-Belda, 1999; Durazo *et al.*, 2010; Durazo, 2015).

A pesar de que el plan de manejo del recurso aún se encuentra en desarrollo, la pesquería se encuentra regulada bajo varias normas, incluidas zonas administrativas, concesiones por cooperativas, tallas mínimas legales y temporadas de vedas, las cuales se establecen con base en la variación latitudinal de sus características de historia de vida (tallas de primera madurez y temporadas de reproducción) (DOF, 1993, 1994; 2018; Sierra-Rodríguez *et al.*, 2006). Adicionalmente, se realizan evaluaciones de stock por bancos y concesiones de cooperativas y desde el 2000 estas regulaciones se han enfocado principalmente en las cuotas de capturas asignadas por cooperativa pesquera considerando sus tasas de crecimiento anual (Sierra-Rodríguez *et al.*, 2006).

A pesar de las diversas medidas de manejo y el gran esfuerzo de los pescadores locales para la restauración de los stocks, la biomasa del recurso muestra una tendencia marcada de disminución similar al observado en otras pesquerías de Haliótidós (Karpov *et al.*, 2000; Dowling *et al.*, 2004; Cook, 2016; Mundy *et al.*, 2020; Rogers-Bennett *et al.*, 2021). Estas disminuciones se han atribuido una combinación de factores tales como manejo inadecuado de los recursos, lenta recuperación, sobrepesca, pesca ilegal, enfermedades y efectos del ambiente (Guzmán-del Proo *et al.*, 2003; Prince, 2005; Morales-Bojórquez *et al.*, 2008; Ponce-Díaz *et al.*, 2013; Cook, 2019). Actualmente se encuentra catalogado como en deterioro con sus poblaciones por debajo del nivel óptimo de recuperación (DOF, 2018).

Desde el punto de vista genético un estudio previo realizado con 4 marcadores tipo microsatélites en *H. fulgens*, concluye panmixia para las poblaciones de la parte centro-oeste de la PBC, solamente diferenciadas de la localidad oceánica Isla Guadalupe (Gutiérrez-González *et*

*al.*, 2007). Sin embargo, no se han realizado estudios que evalúen la estructura genética y/o genómica en las localidades de abulón azul más al norte de Punta Eugenia.

El presente trabajo amplía el rango de muestreo al analizar individuos de localidades al norte de Punta Eugenia, así como el uso de miles de marcadores genéticos para obtener una mayor resolución. Por lo tanto, el principal objetivo fue evaluar la estructura genómica poblacional del abulón azul (*Haliotis fulgens*, Philippi 1845) a lo largo de gran parte de su rango de distribución en la PBC mediante marcadores tipo SNP derivados de técnicas de secuenciación de ADN asociada al sitio de Restricción de doble digestión (ddRAD-seq). El segundo objetivo del estudio fue desarrollar un panel reducido de SNPs para la asignación de poblaciones que describan con precisión los grupos genéticos detectados con los análisis de estructuración poblacional a lo largo de la PBC.

## 2. ANTECEDENTES

### 2.1 Biología de la especie

El abulón azul (*Haliotis fulgens*, Phillipi, 1845) es un molusco gasterópodo perteneciente a la familia Haliotidae. Se distribuye desde Bahía Concepción, California, EUA (34.5° N), hasta Bahía Magdalena, BCS, México (25.5 °N), con máximas densidades desde Isla Cedros hasta Punta Abreojos y con una distribución batimétrica que comprende desde el intermareal hasta los 24 m de profundidad (Guzmán-del Proo, 1992). Es un herbívoro de hábitos nocturnos asociados a fondos rocosos y mantos de algas pardas y rojas (e.g. *Macrocystis*, *Eisenia* y *Gelidium*). Son organismos longevos de lento crecimiento que llegan a vivir de 20 a 25 años, alcanzan la talla de madurez sexual de los 6 a los 12 cm de longitud (5-7 años) y se reporta una variación latitudinal en las tallas con un gradiente que disminuye de norte a sur (Sierra-Rodríguez *et al.*, 2006;).

Los estudios realizados en *H. fulgens* acerca de aspectos reproductivos (fecundidad y temporadas de desove; Sevilla, 1971; Belmar-Pérez y Guzmán-del Proo, 1992), sugieren que el ciclo reproductivo en los abulones varía ampliamente dependiendo del área geográfica, pudiéndose identificar de manera general dos grandes picos reproductivos, uno entre el verano y el otoño (Ortiz-Quintanilla & Leon-Carballo, 1988; Vélez-Arellano *et al.*, 2017) y otro entre invierno y primavera, siendo los meses de Febrero a abril los meses con mayor actividad de desove reportado (Vélez-Arellano *et al.*, 2020). No obstante, este aspecto aún no está bien entendido dado que los estudios se han enfocado principalmente en la zona centro sin considerar el rango entero de distribución.

Los huevos fertilizados tienen flotabilidad negativa y estudios realizados en laboratorio concluyen que el desarrollo larvario es dependiente de la temperatura y que la fase nadadora tiene una corta duración (4 a 15 días) en la columna de agua con un rango de temperatura óptimo de 20 °C a 23 °C hasta que comienzan su vida bentónica como postlarvas (Leighton, 2000). Estas condiciones podrían limitar su capacidad de dispersión a lo largo de largas distancias. Su dispersión depende casi exclusivamente de su fase larvaria ya que una vez que se asientan los organismos hasta su etapa adulta presentan una movilidad limitada.

## 2.2 Pesquería, medidas de regulación, estado del recurso y problemáticas

La pesquería del abulón en México se inició desde mediados del siglo XIX por compañías extranjeras y hasta hace 70 años por cooperativas mexicanas (Ponce-Díaz & Chávez, 2000). De las 8 especies de haliótidos de Norte América, México cuenta con 7 (*H. fulgens*, *H. corrugata*, *H. rufescens*, *H. sorenseni*, *H. cracherodii*, *H. assimilis* y *H. walallensis*) de las cuales *H. fulgens* y *H. corrugata* soportan casi en su totalidad a la pesquería en México, la cual ocurre desde Isla Coronados, B.C. hasta Isla Margarita, B.C.S. en la PBC (INAPESCA, 2006). Su alta demanda y valor económico en el mercado internacional hacen de su pesca una actividad económica importante para la costa Occidental de la PBC (DOF, 2018), generando una derrama económica anual para las comunidades pesqueras de la región de alrededor de 67.3 millones de pesos (datos del 2000 al 2018; INAPESCA, 2021).

La pesca de este recurso se encuentra regulada por la NOM-005-PESC-1993 (DOF, 1993) la cual es una Norma Oficial Mexicana que regula el aprovechamiento de las poblaciones de las distintas especies de abulón en aguas de jurisdicción federal de la PBC y la NOM-09-PESC-2993 (DOF, 1994), la cual establece el procedimiento para determinar las épocas y zonas de veda para la captura de diferentes especies acuáticas en México.

A pesar de que no existe un plan de manejo para el recurso (en desarrollo), desde 1981 se han implementado varias medidas de manejo para regular la pesca y recuperar la abundancia del recurso: se consideran cuatro zonas administrativas las cuales están establecidas con base en las variaciones latitudinales de sus características de historia de vida, específicamente talla de primera madurez y temporada reproductiva (DOF, 1993). Cada zona tiene diferentes periodos de veda y tallas mínimas de captura: zona I (150 mm, 1 Jul-30 Nov); zona II (145 mm, 1 Ago-31 Dic), zona III (140 mm, 1 Ago-31 Dic) y zona IV (120 mm, 1 Sep-1 Ene) (DOF, 1993; 1994; 2018). No se permite la captura en bajamar y se debe entregar el producto vivo en las plantas procesadoras (León-Carballo y Muciño-Díaz, 1996; Searcy-Bernal *et al.*, 2010; DOF, 2012). Además, el bajo reclutamiento del recurso ha dado como resultado el cierre indefinido de la pesquería en algunas zonas, e incluso se han establecido de manera voluntaria zonas de no pesca mediante reservas marinas (Micheli *et al.*, 2012).

A partir del 2000 se puso en marcha un programa entre el Instituto Nacional de Pesca (INAPESCA) y las cooperativas pesqueras para la evaluación de los stocks por bancos y concesiones de las cooperativas y se establecieron las cuotas de captura como principal medida de regulación. Dichas cuotas son asignadas por cooperativa pesquera teniendo en cuenta la tasa anual de crecimiento de cada banco (Sierra-Rodríguez *et al.*, 2006). Sin embargo, a pesar de las diversas medidas de manejo y el gran esfuerzo realizado por parte de los pescadores para repoblar los stocks, esta pesquería aún se encuentra catalogada como en deterioro con sus poblaciones por debajo del nivel óptimo de recuperación (DOF, 2018).

Recientemente, Guzmán-del Proo y del Monte-Luna (2017) sugirieron un cambio para el manejo de la pesquería en donde las evaluaciones deberían de realizarse por grupos de bancos que tenga tasas similares de productividad, en lugar de considerar a cada banco por separado. Sin embargo, ya sean las estrategias actuales o las propuestas (zonas administrativas, concesión por cooperativas, bancos por separado o grupos de bancos con características similares), la escala espacial de la administración sigue siendo un problema de escala indefinido y requiere de la delimitación de unidades biológicas para apoyar el manejo del recurso. De hecho, las autoridades de manejo pesquero reconocen la necesidad de definir los stocks como unidades poblacionales en polígonos bien delimitados, con estrategias de manejo que dependan del estado poblacional de cada banco (DOF, 2012).

Otro problema al que se enfrenta la regulación de la pesquería es la pesca ilegal, ya que se estima que representa alrededor del 27% de la captura comercial (Ponce-Díaz, *et al.*, 2013). Aunado a esto, es probable que las preocupaciones sobre el etiquetado incorrecto (fraude comercial) y la creciente concientización de los pescadores y consumidores sobre la vulnerabilidad de las poblaciones, ha motivado a las pesquerías locales a buscar certificaciones sostenibles (eco-certificaciones) por parte de organizaciones como el Consejo de Administración Marina (MSC por sus siglas en inglés).

Como una medida alterna para la recuperación de los bancos de abulón, a partir de los años 90 y hasta el 2020 era obligatorio realizar acciones de repoblamiento para mantener las concesiones otorgadas; Los programas de repoblamiento se llevaban a cabo mediante

liberación de larvas criadas en laboratorio, sobre todo en la parte central de la PBC (Gutiérrez-González y Pérez-Enriquez, 2005). Sin embargo, las consecuencias de estos programas no han sido evaluadas y persiste una preocupación inherente al posible impacto genético en las poblaciones silvestres (Pérez-Enriquez *et al.*, 2013), el cual puede ir desde pérdida de la diversidad genética y adecuación hasta cambios en la composición de la población o en la estructura poblacional (por ejemplo, cuando existe una estructuración poblacional significativa dentro de la distribución de la especie, la repoblación puede conducir inadvertidamente a la mezcla de distintos linajes evolutivos y a la homogenización de la diversidad genética) (Laikre *et al.*, 2010).

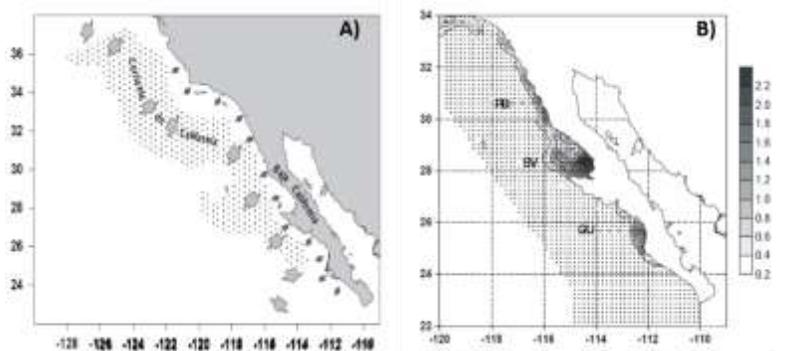
Comprender los elementos que están aportando a la biocomplejidad de *H. fulgens* como son la delimitación de unidades biológicas, patrones de conectividad, grupos con adaptaciones locales, diferencias en tasas de reclutamiento, fecundidad, tallas de madurez; es una tarea clave para aportar al manejo del recurso ya que son estas variaciones las que permiten que los stocks sean resilientes ante cambios ambientales (Hilborn *et al.*, 2003; Ruzzante *et al.*, 2006).

### **2.3 Características del área de estudio**

El área de distribución del abulón azul (~10° de latitud) está influenciada por distintas condiciones oceanográficas y de circulación. En el área que corresponde a México, las características oceanográficas que influyen sobre el recurso son muy particulares y se describen a continuación. La costa occidental de la PBC incluye la porción más sureña del Sistema de la Corriente de California, la cual transporta agua subártica del Pacífico hacia el ecuador. También está influenciada por una contracorriente profunda que viene del sur llevando agua del Pacífico norecuatorial y fluye a lo largo de la costa hacia el norte durante los meses de otoño e invierno (Fig. 1A). A lo largo de la PBC, se forman zonas importantes de surgencias costeras causadas por el viento (incorporación de aguas profundas a la superficie, frías, ricas en nutrientes), sin embargo, se han localizado zonas con mayor concentración biológica, asociadas a estos procesos costeros las cuales tienen altas concentraciones de fitoplancton que coinciden con altas densidades de larvas de peces. Los principales centros de actividad biológica (BACs por sus siglas en inglés) se localizan en Punta Baja (PB), Bahía Sebastián Vizcaíno (SV) y el Golfo de Ulloa

(GU), en los cuales existe diferencias en las temporadas de mayor y menor abundancia (Fig. 1B) (Lluch-Belda *et al.*, 2000).

La dinámica de las corrientes superficiales (< 200 m) a lo largo de la PBC son predominantes hacia el sur con una gran influencia de la Corriente de California durante la mayor parte del año en la región norte (Lynn y Simpson, 1987; Durazo, 2015). Sin embargo, se presentan ciertos cambios estacionales a lo largo del área, los cuales son favorecidos en la región sur durante verano y otoño (Lluch-Belda *et al.*, 2000). También se ha reportado una contracorriente superficial en regiones someras cercanas a la costa la cual puede actuar como banda transportadora para organismos marinos que se encuentren en la columna de agua (Durazo *et al.*, 2010; Durazo, 2015).



**Figura 1.** A) Sistema de corrientes en la Península de Baja California. Se muestra en punteado la Corriente de California y en flechas más pequeñas la Contracorriente Costera. B) Concentración de pigmento promedio fotosintético anual ( $\text{mg}/\text{m}^3$ ) mostrando los principales Centros de Actividad Biológica (BACs). Tomado y modificado de Lluch-Belda *et al.* (2000).

En general, se puede dividir el área en dos grandes regiones teniendo su límite a la altura de Punta Eugenia ( $28.5^\circ\text{N}$ ) (Durazo *et al.*, 2010), la cual se ha sugerido que funciona como una barrera importante para la dispersión de algunos organismos marinos (Briggs, 1974) (e.g., *Fundulus parvipinnis*, Bernardi y Talley, 2000; *Girella nigricans*, Terry *et al.*, 2000; *Embiotoca jacksoni*, Bernardi, 2000). Además, la región norte de la PBC está caracterizada como una subregión templada con fuertes surgencias costeras, mientras que la región sur es subtropical influenciada por el agua superficial ecuatorial y temperaturas más elevadas (Durazo-Arvizu y Baumgartner-Mcbride, 2002).

## 2.4 Genómica y manejo de recursos pesqueros naturales

### 2.4.1 Polimorfismos de una sola base (SNPs) en estudios de genómica poblacional

En este estudio se utilizó la técnica de ddRAD-seq por sus siglas en inglés (*double-digest Restriction site-Associated DNA sequencing*), el cual es un método de baja representación del genoma que utiliza dos enzimas de restricción (generalmente una de corte frecuente y una de corte raro) y la selección del tamaño de los fragmentos RAD distribuidos de manera aleatoria a lo largo del genoma, lo cual le confiere al método una mayor uniformidad entre muestras en la selección de fragmentos para la secuenciación (Peterson *et al.*, 2012; Bhatia *et al.*, 2013; Valencia *et al.*, 2018).

Los polimorfismos de una sola base SNPs por sus siglas en inglés (*Single Nucleotide Polymorphism*), son cambios en una sola base nucleotídica que existen a lo largo de todo el genoma, con mayor abundancia en regiones no codificantes, sin embargo, también existen SNPs en regiones que influyen en la expresión génica los cuales tienen el potencial de detectar la variación genética adaptativa (De Wit *et al.*, 2012). Se ha reportado que los marcadores tipo SNP pueden ofrecer un mayor poder estadístico, mayor cobertura a lo largo del genoma y datos de mayor calidad en comparación de marcadores comúnmente utilizados como son los microsatélites y ADNmt, posicionando a los SNPs como una herramienta genética eficiente y con un mayor costo-beneficio (Morin *et al.*, 2004).

Los estudios de genómica de poblaciones en recursos marinos, así como de importancia comercial han ido en aumento. Algunas aplicaciones prácticas de la genómica aplicada al manejo de recursos pesqueros es la delimitación espacial de stocks, identificación de grupos con adaptaciones locales y asignación de origen (trazabilidad) (*Homarus americanus*, Benestan *et al.*, 2015<sup>a</sup>; *Gadus morhua*, Bernatchez *et al.*, 2017; *Scophthalmus maximus*, do-Prado *et al.*, 2018).

Por ejemplo, Benestan *et al.* (2015<sup>a</sup>) concluyen que utilizar un gran número de marcadores (8,144 SNPs en *Homarus americanus*) les brinda gran resolución para delinear estructura genética poblacional a escala fina en especies marinas altamente conectadas genéticamente.

Mullins *et al.* (2018) evaluaron la concordancia entre las unidades de manejo del atún aleta amarilla (*Thunnus albacares*) en el sur de África y la delimitación genética real de los stocks mediante marcadores tipo SNP neutrales y adaptativos (también conocidos como *outliers*). Estos autores encontraron dos grupos genéticos distintos los cuales no coinciden con los stocks operacionales actuales y concluyen que existe una pesquería mixta, en donde las capturas de un área de manejo pertenecen totalmente al stock reproductor de la otra área administrativa, sugiriendo la necesidad de un reajuste en el manejo debido a las tendencias de decline que presenta la especie.

Recientemente, Sandoval-Castillo *et al.* (2018) realizaron un estudio con 9,109 SNPs (8,786 neutrales y 323 adaptativos) en el abulón de labio verde (*Haliotis laevis*), un importante recurso para las pesquerías en Australia, a lo largo de ~800 km de muestreo. Estos autores concluyen, desde una perspectiva neutral que la población está representada como una sola unidad panmíctica con valores muy bajos de diferenciación poblacional ( $F_{ST}=0.0081$ ) y un ligero patrón de aislamiento por distancia. Sin embargo, al utilizar el set de marcadores adaptativos encuentran 5 grupos genéticamente distintos, en donde la variación adaptativa se vio fuertemente correlacionada con temperaturas mínimas superficiales del mar y concentración de oxígeno.

#### **2.4.2 Estudios de asignación poblacional usando marcadores SNPs**

Los métodos de asignación poblacional apoyan a la conservación y el manejo de las especies ya que permiten la discriminación entre individuos residentes e inmigrantes en poblaciones locales y la identificación de las poblaciones de origen de individuos inmigrantes (Manel *et al.*, 2005). El uso de paneles reducidos de SNPs en los análisis de asignación se ha aplicado para la gestión pesquera y acuícola con la finalidad de identificar stocks pesqueros y poblaciones mixtas, detectar pesca ilegal y fraude de especies, trazabilidad, promover eco-certificación y en estudios de dispersión (Berry *et al.*, 2004; Shackell y Dodds, 2008; Nielsen *et al.*, 2012; Jenkins *et al.*, 2019; Reynes *et al.*, 2021).

A pesar de que existe una variedad de enfoques para el diseño de los paneles (McKinney *et al.*, 2017; Sylvester *et al.*, 2018; Jeffery *et al.*, 2018), parece que seleccionar los marcadores con valores de  $F_{ST}$  más altos y que no están en desequilibrio de ligamiento permite una línea base más adecuada, al reducir la redundancia en el conjunto de datos (Jeffery *et al.*, 2018, Jenkins *et al.*, 2019; May *et al.*, 2020).

Benestan *et al.* (2015<sup>b</sup>) evaluaron el poder de asignación en la langosta americana (*Homarus americanus*) de varios paneles de marcadores, obteniendo precisiones de asignación del 94% con un panel de los 3,000 SNPs más diferenciados según su valor de  $F_{ST}$ . Sin embargo, al evaluar la precisión de asignación a una escala más fina, ésta disminuyó a un 34% aun cuando utilizaron todos los marcadores (10,156). Los autores atribuyen esta disminución en la precisión de asignación al bajo tamaño de muestra y a los bajos niveles de diferenciación a escala fina ( $F_{ST}=0.0018$ ,  $P < 0.001$ ).

Un patrón similar se observa para la langosta europea (*H. gammarus*) en donde utilizan los SNP más diferenciados a diferentes escalas (cuenca, región y localidad de muestreo). Estos autores asignaron correctamente el 99.7% de los individuos a su cuenca de origen utilizando un panel de 79 SNPs. Sin embargo, el éxito de la asignación disminuyó a una escala más fina (localidad de muestreo) del 27 % al 58 % como máximo utilizando el panel de 79 SNP según la localidad de muestreo y la proporción de individuos utilizados en el conjunto de datos de entrenamiento (Jenkins *et al.*, 2019).

En otro estudio realizado con salmón rojo de Alaska (*Oncorhynchus nerka*) (May *et al.*, 2020) se obtuvieron precisiones de asignación del 93 al 96% con un panel reducido de 172 SNPs. Por otra parte, Marie *et al.* (2019) evaluaron la asignación de población en 174 individuos de tiburón martillo (*Sphyrna lewini*) con baja diferenciación genética pero significativa ( $F_{ST}=0.004$ ,  $P<0.001$ ) logrando probabilidades de asignación de 0.81 (s.d.= 0.03) a 0.98 (s.d.=0.01) usando 1,317 SNP neutros putativos.

### 2.4.3 Estudios de genética y trazabilidad de poblaciones en haliótidos en el Pacífico nororiental

Existen diversos estudios de genética de poblaciones realizados para haliótidos en el Pacífico nororiental los cuales reportan diferentes resultados en función del marcador utilizado. En el abulón rojo (*H. rufescens*), utilizando aloenzimas, el gen mitocondrial COI y un microsatélite, se reporta homogeneidad genética (Gaffney *et al.*, 1996; Burton y Tegner, 2000; Kirby *et al.*, 1998), mientras que tanto en el abulón negro (*H. cracherodii*) (Hamm y Burton, 2000) como en el abulón amarillo (*H. corrugata*) utilizando aloenzimas y microsatélites (Díaz-Viloria *et al.*, 2013; del Río-Portilla y González-Avilés, 2001; Díaz-Viloria *et al.*; 2009) se reporta diferenciación genética. En un estudio reciente realizado con SNPs en *H. corrugata* se reporta una ligera pero significativa diferenciación genética ( $F_{ST}$  pareados de 0 a 0.008,  $P < 0.001$ ) a lo largo de la PBC cuando utilizan 1,733 marcadores neutrales, encontrando dos grupos genéticos divididos a la altura de Punta Eugenia, mientras que el patrón adaptativo hace más evidente esta diferenciación ( $F_{ST}$  pareados de 0 a 0.492,  $P < 0.001$ ) (Mares-Mayagoitia *et al.*, 2021). Este marco de evidencias, permiten establecer que los patrones de estructuración inferidos para haliótidos no se encuentran bien establecidos ya sea por la diferencia en el tipo de marcador, escala geográfica y/o especies estudiadas.

De Wit y Palumbi (2013) analizan la estructura genómica poblacional de *H. rufescens* en tres regiones con distintas características ambientales en la parte norte del área de distribución de la especie (entre la Bahía de Monterey y la frontera de Óregon) utilizando 37 individuos y 21,579 marcadores SNPs generados mediante RNA-seq, no encontrando diferenciación genética entre las localidades de estudio ( $F_{ST}= 0.0003$ ). No obstante, al analizar los loci adaptativos detectan 691 SNPs con una diferenciación mayor de la esperada, asociada a genes que podrían conferirles ciertas ventajas adaptativas a los organismos, sugiriendo que, a pesar del patrón de alto flujo genético observado, existe la firma genética de la adaptación local en algunos genes.

Los estudios de genética de poblaciones que se han realizado para *H. fulgens* utilizando marcadores convencionales (aloenzimas y microsatélites) sugieren que forma una población panmíctica en la parte Centro-Sur de la PBC (desde Isla Cedros, B.C. hasta Isla Magdalena,

B.C.S.), mostrando diferencias solamente entre el sitio más alejado (Isla Guadalupe, cerca de 250 km alejada de la costa) (Zúñiga *et al.*, 2000; Gutiérrez-González *et al.*, 2007).

Por otra parte, en un estudio realizado con abulón azul en la parte sur de la Bahía de California, EUA, Gruenthal *et al.* (2013) no encuentran diferenciación genética neutral con 1,209 SNPs neutrales, concluyendo que la especie tiene un alto flujo genético por acción de las corrientes en el área de estudio, sin embargo, hacen énfasis en la necesidad de estudios de patrones adaptativos en un área geográfica más amplia.

A pesar de los trabajos genéticos realizados hasta ahora en *H. fulgens*, aún no está claro cuáles son los principales factores geográficos o ambientales que afectan la estructura genética de la especie. Además, los trabajos enfocados en la trazabilidad solo se han enfocado en la identificación de especies de abulones (Aranceta-Garza *et al.*, 2011; Vargas-Peralta *et al.*, 2022), y no existen reportes de marcadores para la asignación de origen de organismos silvestres. En el presente trabajo se realizó un escaneo del genoma mediante el uso de marcadores tipo SNP con el objetivo de definir con mayor robustez la estructura genética neutral de *H. fulgens*, para la identificación de stocks a lo largo de tres de las cuatro zonas administrativas. La información generada puede ser incorporada en las decisiones de manejo del recurso.

### 3. JUSTIFICACIÓN

El abulón azul (*H. fulgens*) es un valioso recurso pesquero en la costa Occidental de la Península de Baja California, no obstante, a pesar de las diversas medidas de regulación implementadas, es una pesquería catalogada en deterioro. La lenta recuperación de los stocks en función al manejo pesquero pudiera estar relacionada con algunos aspectos biológico-pesqueros del recurso ya que existe poca claridad acerca de la coincidencia entre las zonas administrativas del recurso y la delimitación real de los stocks o unidades biológicas, en donde las cuotas son asignadas por banco para cada cooperativa. Como propuesta alternativa, se ha sugerido un manejo de la pesquería por grupos de bancos que tengan tasas similares de productividad. Sin embargo, la escala espacial de la administración ya sea con las estrategias actuales o las propuestas recientemente, sigue siendo un problema de escala indefinido y requiere de la delimitación de unidades biológicas para apoyar el manejo del recurso. Al no tener claridad en la delimitación de las fronteras poblacionales, la evaluación de stock que se realiza puede contener sesgos cuantitativos para la asignación de cuotas y el seguimiento temporal del estado de las poblaciones de abulón azul. De hecho, las autoridades de manejo pesquero reconocen la necesidad de definir los stocks como unidades poblacionales en polígonos bien delimitados, con estrategias de manejo que dependan del estado poblacional de cada banco (DOF, 2012). Dado lo anterior, las estrategias de manejo del recurso pueden beneficiarse de un análisis de la estructura genómica poblacional utilizando técnicas genómicas modernas para proporcionar información sobre las poblaciones de abulón azul en varias escalas espaciales. Aunado a los problemas de escala en el manejo, esta pesquería se ve afectada por pesca ilegal y el falso etiquetado en productos alimenticios, por lo que surge la necesidad de desarrollar herramientas genómicas útiles para discriminar entre poblaciones y origen de productos. Así mismo, los paneles reducidos de SNPs pueden beneficiar a las cooperativas pesqueras ya que les permitiría identificar la procedencia de sus productos aumentando las probabilidades de certificar sus productos u obtener denominación de origen.

Con este trabajo se pretende: 1) Aportar conocimiento acerca de los patrones genéticos de *H. fulgens*, tales como el grado de estructuración y flujo genéticos (intensidad y dirección) entre

localidades; 2) Evaluar la concordancia entre las unidades de manejo actuales del recurso y las unidades biológicas definidas con los análisis genéticos y 3) Diseñar un panel reducido de marcadores SNPs que permitan la asignación precisa de individuos a sus poblaciones de origen. De esta manera, se espera que los resultados obtenidos aporten información valiosa que ayude a delimitar de manera más precisa las fronteras genéticas de los stocks pesqueros para un manejo integral del recurso entre las cuatro zonas administrativas. Además de generar herramientas de vanguardia que permitan apoyar las políticas y decisiones de manejo del recurso.

#### **4. HIPÓTESIS**

Dadas las características biológicas de la especie y las diferencias reportadas en sus características de historia de vida a lo largo del gradiente latitudinal se espera encontrar división poblacional a lo largo de la PBC al incluir individuos de la parte norte de la PBC. Además, se espera que las características costeras (tales como el promontorio de Punta Eugenia) y las condiciones oceanográficas heterogéneas estén moldeando dichos patrones de estructuración.

## 5. OBJETIVOS

### 5.1 Objetivos general

Determinar la estructura genómica neutral de *H. fulgens* en la costa Occidental de la Península de Baja California y evaluar si la distancia geográfica y/o las condiciones oceanográficas pueden explicar los patrones de variabilidad genómica.

### 5.2 Objetivos particulares

1. Determinar la estructura genómica neutral de *H. fulgens* en la costa Occidental de la Península de Baja California.
2. Determinar magnitud y sentido de la conectividad entre localidades.
3. Evaluar la influencia de los factores geográficos en los patrones de variabilidad neutral en *H. fulgens*.
4. Diseñar un panel de SNPs como herramienta de trazabilidad y asignación poblacional.

## 6. MATERIAL Y MÉTODOS

### 6.1 Área de estudio y toma de muestras

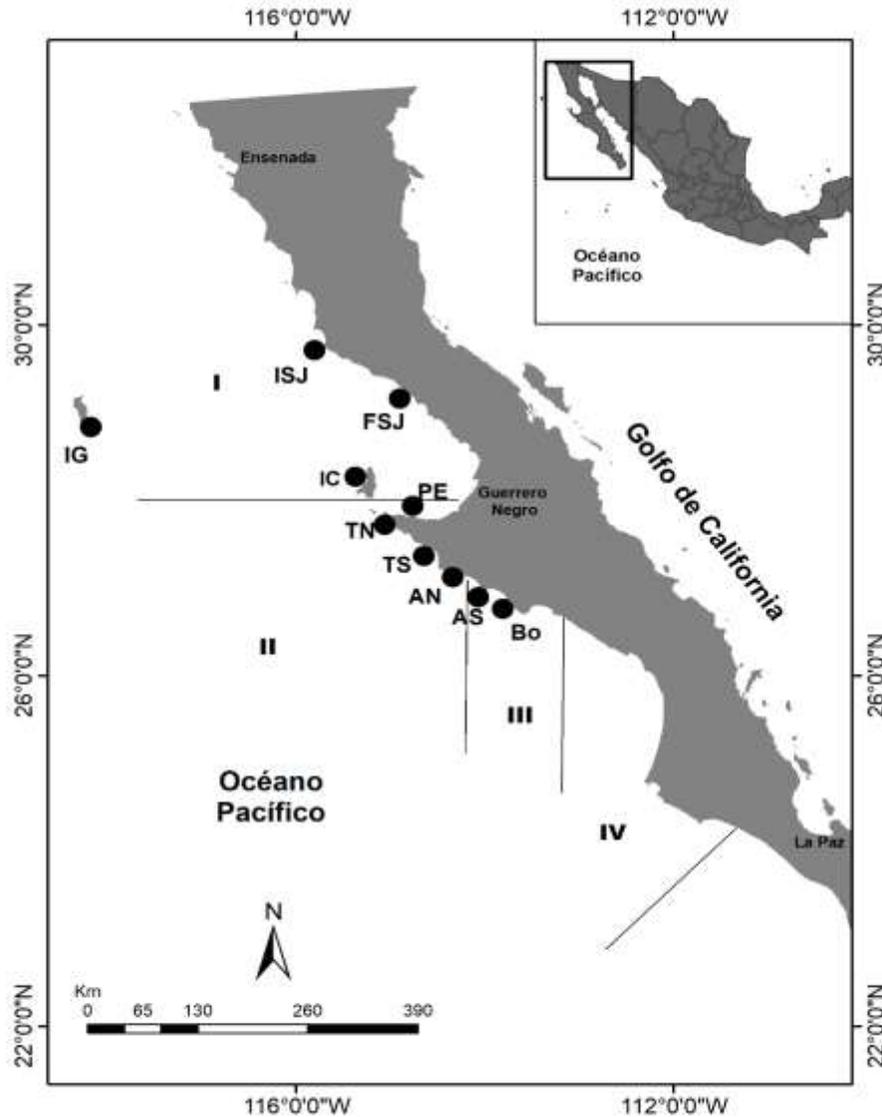
Las muestras de *H. fulgens* se obtuvieron en colaboración con las cooperativas pesqueras de la Península de Baja California (PBC) (Fig. 2). Se realizaron 2 muestreos en 11 cooperativas a lo largo de la PBC durante las temporadas de pesca del 2017 (n= 267) y 2018 (n= 131).

Se colectaron muestras de tejido de un par de epipodios de aproximadamente 30 individuos por localidad y se preservaron en alcohol al 96% para los análisis genéticos. En la Tabla 1 se muestran los nombres de las principales sociedades cooperativas pesqueras de abulón en la región, además se detallan los tamaños de muestra por cooperativa.

**Tabla 1.** Localidades de muestreo de abulón azul (*Haliotis fulgens*) durante las temporadas de pesca del 2017 y 2018 y temporada de evaluación del recurso 2018. Se muestran el ID de las localidades, la localidad y nombre de las Sociedades Cooperativas Pesqueras (SCPP), así como el tamaño de muestra total y por localidad n.

ID	Localidad	SCPP	Fecha de colecta	n
IG	Isla Guadalupe	Abuloneros y Langosteros	2018	31
ISJ	Isla San Jerónimo	El Rosario	2018	21
FSJ	Faro San José	El Rosario	2018	25
IC	Isla Cedros	Pescadores Nacionales de Abulón	2018	54
<b>*IN</b>	Isla Natividad	Buzos y Pescadores	2017	6
PE	Punta Eugenia	La Purísima	2017	49
TN	Tortugas Norte	Bahía Tortugas	2017	40
TS	Tortugas Sur	Emancipación	2017	49
AN	Asunción Norte	California de San Ignacio	2017	25
AS	Asunción Sur	Leyes de Reforma	2017	50
Bo	La Bocana	El Progreso	2017	48
<b>Total</b>				<b>398</b>

\* Los individuos de Isla Natividad fueron removidos de los análisis posteriores debido al bajo número de muestra.



**Figura 2.** Mapa de los sitios de muestreo a lo largo de la Península de Baja California, México. GI, Isla Guadalupe; SJ, Isla San Jerónimo; FSJ, Faro San José; IC, Isla Cedros; PE, Punta Eugenia; TN, Tortugas Norte; TS, Tortugas Sur; AN, Asunción Norte; AS, Asunción Sur; Bo, La Bocana. Las líneas negras y los números romanos indican las cuatro zonas administrativas de la pesquería de abulón: Zona I (Islotes Coronado, B.C. a Punta Malarrimo, B.C.S.), zona II (Punta Malarrimo a Bahía Asunción, B.C.S), zona III (Bahía asunción a Punta Holcomb) y zona IV (Pulta Holcomb a Arroyo El Conejo, B.C.S).

## 6.2 Preparación de librerías genómicas y secuenciación

### 6.2.1 Extracción de ADN genómico

Se extrajo el ADN genómico a partir de los epipodios preservados en etanol al 96% utilizando el método estándar de cloroformo-alcohol isoamílico siguiendo el protocolo de Sambrook *et al.* (1989) y posteriormente todas las muestras fueron tratadas con RNAasa. Se cuantificó el ADN por métodos de espectrofotometría (Nanodorp™ 2000) y fluorometría (Qubit Thermo Scientific®). Se verificó su integridad mediante geles de agarosa-synergel al 1% usando de intercalante SYBER Gold (Thermo Fisher Scientific) y como referencia el marcador de peso molecular 1 kb Plus (Thermo Fisher Scientific); los geles fueron visualizados en un fotodocumentador ChemiDoc™ (BIO-RAD). La concentración de ADN de las muestras se homogeneizó a 50 ng/μL, para los pasos siguientes.

### 6.2.2 Preparación de librerías ddRAD

La preparación de las librerías se llevó a cabo siguiendo el protocolo de Peterson *et al.* (2012) de secuenciación de ADN asociada al sitio de Restricción de doble digestión (ddRAD-seq) con ligeras modificaciones. El Anexo A contiene información detallada del protocolo. De manera breve, de cada individuo se digirieron 500 ng de ADN genómico utilizando dos enzimas de restricción (*EcoRI-HF* y *MspI*, New England Biolabs®). Se realizó la ligación de adaptadores específicos para unirse a los extremos resultantes de la digestión, dichos adaptadores son *P1* que contiene un *barcode* específico para cada individuo (48 diferentes) y un adaptador *P2* que es común. Se incluyeron réplicas de tres individuos para estimar el error de genotipificación. Se realizaron pools equimolares, los cuales consisten en un conjunto de individuos con diferentes *barcodes* (48 individuos por pool), para evitar el efecto de lotes o “batch effect” en inglés, se colocaron en diferentes pools y librerías a los individuos de las mismas localidades. Se seleccionaron tamaños de 376 a 408 pb mediante Pippin Prep (Sage Science). Posterior a la selección de fragmentos, se realizó un PCR con el fin de enriquecer las librerías, agregar un *index* específico a cada pool y los adaptadores de secuenciación Illumina. Se prepararon dos librerías finales, cada una compuesta por dos pools equimolares con 92 individuos en total (con una combinación de 48 *barcodes* y 2 *index*).

Las librerías genómicas se secuenciaron mediante una plataforma de Secuenciación de Nueva Generación Illumina Hi-seq 4000 para producir secuencias pareadas (PE) de 150 pb. La secuenciación fue llevada a cabo por la compañía Novogene (Sacramento, CA).

## 6.3 Análisis bioinformático

### 6.3.1 Limpieza, filtrado y control de calidad de secuencias crudas

Las secuencias obtenidas por Novogene se recibieron en archivos en formato fastq. Se evaluó la calidad de las lecturas crudas utilizando el programa FastQC (Andrews 2010). La limpieza y los análisis posteriores de las secuencias crudas fueron realizados con *STACKS* v2.0 (Catchen *et al.* 2013), el cual fue desarrollado para construir loci a partir de secuencias de lectura corta (por ejemplo, aquellas generadas en la plataforma Illumina), para construir mapas genéticos, realizar estudios de filogeografía y genómica poblacional.

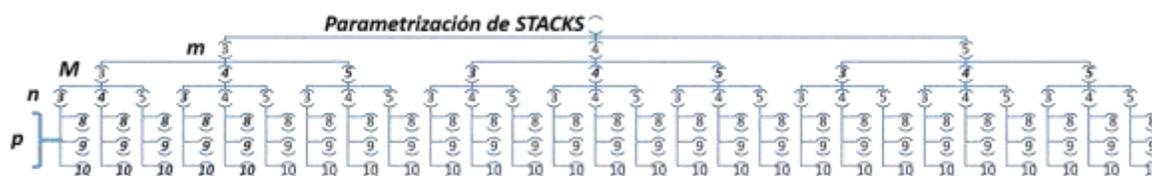
En la primera etapa del pipeline *STACKS* v2.0 se utilizó el módulo *process\_radtags* para demultiplexar las lecturas crudas con los siguientes filtros: *(-q)* Phred33 score >10, *(-t)* trimming 140 pb, eliminación de *barcodes* ambiguos, *(-c)* remoción de bases que no fueron llamadas N, *(-r)* identificación y recuperación de *barcodes*, así como asignación de ID a cada individuo por combinación de *barcode-index*. Por último, se estandarizó el número de lecturas de cada muestra a un millón de lecturas por muestra.

### 6.3.2 Optimización de parámetros, identificación de SNPs y obtención de genotipos de novo

Debido a que el genoma de referencia para esta especie estuvo disponible hasta el último semestre del presente trabajo, los análisis de estructura genética, migración, así como el primer artículo se llevaron a cabo con loci obtenidos *de novo* utilizando el módulo *denovo\_map.pl* en el pipeline *STACKS* v2.0. Este módulo ejecuta de manera consecutiva las etapas principales del pipeline (*Ustacks*, *Cstacks*, *Sstacks*, *tsv2bam* y *Gstacks*) y dentro de ellas existen 3 parámetros principales (*-m*: profundidad mínima de secuencias idénticas para formar un stack, *-M*: controla el número máximo de disimilitudes entre secuencias para considerarlo un locus putativo y *-n*:

controla el número de disimilitudes permitidas entre los stacks formados), los cuales determinan las características de los stacks, el catálogo y obtención de genotipos. El Anexo B contiene una descripción detallada del funcionamiento de los principales módulos del pipeline *STACKS* v2.0 y sus principales parámetros.

Para evaluar el efecto que tiene la modificación de los 3 parámetros principales dentro del módulo *denovo\_map.pl* sobre el número de loci y SNPs obtenidos se realizaron distintas corridas variando uno de los parámetros (Fig. 3). Además, para cada combinación de *-m*, *-M* y *-n*, se llevó cabo un filtrado de los loci con el módulo *populations*, variando el valor para el parámetro *-p* 8, 9 y 10 (número mínimo de localidades en donde cada locus debe estar presente) (Fig. 3), mientras que *r* (% mínimo de individuos en los que cada locus debe estar presente) se dejó fijo a 0.8 y *min\_maf* (frecuencia alélica del alelo menor en todas las localidades) en 0.05. Finalmente, para prevenir loci físicamente ligados se utilizó la opción *write\_single\_SNP* con la finalidad de retener el primer SNP de cada RAD-tag.



**Figura 3.** Pruebas de parametrización. Combinación de parámetros utilizados para la prueba de parametrización en el módulo *denovo\_map.pl* (*-m*, *-M* y *-n*) y *populations* (*-p*) del pipeline *STACKS*.

Se evaluó el porcentaje de genotipos faltantes por individuos con *vcftools* v3.0 (Danecek *et al.*, 2011) utilizando el comando *missing-indv* y se excluyeron de los análisis posteriores aquellos que presentaron más del 10%.

Los archivos *\*.vcf* filtrados se convirtieron a los formatos necesarios para los siguientes análisis utilizando el programa *PGDSPIDER* v2.1.1.5 (Lischer y Excoffier, 2012).

### 6.3.3 Estimación del error de genotipificación

Para evaluar el error de genotipificación se utilizaron 3 réplicas con diferentes *index* y *barcode*; posteriormente, se evaluó el porcentaje de concordancia entre los genotipos de los individuos replicados.

### 6.4 Detección de loci candidatos a selección (*loci* adaptativos)

Con la finalidad de identificar loci candidatos a selección (loci con un nivel de diferenciación poblacional que exceda lo que se esperaría bajo neutralidad) se utilizaron dos métodos independientes, los SNPs que fueron identificados como adaptativos fueron removidos de la base de datos.

Primero se utilizó el programa ARLEQUIN v.3.5.2.2 (Excoffier y Lischer, 2015). Este programa utiliza simulaciones coalescentes para crear una distribución nula de estadísticos  $F$  y genera valores de probabilidad (P) para cada locus basado en su distribución y heterocigosidad a través de los loci (Excoffier *et al.*, 2009). Se consideraron loci adaptativos aquellos que estuvieron sobre el cuantil del 99% de la distribución de  $F_{ST}$  ( $P < 0.01$ ). La corrida se realizó con el modelo finito de islas, 100,000 simulaciones y 1,000 repeticiones por grupo.

Segundo, se utilizó el programa BAYESCAN 2.1 (Foll y Gaggiotti, 2008), un método Bayesiano basado en un modelo de regresión logístico que separa los efectos locus-específicos de la selección (variación genética 'adaptativa') de los efectos poblacionales-específicos de la demografía (variación genética 'neutral') con los siguientes parámetros (-n 10,000; -thin 10; -nbp 20; -pilot 5,000; -burn 50,000; -pr\_odds 100). El criterio de decisión para determinar si era probable que un locus estuviera bajo fuerte selección fue el valor de  $q$  (Foll, 2012) análogo a un valor de  $p$  de FDR que debe estar por debajo de 0.05.

## 6.5 Equilibrio de Hardy-Weinberg

Para determinar si los loci se encontraban en E-HW se utilizó la prueba exacta de Fisher a partir del coeficiente de endogamia ( $F_{IS}$ ) con el programa ARLEQUIN v.3.5.2.2 (Excoffier y Lischer, 2015) con el algoritmo de cadenas de Markov por 1,000,000 cadenas y 100,000 dememorizaciones por locus por localidad. Se removieron loci que se encontraron en desequilibrio en 5 (50%) o más localidades de muestro ( $P < 2.3 \times 10^{-5}$  después de corrección de Bonferroni (BF)).

## 6.6 Desequilibrio de ligamiento

Se estimó desequilibrio de ligamiento entre todos los pares de marcadores utilizando el método de  $r^2$  implementado en plink v1.07 (Purcell *et al.*, 2007). Los loci ligados ( $r^2 > 0.25$ ) fueron descartados para los análisis posteriores.

## 6.7 Estructura genómica poblacional neutral

### 6.7.1 Diferenciación genética poblacional ( $F_{ST}$ )

El grado de diferenciación entre pares de localidades se cuantificó utilizando el estimador imparcial  $F_{ST}$  (Weir y Cockerham, 1984) el cual mide la variación de las frecuencias alélicas entre poblaciones mediante la comparación de las frecuencias alélicas de los individuos de una población con respecto a toda la población y la medición de la varianza de las frecuencias alélicas entre poblaciones. Los valores de  $F_{ST}$  pueden variar entre 0 y 1 en donde valores cercanos a 0 significa poca o nula diferenciación genética mientras que los cercanos a 1 significan fijación total, es decir total diferenciación genética. Para evaluar si las estimaciones de  $F_{ST}$  difirieron significativamente de cero, se realizó un proceso de 10,000 permutaciones con un valor de significancia de  $P \leq 0.05$ , con el programa ARLEQUIN v3.5.2.2 (Excoffier y Lischer, 2015). Los valores de P se ajustaron realizando una corrección de Bonferroni de pruebas múltiples (Rice, 1989). El valor global de  $F_{ST}$  entre los sitios de muestreo se calculó con la función *basic.stats* del paquete *hierfstat* v0.04-22 (Goudet y Jombart, 2015) en R.

### 6.7.2 Análisis Discriminante de Componentes Principales (DAPC)

Se realizaron análisis discriminantes de componentes principales (DAPC por sus siglas en inglés) utilizando el paquete *adegenet* v.2.1.1 (Jombart, 2008) en *R*, el cual está diseñado para determinar la estructura genética de las poblaciones maximizando las diferencias entre grupos mientras que minimiza la variación dentro de los grupos sin información previa sobre las poblaciones. Para identificar el número óptimo de componente principales que dan lugar al modelo con la capacidad predictiva más alta se utilizó la función *optim.a.score*. Se utilizó la función *find.clusters* con el método del criterio de información Bayesiano (BIC) para identificar el número de *clusters* o grupos en función del número de componentes retenidos sugeridos por *optim.a.score*.

### 6.7.3 Análisis de clusters

Se utilizó el programa *STRUCTURE* v.2.3.4 (Pritchard *et al.*, 2000) el cual es un algoritmo para inferir el mejor valor de  $K$  (número putativo de grupos genéticos) a partir de datos genotípicos no ligados. Asume un modelo en el cual hay  $K$  poblaciones (en donde  $K$  se desconoce) y cada una está caracterizada por un set de frecuencias alélicas para cada locus. Los individuos son asignados (probabilísticamente) a los grupos, o uno dos o más grupos si sus genotipos indican que están mezcladas mediante un índice de membresía. El modelo no asume algún modelo de mutación en particular y supone que los marcadores genéticos no están en desequilibrio de ligamiento y están en Equilibrio de Hardy-Weinberg (*E-HW*).

Se utilizó el modelo de mezcla (*admixture*) puesto que permite que los individuos tengan múltiples orígenes poblacionales. El rango de posibles grupos ( $K$ ) fue de entre 1 y 10 y para cada  $K$  se repitió 5 veces, con un proceso de calentamiento de 10,000 iteraciones y una longitud de iteraciones de cadenas de Markov Monte Carlo (MCMC) 100,000 iteraciones. Con los resultados obtenidos se utilizó el programa *STRUCTURE HARVESTER* v0.6.94 (Earl y VonHoldt, 2012) (<http://taylor0.biology.ucla.edu/structureHarvester/>) para evaluar la probabilidad acumulada y

la probabilidad posterior para cada valor de  $K$ , e inferir el número de grupos para datos simulados, empleando el estadístico de  $\Delta K$  de Evanno (Evanno *et al.*, 2005).

Después de determinar el número de  $K$  más probable, se utilizó el programa CLUMMP v1.1.2 el cual promedia las réplicas de  $K$  más probable y define los grupos genéticos (Jakobsson y Rosenberg, 2007), posteriormente se graficaron utilizando el programa DISTRUCT v1.1 (Rosenberg, 2004).

#### **6.7.4 Análisis de Varianza Molecular (AMOVA)**

Se realizaron análisis de varianza jerarquizado de las frecuencias alélicas con una matriz de distancias para determinar la cantidad de variabilidad genética contenida dentro de los subgrupos, dentro grupos y entre ellos en el programa ARLEQUIN v.3.5.2.2 (Excoffier y Lischer, 2015). Esto permite poner a prueba hipótesis de las potenciales agrupaciones jerárquicas: (1) zonas administrativas como grupos genéticos; (2) dos grupos basados en los resultados de STRUCTURE; (3) tres grupos basados en los resultados de  $F_{ST}$ ; (4) tres grupos basados en los resultados de DAPC y (5) cuatro grupos basados en los resultados visuales del DAPC.

#### **6.8 Análisis de diversidad genética**

Para estimar la diversidad genómica en términos de riqueza alélica, riqueza alélica rarificada, heterocigosidad observada ( $H_O$ ) y esperada ( $H_E$ ) así como el coeficiente de endogamia ( $F_{IS}$ ) se utilizó el paquete *PopGenReport* v3.4.0.4 (Adamack y Gruber, 2014) en *R*. Se realizaron pruebas de Kruskal-Wallis y una prueba a posteriori de Gao para examinar las diferencias entre los valores de diversidad genética dentro de las localidades con el paquete *stats* y *nparcomp* v3.0 (Konietschke *et al.*, 2015) en *R*.

## 6.9 Tasas de migración relativa

Para estimar la tasa relativa de migración direccional entre localidades se generaron redes de migración utilizando la función *divMigrate* (Sundqvist *et al.*, 2016) del paquete *diveRsity* en *R* (Keenan *et al.*, 2013) con *Gst* (Nei, 1973) como una medida de distancia genética con un proceso de calentamiento de 1000 iteraciones y un umbral de filtrado arbitrario de 0.6. Los patrones de flujo genético se visualizaron utilizando gráficos producidos con el paquete *qgrph* en *R* (Epskamp *et al.*, 2012). *divMigrate* identifica poblaciones putativas de fuente y sumidero ya que calcula los niveles relativos de migración evaluando la diferenciación genética entre dos poblaciones y un grupo hipotético de migrantes.

## 6.10 Relación entre estructura genómica y distancia geográfica

Se evaluó la hipótesis de asilamiento por distancia (IBD por sus siglas en inglés) mediante una prueba de Mantel (Mantel, 1967) con el paquete *adegenet* v2.1.1 (Jombart *et al.*, 2010) en *R* con la finalidad de evaluar si el patrón espacial de variación genética neutral se relaciona con la distancia geográfica. La prueba utiliza una matriz de distancias genéticas ( $F_{ST}$  linealizado de Slatkin  $\frac{F_{ST}}{(1-F_{ST})}$ ) y otra de distancias geográficas (la distancia lineal marítima más corta entre pares de localidades) y realiza una prueba de correlación entre ambas matrices utilizando la función *mantel.randtest* en *R* y una simulación de Monte-Carlo con 999 iteraciones. Las pruebas se realizaron en tres niveles jerárquicos (1) incluyendo todas las localidades; (2) usando sólo las localidades costeras (sin Isla Guadalupe) y (3) entre las localidades costeras sur (excluyendo Isla Guadalupe, Isla San Jerónimo y Faro San José). Para probar si las correlaciones entre las distancias genéticas y geográficas eran el resultado de una diferenciación genética continua o una clina por parches, se aplicó un estimador de densidad Kernel bidimensional con la función *kde2d* del paquete *MASS* (Venables & Ripley, 2002) en *R*.

## 6.11 Identificación de SNPs y obtención de genotipos utilizando genoma de referencia

### 6.11.1 Limpieza, filtrado y control de calidad de secuencias crudas

Recientemente se ensambló y publicó el genoma de *Haliotis fulgens* por lo que se llevó a cabo el llamado de SNPs utilizando el genoma de referencia lo que permite el descubrimiento de un mayor número de SNPs en comparación con el proceso *de novo*. Los SNPs obtenidos con genoma de referencia se utilizaron para realizar el segundo artículo y los análisis de asignación poblacional para identificar un panel reducido de SNPs candidatos para trazabilidad.

Las lecturas crudas (ver sección 6.3.1) fueron demultiplexadas nuevamente en el módulo *process\_radtags* de *STACKS* v2.54 (Catchen *et al.*, 2013) con los siguientes filtros: *(-q)* Phred33 score >30, *(-t)* sin *trimming*, *(-c)* eliminación de *barcodes* ambiguos y remoción de bases que no fueron llamadas N, *(-r)* identificación y recuperación de *barcodes* así como asignación de ID a cada individuo por combinación de *barcode-index*. Para estos análisis no se estandarizó el número de lecturas de cada muestra a un millón de lecturas por muestra.

### 6.11.2 Alineamiento contra genoma de referencia

Las lecturas filtradas y demultiplexadas fueron alineadas con el programa “Burrows-Wheeler Aligner” (BWA) el cual es utilizado para mapear secuencias de baja divergencia contra un genoma de referencia, para ello se utilizó el algoritmo MEM con los parámetros default, este algoritmo se recomienda para secuencias entre 70 pb y 1 Mpb y se utiliza para consultas de alta calidad ya que es más rápido y preciso que los otros dos algoritmos disponibles (Li y Durbin, 2009). El porcentaje medio de lecturas pareadas mapeadas con éxito se estimó con el comando *flagstat* en Samtools (Li *et al.*, 2009). Se aplicaron filtros de calidad estrictos antes del llamado de los loci utilizando Samtools: (1) se removieron alineamientos no mapeados, alternativos y suplementarios y (2) se retuvieron lecturas pareadas con valores de calidad de alineamiento mayores a 30 ( $Q > 30$ ) lo cual significa que la probabilidad de un alineamiento correcto es cercana al 0.999. Los individuos con menos de 1 millón de alineamientos fueron eliminados.

### 6.11.3 Identificación de SNPs y obtención de genotipos con genoma de referencia

La construcción de los loci y el llamado de los SNPs se realizó utilizando el pipeline *ref\_map* en *STACKS* v2.54. Los loci obtenidos fueron filtrados posteriormente con el módulo *populations*. El módulo *populations* se corrió a dos escalas jerárquicas (1) global que incluía todas las localidades y (2) una costera que excluía a Isla Guadalupe con la finalidad de maximizar el potencial de encontrar SNPs informativos para detectar diferencias más marcadas en los individuos de las localidades costeras. Para ambas escalas se retuvieron loci presentes en al menos 80% de los individuos en cada localidad de muestreo ( $-r= 0.8$ ) y en el 100% de los sitios de muestreo ( $-p 9$  y  $10$ ), *min\_maf* (frecuencia alélica del alelo menor en todas las localidades) de 0.05 y se retuvo el primer SNP de cada RAD-tag con la opción *--write\_single\_SNP*. Se evaluó la cantidad de datos faltantes por individuo (*--missing-indv*) y por loci (*--max-missing*) utilizando *vcftools* v3.0 y se eliminaron individuos y loci con más del 15% y 10% de datos faltantes respectivamente. Los archivos \*.vcf filtrados fueron convertidos en los formatos necesarios para los análisis posteriores utilizando PGDSPIDER v2.1.1.5 (Lischer y Excoffier, 2012).

Se probó E-HW con la prueba exacta de Fisher a partir del coeficiente de endogamia ( $F_{IS}$ ) en el programa ARLEQUIN v.3.5.2.2 (Excoffier y Lischer, 2015) con el algoritmo de cadenas de Markov por 1,000,000 cadenas y 100,000 dememorizaciones por locus por localidad. Se removieron loci que se encontraron en desequilibrio en 5 (50%) o más localidades de muestro ( $P < 4.3 \times 10^{-6}$  después de corrección de BF). Finalmente, se estimó desequilibrio de ligamiento por pares entre todos loci utilizando el método de  $r^2$  implementado en *plink* v1.07 (Purcell *et al.*, 2007) y se eliminaron los loci con un valor de  $r^2 > 0.25$ .

## 6.12 Asignación poblacional

### 6.12.1 Selección de SNPs altamente informativos

Para prevenir sesgo en la precisión de asignación y evitar un artefacto en los resultados (Anderson, 2010; Waples, 2010), se dividió a los individuos de manera aleatoria para crear bases de datos de entrenamiento y de prueba. Los datos de prueba se utilizaron exclusivamente para

seleccionar a los loci más informativos mientras que los de entrenamiento se utilizaron para evaluar la precisión de asignación de los individuos a su población de origen.

La selección de los loci más informativos para ambas escalas se agrupó a los individuos de la base de datos de prueba en los tres grupos detectados en los análisis de estructura genética previos (ver sección 6.7). Posteriormente, cada SNP fue jerarquizado utilizando los valores de  $F_{ST}$  obtenidos con *vcftools* v3.0 con la finalidad de maximizar la diferenciación genética entre grupos. Para cada escala se crearon seis paneles de SNPs (50, 100, 200, 300, 400 y 500 SNPs), comenzando con aquel que presentó los valores más altos de  $F_{ST}$ . El valor de  $F_{ST}$  global entre estos tres grupos se calculó utilizando la función *basic.stats* del paquete *hierfstat* v0.04-22 (Goudet y Jombart, 2015) en *R*.

### 6.12.2 Análisis de asignación y prueba de poder in silico

Para evaluar el desempeño de la asignación poblacional teórica de cada panel de SNPs, se probó la precisión de asignación de los individuos a su población de origen con el paquete *assignPOP* v1.1.4 (Chen *et al.*, 2018) en *R*. *assignPOP* divide aleatoriamente el conjunto de datos de entrenamiento (aquellos que no se usaron para clasificar a los loci) en conjuntos de datos independientes de entrenamiento y prueba evitando la introducción de un sesgo de calificación alto (Anderson, 2010; Waples, 2010). Para cada conjunto de datos de entrenamiento, se reduce la dimensionalidad de un subconjunto de características de entrenamiento (genotipos) utilizando Análisis de Componentes Principales. Posteriormente, la salida es usada para construir las funciones de clasificación de aprendizaje automático y asignar a los individuos de prueba a una población de origen. El proceso se puede repetir varias veces mediante un procedimiento de validación cruzada (Monte-Carlo o *K*-fold) donde el que el usuario especifica el número o la proporción de individuos de cada “población” fuente que se utilizará para entrenar el modelo. Los procedimientos de remuestreo de validación cruzada permiten la creación imparcial de conjuntos de datos de entrenamiento “equilibrados” (es decir que tienen el mismo tamaño de muestra) entre las poblaciones de origen, así como una evaluación de su precisión predictiva (Chen *et al.*, 2018).

Para cada uno de los paneles de SNPs, se realizaron procedimientos de validación cruzada de Monte-Carlo con la función *assign.MC* de *assignPOP* con dos proporciones de individuos (0.75 y 0.8) de cada grupo y todos los loci para entrenar los modelos. Cada combinación de entrenamiento y prueba fueron iteradas 100 veces. Los modelos predictivos se construyeron utilizando la función de clasificación “Support Vector Machine” (SVM) ya que ha mostrado ser la que genera precisiones de asignación más altas que las otras funciones disponibles (Chen *et al.*, 2018).

Debido a que la precisión de asignación es mayor cuando se utilizan todos los marcadores disponibles que cuando sólo se usa un panel reducido de SNPs altamente informativos (Anderson, 2010), se realizaron 4 pruebas de asignación utilizando todos los marcadores en ambas escalas: dos usando todos los individuos por grupo y otras dos usando sólo los individuos de entrenamiento, de esta manera se pudo inferir si alguno de los paneles de SNPs llevaba a un sesgo alto de asignación.

### **6.12.3 Análisis de estructura genética poblacional**

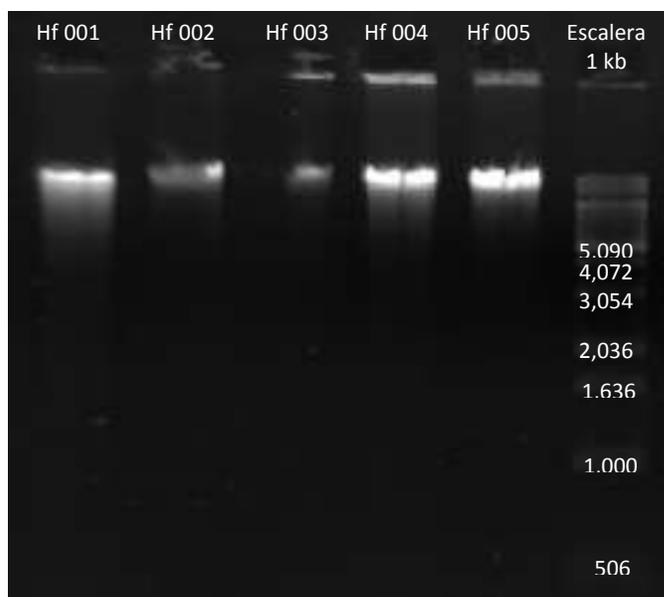
Para asegurar que los paneles de SNPs reflejaran la estructura genética poblacional obtenida *de-novo*, se realizaron DAPCs para ambas escalas usando las bases de datos *de-novo* y con genoma de referencia, así como con cada uno de los paneles de SNPs usados para las pruebas de asignación poblacional con los individuos separados por sitio de muestreo. Estos análisis se realizaron con la función *dapc* del paquete *adegenet* (Jombart *et al.*, 2010) en R. El número óptimo de grupos o *K* se evaluó con la función *find.clusters* con el método de Criterio de Información Bayesiano (BIC).

## 7. RESULTADOS

### 7.1 Preparación de librerías genómicas y secuenciación

#### 7.1.1 Extracción de ADN genómico

Se realizó la extracción de 398 individuos y se obtuvo un promedio de 425 ng/ $\mu$ L. En general, el ADN extraído tuvo alto peso molecular e integridad adecuada (Fig. 4).



**Figura 4.** Muestras de ADN de individuos de *H. fulgens*. Gel de agarosa-Synergel al 1% de las diluciones de trabajo, carril derecho muestra la escalera de 1 kb (Thermo Fisher®). ADN de *H. fulgens* a 250 ng/ $\mu$ L para la elaboración de librerías genómicas.

#### 7.1.2 Preparación de librerías ddRAD

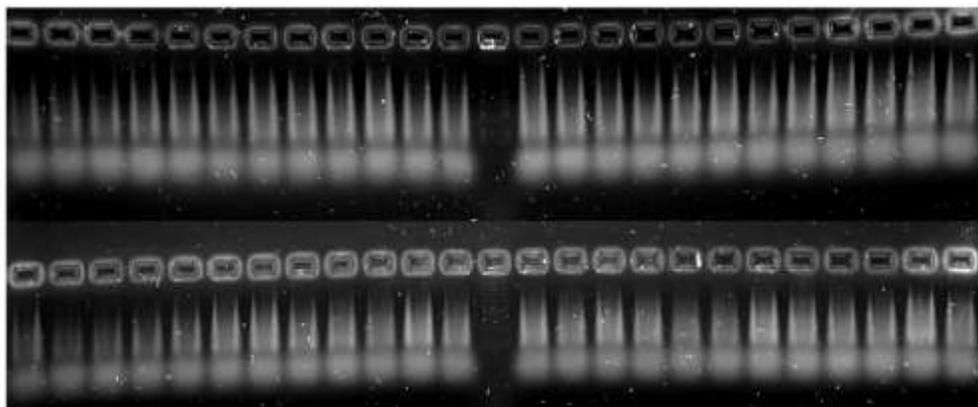
Las librerías se conformaron de individuos divididos en dos placas (Tabla 2), cada placa contenía 96 individuos y dos *index* (uno por cada 48 muestras). Al final se utilizó una combinatoria 4 *index* y 48 *barcodes* (descritos en el protocolo de Peterson *et al.*, 2012).

**Tabla 2.** Abreviaciones de las localidades de muestreo (ID), número de individuos de cada localidad de muestreo utilizados para elaborar las librerías (n), muestras duplicadas para estimar el error de genotipificación (eg).

ID	Localidad	n	eg
IG	Isla Guadalupe	21	1
ISJ	Isla San Jerónimo	14	
FSJ	Faro San José	11	
IC	Isla Cedros	26	1
*IN	Isla Natividad	3	
PE	Punta Eugenia	20	
TN	Bahía Tortugas Norte	20	
TS	Bahía Tortugas Sur	20	
AN	Bahía Asunción Norte	17	
AS	Bahía Asunción Sur	20	1
Bo	La Bocana	20	
<b>Total</b>		<b>192</b>	<b>3</b>

\*Isla Natividad se eliminó de todos los análisis debido al bajo número de muestra.

Posterior a los procesos de digestión y ligación de los adaptadores se realizó una electroforesis en gel de agarosa al 2% para comprobar que la ligación de los adaptadores se hubiera realizado de manera exitosa, lo cual se puede apreciar con el barrido de las muestras (Fig. 5).



**Figura 5.** Productos de ligación de librería genómica. Gel agarosa-Synergel al 2%. Se muestran los productos de ligación después de realizar el PCR1 de la librería genómica, al centro se muestra la escalera 1kb plus (Thermo Fisher®).

La selección de tamaños realizada mediante el Pippin Prep tuvo como objetivo un tamaño de fragmento de 376 pb, comenzando la recolección a los 344 pb y terminando a los 408 pb. Por

último, se comprobó que la selección de tamaños se llevara a cabo de manera correcta mediante el PCR2 el cual sirve para el enriquecimiento de los fragmentos de interés. Finalmente se realizó un gel de agarosa al 2% con los *pools* utilizados para hacer las librerías, se observa una banda nítida menor a 500 pb (Fig. 6).



**Figura 6.** Pools para formar librerías genómicas. Gel agarosa-Synergel al 2%, a la izquierda se muestra la escalera 1kb plus (Thermo Fisher®). Muestra los dos pools antes de formar la librería, cada uno con un index distinto; se observa que su peso molecular está por debajo de las 500 pb.

## 7.2 Análisis bioinformático

### 7.2.1 Limpieza, filtrado y control de calidad de secuencias crudas

Se obtuvieron 1,420,020,148 lecturas crudas totales *pair-end* (100%) provenientes de dos líneas de secuenciación. Después de aplicar los criterios de calidad con el programa *process\_radtags* del *pipeline STACKS* se eliminó el 1.3% y 1.9% debido a que no se reconoció el *barcode* y el sitio de corte RAD respectivamente, además se eliminó el 0.1% de lecturas con baja calidad (Phred33 score <10), quedando en total 1,373,579,489 secuencias *pair-end* (96.8%) con un promedio de 7,458,559 lecturas por muestra (rango: 0.33 M-23.3 M).

Se eliminaron 8 individuos debido a que tuvieron menos de 1,000,000 de lecturas. Los individuos de Isla Natividad fueron eliminados debido a que el tamaño de muestra fue muy pequeño (Tabla 3).

**Tabla 3.** Se muestra el número de individuos por localidad después de eliminar aquellos con <1,000,000 de lecturas, número de lecturas promedio retenidas por localidad obtenidas de novo y número de individuos por localidad después de eliminar aquellos con > 10% de genotipos faltantes.

ID Localidad	n individuos eliminados (<1,000,000 lecturas)	n final ( <i>de novo</i> )	n promedio de lecturas ( <i>de novo</i> )	* n individuos > 10 % de genotipos ( <i>de novo</i> )
GI	0	20	12,349,612	20
SJI	1	13	4,254,695	13
FSJ	0	11	10,109,543	11
CI	1	24	11,148,538	23
PE	0	20	6,689,349	19
TN	3	17	5,782,413	16
TS	0	20	5,905,012	20
AN	2	15	6,256,017	15
AS	0	19	5,731,166	19
Bo	1	19	4,033,914	19
<b>Total</b>	<b>8</b>	<b>178</b>		<b>175</b>

\*Filtro aplicado en la sección 8.2.2.

### 7.2.2 Optimización de parámetros, Identificación de SNPs y obtención de genotipos de novo

Para elegir el set de parámetros a utilizar en el ensamblado de loci y obtención de SNPs se realizaron 81 pruebas modificando un solo parámetro en cada una de ellas (Fig. 3).

Con las pruebas realizadas se decidió utilizar los siguientes parámetros ( $-M\ 3$ ,  $-m\ 5$ ,  $-n\ 4$ ,  $-p\ 10$ ,  $-r\ 0.8$ ) ya que fue con los que se obtuvo el set de datos con mayor número de SNPs (2,216) con un valor de profundidad promedio de  $32\times$  y un nivel de genotipos faltantes menor al 10% (ver Anexo C para mayor detalle sobre los resultados de parametrización).

Del set de datos obtenido con la combinación de parámetros seleccionada, para las siguientes pruebas se eliminaron los individuos que tuvieron un valor de genotipos faltantes mayor al 10% (ver Tabla 3 para detalles de los tamaños de n por localidad después de los filtros).

### **7.2.3 Estimación de error de genotipificación**

Al comparar la concordancia entre los genotipos de los individuos replicados se obtuvo un 0.04% de error.

### **7.3 Detección de loci candidatos a selección (*loci* adaptativos)**

Con el programa ARLEQUIN v.3.5.2.2 (Excoffier y Lischer, 2015) se encontraron 45 loci candidatos a selección. Se consideraron loci adaptativos aquellos que se detectaron por arriba cuantil del 99% de la distribución de  $F_{ST}$  ( $P=0.01$ ). Al utilizar el programa BAYESCAN v2.1 (Foll y Gaggiotti, 2008), se encontraron 5 loci candidatos a selección los cuales fueron compartidos entre ambos métodos. Los loci detectados como adaptativos fueron removidos de la base de datos general para las pruebas posteriores dejando una base de 2,171 loci.

### **7.4 Equilibrio de Hardy-Weinberg (E-HW)**

La prueba de E-HW mostró que 1 locus se encontraba en desequilibrio ( $P \leq 0.05$ ) en cinco o más localidades por lo que fue excluido del set de datos de marcadores neutrales.

### **7.5 Desequilibrio de ligamiento**

Después de la corrección de FDR no se observó evidencia de desequilibrio de ligamiento significativo entre ninguno de los pares de loci dejando una base de datos putativamente neutral de 2,170 SNPs y 175 individuos.

En la Tabla 4 se muestran los SNPs retenidos durante cada filtro, desde el filtrado con el pipeline STACKS hasta el set de datos neutrales utilizados en los análisis.

**Tabla 4.** Resumen del número de SNPs obtenidos de novo retenidos después de cada filtro para *Haliotis fulgens*.

<b>Paso de filtrado</b>	<b># de SNPs</b>
<b>STACKS</b>	
<i>denovo_map</i> (-M 3 -m 5 -n 4)	353,332
<i>Population</i> Presentes en al menos el 80% de los individuos (-r 0.08) en el 100% de las poblaciones (-p 10), Frecuencia alélica del alelo menor (-min_maf 0.05)	2,216
<b>Detección de loci candidatos a selección (loci adaptativos)</b>	
<i>Loci</i> adaptativos de Arlequin	45
<i>Loci</i> adaptativos de Bayescan	5
<i>Loci</i> adaptativos compartidos por ambos métodos	45
<i>Loci</i> en desequilibrio de Hardy-Weinberg (> 50% de las poblaciones)	1
<i>Loci</i> neutrales	2,170

## 7.6 Estructura genómica poblacional neutral

### 7.6.1 Diferenciación genética poblacional ( $F_{ST}$ )

En la Tabla 5, en donde las localidades están ordenadas de norte a sur, se observa que los valores del estadístico  $F_{ST}$  variaron desde 0 hasta 0.021, con un  $F_{ST}$  global de 0.005 estadísticamente significativo ( $P= 0.006$ ).

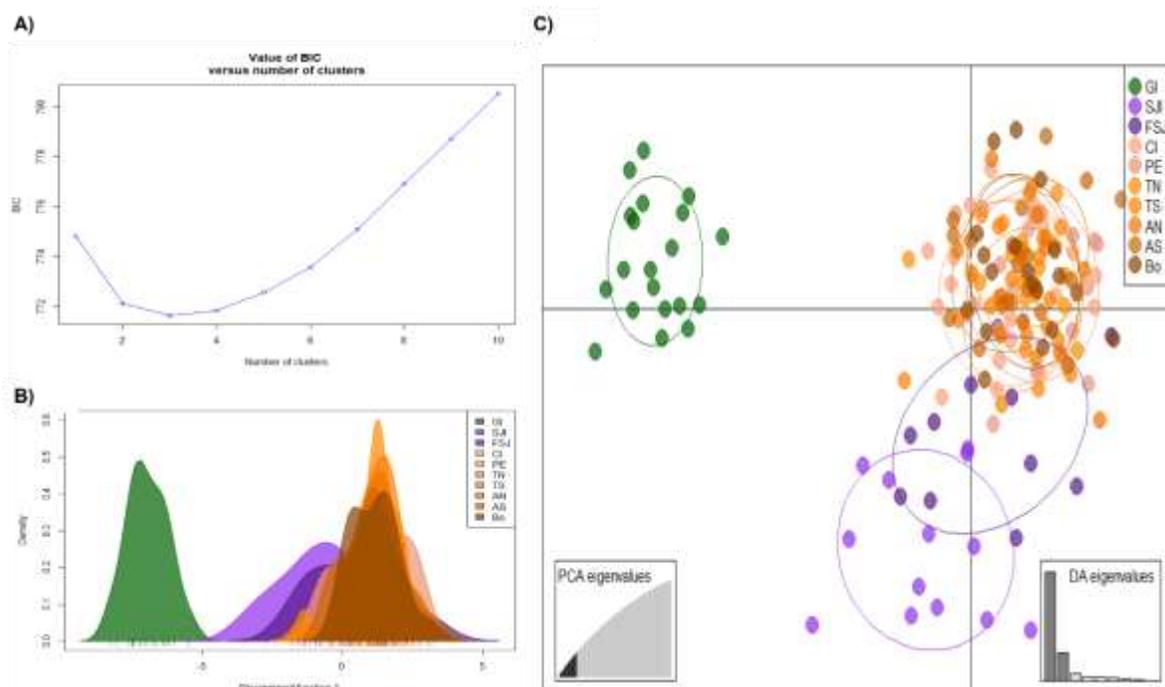
Se observa que sólo las localidades geográficamente más alejadas mostraron diferencias significativas. Isla Guadalupe mostró valores significativos de  $F_{ST}$  ( $P < 0.001$ ) con respecto a todas las localidades. También se observa que la localidad más norteña costera Isla San Jerónimo muestra diferencias significativas con todas las localidades ( $P < 0.001$ ), excepto con Faro San José, aunque el valor de diferenciación entre estas dos últimas localidades está cercano al umbral para considerarse significativo ( $P= 0.002$ ). Las localidades dentro de la región costera sur no mostraron valores de diferenciación genética significativos excepto por las localidades de Faro San José con Tortugas Sur ( $P < 0.001$ ).

**Tabla 5.** Valores de  $F_{ST}$  (Weir y Cockerman, 1984) para comparaciones pareadas por localidades de muestreo utilizando el set de 2,170 SNPs neutrales en 175 individuos de *H. fulgens* pertenecientes a 10 localidades (triángulo inferior). Valores de P para valores pareados de  $F_{ST}$  (triángulo superior). En negrita se muestran los valores significativos ( $P \leq 0.001$  después de la corrección de BF).

	<i>GI</i>	<i>SJI</i>	<i>FSJ</i>	<i>CI</i>	<i>PE</i>	<i>TN</i>	<i>TS</i>	<i>AN</i>	<i>AS</i>	<i>Bo</i>
<i>GI</i>	0	<b>0.000</b>								
<i>SJI</i>	<b>0.018</b>	0	0.002	<b>0.000</b>						
<i>FSJ</i>	<b>0.021</b>	0.007	0	0.025	0.012	0.005	<b>0.000</b>	0.009	0.010	0.012
<i>CI</i>	<b>0.015</b>	<b>0.007</b>	0.004	0	0.783	0.303	0.803	0.882	0.104	0.960
<i>PE</i>	<b>0.017</b>	<b>0.008</b>	0.004	-0.001	0	0.787	0.600	0.793	0.394	0.945
<i>TN</i>	<b>0.019</b>	<b>0.010</b>	0.006	0.001	-0.001	0	0.017	0.400	0.048	0.225
<i>TS</i>	<b>0.018</b>	<b>0.008</b>	<b>0.006</b>	-0.001	0.000	0.003	0	0.098	0.287	0.131
<i>AN</i>	<b>0.017</b>	<b>0.011</b>	0.005	-0.001	-0.001	0.000	0.002	0	0.010	0.157
<i>AS</i>	<b>0.017</b>	<b>0.012</b>	0.004	0.002	0.000	0.003	0.001	0.003	0	0.024
<i>Bo</i>	<b>0.016</b>	<b>0.010</b>	0.004	-0.002	-0.002	0.001	0.001	0.001	0.002	0

### 7.6.2 Análisis Discriminante de Componentes Principales (DAPC)

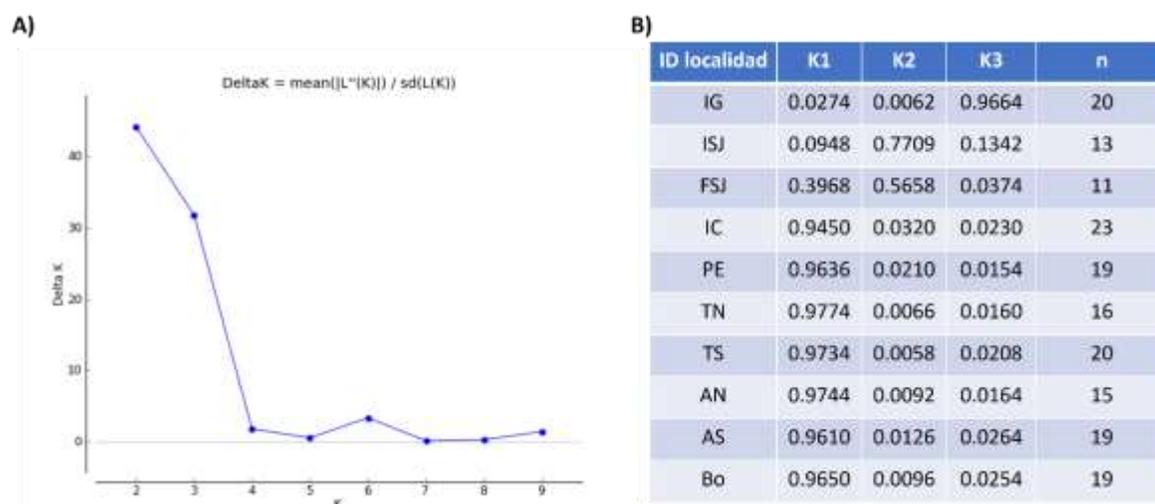
La visualización de la estructura poblacional utilizando un DAPC con el set de marcadores neutrales mostró una agrupación similar a la encontrada con las pruebas de diferenciación genética pareadas ( $F_{ST}$ ) en donde se aprecian 3 grupos genéticos distintos. En la Fig. 7 se observa que la localidad de Isla Guadalupe se separa completamente del resto de las localidades, al igual que Isla San Jerónimo y Faro San José con un poco de mezcla de las localidades pegadas a la costa, sin embargo, se observa una mezcla entre el resto de las localidades. La variación está explicada en los primeros 28 componentes principales y se guardaron las primeras 3 funciones discriminantes.



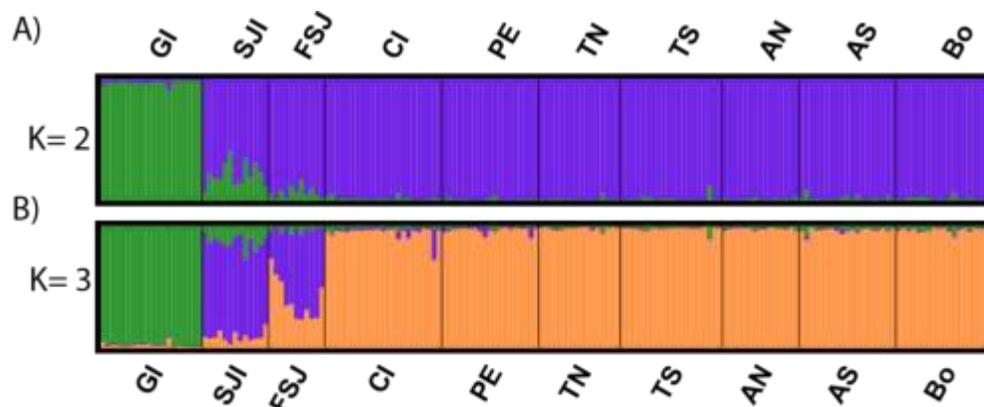
**Figura 7.** Análisis Discriminante de Componentes Principales (DAPC) utilizando un set de 2,170 SNPs neutrales en 175 individuos de *H. fulgens* pertenecientes a 10 localidades. **A)** Valores del Criterio de Información Bayesiano (BIC) versus el número de grupos ( $K$ ). **B)** Gráfica de la distribución poblacional basada en la primera función discriminante. **C)** Gráfica de DAPC, los puntos representan los individuos con colores que indican una localidad de muestreo y la inclusión de la elipse del 95 % de probabilidad de pertenencia al clúster. Las leyendas están ordenadas de norte a sur GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.

### 7.6.3 Análisis de clusters

El análisis de clusters realizado con el set de marcadores neutrales (2,170 SNPs) mostró que el valor más probable de  $K$  es 2 mediante el método de Evanno a partir del valor de  $\Delta K$  (Fig. 8 A). Sin embargo, una inspección visual de  $K$  3 mostró una subdivisión adicional con altas probabilidades de asignación de los individuos (>90% en algunos casos) (Fig. 8B) dando los mismos patrones obtenidos por los dos análisis anteriores ( $F_{ST}$  y DAPC) en donde Isla Guadalupe está separada de las otras localidades, seguido de Isla San Jerónimo y Faro San José, mientras que el resto de las localidades forman un grupo con mayor mezcla genética entre ellos (Fig. 9 B).



**Figura 8.** Numero de grupo más probable (K) y probabilidades de pertenencia. **A)** Número de grupos (K) más probable encontrados con el método de Evanno et al. (2005) a partir del valor de  $\Delta K$  utilizando un set de 2,170 SNPs neutrales en 175 individuos de *H. fulgens* pertenecientes a 10 localidades de muestreo. **B)** Probabilidades de pertenencia de cada individuo a los 3 grupos.



**Figura 9.** Coeficientes de asignación para **A)** K=2 y **B)** K=3 como grupo genéticos distintos más probables. Cada línea vertical representa el genotipo determinado por las frecuencias alélicas para cada individuo y cada color corresponde a la proporción de probabilidad de asignación de cada grupo. El eje horizontal superior e inferior corresponde a las abreviaciones de las 10 localidades de muestreo ordenadas de norte a sur GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.

#### 7.6.4 Análisis de Varianza Molecular (AMOVA)

Las pruebas de AMOVA jerárquico en donde sólo las agrupaciones 3 (GI, SJI, el resto de las localidades), 4 (GI, SJI + FSJ, el resto de las localidades) y 5 (GI, SJI, FSJ, el resto de las localidades) resultaron en una variación significativa entre grupos ( $F_{CT} = 0.0104 - 0.0148$ ,  $P <$

0.05). Sin embargo, sólo la última agrupación (5), la cual considera a GI, SJI, FSJ y el resto de las localidades como grupos separados, mostró una diferenciación entre poblaciones dentro de grupos no significativa ( $F_{SC} = 0.0005$ ,  $P = 0.22$ ). Cuando se tomaron en cuenta a las zonas administrativas como potenciales agrupaciones regionales (Zona 1= GI, SJI, FSJ, CI; Zona 2= PE, TN, TS, AN y Zona 3= AS, Bo) se observó un valor de variación entre grupos bajo y no significativo ( $F_{CT} = 0.0003$ ,  $P = 0.29$ ) (Tabla 6).

**Tabla 6.** Análisis de Varianza Molecular (AMOVA) con 2,170 loci neutrales en 175 individuos de *H. fulgens*. AMOVA para examinar la partición de la variación genética en cuatro niveles jerárquicos entre grupos regionales de poblaciones identificadas por los análisis previos y con diferentes estructuras hipotéticas. Se muestra la fuente de variación, los grados de libertad (gl), Índices de fijación ( $F_{CT}$ : variación debido a diferencias entre los grupos;  $F_{SC}$ : variación debido a diferencias entre poblaciones dentro de los grupos;  $F_{IS}$ : variación debido a diferencias entre individuos dentro de las poblaciones y  $F_{IT}$ : variación debido a diferencias entre individuos) con su valor de probabilidad ( $P \leq 0.05$ ) y el porcentaje de variación.

Niveles jerárquicos	Gl	Índice de fijación	Variación (%)	P
<b>1) Zonas administrativas</b>				
Entre grupos ( $F_{CT}$ )	2	0.0003	0.03	0.297
Entre poblaciones dentro de grupos ( $F_{SC}$ )	7	0.0055	0.55	<b>0.000</b>
Entre individuos dentro de poblaciones ( $F_{IS}$ )	165	0.0041	0.41	0.387
Entre individuos ( $F_{IT}$ )	175	0.0099	99.01	0.243
<b>2) 2 Grupos (IG, resto de localidades) basado en resultados de STRUCTURE</b>				
Entre grupos ( $F_{CT}$ )	1	0.0148	1.48	0.096
Entre poblaciones dentro de grupos ( $F_{SC}$ )	8	0.0023	0.23	<b>0.000</b>
Entre individuos dentro de poblaciones ( $F_{IS}$ )	165	0.0041	0.41	0.380
Entre individuos ( $F_{IT}$ )	175	0.0212	97.88	0.240
<b>3) 3 Grupos (IG, ISJ, resto de las localidades) basado en resultados de FST</b>				
Entre grupos ( $F_{CT}$ )	2	0.0126	1.26	<b>0.0230</b>
Entre poblaciones dentro de grupos ( $F_{SC}$ )	7	0.0011	0.11	<b>0.0241</b>
Entre individuos dentro de poblaciones ( $F_{IS}$ )	165	0.0041	0.41	0.3745
Entre individuos ( $F_{IT}$ )	175	0.0178	98.22	0.2423
<b>4) 3 Grupos (IG, ISJ + FSJ, resto de las localidades basado en resultados de DAPC</b>				
Entre grupos ( $F_{CT}$ )	2	0.0104	1.04	<b>0.0059</b>
Entre poblaciones dentro de grupos ( $F_{SC}$ )	7	0.0009	0.09	<b>0.0274</b>
Entre individuos dentro de poblaciones ( $F_{IS}$ )	165	0.0041	0.41	0.3979

Entre individuos ( $F_{IT}$ )	175	0.0155	98.45	0.2454
<b>5) 4 Grupos (IG, ISJ, FSJ, resto de las localidades basado en división visual de resultados de DAPC)</b>				
Entre grupos ( $F_{CT}$ )	3	0.0112	1.12	<b>0.0088</b>
Entre poblaciones dentro de grupos ( $F_{SC}$ )	6	0.0005	0.05	0.2227
Entre individuos dentro de poblaciones ( $F_{IS}$ )	165	0.0041	0.41	0.3879
Entre individuos ( $F_{IT}$ )	175	0.0158	98.42	0.2500

Los análisis realizados previamente para detectar estructura poblacional indican la existencia de tres grupos genéticos principales, uno formado por Isla Guadalupe, y dos grupos costeros: el norte o grupo costero Norte (GCN) formado por las localidades de SJI y FSJ y el grupo costero sur (GCS) formado por el resto de las localidades. Además, los resultados del AMOVA sugieren una ligera diferenciación entre los grupos del GCN.

### 7.7 Análisis de diversidad genética

Las estimaciones de diversidad genética global fueron  $H_o = 0.284$ ,  $H_e = 0.300$ ,  $Ra = 1.877$ . Las muestras de la región norte (GI, SJI y FSJ) fueron las menos diversas, mientras que PE, CI y TS fueron los sitios con mayor diversidad (Tabla 7). Se observaron diferencias significativas en los valores de diversidad genética entre las localidades, siendo Isla Guadalupe e Isla San Jerónimo los sitios con más diferencias significativas por pares (Anexo D).

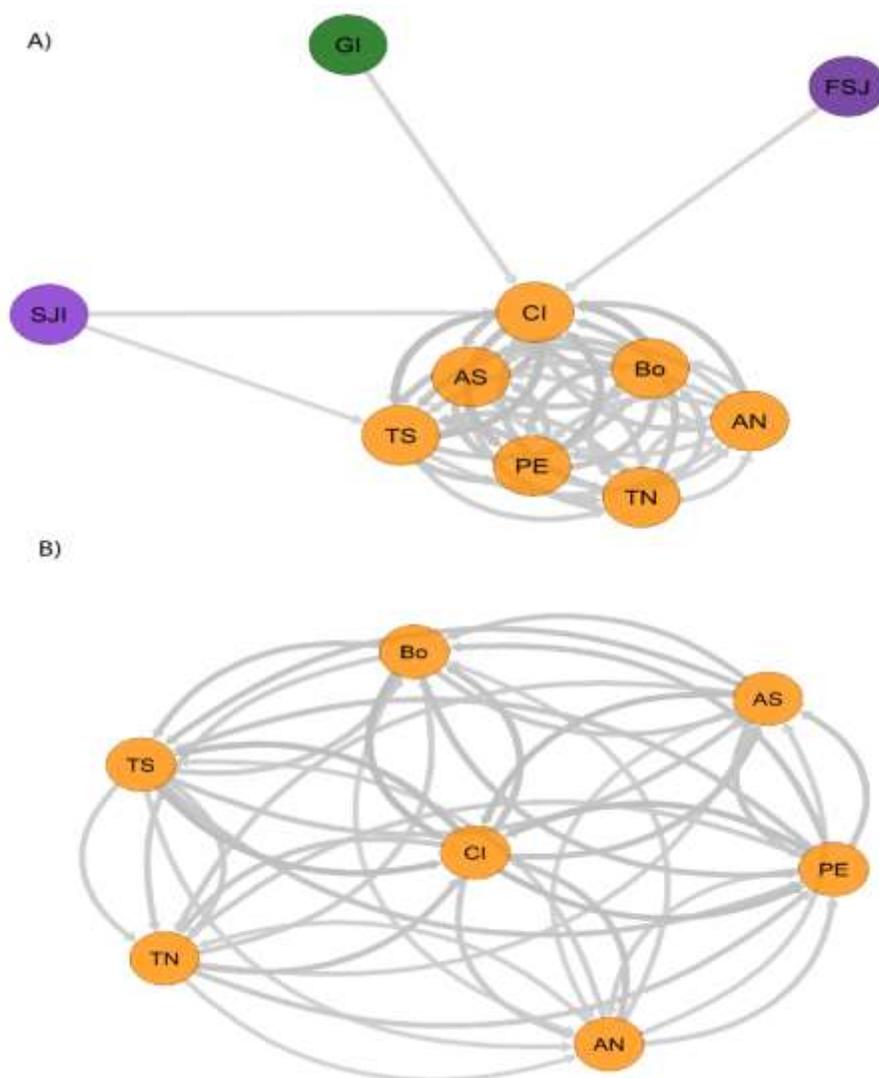
**Tabla 7.** Tabla resumen de la diversidad genética para el set de marcadores neutrales (2,170 SNPs) en 175 individuos de *H. fulgens* pertenecientes a 10 localidades. Se muestran los valores de heterocigosidad observada ( $H_o$ ) y esperada ( $H_e$ ) para cada localidad y global. Valores del coeficiente de endogamia ( $F_{IS}$ ) para cada localidad y el promedio.

<b>Localidad</b>	<b>ID</b>	<b><i>Ra</i></b>	<b><i>Ho</i></b>	<b><i>He</i></b>	<b><i>F<sub>IS</sub></i>*</b>
Isla Guadalupe	GI	1.850	0.272	0.282	0.020
Isla San Jerónimo	SJI	1.847	0.274	0.278	0.003
Faro San José	FSJ	1.860	0.285	0.284	0.008
Isla Cedros	CI	1.895	0.286	0.295	0.017
Punta Eugenia	PE	1.893	0.290	0.295	0.000
Tortugas Norte	TN	1.887	0.286	0.292	-0.005
Tortugas Sur	TS	1.893	0.292	0.296	0.007
Asunción Norte	AN	1.882	0.287	0.292	0.001
Asunción Sur	AS	1.881	0.284	0.291	0.006
Bocana	Bo	1.880	0.286	0.289	-0.019
<b>Global</b>		<b>1.877</b>	<b>0.284</b>	<b>0.300</b>	<b>0.004</b>

\* Ningún valor de  $F_{IS}$  fue estadísticamente significativo.

## 7.8 Tasas de migración relativa

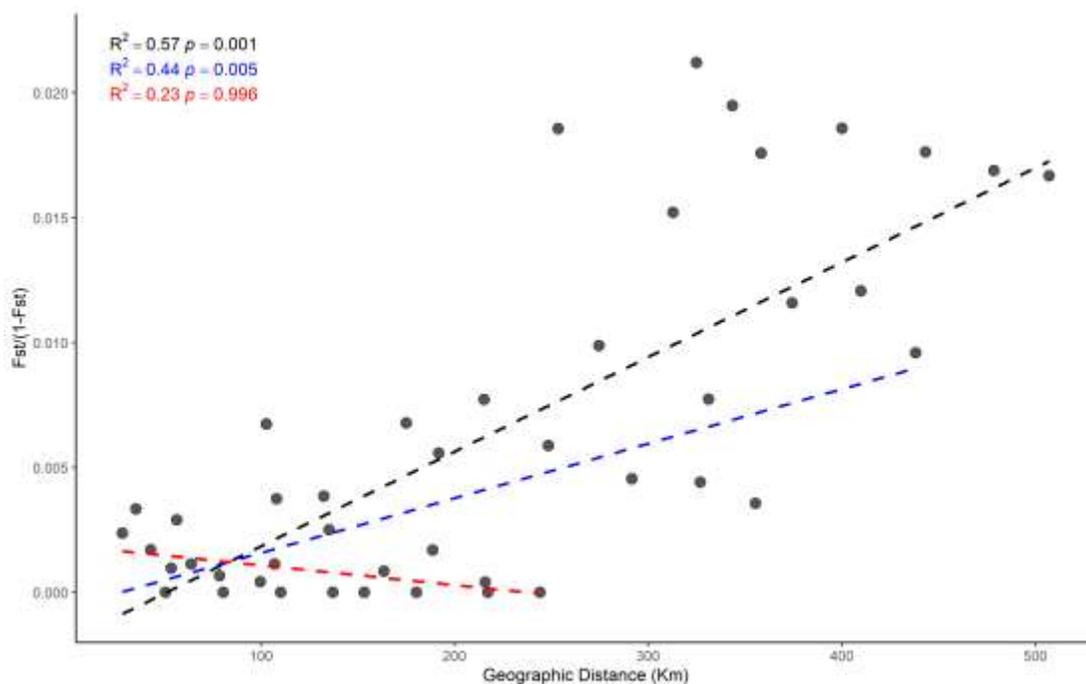
El análisis de migración direccional relativa utilizando *divMigrate* (Fig. 10 A) mostró un flujo predominante de las localidades norteñas (GI, SJI y FSJ) hacia el resto de las localidades. A pesar de la proximidad entre las localidades de la región costera norte (SJI y FSJ) la migración entre estas dos localidades fue más baja en comparación con el resto de las localidades. Una examinación de la conectividad a menor escala entre las localidades de la región costera sur mostró que no existe un flujo direccional aparente (Fig. 10 B).



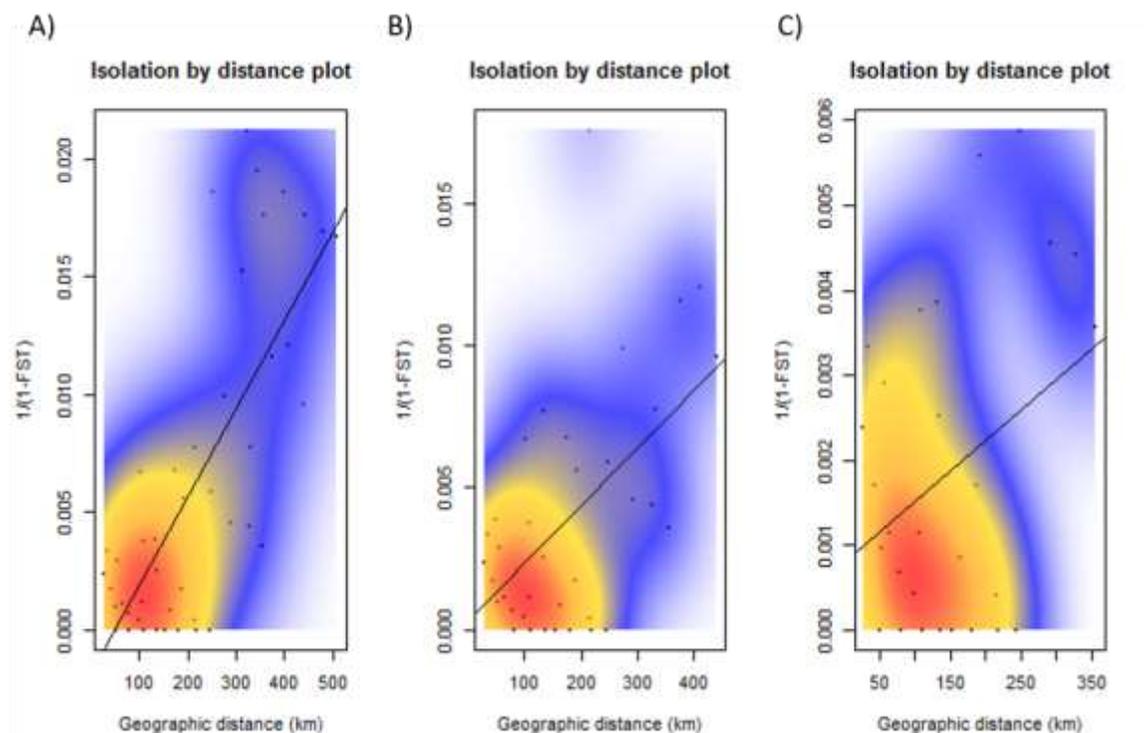
**Figura 10.** Redes de migración relativa para el abulón azul (*Haliotis fulgens*) basadas en 2,170 SNPs neutrales entre 10 localidades de muestreo usando la estimación basada en el  $G_{ST}$  de Nei en el paquete divMigrate, Sólo se presentan las tasas de migración  $> 0.6$ , los círculos representan las localidades, mientras que las flechas indican la dirección y magnitud de los niveles relativos de migración. Las flechas más oscuras indican una relación de migración más fuerte en comparación con las flechas más claras. **A)** todas las localidades y **B)** región costera sur. GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.

### 7.9 Relación entre la estructura genómica neutral y la distancia geográfica

La prueba de Mantel mostró una fuerte relación entre el valor linealizado de  $F_{ST}$  y la distancia geográfica utilizando todos los individuos (Mantel  $r= 0.76$ ,  $P = 0.001$ ) (Fig. 11). El análisis de regresión mostró una relación positiva y lineal ( $R^2= 0.57$ ,  $P < 0.001$ ). Este mismo patrón se observó cuando se eliminó la localidad más alejada (GI) del análisis (Mantel  $r= 0.54$ ,  $P < 0.017$ ;  $R^2= 0.44$ ,  $p=0.005$ ), sin embargo, dejó de ser significativa cuando se evaluó el patrón de aislamiento sólo en la región costera centro-sur (Mantel  $r= 0.34$ ,  $p= 0.129$ ;  $R^2= 0.23$ ,  $p= 0.996$ ). Finalmente, cuando se analizó el patrón de aislamiento con una gráfica de densidad de kernel bi-dimensional utilizando todas las localidades se observó una discontinuidad, lo cual indica un patrón genético parchado entre las poblaciones genéticas (Fig. 12A). Sin embargo, las otras gráficas mostraron un solo parche (Fig. 12B y Fig. 12C). Estos resultados apoyan una estructura poblacional regional a lo largo de la Península de Baja California.



**Figura 11.** Gráficas de aislamiento por distancia de los valores de  $F_{ST}$  linealizados contra la distancia geográfica del abulón azul (*Haliotis fulgens*) basado en 2,170 SNPs neutrales y 10 localidades de muestreo. Línea negra: aislamiento por distancia para todas las localidades, línea azul: aislamiento por distancia sin Isla Guadalupe y línea roja: aislamiento por distancia para las localidades costeras centro-sur.



**Figura 12.** Prueba de Mantel con estimador de densidad de kernel bidimensional. Gráficas de aislamiento por distancia de valores de FST linealizados contra la distancia geográfica por pares de poblaciones para las 10 localidades de *H. fulgens* basada en 2,170 SNPs neutrales A) usando todas las localidades, B) sin Isla Guadalupe y C) usando sólo las localidades costeras de la región centro-sur.

## 7.10 Identificación de SNPs y obtención de genotipos usando genoma de referencia

### 7.10.1 Limpieza, filtrado y control de calidad de secuencias crudas

Del total de 1,420,020,148 lecturas crudas pair-end después de aplicar los filtros de calidad con el pipeline *process\_radtags* en *STACKS* se eliminó el 5.4% y 1.8% debido a que no se reconoció el *barcode* y el sitio de corte RAD respectivamente, además se eliminó el 13.9% de lecturas con baja calidad (Phred33 score <30), quedando en total 1,121,245,268 secuencias *pair-end* (79%) con un promedio de 6,370,712 lecturas por muestra (rango: 1.5 M-19.9 M) (Tabla 8). Se eliminaron 8 individuos debido a que tuvieron menos de 1,000,000 de lecturas.

### 7.10.2 Alineamiento contra genoma de referencia

El mapeo contra el genoma de referencia utilizando BWA-MEM con los parámetros predeterminados retuvo el 99% de las lecturas.

### 7.10.3 Identificación de SNPs y obtención de genotipos con genoma de referencia

El pipeline *ref\_map* en *STACKS* creó un catálogo inicial de 238,246 RAD loci con 484,332 SNPs y una profundidad promedio de 65.5x. El módulo de *populations* utilizando todos los individuos (escala global) generó una base de datos de 10,756 SNPs. Los filtros de loci e individuos por datos faltantes dejaron una base global de 10,523 SNPs con menos del 15% de datos faltantes y 176 individuos con menos del 10% de datos faltantes (Tabla 8).

**Tabla 8.** Se muestra el número de individuos por localidad después de eliminar aquellos con <1,000,000 de lecturas, así como el número de lecturas promedio retenidas por localidad obtenidas con genoma de referencia.

ID Localidad	n individuos eliminados (<1,000,000 lecturas y alineamientos)	n final	n promedio de lecturas (genoma referencia)
GI	0	20	10,705,364
SJI	1	13	3,928,099
FSJ	0	11	8,683,497
CI	1	24	9,471,761
PE	0	20	5,664,779
TN	4	16	5,146,899
TS	0	20	5,004,586
AN	2	15	5,297,097
AS	0	19	4,956,513
Bo	2	18	3,548,006
<b>Total</b>	10	176	

Ningún loci mostró desviaciones significativas del E-HW ( $P < 4.7 \times 10^{-6}$ ) en más del 50% de las localidades. Se eliminaron 569 loci de la base de datos debido a que mostraron desequilibrio de ligamiento ( $r^2 > 0.25$ ) dejando una base global final de 9,954 SNPs y 176 individuos. La escala global incluye los tres grupos genéticos descritos en la sección anterior (ver sección 8.6) y están

constituidos por 1) Isla Guadalupe (GI), 2) Grupo Costero Norte (SJI y FSJ) y 3) Grupo Costero Sur (CI, PE, TN, TS, AN, AS y Bo).

El módulo *populations* a escala costera generó un set de datos inicial de 10,766 SNPs. Ningún loci mostró desviaciones significativas del E-HW ( $P < 4.6 \times 10^{-6}$ ) en más del 50% de las localidades. 897 loci fueron eliminados de la base de datos costera debido a que se encontraron en desequilibrio de ligamiento ( $r^2 > 0.25$ ) dejando una base de datos costera final de 9,869 SNPs y 156 individuos. La escala costera incluye los dos grupos genéticos Grupo Costero Norte (SJI y FSJ) y 3) Grupo Costero Sur (CI, PE, TN, TS, AN, AS y Bo).

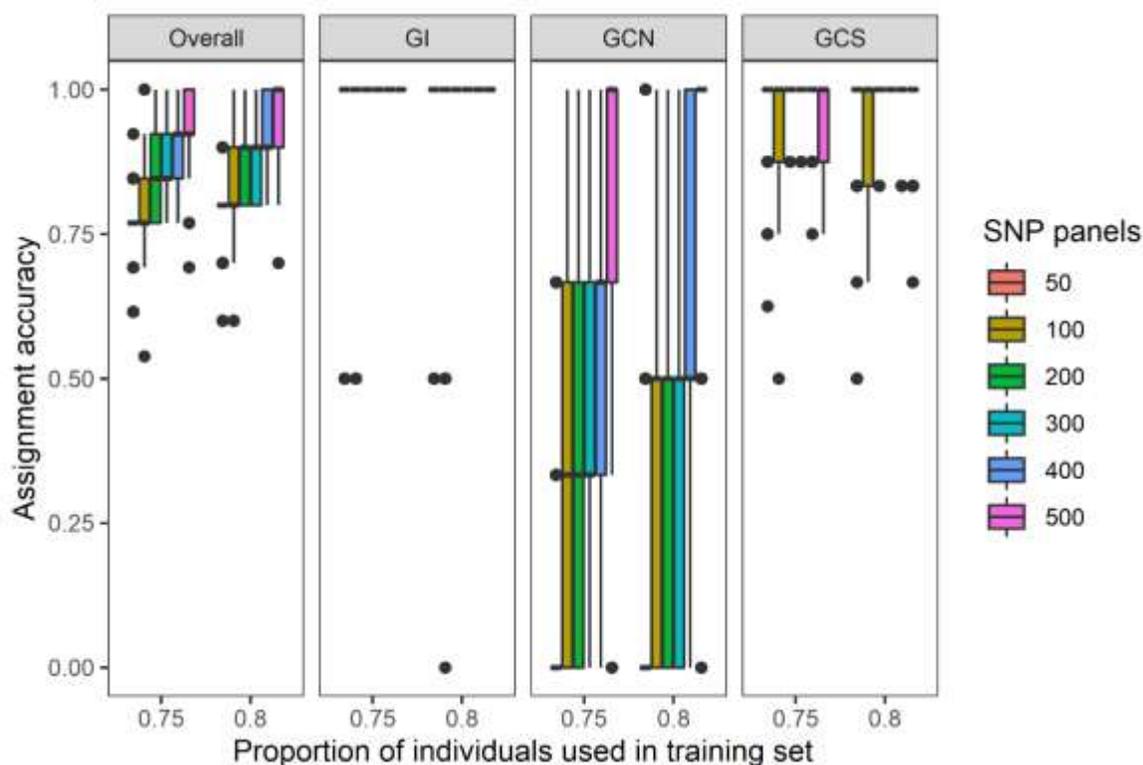
## **7.11 Asignación poblacional**

### **7.11.1 Selección de SNPs altamente informativos**

El  $F_{ST}$  global para los 9,954 loci entre los tres grupos (escala global) Isla Guadalupe (GI), Grupo Costero Norte (GCN) y el Grupo Costero Sur (GCS) fue de 0.011, con valores de  $F_{ST}$  para los 6 paneles de SNPs en un rango desde 0.10 a 0.62. El  $F_{ST}$  global para la escala costera (GCN y GCS) fue de 0.0032, con un rango de 0.095 a 0.53 para los 6 paneles de SNPs probados.

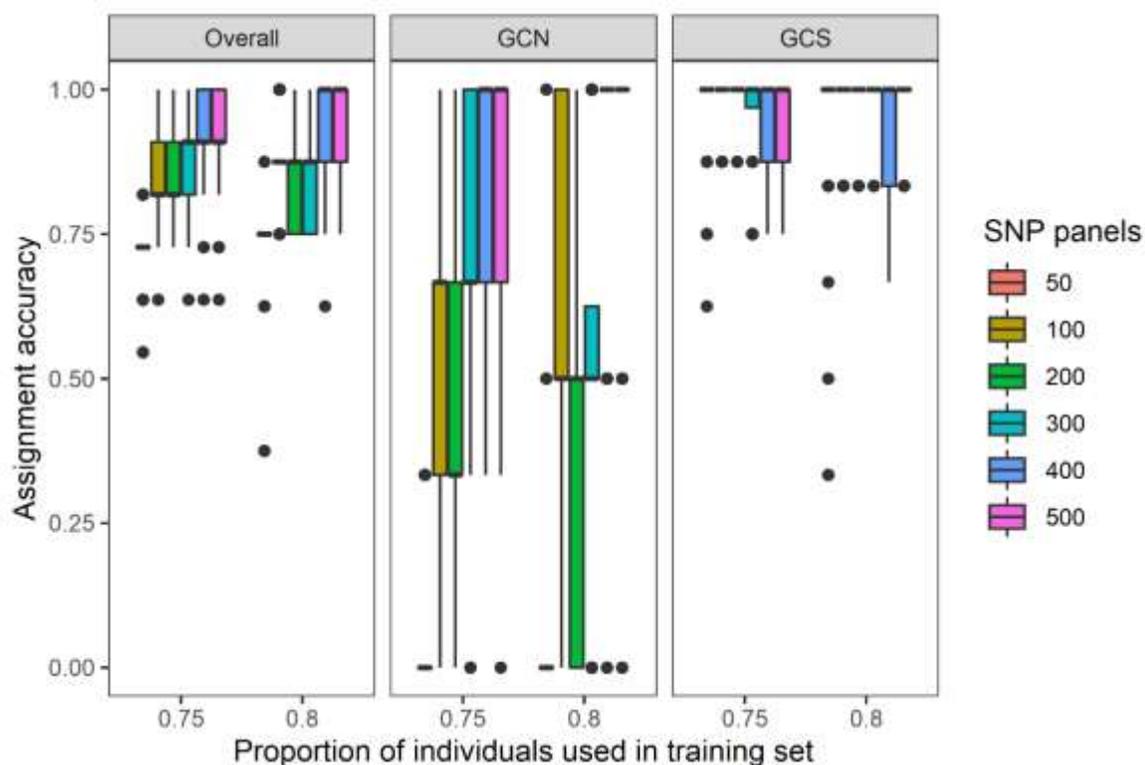
### **7.11.2 Análisis de asignación y prueba de poder**

Los resultados de precisión de asignación a escala global evaluados con el paquete assignPOP fueron similares y mostraron alto poder de discriminación para dos de los tres grupos (GI y GCS) con la mayoría de los paneles de SNPs (Fig. 13), con valores de 0.95, (sd= 0.09) a 1 (Anexo F). Los individuos del GCN mostraron los valores de asignación más bajos yendo desde 0.09 (sd= 0.21), cuando se utilizó el panel de 50 SNPs a 0.88 (sd= 0.24) con el panel de 500 SNPs. En general, al utilizar el 80% de los individuos para entrenar los modelos predictivos se obtienen mayores precisiones de asignación en este análisis (Fig. 13).



**Figura 13.** Resultados del modelo predictivo para la escala global. Precisiones de asignación estimadas vía Monte-Carlo y la función de clasificación SVM (Support Vector Machine), con dos proporciones de individuos para entrenar al modelo (0.75 y 0.8) de la base de datos de entrenamiento ( $n=52$ ) y seis paneles de SNPs (50 a 500 SNPs más diferenciados con base en su valor de  $F_{ST}$ ) representados por diferentes colores. Detalles de la gráfica de cajas: la línea dentro de las cajas es la mediana; los bordes superior e inferior de las cajas son los percentiles 25 y 75; los extremos de los bigotes son los valores mínimos y máximos de los datos no atípicos; los puntos negros son los valores atípicos.

Cuando se quitó a GI de los análisis (escala costera) se obtuvieron valores de precisión similares a los obtenidos con la base de datos global, sin embargo, estos valores aumentaron ligeramente para el GCS (Fig. 14), mostrando valores de 0.97 ( $sd=0.12$ ) desde el primer panel (50 SNPs) con ligeras variaciones entre los paneles desde 0.94 ( $sd=0.08$ ) a 1 (Anexo G). La mejora fue más evidente para el GCN el cual alcanzó precisiones de asignación de 0.91,  $sd=0.23$  al usar el panel de 500 SNPs.

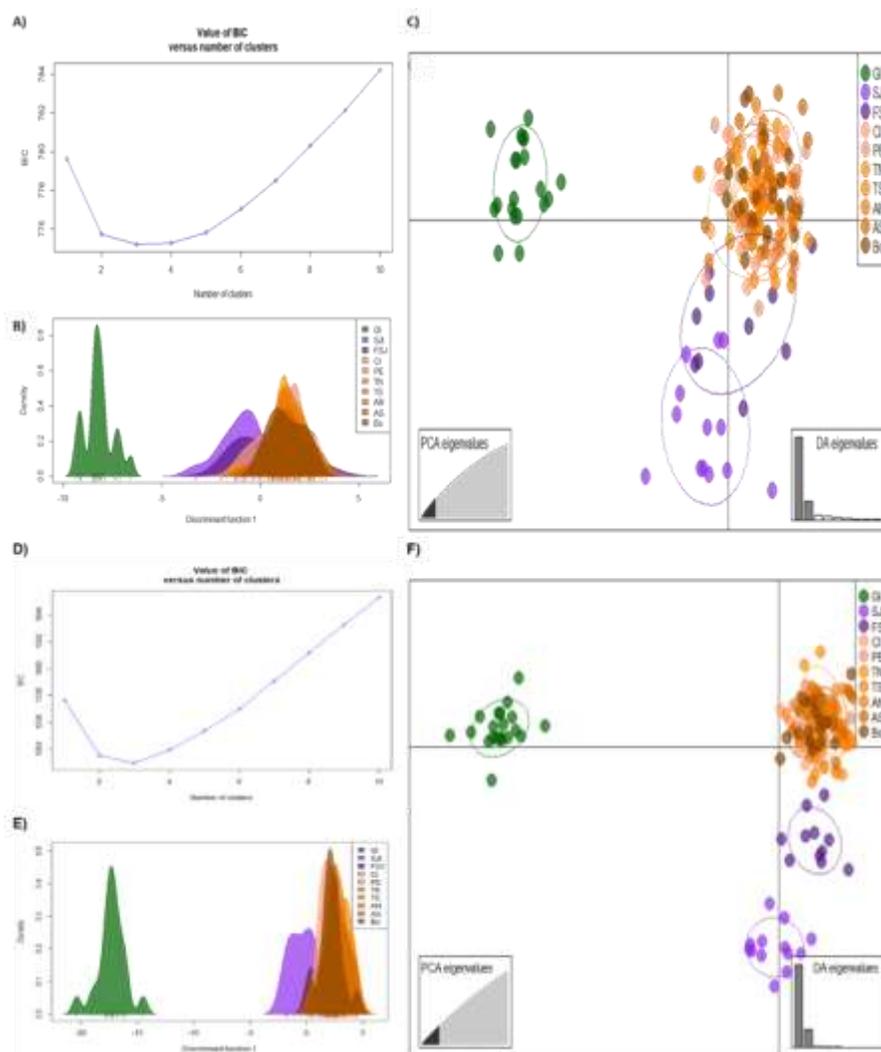


**Figura 14.** Resultados del modelo predictivo para la escala costera. Precisiones de asignación estimadas vía Monte-Carlo y la función de clasificación SVM (Support Vector Machine), con dos proporciones de individuos para entrenar al modelo (0.75 y 0.8) de la base de datos de entrenamiento ( $n=42$ ) y seis paneles de SNPs (50 a 500 SNPs más diferenciados con base en su valor de  $F_{ST}$ ) representados por diferentes colores. Detalles de la gráfica de cajas: la línea dentro de las cajas es la mediana; los bordes superior e inferior de las cajas son los percentiles 25 y 75; los extremos de los bigotes son los valores mínimos y máximos de los datos no atípicos; los puntos negros son los valores atípicos.

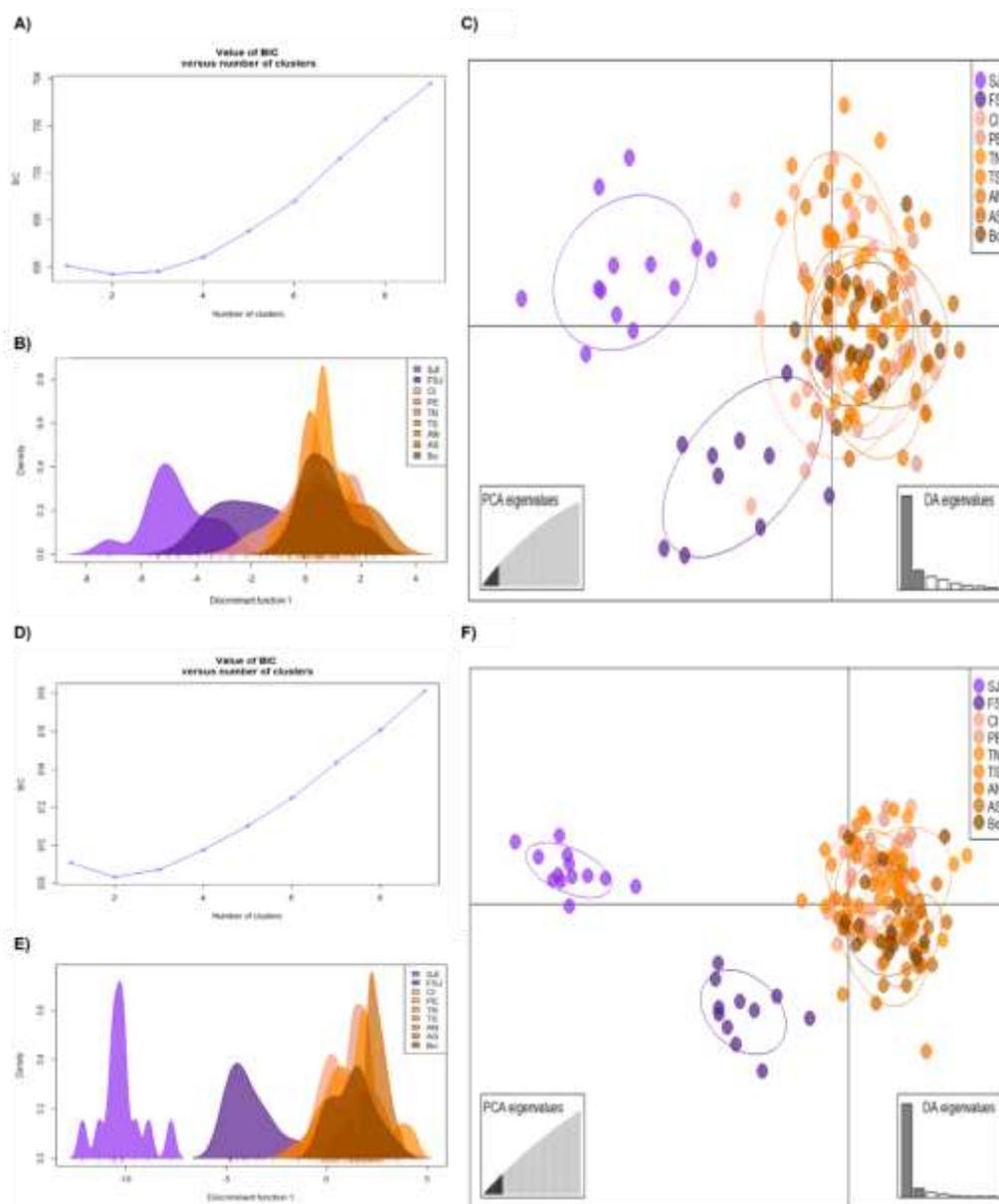
Las pruebas de precisión de asignación utilizando todos los marcadores en ambas escalas usando todos los individuos por grupo y usando sólo los individuos de entrenamiento mostraron valores similares de asignación cuando se utilizó el 80% de los individuos para entrenar los modelos (Anexo H). Nuevamente los individuos del GCN mostraron los valores de asignación más bajos y estos fueron menores a escala costera (0.77,  $sd=0.33$ ) que a escala global (0.89,  $sd=0.25$ ). Cuando se utilizan sólo los individuos de entrenamiento, se obtienen valores del 100% para los grupos GI y GCS en ambas escalas (Anexo H).

### 7.11.3 Análisis de estructura genética poblacional

Los análisis discriminantes de componentes principales (DAPC) utilizando todos los marcadores en ambas escalas tanto con la base de SNPs obtenida *de novo* y la de genoma de referencia mostraron un patrón de estructuración similar al descrito en la sección 8.6, aunque, al utilizar el genoma de referencia la diferenciación entre los grupos costeros SJI y FSI fue más evidente. La Fig. 15 y 16 muestran los resultados de los DAPC utilizando todos los marcadores a escala global y costera respectivamente.



**Figura 15.** Análisis discriminante de componentes principales (DAPC) en *H. fulgens*. Figuras **A**, **B** y **C** muestran los resultados utilizando los 2,216 SNPs derivados de novo a **escala global**. Figuras **D**, **E** y **F** muestran los resultados usando los 9,954 SNPs derivados del alineamiento contra el genoma de referencia del abulón azul a **escala global**. **A**, **D**) Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos (K). **B**, **E**) Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C**, **F**) Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 28 y 35 componentes principales y dos funciones discriminantes (dimensiones) para los datos de novo y con genoma de referencia respectivamente durante los análisis. Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



**Figura 16.** Análisis discriminante de componentes principales (DAPC) en *H. fulgens*. Figuras **A**, **B** y **C** muestran los resultados utilizando los 2,216 SNPs derivados de novo a **escala costera**. Figuras **D**, **E** y **F** muestran los resultados usando los 9,869 SNPs derivados del alineamiento contra el genoma de referencia del abulón azul a **escala costera**. **A**, **D**) Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos (K). **B**, **E**) Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C**, **F**) Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 27 y 30 componentes principales y dos funciones discriminantes (dimensiones) para los datos de novo y con genoma de referencia respectivamente durante los análisis. Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.

En general, los resultados de DAPC utilizando los paneles de SNPs para asignación poblacional no sólo reflejaron la estructura poblacional obtenida con los SNPs *de novo*, sino que la mejoraron al utilizar el panel de 500 SNPs para la escala global (Anexo I) y a partir del panel de 100 SNPs en la escala costera (Anexo J).

## 8. DISCUSIÓN

La presente investigación contó con una cobertura espacial mayor que estudios anteriores, incluye 10 localidades (8 localidades previamente evaluadas y dos más norteñas sin evaluar) pertenecientes a 3 de las cuatro zonas administrativas, abarcando un rango de ~500 km incluidas 3 islas (Isla Guadalupe, e Isla Cedros), siendo el primer estudio que emplea técnicas de secuenciación masiva y escaneo reducido del genoma (ddRAD-seq) en el recurso. Utilizando marcadores neutrales se detectaron tres grupos genéticos integrados por 1) Isla Guadalupe, y la costa de la península dividida en dos grupos 2) Isla San Jerónimo y Faro San José y 3) el resto de las localidades costeras del Sur incluida Isla Cedros. El patrón de estructuración neutral parece estar determinado por la distancia geográfica entre localidades y los patrones de circulación oceánica. De manera general, las unidades biológicas no concuerdan con las unidades administrativas del recurso abulón. Los paneles reducidos de SNPs parecen ser una herramienta prometedora para una mejor aplicación de las normas y vigilancia del recurso.

### 8.1 Preparación de librerías

El uso de ADN degradado en estudios de ecología molecular es muy común, sin embargo, se ha observado que su uso en técnicas de secuenciación asociadas al sitio de restricción tiene impactos negativos en la calidad de las librerías (número de lecturas crudas por individuo, número de sitios variables, porcentaje de secuencias RAD idénticas retenidas) (Puritz *et al.*, 2014; Graham *et al.*, 2015). En el presente estudio las muestras fueron preservadas en etanol al 96% y se realizaron varios recambios durante las primeras 48 horas de captura para asegurar la buena preservación del tejido, además se seleccionaron sólo aquellas muestras con un alto peso molecular e integridad (Fig. 4) y que tuvieran valores de 260/280 arriba de 1.91 y valores de 260/230 menores a 2.30.

Al igual que la calidad del ADN inicial, el proceso durante la preparación de las librerías es de suma importancia para la obtención de marcadores de calidad, por lo que en el presente estudio se realizaron puntos de revisión de calidad durante cada paso.

El número de lecturas obtenidas por individuo varió de 4 M a 12 M. Esta variación es comúnmente observada en estudios de secuenciación de ddRAD (e.g. Ba *et al.*, 2017; Gamboa y Watanabe, 2019; Magris *et al.*, 2022) en donde la preparación de las librerías se hace con muchos individuos en un mismo pool. A pesar de que los pools se tratan de hacer de manera equimolar, siempre existe la posibilidad de un error en el pipeteo. Sin embargo, el número de lecturas por individuo parece no afectar las conclusiones de estructuración en este estudio ya que no se observa gran diferencia cuando se homogenizan las lecturas a 1 M (análisis con datos *de novo*) a cuando se utilizan todas (análisis con genoma de referencia).

## **8.2 Análisis bioinformático**

### **8.2.1 Limpieza, filtrado y control de calidad de secuencias crudas**

Las lecturas crudas obtenidas contaron con una buena calidad, siendo el valor de Phred score mayor a 40 para la mayoría de los casos (lo cual indica que el error por base es de 1 en 10,000 bases o que tuvo una correcta asignación de las bases de un 99.99%). Después de aplicar los criterios de calidad en el programa *process\_radtags* de *STACKS* se retuvieron 96.8% de las lecturas cuando se aplicó un filtro de calidad  $-q > 10$  (error por base de 1 en 10) y el 79% al aplicar un filtro más estricto  $-q > 30$  (error por base de 1 en 1,000) de las lecturas crudas, además se observó una distribución homogénea de las lecturas a lo largo de todos los individuos, esto puede indicar que la preparación de las librerías se llevó a cabo de manera correcta.

### **8.2.2 Optimización de parámetros, Identificación de SNPs y Obtención de genotipos**

Las técnicas de secuenciación de representación-reducida del genoma, como ddRAD-seq han facilitado el cambio en el uso de marcadores moleculares convencionales para hacer inferencias biológicas (e.g. Microsatélites) con grandes sets de marcadores (e.g. SNPs) obtenidos de centenas a millares de loci, sin embargo, este cambio en las técnicas de genotipificación hacia análisis genómicos más complejos también ha traído consigo varios retos. Uno de ellos es la capacidad de manejar correctamente el análisis bioinformático de los datos generados. Como se mencionó anteriormente, en este estudio se utilizó *STACKS* para procesar los datos obtenidos

de la secuenciación y ensamblar los loci *de novo* y utilizando genoma de referencia, sin embargo, se ha reportado que la selección de los principales parámetros que controlan los diferentes programas del pipeline *de novo* determinan las características de los *stacks*, el catálogo y obtención de genotipos.

Algunos autores mencionan que no existe una combinación única de parámetros para realizar los análisis y sugieren que cada estudio debería realizar pruebas de parametrización con su set de datos (Paris *et al.*, 2017). En el presente estudio se realizaron pruebas de optimización con los 3 principales parámetros ( $-m$ ,  $-M$  y  $-n$ ) en el programa *denovo\_map* para evaluar la calidad de los marcadores obtenidos y el efecto que tenían sobre la estructura genética del abulón azul. Se decidió utilizar la combinación  $-m$  5  $-M$  3  $-n$  4, dado que fue con la que más número de marcadores se recuperaba (Fig. 18). Posteriormente se eligieron los loci que estuvieran en el 100% de las localidades  $-p$  10 dado que al realizar los análisis de estructura genómica con los 3 sets de marcadores obtenidos con los diferentes valores de  $-p$  (8, 9 y 10) no se encontraron diferencias, además fue la del valor con el que se mantenía un número alto de marcadores (Fig. 19) y un bajo porcentaje de genotipos faltantes por individuo ( $<10\%$ ) (Fig. 20).

Una de las ventajas de la técnica de ddRAD-seq, es que permiten generar cientos a miles de marcadores sin necesidad de tener un genoma de referencia ( $> 99.9\%$  de las especies, Puritz *et al.*, 2014). Además, la técnica de ddRAD permite maximizar la capacidad de multiplex, es decir, permite incrementar el número de individuos en una misma línea de secuenciación (Peterson *et al.*, 2012). Sin embargo, existe una discrepancia en si es más beneficioso aumentar el número de individuos secuenciados o la profundidad de secuenciación por muestra para estimar la variación genética y definir con precisión la estructura poblacional.

Crawford y Lazzaro (2012), a partir de simulaciones de modelos genéticos poblacionales, reportan que las estimaciones de diferenciación genética poblacional se ven sesgadas cuando utilizan datos con baja profundidad de lecturas ( $<10X$ ). De igual manera, mencionan que el poder para identificar huellas de selección depende de la profundidad de las lecturas y de la fuerza de la selección, ya que si se trata de una selección débil se requieren lecturas con mayor

profundidad para una detección confiable. Paris y *et al.* (2017), recomiendan evaluar la profundidad de las lecturas como primer paso para determinar la confiabilidad de set de datos RAD-seq, ya que se han observado mayores tasas de error de genotipificación cuando se trabaja con profundidades de entre 5X y 10X (Fountain *et al.*, 2016).

Otro reto asociado a los métodos de secuenciación basados en enzimas de restricción es que aún se desconocen las tasas de error de genotipificación. Existen estudios que evalúan el error de genotipificación utilizando ddRAD-seq utilizando genoma de referencia y de-novo, observando que las tasas de error son más sensibles a la cobertura, en donde bajas profundidades llevan a altas tasas de error, particularmente en set de datos obtenidos *de novo* (Fumagalli *et al.*, 2013).

En el presente estudio se utilizaron dos líneas de secuenciación con 96 individuos (equimolarmente) cada una con una combinatoria de 48 *barcodes* y 2 *index* con la finalidad de obtener una mayor profundidad de las lecturas y una buena calidad de los marcadores moleculares. Después de la secuenciación y de aplicar los filtros en el programa *STACKS* se obtuvieron loci con profundidades promedio de 32× cuando se ensamblaron *de novo* y de 65× con genoma de referencia. Finalmente, al evaluar el error de genotipificación se observó que fue constante y bajo (0.04%) en los tres individuos repetidos (datos no mostrados).

El llamado de los loci es mejor cuando se cuenta con un genoma de referencia de la especie o una especie cercana, ya que permite evitar la mezcla de lecturas pertenecientes a loci parálogos al momento de formar el catálogo (Casanova *et al.*, 2021), y reduce el sesgo en las proporciones alélicas debido a errores de mapeo (Olson *et al.* 2015). En el presente estudio, el llamado de loci *de novo* utilizando el módulo *denovo\_map* en *STACKS* generó 116,433 RAD loci mientras que al utilizar el genoma de referencia con el módulo *ref\_map* en *STACKS* se obtuvo una mayor cantidad de RAD loci (238,246) a pesar de que el número de lecturas crudas retenidas en el demultiplexado (debido a que se usó un parámetro de calidad más estricto de  $-q>30$ ) fue mucho menor. Finalmente, el número de SNPs retenidos después de los filtros en el módulo *populations* de *STACKS* fue cuatro veces mayor utilizando el genoma de referencia de la especie.

A pesar de que las conclusiones de estructura genética poblacional neutral no parecen verse afectadas fuertemente con las diferentes bases de datos (*de novo* y con genoma de referencia) en el presente estudio, contar con un mayor número de marcadores permite detectar patrones de diferenciación genética más sutiles como lo observado con los análisis de DAPC a escala costera (Anexo J). Además, un mayor número de marcadores podría permitir detectar más loci adaptativos. Por lo anterior, es necesario realizar futuros estudios que abarquen una mayor área de distribución de la especie, así como realizar una evaluación más profunda sobre los marcadores adaptativos y la influencia de las variables ambientales sobre sus frecuencias en las poblaciones.

Finalmente, con los análisis y filtros anteriores podemos asumir que la calidad de los datos es adecuada e incluso mayor que la reportada en la literatura para estudios de estructuración poblacional.

### **8.3 Estructura genómica poblacional**

Los resultados de estructura poblacional soportan tres grupos genéticos regionales mostrando una clara diferenciación entre la isla oceánica (Isla Guadalupe, ~250 km alejada de la costa) y el resto de las localidades: el segundo grupo está formado por las localidades costeras norteñas, el cual incluye Isla San Jerónimo y Faro San José; dentro de este grupo es posible que exista una sub-estructuración como lo muestran los resultados del AMOVA (Tabla 6); y el tercer grupo incluye las localidades costeras sureñas (Isla Cedros, Punta Eugenia, Tortugas Norte, Tortugas Sur, Asunción Norte, Asunción Sur y la Bocana).

Existen reportes previos sobre la diferenciación genética entre las localidades costeras e isla Guadalupe, así como un patrón de panmixia a lo largo de la costa centro oeste de la PBC (Zúñiga *et al.*, 2000; Gutiérrez-González *et al.*, 2007) los cuales apoyan los resultados obtenidos. El presente estudio amplía estos trabajos previos cubriendo una mayor escala espacial lo que permite detectar estructura poblacional entre la parte norte y sur de la PBC.

En un estudio realizado por Gruenthal *et al.* (2013) se reportan patrones de homogeneidad genética para *H. fulgens* a lo largo del límite norte de la distribución de la especie (sur de la bahía de California, EUA) utilizando un panel de 1,209 SNPs neutrales. Sin embargo, dicho estudio se realizó en un área reducida del rango de distribución de la especie y podrían existir diferencias entre las localidades de Estados Unidos de Norte América y las de México, como lo observado en estudios en *H. corrugata* utilizando microsatélites (Díaz-Viloria *et al.*, 2009; Munguía-Vega *et al.*, 2015). Parece que *H. fulgens* tiene un extenso flujo genético entre las localidades costeras, y es muy probable que la escala espacial de los estudios previos sea un factor importante para concluir panmixia. Basado en los resultados previos, es necesario realizar un muestreo a lo largo de todo el rango de distribución de la especie para capturar una estructura genética más amplia y lograr una mejor comprensión de la biocomplejidad de este recurso pesquero.

En especies marinas semi-sésiles, como el abulón azul, se espera que el principal mecanismo de dispersión se dé durante el estadio de larva pelágica, ya que dependiendo de la duración de esta fase, será en mayor o menor medida el grado de conectividad que puedan tener las poblaciones de dicha especie (Thorson, 1950). En general, en invertebrados marinos, incluidos algunos Haliótidos, se han estimado bajas tasas de dispersión (e.g. 7-10 km en *H. cracherodii*, Gruenthal y Burton, 2008; *Megastrea undosa*, Haupt *et al.*, 2013). La información sobre los patrones de dispersión y conectividad larvaria entre poblaciones locales de *H. fulges* es escasa; Leighton (2000) reporta que en condiciones de laboratorio las larvas pueden durar en la columna de agua entre 4 y 15 días con una temperatura óptima de 20 a 23 °C. Basado en un experimento de deriva de partículas realizado en el sur de la Bahía de California, estimaron que las larvas de *H. fulgens* pueden moverse ~46 km costa abajo en primavera y ~12 km costa arriba durante el otoño (Tegner y Butler, 1985).

Tomando en cuenta la hidrodinámica regional en la costa central de la PBC también se ha planteado la hipótesis de que la dispersión estaría limitada y restringida a distancias cortas debido al debilitamiento de la Corriente de California causado por las camas de kelp y la complejidad de la costa (Guzmán-del Proo *et al.*, 2000). Los resultados en el presente estudio

sugieren que el potencial de dispersión de *H. fulgens*, puede no ser tan limitado como se ha asumido anteriormente (Leighton, 2000; Morgan y Shepherd, 2006; Guzmán- del Proo *et al.*, 2000), particularmente en el centro de la PBC, debido a que el presente trabajo evidenció una amplia conectividad genética en esta zona.

Los patrones de aislamiento por distancia a menudo muestran resultados sesgados cuando existe estructuración jerárquica (Merimans, 2012). Las pruebas de Mantel resultaron en una relación significativa cuando se consideraron todas las localidades, incluso cuando se analizaron sólo las localidades costeras, sin embargo, esta relación no fue significativa dentro del GCS. Teniendo en cuenta los grupos regionales identificados por STRUCTURE y DAPC, el patrón de aislamiento por distancia parece estar sesgado por la estructura poblacional regional. Por lo tanto, el aislamiento por distancia solo juega un papel importante a una escala espacial global, pero no al interior de los grupos regionales. Esta situación sugiere un flujo genético restringido entre los GCN y GCS y altos niveles de flujo genético entre las localidades del centro y el GCS, en lugar de una estructura tipo metapoblacional.

Existen diversos trabajos en el área de estudio que concuerdan parcialmente con los resultados del presente estudio. Haupt *et al.* (2013) realizaron un estudio en el caracol panocha (*M. undosa*), gasterópodo con una distribución y características de historia de vida similares a las del abulón azul (desovador masivo, fase larvaria de 5 a 10 días, edad de madurez a partir de los 4 años). Mediante marcadores de ADNmt encuentran diferenciación genética significativa entre la región norte y sur de la PBC (dividiendo estas regiones a partir de Isla San Martín, 30°N). Si bien, en este estudio no se cuentan con muestras al norte de esa latitud (30°N), la distribución del abulón azul si abarca hasta Bahía Concepción en California. Dado que algunos análisis en este estudio sugieren una ligera sub-estructuración en la región norte de la PBC (sección 7.6.4), y debido a que estas dos localidades tuvieron los menores tamaños de muestra en todo el estudio, es posible que un muestreo más extenso a lo largo de la parte norte de la PBC mostrara patrones de diferenciación genética por lo que es necesario tener en cuenta esto para futuros estudios.

Por otra parte, Haupt *et al.* (2013), al igual que el presente trabajo, observan un patrón de aislamiento por distancia a lo largo de toda el área de estudio. Sin embargo, al evaluar lo que ocurre a nivel regional (norte y sur) el patrón difiere con nuestros resultados, reportando un patrón de aislamiento por distancia en la región sur, no así en la parte norte. Esto probablemente se debe a que incluyeron localidades que se encuentran al norte de Punta Eugenia dentro del grupo de la región sur, aunque también atribuyen estas diferencias a procesos de menor escala de las corrientes marinas.

García-Rodríguez y Pérez-Enríquez (2006), analizan los patrones de estructuración genética en la langosta espinosa (*Panulirus interruptus*) a lo largo de la PBC (desde Ensenada, B.C., hasta Bahía Magdalena B.C.S.) mediante el uso de marcadores de ADNmt. A pesar de ser un invertebrado con un rango de distribución similar al del abulón azul, no encuentran diferenciación genética a lo largo del área de estudio, sin embargo la langosta espinosa tiene una duración larvaria mucho mayor que *H. fulgens* y es probable que el patrón de homogeneidad pueda estar asociado a una mayor dispersión de las larvas mediante las corrientes oceanográficas.

Además de la distancia geográfica, las características costeras tal como los promontorios o prominencias costeras generan zonas de convergencia, sistemas de giros con contracorrientes y zonas de estancamiento, las cuales pueden concentrar o retener larvas (Roughan *et al.*, 2005; Morgan y Shepherd, 2006). Se ha propuesto que el promontorio de Punta Eugenia es una zona de transición biogeográfica para especies tropicales y templadas limitando la dispersión larval (Briggs, 1974; Hewitt, 1981; Lavaniegos, 2014). Además, esta región ha sido considerada como un quiebre filogeográfico para muchas especies marinas (Bernardi *et al.*, 2003) (e.g. *Fundulus parvipinnis*, Bernardi y Talley, 2000; *Girella nigricans*, Terry *et al.*, 2000; *Embiotoca jacksoni*, Bernardi, 2000; *Gillichthys mirabilis*, Huang y Bernardi, 2001). Sin embargo, esta ruptura genética no se limita a un área estrecha y podría cubrir una región más amplia hacia el norte (Haupt *et al.*, 2013).

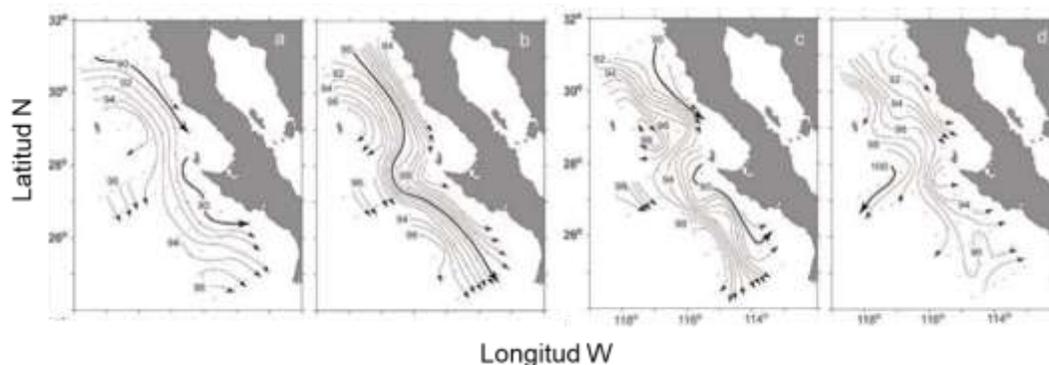
La partición regional detectada para *H. fulgens* en este estudio coincide parcialmente con la región donde esos estudios previos detectaron divergencia poblacional. Por lo tanto, los resultados sugieren que la diferenciación genética entre los grupos GCN y GCS podría estar

relacionada con esta ruptura biogeográfica. No obstante, es necesario realizar más muestreos para determinar con precisión el área donde la división costera de *H. fulgens* ocurre a lo largo de la PBC. El mecanismo subyacente que mantiene la partición regional podría estar relacionada con las características oceanográficas de la región y con la disponibilidad de hábitat que limita la conectividad entre regiones. La parte centro sur de la PBC tiene una zona larga y continua de arrecifes rocosos; sin embargo, la región a lo largo de la costa norte de Punta Eugenia (Guerrero Negro a Isla San Jerónimo) es una extensa región arenosa (Terry *et al.*, 2000) la cual no es un hábitat adecuado para la especie.

Recientemente se reportó un patrón de estructuración genética similar a lo largo de la PBC para otra especie de abulón (*H. corrugata*) utilizando marcadores SNPs neutrales y adaptativos (Mares-Mayagoitia *et al.*, 2021). Los autores encuentran dos grupos genéticos con su división a la altura del promontorio de Punta Eugenia, siendo más evidente con los loci adaptativos. Sin embargo, a diferencia de nuestros resultados, los individuos de Isla Guadalupe no se diferenciaron del grupo norteño.

Los patrones de flujo asimétrico contemporáneo mostraron un fuerte flujo bidireccional dentro del GCS y un flujo predominante del GCN al GCS. Estos resultados son consistentes con las características oceanográficas locales donde la dinámica de las corrientes superficiales del mar (<200 m) a lo largo de la PBC es predominantemente hacia el sur con una gran influencia de la Corriente de California durante el año en la región norte (Lynn y Simpson, 1987; Durazo, 2015). Sin embargo, los cambios estacionales en la circulación oceánica se ven favorecidos en la región sur de la PBC durante el verano y el otoño (Lluch-Belda *et al.*, 2000) en donde el debilitamiento de los vientos y el calentamiento de las masas de agua favorecen la existencia de estructuras o giros de mesoescala (Fig. 17). También se ha reportado una contracorriente superficial en regiones someras cercanas a la costa, donde los procesos de pequeña escala pueden haber servido como mecanismo físico que favoreció el transporte y reclutamiento larvario (Durazo *et al.*, 2010; Durazo, 2015). Este proceso de meso escala en la región sur coinciden con los picos de desove (otoño e invierno) reportados para *H. fulgens* (Ortíz-Quintanilla y León-Carballo, 1996; Vélez-Arellano *et al.*, 2017).

Por lo tanto, el transporte de larvas (y por consecuencia de flujo genético) puede haber sido promovido dentro de las zonas central y sur de la PBC. La circulación oceanográfica estacional también podría aumentar la dispersión de larvas entre regiones durante los picos de desove cuando la Corriente de California es más intensa (invierno) con una dirección hacia el sur, pero la gran discontinuidad del hábitat al norte de Punta Eugenia podría tener un papel crucial en la limitación de la conectividad larval entre regiones. Sin embargo, se necesita un enfoque de paisaje marino para probar el efecto de las corrientes oceánicas en la estructura poblacional de *H. fulgens*.



**Figura 17.** Patrones de circulación oceánica superficial (<200m) en la porción sur del Sistema de Corrientes de California (SCC) a lo largo de las cuatro estaciones del año, las flechas indican la dirección de la corriente. **a)** invierno, **b)** primavera, **c)** verano y **d)** otoño. Tomado y modificado de Gaxiola-Castro y Durazo (2010).

#### 8.4 Diversidad genética

Actualmente no se discute mucho acerca de los valores de diversidad genética en estudios realizados con SNPs, sin embargo, los valores globales de  $H_O$ ,  $H_E$  y  $F_{IS}$  encontrados en el presente estudio para el set de marcadores neutrales ( $H_O= 0.284$ ;  $H_E= 0.300$ ;  $F_{IS}= 0.004$ ,  $P = 0.38$ ) (Tabla 7) son similares a los reportados para otras especies marinas (Tabla 9) e incluso mayores que algunas especies cuyas pesquerías se encuentran colapsadas (e.g. *Scophthalmus maximus*; do-Prado *et al.*, 2018). En trabajos anteriores realizados en *Haliotis fulgens* con marcadores tipo microsatélites (Gutiérrez-González *et al.*, 2007) se han reportado valores relativamente altos de diversidad genética en varias de las localidades muestreadas en este estudio.

**Tabla 9.** Valores de diversidad genética reportados en especies marinas obtenidos con marcadores moleculares tipo SNPs. Se muestra la cita del trabajo, la especie, el tipo de RAD-seq utilizado, tamaño de muestra (n), # de SNPs utilizados para la estimación de los índices heterocigosidad observada ( $H_O$ ) y Heterocigosidad esperada ( $H_E$ ), valores de diferenciación genética ( $F_{ST}$ ) con su valor de probabilidad asociado (P), coeficiente de endogamia ( $F_{IS}$ ) o su análogo ( $G_{IS}$ ) y resultados de pruebas de asilamiento por distancia (IBD) con su respectivo valor de probabilidad asociado (P).

Trabajo	especie	RAD-seq	n	# SNPs	$H_O$	$H_E$	$F_{ST}$	p	$F_{IS}$	IBD
Saenz-Agudelo et al., 2015	<i>Amphirion bicinctus</i>	ddRAD	115	4,559 (todos)	0.218	0.246	0.054	----	0.115	----
Benestan et al., 2015 <sup>a</sup>	<i>Homarus americanus</i>	RAD-seq	586	8,144 (todos)	----	----	0.002	<b>0.0002</b>	----	$r^2=0.56$ , <b>p=0.00009</b>
Wyngaarden et al., 2016	<i>Placopecten magellanicus</i>	RAD-seq	245	7,163 (todos)	----	0.271	0.005	<b>&lt;0.001</b>	----	$r^2=0.157$ , <b>p&lt;0.05</b>
Xurebe et al., 2017	<i>Parastichopus californicus</i>	RAD-seq	717	3,699 (todos)	0.103- 0.115	0.110- 0.117	0.004	<b>0.001</b>	$G_{IS}= 0.014$ - 0.081	----
Sandoval-Castillo et al., 2018	<i>Haliotis laevis</i>	ddRAD	371	8,786 (neutral)	0.291	0.263	0.008	NS	----	<b>p&lt;0.001</b>
Zhao et al., 2018	<i>Lateolabrax maculatus</i>	ddRAD	219	10,297 (neutral)	0.214 - 0.239	0.263- 0.277	0.015- 0.057	<b>&lt;0.001</b>	0.115- 0.204	----
do Prado et al., 2018	<i>Scophthalmus maximus</i>	ddRAD	672	755 (todos)	----	0.093- 0.189	0.090	<b>&lt;0.001</b>	----	----
Mullins et al., 2018	<i>Thunnus albacares</i>	tGBS	89	11,101 (todos)	0.232	0.261	0.002	<b>&lt;0.0001</b>	0.114	----
Presente trabajo	<i>Haliotis fulgens</i>	ddRAD	175	2,170 (neutral)	0.284	0.300	0.005	<b>0.006</b>	$F 0.004$	Global ( $r^2= 0.57$ , P = 0.001) Costero ( $r^2= 0.44$ , P = 0.005) GCS ( $r^2= 0.23$ , P = 0.996)

## **8.5 Implicaciones de los patrones de estructuración poblacional en el manejo del recurso**

Los resultados en el presente estudio sugieren que las unidades biológicas de *H. fulgens* no soportan las áreas administrativas de la pesquería del recurso, por lo que las áreas de evaluación y gestión de las poblaciones deberían de ser revisadas. Por ejemplo, la Isla Guadalupe debería ser considerada como un stock independiente, separado de cualquiera de las localidades costeras (Zona putativa I). Las localidades en la parte norte de la PBC (al norte de Guerrero Negro, abarcando a Isla San Jerónimo y el Faro San José) formarían una segunda subpoblación, a reserva de una evaluación más extensa que abarque una mayor área dentro de la parte norte para descartar sub-estructuración (Zona putativa II). Todas las localidades desde Isla Cedros a La Bocana y Punta Abreojos conformarían el tercer grupo (Zona putativa III). Las localidades al sur de la laguna de San Ignacio no fueron muestreadas en este estudio, mantendrían su clasificación actual como Zona IV hasta que se realicen más muestreos y el análisis pertinente.

Sin embargo, dado que hay diferencias evidentes en algunos aspectos biológicos, características morfológicas y de productividad de los bancos de *H. fulgens* dentro y entre las cuatro zonas administrativas actualmente dispuestas (Ortíz-Quintanilla y León-Carballo, 1996; Vélez-Arellano, 2016; Guzmán-del Prío y Del Monte-Luna, 2017), los resultados genéticos deberían ser complementados con otras técnicas (morfológicas, tasas reproductivas y de reclutamiento, etc.) con la finalidad de mejorar las decisiones de manejo acerca del recurso y ayudar a delinear mejor los stocks pesqueros.

## **8.6 Asignación poblacional**

### **8.6.1 Selección de SNPs altamente informativos, análisis de asignación y pruebas de poder**

La asignación de los individuos a su población de origen es relevante para el manejo y la conservación de las especies (Manel *et al.*, 2005). Los métodos modernos de trazabilidad genética para especies marinas que aplican enfoques del genoma completo para identificar pequeños paneles de SNPs con alto poder de diferenciación, representan una herramienta

prometedora para abordar cuestiones específicas de asignación y apoyar al manejo y conservación de las especies (Manel *et al.*, 2005; Weist *et al.*, 2019). Sin embargo, los estudios genómicos en haliótidos se centran en la caracterización de la estructura genética (e.g. *H. fulgens* Zúñiga *et al.*, 2000; Gutiérrez-González *et al.*, 2007; Gruenthal *et al.*, 2013; *H. corrugata* Díaz-Viloria *et al.*, 2009; Munguía-Vega *et al.*, 2015; Mares-Mayagoitia *et al.*, 2021; *H. midae* Merwe *et al.*, 2009; *H. discus* Nam *et al.*, 2021) así como el desarrollo de herramientas para la identificación de especies (Aranceta-Garza *et al.*, 2011; Vargas-Peralta *et al.*, 2022) y las herramientas genéticas para la identificación de las poblaciones de origen son escasos o nulos (e.g. Miller *et al.*, 2008 usa microsatélites para identificar población de origen en *H. rubra* alrededor de Tasmania).

Este es el primer estudio que aplica marcadores SNPs para investigar los patrones de asignación poblacional en el abulón azul. Como se mencionó anteriormente, el uso del genoma de referencia de alta calidad de la especie incrementó cuatro veces el número de marcadores obtenidos (de 2,216 a 9,954 y 9,869 para las escalas global y costera respectivamente). De estos nuevos datos de SNPs, se seleccionaron los 500 loci con mayor valor de  $F_{ST}$  y se desarrollaron seis paneles de SNPs para cada escala (de 50 a 500 SNPs).

Las pruebas de asignación poblacional suelen estar afectadas por una sobre estimación en la precisión de asignación (conocido normalmente en la literatura como “*high-grading bias*”). Con la finalidad de evitar este sesgo, se recomienda utilizar un enfoque de entrenamiento/prueba o doble validación cruzada en donde se separan los individuos en un grupo de entrenamiento y otro de prueba, el primero es utilizado para la selección de los marcadores más diferenciados y el segundo para validar el poder de los marcadores en los análisis de asignación (Anderson, 2010; Waples, 2010). No obstante, cumplir con este enfoque muchas veces resulta poco viable debido al bajo tamaño de muestra por localidad en los estudios genéticos poblacionales.

En este estudio, dos de los tres grupos genéticos tuvieron bajos tamaños de muestra ( $G_I= 20$  y  $G_{CN}= 24$ ), por lo que, siguiendo las recomendaciones en Benestan *et al.* (2015<sup>b</sup>), se realizaron pruebas de asignación poblacional utilizando todos los marcadores con la finalidad de prevenir

el problema de sobre estimación. También se realizaron corridas separando y sin separar a los individuos en datos de entrenamiento y prueba para probar el impacto que tiene dividir el número total de individuos sobre la precisión de asignación. Estos resultados también sirvieron como control positivo para comparar el máximo porcentaje de asignación esperado (ya que no se espera obtener mayor porcentaje de asignación con un set reducido de marcadores que al utilizar el total de marcadores disponibles; Anderson, 2010).

Los resultados obtenidos con estas pruebas preliminares indican que dividir los individuos para generar las bases de entrenamiento y prueba no tuvo un gran impacto en la precisión en estos análisis, excepto para el GCN que mostró mayores precisiones de asignación cuando se utilizaron todos los individuos para las pruebas de asignación (Anexo H). Dado lo anterior, podemos inferir que los resultados obtenidos utilizando sólo los individuos de prueba (GI=10, GCN=12 y GCS=30) en los paneles de SNPs son confiables. Estos resultados también indicaron que en ambas escalas el nivel máximo de precisión de asignación que se puede obtener para el GCN no puede ser mayor al 89% a escala global y al 77% para la escala costera con ninguno de los paneles, ya que estos valores fueron los máximos alcanzados al utilizar todos los marcadores y el grupo de datos de entrenamiento (Anexo H).

Es un hecho que la precisión de asignación es mayor cuando se utilizan todos los marcadores disponibles que con un subconjunto de los marcadores más discriminantes (Anderson, 2010); sin embargo, para que tengan un valor práctico para las aplicaciones pesqueras, es necesario un panel reducido para lograr que los análisis sean eficientes y rentables (Sylvester *et al.*, 2018).

Las precisiones de asignación de población para GI y GCS con los seis paneles tuvieron un rango de 0.94 (s.d.= 0.08) a 1 (Figs. 13 y 14) fueron similares a las obtenidas con todos los marcadores (Anexo H) en ambas escalas. Estos resultados indican que cualquiera de estos seis paneles puede resultar una herramienta útil para asignar a los individuos de GI y del GCS con alta precisión >95% y >94% respectivamente (Anexos F y G). En general, las precisiones de asignación para los individuos de GI y del GCS fueron más altas que para los del GCN; esto puede deberse a que el número de individuos utilizados en ambos sets de datos (entrenamiento

y prueba) fue mayor para el GCS (n= 30) resultando en una estimación más precisa de los valores de  $F_{ST}$  para ese grupo al hacer la jerarquización de los marcadores y la estimación de las frecuencias alélicas para las pruebas de asignación poblacional. Además, GI fue el grupo que presentó los mayores valores de diferenciación pareados entre las localidades de muestreo (de 0.16 a 0.21  $P < 0.001$ , ver Tabla 5).

En contraste, las precisiones de asignación para el GCN mostraron los valores más bajos en todos los casos, disminuyendo drásticamente del 81% cuando se usan todos los marcadores a escala global (Anexo 7) a un 3% cuando se utiliza el panel de 50 SNPs (Anexo 5). Este grupo también mostró los valores más altos de precisión usando el panel de 500 SNPs (0.91, s.d.= 23) en comparación que cuando se utilizaron todos los marcadores y los individuos de entrenamiento a escala costera (0.77, s.d.= 0.33). Waples (2010) menciona que los problemas de sesgo por sobre estimación en la precisión son más severos cuando los tamaños de muestra son bajos y/o las diferencias genéticas son bajas. El hecho de que el GCN tenga un bajo número de individuos (n= 12) aunado a la posibilidad de una sub-estructuración podría estar explicando el patrón de asignación observado, ya que de existir sub-estructuración dentro de la región norte de la PBC sería necesario aumentar el número de individuos en este grupo para realizar las pruebas con la finalidad de obtener precisiones de asignación más confiables. Además, este grupo es el que presenta una menor diferenciación con respecto a GI y GCS

Las precisiones de asignación a ambas escalas reportadas en el presente estudio son similares a las reportadas para otras especies marinas. Por ejemplo; Benestan *et al.* (2015<sup>a</sup>) obtuvieron precisiones del 94% para la langosta americana (*Homarus americanus*) a escala regional ( $F_{CT} = 0.0011$ ,  $P < 0.001$ ) utilizando los 3,000 SNPs más diferenciados de acuerdo con su valor de  $F_{ST}$ . Sin embargo, las precisiones de asignación a escala fina ( $F_{ST} = 0.0018$ ,  $P < 0.001$ ) resultaron en un valor promedio máximo de 34% utilizando todos los marcadores (10,156 SNPs) (Benestan *et al.*, 2015<sup>b</sup>). Los autores mencionan que probablemente el bajo nivel de diferenciación y el bajo número de muestras para el set de datos de entrenamiento y prueba resultaran en bajos valores de precisión de asignación (Waples, 2010).

En otro estudio realizado con el salmón de Alaska (*Oncorhynchus nerka*) por May *et al.* (2020) reportaron precisiones de asignación del 93% al 96% utilizando un panel reducido de 172 SNPs. Por otra parte, Marie *et al.* (2019) evaluaron la asignación poblacional en 174 individuos de tiburón martillo (*Sphyrna lewini*) con baja diferenciación genética ( $F_{ST} = 0.004$ ,  $P < 0.001$ ) reportando probabilidades de asignación de 0.81 (s.d= 0.03) a 0.98 (s.d.= 0.01) con un panel de 1,317 SNPs presuntamente neutrales.

Jenkins *et al.* (2019) utilizaron los SNPs más diferenciados para asignar a individuos de langosta europea (*Homarus gammarus*) a su localidad de origen en diferentes escalas (cuencas, regional y localidad de muestreo). Estos autores reportan una correcta asignación a su cuenca de origen del 99.7% utilizando un panel de 79 SNPs. Sin embargo, los valores de asignación disminuyen cuando intentan asignar individuos a su localidad de muestreo, yendo de 27% a 58% máximo, dependiendo de la proporción de individuos utilizados para entrenar los modelos.

Con los resultados obtenidos en el presente trabajo podemos concluir que el uso de loci basados en sus valores de máxima divergencia es un enfoque viable para obtener un alto poder de asignación para los individuos de *H. fulgens* de las regiones de GI y del GCS, en consecuencia, el poder de resolución del conjunto de datos de SNPs aumentó, mientras que al mismo tiempo se reduce el número de marcadores necesarios para obtener la misma estructura genética que con el panel completo o el obtenido *de novo*. A pesar de que en este trabajo los paneles de SNPs propuestos no se validaron experimentalmente, validársela validación podría realizarse utilizando bibliotecas de representación reducida (e.g. GT-Seq, Campbell *et al.*, 2015).

Cualquiera de los paneles de 50 SNPs obtenidos para las dos escalas parece ser suficiente para lograr altas precisiones de asignación para los individuos de GI y del GCS. May *et al.* (2020) recomiendan incluir al menos el doble de loci objetivo para el diseño y optimización de *primers* o cebadores con la finalidad de amortiguar la pérdida de loci en los pasos de filtrado. Dado lo anterior, para futuras investigaciones se recomienda realizar el diseño de cebadores con el panel de 500 SNPs en ambas escalas.

### 8.6.2 Análisis de estructura genética poblacional

Los resultados de DAPC utilizando todos los marcadores mostraron que la resolución de la estructura poblacional mejora cuando se utiliza un genoma de referencia de la especie (Figs. 15 y 16), incluso cuando se usa un número reducido de marcadores a escala costera (Anexo J). Esto, a pesar de que las diferencias observadas en el número de SNPs entre un enfoque y otro, parece no afectar de manera substancial las conclusiones de patrones de estructuración genéticos, sin embargo en el presente trabajo no se realizaron análisis de estructura poblacional ( $F_{ST}$ , AMOVA, STRUCTURE) con el panel obtenido con genoma de referencia. Resultados similares se han observado en otras especies, y a pesar de que la secuenciación y ensamblaje de genomas completos se está volviendo cada vez más accesible gracias a los métodos de secuenciación de lecturas largas y las estrategias de ensamblaje, la mayoría de las especies aún no cuentan con un genoma de referencia; no obstante, de manera general para algunos parámetros poblacionales, tener un menor número de marcadores no implica una pérdida relevante de información para las inferencias biológicas (Casanova *et al.*, 2021).

### 8.6.3 Potenciales usos y alcances de los paneles de SNPs para el manejo del recurso

Debido a que los paneles diseñados en este trabajo contienen tanto loci putativos neutrales y adaptativos, es posible que los loci adaptativos se pudieran utilizar para cumplir objetivos específicos en un futuro, tales como el monitoreo de los cambios en sus frecuencias alélicas, lo cual implicaría cambios en el potencial adaptativo debido a cambios ambientales o acciones de manejo (Flanang *et al.*, 2018), (e.g. las prácticas de repoblamiento). Dichos paneles pueden ser monitoreados periódicamente para asegurar que exista suficiente variación genética (tanto en las poblaciones de cultivo como en las silvestres) y utilizar los resultados para actualizar las prácticas de manejo de manera continua (Hohenlohe *et al.*, 2019).

Otro problema que enfrenta esta pesquería en México es la pesca ilegal y el fraude comercial o de alimentos. Se estima que el primero representa el 27% de la captura comercial (Ponce-Díaz *et al.*, 2013). Para el segundo, Aranceta-Garza *et al.* (2011) reportaron el etiquetado erróneo de

abulón en algunos productos enlatados utilizando un análisis con marcadores de ADNmt (18S). Recientemente, Vargas- *et al.* (2022) desarrollaron un panel de SNPs que permite la correcta identificación de cinco especies importantes de abulón que habitan las costas occidentales de la PBC en México. Si bien estos métodos brindan a los productores y autoridades una herramienta para certificar su autenticidad y mejorar el control en su pesca ilegal, estas herramientas no pueden distinguir la población de origen dentro de las poblaciones de la especie.

El uso de herramientas genéticas para uso forense de recursos pesqueros con la finalidad de identificar especies es cada vez más común, sin embargo, su aplicación para determinación de la población de origen tanto de individuos como productos es menos avanzada (Cusa *et al.*, 2021), por lo que se dificulta el control y cumplimiento de las normas pesqueras y en la cadena de alimentos en el mercado (Martinson y Ogden, 2009). Así mismo, Von der Hyden *et al.* (2014) mencionan que la identificación de la población de origen más que simplemente la identificación de especie asegurará una mejor vigilancia de los recursos marinos, aunque su aplicación sigue siendo un reto a nivel mundial. Un ejemplo del uso de paneles de SNPs para identificar poblaciones de origen para trazabilidad y aplicaciones de regulación es el realizado con éxito por el consorcio FishPopTrace (Martinson y Odgen, 2009) en cuatro especies de peces con importancia económica (*Gadus morhua*, *Merluccius merluccius*, *Clupea harengus* y *Solea solea*) (Martinson y Odgen, 2009).

Algunas pesquerías de abulón en México están empezando a buscar certificaciones sostenibles o eco-certificación por parte de organizaciones como el Consejo de Administración Marina (MSC por sus siglas en inglés). Este tipo de certificaciones requiere que todos los distribuidores, procesadores y comerciantes minoristas del producto tengan un sistema de rastreo documentado; también requieren que los productos hayan sido recolectados en el lugar y por los pescadores que poseen el certificado que lo acredite (Barendse *et al.*, 2019; Cusa *et al.*, 2021). Por lo tanto, el uso de los paneles desarrollados aquí para asignación poblacionales podría aplicarse en este contexto de eco-certificación permitiendo la trazabilidad desde los pescadores hasta los consumidores, lo que podría permitir a las cooperativas de las zonas de GI y GCS definir marcas territoriales y aumentar su valor comercial (Aprile, *et al.*, 2012; Cusa *et al.*,

2021; Lamas *et al.*, 2022). Sin embargo, antes de que esto pueda lograrse, es necesario el diseño de los cebadores y la validación de los paneles para corroborar su precisión y confiabilidad.

Estos paneles representan una herramienta interesante y prometedora para mejorar la utilidad del monitoreo genético en las poblaciones de abulón ya que se pueden genotipificar de manera eficiente en muchos individuos, utilizando un muestreo poco invasivo. Es necesario un análisis más detallado con un mayor número de muestras por localidad y que incluya individuos de los límites norte y sur en la PBC beneficiarían la información aquí recaudada.

## 9. CONCLUSIONES

- Los resultados de la estructura poblacional del abulón azul a lo largo de la PBC respaldan tres grupos genéticos regionales que muestran una clara diferenciación entre la isla oceánica (Isla Guadalupe, alejada ~250 km de la costa) y el resto de las localidades costeras; el segundo grupo está formado por el GCN que incluyen la isla San Jerónimo y Faro San José; y el tercer grupo incluye las localidades del GCS (Isla Cedros, Punta Eugenia, Tortugas Norte, Tortugas Sur, Asunción Norte, Asunción Sur y La Bocana).
- Las diferencias genéticas muestran una correlación positiva con la distancia geográfica (aislamiento por distancia), sin embargo, más que la distancia geográfica, las corrientes oceanográficas están jugando un papel importante en la estructuración genética actuando como limitantes al flujo genético entre individuos de la zona norte y sur de la PBC y como mecanismo de dispersión larval en la región sur ya que se observó de baja a nula diferenciación genética en las localidades de dicha región, lo cual puede estar explicado por los procesos de mesoescala que suceden durante los meses de máxima madurez del abulón azul. No obstante, la región del norte (arriba de Punta Eugenia) no se ve tan influenciada por los cambios estacionales de las corrientes oceánicas superficiales, manteniendo un flujo predominante de norte a sur evitando así un gran flujo de individuos (flujo genético) del sur al norte.
- Los valores de diversidad genética encontrados en este estudio tanto con el set de marcadores neutrales son similares a los reportados en otras especies marinas que utilizan SNPs derivados de métodos RAD-seq.
- Los resultados de la estructura genómica neutral no respaldan las áreas de manejo administrativo actuales de la pesquería de abulón verde en el BCP. Este estudio mostró patrones sólidos de estructura de la población que deben considerarse para delinear los límites de las poblaciones, la evaluación y las acciones de gestión para garantizar la resiliencia de la pesca. Sin embargo, se requiere un muestreo adicional que incluya sitios a lo largo de todo el rango de distribución de la especie para tener una mejor comprensión de la diversidad genómica espacial. Además, la integración futura de la variación adaptativa y

el modelado ambiental y oceanográfico mejorará la comprensión de la biocomplejidad de este recurso pesquero.

- Los resultados de estructura poblacional mejoraron utilizando un genoma de referencia de la especie; sin embargo, las conclusiones sobre la estructura genética poblacional obtenidas con el enfoque *de novo* también son precisas.
- La asignación individual a la población de origen mediante un perfil genético proporciona nuevas herramientas informativas para el manejo, vigilancia y comercialización del abulón azul.
- El presente estudio sirve como una base de datos de referencia de alta calidad con muestras de origen conocido en un amplio rango de la distribución de la especie en México la cual podría utilizarse en futuros estudios con la finalidad de evaluar la trazabilidad o asignación de origen en individuos con origen desconocido.

## 10. LITERATURA CITADA

- Adamack, E. T. y Gruber, B. (2014). PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, 5: 384-387. <https://doi.org/10.1111/2041-210X.12158>
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Allendorf, F. W., Hohenlohe, P. A. y Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11: 697-709. <https://doi.org/10.1038/nrg2844>
- Anderson, C. E. (2010). Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, 20: 701-710. doi: 10.1111/j.1755-0998.2010.02846.x
- Aprile, C. M., Caputo V. y Nayga. M. R. (2012). Consumers' valuation of food quality labels: the case of the European geographic indication and origin farming labels. *International Journal of Consumer Studies*, 36: 158-165. doi: 10.1111/j.1470-6431.2011.01092.x
- Aranceta-Garza, F., Pérez-Enriquez, R. y Cruz, P. (2011). PCR-SSCP method for genetic differentiation of canned abalone and commercial gastropods in the Mexican retail market. *Food Control*, 22: 1015-1020. doi:10.1016/j.foodcont.2010.11.025
- Ba, H., Jia, B., Wang, G., Yang, Y., Kedem, G. y Li, C. (2017). Genome-Wide SNP Discovery and Analysis of Genetic Diversity in Farmed Sika Deer (*Cervus nippon*) in Northeast China Using Double-Digest Restriction Site-Associated DNA Sequencing. *G3 (Bethesda)*, 7(9): 3169-3176. <https://doi.org/10.1534/g3.117.300082>
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, A. Z., Selker, U. E., Cresko, A. W. y Johnson, A. E. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One*, 3(10): e3376. <https://doi.org/10.1371/journal.pone.0003376>
- Barendse, J., Roel, A., Longo, C., Andriessen, L., Webster, M. I. L., Odgen, R. y Neat, F. (2019). DNA barcoding validates species labelling of certified seafood. *Current Biology*, 29(6): R183-R199. <https://doi.org/10.1016/j.cub.2019.02.014>.
- Belmar-Pérez, J. y Guzmán-del Proo, S. (1992). Madurez sexual y ciclo gonádico en *Haliotis fulgens* y *Astraea undosa* en Bahía Tortugas. 121-130pp. En: Guzmán del Proo (ed.), Memorias del Taller México-Australia sobre reclutamiento de recursos bentónicos de Baja California. IPN-Secretaría de Pesca. 212 pp.
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R. y Bernatchez, L. (2015<sup>a</sup>). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, 24(13): 3299-3315. doi: 10.1111/mec.13245. Epub 2015 Jun 15.
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R. y Bernatchez, L. (2015<sup>b</sup>). Erratum (2016). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, 25: 1626-1629. doi: 10.1111/mec.13600

- Bernardi, G., Lloyd F. y Axayacatl R. O. (2003). Vicariance and dispersal across Baja California in disjunct marine fish populations. *Evolution International Journal of Organic Evolution*, 57(7): 1599-1609. <https://doi.org/10.1111/j.0014-3820.2003.tb00367.x>
- Bernardi, G. y Talley, D. (2000). Genetic evidence for limited dispersal in the coastal California killifish, *Fundulus parvipinnis*. *Journal of Experimental Marine Biology and Ecology*, 255(2): 187-199. doi: 10.1016/s0022-0981(00)00298-7. PMID: 11108851.
- Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., Maes, G. E., Martinsohn, J. T., Miller, K. M., Naish, K. A., Ovenden, J. R., Primmer, C. R., Young Suk, H., Therkildsen, O. N. y Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, 32(9): 665-680. doi: 10.1016/j.tree.2017.06.010. Epub 2017 Aug 14. PMID: 28818341.
- Berry, O., Tocher, D. y Sarre, S. D. (2004). Can assignment test measure dispersal?. *Molecular ecology*, 13(3): 551-561. doi: 10.1046/j.1365-294x.2004.2081.x. PMID: 14871360.
- Bhatia, D., Wing, R. A. y Singh, K. (2013). Genotyping by sequencing, its implications and benefits. *Journal of Crop Improvement*, 40(2): 101-111
- Briggs, J. C. (1974). Marine zoogeography. 1<sup>a</sup> ed. McGraw-Hill Inc., Nueva York.
- Burton R. S. y Tegner, M. J. (2000). Enhancement of red abalone *Haliotis rufescens* stocks at San Miguel Island: reassessing a success story. *Marine Ecology Progress Series*, 202: 303-308. doi:10.3354/meps202303
- Campbell, N. R., Harmon, S. A. y Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15: 855-867. doi:10.1111/1755-0998.12357
- Casanova, A., Maroso, F., Blanco, A., Hermida, M., Ríos, N., García, G., Manuzzi, A., Zane, L., Verissimo, A., García-Marín, J. L., Bouza, C., Vera, M., y Martínez, P. (2021). Low impact of different SNP panels from two building-loci pipelines on RAD-Seq population genomic metrics: case study on five diverse aquatic species. *BMC Genomics*, 22(150). <https://doi.org/10.1186/s12864-021-07465-w>
- Catchen J, P., Hohenlohe, A., Bassham, S., Amores, A. y Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11): 3124-3140. doi: 10.1111/mec.12354.
- Chelton, B. D. (1982). Large-scale response of the California current to forcing by the wind stress curl. *CalCofi Rep*, XXIII: 130-148.
- Chen, K., Marschall, A. E., Sovic, G. M., Fries, C. A., Gibbs, L. H., y Ludsin, A. S. (2018). assignPOP: An R package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution*, 9: 439-446. doi: 10.1111/2041-210X.12897
- Cook, A. P. (2016). Recent trends in worldwide abalone production. *BioOne Research Evolved*, 35(3): 581-583. doi: <http://dx.doi.org/10.2983/035.035.0302>
- Cook, P. A. (2019). Worldwide abalone production statistics. *Journal of Shellfish Research*, 38(2): 401-404. <https://doi.org/10.2983/035.038.0222>

- Crawford, J. E. y Lazzaro, P. B. (2012). Assessing the accuracy and power of population genetic inference from low-pass next-generation sequencing data. *Frontier in Genetics*, 24: 3-66. doi: 10.3389/fgene.2012.00066
- Cusa, M., Glew, St J. K., Trueman, C., Mariani, S., Buckley, L., Neat, F. y Longo, C. (2021). A future for seafood point-of-origin testing using DNA and stable isotope signatures. *Reviews in Fish Biology and Fisheries*, 32: 597-621. <https://doi.org/10.1007/s11160-021-09680-w>.
- Danecek, P., Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*. 27(15): 2156-8. doi: 10.1093/bioinformatics/btr330. Epub 2011 Jun 7. PMID: 21653522; PMCID: PMC3137218.
- del Río-Portilla, M. A., y González-Avilés, J. G. (2001). Population genetics of the yellow abalone, *Haliotis corrugata*, in Cedros and San Benito Islands: A preliminary survey. *J. Shellfish Res.* 20(2): 765-770
- De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, J. D., Seneca, F., Jaris, H., Therkildsen, O. N., Morikawa, M. y Palumbi, R. S. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, 12: 1058-1067. <https://doi.org/10.1111/1755-0998.12003>
- De Wit, P. y Palumbi, S. R. (2013). Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, 22(11): 2884-2897. doi: 10.1111/mec.12081. Epub 2012 Oct 29. PMID: 23106543.
- DeWoody, J. A. y Avise, J. C. (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology*, 56: 461-473. <https://doi.org/10.1111/j.1095-8649.2000.tb00748.x>
- Díaz-Viloria, N., Cruz-Hernández, P., Guzmán-del Proo, S. A. y Perez-Enriquez, R. (2009). Genetic connectivity among pink abalone *Haliotis corrugata* populations. *Journal of Shellfish Research*, 28(3): 599-608. doi: 10.2983/035.028.0324
- Díaz-Viloria, N., Perez-Enriquez, R. Cruz-Hernández, P. y Aguilar-Osuna, D. (2013). Herencia Mendeliana en microsatélites de abulón Amarillo *Haliotis corrugata*. *Interciencia*, 38(6): 455-460. <https://www.redalyc.org/articulo.oa?id=33928571005>
- Diario Oficial de la Federación (DOF). 1993. Proyecto de Norma Oficial Mexicana NOM-005-PESC-1993, para regular el aprovechamiento de las poblaciones de las distintas especies de abulón en aguas de jurisdicción federal de la península de Baja California. Diario Oficial de la Federación, Miércoles 1 de septiembre de 1993. Disponible en: [http://dof.gob.mx/nota\\_detalle.php?codigo=4779603&fecha=01/09/1993&print=true](http://dof.gob.mx/nota_detalle.php?codigo=4779603&fecha=01/09/1993&print=true).
- Diario Oficial de la Federación (DOF). 1994. Aviso por el que se da a conocer el establecimiento de épocas y zonas de veda para la pesca de diferentes especies de la fauna acuática en aguas de jurisdicción federal de los Estados Unidos Mexicanos. Diario Oficial de la Federación, Miércoles 16 de marzo de 1994. Disponible en: [http://www.dof.gob.mx/nota\\_detalle.php?codigo=4678590&fecha=16/03/1994](http://www.dof.gob.mx/nota_detalle.php?codigo=4678590&fecha=16/03/1994).
- Diario Oficial de la Federación (DOF). 2012. Acuerdo por el que se da a conocer la actualización de la Carta Nacional Pesquera. Diario Oficial de la Federación. 24 de Agosto del 2012.

- Segunda sección. Secretaría de Agricultura Ganadería, Desarrollo Rural, Pesca y Alimentación. México. 21-128. Disponible en: [https://www.dof.gob.mx/nota\\_detalle.php?codigo=5265388&fecha=24/08/2012](https://www.dof.gob.mx/nota_detalle.php?codigo=5265388&fecha=24/08/2012)
- Diario Oficial de la Federación (DOF). 2018. Acuerdo por el que se da a conocer la actualización de la Carta Nacional Pesquera. Diario Oficial de la Federación, Lunes 11 de junio de. Disponible en: [http://dof.gob.mx/nota\\_detalle.php?codigo=5525712&fecha=11/06/2018](http://dof.gob.mx/nota_detalle.php?codigo=5525712&fecha=11/06/2018)
- do-Prado, F. D., Vera, M., Hermida, M., Bouza, C., Pardo, B. G., Vilas, R., Blanco, A., Fernández, C., Maroso, F., Maes, G. E., Turan, C., Volckaert, F. A. M., Taggart, J. B., Carr, A., Ogden, R., Nielsen, E. E., Consortium, T. A. y Martínez, P. (2018). Parallel evolution and adaptation to environmental factors in a marine flatfish: Implications for fisheries and aquaculture management of the turbot (*Scophthalmus maximus*). *Evolutionary Applications*, 11: 1322-1341. <https://doi.org/10.1111/eva.12628>
- Dowling, A. N., Stephen, J. H. y McGarvey, R. (2004). Assessing population sustainability and response to fishing in terms of aggregation structure for greenlip abalone (*Haliotis laevigata*) fishery management. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(2): 247-259 doi: 10.1139/f03-165.
- Durazo, R. (2015). Seasonality of the transitional region of the California Current System off Baja California. *Journal of Geophysical Research: Oceans*, 120: 1173-1196.
- Durazo-Arvizu, R. y Baumgartner-Mcbride, T. R. (2002). Evolution of oceanographic conditions of Baja California: 1997-1999. *Progress in Oceanography*, 54:7-31.
- Earl, D. A. y VonHoldt, B. M. (2012). Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources*, 4: 359-361
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D. y Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(4): 1-18. <https://doi.org/10.18637/jss.v048.i04>
- Evanno, G., Regnaut, S. y Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8): 261-2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., Hofer, T. y M. Foll (2009). Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285-298. <https://doi.org/10.1038/hdy.2009.74>
- Excoffier, L. y Lischer, H. E. L. (2015). Arlequin suite ver 3.5.2.2: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3): 564-567. doi: 10.1111/j.1755-0998.2010.02847.x
- Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N. y Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, 11: 1035-1052. 10.1111/eva.12569
- Foll, M. (2012). BayeScan v2.1 User manual. 10 pp. Disponible en: [http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.1\\_manual.pdf](http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.1_manual.pdf)

- Foll, M. y Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2): 977-993. doi: 10.1534/genetics.108.092221.
- Fountain, E. D., Pauli, J. N., Reid, B. N., Palsbøll, P. J. y Perry, M. Z. (2016). Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Molecular Ecology Resources*, 16(4): 966-978. doi: 10.1111/1755-0998.12519
- Frank, T. K. y Brickman, D. (2000). Allee effects and compensatory population dynamics within a stock complex. *Canadian Journal of Fisheries and Aquatic Sciences*, 57(3): 513-517. <http://dx.doi.org/10.1139/cjfas-57-3-513>
- Fumagalli, M. (2013). Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. *Plos One*, 8(11): e79667. <https://doi.org/10.1371/journal.pone.0079667>
- Funk, C. W., McKay, K. J., Hohenlohe, A. P. y Allendorf, W. F. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, 27(9): 489-496. DOI: 10.1016/j.tree.2012.05.012
- Gaffney, P. M., Rubin, V. P., Hedgecock, D., Powers, D. A., Morris, G. y Hereford, L. (1996). Genetic effects of artificial propagation: signals from wild and hatchery populations of red abalone in California. *Aquaculture*, 143(3-4): 257-266. [https://doi.org/10.1016/0044-8486\(96\)01278-1](https://doi.org/10.1016/0044-8486(96)01278-1)
- Gamboa M, y Watanabe K (2019). Genome-wide signatures of local adaptation among seven stoneflies species along a nationwide latitudinal gradient in Japan. *BMC Genomics*, 20(1):84. doi: 10.1186/s12864-019-5453-3. PMID: 30678640; PMCID: PMC6346529.
- García-Rodríguez, F. J. y Perez-Enriquez, R. (2006). Genetic differentiation of the California spiny lobster *Panulirus interruptus* (Randall, 1840) along the west coast of the Baja California Peninsula, Mexico. *Marine Biology*, 148: 621-629. <https://doi.org/10.1007/s00227-005-0101-7>
- Gilbey, J., Cauwelier, E., Coulson, W. M., Stradmeyer, L., Sampayo, N. J., Armstrong, A., Verspoor, E., Corrigan, L., Shelley, J. y Middlemas, S. (2016). Accuracy of assignment of Atlantic salmon (*Salmo salar* L.) to rivers and regions in Scotland and Northeast England based on Single Nucleotide Polymorphism (SNP) markers. *PLoS One*, 11(10): e0164327. <https://doi.org/10.1371/journal.pone.0164327>
- Goudet, J. y Jombart, T. (2015). Hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04-22. Available at <https://CRAN.R-project.org/package=hierfstat>
- Graham, C., F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R. G., Martino, J. A., Pierson, T., Rogers, S. M., Wilson, J. Y. y Somers, C. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, 15: 1304-1315. <https://doi.org/10.1111/1755-0998.12404>
- Grewe, M. P., Feutry, P., Hill, L. P., Gunasekera, M. R., Shaefer, M. K., Itano, G. D., Fuller, W. D., Foster, D. S. y Davies, R. C. (2015). Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Scientific Reports*, 5: 16916. doi: 10.1038/srep16916

- Gruenthal, K. M. y Burton, R. S. (2008). Genetic structure of natural populations of the California black abalone (*Haliotis cracherodii* Leach, 1814), a candidate for endangered species status. *Journal of Experimental Marine Biology and Ecology*, 355(1): 47-58. <https://doi.org/10.1016/j.jembe.2007.11.013>
- Gruenthal, K. M., Witting, D. A., Ford, T., Neuman, M. J., Williams, J. P., Pondella II, D. J., Bird, A., Caruso, N., Hyde, R. J., Seeb, W. L. y Larson, A. W. (2013). Development and application of genomic tools to the restoration of green abalone in southern California. *Conservation Genetics*, 15: 109-121. <https://doi.org/10.1007/s10592-013-0524-5>
- Gutiérrez-González, J. L., Cruz-Hernández, P. del Río-Portilla, M. A. y Perez-Enriquez, R. (2007). Genetic structure of green abalone *Haliotis fulgens* populations of Baja California, México. *Journal of Shellfish Research*, 26(3): 839-846. [https://doi.org/10.2983/0730-8000\(2007\)26\[839:GSOGAH\]2.0.CO;2](https://doi.org/10.2983/0730-8000(2007)26[839:GSOGAH]2.0.CO;2)
- Gutiérrez-González, J. L. y Perez-Enriquez, R. (2005). A genetic evaluation of stock enhancement of blue abalone *Haliotis fulgens* in Baja California, México. *Aquaculture*, 247: 233-242. doi:10.1016/j.aquaculture.2005.02.021
- Guzmán del-Próo, S. A. (1992). A review of the biology of abalone and its fishery in Mexico. En: Shepard, S. A., Tegner, M. J. y Guzmán del Próo, S. A. Abalone of the world. Biology, Fisheries and Culture. Fishing News Books. EUA.
- Guzmán del-Próo, S. A., Salinas, F., Zaytsev, O., Belmar-Perez, J. y Carrillo-Laguna, J. (2000). Potential dispersion of reproductive products and larval stages of abalone (*Haliotis* spp.) as a function of the hydrodynamics of Bahía Tortugas, Mexico. *Journal of Shellfish Research*, 19(2): 869-881.
- Guzmán del-Próo, S. A., Palau, L. C., Laguna, J. B. y Fregoso, R. H. (2003). Effects of the 'El Niño' event on the recruitment of benthic invertebrates in Bahía Tortugas, Baja California Sur. *Geofísica Internacional*, 42(3): 429-438. <http://www.redalyc.org/articulo.oa?id=56842314>
- Guzmán-del Próo, S. A. y del Monte-Luna, P. (2017). Abalone reef productivity and the problem of scale in the management of the Mexican abalone fishery. *Ocean & Coastal Management*, 144: 1-6. <https://doi.org/10.1016/j.ocecoaman.2017.04.005>
- Hamm, D. E y Burton, R. S. (2000). Population genetics of black abalone, *Haliotis cracherodii*, along the central California coast. *Journal of Experimental Marine Biology and Ecology*, 254(2): 235-247. [https://doi.org/10.1016/S0022-0981\(00\)00283-5](https://doi.org/10.1016/S0022-0981(00)00283-5)
- Hammer-Hansen, J., Therkildsen, N. O. y Pujolar, J. M. (2014). Population genomics of marine fishes: Next-generation prospects and challenges. *Biological Bulletin*, 227(2): 117-132. doi: 10.1086/BBLv227n2p117
- Haupt, A. J., Micheli, F. y Palumbi, S. R. (2013). Dispersal at a Snail's Pace: Historical Processes Affect Contemporary Genetic Structure in the Exploited Wavy Top Snail (*Megastraea undosa*). *Journal of Heredity*, 104(3): 327-340. doi: 10.1093/jhered/est002
- Hauser, L. y Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, 9(4): 333-362. <https://doi.org/10.1111/j.1467-2979.2008.00299.x>

- Hedgecock, D., Launey, S., Pudovkin, A. I., Naciri, Y., Lapègue, S. y Bonhomme, F. (2007). Small effective number of parents ( $N_b$ ) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. *Marine Biology*, 150(1): 173-182. <https://doi.org/10.1007/s00227-006-0441-y>
- Hewitt, R. P. (1981). Eddies and speciation in the California Current. *CalCOFI Rep.*, XXII: 96-98.
- Hilborn, R., Quinn, T. P., Schindler D. E. y Rogers, D. E. (2003). Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences of the United States of America*, 100(11): 6564-6568. <https://doi.org/10.1073/pnas.1037274100>
- Hohenlohe, A. P., Funk, C. W. y Rajora, P. O. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1): 62-82. doi: 10.1111/mec.15720
- Huang, D. y Bernardi, G. (2001). Disjunct Sea of Cortez-Pacific Ocean *Gillichthys mirabilis* populations and evolutionary origin of their Sea of Cortez endemic relative, *Gillichthys seta*. *Marine Biology*, 138: 421-428. <https://doi.org/10.1007/s002270000454>
- INAPESCA. (2006). Sustentabilidad y Pesca Responsable en México Evaluación y Manejo. México. 256. Disponible en: <https://comisiones.senado.gob.mx/agricultura/docs/bibliografia/sustentabilidad.pdf>
- INAPESCA. (2021). Informe de autoevaluación primer trimestre 2021. 136 pp. Disponible en: <https://www.inapesca.gob.mx/portal/documentos/transparencia/planes-programas-informes/Informe-de-Autoevaluacion-1T2021.pdf>
- Jakobsson, M. y Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14): 1801-1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Jeffery, W. N., Wringe, F. B., McBride, C. M., Hamilton, C. L., Stanley, R. E. R., Bernatchez, L., Kent, M., Clément, M., Gilbey, J., Sheehan, F. T., Bentzen, P. y Bradbury, R. I. (2018). Range-wide regional assignment of Atlantic salmon (*Salmo salar*) using genome wide single-nucleotide polymorphisms. *Fisheries Research*, 206: 163-175. <https://doi.org/10.1016/j.fishres.2018.05.017>
- Jenkins, T. L., Ellis, C. D., Triantafyllidis, A. y Stevens J. R. (2019). Single nucleotide polymorphisms reveal a genetic cline across the north-east Atlantic and enable powerful population assignment in the European lobster. *Evolutionary applications*, 12: 1881-1899. doi: 10.1111/eva.12849
- Jombart, T. (2008). ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403-1405. doi:10.1093/bioinformatics/btn129/
- Jombart, T., Devillard, S. y Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1): 94. doi 10.1186/1471-2156-11-94
- Karpov, A. K., Haaker, L. P., Taniguchi, K. I. y Rogers-Bennett, L. (2000). Serial depletion and the collapse of the California abalone (*Haliotis* spp.) fishery. Workshop on Rebuilding Abalone Stocks in British Columbia. 11p. <http://site.ebrary.com/lib/ucsd/Doc?id=10006181&ppg=25>

- Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W. y Prodöhl, P. A. (2013). DiveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4(8): 782-788. <https://doi.org/10.1111/2041-210X.12067>
- Kenchington, E. L. (2003). The effects of fishing on species and genetic diversity. En: Responsible Fisheries in the Marine Ecosystem, Capítulo: 14. M. Sinclair, G. Valdimarson, Reino Unido.
- Kirby, V. L., Villa, R. y Powers, D. A. (1998). Identification of microsatellites in the California red abalone, *Haliotis rufescens*. *Journal of Shellfish Research*, 17(3): 801-804. <https://eurekamag.com/research/003/468/003468019.php>
- Konietschke, F., Placzek, M., Schaarschmidt, F., Ludwig, A. y Hothorn. (2015). nparcomp: An R Software Package for Nonparametric Multiple Comparisons and Simultaneous Confidence Intervals. *Journal of Statistical Software*, 64(9): 1-17. <http://www.jstatsoft.org/v64/i09/>.
- Laconcha, U., Iriondo, M., Arrizabalaga, H., Manzano, C., Markaide, P., Montes, I., Zarraonaindia, I., Velado, I., Bilbao, E., Goñi, N., Santiago, J., Domingo, A., Karakulak, S., Oray, I. y Estonba, A. (2015). New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*). *PLoS One*, 10(6): e0128247. <https://doi.org/10.1371/journal.pone.0128247>
- Laikre, L., Palm, S. y Ryman, N. (2005). Genetic population structure of fishes: Implications for coastal zone management. *Ambio*, 34(2): 111-119. PMID: 15865307.
- Lamas, A., Barreiro, R., Regal, P., Vázquez, B., Miranda, J. M., Cepeda, A. y Franco, C. M. (2022). Development and validation of a SNPs panel used for beef traceability throughout the food chain. *Food Science and Technology*, v42, e07221. doi: <https://doi.org/10.1590/fst.07221>
- Lavaniegos, B. E. 2014. Pelagic amphipod assemblage associated with subarctic water off the West Coast of the Baja California peninsula. *Journal of Marine Systems*, 132: 1-12.
- Leighton L. D. (2000). The biology and culture of the California abalones. Dorrance Publishing Co, Pittsburgh, Pennsylvania.
- León-Carballo, G. y Muciño-Díaz, M. (1996). Pesquería de abulón. En: Estudio del Potencial Pesquero y Acuícola de Baja California Sur. 1ª ed. Casas-Valdéz, M. y Ponce-Díaz, G, México.
- Li, H., Handsaker, B., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. y 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 15;25(16): 2078-9. doi: 10.1093/bioinformatics/btp352.
- Li, H. y Durbin, R. (2009). Fast accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14): 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lischer, L. E. H. y Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 15;28(2): 298-9. doi: 10.1093/bioinformatics/btr642.

- Lluch-Belda, D., Elorduy-Garay, J., Lluch-Cota, S. E. y Ponce-Díaz, G. (2000). BAC Centros de Actividad Biológica del Pacífico Mexicano. Centro de Investigaciones Biológicas del Noroeste, S.C., 1ª ed. México.
- Lluch-Belda, (1999). Coastal upwelling in the eastern Gulf of California. *Oceanologica Acta*, 23(6): 731-740. [https://doi.org/10.1016/S0399-1784\(00\)00121-3](https://doi.org/10.1016/S0399-1784(00)00121-3)
- Luikart, G., England, P. R., Tallmon, D., Jordan, S. y Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews: Genetics*, 4: 981-994. <https://doi.org/10.1038/nrg1226>
- Lynn, R. J. y Simpson, J. J. (1987). The California current system: The seasonal variability of its physical characteristics. *Journal of Geophysical Research*, 92(C12): 12947-12966. <https://doi.org/10.1029/JC092iC12p12947>
- Manel, S., Gaggiotti, O. E. y Waples R. S. (2005). Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, 20(3): 136-142. <https://digitalcommons.unl.edu/usdeptcommercepub/481>
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2): 209-220. PMID: 6018555.
- Mares-Mayagoitia, J. A., De-Anda-Montañez, J. A., Perez-Enriquez, R., Mejía-Ruíz, P., Gutiérrez-González, J. L. y Valenzuela-Quiñonez, F. (2021) Neutral and adaptive population structure of pink abalone (*Haliotis corrugata*): fishery management implications. *ICES Journal of Marine Science*, 1-13. doi:<https://doi.org/10.1093/icesjms/fsab098>
- Magris, G., Marroni, F., D'Agaro, E., Vischi, M., Chiabá, C., Scaglione, D., Kijas, J., Messina, M., Tibaldi, E. y Morgante, M. (2022). ddRAD-seq reveals the genetic structure and detects signals of selection in Italian brown trout. *Genetics Selection Evolution*, 54(8). <https://doi.org/10.1186/s12711-022-00698-7>
- Marie, D. A., Stockwell, L. B. y Rico, C. (2019). DNA analysis of juvenile scalloped hammerhead sharks *Sphyrna lewini* (Griffith, 1834) reveals multiple breeding population and signs of adaptive divergence in the South Pacific. *Frontiers in Marine Science*, 6(718). doi: 10.3389/fmars.2019.00718
- Martinsohn, Th. J. y Odgen, R. (2009). FishPopTrace-Developing SNP-based population genetic assignment methods to investigate illegal fishing. *Forensic Science International: Genetics supplement Series*, 2: 294-296. doi:10.1016/j.fsigs.2009.08.108
- May, A. S., McKinney, J. G., Hilborn, R., Hauser, L. y Naish, A. K. (2020). Power of a dual-use SNP panel for pedigree reconstruction and population assignment. *Ecology and Evolution*, 10: 9522-9531. doi: 10.1002/ece3.6645
- McKinney, G. J., Seeb, J. E. y Seeb, L. W. (2017). Managing mixed-stock fisheries: Genotyping multi-SNP haplotypes increases power for genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, 74: 429-434. <https://doi.org/10.1139/cjfas-2016-0443>
- Merimans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21: 2839-2846. <https://doi.org/10.1111/j.1365-294X.2012.05578.x>
- Micheli, F., Saenz-Arroyo, A., Greenley, A., Vázquez, L., Espinoza-Montes, J. A., Rossetto, M., y de Leo-Giulio, A. (2012). Evidence That Marine Reserves Enhance Resilience to Climatic Impacts. *PLoS One*, 7(7): e40832. doi: 10.1371/journal.pone.0040832

- Miller, A. D., Rooney, A. V., Rasic, G., Ierodiaconou, D. A., Gorfine, H. K., Day, R., Wong, C., Hoffman, A. A. y Weeks, R. (2016). Contrasting patterns of population connectivity between regions in a commercially important mollusc *Haliotis rubra*: integrating population genetics, genomics and marine LiDAR data. *Molecular Ecology*, 25: 3845-3864. <https://doi.org/10.1111/mec.13734>
- Morales-Bojórquez, E., Muciño-Díaz, M. O., Vélez-Barajas, J. A. (2008). Analysis of the Decline of the Abalone Fishery (*Haliotis fulgens* and *H. corrugata*) along the Westcentral Coast of the Baja California Peninsula, Mexico. *Journal of Shellfish Research*, 27(4): 865-870. [https://doi.org/10.2983/0730-8000\(2008\)27\[865:AOTDOT\]2.0.CO;2](https://doi.org/10.2983/0730-8000(2008)27[865:AOTDOT]2.0.CO;2)
- Morgan, L. E. y Shepherd, A. S. (2006). Population and spatial structure of two common temperate reef herbivores: abalone and sea urchins. En: Kritzer, P. J y Sale, F. P. *Marine metapopulations*. Elsevier, Amsterdam.
- Morin, P. A., Luikart, G., Wayne, R. K. y SNP workshop group. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, 19(4): 208-216. <https://doi.org/10.1016/j.tree.2004.01.009>
- Mullins, R. B., McKeown, N. J., Sauer, W. H. H., Shaw, P. W. y Grant, W. S. (2018). Genomic analysis reveals multiple mismatches between biological and management units in yellowfin tuna (*Thunnus albacares*). *ICES Journal of Marine Science*, 75(6): 2145-2152. <https://doi.org/10.1093/icesjms/fsy102>
- Mundy, C., Chik, C. R., Stobart, B. y Strain, L. (2020). Status of Australian fish stocks report, blacklip abalone. Disponible en: [https://www.fish.gov.au/2020-Reports/blacklip\\_abalone](https://www.fish.gov.au/2020-Reports/blacklip_abalone)
- Munguía-Vega, A., Sáenz-Arroyo, A., Greenley, A. P., Espinoza-Montes, J. A., Palumbi, S. R., Rossetto, M. y Micheli, F. (2015). Marine reserves help preserve genetic diversity after impacts derived from climate variability: Lessons from the pink abalone in Baja California. *Global Ecology and Conservation*, 4: 264-276. <https://doi.org/10.1016/j.gecco.2015.07.005>
- Nam, B. H., Kim, H., Seol, D., Kim, H., Noh, E. S., Kim, E.M. Noh, S. E., Kim, M. E., Noh, K. J., Kim, O. Y., Park, Y. J. y Kwakl W. (2021). Genotyping-by-Sequencing of the regional Pacific abalone (*Haliotis discus*) genomes reveals population structures and patterns of gene flow. *PLoS One*, 16(4): e0247815. <https://doi.org/10.1371/journal.pone.0247815>
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, 70(12): 3321-3323. <https://doi.org/10.1073/pnas.70.12.3321>
- Nielsen, E. E., Cariani, A., Aoidh, E. M., Maes, G. E., Milano, I., Ogden, R., Taylor, M., Hemmer-Hansen, J., Babbucci, M., Bargelloni, L., Bekkevold, D., Diopere, E., Grenfell, L., Helyar, S., Limborg, T. M., Martinsohn, T. J., McEwing, R., Panitz, F., Patarnello, T., Tinti, F., Van Houdt, K. J. J., Volckaert, A. M. F., Waples, S. R., FishPopTrace Consortium y Carvalho, R. G. 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*, 3, 851: <https://doi.org/10.1038/ncomm.s1845>.
- Olson, N. D., Lund, S. P., Colman R. E., Foster J. T., Sahl J. W., Schupp J. M., Keim, P., Morrow J. B., Salit M. L., Zook J. M. (2015). Best practices for evaluating single nucleotide variant

- calling methods for microbial genomics. *Frontiers in Genetics*, 6. <https://doi.org/10.3389/fgene.2015.00235>
- Ortiz-Quintanilla, M. y Leon-Carballo, G. (1988). Recursos pesqueros de México y sus perspectivas, recurso abulón *Haliotis* spp. En: Recursos Pesqueros del País. Secretaria de Pesca-INP. 1:11-51
- Paris, J. R., Stevens, J. R. y Catchen, M. J. (2017). Lost in parameter space: a road map for STACKS. *Methods in Ecology and Evolution*, 8: 1360-1373. <https://doi.org/10.1111/2041-210X.12775>
- Pecoraro, C., Babbucci, M., Franch, R., Rico, C., Papetti, C., Chassot, E., Bodin, N., Cariani, A., Bargelloni, L. y Tini, F. (2018). The population genomics of yellowfin tuna (*Thunnus albacares*) at global geographic scale challenges current stock delineation. *Scientific Reports*, 8: 13890. <https://doi.org/10.1038/s41598-018-32331-3>
- Perez-Enríquez, R., Díaz-Viloria, N., Cruz-Hernández, P., Aranceta-Garza, F., Gutiérrez-González, J. L., Arciniega de los-Santos, A. y Max-Aguilar, A. 2016. Estudios de genética en poblaciones de abulón y sus aplicaciones en ordenamiento pesquero. *Recursos Naturales y Sociedad*, 2(2): 24-39. <http://dspace.cibnor.mx:8080/handle/123456789/600>
- Peterson, B. K., Weber, N. J., Kay, H. E., Fisher, S. H., Hoekstra, E. H. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS One*, 7(5): e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Philippi, R. A. (1845). Diagnoses testaceorum quorundam novorum. *Zeitschrift für Malakozoologie*, 2(10): 147-152. <https://www.biodiversitylibrary.org/page/16291735>
- Pritchard, J.K., Stephens, M. y Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2): 945-959. doi: 10.1093/genetics/155.2.945.
- Ponce-Díaz, G. y Chávez, A. E. (2000). Evaluation of the green abalone *Haliotis fulgens* fishery in Bahía Asuncion, Baja California Sur, Mexico. *Ciencias Marinas*, 26(3): 393-412. <https://doi.org/10.7773/cm.v26i3.598>
- Ponce-Díaz, G., Arreguín-Sánchez, F., Hernández-Llamas, A., del Monte-Luna, P. y Ramade-Villanueva, M. (2013). Estimación de la captura ilegal del abulón y la langosta en la costa de la Península de Baja California. *Ciencias Marinas*, 39(3): 323-329. ISSN 0185-3880
- Prince, J. (2005). Combating the tyranny of scale for haliotids: micro-management for microstocks. *Bulletin of Marine Science*, 76(2): 557-577.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. y Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3): 559-75. doi: 10.1086/519795
- Puritz, J. B., Matz, V. M., Toonen, J. R., Webe, N. J., Bolnick, I. D. y Bird, E. C. (2014). Demystifying the RAD fad. *Molecular Ecology*, 23: 5937-5342. <https://doi.org/10.1111/mec.12965>
- R core Team. (2014). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

- Reiss, H., Hoarau, G., Dickey-Collas, M. y Wolff, J. W. (2009). Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, 10: 361-395. <https://doi.org/10.1111/j.1467-2979.2008.00324.x>
- Reynes, L., Aurelle, D., Chevalier, C., Pinazo, C., Valero, M., Mauger, S. y Thibaut, T. (2021). Population genomics and modeling shed light on dispersal events in the Mediterranean endemic *Ercaria zosteroides* (=Cystoseira zosteroides) (Fucales). *Frontiers in Marine Science*, 8: 1-18. doi: 10.3389/fmars.2021.683528
- Rice, W.R. (1989). Analysis tables of statistical tests. *Evolution*. 43(1): 223-225. <https://doi.org/10.2307/2409177>
- Rogers-Bennett, L., Klamt, R. y Catton, C. A. (2021). Survivors of climate driven abalone mass mortality exhibited declines in health and reproduction following kelp forest collapse. *Frontiers in Marine Science*, 8: 725134. doi: 10.3389/fmars.2021.725134
- Rosenberg, N. A. (2004). Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4: 137-138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Roughan, M., Mace, A., Largier, J., Morgan, S., Fisher, J. y Carter, M. (2005). Subsurface recirculation and larval retention in the lee of a small headland: A variation on the upwelling shadow theme. *Journal of Geophysical Research*, 110(C10). 10.1029/2005JC002898.
- Ruzzante, E. D., Mariani, S., Bekkevold, D., André, C., Mosegaard, H., Clausen, A. W. L., Dahlgren, G. T., Hutchinson, F. W., Hatfield, M. C. E., Torstensen, E., Brigham, J., Simmonds, J. E., Laikre, L., Larsson, C. L., Stet, J. M. R., Ryman, N. y Carvalho, R. G. (2006). Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proceedings of the Royal Society B*, 273: 1459-1464. <https://doi.org/10.1098/rspb.2005.3463>
- Sandoval-Castillo, J., Robinson, A. N., Hart, M. A., Strain, W. L. y Beheregary, B. L. (2018). Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevis*), along a longitudinal environmental gradient. *Molecular Ecology*, 27(7): 1603-1620. doi: 10.1111/mec.14526
- Sambrook, J., Fritsch, F. E. y Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. 3<sup>a</sup> ed. Cold Spring Harbor Laboratory Press, Nueva York.
- Sevilla, M. L. (1971). Desarrollo gonádico del abulón azul *Haliotis fulgens* Philippi. *Revista de la Sociedad Mexicana de Historia Natural*, 32: 129-139
- Shackell, G. H. y Dodds, K. G. (2008). DNA-based traceability of meat. En: Toldrá, F. Meat biotechnology. Springer. España.
- Sierra-Rodríguez, P., Muciño-Díaz, M., Gutiérrez-González, J. L., Turrubiates-Morales J. (2006). La Pesquería de Abulón. En: Sustentabilidad y pesca responsable en México: Evaluación y Manejo. México: Instituto Nacional de la Pesca, México.
- Sundqvist, L., Keenan, K., Zackrisson, M., Prodöhl, P. y Kleinhans, D. (2016). Directional genetic differentiation and relative migration. *Ecology and Evolution*, 6(199): 3461-3475. doi: 10.1002/ece3.2096
- Sylvester, A. V. E., Bentzen, P., Bradbury, R. I., Clément, M., Pearce, J., Horne, J. y Beijo, G. R. (2018). Applications of random forest features selection for fine-scale genetic population assignment. *Evolutionary Applications*, 11: 153-165. doi: 10.1111/eva.12524

- Tegner, M. J. y Butler, R. A. (1985). Drift-tube study of the dispersal potential of green abalone (*Haliotis fulgens*) larvae in the southern California Bight: implications for recovery of depleted populations. *Marine Ecology Progress Series*, 26(1/2): 73-84. <http://www.jstor.org/stable/24817609>
- Terry, A., Bucciarelli, G. y Bernardi, G., (2000). Restricted gene flow and incipient speciation in disjunct Pacific Ocean and Sea of Cortez populations of a reef fish species, *Girella nigricans*. *Evolution*, 54: 652-659. doi: 10.1111/j.0014-3820.2000.tb00066.x.
- Thorson, G. (1950). Reproductive and larval ecology of marine bottom invertebrates. *Biological Reviews*, 25: 1-45. <https://doi.org/10.1111/j.1469-185X.1950.tb00585.x>
- Valencia, L. M., Martins, A., Ortiz, E. M. y Di Fiore, A. (2018). A RAD-sequencing approach to genome-wide marker discovery, genotyping, and phylogenetic inference in a diverse radiation of primates. *PLoS One*, 13(8): e0201254. <https://doi.org/10.1371/journal.pone.0201254>
- Van der Merwe, A. E. (2009). Population genetic structure and demographical history of South African abalone, *Haliotis midae*, in a conservation context. *Tesis Doctoral*. Universidad de Stellenbosch, Departamento de Genética, Sudáfrica.
- Vargas-Peralta, C. E., Araneda, C., Galindo-Sánchez, C. E., Larraín M. A., Del Río-Portilla, M. A. y Lafarga-De la Cruz, F. (2022). Species identification in *Haliotis* genus from the northeastern Pacific Ocean using genome-wide RAD-SNPs. *Food Control*, 138:108979. <https://doi.org/10.1016/j.foodcont.2022.108979>
- Vélez-Arellano, N. M., García-Domínguez, A. F., Lluch-Cota, B. D., Gutiérrez-González, J. L., Salcido-Guevara, L. A., Sanchez-Cardenas, R. (2016). Morphophysiological Indices of the Green Abalone *Haliotis fulgens* Philippi, 1845 at Mexican Ocean Pacific Coast. *Turkish Journal of Fisheries and Aquatic Sciences*, 17: 31-39. <http://dspace.cibnor.mx:8080/handle/123456789/1902>
- Vélez-Arellano, N. M., Valenzuela-Quiñonez, F., García-Domínguez, F. A., Lluch-Cota, D. B., Gutiérrez-González, J. L. y Martínez-Rincón, R. O. (2020). Long-term analysis on the spawning activity of green (*Haliotis fulgens*) and pink (*Haliotis corrugata*) abalone along the central west coast of Baja California. *Fisheries Research*, 228: 105588. <https://doi.org/10.1016/j.fishres.2020.105588>
- Vélez-Arellano, N. M. (2016). Tácticas reproductivas de *Haliotis fulgens* Philippi, 1845 y *Haliotis corrugata* W. Wood, 1828 (Gastropoda: Archeogastropoda) en zonas de pesca en la costa occidental de Baja California Sur. Tesis de Doctorado, Departamento de Ciencias Marinas, Centro de Investigaciones Biológicas del Noroeste, México.
- Venables, N. W. y Ripley, D. B. (2002). *Statistics and Computing*. 4a ed. Springer, Nueva York.
- Von der Hyden, S., Beger, M., Toonen, J. R., van Herwerden, L., Junino-Meñez, M. A., Ravago-Gotanco, R., Fauvelot, C. y Bernardi, G. (2014). The application of genetics to marine management and conservation: examples from the Indo-Pacific. *Bulletin of Marine Science*, 90(1): 123-158. <https://doi.org/10.5343/bms.2012.1079>
- Waples, S. R. (2010). High-grading bias: subtle problems with assessing power of selected subset of loci for population assignment. *Molecular Ecology*, 19: 2599-2601. doi: 10.1111/j.1365-294X.2010.04675.x

- Weir, B. S., y Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population-structure. *Evolution*, 38: 1358-1370. <https://doi.org/10.2307/2408641>
- Weist, P., Schade, M. F., Damerau, M., Barth, I. J. M., Dierking, J., André, C., Petereit, C., Reusch, T., Jentoft, S., Hanel, R. y Krumme, U. (2019). Assessing SNP-markers to study population mixing and ecological adaptation in Baltic cod. *PLoS One*, 14(6): e0218127. doi: 10.1371/journal.pone.0218127
- Zúñiga, G., Guzmán Del Proó, S. A., Cisneros, R. y Rodríguez, G. (2000). Genetic population analysis of abalone *Haliotis fulgens* (Mollusca: Gastropoda) in Baja California, Mexico. *Journal of Shellfish Research*, 19(2): 540-547

## 11. ANEXOS

### Anexo A. Preparación de librerías ddRAD

A continuación, se describe el protocolo de secuenciación de ADN Asociado a Sitios de Restricción de Doble Digestión (ddRAD-Seq) descrito por Peterson *et al.*, (2012) con ligeras modificaciones.

Primero se realiza una doble digestión del ADN de cada individuo utilizando dos enzimas de restricción EcoR1-HF y Msp1 (New England BioLabs®). Las digestiones se llevaron a cabo en un volumen de 30 µl con 500 ng de ADN, 1X de Buffer CutSmart y 10 U de cada enzima y se dejaron en digestión toda la noche a temperatura ambiente. Se realizó una purificación con 1.5X AMPure XP beads (Beckman Coulter®) y se eluyó en 30 µl 0.1X TE.

Posteriormente, se realizó la ligación de adaptadores específicos para unirse a los extremos resultantes de la digestión, dichos adaptadores son **P1** que contiene el cebador de amplificación (*Forward*) para la plataforma Illumina, el *barcode* (48 diferentes) y es complementario al corte de enzima EcoRI-HF, y el adaptador **P2** que contiene el extremo complementario al corte de la enzima Msp1 y es común para todas las muestras. La mezcla de ligación se llevó a cabo en un volumen de 40 µl con 30 µl de ADN digerido, 1x de buffer T4 ligasa, 100 U de T4 DNA ligasa (New England BioLabs®) y 2 mM de cada adaptador. Las reacciones se dejaron toda la noche a temperatura ambiente. El proceso fue detenido mediante calor (10 minutos a 65°C). Los productos ligados fueron purificados con 1.5x AMPure XP beads (Beckman Coulter®) y se eluyeron en un volumen de 30 µl de 0.1x TE. Para verificar la correcta ligación de los adaptadores, se realizó una PCR (PCR 1) la cual se llevó a cabo en un volumen de 12.5 µl con 1x Gotaq® (Promega), 0.8mM de cada *primer* (Forward PCR1) (Reverse Msp1\_P2) y 1µl del ADN ligado. El programa de PCR fue una desnaturalización inicial 95°C por 2 min, 20 ciclos de desnaturalizado a 94°C por 30 seg, alineamiento con 60°C por 30 seg, y una extensión a 72°C por 45 seg, Los productos de PCR se visualizaron en geles de agarosa-synergel al 2%.

Posterior a la verificación, se realizó una cuantificación por fluorometría en Qubit™ 4

Fluorometer (Thermo Scientific®), para realizar un pool equimolar de las muestras. Los pools se crearon al mezclar los 48 individuos que contienen un *barcode* distinto, en total se crearon 4 pools. Cada pool fue concentrado con 1.5× AMPure XP beads (Beckman Coulter®) y eluido en 30 µl de TE 0.1×. A cada pool se le realizó una selección de fragmentos (de 344 pb a 408 pb) mediante Pippin Prep (Sage Science) siguiendo el protocolo del fabricante. Con los productos obtenidos en la selección de fragmentos, se llevó a cabo un PCR (*PCR 2*) con el fin de enriquecer las librerías, agregar un index específico a cada pool y así como los adaptadores de secuenciación Illumina. Se realizaron 8 reacciones de PCR por cada pool, las reacciones de PCR se llevaron a cabo en un volumen de 20 µl con 1× de Phusion HF Buffer, 0.5 U de enzima Phusion® DNA polimerase (New England Biolabs®), 0.5 mM de cada primer (Forward PCR1) (Reverse PCR2\_Idx\_4; PCR2\_Idx\_5; PCR2\_Idx\_1; PCR2\_Idx\_6) y 3µl de ADN; Las condiciones de PCR fueron una desnaturalización inicial 95°C por 2 min, 10 ciclos de desnaturalización a 94°C por 30 seg, alineamiento con 60°C por 30 seg, y extensión a 72°C por 45 seg.

Los productos de cada reacción de PCR fueron mezclados y concentrados por cada pool respectivamente con 1.5× AMPure XP beads (Beckman Coulter®) y eluidos en 40 µl. Los productos de PCR de cada pool fueron verificados en geles de agarosa-synergel al 2%. Posteriormente, cada pool enriquecido fue cuantificado en Qubit™ 4 Fluorometer (Thermo Scientific®) para poder realizar una mezcla equimolar en la librería final. La librería final consistió en dos pools con dos index cada uno y 96 individuos en total. Se construyeron 2 librerías finales, las cuales fueron enviadas a secuenciar a un prestador de servicios NOVOGENE® (Sacramento, CA) en dos líneas de secuenciación Illumina Hiseq-4000 con lecturas pareadas de 150 pb.

#### **Anexo B.** Descripción del pipeline STACKS y los principales parámetros

El pipeline de *STACKS* está formado por varios módulos. Primero se utilizó el programa *process\_radtags* para *demultiplexar* las lecturas crudas; en este paso se eliminan las lecturas de baja calidad (utilizando puntajes Phred 33 provistos en los archivos FASTQ), revisa que el *barcode* y el sitio de corte RAD estén intactos, así mismo compila y ordena las lecturas en

archivos separados para recuperar las lecturas por individuo de la librería. En este paso, también se cortan las lecturas (*trimming*) ya que es común que al final de la secuenciación se presente un mayor error de secuenciación. Para esto, es importante introducir en el programa información tal como las enzimas utilizadas y una relación de los *barcodes-index* e *ID* por individuo.

Debido a que no existía genoma de referencia para esta especie cuando se realizaron los primeros análisis, se ensamblaron los loci *de novo* utilizando el módulo *denovo\_map.pl* en el *pipeline STACKS*. Este programa ejecuta de manera consecutiva las tres etapas principales del *pipeline* y dentro de ellas existen 3 parámetros principales (*-m*, *-M* y *-n*), los cuales determinan las características de los stacks, el catálogo y obtención de genotipos.

De manera general, *denovo\_map.pl* primero ejecuta el módulo *ustacks* en donde toma el conjunto de secuencias (resultado de *process\_radtags*) y las alinea en pilas o “stacks” de correspondencia exacta formando los alelos putativos. El número de secuencias idénticas requeridas para formar un stack está gobernado por *-m* (profundidad mínima de secuencias idénticas para formar un stack). Después de que se forman los alelos putativos, se comparan los stacks para emparejar los alelos putativos en loci putativos, este paso está gobernado por el parámetro *-M*, el cual controla el número máximo de disimilitudes entre secuencias para considerarlo un locus putativo.

El segundo módulo es *cstacks*, en el cual después de que los loci son ensamblados en cada muestra individual estos son comparados a través de las muestras para emparejar loci homólogos para crear un catálogo de loci putativos para la población. Este módulo está gobernado por el parámetro *-n*, el cual controla el número de disimilitudes permitidas entre los stacks formados.

Los siguientes módulos consisten en: *sstacks*, un módulo en el cual todos los loci putativos construidos en *ustacks* son comparados contra los loci del catálogo. *tsv2bam* que transpone la

matriz de stacks de muestras a locus y por último *gstacks* que ensambla las secuencias pair-end y single-end, alinea las secuencias al locus y llama los SNPs.

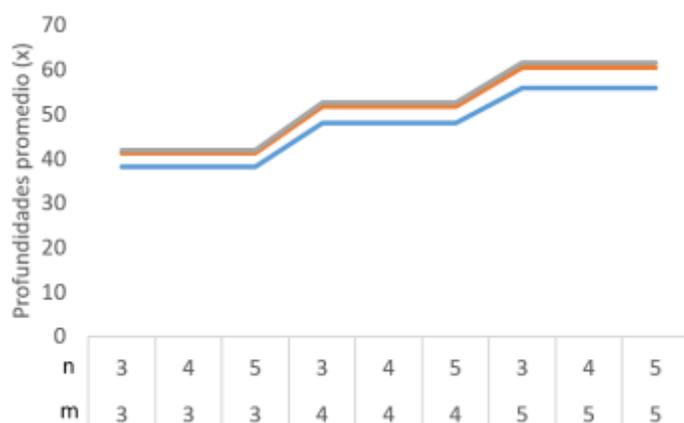
Posteriormente, se utilizó el programa *populations* realiza el filtrado de los loci y realiza análisis de genética de poblaciones, además de que exporta los resultados en varios formatos de salida estándar (*genepop*, *.vcf*, *structure*, etc). En este paso se puede incluir un mapa de poblaciones en donde se especifica qué individuos pertenecen a qué localidades para realizar los estadísticos de genética de poblaciones por localidades o global. Este programa está gobernado principalmente por los parámetros *-p* (número mínimo de localidades en donde cada locus debe estar presente), *-r* (% mínimo de individuos en los que cada locus debe estar presente) y *min\_maf* (frecuencia alélica del alelo menor en todas las localidades).

### **Anexo C.** Optimización de parámetros, Identificación de SNPs y obtención de genotipos

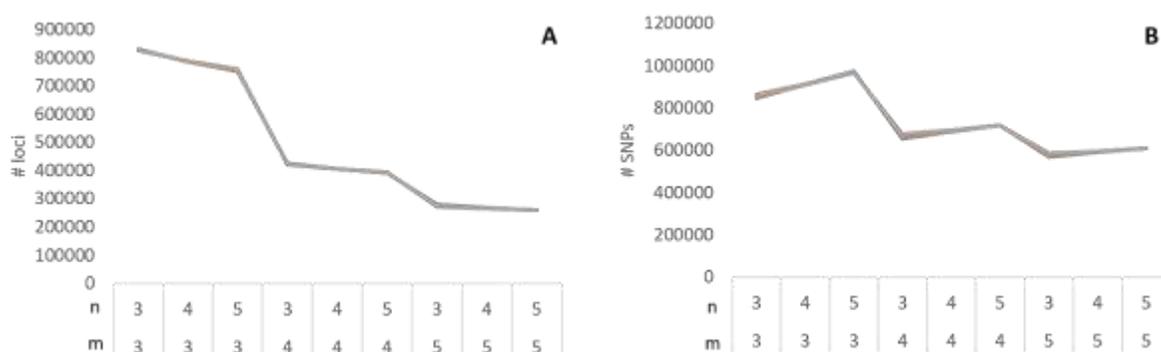
Para elegir el set de parámetros a utilizar en el ensamblado de loci y obtención de SNPs se realizaron 81 pruebas modificando un solo parámetro en cada una de ellas (Fig. 3).

Primero, se realizaron pruebas con el módulo *denovo\_map.pl* para ver cómo se comportaban las profundidades promedio, el número de loci y número de SNPs obtenidos a nivel individual. En la Fig. 18 se puede observar que conforme se utilizan valores más altos de *-m* y *-M* la profundidad promedio incrementa y no se observa un efecto de *-n* sobre esta.

Sin embargo, se observa que el número de loci ensamblados disminuye conforme aumenta el valor de *-m* y *-n* (Fig. 19A) al igual que disminuye la cantidad de SNPs obtenidos al aumentar el valor de *-m* y aumenta ligeramente con el aumento de *-n* (Fig. 19B). También se puede observar que este valor no está afectado por los valores de *-M* puesto que la tendencia es prácticamente la misma con los 3 valores probados. En general al ir aumentando la *-m* de 3 a 4 se observa una caída dramática en el número de loci ensamblados, sin embargo, la pendiente se suaviza cuando el valor de *-m* pasa de 4 a 5.



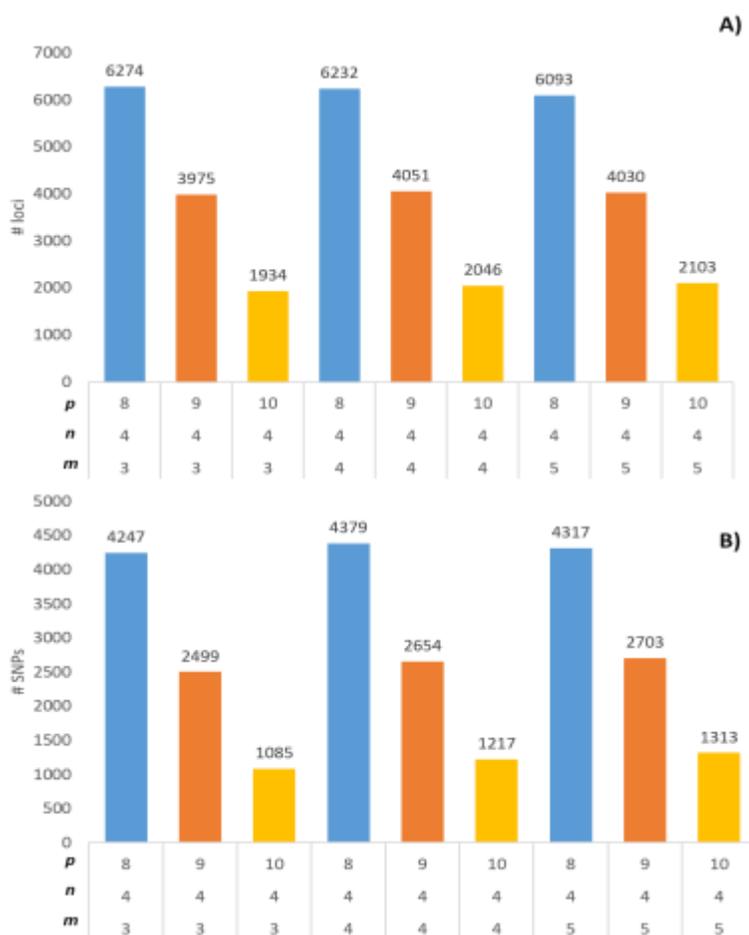
**Figura 18.** Profundidades promedio ( $x$ ) para las pruebas de parametrización ( $-m$ ,  $-M$  y  $-n$ ). La línea azul representa los valores obtenidos con  $-M$  1, en naranja los valores con  $-M$  3 y en gris con  $-M$  4.



**Figura 19.** A) Número de loci ensamblados y B) número de SNPs obtenidos con las pruebas de parametrización ( $-m$ ,  $-M$  y  $-n$ ). En línea azul se representan los valores obtenidos con  $-M$  1, en naranja los valores con  $-M$  3 y en gris con  $-M$  4.

Posterior al análisis realizado a nivel individual, se evaluó el efecto que tiene  $-p$  a nivel de localidades con el programa *populations* (Fig. 3). Dado que  $-M$  tuvo mayor efecto en los valores de profundidades promedio y no en el número de loci y SNPs se presentan los resultados sólo para las pruebas de  $-M$  3. En cuanto a los valores de  $-m$  y  $-n$  se presentan sólo los resultados obtenidos con  $-m$  de 3 a 5 y  $-n$  4 (respetando la regla sugerida de  $-n = M+1$ ).

En la figura 20 se observa que a pesar de que el número de loci y SNPs es mucho mayor con valores de  $-m$  más pequeños estos loci no son compartidos a lo largo de todas las localidades por lo que se pierden y la diferencia se vuelve imperceptible al realizar los *populations*. Lo que destaca es una notoria disminución en el número de loci y SNPs conforme aumenta el valor de  $-p$  sin importar el valor de  $-m$  y/o  $-n$ .

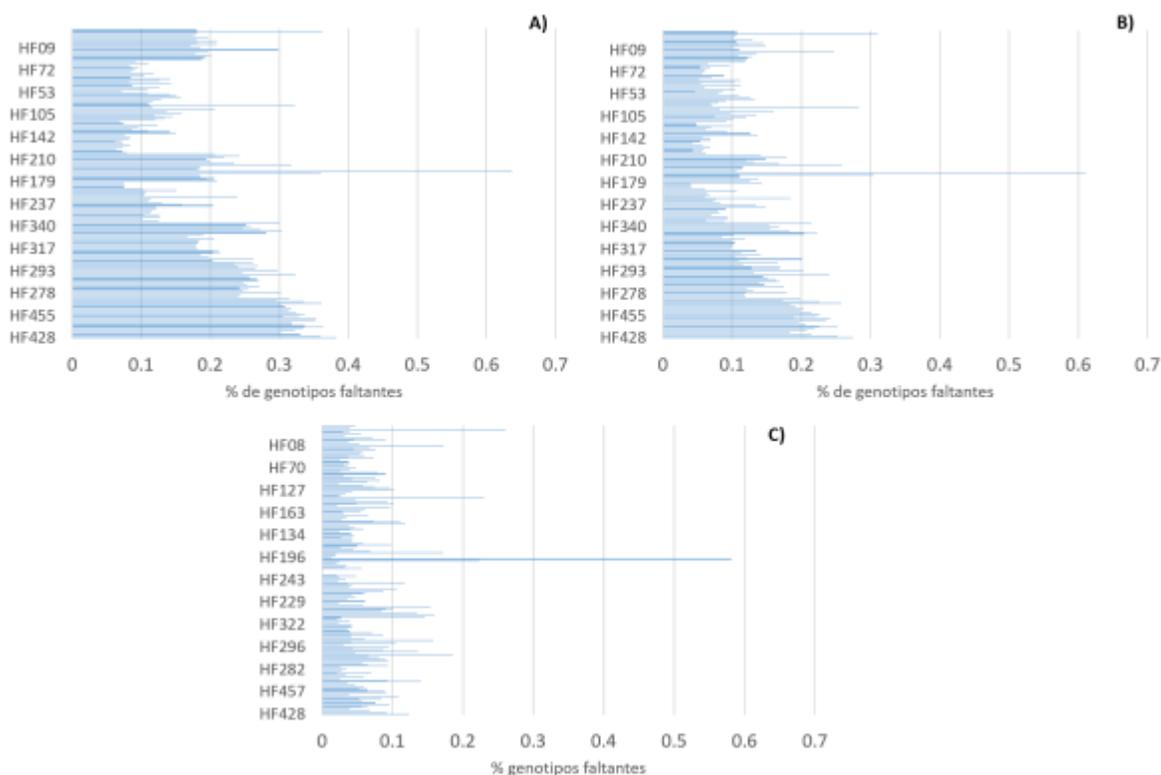


**Figura 20.** A) Número de loci ensamblados y B) número de SNPs obtenidos con las pruebas de parametrización utilizando  $-M$  3,  $-n$  4 y distintos valores de  $-m$  y  $-p$ . Los colores de las barras representan los valores para cada  $p$ : azules  $-p$  8, naranjas  $-p$  9 y amarillas  $-p$  10.

Con las pruebas anteriores se seleccionó el set de datos obtenidos con los siguientes parámetros ( $-M$  3,  $-m$  5 y  $-n$  4), ya que con estos parámetros se obtuvieron mayor número de SNPs en la mayoría de las  $-p$  ( $-p$  8, 4313;  $-p$  9, 2703 SNPs;  $-p$  10, 1313 SNPs).

*Filtrado de genotipos faltantes por individuo*

Antes de realizar pruebas de estructura genética poblacional se utilizó el programa vcfTools (Danecek *et al.*, 2011) para evaluar el porcentaje de genotipos faltantes para cada individuo con los SNPs obtenidos para  $-p$  8, 9 y 10. En la figura 21 se observa que a pesar de recuperar un mayor número de marcadores utilizando una  $-p$  de 8 (A) el valor de genotipos faltantes es mayor al 30% el cual disminuye al 20% y 10% con una  $-p$  de 9 (B) y 10 (C) respectivamente.



**Figura 21.** Porcentaje de genotipos faltantes por individuos con distintos valores de  $-p$ . **A)**  $-p$  8, **B)**  $-p$  9 y **C)**  $-p$  10.

A partir de estos resultados se decidió evaluar si la estructura poblacional se mantenía constante independientemente del set de SNPs utilizado. Con estas pruebas no se encontraron diferencias en los patrones de estructuración con alguno de los sets de datos, por lo que se decidió utilizar los parámetros y el set de datos con menor número de datos faltantes ( $-M$  3,  $-m$  5,  $n$  4,  $-p$  10,  $-r$  0.8).

Del set de datos obtenido con la combinación de parámetros seleccionada se eliminaron para las siguientes pruebas los individuos que tuvieron un valor de genotipos faltantes mayor al 10%. Ver Tabla 9 para detalles de los tamaños de n por localidad después de los filtros.

**Anexo D.** Pruebas no paramétricas de Gao para comparaciones múltiples de valores de Heterocigosidad observada (*Ho*) en 10 localidades de *H. fulgens*

**Prueba de Gao.** Valores de P para múltiples comparaciones no paramétricas (triángulo inferior) con su valor de alfa (triángulo superior) para *Ho* usando 1,270 SNPs en 10 localidades de *H. fulgens*. En negrita se muestran las comparaciones significativas ( $P < \alpha$ ).

<i>Ho</i>	GI	SJI	FSJ	CI	PE	TN	TS	AN	AS	Bo
<b>GI</b>	*	0.010	0.010	0.035	0.040	0.015	0.050	0.025	0.020	0.030
<b>SJI</b>	0.772	*	0.015	0.040	0.050	0.020	0.050	0.030	0.025	0.035
<b>FSJ</b>	0.228	0.059	*	0.030	0.035	0.010	0.040	0.020	0.015	0.025
<b>CI</b>	0.035	<b>0.027</b>	0.814	*	0.010	0.025	0.015	0.015	0.020	0.010
<b>PE</b>	<b>0.006</b>	<b>0.005</b>	0.588	0.645	*	0.030	0.010	0.020	0.025	0.030
<b>TN</b>	0.027	0.023	0.224	0.980	0.971	*	0.035	0.015	0.010	0.020
<b>TS</b>	<b>0.001</b>	<b>0.001</b>	0.319	0.535	0.862	0.849	*	0.025	0.030	0.020
<b>AN</b>	0.038	0.066	0.658	0.944	0.933	0.786	0.778	*	0.010	0.010
<b>AS</b>	0.031	> 0.025	0.579	0.925	0.893	0.650	0.773	0.665	*	0.015
<b>Bo</b>	0.046	0.046	0.862	0.789	0.823	1.000	0.762	0.846	0.949	*

**Anexo E.** Pruebas no paramétricas de Gao para comparaciones múltiples de valores de *Riqueza alélica* (*Ra*) en 10 localidades de *H. fulgens*

**Prueba de Gao.** Valores de P para múltiples comparaciones no paramétricas (triángulo inferior) con su valor de alfa (triángulo superior) para *Ra* usando 1,270 SNPs en 10 localidades de *H. fulgens*. En negrita se muestran las comparaciones significativas ( $P < \alpha$ ).

<i>Ra</i>	GI	SJI	FSJ	CI	PE	TN	TS	AN	AS	Bo
<b>GI</b>	*	0.0153	0.05	0.0102	0.0353	0.0402	0.0303	0.05	0.0253	0.0203
<b>SJI</b>	0.2817	*	0.0402	0.0102	0.0253	0.0303	0.0203	0.0353	0.0153	0.0102
<b>FSJ</b>	<b>0.0001</b>	0.0876	*	0.05	0.0203	0.0153	0.0253	0.0102	0.0303	0.0353
<b>CI</b>	0.0957	0.8338	0.0664	*	0.0303	0.0353	0.0253	0.0402	0.0203	0.0153
<b>PE</b>	0.0503	0.8542	0.2625	0.933	*	0.0102	0.0102	0.0153	0.0153	0.0203
<b>TN</b>	<b>0.0062</b>	0.5187	0.3281	0.5208	0.4019	*	0.0153	0.0102	0.0203	0.0253
<b>TS</b>	0.0814	0.8888	0.1367	0.9517	0.6771	0.5307	*	0.0203	0.0102	0.0153
<b>AN</b>	<b>0.0015</b>	0.3205	0.5028	0.2915	0.3052	0.316	0.3483	*	0.0253	0.0303
<b>AS</b>	0.1735	0.952	0.1115	0.9979	0.7839	0.4574	0.7805	0.2532	*	0.0102
<b>Bo</b>	0.2464	0.9849	0.0663	0.9878	0.7598	0.3451	0.8192	0.1646	0.7675	*

**Anexo F.** Matriz de asignación poblacional con los seis paneles de SNPs a escala global  
Valores de media, mediana y desviación estándar (SD) de las pruebas de precisión de asignación de los individuos a su población de origen a escala global (GI, GCN y GCS) utilizando dos proporciones de individuos para entrenar a los modelos (0.75 y 0.8).

panel	prop.ind	GI			GCN			GCS		
		media	mediana	SD	media	mediana	SD	media	mediana	SD
50	0.75	0.97	1	0.12	0.09	0	0.19	0.97	1	0.07
	0.8	0.97	1	0.12	0.09	0	0.21	0.96	1	0.09
100	0.75	0.98	1	0.11	0.30	0.33	0.31	0.96	1	0.08
	0.8	0.95	1	0.17	0.40	0.5	0.35	0.95	1	0.09
200	0.75	1	1	0	0.31	0.33	0.27	1	1	0.01
	0.8	1	1	0	0.38	0.50	0.33	1	1	0.02
300	0.75	1	1	0	0.41	0.33	0.29	1	1	0.02
	0.8	1	1	0	0.36	0.50	0.38	1	1	0
400	0.75	1	1	0	0.62	0.67	0.27	0.97	1	0.06
	0.8	1	1	0	0.62	0.50	0.37	0.99	1	0.04
500	0.75	1	1	0	0.83	1.00	0.25	0.96	1	0.06
	0.8	1	1	0	0.88	1.00	0.24	0.96	1	0.08

**Anexo G.** Matriz de asignación poblacional con los seis paneles de SNPs a escala costera  
Valores de media, mediana y desviación estándar (SD) de las pruebas de precisión de asignación de los individuos a su población de origen a escala costera (GCN y GCS) utilizando dos proporciones de individuos para entrenar a los modelos (0.75 y 0.8).

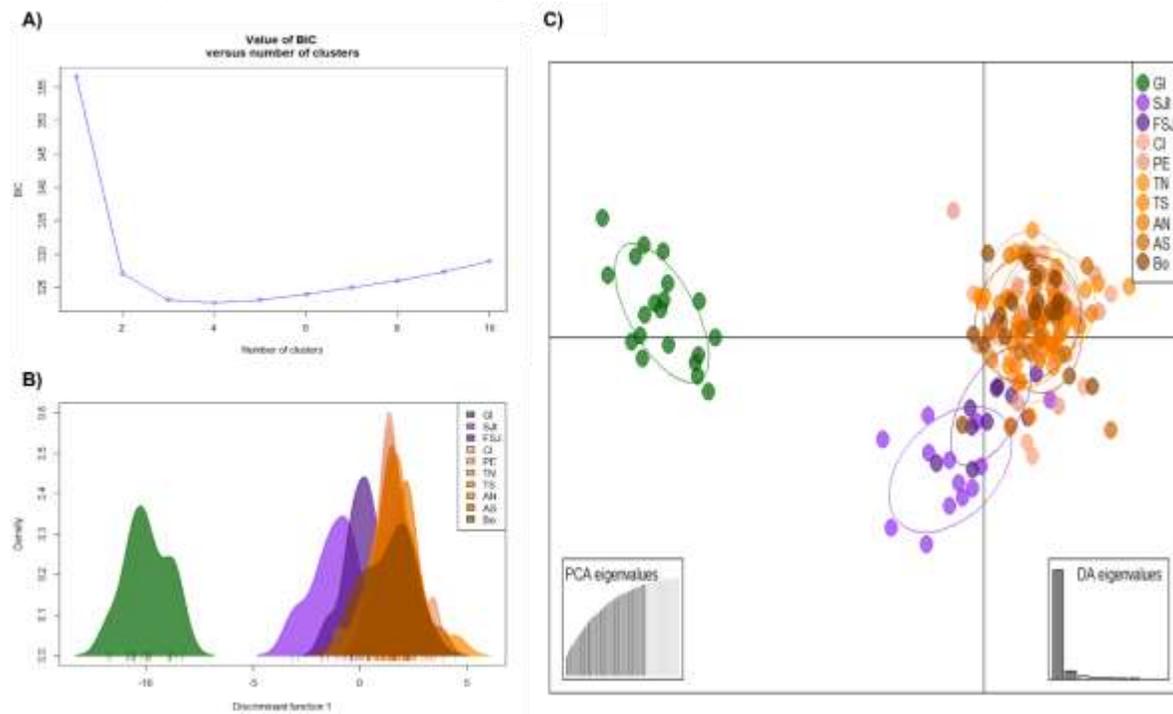
panel	prop.ind	GCN			GCS		
		media	mediana	SD	media	mediana	SD
50	0.75	0.04	0	0.11	0.98	1	0.06
	0.8	0.03	0	0.13	0.97	1	0.12
100	0.75	0.52	0.67	0.29	0.99	1	0.03
	0.8	0.58	0.5	0.34	0.98	1	0.05
200	0.75	0.42	0.33	0.31	1	1	0.02
	0.8	0.44	0.5	0.35	0.99	1	0.03
300	0.75	0.70	0.67	0.25	0.96	1	0.08
	0.8	0.51	0.5	0.35	0.99	1	0.05
400	0.75	0.87	1	0.21	0.94	1	0.08
	0.8	0.88	1	0.25	0.94	1	0.09
500	0.75	0.86	1	0.23	0.95	1	0.07
	0.8	0.91	1	0.23	0.96	1	0.07

**Anexo H.** Matriz de asignación poblacional utilizando todos los marcadores a escala global y costera

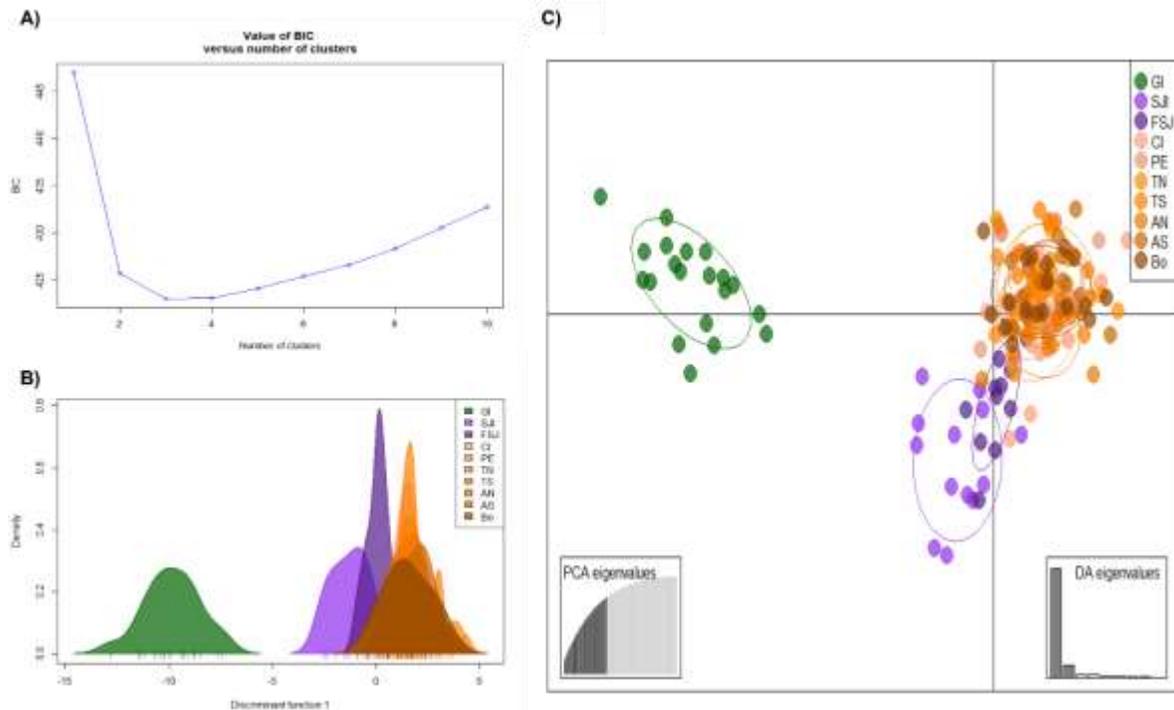
Valores de media, mediana y desviación estándar (SD) de las pruebas de precisión de asignación de los individuos a su población de origen a escala global (GI, GCN y GCS) y costera (GCN y GCS) utilizando todos los marcadores y dos proporciones de individuos para entrenar a los modelos (0.75 y 0.8).

	prop.ind	GI			GCN			GCS			
		media	mediana	SD	media	mediana	SD	media	mediana	SD	
Global	Todos los individuos	0.75	1	1	0	0.92	1	0.93	1	1	0.03
		0.8	1	1	0	0.93	1	0.10	1	1	0.03
	Individuos de entrenamiento	0.75	1	1	0	0.82	1	0.25	1	1	0.01
		0.8	1	1	0	0.89	1	0.25	1	1	0.01
Costera	Todos los individuos	0.75				0.92	1	0.10	1	1	0.01
		0.8				0.91	1	0.11	1	1	0
	Individuos de entrenamiento	0.75				0.77	1	0.33	1	1	0.02
		0.8				0.77	1	0.33	1	1	0

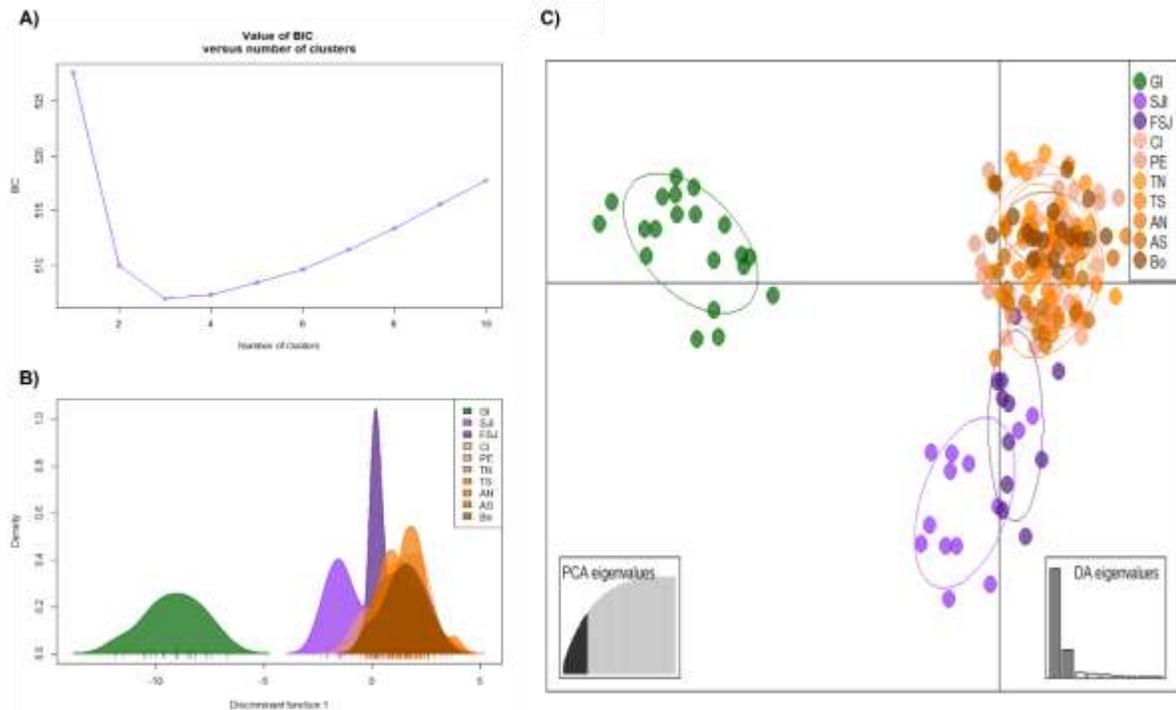
**Anexo I.** Resultados de Análisis discriminante de componentes principales (DAPC) en *H. fulgens* con los seis paneles de SNPs a escala global.



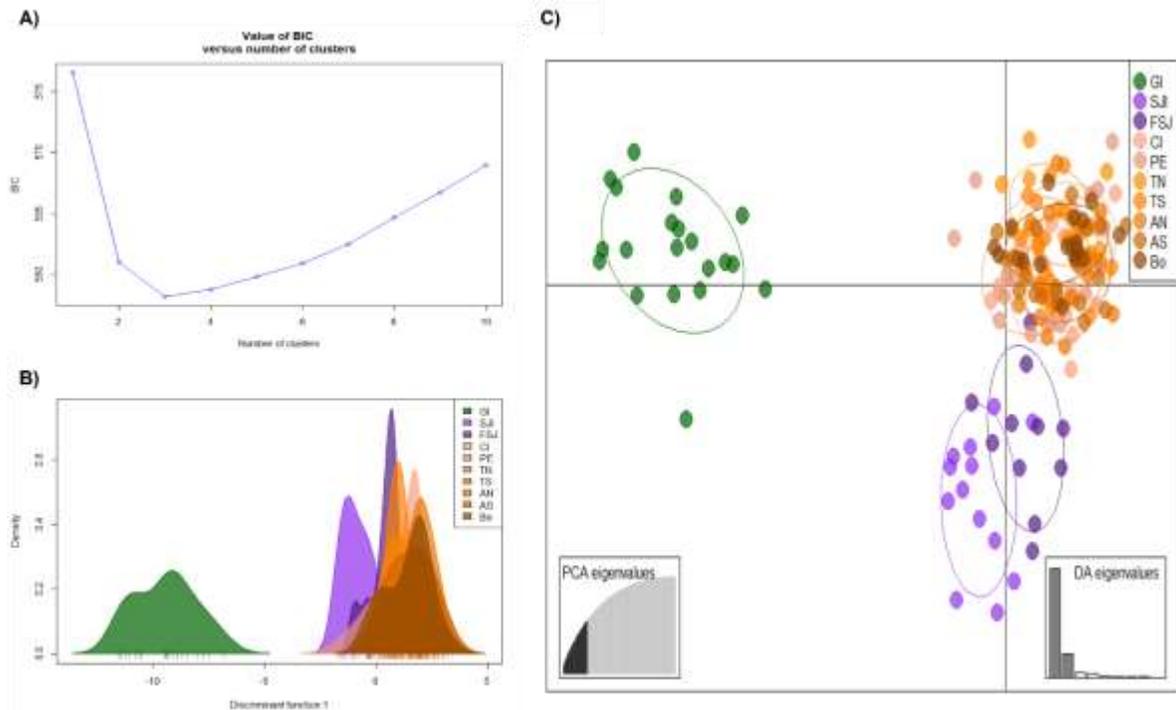
**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 50 SNPs derivados del alineamiento contra el genoma de referencia y 176 individuos a **escala global**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 35 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando tres grupos genéticos ( $K=3$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



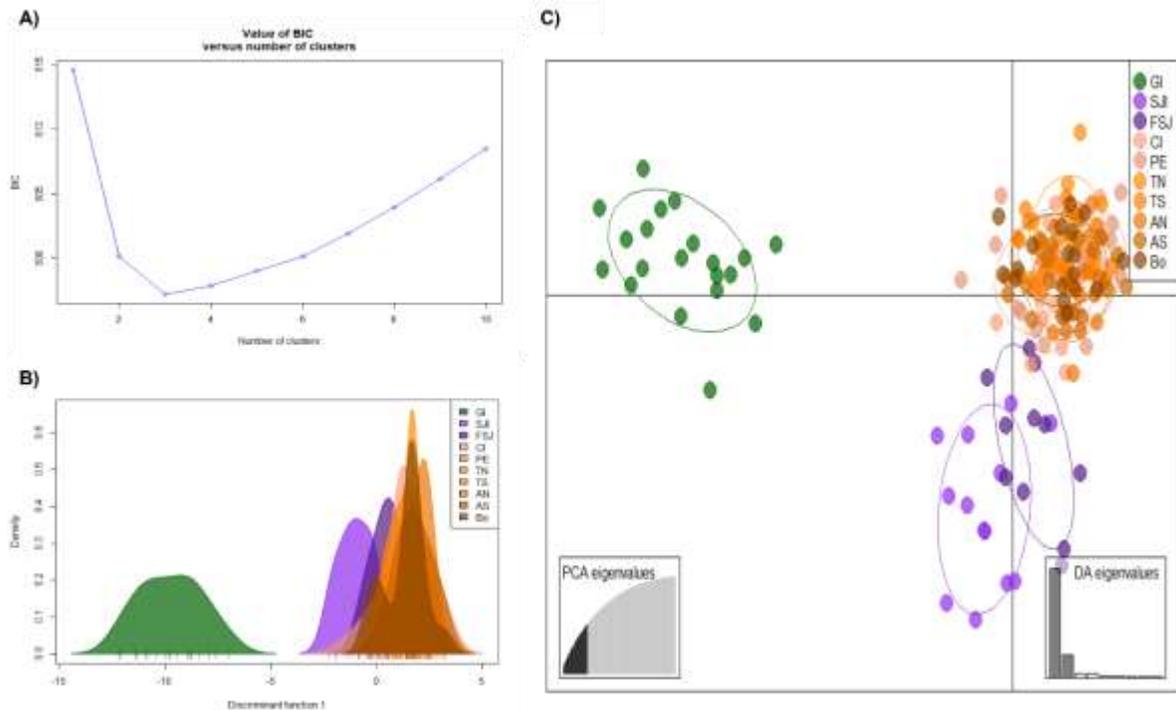
**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 100 SNPs derivados del alineamiento contra el genoma de referencia y 176 individuos a **escala global**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 38 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando tres grupos genéticos ( $K=3$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



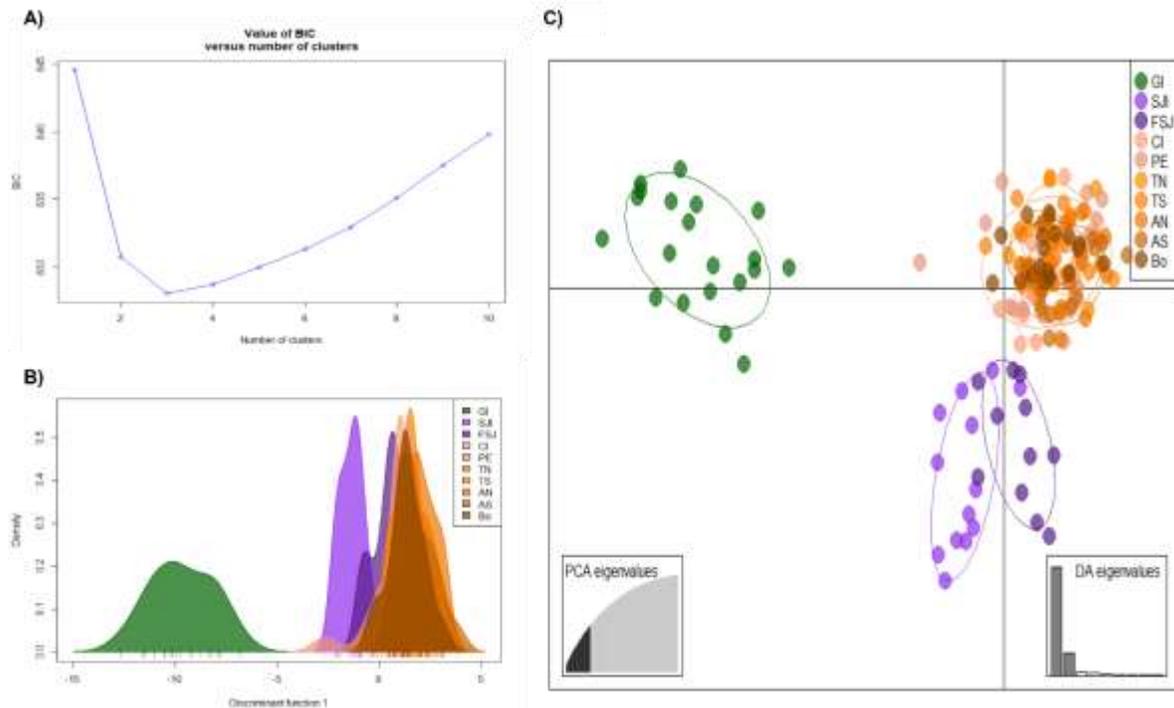
**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 200 SNPs derivados del alineamiento contra el genoma de referencia y 176 individuos a **escala global**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 38 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando tres grupos genéticos ( $K=3$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJ= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 300 SNPs derivados del alineamiento contra el genoma de referencia y 176 individuos a **escala global**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 38 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando tres grupos genéticos ( $K=3$  de acuerdo al valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.

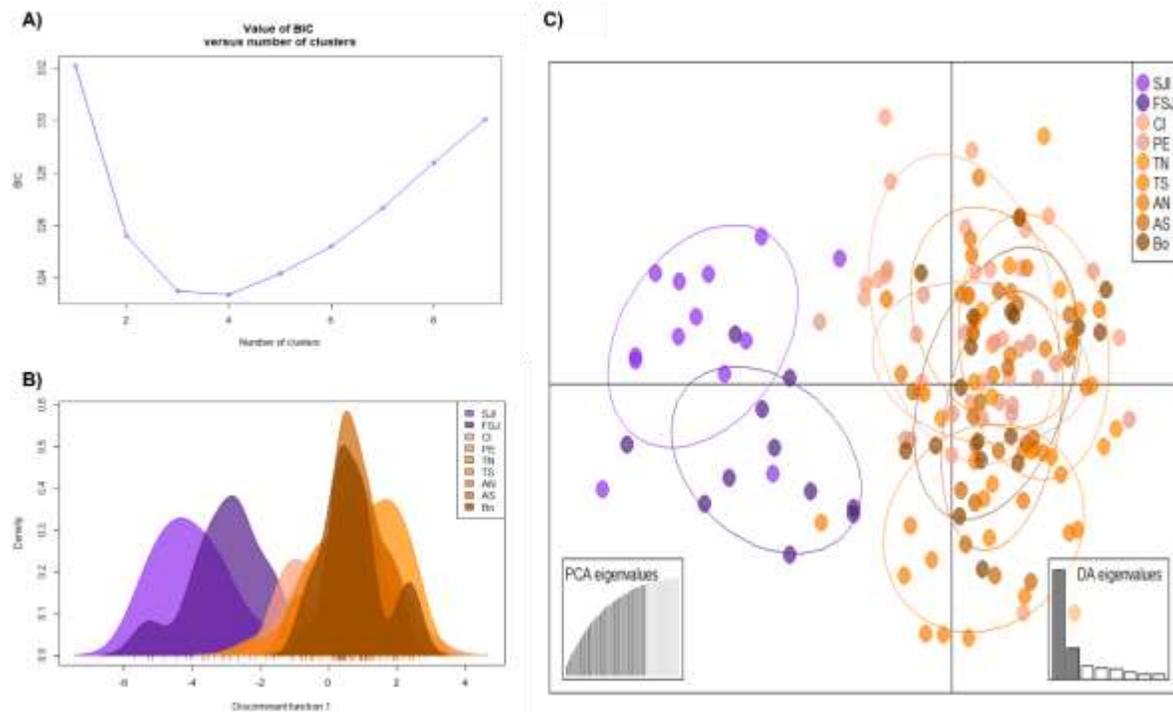


**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 400 SNPs derivados del alineamiento contra el genoma de referencia y 176 individuos a **escala global**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 38 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando tres grupos genéticos ( $K=3$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJ= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.

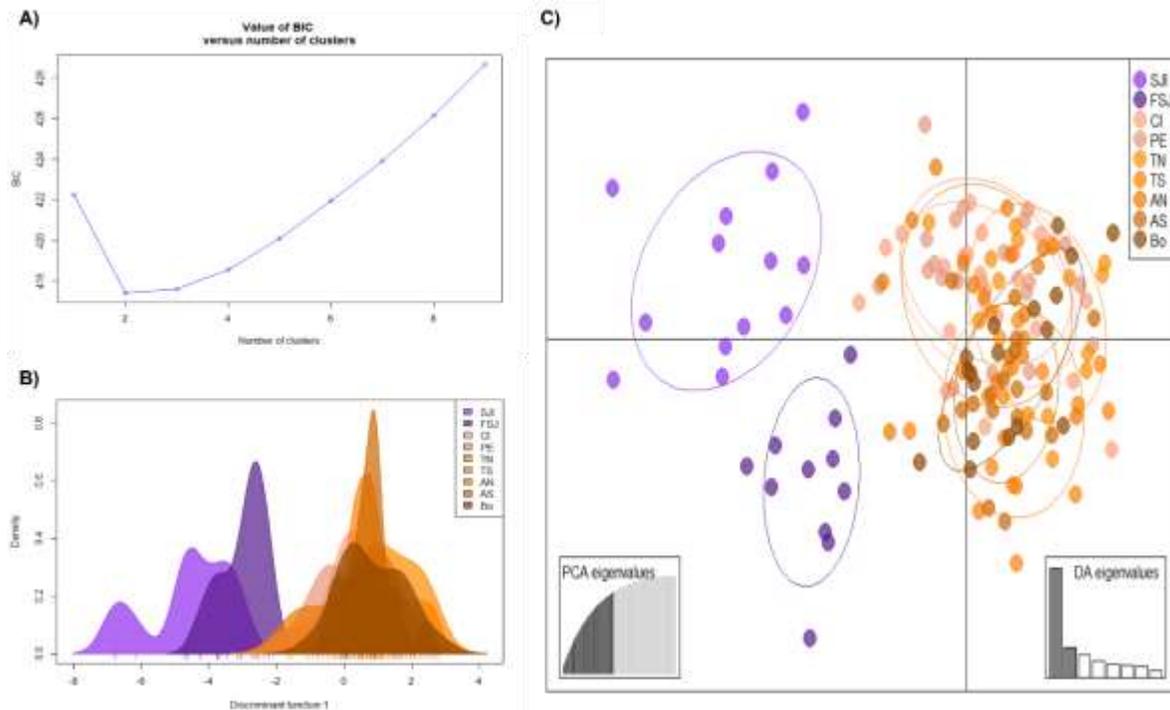


**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 500 SNPs derivados del alineamiento contra el genoma de referencia y 176 individuos a **escala global**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 38 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando tres grupos genéticos ( $K=3$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.

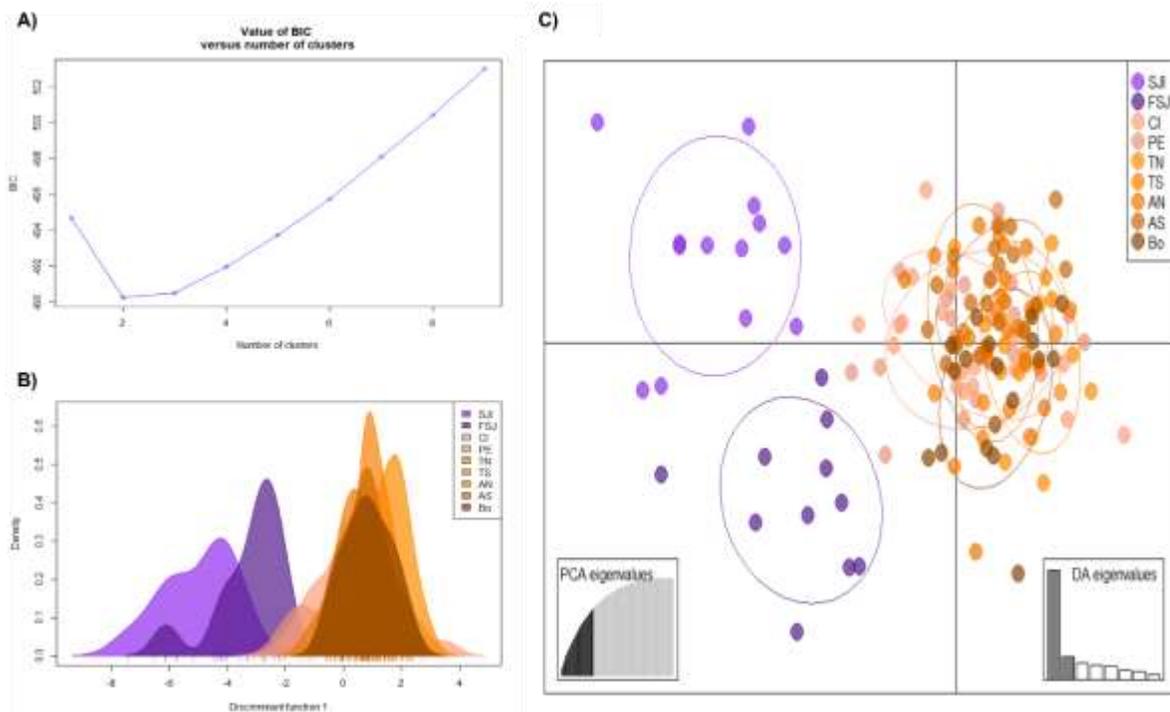
**Anexo J.** Resultados de Análisis discriminante de componentes principales (DAPC) en *H. fulgens* con los seis paneles de SNPs a escala costera.



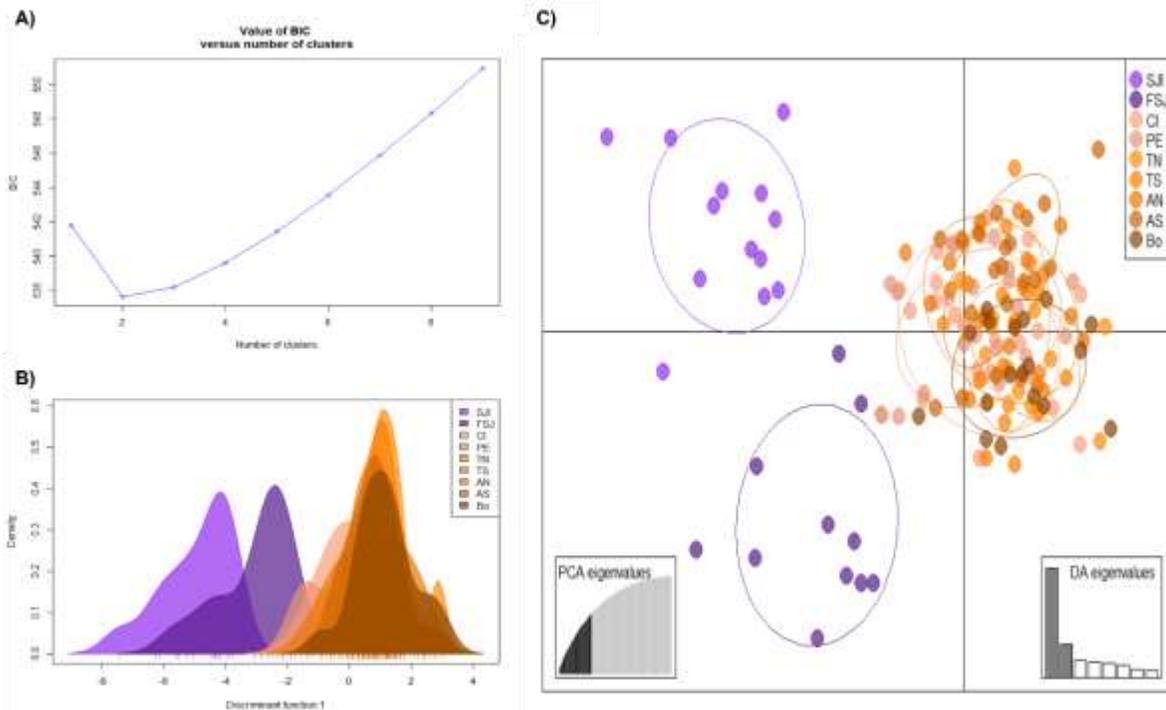
**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 50 SNPs derivados del alineamiento contra el genoma de referencia y 156 individuos a **escala costera**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 35 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis. No fue posible determinar el mejor valor de  $K$  con el valor más bajo de BIC para este set de datos. Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



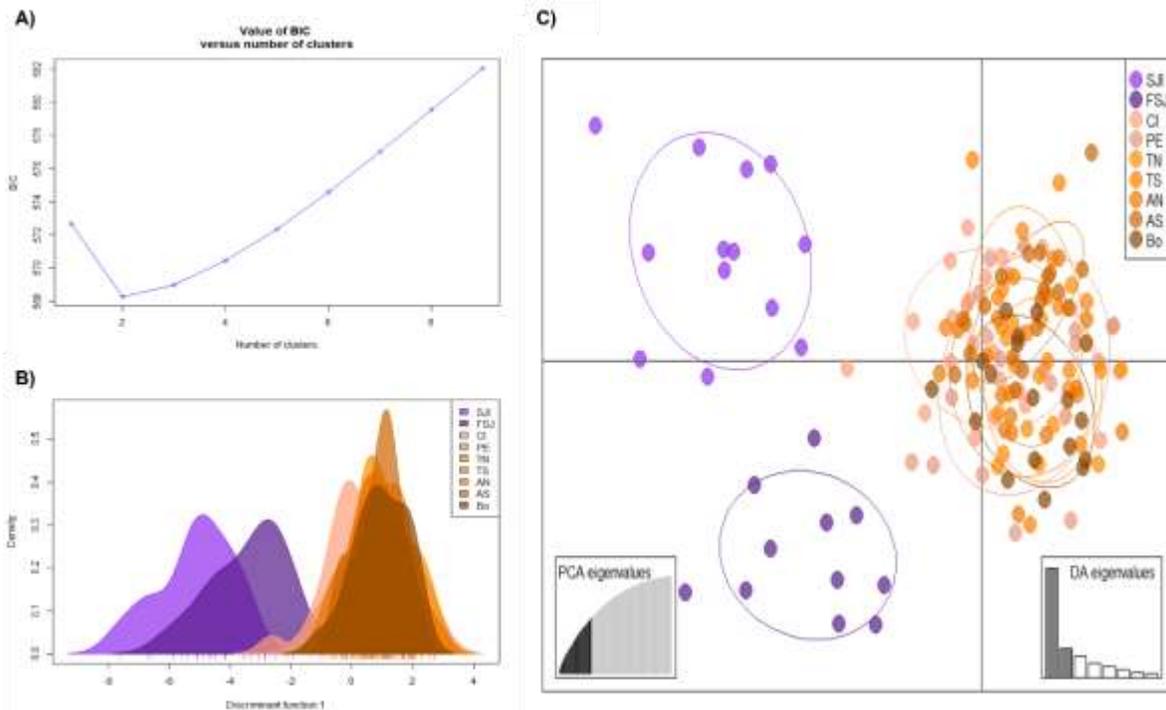
**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 100 SNPs derivados del alineamiento contra el genoma de referencia y 156 individuos a **escala costera**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 45 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando dos grupos genéticos ( $K=2$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



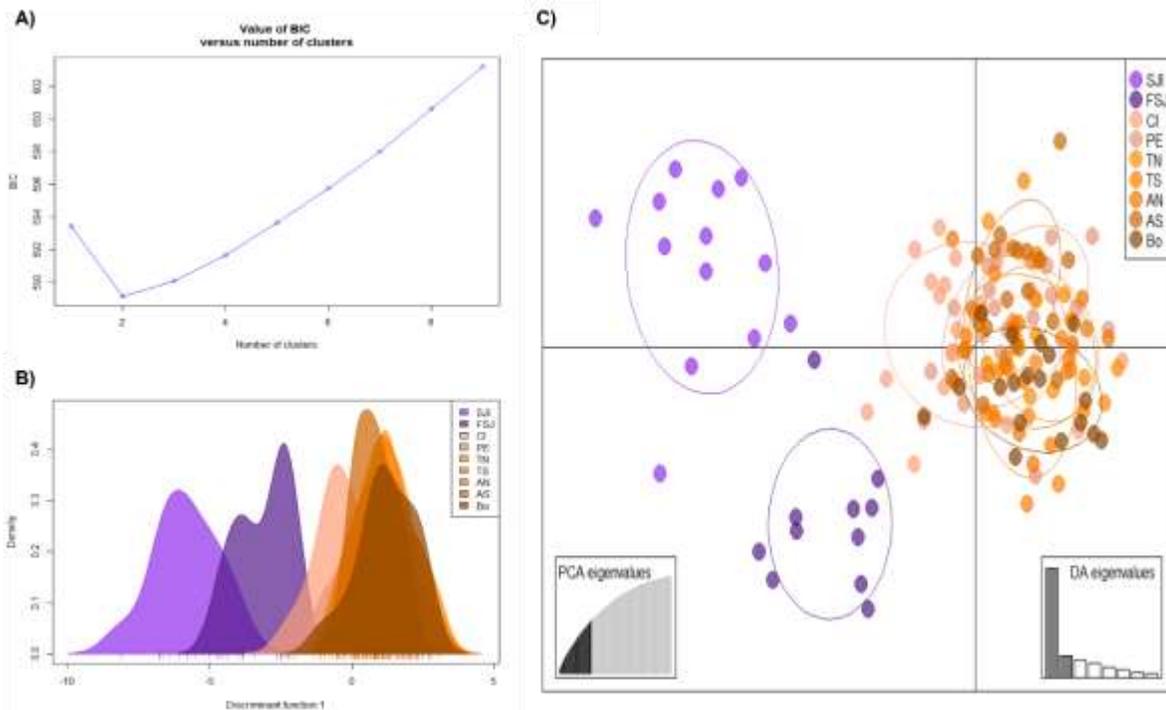
**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 200 SNPs derivados del alineamiento contra el genoma de referencia y 156 individuos a **escala costera**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 45 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando dos grupos genéticos ( $K=2$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 300 SNPs derivados del alineamiento contra el genoma de referencia y 156 individuos a **escala costera**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 45 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando dos grupos genéticos ( $K=2$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de **400 SNPs** derivados del alineamiento contra el genoma de referencia y 156 individuos a **escala costera**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 45 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando dos grupos genéticos ( $K=2$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.



**Análisis discriminante de componentes principales (DAPC) en *H. fulgens*.** Resultados utilizando el panel de 500 SNPs derivados del alineamiento contra el genoma de referencia y 156 individuos a **escala costera**. **A)** Valores del Criterio de Información Bayesiana (BIC) contra el número de grupos ( $K$ ). **B)** Gráficas de la distribución de las poblaciones basadas en la primera función discriminante. **C)** Gráficas de dispersión de DAPC, cada punto representa a un individuo. Se retuvieron 45 componentes principales y dos funciones discriminantes (dimensiones) durante los análisis identificando dos grupos genéticos ( $K=2$  de acuerdo con el valor más bajo de BIC). Las leyendas están ordenadas de norte a sur: GI= Isla Guadalupe, SJI= Isla San Jerónimo, FSJ= Faro San José, CI= Isla Cedros, PE= Punta Eugenia, TN= Tortugas Norte, TS= Tortugas Sur, AN= Asunción Norte, AS= Asunción Sur y Bo= Bocana.