



CENTRO DE INVESTIGACIONES BIOLÓGICAS  
DEL NOROESTE, S.C.

---

---

Programa de Estudios de Posgrado

**ESTRUCTURA GENÉTICA MEDIANTE SNPS EN  
BANCOS DE ALMEJA GENEROSA (*Panopea globosa*;  
DALL 1898) DEL ALTO GOLFO DE CALIFORNIA**

**TESIS**

Que para obtener el grado de

**Maestro en Ciencias**

Uso, Manejo y Preservación de los Recursos Naturales

(Orientación en Biología Marina)

P r e s e n t a

**Karina Alejandra Vázquez Rojas**

La Paz, Baja California Sur, agosto 2018

## ACTA DE LIBERACIÓN DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 12 horas del día 13 del Mes de Agosto del 2018, 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 genética mediante SNPs en bancos de almeja generosa (*Panopea globosa*; Dall 1898) del Alto Golfo de California"**

Presentada por el alumno:

**Karina Alejandra Vázquez Rojas**

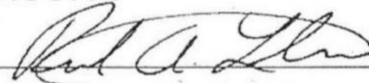
Aspirante al Grado de MAESTRO EN CIENCIAS EN EL USO, MANEJO Y PRESERVACIÓN DE LOS RECURSOS NATURALES CON ORIENTACIÓN EN **Biología Marina**

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:



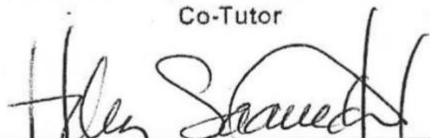
Dr. Pedro Cruz Hernández  
Director de Tesis



Dr. Raúl Antonio Llera Herrera  
Co-Tutor



Dr. Fidel Alejandro Sánchez Flores  
Co-Tutor



Dra. Norma Yolanda Hernández Saavedra,  
Directora de Estudios de Posgrado y  
Formación de Recursos Humanos

## **COMITÉ TUTORIAL**

### **Director de tesis**

Dr. Pedro Cruz Hernández

Centro de Investigaciones Biológicas del Noroeste

### **Co-tutor de tesis**

Dr. Raúl Antonio Llera Herrera

Centro de Investigación en Alimentación y Desarrollo

### **Co-tutor de tesis**

Dr. Fidel Alejandro Sánchez Flores

Instituto de Biotecnología UNAM

## **COMITÉ REVISOR DE TESIS**

Dr. Pedro Cruz Hernández

Dr. Raúl Antonio Llera Herrera

Dr. Fidel Alejandro Sánchez Flores

## **JURADO DE EXAMEN DE GRADO**

Dr. Pedro Cruz Hernández

Dr. Raúl Antonio Llera Herrera

Dr. Fidel Alejandro Sánchez Flores

**SUPLENTE:** Dr. Fausto Valenzuela Quiñonez

## Resumen

La almeja generosa (*Panopea globosa*) es un bivalvo sésil de amplia distribución en el Golfo de California y costa occidental de la península de B.C.S., la cual sostiene una pesquería creciente de gran importancia. Debido a ello, ha sido el objeto de estudio de múltiples trabajos de investigación, entre los que destaca la variabilidad genética con marcadores neutros sencillos (p.ej. microsatélites). Sin embargo, no se ha realizado un estudio a escala microgeográfica que permita identificar eventos en las poblaciones con marcadores genéticos capaces de diferenciar entre bancos, y mejorar el manejo de este recurso a nivel local. Esto se puede lograr con marcadores moleculares que tengan una elevada frecuencia en el genoma pues son los que tienen mayor potencial para evaluar la variabilidad genética. En el presente trabajo se realizó un acercamiento a la estructura genética dentro de las poblaciones del Alto Golfo de California (San Felipe, B.C., y Puerto Peñasco, Sonora), evaluando las frecuencias alélicas en bancos de *P. globosa*, mediante el uso de polimorfismos de una sola base (SNP), provenientes de la técnica de secuenciación de ADN asociada a dos sitios de restricción (ddRAD-seq). En total se obtuvieron 14,491 SNPs distribuidos en 8,326 *loci*, en los que se encontró una menor diferenciación entre bancos de la misma localidad (San Felipe). También se realizó una simulación de transporte de partículas con el modelo HAMSOM para evaluar la conectividad entre las dos poblaciones. Además, se detectaron 199 SNPs *outlier* con potencial para futuros análisis de asignación de origen (bancos). Con este trabajo se puso en evidencia que los bancos de almeja generosa en las localidades del Alto Golfo de California presentaron diferencias en las frecuencias alélicas en SNPs característicos entre los bancos de una misma localidad, y se observó una conectividad asimétrica entre las localidades, con una mayor retención de larvas en la localidad de San Felipe, y en menor grado un flujo de ésta hacia Puerto Peñasco.

**Palabras clave:** *Panopea globosa*, dd-RADseq, genómica poblacional, SNP outlier, modelo de conectividad HAMSOM.




Vo.Bo

Dr. Pedro Cruz Hernández  
Director de tesis

## Summary

The Corte Geoduck (*Panopea globosa*) is a sessile bivalve widely distributed in the Gulf of California and west coast of the peninsula of B.C.S., which supports a growing fishery of great importance. Due to this, it has been the object of study of multiple research works, among them the assessment of genetic variability with simple neutral markers (e.g. microsatellites). However, a microgeographic study has not been carried out to identify events within the populations, with genetic markers capable of differentiating between stocks, and to improve the management of this resource at the local level. This can be achieved with molecular markers that have a high frequency along the genome, these DNA markers have a great potential to evaluate genetic variability. In the present study, we make an approach to the genetic structure within the populations of the Upper Gulf of California (San Felipe, BC, and Puerto Peñasco, Sonora), evaluating the allelic frequencies in *P. globosa* stocks, through the use of Single Nucleotide Polymorphisms (SNP), from the technique double digest Restriction site Associated DNA sequencing (ddRAD-seq). A total of 14,491 SNPs distributed in 8,326 *loci* were identified, in which a lower differentiation was found between banks of the same locality (San Felipe). A particle transport simulation was performed with the HAMSOM model to evaluate the connectivity between the two populations. In addition, 199 *outlier* SNPs with potential for future analysis of origin assignment (stocks) were detected. With this work it was proved that the generous clam banks in the Upper Gulf of California localities presented differences in the allele frequencies in characteristic SNPs between the banks of the same locality, and an asymmetric connectivity between the localities was observed, with a greater retention of larvae in San Felipe, and flow from San Felipe to Puerto Peñasco in lesser degree.

**Keywords:** *Panopea globosa*, *dd-RADseq*, *population genomics*, *outlier SNP*, *HAMSOM model of connectivity*.



Vo.Bo

Dr. Pedro Cruz Hernández

Director de tesis

## **Dedicatoria**

A las tardes que compartí con mi padre, tomando café, hablando de bucles e infinitos.

A los momentos que descansé en los brazos de mi madre, soñando deidades.

Al mes de abril, sus canciones y a mi hermana.

Al instante que sostengo su mano.

Al tiempo.

## Agradecimientos

Agradezco al CIBNOR porque ha sido mi casa desde antes de iniciar la maestría, y sin saberlo en ese entonces, me fue orientando hacia esta parte de la ciencia que disfruto tanto.

Al Posgrado de esta misma institución, que permitió mi desarrollo integral. Sin duda ésta ha sido mi mejor etapa, en cuanto a formación académica, hasta el momento.

Agradezco a CONACYT por la beca otorgada (No. Becario 590763).

Al proyecto 097-C de CONACYT “Genómica poblacional y polimorfismos de una sola base (SNPs) con potencial valor adaptativo en almeja generosa (*Panopea globosa*)”.

La modelación de dispersión larvaria con el modelo oceanográfico tridimensional HAMSOM fue financiada por el proyecto institucional CONACyT Fronteras de la Ciencia 292/2016 "ESTUDIO INTEGRATIVO DE LA BIODIVERSIDAD Y LA CONSERVACIÓN DEL GOLFO DE CALIFORNIA, BAJO UN ENFOQUE DE PAISAJE GENÉTICO MARINO Y CONECTIVIDAD".

Agradezco de una manera muy especial a mis co-tutores, el Dr. Raúl Llera y el Dr. Alejandro Sánchez, por trabajar en equipo conmigo, aconsejarme y asesorarme en los puntos más críticos, así como por las oportunidades que me han brindado a lo largo de esta etapa.

Al Dr. Adrián Munguía por asesorar este proyecto, pero sobre todo, por el ejemplo que ha sido en mi vida académica. Ha sido un honor para mí por fin trabajar con él.

Sobre todo, doy gracias al Dr. Pedro Cruz que me dio la oportunidad de trabajar en este proyecto y lo que ha resultado de él: las pláticas, el conocimiento, y la amistad.

Al Dr. Fausto Valenzuela, y al equipo RADseq, por encaminarnos en el proceso de aprendizaje.

A mi compañero de trabajo, el M.C. José Víctor López, y a la técnico del Laboratorio de Genómica y Bioinformática, M.C. Gabriela Mendoza, agradezco el apoyo, el aprendizaje y la compañía que me han brindado.

A mis compañeros de maestría que se convirtieron en amistades y que han sido una parte importante de este proceso. En especial a Nuria, Caroli, Leila y Yaredh, por formar esa sociedad de *ladies* tan singular.

A mis amigas de toda la vida, Ana Karina y Nathalie, quienes me sostienen y me toman de la mano en las buenas y en las malas. Agradezco que sean parte de mi vida.

A Chuy y a Rigel, por las largas pláticas y la amistad que me han ofrecido.

A Clari, Memo, Juan, Esme, Dany y Alex, pues sin su presencia, estos años habrían sido libres de carcajadas.

A mi hermanita Abril, que es y siempre seguirá siendo mi principal motivación para crecer como persona, que sin darse cuenta me enseña sobre la vida y llena mi corazón.

Agradezco infinitamente a la vida por mis padres, mi principal ejemplo y guía, mis mejores amigos. Agradezco a mi mamá por sus palabras tan certeras, su amor tan grande, y su fuerza inquebrantable, pues con esto nos levanta y nos impulsa a ser mejores. Agradezco a mi papá, por su corazón de poeta, su mente sabia y su espíritu de aventura, es quien me motiva a soñar en grande y quien me da las herramientas para alcanzar mis metas.

Por último, agradezco poder compartir esta etapa con mi esposo, nada ha sido más grato que caminar lado a lado, crecer y construir un futuro juntos.



## Contenido

<b>Resumen</b> .....	i
<b>Summary</b> .....	ii
<b>Dedicatoria</b> .....	iii
<b>Agradecimientos</b> .....	iv
<b>Contenido</b> .....	vi
<b>Lista de figuras</b> .....	viii
<b>Lista de tablas</b> .....	ix
<b>1. INTRODUCCIÓN</b> .....	1
<b>2. ANTECEDENTES</b> .....	3
2.1 Poblaciones de <i>Panopea globosa</i> en el Noroeste de México como modelo biológico. 3	
2.1.1 Estudios genéticos poblacionales en <i>Panopea spp.</i> .....	4
2.2 Genómica poblacional.....	5
2.2.1 Polimorfismos de una sola base (SNPs): aplicación en estudios de variabilidad genética.....	6
2.2.2 Técnica de Secuenciación Asociada a Sitios De Restricción Con Doble Digestión (dd-RADseq) .....	7
2.2.3 Genómica poblacional con RADseq en otras especies.....	9
2.3 Conectividad genética en el Golfo de California .....	10
2.3.1 Modelo HAMSOM.....	12
2.3.2 Estudios de conectividad con el modelo HAMSOM en el Golfo de California ..	13
<b>3. JUSTIFICACIÓN</b> .....	16
<b>4. HIPOTESIS</b> .....	18
<b>5. OBJETIVOS</b> .....	19
5.1 Objetivo general .....	19
5.2 Objetivos particulares.....	19
<b>6. MATERIAL Y MÉTODOS</b> .....	20
6.1 Colecta de organismos .....	20
6.2 Preparación de librerías genómicas.....	21
6.2.1 Extracción de ADN .....	21
6.2.2 Selección de enzimas de restricción .....	22
6.2.3 Elaboración de librerías con el método ddRAD-Seq .....	23
6.3 Análisis bioinformático.....	24
6.3.1 Análisis exploratorio .....	24
6.3.2 Obtención de SNPs.....	26
6.3.2.1 Preparación y limpieza de datos .....	26
6.3.2.2 Generación del catálogo de secuencias.....	26
6.3.3 Estimación de parámetros genéticos .....	27
6.3.3.1 Cálculo de frecuencias alélicas .....	27
6.3.3.2 Análisis de diferenciación genética (FST).....	27
6.3.4 Selección de <i>loci</i> para análisis futuros de asignación de origen (bancos) .....	30
6.3.4.1 Selección de <i>loci outlier</i> .....	30
6.3.4.2 Selección de SNPs diseñados para futuros análisis de asignación .....	30

6.3.4.3	Análisis de estructura genética con <i>loci outlier</i> .....	30
6.3.5	Simulación de movimiento de partículas con el modelo HAMSOM.....	31
7.	<b>RESULTADOS</b> .....	32
7.1	Colecta de organismos .....	32
7.2	Librerías genómicas RADseq .....	32
7.2.1	Extracción de ADN y elaboración de librerías .....	32
7.2.2	Secuenciación de las librerías .....	33
7.2.3	Obtención de SNPs.....	35
7.3	Estructura poblacional.....	37
7.3.1	Frecuencias alélicas .....	37
7.3.2	Diferenciación genética (FST) .....	40
7.3.3	SNPs con potencial para análisis de asignación .....	41
7.3.3.1	Selección de SNPs con potencial para análisis de asignación .....	41
7.3.3.2	Análisis de estructura poblacional con <i>loci outlier</i> .....	43
7.4	Conectividad entre bancos .....	46
7.4.1	Modelo HAMSOM.....	46
8.	<b>DISCUSIÓN</b> .....	48
8.1	Diseño experimental.....	48
8.1.1	Colecta de organismos.....	48
8.1.2	Purificación de ADN y elaboración de librerías.....	49
8.1.3	Obtención de SNPs.....	50
8.2	Estructura genética .....	50
8.3	SNPs con potencial para análisis de asignación por grupo de origen .....	52
8.3.1	Estructura entre bancos a partir de los <i>loci outlier</i> .....	53
8.4	Conectividad genética .....	55
9.	<b>CONCLUSIONES</b> .....	56
10.	<b>LITERATURA CITADA</b> .....	58
11.	<b>ANEXOS</b> .....	65
Anexo A.	Concentración en las muestras de ADN genómico. ....	65
Anexo B.	Selección de enzimas de restricción. ....	68

## Lista de figuras

<b>Figura 1.</b> Mapa de sitios de colecta. ....	20
<b>Figura 2.</b> Diagrama de flujo del análisis de datos. ....	29
<b>Figura 3.</b> Frecuencia de tamaño de los fragmentos <i>sinle-end</i> y <i>pair-end</i> unidos. ....	34
<b>Figura 4.</b> Frecuencia de SNPs por posición en la lectura. ....	37
<b>Figura 5.</b> Diferenciación en las frecuencias alélicas entre los tres bancos. ....	39
<b>Figura 6.</b> Frecuencia de los valores de FST en todos los <i>loci</i> . ....	41
<b>Figura 7.</b> Histograma de frecuencia de las diferencias en las frecuencias alélicas en todos los SNPs e identificación de <i>outliers</i> . ....	42
<b>Figura 8.</b> Frecuencia de los valores de FST pareados estimados para los <i>loci outliers</i> . ....	43
<b>Figura 9.</b> Valores de FST de los <i>loci outliers</i> y clusterización por similitud. ....	45
<b>Figura 10.</b> Probabilidades de transporte de partículas entre poblaciones de San Felipe y Puerto Peñasco en la temporada de invierno y con duración larval de dos semanas. ....	47
<b>Figura 11.</b> Boxplot de distribución de edades de una submuestra de 15 organismos por banco. ....	54

## Lista de tablas

<b>Tabla I.</b> Parámetros utilizados para la detección de SNPs.....	25
<b>Tabla II.</b> Sitios de colecta en cada banco, por localidad.....	32
<b>Tabla III.</b> Concentraciones y tamaños de fragmentos de la librería genómica .....	33
<b>Tabla IV.</b> Filtrado de datos. ....	36
<b>Tabla V.</b> Frecuencias alélicas .....	38
<b>Tabla VI.</b> Valores de FST por <i>locus</i> y por SNP. ....	40

## 1. INTRODUCCIÓN

La almeja generosa (*Panopea globosa*) es un bivalvo sésil de amplia distribución en el Golfo de California y costa occidental de la península de BCS. Por su reciente explotación y alto valor comercial se ha destinado un gran esfuerzo al estudio de la biología de la especie. Entre ellos se ha estudiado la variabilidad genética con marcadores neutros sencillos, demostrándose que existen diferencias genéticas en base al análisis de regiones mitocondriales (Suárez-Moo *et al.*, 2013), microsatélites (Munguia-Vega *et al.*, 2015), así como diferencias morfométricas en la concha entre poblaciones (Leyva-Valencia *et al.*, 2012), que bien podrían deberse a la respuesta adaptativa a diferentes condiciones ambientales o a cambios acumulados a lo largo de la evolución de la especie por otras fuerzas evolutivas. A su vez, el Golfo de California, por ser uno de los ambientes marinos más diversos y productivos, también ha sido un objeto de estudio recurrente, sobre todo para caracterizar la hidrodinámica y su relación con la distribución de las especies en este sistema, estableciendo patrones claros de dispersión dependientes de las características biológicas de cada especie (Marinone *et al.*, 2008; Peguero-Icaza *et al.*, 2011). Por tanto, es imperante evaluar la estructura genética y conectividad de *P. globosa* desde una perspectiva de genómica poblacional, analizando polimorfismos tanto en regiones codificantes o funcionales del genoma, como en sitios neutros.

La dispersión de *P. globosa* es limitada, pues sólo ocurre cuando se encuentra en su fase de larva, por lo que su movimiento es mayormente influenciado por las corrientes marinas. Además, se encuentra en ambientes con distintas condiciones, esto indica que presenta posibles adaptaciones locales. Por ello, es un modelo ideal para evaluar la conectividad mediante dispersión larval de *P. globosa*, así como para caracterizar regiones genómicas que han sido moldeadas por la presión de selección de los diferentes ambientes, y las regiones neutrales que responden a otras fuerzas evolutivas, como la deriva génica, que proporcionan información sobre parámetros poblacionales, por ejemplo: tamaños efectivos poblacionales, migración, estructura genética, entre otros. El uso de las novedosas metodologías de secuenciado de siguiente generación (NGS por sus siglas en inglés) permiten obtener información representativa de todo el genoma o, en su caso, de genomas completos.

Partiendo de una muestra de cada población y mediante análisis bioinformático, es posible detectar los polimorfismos de una sola base (*Single Nucleotide Polimorfism*, o SNPs) que puedan ocurrir entre secuencias homólogas; esto permite definir genotipos para cada individuo, y así evaluar la variabilidad y diferenciación genética a distintos niveles. Es decir, en una escala mayor se puede tener el panorama de la estructura genética poblacional a lo largo de toda la distribución de la especie, y a una escala menor, detectar SNPs capaces de diferenciar entre agrupaciones de individuos, por ejemplo bancos, dentro de una misma población (Bradbury *et al.*, 2013).

Los resultados del trabajo, aportarán el conocimiento básico bajo diferentes perspectivas: 1) Definir marcadores moleculares que permitan a futuro análisis de asignación de organismos a localidades específicas, con uso potencial para trazabilidad. 2) Patrones de conectividad genética que sirvan para gestionar la pesquería de la especie. 3) Información relevante sobre la estructura genética meta-poblacional que contribuya a un mejor manejo y conservación del recurso.

## 2. ANTECEDENTES

### 2.1 Poblaciones de *Panopea globosa* en el Noroeste de México como modelo biológico.

Se le llama almeja generosa al complejo de especies pertenecientes al género *Panopea spp.* Éstas se distribuyen en esteros y lagunas, y en México ha sostenido una pesquería por más de una década (DOF, 2012). En México se distribuyen dos especies: *Panopea generosa*, con distribución a lo largo de las costas del Pacífico, desde Alaska hasta Ensenada, B. C. (Goodwin y Pease, 1989; Orensanz *et al.*, 2000; Sloan y Robinson, 1984), y *P. globosa* la cual es exclusiva de México y se distribuye en la parte central y norteña del Golfo de California (Aragón-Noriega *et al.*, 2007; Arámbula-Pujol *et al.*, 2008) hasta la costa occidental de Baja California Sur (González-Peláez *et al.*, 2013; Leyva-Valencia *et al.*, 2012; Suárez-Moo *et al.*, 2013).

Las almejas del género *Panopea spp* son consideradas unas de las especies excavadoras más grandes en el mundo, con una longitud máxima de concha de 212 mm y un peso vivo, incluyendo concha, de 3.25 kg (Goodwin y Pease, 1989). Son organismos longevos, con una edad media de 20 a 60 años y una edad individual reportada de hasta 146 años (Thórarinsdótti, 1999). Presentan sexos separados, con un periodo de madurez sexual entre los 4 y los 107 años de edad, desovando entre 1 y 2 millones de huevos (con un máximo de 20 millones), generalmente durante el verano en latitudes templadas (12°C); los huevos y el desarrollo larval se lleva a cabo en la columna de agua hasta el asentamiento en el fondo, 47 días después (Goodwin, 1976). Sin embargo, para *P. globosa* se ha observado que el ciclo reproductivo en una población de Guaymas en el Golfo de California fue de noviembre a enero (invierno) (Aragón-Noriega *et al.*, 2007; Arambula-Pujol *et al.*, 2008), y que la duración larval es de 13 a 14 días (Ferreira-Arrieta *et al.*, 2015). Se les encuentra en la infauna, enterradas en el fondo marino hasta un metro de profundidad. Una vez asentadas en fondos, permanecen enterradas el resto de su vida. De esta forma, la distribución de la especie está influenciada por el tipo de sustrato que le permite enterrarse (Campbell *et al.*, 1998). El número de organismos presentes en un área está en función de varios factores ambientales

como la disponibilidad de alimento, la temperatura del agua, el tipo de sustrato, la turbidez y las corrientes (Bradbury *et al.*, 2000).

La explotación de *P. globosa* en el Golfo de California se inició en el 2002 bajo el régimen de pesca de fomento y en 2004 como pesca comercial (Aragón-Noriega *et al.*, 2007). Al norte de la península de Baja California, en las localidades de Ensenada y San Felipe, existe una creciente pesquería, de *P. generosa* y *P. globosa* respectivamente, que cubre la demanda de consumo de los países asiáticos (Perea, 2008). Al sur de la península en B. C. S. solamente se tiene identificada una población de *P. globosa* en Bahía Magdalena, la cual contribuye con el 19% de la producción pesquera (DOF, 2012).

### **2.1.1 Estudios genéticos poblacionales en *Panopea* spp.**

La variabilidad genética de las poblaciones de *Panopea generosa* de Washington y la Columbia Británica ya ha sido evaluada a partir de marcadores neutros (microsatélites) (Miller *et al.*, 2006; Vadopalas *et al.*, 2004). En esos estudios se observó una mayor diferenciación genética a mayor escala geográfica, en donde las poblaciones de Estados Unidos (Washington) se diferenciaron de las canadienses (Columbia Británica), resultados relevantes también para el manejo pesquero.

En México, específicamente para *P. globosa* los primeros análisis genéticos realizados no encontraron diferenciación genética inter poblacional, debido a que solo se secuenciaron fragmentos del gen nuclear ribosomal (ADNr) en unos cuantos organismos de San Felipe y Puerto Peñasco (localidades en el Alto Golfo) (Rocha-Olivares *et al.*, 2010). Un análisis más reciente de morfometría geométrica en las conchas, demostró que se encontraban diferencias significativas entre los organismos de Bahía Magdalena y las localidades al interior del Golfo de California (Leyva-Valencia *et al.*, 2012); en donde además los organismos de Guaymas fueron más pequeños, con conchas más infladas, y un seno palial más profundo que los organismos de San Felipe y Puerto Peñasco, sugiriendo una respuesta adaptativa o fenotípica a las condiciones ambientales de cada sitio. Sin embargo, la diferenciación fenotípica



observada entre poblaciones de *P. globosa* puede deberse a plasticidad fenotípica y/o diferenciación genética, por lo que fue necesario desarrollar marcadores moleculares neutros altamente polimórficos, para determinar la estructura genética poblacional. Cruz-Hernández *et al.* (2014) identificaron 24 marcadores tipo microsatélite a partir de pirosecuenciado, analizando así la variabilidad y diferenciación genética entre: las dos localidades del Alto Golfo (San Felipe, Puerto Peñasco), una en la parte central (Guaymas) y la localidad reportada al límite de su distribución en la costa occidental de la península (Bahía Magdalena) (Mungía-Vega *et al.*, 2015). Sus resultados demostraron que las almejas de Bahía Magdalena se diferenciaron genéticamente de las del Golfo de California, particularmente con las del Alto Golfo; además de observar que el flujo genético y la dispersión larval inferida es predominantemente unidireccional siguiendo la circulación anticiclónica (en sentido de las manecillas del reloj) en la parte norte del Golfo de California durante el otoño e invierno, que es cuando las almejas desovan. Por otro lado, San Felipe, que está localizada al extremo superior del flujo oceanográfico, presentó el tamaño efectivo poblacional más grande, mostrando evidencias de retención larval, y pudiendo actuar como la principal fuente de larvas hacia las poblaciones al sur que muestran un gradiente de menor variabilidad genética siguiendo el flujo prevaleciente de las corrientes.

Con la caracterización de la estructura genética de la especie con los marcadores neutros (microsatélites), se conoce ya una parte de la historia demográfica en tiempos ecológicos, su tamaño efectivo poblacional y magnitud del flujo genético entre poblaciones. Sin embargo, hasta ahora, no se ha evaluado la dinámica meta-poblacional, es decir, los índices de retención, así como los de exportación de larvas, en función de la conectividad oceanográfica.

## **2.2 Genómica poblacional.**

Los ecosistemas marinos contienen una gran diversidad de especies que pueden ser aprovechadas como recurso para el humano. De ello deriva la importancia de conocer el estado en el que se encuentran en su medio natural para diseñar un manejo óptimo, así como la manera en que pueden ser adaptados a un sistema de cultivo. A partir del desarrollo de

nuevas tecnologías de secuenciación de ADN, se pueden obtener gigabases de información con un bajo costo y en poco tiempo, relativamente (Luikart *et al.*, 2003; Rašić *et al.*, 2014). Por ello ahora es posible trabajar con genomas completos (Davey y Blaxter, 2011); sin embargo, las plataformas utilizadas para ello no están adaptadas para el estudio de genética de poblaciones, donde lo que se busca son marcadores moleculares en una gran cantidad de individuos, sobre todo cuando se trabaja con especies “no modelo”, para las que no se encuentran genomas de referencia (Riveiro *et al.*, 2017). Aun así, se puede trabajar a nivel genómico con las nuevas técnicas de genotipificado por secuenciación (GBS por sus siglas en inglés), que es una manera de obtener una representación reducida del genoma (Da Fonseca *et al.*, 2016). Así se ha logrado tener marcadores que se encuentren a lo largo del genoma, tanto en regiones intrónicas como exónicas, por lo que ahora se tiene la posibilidad de identificar los marcadores neutrales para hacer análisis de la variabilidad genética, desde la perspectiva tradicional de la genética de poblaciones, pero de manera más robusta (Emerson *et al.*, 2010); y además, utilizar marcadores de regiones que estén sujetas a selección para cubrir la necesidad de entender la base genética de la respuesta fenotípica bajo diferentes condiciones. El nuevo objetivo de la ecología molecular es complementar el enfoque de marcadores genéticos neutros hacia la caracterización de polimorfismos genéticos selectivamente importantes (Feder y Mitchell-Olds, 2003; Reusch y Wood, 2007).

### ***2.2.1 Polimorfismos de una sola base (SNPs): aplicación en estudios de variabilidad genética***

Los polimorfismos de una sola base (SNPs; por sus siglas en inglés *Single Nucleotide Polymorphisms*), es decir aquellos que se originan por una simple mutación que cambia un nucleótido por otro, representan actualmente los marcadores moleculares con mayor potencial para evaluar la variabilidad genética, debido al desarrollo del secuenciado de siguiente generación y su elevada frecuencia en el genoma, lo que permite además caracterizarlos en regiones codificantes (De Wit *et al.*, 2012). Específicamente para moluscos existen ya algunos estudios que se han enfocado en la búsqueda de SNPs con la

finalidad de evaluar variabilidad genética en poblaciones silvestres y/o cultivadas (Li y Wang, 2017; Lopes-Lima *et al.*, 2018; Mariani y Bekkevold, 2013; Ribeiro *et al.*, 2017; Robledo *et al.*, 2017; Takeuchi, 2017; Wenne y Tsigenopoulos, 2018).

La aproximación de genómica poblacional en base al análisis de SNPs en organismos no modelos ha sido poco abordada, sobresaliendo un estudio reciente en que se compara la diversidad genómica entre cinco especies de invertebrados (incluyendo un molusco) y vertebrados, observándose la influencia del tamaño poblacional sobre la diversidad no-sinónima versus la sinónima (Gayral *et al.*, 2013). Los resultados fueron comparados con los estudios genómicos realizados en organismos modelo como humano (Sabeti *et al.*, 2007) y *Drosophila* (Mackay *et al.*, 2012); en donde la diversidad no-sinónima/sinónima ha sido menor en vertebrados con tamaños poblacionales reducidos en comparación con los tamaños poblaciones elevados en insectos. Específicamente para la comparación de poblaciones de una sola especie de moluscos, existe un estudio (Chu *et al.*, 2014) en un caracol inter-mareal del Atlántico el cual se distribuye a lo largo de un gradiente latitudinal de temperatura en el que a partir de RAD-seq se identificaron SNPs, en donde se han fijado alelos característicos en genes asociados a la tolerancia a choque térmico, siendo *loci* candidatos con elevado valor adaptativo (sujetos a selección).

### ***2.2.2 Técnica de Secuenciación Asociada a Sitios De Restricción Con Doble Digestión (dd-RADseq)***

Una reciente metodología llamada “Secuenciado de ADN Asociado a Sitios de Restricción” (RAD-seq, por sus siglas del inglés *Restriction site Associated DNA sequencing*) permite el secuenciado de fragmentos cortos de ADN, adyacentes a cada sitio de reconocimiento de una enzima de restricción particular, a lo largo del genoma de múltiples organismos para obtener así miles de nuevos marcadores polimórficos informativos útiles para mapeos genéticos y estudios poblacionales (Baird *et al.*, 2008; Willing *et al.*, 2011). El método de RADseq se desarrolló para la genotipificación simultánea de decenas de miles de marcadores genómicos mediante un protocolo de representación reducida. Estudios sobre genética de poblaciones

en especies no-modelo han recurrido frecuentemente a esta técnica principalmente por dos características: i) no requiere un genoma de referencia y ii) provee una manera eficiente y de bajo costo para evaluar marcadores a lo largo de todo el genoma en muchos individuos (DaFonseca *et al.*, 2016). Este método es flexible en el sentido de que la enzima de restricción utilizada se puede elegir en base a la frecuencia de corte que se requiera para ajustar el factor de reducción deseado, es decir, la proporción del genoma que será secuenciado. Esto permite jugar con el balance de la capacidad de secuenciar menor número de marcadores con una mayor profundidad o aumentar el número de marcadores con una enzima de restricción de corte más frecuente, sacrificando la profundidad de secuenciación.

Sin embargo, el método de Baird (2008) presentó algunas desventajas, por ejemplo, ha sido demostrado que el polimorfismo en el sitio de restricción puede causar el descarte de alelos (sobrestimando la heterocigocidad), o los duplicados de PCR pueden introducir sesgo en la profundidad de secuenciación por *locus*, así como otros artefactos analíticos. Para evitar o reducir el sesgo por las cuestiones antes mencionadas se han creado variaciones del protocolo original de RADseq, por ejemplo: ddRAD (Peterson *et al.*, 2012), ezRAD (Toonen *et al.*, 2013), 2b-RAD (Wang *et al.*, 2012), y además la lista de las mejoras al flujo de trabajo de RADseq ha ido creciendo, tomando ventaja de las tecnologías genómicas para atender preguntas de investigación que han existido desde hace mucho tiempo, así como otras de generación reciente debido a la disposición de una gran cantidad de información genómica (Ribeiro *et al.*, 2017).

Para aumentar el alcance de la aplicación de la metodología RADseq, Peterson *et al.* (2012) elaboraron un método basado en el RADseq, llamado Secuenciación Asociada a Sitios De Restricción Con Doble Digestión (dd-RADseq), eliminando cortes aleatorios y reparación de extremos de fragmentos de ADN, en vez de ello se realiza una doble digestión con dos enzimas de restricción, y de esta manera se puede reducir, en costo y cobertura, la librería genómica. Además, añaden al método una selección de tamaño de fragmentos para recuperar un número de regiones ordenables que están distribuidas aleatoriamente a lo largo del genoma. Además, para incrementar el número de muestras por línea de secuenciación, también desarrollaron un acercamiento de marcaje por combinación por pares de adaptadores

(*barcodes* con oligonucleótidos). Se ha probado que este método es de bajo costo, rápido, requiere poco material de inicio, y es adecuado para aplicaciones de alto rendimiento, manteniendo la ventaja de que no se necesita un genoma de referencia. Al combinar estas dos características, el método ddRADseq produce librerías que consisten en fragmentos cortados por ambas enzimas de restricción, una en cada extremo del fragmento. Además, entran dentro de la ventana de selección de tamaño y se pueden organizar para representar una fracción del genoma completo. Otras ventajas del proceso de selección del tamaño del fragmento son que sólo una pequeña parte de todos los fragmentos entrarán en esta ventana de selección, por lo que la probabilidad de muestrear la muestra en dos direcciones desde el mismo sitio de restricción será muy baja. Esto reduce el muestreo duplicado de una región, por lo que sólo se requiere la mitad de lecturas para alcanzar efectivamente altos niveles de confianza para cada SNP asociado a un sitio de corte de la enzima de restricción. Todas estas propiedades hacen al método ddRADseq robusto, lo cual permite buscar un número de lecturas menor, que para otros métodos es un problema común derivado de la representación desigual en este número de lecturas, a lo largo de muestras individuales en experimentos de secuenciación.

### ***2.2.3 Genómica poblacional con RADseq en otras especies***

Los avances antes mencionados en las técnicas de secuenciación han llevado a reevaluar la dispersión y conectividad en las especies marinas. Lo que sobresale de éstas actualizaciones es el incremento dramático en el número de marcadores, así como en el tipo de *loci* disponibles para estudiar en las especies marinas, donde es frecuente que la estructura se vea enmascarada por un flujo abundante de genes a falta de barreras físicas obvias.

La ventaja de la secuenciación asociada a sitios de restricción es que permite escanear todo el genoma de los organismos no modelos y de esta manera caracterizar la diversidad genética de las especies marinas, desde peces, *Gasterosteus aculeatus* (Allendorf *et al.*, 2010; Catchen *et al.*, 2013b; Guo *et al.*, 2015), hasta invertebrados *Placopecten magellanicus* (Van Wyngaarden *et al.*, 2016), *Haliotis laevigata* (Sandoval-Castillo *et al.*, 2018), *Pinctada*

*margaritifera* (Lal *et al.*, 2016), *Mytilus chilensis* (Araneda *et al.*, 2016), y *Cepaea nemoralis* (Richards *et al.*, 2013); además de ir a la búsqueda de *loci outlier* (candidatos a selección), demostrando consistentemente que se puede identificar variación genética a escalas pequeñas en esta gran variedad de taxa marinas (Antao *et al.*, 2008).

Esta ventaja de separar los *loci* neutrales y los *outliers* (candidatos a selección) se ha visto reflejada en trabajos como el de Sandoval-Castillo *et al.* (2018), con el abulón australiano (*Haliotis laevigata*) en el que, desde una perspectiva neutral, presentó un aislamiento por distancia muy débil, mejor representado por una metapoblación con muy baja diferenciación, pero al hacer el análisis para los *loci* candidatos, se encontraron cinco poblaciones divergentes, y además relacionaron esta diferenciación con variables ambientales. Un caso similar ocurre con *Mytilus chilensis* (Araneda *et al.*, 2016), que en el panel de 891 SNPs neutrales, la estructura entre localidades que encontraron fue baja, pero al hacer el análisis con 58 SNPs *outlier*, la diferenciación que obtuvieron fue mayor, indicando posibles adaptaciones locales.

### **2.3 Conectividad genética en el Golfo de California**

Es reconocido que los procesos de circulación de corrientes marinas juegan un papel importante en la dispersión del zooplancton, incluyendo los estadíos larvarios y huevos de distintas especies. La circulación a mesoescala puede minimizar la dispersión a larga distancia reteniendo o concentrando los productos de desoves. Estudios de dispersión planctónica han permitido la detección de desove potencial y de áreas de reclutamiento de especies de peces y otros organismos marinos, así como de rutas de conectividad entre estas áreas (dirección y distancia) y sus cambios en función al tiempo (Fogarty y Botsford, 2007; Jones *et al.*, 2007).

El Golfo de California tiene pesquerías de gran importancia (Arvizu-Martínez, 1987), y es rico en biodiversidad y endemismos regionales, en particular para vertebrados (Brusca, 2002). Estas características resultan de la combinación de distintos procesos físicos en el

Golfo, incluyendo fuertes corrientes de marea, grandes variaciones de temperatura estacionales, diversidad geomorfológica y tipos de hábitats. Además, en respuesta al declive de algunas pesquerías y la amenaza a algunos ecosistemas marinos, las reservas marinas han sido utilizadas como una herramienta de conservación, por lo que ha incrementado la presión por realizar un manejo apropiado que requiere de estudios multidisciplinarios.

En particular, es importante entender el grado de asociación o conexión entre diferentes áreas del Golfo con respecto al flujo de nutrientes, contaminantes y estadíos larvarios de varios organismos (Van der Molen *et al.*, 2007). La circulación y su cambio estacional juega un papel importante en la dispersión larval en cualquier océano. Existen varios estudios de conectividad enfocados en el Golfo de California, los primeros se enfocaron en estaciones o áreas específicas, mientras que otros abarcaron regiones que comprenden casi la totalidad del golfo y compararon distintas estaciones. Un ejemplo es el trabajo de Marinone *et al.* (2008) que presentó matrices de conectividad para el Alto Golfo de California y se limitó a la estación de verano, en donde encontraron que la conectividad era dominada por la circulación estacional, y que en el verano el flujo de partículas seguía una ruta ciclónica hacia el sur. Por otro lado, Peguero-Icaza *et al.* (2008), Sánchez-Velasco *et al.* (2009), y Peguero-Icaza *et al.* (2011) encontraron que la distribución de las larvas de peces estaba asociada con áreas de distintas condiciones ambientales. Por ejemplo, durante el verano, la presencia de circulación superficial ciclónica incluye un remolino en el centro y una corriente costera hacia el noroeste en el lado continental que actúa como una continuación de la entrada de corriente del sur y por lo tanto carga nutrientes y larvas al norte del Golfo. Sánchez-Velasco *et al.* (2009) y Peguero-Icaza *et al.* (2011) condujeron estudios de conectividad estacional en el norte del Golfo entre áreas definidas por el ensamblaje de larvas de peces, revelando una diferenciación en el grado de retención de partículas dentro de cada ensamblaje de larvas de peces con áreas que dependen de la estación.

La circulación superficial estacional en el Golfo es dirigida principalmente por la circulación costera geostrófica del este del pacífico tropical, y en menor grado por los vientos superficiales, con una pequeña contribución del flujo de flotabilidad (Beier, 1997; Marinone, 2003). Observaciones directas de las corrientes realizadas por Lavin *et al.* (1997), y Palacios-

Hernández *et al.* (2002), así como modelos numéricos analíticos (Beier, 1997; Marinone, 2003; Zamudio *et al.*, 2008) han establecido que todo el Golfo de California desde finales del otoño hasta iniciando la primavera, la circulación superficial es anticiclónica, es decir, en el sentido de las manecillas del reloj, y por el contrario en verano, tiene circulación ciclónica.

### **2.3.1 Modelo HAMSOM**

Uno de los métodos más prometedores para evaluar conectividad de ecosistemas marinos, son modelos computacionales que predigan la dirección e intensidad de las corrientes oceanográficas en función de tiempo y distancia. Estos modelos han logrado esclarecer los mecanismos subyacentes que dirigen los patrones de exportación e importación de larvas a lo largo de distintos sitios, así como los niveles de retención local larval dentro de los sitios (Marinone, 2012; Metaxas y Saunders, 2009; Treml *et al.*, 2012). Los modelos oceanográficos son modelos numéricos espaciales explícitos que simulan la dispersión de partículas que se consideran análogas a larvas y que son dirigidas por corrientes basadas en descripciones precisas sobre la topografía, batimetría, temperatura, salinidad, mareas, viento, afluencia de agua dulce, entre otros parámetros, y a su vez incorporando información propia de las especies de estudio, que son las temporadas de desove y duración del estado larval (Metaxas y Saunders, 2009).

Marinone (2003) adaptó el modelo numérico integrado verticalmente en capas del océano de la plataforma de Hamburgo (HAMSOM). Este modelo es descrito a detalle en los trabajos de Backhaus (1983, 1985), Crean *et al.* (1988), Stronach *et al.* (1993), Marinone *et al.* (1996), y Rady *et al.* (1998). Las ecuaciones del modelo se resuelven de manera semi-implícita con campos de temperatura y salinidad totalmente pronósticos, lo que permite el movimiento baroclínico dependiente del tiempo. Éstas se encuentran descritas en el trabajo de Marinone (2003). De ello resultó un modelo numérico tridimensional para simular la circulación estacional del Golfo de California. El modelo fue forzado en la abertura del golfo (en el Océano Pacífico) especificando la elevación superficial marina, conteniendo las mareas y los



cambios estacionales de temperatura y campos de salinidad. En la superficie marina, se utilizaron variables estacionales meteorológicas para forzar el modelo también.

Munguía-Vega *et al.* (2014) generaron una matriz de conectividad en el Golfo de California a partir de la modelación de la conectividad espacial estudiando la cabrilla sardinera *Mycteroperca rosacea*. Se integró la información de los límites físicos y políticos, incorporando líneas de costa, batimetría (desarrolladas por el Centro Nacional de Datos Geofísico) y la delimitación de áreas marinas protegidas en México, y junto con la función de unión espacial, se establecieron las unidades espaciales (polígonos) del modelo. El tamaño de estas unidades varió de 13 a 812 km<sup>2</sup>.

Para generar las matrices se realizó una simulación por computadora, donde se liberaron cuatro mil partículas a una profundidad de 5 m en cada uno de los 17 sitios de agregación de desove de *M. rosacea*. Las partículas se rastrearon durante fechas que representan las mareas vivas y mareas muertas, que según sus observaciones cubre distintos períodos de desove máximo alrededor de las regiones del Alto Golfo. Dado que los regímenes oceánicos estacionales son consistentes a través de los años en el Golfo de California (Marinone, 2003; Soria *et al.*, 2014), el año de simulación se elige arbitrariamente, y la fase del ciclo estacional simplemente cambia de año a año.

Para la modelación utilizaron el campo de velocidad de la implementación de Golfo de California del modelo Baroclínico tridimensional Hamburg Shelf Ocean Model (HAMSOM) (Backhaus, 1985) para calcular las trayectorias de las partículas. El modelo ha sido descrito en detalle para el GC (Marinone, 2003; Marinone *et al.*, 2008).

### ***2.3.2 Estudios de conectividad con el modelo HAMSOM en el Golfo de California***

Especies marinas de peces e invertebrados que son explotados comercialmente, están compuestas por metapoblaciones que están conectados principalmente por la dispersión larval (Cowen *et al.*, 2000). Hasta el momento, sólo se han realizado unos cuantos intentos para incorporar estudios multidisciplinarios para comprender de mejor manera la

conectividad marina (Munguía-Vega *et al.*, 2014). Aún es un reto encontrar especies clave que representen patrones de conectividad marina para los factores influyentes que son intrínsecos de la especie, como el lugar y época del año de temporadas de desove, y la duración del estadio larvario pelágico (DLP). En el Golfo de California ya se han observado patrones clave de conectividad que utiliza este modelo oceanográfico junto con datos genéticos empíricos, y que sirven como referencia para modelos biológicos similares.

Tal es el caso de un estudio realizado con la cabrilla sardinera, *Mycteroperca rosacea*, que es una especie que desova en la temporada de abril-junio, y tiene una DLP de 28 días (Munguía-Vega *et al.*, 2014). Evaluaron con dos marcadores de ADN mitocondrial (Cyt b y ATPasa) y lo compararon con el modelo de conectividad en el Alto Golfo de California. Los resultados respaldaron el modelo de flujo genético altamente relacionado con giro ciclónico del Alto Golfo durante el periodo de desove que transporta larvas de la península (fuente de larvas), a través del golfo, hasta las costas de Sonora (sumidero de larvas).

Peguero-Icaza *et al.* (2010), también evalúan la conectividad en larvas de peces, en donde se observan cambios estacionales en las rutas de conectividad en el Alto Golfo de California. Observaron una retención de partículas alta (>55%) para el conjunto de larvas de peces, después de 30 días de dispersión entre junio y agosto, cuando hay un remolino ciclónico y la corriente costera del noroeste en la plataforma continental estuvieron presentes. Esto sugiere que las larvas de la mayoría de las especies de peces que habitan el golfo norte se mantienen cerca de sus áreas de desove: en junio, especies demersales costeras como *Gobulus crescentalis*, *Etropus crossotus* y *Lythrypnus dalli*, y en agosto especies epipelágicas costeras y con un desove extenso como *Opisthonema libertate* y *Anchoa* spp. Por el contrario, la retención de partículas entre los ensambles de larvas de peces fue baja (< 35%) entre diciembre y febrero, cuando la circulación es anticiclónica y la corriente costera sobre la plataforma continental fluye hacia el sur; el transporte de partículas desde el golfo norte al golfo sur es alto (>50%). Por lo tanto, el reclutamiento podría ocurrir lejos de las ubicaciones de desove, es decir, en el golfo sur. Esta ruta de conectividad es congruente con la amplia distribución espacial y la adaptabilidad ambiental de las larvas de *Engraulis mordax* y *Benthoosema panamense*, que se reclutan en mar abierto. La excepción fue *Merluccius*

*productus*, que se retuvo en el área del remolino en febrero, cuando éste ya estaba bien formado.

Si bien, hay ejemplos de la conectividad en otras especies, éstos no evalúan a profundidad un modelo como el que se presenta en este trabajo. Sin embargo, dejan claro la utilidad de evaluar la conectividad larval en los corredores del golfo mediante modelos hidrológicos y corroborarlo con marcadores genéticos.

### 3. JUSTIFICACIÓN

El derrame económico que representa la almeja generosa, *Panopea globosa*, resulta muy atractivo para los productores locales generando una importante fuente de empleos para la región (DOF, 2012). La pesquería de esta especie representa ganancias mayores a 30 millones de dólares en los últimos años (Rocha-Olivares *et al.*, 2010). Este recurso es de interés para quienes desean su explotación como cooperativas, Sociedades de Producción Rural, Uniones y personas físicas y morales, principalmente, la CONAPESCA, el INAPESCA a través de su Centro Regional de Investigación Pesquera (CRIP) en La Paz, BCS, Guaymas, Son. y en Ensenada, BC, la Subdelegación Federal de Pesca de Sonora y Baja California, dependientes de las Delegaciones de SAGARPA, PROFEPA, y los Gobiernos de Baja California Sur, Baja California y Sonora. En el ACUERDO por el que se da a conocer el Plan de Manejo para la Pesquería de Almeja Generosa (*P. globosa*) en las costas de Sonora, México, mencionan que “El establecimiento de una estrategia de manejo para esta especie debe ser considerado con un enfoque completamente precautorio, la principal causa se relaciona con el grado de desconocimiento biológico y pesquero para la almeja generosa[...]", para poder explotar el recurso de manera óptima, debe haber información sobre el recurso genético con el que se cuenta, y al hacer un estudio de este grado de resolución, permitirá establecer aspectos básicos sobre su biología, y como menciona “[...]En todo caso, deben existir reglas de control que regulen el esfuerzo de pesca, así como evitar que sea una pesquería de acceso abierto[...]”. Lo anterior es evidencia de la importancia de generar conocimiento biológico/genético del recurso que se está evaluando en este trabajo.

Ya es conocido que existen diferentes poblaciones de almeja generosa en México, en base a análisis previamente realizados con marcadores neutros (microsatélites y genes mitocondriales), debido a que se encuentran distribuidas a grandes distancias y con flujo genético reducido. Sin embargo, no se cuenta con marcadores que nos puedan explicar los procesos que ocurren a una escala microgeográfica, ni capaces de evaluar regiones genómicas con diferencias genéticas asociadas a procesos de la selección, o de distinguir entre grupos (por ejemplo, metapoblaciones, bancos, o morfotipos) que presentan un alto flujo genético.

La generación de este tipo de conocimiento genómico será fundamental para conocer la estructura genética y conectividad entre sitios. Tal aproximación será relevante para comprender su adaptación a un ambiente cambiante, así como para generar un conocimiento que permita el manejo oportuno de la pesquería y futuro cultivo. Así como identificar marcadores para trazabilidad de productos (origen de los organismos), los cuales están siendo demandados para evitar actividades fraudulentas al momento de la comercialización de los organismos. Existe la necesidad de estrategias de gestión que incorporen la dinámica de metapoblaciones en la pesquería, de ello deriva la importancia de hacer un estudio de alta resolución a escala microgeográfica, en especial en la localidad de San Felipe, pues alberga la mayor población de almeja generosa. El uso de múltiples metodologías podría ayudar a aumentar el interés por incorporar información genética en la planeación del uso del recurso marino.

La comprensión de cómo la conectividad larval influye en la dinámica de la metapoblación fuente-sumidero y el reabastecimiento local de larvas podría ayudar a aumentar la efectividad de las herramientas de gestión espacial marina. Por ejemplo, durante el diseño de redes de reservas marinas o zonas de prohibición de extracción, los patrones de dispersión de larvas para las especies focales podrían informar sobre la ubicación óptima de la reserva y el espaciamiento (Dubois *et al.*, 2016; Green *et al.*, 2015; Munguia-Vega *et al.*, 2014).

#### **4. HIPOTESIS**

En las localidades del Alto Golfo de California, los bancos de almeja generosa (*P. globosa*) presentarán diferencias en las frecuencias alélicas en SNPs característicos entre los bancos de una misma y diferente localidad, no antes detectadas con marcadores nucleares como microsatélites; además de que será posible distinguir *loci outlier*, posiblemente sujetos a selección, que nos generen una mayor estructuración genética y por ende tengan la potencial utilidad para futuros análisis de asignación (trazabilidad) de origen de bancos. En lo que respecta a un análisis de simulación de transporte de partículas, se observará que la conectividad entre las localidades es asimétrica, siendo mayor de San Felipe hacia Puerto Peñasco y con retención en ambas.

## **5. OBJETIVOS**

### **5.1 Objetivo general**

Definir la estructura genética de *P. globosa* en el Alto Golfo de California, a partir del análisis de SNPs identificados *de novo*, entre bancos de San Felipe, B. C. y Puerto Peñasco, Sonora.

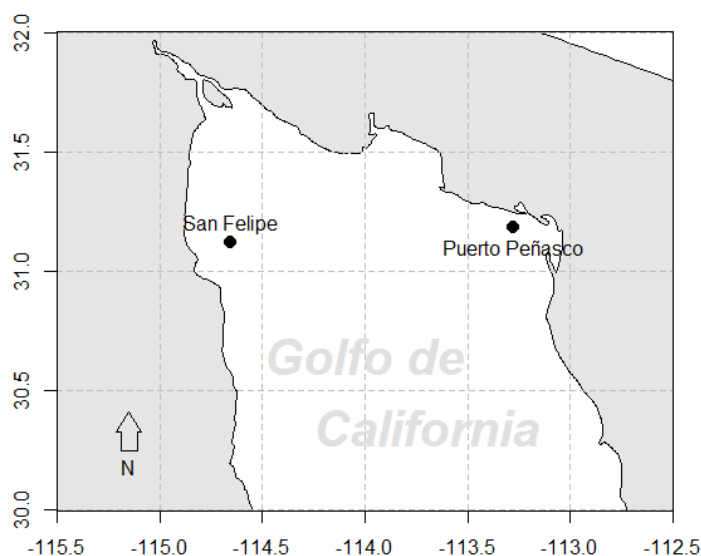
### **5.2 Objetivos particulares**

1. Identificar polimorfismos de una sola base (SNPs) al comparar secuencias de diferentes bancos de las localidades de San Felipe y Puerto Peñasco.
2. Determinar diferencias genéticas mediante análisis de frecuencias alélicas y  $F_{ST}$ , entre bancos a partir de SNPs.
3. Seleccionar SNPs *outliers* potenciales para futuros análisis de asignación de origen (bancos).
4. Evaluar la conectividad genética a partir de un análisis de simulación de transporte de partículas.

## 6. MATERIAL Y MÉTODOS

### 6.1 Colecta de organismos

Los organismos fueron colectados en las localidades de San Felipe, Baja California y en Puerto Peñasco, Sonora (Fig. 1). Se realizó una primera colecta en el mes de abril de 2015, muestreando 2 bancos de San Felipe, Baja California, separados por 2.5 km; y la segunda colecta en el mismo mes en un banco de Puerto Peñasco. A manera general la extracción de la almeja generosa se realizó por buceo semiautónomo, en donde se empleó una manguera conectada a una motobomba para arrojar agua a presión dirigida al sitio donde se encontraba enterrada la almeja (normalmente hasta un metro de profundidad del sustrato) y se extrajo al individuo completo. Se cortó un pequeño fragmento (2-3 g) de tejido disponible y se preservó en alcohol al 96% para el posterior análisis de ADN. En el momento de la colecta se tomaron las coordenadas geográficas mediante un GPS, se marcó en cada banco el punto de inmersión del buzo, y el punto de salida de éste. Esto con el fin de poder determinar la longitud del transecto muestreado y la distancia entre bancos. Además, se midió la profundidad en cada banco con una cinta métrica que en un extremo tenía colocado un plomo para ayudar a sumergirla y se tomó la medida al nivel del mar.



**Figura 1.** Mapa de sitios de colecta.



## 6.2 Preparación de librerías genómicas

### 6.2.1 Extracción de ADN

Se tomó un fragmento de la muestra de aproximadamente 50 mg, posteriormente fue enjuagada con agua destilada y secada completamente para eliminar restos de etanol. El tejido fue incubado a 56°C por 18 horas en 400 µL de Buffer de sales (0.4 M NaCl 10 mM Tris-HCl pH 8.0 y 2 mM EDTA pH 8.0), 40 µL de SDS al 20%, 20 µL de proteinasa K (20 mg/mL), y 2 µL de RNasa A (Promega). Para completar la lisis, se hizo una homogenización mecánica con pistilo. Después fue agregado 200 µL de solución de NaCl 6M a cada muestra, ésta hace una red iónica que al momento de centrifugar, filtra la “basura” llevándola al fondo, dejando el ADN en el sobrenadante. Para que fuera más efectivo se enfrió por 10 minutos a -20°C, se vortexeó y posteriormente se centrifugó por 10 minutos a 15,294 xg. Se extrajo el sobrenadante y se pasó a un nuevo tubo, al cual se le agregó un volumen igual (aproximadamente 400 µL) de cloroformo con alcohol isoamílico. Con éste se formó un gradiente de densidad, para que en caso de restos de *debri* se fuera al fondo y en el sobrenadante sólo quedara ADN. Se centrifugó por 2 minutos a 15,294 xg. Se tomó la fase superior de la muestra (aproximadamente 300 µL) y se colocó en un tubo nuevo, evitando tocar el anillo que se forma en el centro, que corresponde a remanentes de grasa y basura. Posteriormente se procedió al primer precipitado agregando 400 µL de isopropanol y enfriando a -20°C por 1 hora. Se centrifugó después por 15 minutos a 15,294 xg para formar el *pellet*. Para lavar cada *pellet* se decantó el sobrenadante con cuidado y se agregó 400 µL de etanol al 70% a cada muestra para resuspenderlo. Se volvió a centrifugar por 15 minutos a 15,294 xg y se decantó una vez el sobrenadante con mucho cuidado de no tirar el *pellet*, que se dejó secando para finalmente agregar 100 µL de agua MiliQ y disolver el ADN.

El ADN se cuantificó en NanoDrop™ 2000, y se evaluó la pureza de éste tomando en cuenta la relación 260/280 que indica presencia de proteínas, fenol y otros contaminantes que pueden ser leídos a 280 nm; así como la relación 260/230, que indica que hay contaminantes como EDTA, carbohidratos y fenol que absorben a 230 nm.

Se corroboraron estos resultados por medio de electroforesis en geles de agarosa al 1%, contrastando la concentración con la escalera *low DNA mass™ ladder* (Thermo Fisher Scientific). Las muestras que no cumplieron con la concentración esperada mínima de 100 ng/ $\mu$ L, y que además presentaran la relación 260/280 menor a 1.8, y la relación 260/230 menor a 1.5, se consideraron no adecuadas, por lo que se les realizó una segunda precipitación y lavado. Se les agregó a cada muestra 3 volúmenes de etanol absoluto y 10% del volumen de la muestra de Acetato de sodio (en este caso: 300  $\mu$ L de EOH y 10  $\mu$ L de acetato de sodio). Se enfriaron a -20°C por toda la noche y posteriormente se centrifugó por 15 minutos a 4°C a 15,294 xg. Se lavó el *pellet* con EOH al 70%, y se volvió a centrifugar y decantar. El *pellet* se lavó con EOH absoluto para centrifugarse por última vez, decantarse y secarse. Esta vez se diluyó en 50  $\mu$ L de agua para aumentar la concentración.

Se repitieron la cuantificación y evaluación de pureza de la misma manera, y si la muestra no presentó la pureza deseada, se realizó la extracción desde una nueva fracción del tejido colectado.

### **6.2.2 Selección de enzimas de restricción**

Previo a la preparación de las librerías, se hizo un análisis *in silico* para establecer el diseño experimental. Se utilizó el genoma de *Crassostrea gigas* publicado por Zhang *et al.* (2012), y una secuenciación 454 de *P. globosa* que se realizó para la obtención de marcadores microsatélites (Cruz-Hernández *et al.*, 2014). Se utilizó la secuencia del sitio de reconocimiento de distintas enzimas de restricción (ER), y se buscó el número de coincidencias que tuvieron en cada base de datos, por cada combinación de par de ER. Posteriormente se calculó, de todos los fragmentos obtenidos, cuántos tendrían un tamaño aproximado de 300 pb ( $\pm$ 50pb). De esta manera se obtuvo un estimado del número de *loci* que se obtendrían con cada combinación de enzimas, y se seleccionaron las que dieran un número mayor.

### 6.2.3 Elaboración de librerías con el método ddRAD-Seq

El método ddRAD-Seq (Secuenciado de ADN Asociado a Sitios de Restricción de Doble Digestión) (Peterson *et al.*, 2012) fue empleado para descubrir y obtener los genotipos de miles de marcadores SNPs distribuidos a lo largo del genoma de *P. globosa* simultáneamente. Se inició con la doble digestión de 1 mg de ADN de cada individuo, con las enzimas seleccionadas previamente (Anexo B): EcoRI y MspI (New England BioLabs). Las muestras se incubaron por 3 horas a 37°C. Las digestiones se purificaron con un volumen en una proporción de 1.8X de perlas (*AMPure XP beads*; New England BioLabs) empleando un soporte magnético para separarlas. Para cada paso clave del proceso se hizo una verificación midiendo la concentración del ADN mediante cuantificación fluorométrica con Qubit® (Thermo Fisher Scientific) y con una electroforesis para documentar el estado del ADN procesado. Posteriormente se ligaron adaptadores específicamente diseñados para unirse a los extremos cohesivos resultantes de la digestión con el par de enzimas de restricción mencionadas. Se emplearon adaptadores con un *barcode* único para cada individuo (P1, alineando los oligos 1.1 y 1.2-fosforilado en un radio 1:1 y calentándolos a 97° C 3 min) y el adaptador común P2 (alineando los oligos 2.1 y 2.2-fosforilado). Los adaptadores se ligaron en una concentración 10 veces mayor que los fragmentos de ADN producidos con extremos cohesivos en presencia de la enzima *T4 DNA ligase* (New England BioLabs) a una temperatura de 23°C por 30 min, seguidos de 65°C por 10 min para inactivar la enzima. Los ADNs de cada individuo marcados con un *barcode* distinto se mezclaron en concentraciones equimolares para crear un *pool* por población. En cada *pool* distinto, los fragmentos de ADN cortos y el exceso de adaptadores se removieron con perlas magnéticas (*AMPure XP*) en una proporción de 1.5x y los fragmentos resultantes fueron eluidos en 30 µL. Posteriormente se realizó la selección de fragmentos de aproximadamente 300 pb mediante el protocolo del equipo Pippin Prep (Sage Science). Con los adaptadores ya ligados, la media del tamaño seleccionado fue de 386 pb ± 26 pb. Para eliminar aquellos fragmentos con adaptadores P1 en ambos extremos 5' y 3', se realizó la limpieza de las librerías con perlas magnéticas. Las librerías resultantes fueron amplificadas por cuadruplicado mediante 10 ciclos de PCR con el kit Phusion Polymerase (NEB) para aumentar su concentración e incluir las secuencias de

alienación de celdas y los *primers* de secuenciación Illumina. Además, el primer reverse (PCR2) incluye un índice adicional capaz de identificar cada uno de los *pools* o bancos que contienen los *barcodes* únicos, permitiendo así la identificación de las variantes de cada uno de los individuos. Las reacciones por cada *pool* fueron combinadas, y purificadas con perlas magnéticas (AMPure XP), cuantificadas en el Agilent Bioanalyzer para determinar molaridad y la distribución del tamaño de los fragmentos y posteriormente las muestras con distintos índices introducidos mediante PCR se combinaron en concentraciones equimolares. La librería resultante fue secuenciada usando el servicio de, Genewiz (Nueva Jersey, Estados Unidos de América) en una línea de secuenciación Illumina HiSeq2000, para producir secuencias pareadas de ~150 pb por fragmento.

## **6.3 Análisis bioinformático**

### **6.3.1 Análisis exploratorio**

Se utilizó STACKS v1.47 (Catchen *et al.*, 2013b), que es un programa desarrollado para el análisis de datos provenientes de RADseq. Éste se implementa en una serie de algoritmos útiles para aplicaciones genómicas que van desde mapas de ligamiento, hasta estudios filogeográficos y de genómica de poblaciones.

Se secuenciaron los fragmentos en ambos sentidos (*single-end* y *pair-end*). Para unir los dos sentidos se utilizó el programa FLASH (*Fast Length Adjustment of Short reads*), el cual está diseñado para ensamblar los pares de lecturas cuando los fragmentos originales de ADN son más cortos que dos veces la longitud de la lectura (Magoč *et al.*, 2011).

Las secuencias combinadas se utilizaron para hacer la identificación de SNPs. Para ello se utilizaron paralelamente dos programas: AftrRAD v5.0 (Sovic *et al.*, 2015) y STACKS v1.46 (Catchen *et al.*, 2013b). Los parámetros que se establecen para el análisis de las secuencias en ambos programas se presentan en la tabla I.

**Tabla I.** Parámetros utilizados para la detección de SNPs.

Parámetro	AftrRADv5.0	STACKSv1.46
Enzima de restricción	AATTC	ecoRI
Calidad mínima (Phred)	17	20
Profundidad mínima	5	5
Identidad mínima (% / # mismatches)	90	3
Distancia máxima entre Stacks	No se especificó	3
# mínimo de lecturas para generar un stack	10	5
Barcode P2	CGAGATCGGAAG	No se especificó
Phred	33	33
Homopolimeros (longitud)	15	No se especificó
# de indels permitidos	3	No se especificó

En el programa STACKS los datos se analizaron como *single-end*. El *pipeline* general inicia con el procesamiento de las secuencias de las librerías, el objetivo de este primer paso es eliminar las lecturas de baja calidad, que tengan índices y códigos de barra indefinidos, recuperar las lecturas que se puedan, así como agrupar las lecturas por individuo (*demultiplexing* es el término en inglés). Para ello es importante introducir al programa información sobre las enzimas de restricción utilizadas “EcoRI” y “MSpI”, y la relación de los adaptadores y *barcodes* utilizados.

El programa AftrRADv5.0 puede analizar lecturas con longitudes diferentes, pues toma en cuenta las inserciones y deleciones como un paso mutacional, y potencialmente un SNP. Sin embargo, este programa requiere una gran capacidad computacional y espacio de almacenamiento, por ello se realizó sólo el *demultiplexing* y limpieza de todas las secuencias con este programa.

Una vez teniendo las lecturas por individuo, se seleccionó el individuo por *pool* con más lecturas, y se hizo un análisis de *clusters* con el programa CD-HIT (Li y Godzik, 2006). Con este programa se alinearon las secuencias, para posteriormente agrupar aquellas con un índice

de similitud mayor o igual a 96%, para permitir encontrar máximo un SNP por cada 25 bases. Se espera encontrar dos variantes en cada cluster, representando la diploidía del individuo, en caso contrario, se estaría asumiendo que el *barcode* no estaría representando a un solo individuo, y probablemente se trataría de un *pool*.

### 6.3.2 Obtención de SNPs

#### 6.3.2.1 Preparación y limpieza de datos

Se utilizó STACKS v1.47 (Catchen *et al.*, 2013b), que es un *software* desarrollado para el análisis de datos provenientes de RADseq, en específico el programa `process_radtags`, para separar las secuencias por individuo, eliminar las que tengan barcodes ambiguos y un puntaje de calidad menor a 20.

Una vez separadas las secuencias por individuo, se hizo de nuevo el *pool*, pero ahora de manera *in silico*, con líneas de código en Shell, además se generó un archivo maestro con todas las secuencias limpias de todos los individuos, conservando el formato original FASTQ, y haciendo una copia en formato FASTA (Fig. 2).

#### 6.3.2.2 Generación del catálogo de secuencias

Para poder identificar todos los *loci* muestreados, y las secuencias de cada banco que pertenecen a cada *locus*, es necesario crear un catálogo que correlacione las secuencias. Con el archivo FASTA se realizó un análisis de *clusters* para formar el catálogo de secuencias, la clusterización se llevó a cabo en el programa VSEARCH v2.7.1 (Rognes *et al.*, 2016). Teniendo como resultado un archivo multifasta con las secuencias consenso representativas de cada *cluster* formado, el cual se utilizó como referencia.

Para tener un catálogo por banco, se alinearon las secuencias de los archivos FASTQ de cada *pool* con el archivo de referencia con el programa BWA v0.7.17 (Li *et al.*, 2009a). Para vincular los tres catálogos se generó un archivo sincronizado (*\*.mpileup*) con el programa SAMTOOLS v1.7 (Li *et al.*, 2009b) (Fig. 2).

### 6.3.3 Estimación de parámetros genéticos

#### 6.3.3.1 Cálculo de Frecuencias alélicas

Se calcularon las frecuencias alélicas con la paquetería de PoPoolation2 (Kofler *et al.*, 2011), con el programa `snp-frequency-diff.pl`, y posteriormente se hizo una prueba exacta de Fisher (FET) con el programa `fisher-test.pl` de la misma paquetería, para evaluar el nivel de significancia de las diferencias entre las frecuencias alélicas calculadas.

Para los análisis realizados con PoPoolation2 se determinó una cuenta mínima de 5 para validar al alelo con menor frecuencia con el parámetro `--min-count`; una cobertura mínima de 10x que debe tener cada SNP en total en todos los grupos (bancos) con el parámetro `--min-coverage`; y una cobertura máxima de 100x con el parámetro `--max-coverage`.

Para evaluar si los valores de frecuencias alélicas fueron significativamente mayores en al menos un banco, se realizó un análisis de varianza (ANOVA) teniendo como hipótesis nula que todos los valores de frecuencia alélica son iguales entre sí. Posteriormente se realizó una prueba de Tukey de comparaciones pareadas para encontrar el banco con frecuencias significativamente más altas o bajas.

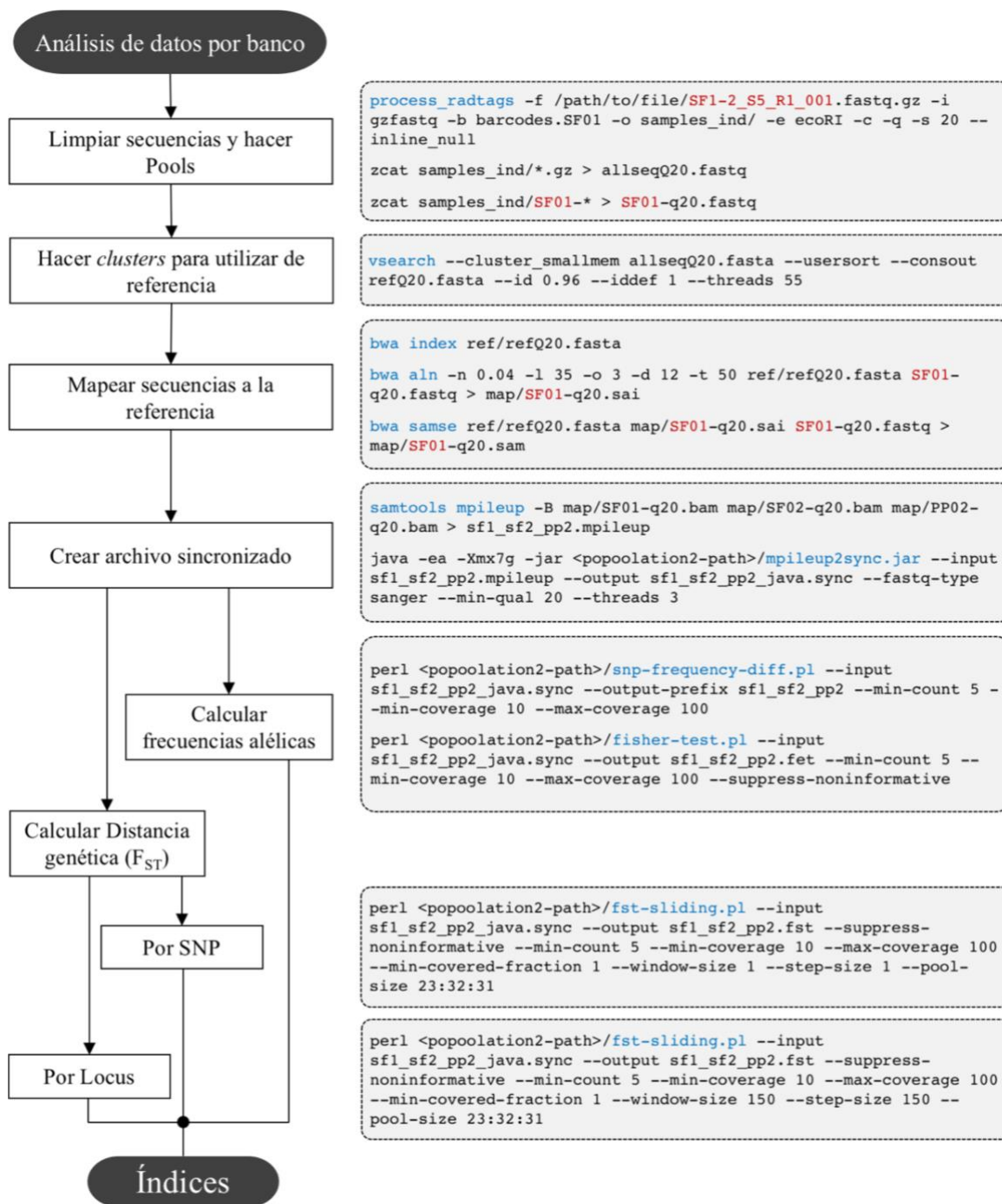
#### 6.3.3.2 Análisis de diferenciación genética (*F<sub>ST</sub>*)

Se calcularon las distancias genéticas con el Índice de Fijación de Wright (*F<sub>ST</sub>*). Se obtuvieron los valores de *F<sub>ST</sub>* por SNP y por *locus* con el Programa `fst-sliding.pl` de la paquetería PoPoolation2 (Kofler *et al.*, 2011). Para calcular el *F<sub>ST</sub>* por SNP se utilizó el parámetro `--window-size 1 --step-size 1`, y para calcularlo por *loci* se utilizó el tamaño de ventana igual a la longitud de fragmento `--window-size 146 --step-size 146` (Fig. 2).

Al igual que en el análisis de frecuencias alélicas, para los *F<sub>ST</sub>* por SNP y por *locus* se realizó un análisis de varianza (ANOVA) teniendo como hipótesis nula que todos los valores de *F<sub>ST</sub>* son iguales entre sí. Posteriormente se realizó una prueba de Tukey de comparaciones pareadas. Sin embargo, al haber más de un SNP por *locus*, estos se encuentran en ligamiento, y se considera que la información que nos brindan es repetitiva, por lo que basta con analizar

uno de ellos. Por ello, para calcular las medias, hacer el ANOVA y la prueba de Tukey, se tomó aleatoriamente un solo SNP por *locus*.





**Figura 2.** Diagrama de flujo del análisis de datos. Del lado izquierdo se encuentra el diagrama de flujo del análisis bioinformático, desde el procesamiento de los archivos recibidos de la secuenciación hasta la obtención de índices de frecuencia alélica y de diferenciación genética. Del lado derecho se encuentran las líneas de código que se utilizaron para cada paso, con los parámetros establecidos; en azul se destaca el programa utilizado; algunos programas se ejecutaron para cada set de datos, en rojo se marca el nombre utilizado para indicar los distintos archivos de entrada/salida.

### **6.3.4 Selección de loci para análisis futuros de asignación de origen (bancos)**

#### *6.3.4.1 Selección de loci outlier*

A partir de la base de datos de las diferencias en frecuencias alélicas de todos los SNPs encontrados, se realizó un primer filtrado seleccionando sólo los datos que resultaron significativos ( $p < 0.05$ ) en la prueba exacta de Fisher.

Para detectar los *loci outliers* se utilizó la función `boxplot.stats()$out` en R versión 3.5.0, la cual utiliza el método de Tukey para identificar los *outliers* distribuidos por fuera de 1.5 veces el rango intercuartílico.

#### *6.3.4.2 Selección de SNPs diseñados para futuros análisis de asignación*

Una vez teniendo la lista de los *loci outliers*, se realizó un nuevo filtrado para eliminar los que tuvieran SNPs en los extremos de cada secuencia. Para poder utilizar los *loci* reportados en este trabajo, para futuros análisis de asignación, es importante considerar el diseño de *primers* para amplificar los marcadores, por ello se seleccionaron los SNPs que se encontraran al centro del *locus*, eliminando así los *loci* con SNPs en las primeras o últimas 18 bases (Lapègue *et al.*, 2014).

#### *6.3.4.3 Análisis de estructura genética con loci outlier.*

Una vez obtenido el set de datos *outlier*, los *loci* seleccionados se extrajeron del archivo sincronizado \*.sync para analizarlo de nuevo con el programa `fst-sliding.pl` de la paquetería de PoPoolation2 (Kofler *et al.*, 2011), con el parámetro de ventana `--window-size 146` para incluir todos los SNPs dentro del *locus* y evitar evaluar por separado dos SNPs que se encuentren ligados, es decir que no se segreguen independientemente, bajo el esquema de desequilibrio de ligamiento. Para representar los valores de  $F_{ST}$  de los SNPs *outliers* se realizó un gráfico tipo *heatmap*, con un análisis de *cluster* por método de distancia en la plataforma R versión 3.5.0.

### ***6.3.5 Simulación de movimiento de partículas con el modelo HAMSOM***

Los sitios de muestreo fueron georeferenciados y se ubicaron dentro de los cuadrantes disponibles para el modelo de simulación de movimiento de partículas en el Alto Golfo de California (Marinone *et al.*, 2008).

Los datos de probabilidad de conectividad entre los cuadrantes muestreados teóricamente por Marinone *et al.* (2008), se adaptaron a las condiciones de *P. globosa*, con un pico de reproducción en invierno y duración larval pelágica de 13 días. Se calculó una matriz cuadrada de probabilidad de transporte de partículas con las dos localidades (San Felipe y Puerto Peñasco) con fuente de larvas/partículas y su probabilidad de transporte hacia éstas mismas. Se considera como índice de retención larval cuando el receptor de larvas/partículas es también la fuente de larvas/partículas.

## 7. RESULTADOS

### 7.1 Colecta de organismos

El número de ejemplares extraídos en cada banco está relacionado con la abundancia de recurso en la localidad, por lo que el tamaño de muestra total de cada banco fue variable (SF1 = 23, SF2 = 32, PP1 = 31), contando con un muestreo total de 86 individuos (Tabla II). La selección del banco fue realizada por el buzo con base en registros de colectas previas. La longitud del transecto y la distancia entre los bancos se obtuvo a partir de las coordenadas geográficas marcadas en los puntos de colecta (Tabla II).

**Tabla II.** Sitios de colecta en cada banco, por localidad.

Localidad	Punto <sup>1</sup>	Latitud	Longitud	<i>n</i>	Profundidad	Longitud transecto
San Felipe	b1, i	31.12272497	-114.657121	23	21.4m	60.42m
	b1, s	31.12269798	-114.657708			
	b2, i	31.10441504	-114.67384	32	19.8m	108m
	b2, s	31.10500999	-114.67469			
Puerto Peñasco	b1, i	31.214094	-113.27487	31	12.3m	25m
	b1, s	31.21428997	-113.274753			

Notas: <sup>1</sup>El Punto se refiere a la marca con GPS, siendo *b1* el banco 1, y *b2* el banco 2 de cada localidad. Se agregó una *i* si se tomó la coordenada en el punto de inmersión, y una *s* en el punto de salida del buzo.

### 7.2 Librerías genómicas RADseq

#### 7.2.1 Extracción de ADN y elaboración de librerías

Los distintos métodos de extracción y purificación de ADN genómico realizada dieron resultados variados, siendo el método de sales el que resultó con mayor rendimiento. Se logró obtener concentraciones adecuadas de ADN genómico de todas las muestras a través del método de sales (Anexo A); estas muestras fueron las seleccionadas para la elaboración de librerías.

El procedimiento de elaboración de librerías lleva pasos en los que se reduce la cantidad de ADN, teniendo una concentración media de 0.57 ng/μL por individuo antes de hacer los

*pools*. El producto final fue un *pool* por cada banco, con una concentración promedio de 39.63 ng/ $\mu$ L entre los tres bancos (Tabla III).

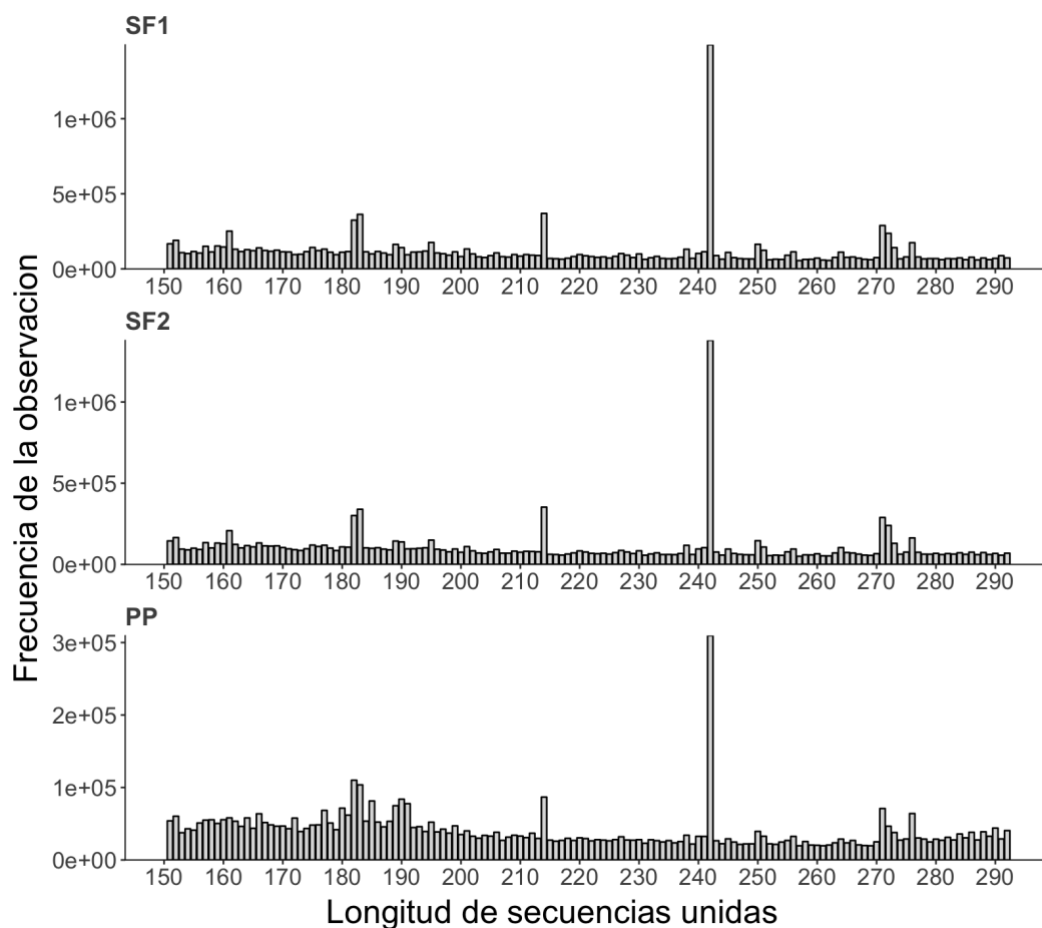
Los resultados del Agilent Bioanalyzer muestran que la selección de fragmentos por tamaño (300 pb) no fue exitosa, por lo que se obtuvieron fragmentos de tamaños desde 73 pares de bases (pb), hasta 2566 pb (Tabla III).

**Tabla III.** Concentraciones y tamaños de fragmentos de la librería genómica

<b>Banco</b>	<b>Agilent Bioanalyzer</b>					<b>Qubit Data</b>			
	Factor de dilución	Long. Min. (pb)	Long. Max. (pb)	Long. Media (pb)	Conc. (ng/ $\mu$ L)	Factor de dilución	Conc. (ng/ $\mu$ L)	Conc. Real (ng/ $\mu$ L)	Conc. (nM)
SF1-2	108	195	2566	762	96.552	3	42.8	128.4	255.31
SF2-2	108	74	2145	669	100.44	3	36.5	109.5	248
PP1-2	108	73	2391	555	167.4	3	39.6	118.8	324.32

### 7.2.2 Secuenciación de las librerías

Se realizó la secuenciación *pair-end* de fragmentos de 300 pb. Se secuenciaron más de 200 millones de fragmentos en ambos sentidos, al unir ambos sentidos se obtuvieron 36'549,819 fragmentos de longitudes desde 151 hasta 292 bases, con el programa FLASH (Fig. 3). Para no eliminar una gran cantidad de lecturas al seleccionar fragmentos con longitudes mayores a 150 bases, se decidió trabajar con las lecturas *single-end*, y de esta manera incluir todos los *loci* muestreados, perdiendo información sobre SNPs que se pudieran encontrar en la segunda mitad del fragmento. Sin embargo, se consideró que, por desequilibrio de ligamiento, la información de estos SNPs no incluidos sería redundante con los encontrados en las lecturas *single-end*, por lo que no se consideró una pérdida importante de información.



**Figura 3.** Frecuencia de tamaño de los fragmentos *sinle-end* y *pair-end* unidos.

Se obtuvo un total de 102 millones de secuencias *single-end* de los 86 individuos provenientes de los tres bancos analizados. Después de aplicar filtros de calidad en STACKS (Phred33 score > 20), eliminando *barcodes* ambiguos y secuencias sobre representadas, quedaron un total de 95.9 millones de secuencias de 146 bases de longitud.

La separación de las lecturas por individuo dentro de cada *pool*, utilizando el marcado con *barcodes*, dio como resultado un número desigual de éstas, con un solo *barcode* sobrerrepresentado dentro de cada banco. Las lecturas del *barcode* sobrerrepresentado fue por arriba del 90% del total de lecturas en cada banco. Éste patrón se vio con los resultados de STACKS y AfrRAD.

Con los *barcodes* sobre-representados se realizó un análisis de *clusters* en CD-HIT, en el cuál se corroboró que el *barcode* no representaba a un individuo, pues cada *cluster* contenía más de 20 variantes, cuando en un organismo diploide, como *P. globosa*, sólo se encontrarían los dos alelos. A partir de este resultado y una revisión exhaustiva de la elaboración de las librerías, se determinó que el marcado de *barcode* por individuo falló, pero el marcado por *pool* fue exitoso. En base a este resultado, los datos se trataron a nivel de banco/*pool*.

### **7.2.3 Obtención de SNPs**

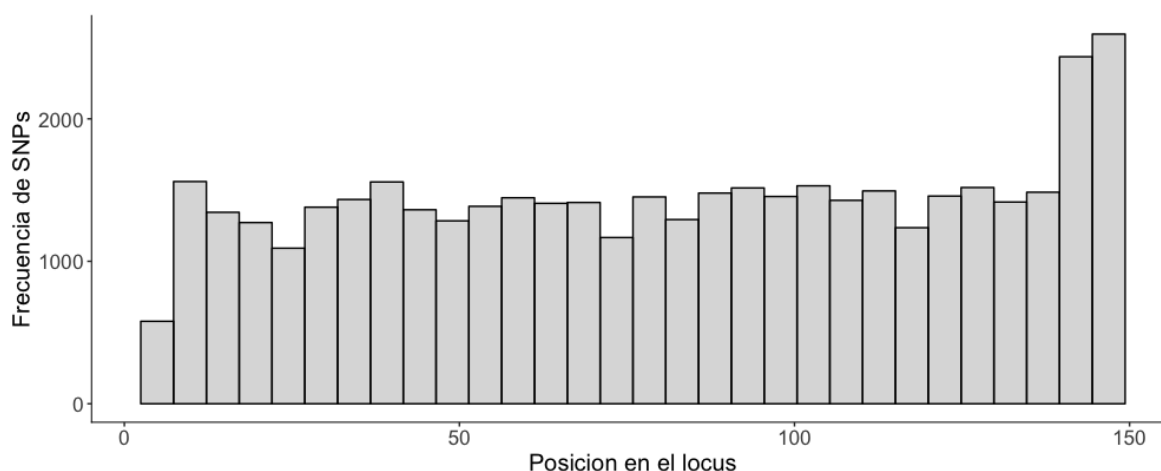
Las secuencias resultantes de la limpieza y *demultiplexing* en STACKS, se volvieron a agrupar en un *pool* por banco. Éstas fueron utilizadas para generar el catálogo de referencia, en el cual se crearon 2.5 millones de *stacks*, cada uno representando un *locus* con una o más secuencias asociadas a él. Al sincronizar los catálogos de los tres bancos, sólo se perdió un 3% de los *stacks* obtenidos anteriormente. Sin embargo, una gran cantidad de esos *stacks*, tuvieron pocas secuencias representándolos, y en el último filtro, quedó un total de 8,326 *loci*, con un promedio de 1.74 SNPs por *locus* (Tabla IV).

**Tabla IV.** Filtrado de datos. Número de individuos, *loci* y SNPs iniciales, y el número de éstos seleccionados después de pasar por filtros de calidad.

<b>Descripción</b>		<b>Porcentaje</b>
Número de individuos	86	
Número de secuencias totales obtenidas	102,758,345	100%
SF1 .....	35,466,520	
SF2.....	35,277,429	
PP.....	32,014,396	
Número de secuencias después de limpieza (Phred33>20, barcodes ambiguos)	95,901,147	93.33%
SF1.....	34,346,703	
SF2.....	32,864,199	
PP.....	28,690,245	
Stacks para catalogo (vsearch)	2,539,096	100%
Stacks sincronizados	2,459,182	96.85%
<i>Loci</i> presente en las 3 poblaciones (Min count 5, min coverage 10)		
SNPs.....	14,491	
<i>Loci</i> .....	8,326	0.33%
#Seqs.....	1,927,256	1.88%

La distribución de los SNPs a lo largo de cada lectura, se espera sea homogénea. Sin embargo, cuando se hace la secuenciación *single-end*, tiende a disminuir la calidad de las últimas bases. Después de pasar por los filtros resumidos en la tabla IV, al final se obtuvo una distribución casi homogénea de los SNPs a lo largo de las lecturas, con un aumento en la frecuencia de SNPs al final de las lecturas (Fig.4).





**Figura 4.** Frecuencia de SNPs por posición en la lectura.

## 7.3 Estructura poblacional

### 7.3.1 Frecuencias alélicas

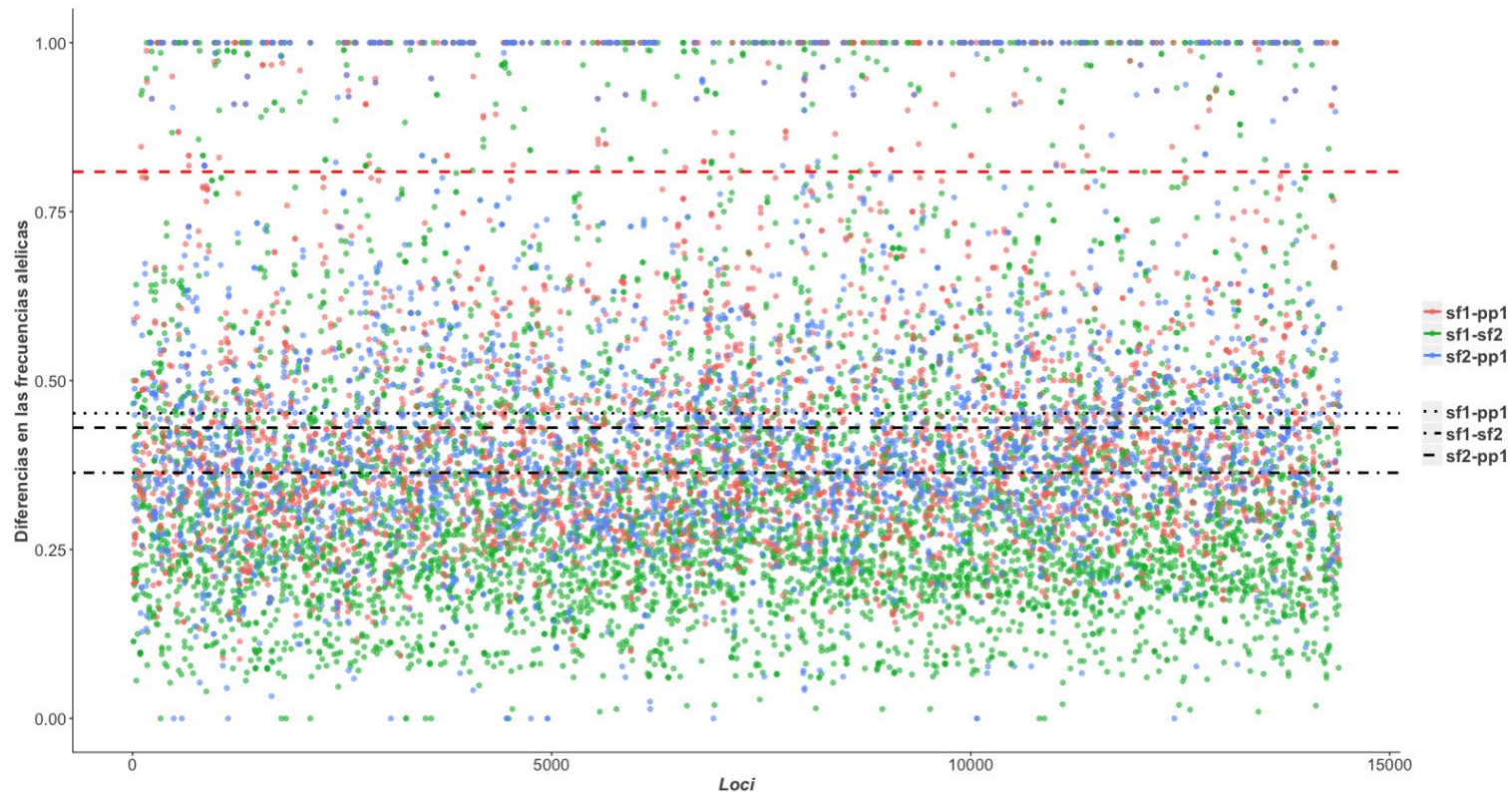
Se calcularon las frecuencias alélicas para el alelo más frecuente y menos frecuente en cada población, en donde el alelo más frecuente en la población debe tener una frecuencia mayor al 50%, y el de menor, inferior a este valor. En los tres bancos se observaron valores muy similares (Tabla V), en donde el valor promedio global para el alelo de mayor frecuencia fue del 73.5%, mientras que el promedio global del alelo menos frecuente fue del 26.4% (Tabla V).

Además, se determinó el alelo más variable por SNP por *locus*, este alelo puede ser el más frecuente de una población y el menos frecuente de otra, por ello, el rango de frecuencias del alelo más variable y del menos variable en cada SNP va de 0 a 1 y tiene una media en la frecuencia del 52.95% y 56.62% respectivamente (Tabla V).

**Tabla V.** Frecuencias alélicas.

	San Felipe B1	San Felipe B2	Puerto Peñasco
En el banco			
Alelo más frecuente	0.7345347	0.7335962	0.7378936
Alelo menos frecuente	0.2651027	0.2660102	0.2616869
En cada SNP			
Alelo más variable	0.5333	0.5139	0.5415
Alelo menos variable	0.5662	0.5520	0.58052

Se seleccionaron las diferencias entre las frecuencias alélicas que fueron estadísticamente significativas ( $p < 0.05$ ) en la prueba exacta de Fisher. En la figura 5 se puede observar la distribución de los valores de diferenciación, y un grupo de valores fuera de la distribución, cercanos a la diferenciación de 1 (separados por la línea roja punteada).



**Figura 5.** Diferenciación en las frecuencias alélicas entre los tres bancos. En una prueba de comparación pareada, se muestra en el eje de las ordenadas, el valor de diferenciación; mientras que a lo largo del eje horizontal se representa cada *locus* ( $n = 8,273$ ). En color rojo se representan las diferencias entre los bancos San Felipe B1 y Puerto Peñasco; en verde, entre San Felipe B1 y San Felipe B2; y en azul, entre San Felipe B2 y Puerto Peñasco. Sólo se muestran los datos que fueron estadísticamente significativos en la prueba exacta de Fisher. Las líneas negras discontinuas representan la media del grupo de datos, SF1-PP1 de 0.452, SF1-SF2 de 0.364 y SF2-PP1 de 0.430. Con una línea roja discontinua se representa el umbral de diferenciación (0.809), por encima de ésta se definen los valores *outliers*.

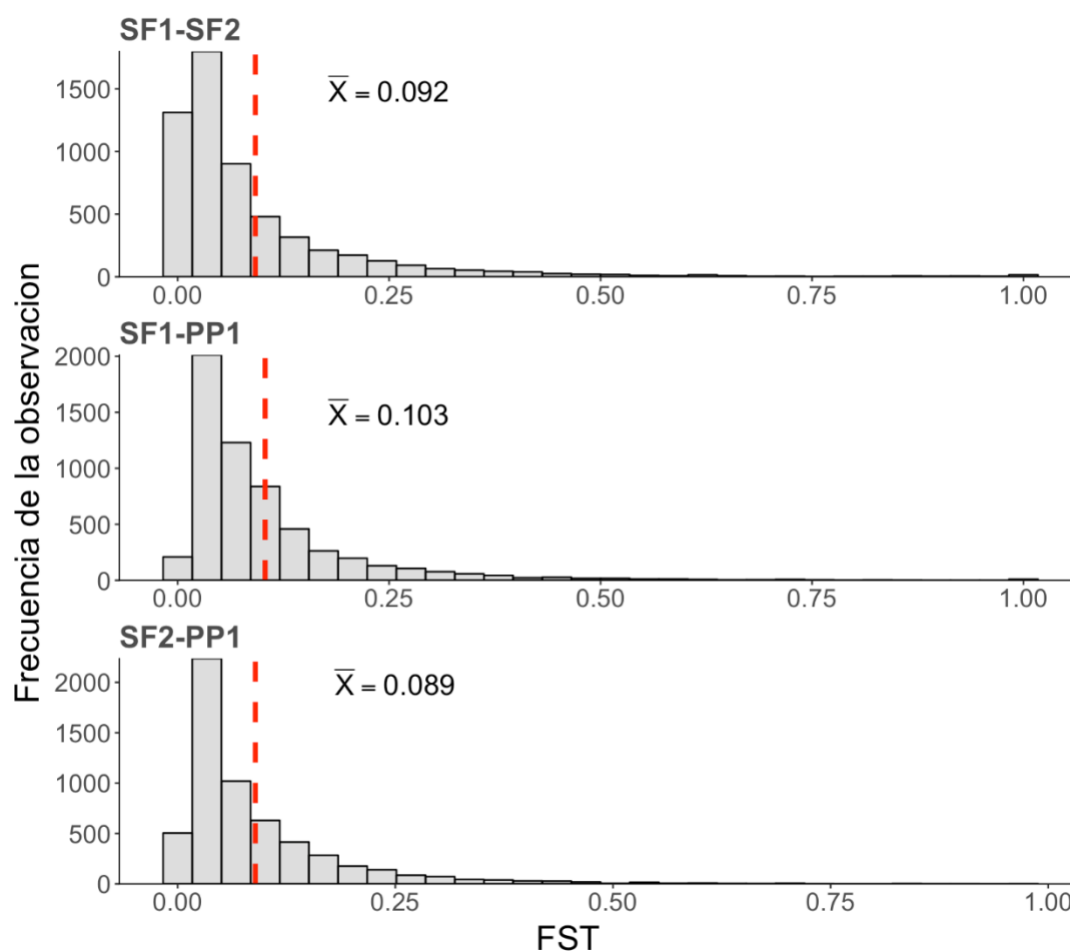
### 7.3.2 Diferenciación genética (*F<sub>ST</sub>*)

La diferenciación genética entre bancos se calculó por *locus* teniendo una ventana de 146 bases, que fue la longitud final de las secuencias. El ANOVA mostró que había diferencia significativa en al menos un grupo de valores de *F<sub>ST</sub>* por *locus* y por SNP. Con la prueba de Tukey se determinó que los valores de *F<sub>ST</sub>* a nivel de *locus*, entre los bancos de San Felipe B1 y Puerto Peñasco B2, fueron significativamente mayores a los valores de *F<sub>ST</sub>* entre los dos bancos de San Felipe, y entre San Felipe B2 y Puerto Peñasco B1. Mientras que todos fueron significativamente distintos a nivel de SNPs (Tabla VI).

**Tabla VI.** Valores de *F<sub>ST</sub>* por *locus* y por SNP. Por debajo de la diagonal se encuentran las medias de los valores de *F<sub>ST</sub>* obtenidos por *locus* ( $n = 8,326$ ). Por encima de la diagonal, están los valores promedio de *F<sub>ST</sub>* calculados por SNP ( $n = 14,491$ ). \* Significativamente diferentes ( $p < 0.01$ ).

	San Felipe B1	San Felipe B2	Puerto Peñasco
San Felipe B1	-	0.1142228*	0.1397516*
San Felipe B2	0.09178177	-	0.1214672*
Puerto Peñasco	0.10339534*	0.08916366	-

La distribución de la frecuencia de los valores de *F<sub>ST</sub>* calculados por *locus* se inclina a valores cercanos a cero, es decir, a una menor diferenciación genética entre bancos (Fig. 6), lo que indica que existe un gran flujo genético entre los bancos, reflejado en la mayor parte de los marcadores.

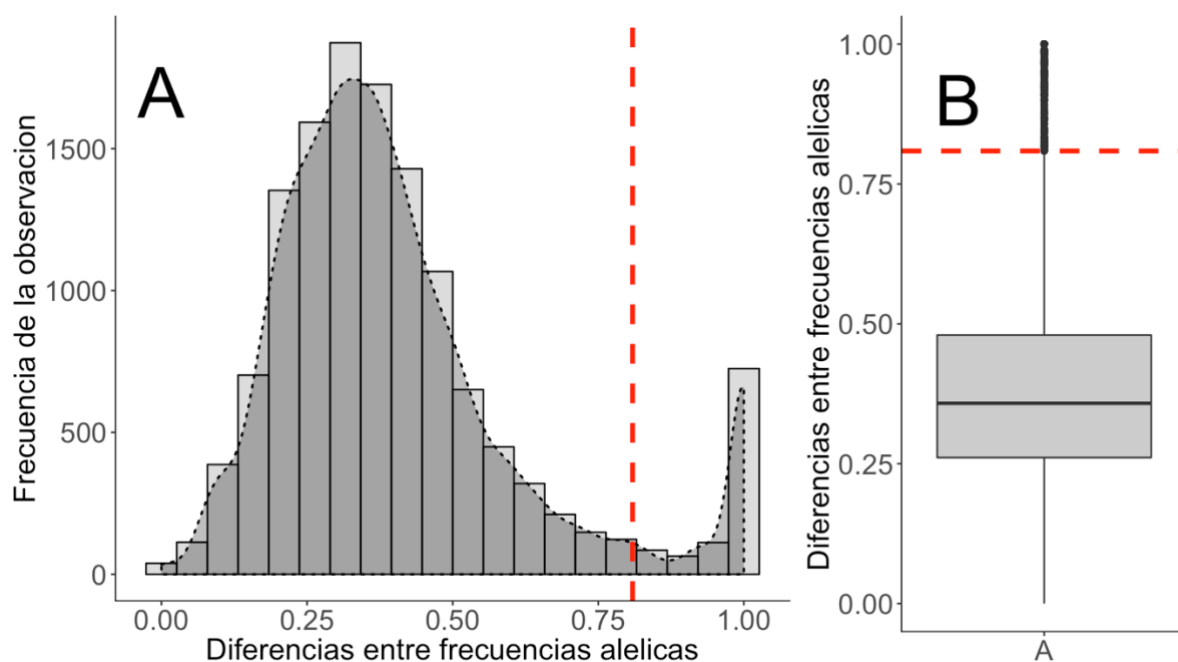


**Figura 6.** Frecuencia de los valores de FST en todos los *loci*.

### 7.3.3 SNPs con potencial para análisis de asignación

#### 7.3.3.1 Selección de SNPs con potencial para análisis de asignación

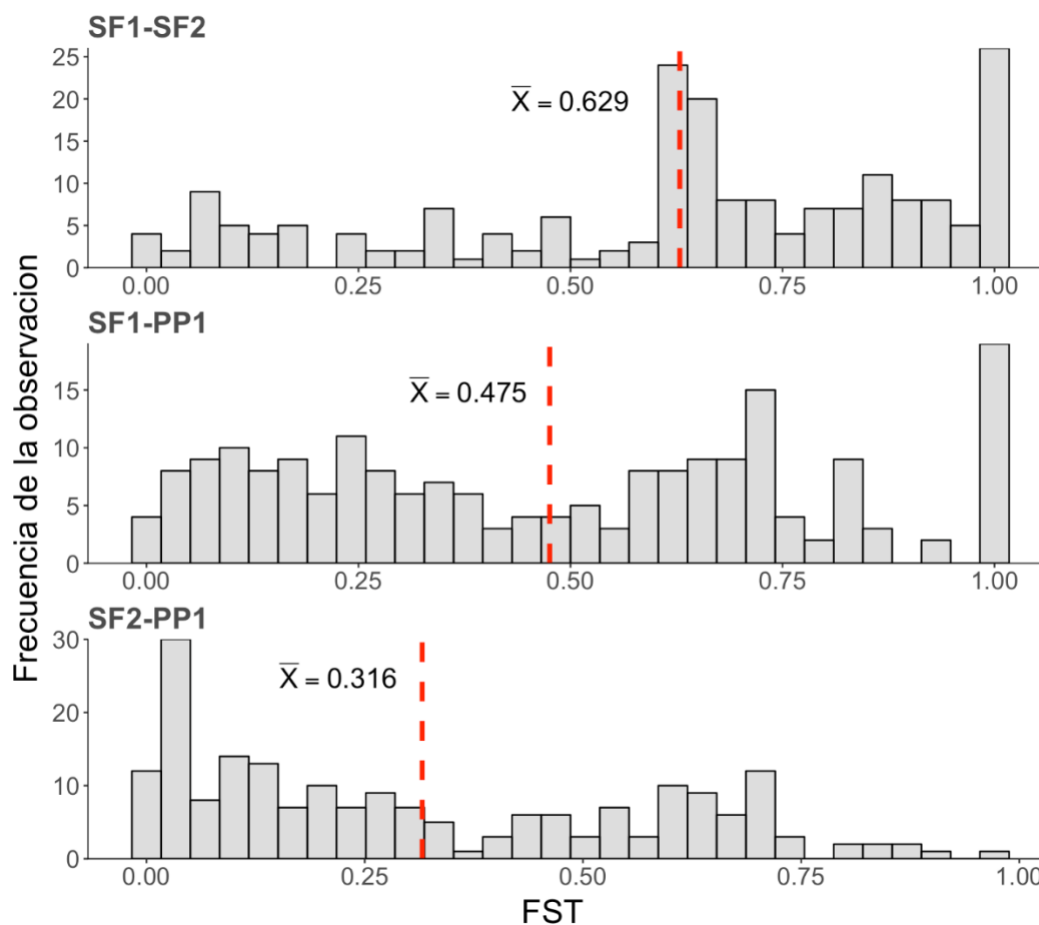
Se obtuvieron en total 199 SNPs *outliers* en base a la diferencia pareada entre las frecuencias alélicas de los bancos analizados. De este grupo de SNPs, el valor mínimo de diferenciación fue de 0.809, y el máximo de 1 (Fig. 7).



**Figura 7.** Histograma de frecuencia de las diferencias en las frecuencias alélicas en todos los SNPs e identificación de *outliers*. Se puede observar una distribución normal en los valores de diferenciación entre las frecuencias alélicas (A), donde los valores de 0 ocurren cuando las frecuencias alélicas son completamente iguales, y 1 cuando son completamente diferentes (eje de las ordenadas). El boxplot (B) representa el análisis de Tukey para identificar el grupo de datos *outliers* dentro de la distribución de éstas diferencias. El valor mínimo de los *outliers* fue igual a 0.809, el cual se representa en ambos gráficos con una línea discontinua roja.

### 7.3.3.2 Análisis de estructura poblacional con loci outlier

Se obtuvieron los valores de FST a partir de los *loci outliers*. Al ser *loci* con frecuencias alélicas muy distintas en al menos un par de bancos, el índice de fijación de Wright es más frecuente en los niveles más altos de estructura ( $F_{ST} \approx 1$ ).

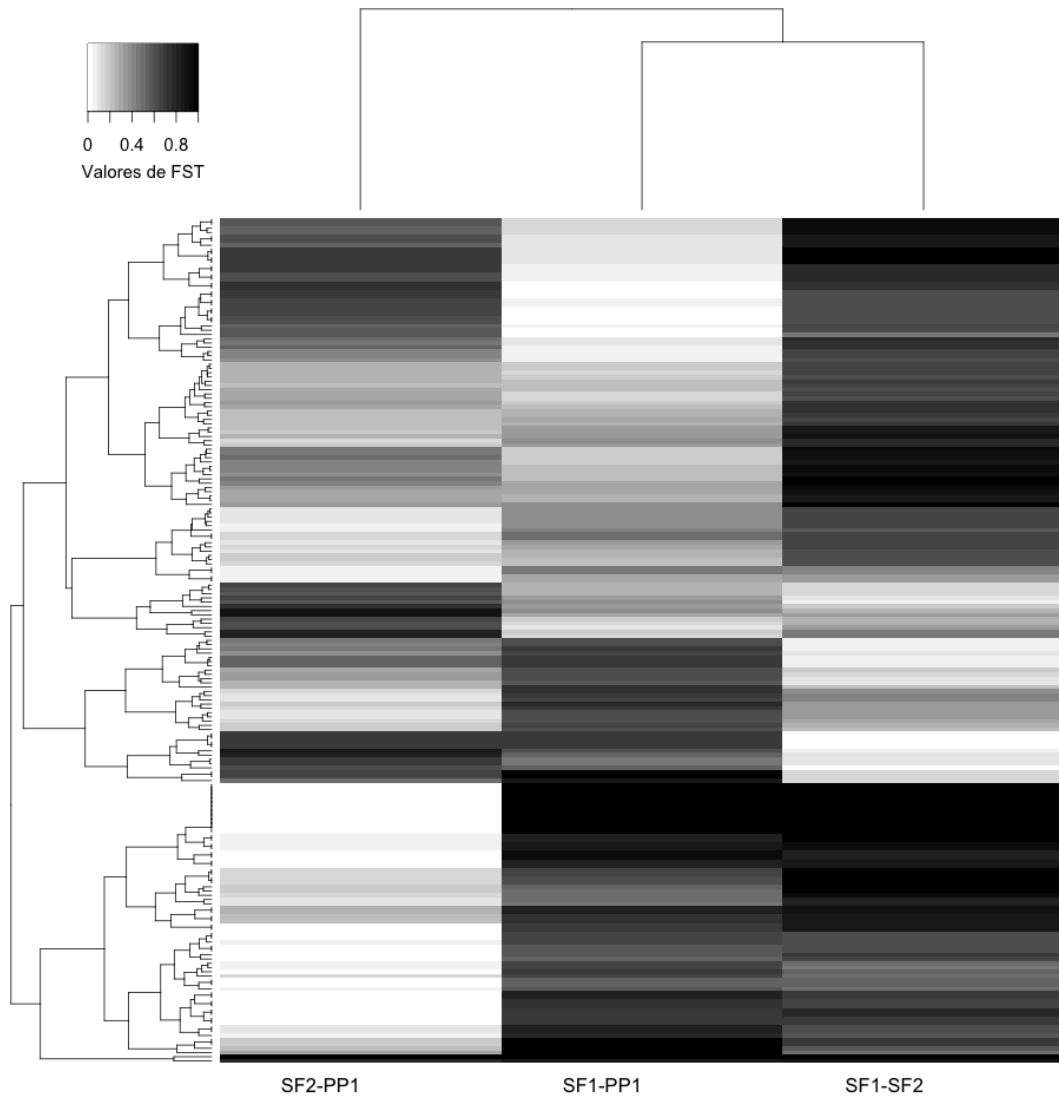


**Figura 8.** Frecuencia de los valores de FST pareados estimados para los *loci outliers*. La línea discontinua roja representa el FST promedio de cada comparación pareada.

Al analizar sólo los *loci outlier*, se encontró en promedio una mayor estructura ( $F_{ST\text{promedio}}=0.629$ ), entre los dos bancos de San Felipe (Fig. 8), y en segundo lugar ( $F_{ST\text{promedio}}=0.475$ ), los bancos de San Felipe B1 y Puerto Peñasco. Esto hace resaltar al Banco de San Felipe B1, que es el que presentó mayor diferencia contra los otros dos bancos.

Para seleccionar los SNPs con potencial para análisis de asignación por banco, son más útiles los *loci* que tengan un valor alto de FST en al menos dos de las comparaciones pareadas. En la figura 9 se presenta un gráfico tipo *heatmap*, en donde se pueden seleccionar los *loci* (renglones) que presentan los valores mas altos en dos de las comparaciones pareadas de FST; sobresalen en el último tercio de los renglones los *loci* que tienen potencial para distinguir el banco de SF1, de los otros dos bancos (SF2 y PP1), lo cual nos corrobora la mayor diferenciación del banco SF1 que había sido observada previamente en el análisis con la totalidad de los *loci*.



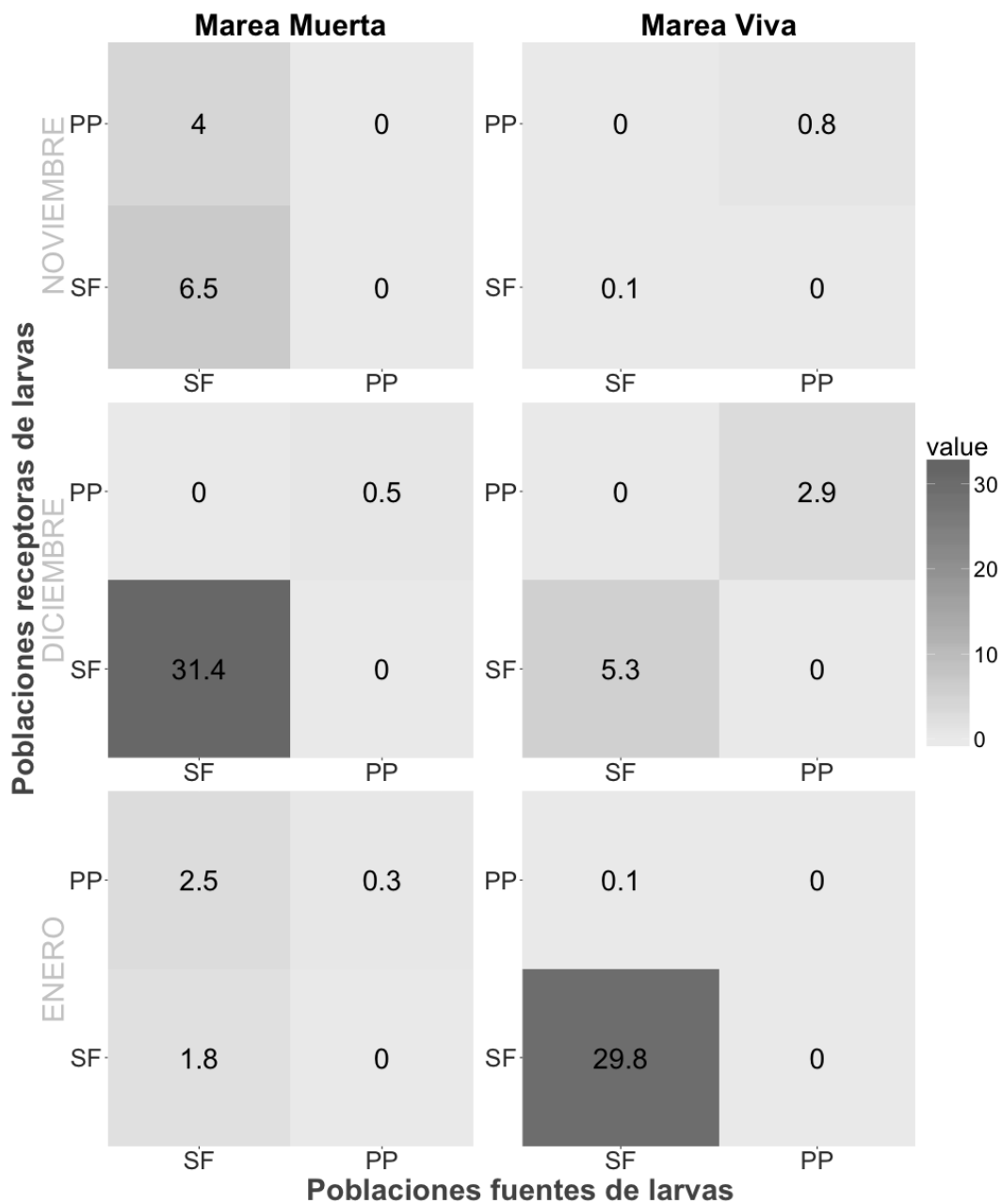


**Figura 9.** Valores de FST de los *loci outliers* y clusterización por similitud. Cada renglón representa un *locus*, el valor está representado sin escala, y ordenado por similitud, calculada con una función de distancia.

## 7.4 Conectividad entre bancos

### 7.4.1 Modelo HAMSOM

El modelo HAMSOM adaptado por Marinone (2008) para el Alto Golfo de California, y en este trabajo para *P. globosa*, con época de desove en invierno y duración larval de 13 días, muestra patrones de dispersión larval contrastantes. La máxima probabilidad de transporte de partículas se encuentra en el mes de diciembre con marea muerta, y en enero con marea viva ( $p = 31.4\%$  y  $p = 29.8\%$  respectivamente), ambas probabilidades representando retención local en la población de San Felipe. Este patrón es el que tiene la mayor probabilidad en todos los casos presentados ( $p$  entre  $0.1\%$ - $6.5\%$ ), con excepción de la marea muerta en enero, cuando hay mayor probabilidad de transporte de partículas desde San Felipe hacia Puerto Peñasco ( $p = 2.5\%$  contra  $p = 1.8\%$ ); este patrón de transporte tiene probabilidades de  $0.1\%$  -  $4\%$ , pero nulas en el mes de diciembre y en la marea viva de noviembre. También existe, en menor medida, probabilidad de retención de partículas en Puerto Peñasco ( $p$  entre  $0.3\%$ - $2.9\%$ ), con probabilidades nulas en la marea muerta de noviembre y la marea viva de enero. Por último, cabe resaltar que no hay probabilidad, en ningún caso, que el transporte sea desde Puerto Peñasco hacia San Felipe (Fig. 10).



**Figura 10.** Probabilidades de transporte de partículas entre poblaciones de San Felipe y Puerto Peñasco en la temporada de invierno y con duración larval de dos semanas.

## 8. DISCUSIÓN

### 8.1 Diseño experimental

#### 8.1.1 *Colecta de organismos*

La densidad de organismos que se encontró en cada localidad representa la abundancia intrínseca del lugar, pues fue necesario un mayor esfuerzo (tiempo de inmersión) en Puerto Peñasco, para poder obtener un tamaño de muestra cercano a lo que se obtuvo en San Felipe. Los buzos reportaron que el sustrato en San Felipe es más arenoso que en Puerto Peñasco, y por tanto favorable para el asentamiento, lo cual puede ser una de las razones por las que varía la abundancia de organismos, sin embargo, no se tomaron datos al respecto.

En el muestreo se buscó que la distancia entre bancos de San Felipe fuera lo más alejada posible para asegurar que se contara con diferentes bancos.

Una de las ventajas de las nuevas técnicas para obtener datos genómicos es que es posible hacer diseños experimentales con tamaños de muestra muy grandes, a muy bajos costos (Li y Wang, 2017; Ribeiro *et al.*, 2017). La técnica de RADseq además agrega la ventaja de poder trabajar con organismos no modelo y obtener numerosos marcadores (Andrews *et al.*, 2016), siendo muy eficiente en costos por individuo en contraste con el número de marcadores que se pueden obtener por marcadores neutrales sencillos, como los microsatélites (Thrasher *et al.*, 2017). En este trabajo se utilizó una muestra total de 86 individuos, que a comparación de otros trabajos que tienen tamaños en el orden de cientos de individuos (Araneda *et al.*, 2016; Catchen *et al.*, 2013a; Guo *et al.*, 2015; Hohenlohe *et al.*, 2010; Lal *et al.*, 2016; Richards *et al.*, 2013; Sandoval-Castillo *et al.*, 2018; Wyngaarden *et al.*, 2017). Cabe mencionar que la línea de secuenciación donde se corrió el ADN de los 86 individuos se compartió con más individuos de *P. globosa*, que se utilizarían para otros análisis poblacionales a una escala mayor, incluyendo toda la distribución, por lo que sí se hizo un aprovechamiento adecuado del recurso de secuenciación.

### 8.1.2 Purificación de ADN y elaboración de librerías

Un paso clave para obtener lecturas de buena calidad en la secuenciación, es la extracción de ADN (Graham *et al.*, 2015). Es importante que éste se haya obtenido por el mismo método para homogenizar el resultado. A pesar de que el método de extracción con sales es de mayor duración, se logró obtener mayor cantidad de ADN por este método, que con el kit *Wizard® SV 96 Genomic* de Promega.

Uno de los puntos clave del diseño experimental en este trabajo fue la elaboración de las librerías genómicas, donde el paso de selección de fragmento por tamaño no se realizó con la eficiencia deseada; y el segundo paso crítico fue el diseño del marcaje de los fragmentos por individuo. El primer punto se reveló primeramente en los resultados del Bioanalyzer, que mostraron tamaños de fragmentos dentro de las librerías desde 73 pb hasta 2566 pb (Tabla III) y se corroboró una vez secuenciado, al alinear y unir ambas lecturas (*pair-end* y *single-end*) bioinformáticamente, con lo que se descartó el 70% de las lecturas que posiblemente eran fragmentos menores a 150 pb, o lecturas que no se pudieron alinear por no traslaparse en al menos 10 bases pues eran de fragmentos mayores a 300 pb. Más, si se toma en cuenta el método original de RADseq (Baird, *et al.*, 2008), este no incluye selección de fragmentos por longitud, y lo que esto implica es una mayor cantidad de *loci* muestreados, pues no limita a seleccionar aquellos que tengan a  $\approx 300$  bases de distancia el sitio de reconocimiento para la segunda enzima de restricción (Andrews, *et al.*, 2016).

El segundo punto clave fue el marcado de los individuos con los *barcodes*. Este puede ser causado por una asignación no adecuada de los *barcodes*, pues no se consideró un equilibrio entre las bases en cada posición dentro de *barcodes*, lo cual al momento de la secuenciación causó una mala asignación de bases en las primeras seis lecturas (longitud del *barcode*). Este problema durante la secuenciación se ha reportado recientemente para las nuevas plataformas de secuenciación como *index hopping* (van der Valk, *et al.*, 2018) el cual es provocado por la longitud de las lecturas (lecturas cortas), el contenido de GC, mientras que la mala asignación en proporciones menores puede explicarse por formaciones quiméricas durante el paso de amplificación del *pool* (Illumina, 2018).

### 8.1.3 Obtención de SNPs

De los 2.5 millones de *loci* putativos, que se encontraron en la secuenciación de las librerías tipo RADseq, se trabajó con un total de 8,326 *loci* con suficiente cobertura, en los que se encontró al menos un polimorfismo. Si tomamos en cuenta el tamaño del genoma completo de un molusco ampliamente estudiado, *Crassostrea gigas*, que es de aproximadamente 558 Mb (Zhang, *et al.*, 2012), la cobertura genómica de la técnica ddRADseq, con las modificaciones realizadas en este trabajo, submuestreó el 0.218% de un genoma de tal tamaño. Este valor coincide con el 0.21% reportado para el único molusco, *Lottia gigantea*, con el par de enzimas que se utilizó en este trabajo (EcoRI y MspI), en el trabajo de Peterson *et al.* (2012) donde publica la técnica de ddRADseq. Al igual que Peterson, que encuentra una mayor cobertura, de 0.64%, con el par de enzimas de restricción NlaIII y MluCI en *L. gigantea*, al hacer las simulaciones de corte *in silico* para seleccionar la enzima, éste último par mencionado dio como resultado una frecuencia de corte mucho mayor (Anexo B). Sin embargo, no se seleccionó este par, pues un mayor número de *loci* dentro de la línea de secuenciación, significaría sacrificar la cantidad de individuos muestreados, o la profundidad de secuenciación (Kenny *et al.*, 2011; Quail *et al.*, 2008).

## 8.2 Estructura genética

Se observó que, en base a las diferencias significativas en las frecuencias alélicas, la diferenciación promedio es menor entre los dos bancos de San Felipe, mientras que la diferenciación de éstos con Puerto Peñasco, es mayor, lo cual es coherente con la distancia geográfica entre los bancos de *P. globosa*.

En cuanto al conjunto de valores de FST pareados entre San Felipe B1 y Puerto Peñasco, resultó significativamente mayor en el ANOVA y la prueba de Tukey (Tabla VI), que es lo esperado considerando que hay una mayor diferenciación entre dos bancos de distintas localidades y se corrobora lo observado con las frecuencias alélicas.

Los valores de  $F_{ST}$  indican el nivel de estructura que hay entre los bancos, y los valores reportados en este trabajo, son los mayores reportados analizando las poblaciones de *P. globosa* del Alto Golfo. Esto puede ser porque los trabajos previos han utilizado sólo marcadores neutrales para evaluar la estructura de las poblaciones, mientras que en este se están utilizando todos los *loci* que se muestrearon a lo largo del genoma, sin embargo, no fue posible seleccionar los SNPs neutrales, pues para evaluar cuáles están en equilibrio de Hardy-Weinberg, es necesario calcular la heterocigocidad, y al no poder identificar el genotipo por individuo, no fue posible hacer tal cálculo.

Sin embargo, el patrón de diferenciación corrobora lo encontrado por Munguía-Vega *et al.* (2015) con un  $F_{ST} = 0.024$  entre las dos poblaciones del Alto Golfo, el cual fue significativamente diferente a 0 y uno de los más elevados (de 0.018 a 0.026) entre las poblaciones evaluadas, incluyendo a Bahía Magdalena que es la población conocida más alejada del Alto Golfo. Esta diferenciación genética de las poblaciones del Alto Golfo también se observó en una evaluación filogenética con el marcador ribosomal ITS-1 (Rocha-Olivares *et al.*, 2010); que separa en dos clados a las poblaciones de San Felipe y Puerto Peñasco. En contraste, con otro estudio (Suárez-Moo *et al.*, 2013) con un marcador mitocondrial (COI) no se pudo diferenciar a las dos poblaciones, aunque hay que tomar en cuenta que este marcador es más conservado que los marcadores nucleares y es usado precisamente para distinguir a nivel de especies.

El incorporar miles de marcadores tipo SNP, esclarece la presencia de flujo genético con un gran número de marcadores que no diferencian en gran medida a las poblaciones del Alto Golfo. Pero también se es capaz de distinguir *loci* con frecuencias alélicas muy distintas, reflejo de otras fuerzas que pueden estar operando, como lo es la selección natural.

Para el Alto Golfo de California, este es uno de los primeros trabajos en incorporar miles de marcadores genómicos para la evaluación de estructura poblacional de un organismo marino sésil. Además de que el grado de estructura denotado por los valores de  $F_{ST}$  con marcadores neutros y *outlier* ( $F_{ST\text{neutral}} = 0.09$  y  $F_{ST\text{outlier}} = 0.47$ ) es superior a otros trabajos que hacen la comparación con *loci* neutrales y candidatos a selección. Por ejemplo, *Pinctada*

*margaritifera* con  $F_{ST}^{neutral} = 0.046$  y  $F_{ST}^{outlier} = 0.15$  (Lal *et al.*, 2016), *Mytilus chilensis* con  $F_{ST}^{neutral} = 0.005$  y  $F_{ST}^{outlier} = 0.114$  (Araneda *et al.*, 2016), *Placopecten magellanicus* con  $F_{ST}^{neutral} = 0.004$  y  $F_{ST}^{outlier} = 0.15$  (Wyngaarden *et al.*, 2017), y *Haliotis laevigata* con  $F_{ST}^{neutral} = 0.0081$  y  $F_{ST}^{outlier} = 0.267$  (Sandoval-Castillo *et al.*, 2018).

### 8.3 SNPs con potencial para análisis de asignación por grupo de origen

Una alta diferenciación de los *loci* entre distintas localidades normalmente es atribuida a una selección adaptativa al ambiente, ya sea directamente (genes involucrados) o por cercanía (ligamiento) a regiones bajo esta selección (Wenne *et al.*, 2018); aunque también se reconocen otros procesos que interfieren altamente en la diferenciación de regiones genómicas o SNPs, como recombinación reducida o reordenamientos cromosómicos (Yeaman, 2013). Este trabajo difiere considerablemente de las aproximaciones a la detección de *loci outlier* basadas en modelos de evolución neutral, que se han usado ampliamente para analizar marcadores anónimos predominantes (Excoffier, *et al.*, 2009; Fischer, *et al.*, 2011). En este caso, por la naturaleza de los datos, la identificación de los SNPs fue validada tomando en cuenta la variación estocástica en la cobertura de las secuencias para cada variante, un aspecto central cuando se utilizan datos NGS. Además, se aplicó un umbral riguroso (diferenciación de frecuencias alélicas  $> 0.809$ ) para elegir SNPs altamente diferenciados. En consecuencia, los valores de FST analizando estos SNPs fueron también más altos. Además, es importante mencionar que se consideraron sólo los SNPs que estuvieran en *loci* que pudieran ser amplificados a futuro mediante PCR, eliminando aquellos que sólo tuvieran SNPs al inicio o final de la secuencia. Así, los 199 SNPs altamente diferenciados, representan el 0.014% del total detectado en este trabajo.

La obtención de un número reducido de SNPs altamente diferenciados, es ventajoso en el sentido de reducir costos para obtener marcadores óptimos para análisis de asignación de banco de origen, con el fin de evaluar el estado de las poblaciones de *P. globosa* a una escala microgeográfica y hacer un mejor manejo de la especie. Además, un set de marcadores de este tipo, tiene un uso potencial para la trazabilidad comercial de *P. globosa*.



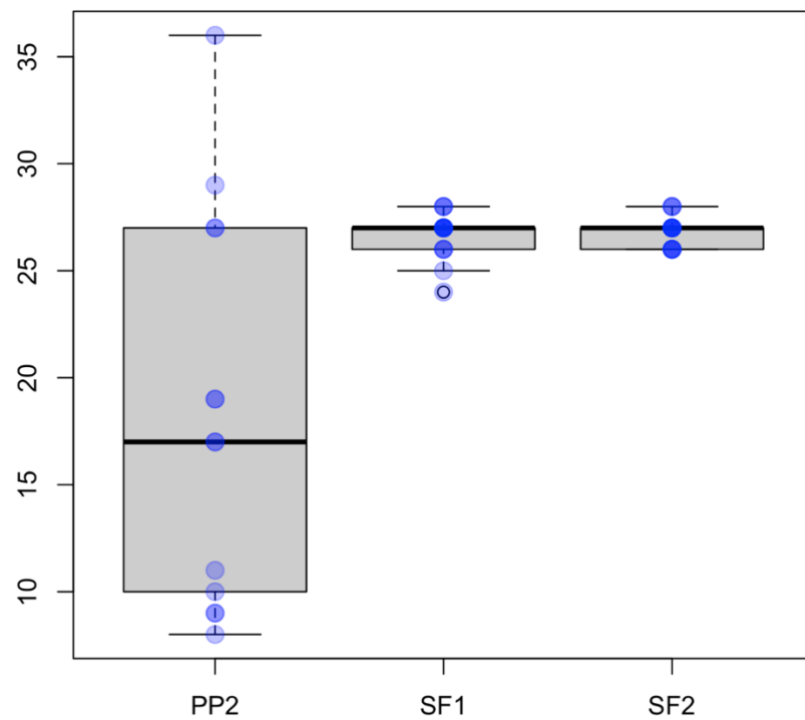
Aún cuando está pendiente la validación de los *loci outlier* seleccionados en este trabajo, se podría esperar que éstos sean tan variables que se puedan utilizar para diferenciar más bancos que se encuentren en las poblaciones del Alto Golfo, o hasta para otras poblaciones de *P. globosa* a lo largo de su distribución. A la par del presente estudio, se construyeron librerías genómicas para otras dos localidades, una en la parte media del Golfo de California (Punta Lobos), y otra en la costa del Océano Pacífico (Bahía Magdalena), por lo que un análisis incorporando estas localidades reforzará la búsqueda de marcadores con fines de trazabilidad.

Sin embargo, para generar un panel de SNPs con todos los posibles usos antes mencionados, sería necesario poner a prueba estos *loci*, pues debido a la limitante de trabajar con *pools* en vez de genotipos individuales, no fue posible realizar los análisis de asignación, los cuales se pueden lograr con programas como STRUCTURE (Pritchard *et al.*, 2010), GenAlEx (Peakall *et al.*, 2012), GeneClass (Piry *et al.*, 2004), entre otros.

### **8.3.1 Estructura entre bancos a partir de los *loci outlier***

La diferenciación entre los tres bancos es alta basándose en los *loci* seleccionados como *outlier*, siendo esta una lista de marcadores con potencial para análisis de asignación. Lo esperado es que el banco de Puerto Peñasco sea el que tenga más valores de estructura alta, lo cual sucede al compararlo con San Felipe B1. Sin embargo, al contrastar Puerto Peñasco con San Felipe B2, se ve una tendencia hacia valores menores en el índice de fijación (Fig. 8). Sin embargo, esta tendencia puede ser el reflejo de características específicas de los individuos muestreados en cada banco. Por ejemplo, vemos un número reducido de cohortes en ambos bancos de San Felipe, mientras que hay muchas más en Puerto Peñasco (Fig. 11). El número reducido de cohortes hace que lo representado en los datos esté siendo más afectado por diferencias entre familias generadas en unos cuantos eventos reproductivos en particular, que por otras fuerzas que moldean a las poblaciones. Además, el número de cohortes en sí, puede estar indicando que el asentamiento de los organismos no sea aleatorio, sino siguiendo un patrón de parcheo genético caótico (*chaotic genetic patchiness*, en inglés),

hay cuatro mecanismos que pudieran generar este parcheo: selección, un sorteo en el éxito reproductivo, dispersión colectiva, o cambios temporales en la dinámica poblacional local (Eldon *et al.*, 2016). Este tipo de parcheo caótico se ha observado en otra especie de almeja generosa, *P. generosa*, analizando 11 aloenzimas, 7 microsatélites y edades en más de 2 mil individuos, encontrando un asentamiento en parches que puede ser explicado por la conectividad hidrológica entre sitios, pero que también es influenciado por el comportamiento de la larva al asentarse, esto último, afectando el éxito en el reclutamiento de organismos en cada sitio, a pesar de la codispersión de larvas (Vadopalas *et al.*, 2012).



**Figura 11.** Boxplot de distribución de edades de una submuestra de 15 organismos por banco. Análisis a partir de la base de datos de edades de Hernández-Muñoz (2017).

#### 8.4 Conectividad genética

El modelo HAMSOM ha sido utilizado para probar hipótesis sobre la dinámica metapoblacional de la fuente y sumidero de larvas, describiendo la dirección y frecuencia de la dispersión larval de acuerdo al modelo oceanográfico propuesto.

El uso de diversas metodologías podría ayudar a avanzar un incremento en el interés de incorporar información genómica para la planeación y el manejo de recursos marinos, por ejemplo, identificando sitios con alta diversidad genética que tenga potencial evolutivo de adaptación a futuros cambios ambientales (Beger *et al.*, 2014). Sin embargo, no se ha reportado un modelo como *P. globosa* con una DLP relativamente corta (14 días) y con desove en la época de diciembre.

Peguero-Icaza *et al.* (2010) también reporta valores de retención menores al 35% para partículas que se dispersan en la temporada de invierno, como es el caso de *P. globosa*.

Las probabilidades de asentamiento de larvas coinciden con las tasas de migración direccional calculadas con microsátélites (Munguía-Vega *et al.*, 2015); la tasa de migración mayor es una retención de 0.987 en San Felipe, al igual que la probabilidad mayor (31.4%) calculada con el modelo oceanográfico. Se observó la retención en Puerto Peñasco, con una tasa de migración de 0.309 mientras que la probabilidad del modelo fue muy baja (0.75%). La tasa de migración de San Felipe a Puerto Peñasco fue de 0.676, pero también fue muy baja para el modelo oceanográfico (1.1%); mientras que en sentido contrario es la tasa menor, de 0.003, cuando la probabilidad calculada con el modelo fue de 0 para esta ruta, en todos los casos de noviembre a enero. Aunque el nivel de conectividad difiere en ambos trabajos, se guarda la proporción entre las posibles rutas.

Este acercamiento provee una explicación general a la dinámica de la conectividad entre poblaciones de *P. globosa* en el Alto Golfo de California. Sin embargo, el modelo oceanográfico HAMSOM para el Golfo de California, aún no provee la resolución para hacer el análisis a nivel de bancos, y posiblemente esté ocurriendo una dinámica distinta que a la escala poblacional, como se observó con los datos *outlier*.

## 9. CONCLUSIONES

En este trabajo se demostró que la caracterización de miles de marcadores SNP, ha sido efectiva para evaluar la estructura de las poblaciones y a una escala geográfica menor en comparación con los marcadores genéticos tradicionales y a escalas geográficas superiores. Y aún más importante, la caracterización de SNPs *outliers* tiene aplicaciones para estudios con una profundidad sin precedentes, ya sea para detectar regiones genómicas bajo selección, para poder distinguir entre grupos de individuos (por ejemplo, poblaciones, bancos o morfotipos) con un gran flujo genético, o para fines comerciales de trazabilidad.

La técnica de secuenciación masiva ddRADseq fue eficiente para el número de marcadores muestreado, pues se obtuvo la cobertura genómica deseada ( $\approx 0.2\%$ ) bajo el diseño experimental establecido. Sin embargo, cabe resaltar la importancia de hacer un marcado a nivel de individuos en las librerías de manera eficiente, para poder tener un mayor alcance en los análisis poblacionales que se puedan realizar, lo cual fue una limitante en este trabajo. Sin embargo, con el correcto marcado de las librerías, fue posible caracterizar 14,491 SNPs de tres bancos del Alto Golfo de California, y a partir de ellos, manejando los datos como grupos (*pools*) se logró calcular frecuencias alélicas a nivel de banco, y proceder a su comparación *inter-bancos* y análisis de estructura genética.

Las diferencias en las frecuencias alélicas encontradas utilizando los 14 mil *loci*, lograron distinguir entre los bancos de San Felipe y el banco de Puerto Peñasco. Además, se logró el objetivo de tener marcadores capaces de diferencias entre los bancos de una misma población (San Felipe).

La estructura poblacional, evaluada con el índice de fijación de Wright (FST) dio una mayor diferenciación promedio estadísticamente significativa entre el banco de San Felipe B1 y Puerto Peñasco.

En base a las diferencias alélicas más elevadas entre pares de bancos, fue posible obtener un set de marcadores altamente diferenciados (199 *loci outlier*), esta característica los hace prometedores para futuros análisis de asignación de origen. De manera interesante, en este

trabajo, al calcular  $F_{ST}$  para éstos *loci outlier*, se observó un patrón de parcheo genético caótico, probablemente relacionado con el número de cohortes que se encontró en los bancos.

La conectividad oceanográfica se ve reflejada en la conectividad genética a nivel de población, pero hasta el momento no es capaz de evaluar el flujo de larvas entre bancos, donde pueden estar influyendo otros factores (ambientales, éxitos reproductivos, dinámicas poblacionales) que permitan diferencialmente el establecimiento de las larvas dentro de los bancos.

## 10. LITERATURA CITADA

- Allendorf, F. W., P. A. Hohenlohe, G. Luikart. 2010. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* 11(10):697–709.
- Antao, T., A. Lopes, R.J. Lopes, A. Beja-Pereira, G. Luikart. 2008. Lositan: A workbench to detect molecular adaptation based on a FST-*outlier* method. *BMC Bioinformatics*, 9(1): 323-327.
- Aragón-Noriega, E. A., J. Chávez-Villalba, P. E. Gribben, E. Alcántara-Razo, A. N. Maeda-Martínez, E. M. Arambula-Pujol, A. R. García-Juárez, R. Maldonado-Amparo. 2007. Morphometric relationships, gametogenic development and spawning of the geoduck clam *Panopea globosa* (Bivalvia: Hiatellidae) in the Central Gulf of California. *J. Shellfish Res.* 26:423-431
- Arambula-Pujol, M., A. R. Garcia-Juarez, E. Alcantara-Razo, E. A. Aragon-Noriega. 2008. Aspectos de biología reproductiva de la almeja de sifón *Panopea globosa* (Dall 1898) en el Golfo de California. *Hidrobiologica.* 18:89-98.
- Araneda, C., M. A. Larraín, B. Hecht, S. Narum. 2016. Adaptive genetic variation distinguishes Chilean blue mussels (*Mytilus chilensis*) from different marine environments. *Ecol Evol*, 6(11):3632–3644.
- Baird N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376.
- Beaumont, M.A., R. A. Nichols. 1996. Evaluating *loci* for use in the genetic analysis of population structure. *P. Roy. Soc. B-Biol Sci.* 263:1619-1626.
- Beier, E. 1997. A numerical investigation of the annual variability in the Gulf of California. *J. Phys. Oceanogr.* 27:615–632.
- Bradbury, A., B. Sizemore, D. Rothaus, M. Ulrich. 2000. Stock Assessment of Subtidal Clams (*Panopea abrupta*) in Washington. Marine Resours Unit. Fish Managements Division. Fish Program. USA.
- Bradbury, I. R., S. Hubert, B. Higgins, S. Bowman, T. Borza, I. G. Paterson, P. Bentzen. 2013. Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evol. Appl.* 6(3):450–461.
- Campbell A., R. M. Harbo, C. M. Hand. 1998. Harvesting and distribution of Pacific geoducks clams, *Panopea abrupta*, in British Columbia. Proceeding of the North Pacific Symposium on Invertebrate Stock Assesment and Management. Editado por G. S.

- Jamienson y A. Campbell. *Can. Spec. Fish. Aquat. Sci.* 125:349–358.
- Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, J. H. Postlethwait. 2011. STACKS: building and genotyping *loci* de novo from short-read sequences. *G3-Genes Genom. Genet.* 1(3):171–82.
- Catchen, J., S. Bassham, T. Wilson, M. Currey, C. O'Brien, Q. Yeates, W. A. Cresko. 2013a. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Mol. Ecol.* 22(11):2864–2883.
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, W. A. Cresko. 2013b. STACKS: An analysis tool set for population genomics. *Mol Ecol*, 22:3124-40.
- Chu, N. D., S. T. Kaluziak, G. C. Trussell, S. V. Vollmer. 2014. Phylogenomic analyses reveal latitudinal population structure and polymorphisms in heat stress genes in the North Atlantic snail *Nucella lapillus*. *Mol. Ecol.*, 23(7):1863–73.
- Cruz-Hernández, P., A. Munguia-Vega, I. Leyva-Valencia, L. Lucero-Burquez, D. B. Lluch-Cota. 2014. Development of 24 tetra-nucleotide microsatellite markers in Cortes Geoduck *Panopea globosa* by next-generation sequencing. *Conserv. Genet. Resour.* 6:531-533
- De Wit, P., M. H. Pespeni, J. T. Ladner, D. J. Barshis, F. Seneca, H. Jaris, N. O. Therkildsen, M. Morikawa, S. R. Palumbi. 2012. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* 12(6):1058–67.
- Diario Oficial de la Federación. 2012. Carta Nacional Pesquera. Secretaría de Agricultura, Ganadería, desarrollo Rural, Pesca y Alimentación. México. 24 de agosto de 2012.
- Emerson, K. J., C. R. Merz, J. M. Catchen, P. A. Hohenlohe, W. A. Cresko, W. E. Bradshaw, C. M. Holzapel. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *P. Natl. A. Sci.* 107:16196.
- Feder, M. E., T. Mitchell-Olds. 2003. Evolutionary and ecological functional genomics. *Nat. Rev. Genet.* 4:649–655.
- Gayral, P., J. Melo-Ferreira, S. Glémin, N. Bierne, M. Carneiro, B. Nabholz, J. M. Lourenco, P. C. Alves, M. Ballenghien, N. Faivre, K. Belkhir, V. Cahais, E. Loire, A. Bernard, N. Galtier. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* 9(4):e1003457.
- González-Peláez, S. S. 2013. Almeja Generosa de Bahía Magdalena: Dinámica Poblacional y consideraciones para el Manejo. Tesis (Doctorado Uso Manejo Y Preservación de los

- Recursos Naturales). La Paz, B.C.S. Centro de Investigaciones Biológicas del Noroeste. 103 p.
- González-Peláez, S. S., I. Leyva-Valencia, S. Pérez-Valencia, D. B. Lluch-Cota. 2013. Distribution Limits of the Geoduck Clams *Panopea generosa* and *P. globosa* on the Pacific Coast of Mexico. *Malacologia*. 56:85-94.
- Goodwin, C. L. 1976. Observations on spawning and growth of subtidal geoduck (*Panopea generosa*, Gould). *Proc. Natl. Shellfisheries Assoc.* 65: 49-58.
- Goodwin, C. L., B. Pease. 1989. Pacific Geoduck Clam. Species Profiles: Life Histories and Environmental Requirements of Coastal Fish and Invertebrates (Pacific Northwest). *Biological Report*. 82:11-120.
- Graham, C. F., T. C. Glenn, A. G. McArthur, D. R. Boreham, T. Kieran, S. Lance, R. G. Manzon, J. A. Martino, T. Pierson, S. M. Rogers, J. Y. Wilson, C. M. Somers. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Mol. Ecol. Resour.* 15(6):1304–1315.
- Guo, B., J. DeFaveri, G. Sotelo, A. Nair, J. Merilä. 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol.* 13(1):19.
- Hohenlohe, P. A., S. J. Amish, J. M. Catchen, F. W. Allendorf, G. Luikart. 2011. Nextgeneration RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol. Ecol. Resour.* 11:117–122.
- Kenny, E. M., P. Cormican, W. P. Gilks, A. S. Gates, C. T. O’Dushlaine, C. Pinto, A. P. Corvin, M. Gill, D. W. Morris. 2011. Multiplex target enrichment using DNA indexing for Ultra-High throughput SNP detection. *DNA Res.* 18(1):31–38.
- Lal, M. M., P. C. Southgate, D. R. Jerry, K. R. Zenger. 2016. Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *Mar. Genom.* 25:57–68.
- Lapègue, S., E. Harrang, S. Heurtebise, E. Flahauw, C. Donnadieu, P. Gayral, M. Ballenghien, L. Genestout, L. Barbotte, R. Mahla, P. Haffray, C. Klopp. 2014. Development of SNP-genotyping arrays in two shellfish species. *Mol. Ecol. Resour.* 14(4):820–30.
- Lavín, M. F., R. Durazo, E. Palacios, M. L. Argote, L. Carrillo. 1997. Lagrangian observations of the circulation in the northern Gulf of California. *J. Phys. Oceanogr.* 27:2298–2305.



- Leyva-Valencia, I., S. T. Álvarez-Castañeda, D. B. Lluch-Cota, S. S. González-Peláez, S. Pérez-Valencia, B. Vadopalas, S. Ramírez-Pérez, P. Cruz-Hernández. 2012. Shell shape differences between two panopea species and phenotypic variation among *P. globosa* at different sites using two geometric morphometrics approaches. *Malacologia*. 55(1):1-13.
- Li, H., R. Durbin. 2009a. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14):1754-1760.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin. 2009b. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078-2079.
- Li, Y. H., H. P. Wang. 2017. Advances of genotyping-by-sequencing in fisheries and aquaculture. *Rev. Fish Biol. Fisher.* 27(3):535-559
- Lopes-Lima, M., L. E. Burlakova, A. Y. Karatayev, K. Mehler, M. Seddon, R. Sousa. 2018. Conservation of freshwater bivalves at the global scale: diversity, threats and research needs. *Hydrobiologia*. 810(1):1-14.
- Luikart, G., P. R. England, D. Tallmon, S. Jordan, P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4(12):981–94.
- Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. H. Anholt, M. Barrón, C. Bess, K. P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L. Perales, L. Pu, C. Qu, M. Ràmia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley, Y. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman, R. A. Gibbs . 2012. The *Drosophila melanogaster* genetic reference panel. *Nature*. 482:173–178
- Magoč, T., S. L. Salzberg. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 27(21):2957-2963.
- Mariani, S., D. Bekkevold. 2013. The Nuclear Genome: Neutral and Adaptive Markers in Fisheries Science. En: Cadrin, S. X., L. A. Kerr, S. Mariani (eds.). *Stock Identification Methods: Applications in Fishery Science*. Academic Press. Segunda Edición. pp. 297–327.
- Marinone, S. G., M. J. Ulloa, A. Parés-Sierra, M. F. Lavín, R. Cudney-Bueno. 2008. Connectivity in the northern Gulf of California from particle tracking in a three-dimensional numerical model. *J. Marine Syst.* 71(1–2):149–158.

- Marinone, S. G. 2012. Seasonal surface connectivity in the Gulf of California. *Estuar. Coast. Shelf S.* 100:133-141.
- Miller, K. M., K. J. Supernault, S. Li, R. E. Withler. 2006. Population structure in two marine invertebrate species (*Panopea abrupta* and *Strongylocentrotus franczscanus*) targeted for aquaculture and enhancement in British Columbia. *J. Shellfish Res.* 25:33-42.
- Munguia-Vega, A., A. Jackson, S. G. Marinone, B. Erisman, M. Moreno-Baez, A. Girón-Nava, T. Pfister, O. Aburto-Oropeza, J. Torre. 2014. Asymmetric connectivity of spawning aggregations of a commercially important marine fish using a multidisciplinary approach. *PeerJ.* 2:e511.
- Munguia-Vega, A., I. Leyva-Valencia, D. B. Lluch-Cota, P. Cruz-Hernandez. 2015. Genetic Structure of the Cortes Geoduck *Panopea globosa* Dall, 1898, from the Mexican Northwest. *J. Shellfish Res.* 34(1):153–161.
- Orensanz, J. M., R. Hilborn y A. M. Parma. 2000. Harvesting Methuselah’s clams - is the geoduck fishery sustainable, or just apparently so? *Can. Stock Assess. Secr. Res. Doc.* No. 2000/175.
- Palacios-Hernández, E., E. Beier, M. F. Lavín, P. Ripa. 2002. The effect of the seasonal variation of stratification on the circulation on the northern Gulf of California. *J. Phys. Oceanogr.* 32:705–728.
- Peguero-Icaza, M., L. Sánchez-Velasco, M. F. Lavín, S. G. Marinone, E. Beier. 2011. Seasonal changes in connectivity routes among larval fish assemblages in a semi-enclosed sea (Gulf of California). *J. Plankton Res.* 33(3):517–533.
- Perea, E. 2008. Almeja “chiluda”, generoso negocio. *La imagen Agropecuaria.* 30 de marzo 2008. <http://www.imagenagropecuaria.com>
- Peterson, B. B. K., J. N. J. Weber, E. H. E. Kay, H. S. Fisher, H. E. Hoekstra. 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.
- Quail, M. A., I. Kozarewa, F. Smith, A. Scally, P. J. Stephens, R. Durbin, H. Swerdlow, D. J. Turner, 2008. A large genome center’s improvements to the Illumina sequencing system. *Nat. Methods.* 5(12):1005–1010.
- Rašić, G., I. Filipović, A. R. Weeks, A. A. Hoffmann. 2014. Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics.* 15(1):275-286.
- Reusch T.B.H, T. E. Wood. 2007. Molecular ecology of global change. *Mol. Ecol.* 16:3973-

3992.

- Ribeiro, Â. M., A. D. Foote, A. Kupczok, B. Frazão, M. T. Limborg, R. Piñeiro, S. Abalde, S. Rocha, R. R. da Fonseca. 2017. Marine genomics: News and views. *Mar. Genom.* 31:1-8
- Richards, P. M., M. M. Liu, N. Lowe, J. W. Davey, M. L. Blaxter, A. Davison. 2013. RAD-Seq derived markers flank the shell colour and banding *loci* of the *Cepaea nemoralis* supergene. *Mol. Ecol.* 22(11):3077–3089.
- Robledo, D., C. Palaiokostas, L. Bargelloni, P. Martínez, R. Houston. 2017. Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquacult.* 0:1-13
- Rocha-Olivares, A., L. E. Calderon-Aguilera, E. A. Aragón-Norieg, N. C. Saavedra-Sotelo, V. M. Moreno-Rivera. 2010. Genetic and Morphological Variation of Northeast Pacific *Panopea* Clams: Evolutionary Implications. *J. Shellfish Res.* 29:327-335.
- Rousset, F. 2008. Genepop'007: A complete re-implementation of the genepop software for windows and linux. *Mol. Ecol. Res.* 8:103-106.
- Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E. H. Byrne, S. A. McCarroll, R. Gaudet, S. F. Schaffner, E. S. Lander, The International HapMap Consortium. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.
- Sandoval-Castillo, J., N. A. Robinson, A. M. Hart, L. W. S. Strain, L. B. Beheregaray. 2018. Seascape Genomics Reveals Adaptive Divergence in a Connected and Commercially Important Mollusc, the Greenlip Abalone (*Haliotis Laevis*), along a Longitudinal Environmental Gradient. *Mol. Ecol.* 27:1603– 1620.
- Sloan, N. A., S. M. C. Robinson. 1984. Age and gonad development in the geoduck clam *Panope abrupta* (Conrad) from southern British Columbia, Canada. *J. Shell. Res.* 4 (2):131-137.
- Sovic, M. G., A. C. Fries, H. L. Gibbs. 2015. AFRAD: a pipeline for accurate and efficient de novo assembly of RADseq data. *Mol. Ecol. Resour.* 15(5):1163-1171.
- Suárez-Moo, P. J., L. E. Calderon-Aguilera, H. Reyes-Bonilla, G. Díaz-Erales, V. Castañeda-Fernandez-de-Lara, E. A. Aragón-Noriega, A. Rocha-Olivares. 2013. Integrating genetic, phenotypic and ecological analyses to assess the variation and clarify the distribution of the Cortes geoduck (*Panopea globosa*). *J. Mar. Biol. Assoc. Uk.* 93(03):809–816.
- Takeuchi, T. 2017. Molluscan genomics: implications for biology and aquaculture. *Curr. Mol. Biol. Rep.* 3(4):297–305.

- Thórarinsdóttir, G. G. 1999. Lifespan of two long-lived bivalves, *Arctica islandica* and *Panopea generosa*. Spec. Publ. Phuket Mar. Biol. Cent. 19(1): 41- 46.
- Vadopalas, B., L. L. LeClair, P. Bentzen. 2004. Microsatellite and allozyme analyses reveal few genetic differences among spatially distinct aggregations of geoduck clams (*Panopea abrupta*, Conrad 1849). J. Shellfish Res. 23:693-706.
- Van Wyngaarden, M., P. V. R. Snelgrove, C. DiBacco, L. C. Hamilton, N. Rodríguez-Ezpeleta, N. W. Jeffery, R. R. E. Stanley, I. R. Bradbury. 2016. Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus*, using RAD-seq derived SNPs. Evol. Appl. 10:1–16.
- Wang, J., H. Qi, L. Li, H. Que, D. Wang, G. Zhang. 2014. Discovery and validation of genic single nucleotide polymorphisms in the Pacific oyster *Crassostrea gigas*. Mol. Ecol. Resour. 15:1–13.
- Wenne, R., C. Tsigenopoulos. 2018. Single nucleotide polymorphism markers with applications in aquaculture and assessment of its impact on natural populations. Aquat. Living Resour. 31:2.
- Willing, E.-M., M. Hoffmann, J. D. Klein, D. Weigel, C. Dreyer. 2011. Paired-end RAD-seq for de novo assembly and marker design without available reference. Bioinformatics. 27(16):2187–93.
- Zamudio, L., P. Hogan, E. J. Metzger. 2008. Summer generation of the southern Gulf of California eddy train. J. Geophys. Res. 113:C06020.
- Zhang, G., X. Fang, X. Guo, L. Li, R. Luo, F. Xu, P. Yang, L. Zhang, X. Wang, H. Qi, Z. Xiong, H. Que, Y. Xie, P. W. H. Holland, J. Paps, Y. Zhu, F. Wu, Y. Chen, J. Wang, C. Peng, J. Meng, L. Yang, J. Liu, B. Wen, N. Zhang, Z. Huang, Q. Zhu, Y. Feng, A. Mount, D. Hedgecock, Z. Xu, Y. Liu, T. Domazet-Lošo, Y. Du, X. Sun, S. Zhang, B. Liu, P. Cheng, X. Jiang, J. Li, D. Fan, W. Wang, W. Fu, T. Wang, B. Wang, J. Zhang, Z. Peng, Y. Li, N. Li, J. Wang, M. Chen, Y. He, F. Tan, X. Song, Q. Zheng, R. Huang, H. Yang, X. Du, L. Chen, M. Yang, P. M. Gaffney, S. Wang, L. Luo, Z. She, Y. Ming, W. Huang, S. Zhang, B. Huang, Y. Zhang, T. Qu, P. Ni, G. Miao, J. Wang, Q. Wang, C. E. W. Steinberg, Ha. Wang, N. Li, L. Qian, G. Zhang, Y. Li, H. Yang, X. Liu, J. Wang, Y. Yin, Z. Huang. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 490(7418):49–54.
- Zhang, L., X. Guo. 2010. Development and validation of single nucleotide polymorphism markers in the eastern oyster *Crassostrea virginica* Gmelin by mining ESTs and resequencing. Aquaculture, 302(1-2):124–129.

## 11. ANEXOS

### Anexo A. Concentración en las muestras de ADN genómico.

Banco	Individuo	Concentración (ng/ $\mu$ L)		
		ADNg precipitado	Post-digestión	Ligación de adaptadores
San Felipe Banco 1	SF21	106.4	4.9	0.306
	SF22	61.7	4.6	0.836
	SF23	285.1	6.3	0.544
	SF24	173.4	6.9	0.54
	SF25	178.4	9.9	0.3805
	SF26	271.7	4.7	0.764
	SF27	281.3	10.5	0.494
	SF28	102.1	5	0.664
	SF29	184.3	6.1	0.462
	SF30	439.1	7.1	0.82
	SF31	378.3	7.1	0.446
	SF32	108.2	7.2	0.46
	SF33	293.1	5.8	0.484
	SF34	194.4	16	0.3895
	SF35	199.8	5.2	0.672
	SF36	284.2	4.9	0.682
	SF37	260.9	4.5	0.766
	SF38	232.3	7.4	0.726
	SF39	855	11.5	0.313
	SF40	242	8.3	0.722
SF41	191.3	5.1	0.772	
SF42	308.4	6.5	0.68	
SF43	158.7	6.1	0.942	
San Felipe Banco 2	SF64	173.3	7.1	0.54
	SF65	172.9	6	0.864
	SF66	161.3	16.1	0.339
	SF67	180.8	3.8	0.534
	SF68	140.8	3	0.772
	SF69	132	6.3	0.353
	SF70	232.3	9.6	0.1775
	SF71	114.7	3.8	0.636
	SF72	113.9	6.4	0.442
	SF73	175.7	5.7	0.358
	SF74	154	5.6	0.281
	SF75	62.5	2.9	0.464

Banco	Individuo	Concentración (ng/ $\mu$ L)		
		ADNg precipitado	Post-digestión	Ligación de adaptadores
	SF76	82.5	4	0.618
	SF77	76.9	9.2	0.2145
	SF78	66	5.9	0.4065
	SF79	113	5.8	0.524
	SF80	72.7	4.2	0.522
	SF81	156.2	6.8	0.48
	SF82	157.2	11	0.355
	SF83	108.5	3.5	0.786
	SF84	119.7	8.1	0.634
	SF85	82.1	7.3	0.526
	SF86	129.8	6.3	0.714
	SF87	139.8	8	0.65
	SF88	120.2	7.7	0.422
	SF89	109.8	7.6	0.554
	SF90	132.1	5.4	0.648
	SF91	80.5	4.8	0.434
	SF92	39.9	4.2	0.546
	SF93	55.8	6.5	0.48
	SF94	76.2	4.5	0.373
	SF95	147	4.1	0.3415
	PP41	166.3	7.8	0.544
	PP42	171.4	9.2	0.59
	PP43	119.8	6.1	0.502
	PP44	121.8	6.2	0.688
	PP45	82.1	8.3	0.476
	PP46	206.9	9.4	0.522
	PP47	271.7	6.4	0.546
	PP48	354.5	8.6	0.488
	PP49	197.2	10.9	0.412
Puerto Peñasco Banco 1	PP50	296.2	8.6	0.706
	PP51	249.7	7.8	0.29
	PP52	162.9	7.7	0.78
	PP53	138.3	7.3	0.89
	PP54	257.7	6.9	0.371
	PP55	110.4	7.3	0.532
	PP56	138.3	13.9	0.438
	PP57	108.7	5	0.844
	PP58	226.5	8.3	0.89
	PP59	220.6	7.6	0.608
	PP60	98.1	6.3	0.594

Banco	Individuo	Concentración (ng/ $\mu$ L)		
		ADNg precipitado	Post-digestión	Ligación de adaptadores
	PP61	153.3	6.8	0.874
	PP62	122.1	8.3	0.762
	PP63	131.1	4.7	0.948
	PP64	98.1	9.2	0.782
	PP65	152.9	12.2	0.48
	PP66	161.2	9	0.341
	PP67	181.4	6.4	0.768
	PP68	115.2	7.6	0.512
	PP69	154.5	7.2	0.836
	PP70	122.4	7.6	0.662
	PP71	133.8	8.4	0.542

**Anexo B. Selección de enzimas de restricción.**

Tipos de combinaciones	Enzima <i>Forward</i>	Enzima <i>Reverse</i>	454-Seq	Transcriptoma	Genoma
			<i>P. globosa</i>	<i>P. generosa</i>	<i>C. gigas</i>
			<i>N</i>	<i>N</i>	<i>N</i>
	NlaIII	MluCI	106,072	122,265	673,734
Adaptadores FLEX	NlaIII	EcoRI	4,745	6,985	42,515
	SphI	MluCI	6,397	6,649	46,393
	SphI	EcoRI	190	266	2,122
Adaptadores específicos	MspI	SbfI	17	85	205
Combinados	EcoRI	MspI	880	1,431	8,839