

**CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN  
SUPERIOR DE ENSENADA, BAJA CALIFORNIA**



---

**PROGRAMA DE POSGRADO EN CIENCIAS  
EN ECOLOGÍA MARINA**

---

**Secuenciación genómica masiva de *Panopea generosa* y  
*Panopea globosa* para el desarrollo de marcadores  
moleculares**

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de  
Maestro en Ciencias

Presenta:

**Celia Isabel Bisbal Pardo**

Ensenada, Baja California, México,

2015

Tesis defendida por  
y aprobada por el siguiente Comité

---

Dr. Axayácatl Rocha Olivares  
Director del Comité

---

Dr. Luis Eduardo Calderón Aguilera  
Miembro del Comité

---

Dr. Miguel Ángel del Río Portilla  
Miembro del Comité

---

Dr. David Alberto Rivas Camargo  
Coordinador Programa de Posgrado en  
Ecología Marina

---

Dr. Jesús Favela Vargas  
Director de Estudios de Posgrado

Febrero, 2015

Resumen de la tesis que presenta Celia Isabel Bisbal Pardo como requisito parcial para la obtención del grado de Maestro en Ciencias en Ecología Marina.

## **Secuenciación genómica masiva de *Panopea generosa* y *Panopea globosa* para el desarrollo de marcadores moleculares**

Resumen elaborado por:

---

Celia Isabel Bisbal Pardo

Los bivalvos del género *Panopea*, conocidos como almejas de sifón o geoduck, se caracterizan por su enorme tamaño y gran longevidad. Varias especies de esta almeja han adquirido un importante valor comercial, entre las que se encuentran *Panopea generosa* y *P. globosa*, que se distribuyen en el Noroeste de México. Recientemente se ha despertado el interés por la biología, ecología y evolución de estas especies, para lo cual se han aplicado aproximaciones moleculares. Sin embargo, los marcadores moleculares disponibles no han sido suficientemente exitosos para esclarecer varias cuestiones sobre la ecología y dinámica poblacional de dichas especies a lo largo de su distribución geográfica. Estos aspectos darán información muy valiosa para determinar planes de manejo específicos, incluso a nivel poblacional. En este estudio se desarrolló una batería de marcadores moleculares microsatelitales polimórficos usando secuenciación de siguiente generación (plataforma Illumina) y aplicando análisis bioinformáticos. Se obtuvieron secuencias de buena calidad con las que se ensamblaron 868,521 contigs en *P. generosa* y 737,851 contigs en *P. globosa*. Se identificaron 30 loci microsatelitales (di-, tri- y tetranucleótidos) candidatos a marcadores moleculares por especie, de los que experimentalmente se asilaron satisfactoriamente 8 loci polimórficos en *P. generosa* y 9 en *P. globosa*. El número de alelos por locus varió de 3 a 22 y los valores de heterocigosidad observada ( $H_o$ ) y esperada ( $H_e$ ) van de 0.286 a 0.818 y de 0.504 a 0.962, respectivamente. Además se obtuvo la anotación mitocondrial completa. Los genomas mitocondriales de ambas especies fueron muy similares, pero difieren de los de otras especies de la subclase Heterodonta. Estos nuevos hallazgos sumados a los recursos genéticos existentes servirán para profundizar en la ecología y biología de estas especies permitiendo el desarrollo de planes de manejo apropiados para asegurar la conservación de las mismas.

Palabras claves: ***Panopea generosa*, *Panopea globosa*, Secuenciación de siguiente generación, microsátélites, anotación.**

Abstract of the thesis presented by Celia Isabel Bisbal Pardo as a partial requirement to obtain the Master of Science degree in Marine Ecology.

**Massive genomic DNA sequencing of the species *Panopea generosa* and *Panopea globosa* for molecular marker development.**

Abstract by:

---

Celia Isabel Bisbal Pardo

Bivalves in the genus *Panopea*, also known as siphon clams or geoduck, are characterized by their large size and longevity. Several species have gained a significant commercial value, such as *Panopea generosa* and *P. globosa*, which are distributed in northwestern México. Recently, interest in their biology, ecology and evolution has increased, for which molecular techniques have been useful. However, available molecular markers have not been sufficient to clarify several issues regarding the ecology and population dynamics of these species. This knowledge is necessary to establish management plans at species level and population levels. In this study, polymorphic microsatellite molecular markers were obtained using next-generation sequencing (Illumina platform) with the help of bioinformatics tools. Good quality reads were used to build an assembly of 868,521 and 737,851 contigs of *P.generosa* and *P.globosa*, respectively. Thirty microsatellite loci (di-, tri- and tetranucleotide) per species were tested and only 8 and 9 were successfully amplified and confirmed as polymorphic markers. The number of alleles per locus ranged from 3 to 22 and the observed and expected heterozygosities ranged from 0.286 to 0.818 and 0.504 to 0.962, respectively. In addition, the complete mitochondrial genome was obtained from each species. These mitogenomes were very similar but different from other available mitochondrial genomes from other clams of the Heterodonta subclass. These findings complement the existing genetic resources and will be useful to advance our knowledge in the ecology and biology of these species, which will be of consequence to develop appropriate management and conservation plans for both species.

Keywords: ***Panopea generosa*, *Panopera globosa*, next generation sequencing, microsatellite, annotation.**

## **Dedicatoria**

A México y a su gente,  
porque son muchas las riquezas naturales que alberga este país y de todos depende  
que perduren...

## **Agradecimientos**

A mi director de tesis, el Dr. Axayáctli Rocha Olivares, por aceptarme como una de sus estudiantes y confiar en mi desempeño. Así como por guiarme en este proyecto y estimular mi desarrollo profesional.

Al Dr. Luis Eduardo Calderón Aguilera por mostrarme todo su apoyo y disponibilidad infinita así como por añadir siempre un toque de humor en el día a día.

Al Dr. Miguel Ángel del Río Portilla por todas sus asesorías matutinas y entera disposición para este proyecto.

Al Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) por admitirme como estudiante en el posgrado de Ecología Marina. Y al mismo posgrado por sus enseñanzas en las clases y seminarios.

Al CONACYT por concederme la beca (CVU 501574) para realizar mis estudios de maestría.

Al Dr. Luis Manuel Enriquez Paredes de la UABC por las asesorías para el análisis de microsatelites.

A todos los profesores del posgrado que aportaron nuevas visiones para completar mi formación.

A todo el personal de CICESE que desde la caseta a telemática han formado parte de esto.

A mis amigos y compañeros del Laboratorio de Ecología Molecular por estar ahí para todas mis dudas, que no fueron pocas, pero sobretodo por crear un ambiente de trabajo en equipo y muchas risas. En especial a Anita, por acompañarme a lo largo de toda la tesis, y compensar con tu energía mis días de desesperación. A Geo, por tu gran disposición a ayudar y por compartir las salidas nocturnas del laboratorio. A Francesco,

por hacerme pensar mucho sobre mi tesis y crearme dilemas continuos. A José, por transmitir tu tranquilidad innata y, junto con tu esposa Dani, ofrecer siempre tu casa para cualquier evento. A Bea, por tu ejemplo como estudiante modelo y demostrar que hay tiempo para hacer muchas actividades. A los técnicos Alejandro e Ivonne, por preocuparse de todo lo que necesitaba el laboratorio y estar al pendiente de que trabajáramos en buenas condiciones. A los que ya se fueron porque quedaron en mi recuerdo sus enseñanzas y buenos ratos compartidos, Pablo, Nancy, Erick y Marianita. Y al recién llegado, Pepe, le deseo mucha suerte.

Gracias a Asier, Anto, Pau, Pablo (fabas), Tati, Lu, Ruben, por cuidarme y tratarme como una hermana, al menos así lo sentí yo =). Gracias a toda la buena vibra recibida por los compis más guays de Cicese: Aide, Andrea, Fer, Ceci, Elena, Edgar, el Yama, Laurita, Laurica, Anita, Alma, Vero, las Karen, Rian, Rafa, los Gera...y a las más peques: Amparo, Bárbara y Magda. A los que me ayudaron a desconectar del Cicese: Cris, Alex, Ian, Luigi, Trucco, Andrés, Elisú, Raquel, Malú. Y a otros que no quiero olvidar: Emi, Luz, Tere, Andrea, Sarita, Alejandro... A David, muchísimas gracias por no salir corriendo mis días de super estrés y comunicarnos por “telequinesis” cuando era necesario. Pero sobre todo por cuidarme mucho muchísimo y apostar por el futuro.

Gracias a mis amig@s españoles, Maite, Manu, Evita, María, Selene, Julia, Vane, Abel, las Maripavis, muchos repartidos por el mundo, por seguir al tanto de todos mis pasos, apoyándome en cada momento y no permitir que la amistad se pierda.

A mis papas, por casi obligarme a vivir esta oportunidad y preocuparse de mí cada día para no sentir la distancia. A Oti y a Josep, gracias por venir de tan lejos para verme, y a Mary, porque sé que está deseando venir y traerse a Olmito. A ellos y a toda la familia, por seguir ahí sin entender muy bien lo que hago xD

# Tabla de contenido

	Página
<b>Resumen español</b> .....	ii
<b>Resumen inglés</b> .....	iii
<b>Dedicatoria</b> .....	iv
<b>Agradecimientos</b> .....	v
<b>Lista de Figuras</b> .....	ix
<b>Lista de Tablas</b> .....	x
<b>1. Introducción</b> .....	1
<b>1.1. Introducción general</b> .....	1
1.1.1 El género <i>Panopea</i> .....	1
1.1.2. Pesquería de <i>Panopea</i> en México.....	3
1.1.3. Marcadores moleculares.....	5
1.1.4. Secuenciación genómica masiva.....	7
1.1.5. Anotación genómica.....	10
<b>1.2. Antecedentes</b> .....	12
1.2.1. Estudios de <i>Panopea generosa</i> y <i>P. globosa</i> con diferentes tipos de marcadores moleculares.....	12
1.2.2. Desarrollo de marcadores microsatelitales mediante plataformas de secuenciación de siguiente generación.....	14
<b>1.3. Justificación</b> .....	16
<b>1.4. Hipótesis</b> .....	17
<b>1.5. Objetivos</b> .....	17
1.5.1. Objetivo general.....	17
1.5.2. Objetivos específicos.....	17
<b>2. Materiales y métodos</b> .....	18
<b>2.1. Extracción de ADN</b> .....	18
<b>2.2. Secuenciación masiva</b> .....	18
<b>2.3. Bioinformática</b> .....	19
2.3.1. Análisis de secuencias.....	19
2.3.2. Identificación y selección de loci microsatelitales candidatos a marcadores moleculares.....	20
<b>2.4. Pruebas experimentales</b> .....	21
2.4.1. Optimización de las condiciones de amplificación de los loci microsatelitales candidatos a marcadores moleculares.....	21
2.4.2. Caracterización.....	22
<b>2.5. Anotación</b> .....	22
2.5.1. Anotación general de los contigs.....	22
2.5.2. Anotación mitocondrial.....	23



## Tabla de Contenido (Continuación)

	Página
<b>3. Resultados</b> .....	24
<b>3.1. Bioinformática</b> .....	24
3.1.1. Análisis de secuencias .....	24
3.1.2. Identificación y selección de loci microsatelitales candidatos a marcadores moleculares .....	26
<b>3.2. Pruebas experimentales</b> .....	27
<b>3.3. Anotación</b> .....	33
3.3.1. Anotación general de los contigs.....	33
3.3.2. Anotación mitocondrial .....	34
<b>4. Discusión</b> .....	36
<b>4.1. Análisis de las secuencias de ambas especies provenientes de la NGS</b> .....	36
<b>4.2. Identificación de loci microsatelitales y selección de los loci candidatos a marcadores moleculares</b> .....	37
4.2.1. Elección de la plataforma de NGS .....	37
4.2.2. Selección de los loci candidatos a marcadores moleculares.....	39
<b>4.3. Optimización de las condiciones de amplificación por PCR de los loci microsatelitales candidatos y evaluación de su nivel de polimorfismo</b> .....	40
<b>4.4. Anotación general de los contigs</b> .....	43
<b>4.5. Anotación mitocondrial</b> .....	44
<b>5. Resultados significativos</b> .....	46
<b>Lista de referencias</b> .....	47

## Lista de Figuras

Figura		Página
1	Representación de <i>P. globosa</i> (Dibujo de Eduardo Vázquez Solórzano).....	1
2	Distribución mundial y regiones específicas de distribución en México de <i>Panopea generosa</i> (línea roja) y <i>Panopea globosa</i> (línea morada). EP-N (Región Norte del Pacífico Oriental, EP-S (Región Sur del Pacífico Oriental), UGC-W (Región Oeste del Alto Golfo de California), UGC-E (Región Este del Alto Golfo de California) y CGC (Región Central del Golfo de California). Tomado y editado de Aragón Noriega <i>et al.</i> (2012).....	3
3	Captura comercial del género <i>Panopea</i> en el Pacífico Nororiental. Desembarco de <i>Panopea generosa</i> en Canadá (Columbia Británica) y Estados Unidos de America (Washington); <i>P. generosa</i> y <i>P. globosa</i> en México. Tomado de Aragón Noriega <i>et al.</i> (2012), p. 735.....	4
4	Cambios relativos en la popularidad de los diferentes tipos de marcadores moleculares (Tomada de Schlötterer, 2004)...	6
5	Evolución del coste de la secuenciación de ADN por megabase desde el año 2001 al 2014.....	8
6	Esquema general de la secuenciación de siguiente generación para las cuatro plataformas disponibles (Loman & Constantinidou, 2012).....	9
7	Representación gráfica de la secuenciación pair-end 2x100.....	18
8	Anotación genética de los contigs totales de <i>P. generosa</i> y <i>P. globosa</i> . Bases de datos: GO: Ontología genética; DOM: Dominio y ENZ: Enzima.....	34
9	Mitogenoma de <i>Panopea generosa</i> .....	35
10	Mitogenoma de <i>Panopea globosa</i> .....	35

## Lista de Tablas

Tabla		Página
1	Valor de calidad Phred y precisión de la secuenciación.....	19
2	Cantidad de secuencias que se obtuvieron en la secuenciación masiva de cada organismo y el total de secuencias que se obtuvo para cada especie. Pgen= <i>Panopea generosa</i> , Pglo = <i>P. globosa</i> , RB =Rincón de Ballenas (Ensenada), SF = San Felipe, BM = Bahía Magdalena.....	24
3	Porcentaje de secuencias no modificadas, modificadas y descartadas en el proceso de limpieza y recorte (control de calidad).....	25
4	Porcentaje total de secuencias ensambladas y porcentaje específico de las secuencias que se ensamblaron pareadas y las no pareadas .....	25
5	Número de loci di-, tri- y tetranucleótidos que se obtuvieron para cada especie de <i>Panopea</i> y el porcentaje de éstos a los que se les pudo diseñar cebadores.....	26
6	Perfil de termociclado de PCR usado para la amplificación de los loci ensayados.....	27
7	Descripción de los cebadores de los loci descartados y causa del descarte en <i>P. generosa</i> .....	28
8	Descripción de los cebadores de los loci que pasaron a la 2ª fase experimental y el resultado de la misma en <i>P. generosa</i> .....	30
9	Descripción de los cebadores de los loci que pasaron a la 2ª fase experimental y el resultado de la misma en <i>P. globosa</i> .....	31
10	Características de los 8 loci polimórficos de <i>P.generosa</i> ....	32
11	Características de los 9 loci polimórficos de <i>P.globosa</i> .....	33
12	Comparación de los resultados obtenidos en dos plataformas diferentes de NGS, 454 e Illumina (IPE-119K), para la especie <i>Python molurus bivittatus</i> . *PALs: potentially amplifiable microsatellite loci.....	38

# 1. Introducción

## 1.1. Introducción general

### 1.1.1 El género *Panopea*

*Panopea* (Ménard, 1807) es un género de bivalvos marinos de la familia Hiatellidae, que engloba a almejas de gran tamaño y muy longevas, conocidas vulgarmente como chiludas o almejas de sifón (Fig. 1). Actualmente este género cuenta con 9 especies existentes: *P. abbreviata* (Valenciennes, 1839), *P. australis* (G. B. Sowerby I, 1833), *P. bitruncata* (Conrad, 1872), *P. generosa* (Gould, 1850), *P. globosa* (Dall, 1898), *P. glycimeris* (Born, 1778), *P. smithae* (Powell, 1950), *P. japonica* (Adams, 1850), *P. zelandica* (Quoy & Gaimard, 1835). Habitan en zonas templadas y viven enterradas en sustratos arenosos, fango-arenosos y fango-grava desde el intermareal hasta los 110 metros de profundidad con una disposición espacial en forma de “parches” (Campbell *et al.*, 2004).

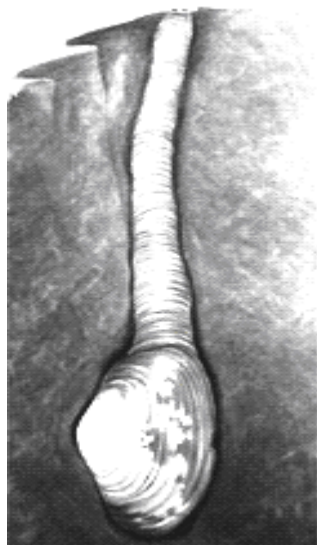


Figura 1. Representación de *P. globosa* (Dibujo de Eduardo Vázquez Solórzano).

Las almejas de sifón tienen la concha formada por dos valvas ovaladas de tamaño similar que pueden presentar una coloración blanquecina o ser más oscuras adquiriendo tonos marrones. La concha puede medir más de 25cm y el sifón más de 1m de longitud de tal manera que no puede ser retraído completamente, pero les

permite vivir enterradas en el sedimento hasta 1 metro de profundidad ya que al extenderlo se alimentan, expulsan desechos y gametos por el mismo. El pie muscular les sirve para enterrarse, pero es vestigial en la fase adulta por lo que si un individuo es desenterrado ya no puede volverse a enterrar resultando vulnerable a la depredación (Feldman *et al.*, 2004). Además esto condiciona la capacidad de dispersión de estas almejas, que depende casi exclusivamente de la fase larvaria, en la cual la larva viaja hasta 47 días en el plancton, y, en menor medida, de la fase juvenil (Straus *et al.*, 2008). Es una especie muy longeva, se tienen registros de organismos de más de 100 años de edad (Bureau *et al.*, 2002). Posee una alta tasa de fecundidad, ya que producen millones de óvulos en un desove y pueden desovar múltiples veces por temporada reproductiva. Sumado a esto, no presentan senescencia reproductiva, por lo que su capacidad reproductiva se mantiene intacta desde su adquisición, registrada en *P. generosa* para machos a los 7 años y para hembras a los 8 años (Sloan & Robinson, 1984). Sin embargo, las tasas de reclutamiento de estas especies son muy bajas lo que indica una tasa de mortalidad natural mayor en los primeros años de vida que se ve reflejado en la estructura de edad de las poblaciones, donde predominan los organismos de más de 10 años de edad. El conjunto de estos caracteres biológicos, gran longevidad, baja tasa de reclutamiento y vida sésil, la convierten en una especie susceptible a la sobreexplotación.

Actualmente cinco de las nueve especies existentes de este género son objeto de pesca para el consumo humano: *P. zelandica* (en Nueva Zelanda), *P. japonica* (en Japón), *P. abbreviata* (en Argentina), *P. globosa* (en México) y *P. generosa* (en Canadá, USA y México). Esta última es la especie de *Panopea* más importante comercialmente, acaparando para el año 2011 más del 80% de la producción global de almeja de sifón (Caplog report, 2013). Cabe destacar que la especie *Panopea generosa* ha sido confundida con *P. abrupta* (Conrad, 1849) hasta que se aclaró recientemente su nomenclatura correcta (Vadopalas *et al.*, 2010).

### 1.1.2. Pesquería de *Panopea* en México

En aguas mexicanas se encuentran dos especies de este género, *P. generosa*, que se distribuye desde el sur de Alaska hasta las costas occidentales de la península de Baja California, y *P. globosa*, que se creía endémica del Golfo de California (Fig. 2), pero también se encuentra en la costa occidental del estado de Baja California Sur, en Bahía Magdalena (Suárez-Moo *et al.*, 2012).

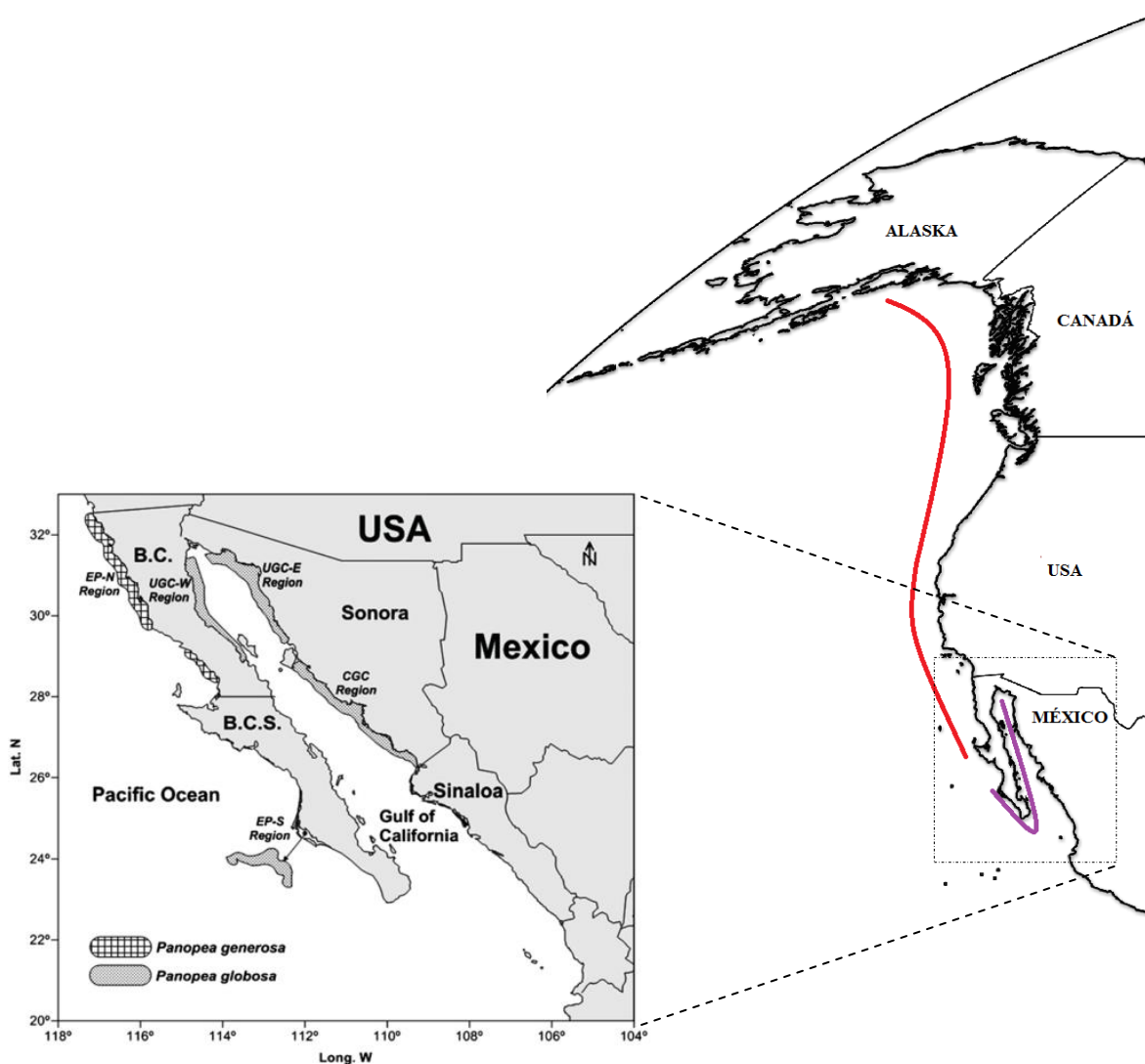
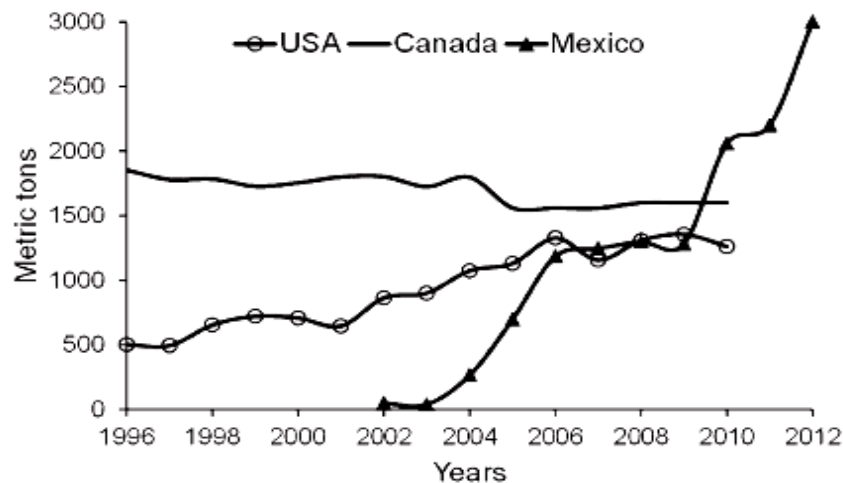


Figura 2. Distribución mundial y regiones específicas de distribución en México de *Panopea generosa* (línea roja) y *Panopea globosa* (línea morada). EP-N (Región Norte del Pacífico Oriental, EP-S (Región Sur del Pacífico Oriental), UGC-W (Región Oeste del Alto Golfo de California), UGC-E (Región Este del Alto Golfo de California) y CGC (Región Central del Golfo de California). Tomado y editado de Aragón Noriega *et al.* (2012).

La explotación de *P. generosa* comenzó en los años 70s en el Noreste del Pacífico, primero en el estado de Washington de los EEUU y poco después se sumó el estado de British Columbia de Canadá (Orensanz *et al.*, 2004). En México, las capturas no se iniciaron hasta el año 2002, primero de manera exploratoria y posteriormente, en 2004, con fines comerciales. Las primeras extracciones fueron de *P. globosa* en la costa oriental de la península de Baja California y más tarde en la costa occidental de la misma se inició la pesca de *P. generosa*. En el año 2006, las capturas de *Panopea spp.* en México alcanzaron las 1200t/año equiparándose a las de EEUU y Canadá. Posteriormente, también se comenzó a explotar a *P. globosa* en la costa occidental de Baja California Sur y en la costa de Sonora, con lo que en 2010 y 2011, las capturas en México sobrepasaban las 2000t/año (Aragón-Noriega *et al.*, 2012) (Fig. 3). El destino final de este recurso es el mercado asiático para el consumo humano y, dado que no existen competidores similares, su demanda y precio en el mercado son altos proporcionando ingresos de más de 30 millones de dólares en México. Por consiguiente, tanto *Panopea generosa* como *P. globosa* son un recurso de alta relevancia para las pesquerías del Noroeste de México.



**Figura 3. Captura comercial del género *Panopea* en el Pacífico Nororiental. Desembarco de *Panopea generosa* en Canadá (Columbia Británica) y Estados Unidos de América (Washington); *P. generosa* y *P. globosa* en México. Tomado de Aragón Noriega *et al.* (2012), p. 735.**

Recientemente, se han publicado dos planes de manejo, uno que engloba ambas especies en el estado de Baja California y uno exclusivo de *P. globosa* en el estado de Sonora (Diario oficial de la Federación, 2012). Las pautas de manejo definidas se basan en un enfoque precautorio, ya que ambos planes de manejo recalcan la necesidad de aumentar el conocimiento local de las poblaciones del género *Panopea* de México para que se maximice su producción a la vez que se asegura la conservación de cada una de las especies.

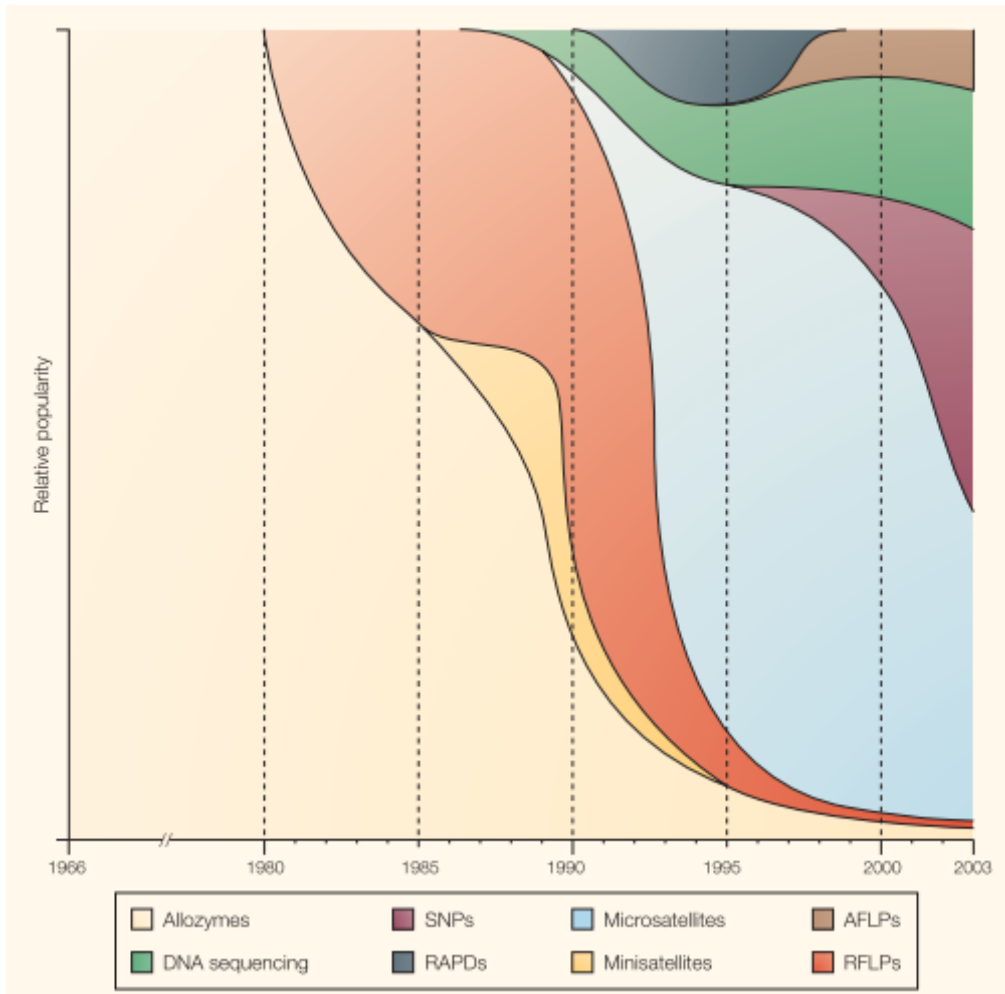
### **1.1.3. Marcadores moleculares**

La explotación sustentable de un recurso vivo debe desarrollarse con base en un conocimiento completo de su biología, ecología y evolución. Hoy en día las herramientas moleculares juegan un papel muy importante en este sentido, ya que con los análisis genéticos de una especie se pueden inferir muchos procesos ecológicos y evolutivos de la misma (Burton, 1996). La mayoría de los estudios genéticos se desarrollan gracias a marcadores moleculares que nos permiten diferenciar el genotipo de los organismos (Schlötterer, 2004). Existe una gran variedad de marcadores moleculares que se diferencian tanto en la cantidad de información que pueden revelar como en las técnicas de laboratorio que precisan para su aplicación. Un buen marcador molecular debe reunir ciertas características para maximizar su utilidad: distribuido en todo el genoma, alto grado de polimorfismo, proceso de análisis rápido y sencillo, reproductibilidad eficiente y menor costo posible. La búsqueda de este ideal junto con los avances científicos y tecnológicos es lo que provoca que la popularidad de los diferentes tipos de marcadores cambie a lo largo de los años y muchos se vean reemplazados por otros casi por completo (Fig. 4).

Destaca la gran popularidad de los marcadores microsatelitales que desde su aparición a finales de los 80 sustituyeron a los RFLPs en importancia y, hoy en día, siguen siendo uno de los más usados en diferentes campos de la biología. Los marcadores microsatélites son motivos de 1 a 6 pares de bases (pb) repetidos en tándem dentro del genoma de todos los organismos eucariotas y son muy prácticos dadas sus características: elevado polimorfismo por sus altas tasas de mutación, herencia



mendeliana simple, codominancia y no precisan secuenciación para su análisis. El mayor inconveniente de estos marcadores es que su desarrollo precisa conocimiento previo de la secuencia genómica de la especie, lo que hace que el proceso de obtención clásico sea largo y costoso.



**Figura 4. Cambios relativos en la popularidad de los diferentes tipos de marcadores moleculares (Tomada de Schlötterer, 2004).**

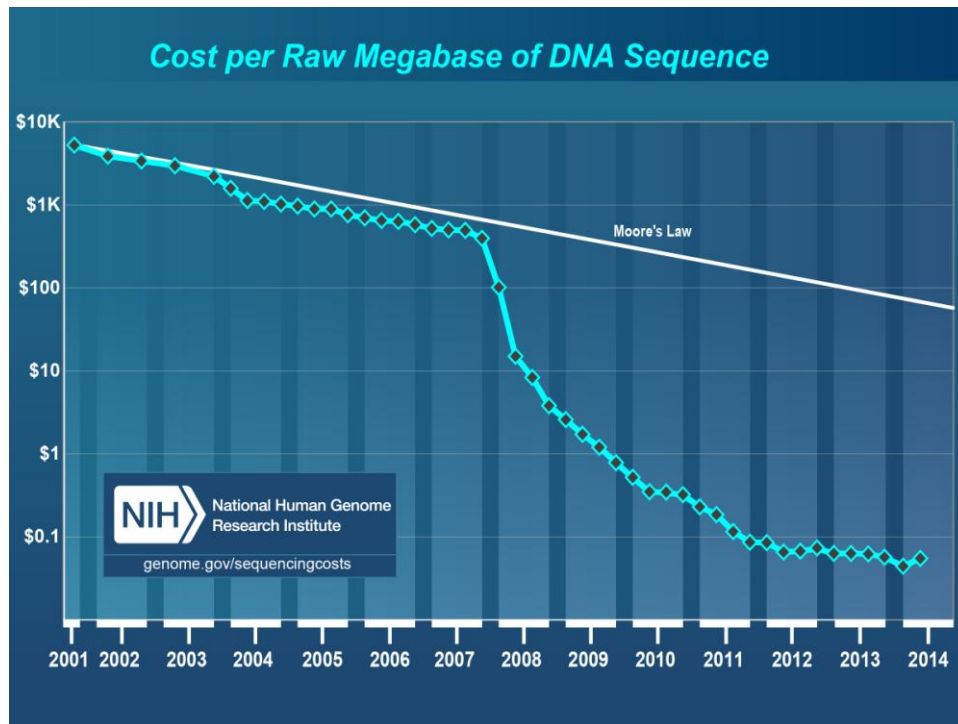
El método clásico consiste en la identificación de loci microsatélites a partir de sondas aplicadas a librerías genómicas (Rassmann *et al.*, 1991). El primer paso es digerir el ADN objetivo con enzimas de restricción y clonarlos en vectores bacterianos. Una vez completada la transformación bacteriana y el crecimiento de las colonias de las mismas, éstas se transfieren a una membrana de nylon para el paso de hibridación con sondas. Estas sondas contienen microsatélites y están marcadas radiactivamente, de tal manera

que las hibridaciones positivas se detectarán indicando aquellas clonas cuyos insertos contengan loci microsatelitales. Estos fragmentos de ADN con loci microsatelitales se secuencian y a partir de la secuencia obtenida se diseñan cebadores para poder realizar el análisis genético de estos loci mediante la técnica de la PCR. Finalmente los resultados se revelan en geles de poliacrilamida. Este proceso tiene una eficacia de detección de loci microsatelitales baja en relación al esfuerzo que supone su aislamiento (Zane *et al.*, 2002). En consecuencia, muchos investigadores han optado por usar cebadores ya diseñados en una especie para estudios genéticos en especies afines (cebadores heterólogos) logrando buenos resultados (Srikwan *et al.*, 1996). Pero frecuentemente suele ocurrir lo contrario, que los microsatélites aislados para una especie no sean útiles en especies afines. Por un lado, es posible que los cebadores diseñados para una especie no amplifiquen nada en el ADN de la otra especie, pero también puede ocurrir que amplifiquen inconsistentemente, lo que genera alelos nulos que van a provocar resultados equívocos y difíciles de interpretar. Otras aproximaciones se enfocaron en maximizar el número de microsatélites aislados usando librerías genómicas enriquecidas, pero este proceso involucra aún más pasos por lo que implica un tiempo aún más largo de obtención (Ostrander *et al.*, 1992; Paetkau 1999; Nunome *et al.*, 2006). Con el surgimiento de la secuenciación de siguiente generación (NGS, por sus siglas en inglés), aparece otra alternativa para aislar microsatélites en especies de novo, ya que estas técnicas permiten la producción de una gran cantidad de secuencias genómicas de cualquier taxón a un costo relativamente bajo. Gracias a esta tecnología se ha disparado el desarrollo de marcadores moleculares como los microsatelitales específicos en especies no modelo (Abdelkrim *et al.*, 2009; Lance *et al.*, 2013).

#### **1.1.4. Secuenciación genómica masiva**

Los últimos avances en el desarrollo de marcadores moleculares vienen de la mano de las nuevas tecnologías de secuenciación, comúnmente conocidas como Secuenciación en paralelo o de siguiente generación que han suplantado al tradicional método de secuenciación Sanger (Sanger & Nicklen, 1977). Esta nueva tecnología abarata enormemente los costes y maximiza el rendimiento de la secuenciación, ya que se efectúan cientos de miles de reacciones de secuenciación al mismo tiempo sobre una

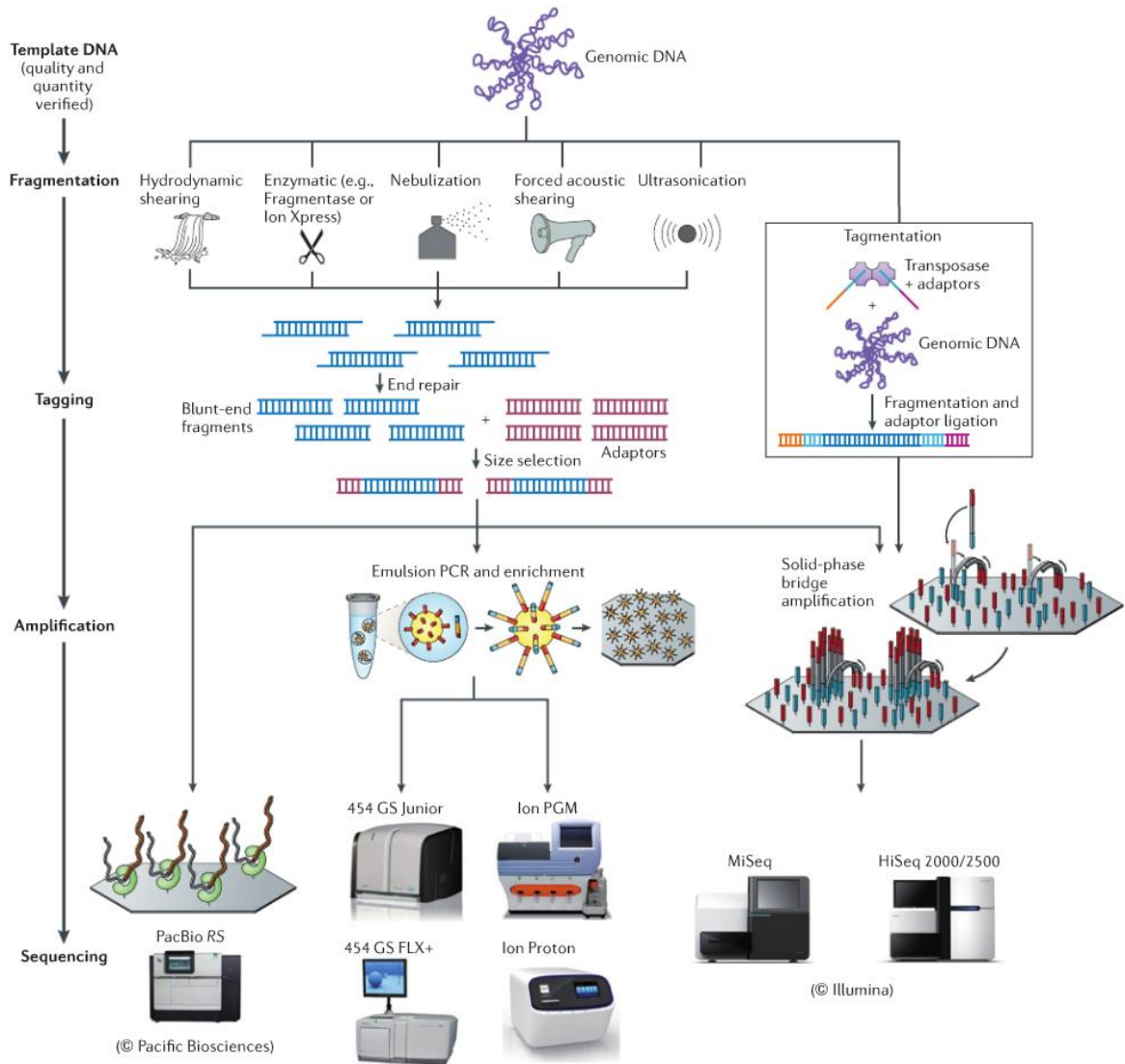
plataforma sólida, permitiendo la obtención de gran parte del genoma de una especie a investigadores individuales (Mardis, 2006; Metzker, 2005; Schuster, 2008; Shendure & Ji, 2008) (Fig.5) . Por lo tanto, la NGS tiene un gran número de aplicaciones que están facilitando un progreso científico en áreas como el análisis de patrones de metilación, los estudios de transcriptómica y el desarrollo de marcadores moleculares.



**Figura 5. Evolución del coste de la secuenciación de ADN por megabase desde el años 2001 al 2014.**

En la última década cinco secuenciadores de siguiente generación han dominado el mercado: Roche GS-FLX 454 Genome Sequencer (originalmente 454 sequencing), the Illumina Genome Analyzer (originalmente Solexa technology), ABI SOLiD analyzer, Ion Torrent y PacBio, pero existen muchas otras plataformas (Van Dijk *et al.*, 2014).

En general, estas plataformas se basan en un flujo de trabajo que consta de tres etapas: preparación de la biblioteca (extracción, fragmentación y etiquetado del ADN), amplificación y secuenciación del ADN (Fig. 6). Sin embargo, en cada secuenciador estos procesos se llevan a cabo de forma diferente, lo que condiciona las características (longitud, calidad, cantidad) de las secuencias resultantes (Fig. 7).



**Figura 6. Esquema general de la secuenciación de siguiente generación para las cuatro plataformas disponibles (Loman & Constantinidou, 2012).**

La plataforma 454/ Roche FLX System fue la primera en salir al mercado en 2005. Esta plataforma tiene como soporte una placa picotituladora (“picotiter plate”) y funciona por reacciones de pirosecuenciación. De tal manera que se llevan a cabo reacciones de polimerización y cada vez que se incorpora un nucleótido la liberación del pirofosfato se manifiesta mediante luz que es captada y traducida por un ordenador dando como resultado la secuencia de ADN (Margulies *et al.*, 2005). Esta plataforma se ha usado en muchos estudios para el desarrollo de marcadores moleculares como los microsatelites,

ya que permite obtener lecturas largas de hasta 700pb donde encontrar loci microsatelitales con suficiente región flanqueante para el diseño de cebadores de amplificación y así probar su eficacia como marcadores moleculares en el laboratorio (Castoe *et al.*, 2012). Sin embargo requiere reactivos muy caros lo que condiciona un costo por base relativamente alto si se compara con otras plataformas de secuenciación. Un año después se comercializó la plataforma Illumina, que se basa también en la técnica de secuenciación por síntesis, pero en ésta los nucleótidos están marcados por fluorescencia y su extremo 3' está obstruido hasta que se une a la cadena de ADN. Esta unión desencadena la liberación del fluoróforo y el extremo queda expuesto para la unión del siguiente nucleótido (Turcatti, Romieu, Fedurco, & Tairi, 2008). La tercera plataforma en comercializarse fue por la compañía *Solid* que usa la técnica de secuenciación por ligación, pero la detección también es mediante fluorescencia. Estas últimas plataformas no consiguen lecturas tan largas como la 454, en la que actualmente se obtienen lecturas de aproximadamente 1000pb. Sin embargo, la cantidad de lecturas por corrida es mayor que en la primera dando como consecuencia un costo por Megabase menor (Van Dijk *et al.*, 2014). Hasta ahora, la plataforma Illumina es la de mayor rendimiento y abarca el 70% del mercado de las plataformas de NGS (Bahassi & Stambrook, 2014). Sin embargo el progreso tan vivaz de esta tecnología puede suponer que otra plataforma la sustituya en un futuro cercano.

### **1.1.5. Anotación genómica**

Consiste en localizar e identificar los genes codificados en las secuencias de ADN y la función de su producto (ARN o proteína). Esto se ha llevado a cabo en los genomas de diferentes organismos modelo como por ejemplo la levadura de la cerveza, *Saccharomyces cerevisiae*, o la especie de nematodo *Caenorhabditis elegans*. El conjunto de varios organismos anotados genéticamente ha manifestado un alto nivel de conservación de la secuencia de genes con funciones biológicas importantes en eucariotas (Tang, Kuo, Shen, & Lin, 2000). Esta homología permite realizar lo que se conoce como anotación genética automatizada a través de programas informáticos que transfieren las anotaciones de especies ya analizadas a la secuencia de estudio por

comparación (Ashburner *et al.*, 2000). Existe un Consorcio de Ontología genética que define un vocabulario universal que clasifica en categorías funcionales cada gen así como su producto génico y que es actualizado constantemente. La ontología genética se divide en tres categorías funcionales: proceso biológico, función molecular y componente celular (Ashburner *et al.*, 2000).

La gran cantidad de secuencias de organismos no modelo que se está generando a partir de la NGS ha llevado a una alta demanda en la anotación automatizada, por lo que este proceso se puede llevar a cabo cada vez más rápido y eficientemente (Chen *et al.*, 2012). Esta aplicación permite, entre otros, verificar si los marcadores moleculares que se van a desarrollar en un organismo están o no sujetos a selección. Por otro lado se ha desarrollado un amplio campo de anotación automatizada del ADN mitocondrial ya que éste es altamente conservado en metazoos. En los cuales esta formado, generalmente, por 13 genes estructurales, 2 ARN ribosómicos y 22 ARN de transferencia. Además, es muy probable obtener la secuencia del ADN mitocondrial completo de un organismo a partir de las lecturas obtenidas por NGS ya que existen múltiples copias del mismo en cada célula. Esto aunado a la gran cantidad de software especializados en la anotación mitocondrial ha desencadenado la rápida anotación del mitogenoma de una gran cantidad de especies (Gan, Schultz, & Austin, 2014).

En este estudio se van a secuenciar masivamente los genomas de *Panopea generosa* y *Panopea globosa* con el propósito de crear una base genómica para el desarrollo de marcadores microsatelitales, anotación funcional y anotación del mitogenoma que sirvan para futuros estudios genéticos sobre ambas especies con el fin de mejorar los planes de gestión, acuicultura y conservación de éstas.

## 1.2. Antecedentes

### 1.2.1. Estudios de *Panopea generosa* y *P. globosa* con diferentes tipos de marcadores moleculares.

Se han realizado diferentes estudios genéticos con el propósito de aclarar la distribución geográfica de *Panopea generosa* y *P. globosa*, así como para analizar la variabilidad genética poblacional, entre otros. Se describen brevemente a continuación en orden cronológico.

VanKoeveringe (1998) secuenció el gen Citocromo Oxidasa III mitocondrial en individuos de *P. generosa* de la Columbia Británica, Canadá, sin encontrar una estructura genética poblacional. Sin embargo, estos resultados pueden ser consecuencia de que el estudio se realizó con pocos individuos de cada población (6 en promedio) y/o una escala geográfica pequeña en relación a la distribución total de la especie.

Vadopalas *et al.* (2004) usaron dos tipos de marcadores moleculares, aloenzimas (11 loci) y microsatélites (7 loci), para un estudio filogeográfico de *P. generosa* evaluando 15 poblaciones de Washington y 1 de Alaska. La mayoría de las poblaciones mostraron homogeneidad genética, pero las que sí se diferenciaron significativamente no cumplen el modelo de aislamiento por distancia. Por lo que parece que la complejidad del área de estudio hace que la dispersión larval, y en consecuencia el flujo génico, dependa de otros factores como la batimetría y la hidrología.

Miller *et al.* (2006) estudiaron la estructura genética de *P. generosa* mediante 8 marcadores microsatelitales en poblaciones que abarcan 1100 km de línea de costa, desde Washington hasta el Norte de la Columbia Británica. La estructura genética encontrada en estas poblaciones responden a un patrón de flujo génico conocido como aislamiento por distancia que sólo se cumple a una escala media (500- 1000km), en cambio a una escala menor (50-300km) se observó que las poblaciones son panmícticas.

Rocha-Olivares *et al.* (2010) compararon las variaciones genéticas y morfológicas entre las especies *P. globosa* y *P. generosa*. Los análisis genéticos, realizados mediante PCR-RFLP (Polymerase Chain Reaction - Restriction Fragment Length Polymorphism) y la secuenciación de DNA, usando la region ITS del ADN ribosomal (ADNr: ITS1-5.8S-ITS2), revelaron que, la primera técnica manifiesta datos limitados para evaluar la estructura genética de cada especie, no obstante es útil para la identificación taxonómica entre estas especies de *Panopea*. En la secuenciación del ADNr (ADN ribosómico) se encontraron más diferencias genéticas tanto intra como interespecíficas que en el análisis RFLP, siendo mayor la diversidad genética en *P. globosa* que en *P. generosa*. Aún así, las muestras y las localidades analizadas no parecen ser suficientes para inferir patrones filogeográficos.

Perez-Valencia (2011) en su tesis de maestría estudia la variabilidad genética de las poblaciones de *P. globosa* mediante 2 marcadores microsatélites polimórficos. Los resultados revelan una posible estructura genética poblacional, formando las almejas de Bahía Magdalena una población diferente de las del Alto Golfo de California, pero estadísticamente los resultados no son robustos debido al bajo número de loci microsatelitales, por lo que propone la necesidad de utilizar más loci microsatélites polimórficos.

Leyva-Valencia (2012) en su tesis de doctoral estudió la filogenia de 5 especies de *Panopea*: *P. abbreviata*, *P. zelandica*, *P. japonica*, *P. generosa* y *P. globosa* usando el gen citocromo oxidasa I de ADNmt (ADN mitocondrial) y los genes 28S y 18S de ADNr. Los resultados mostraron dos parejas de especies más similares genéticamente: *P. japonica*- *P. generosa* y *P. abbreviata*- *P. globosa*, quedando *P. zelandica* como especie relictas.

Vadopalas *et al.* (2012) estimaron la edad de individuos de *P. generosa* contando las bandas de crecimiento de su concha en dos poblaciones de Puget Sound, en las que también analizaron las diferencias genéticas mediante marcadores moleculares, aloenzimas (11) y microsatélites (7), con el objetivo de comprobar la hipótesis del reclutamiento tipo lotería que postularon tras los resultados de un estudio anterior



(Vadopalas *et al.*, 2004). Los datos obtenidos no sostienen esta hipótesis, por lo que la diferencia genética entre poblaciones cercanas se atribuye a otras variables que afectan al proceso de dispersión de la larva.

Suárez-Moo (2012) en su tesis de maestría analizó la estructura genética poblacional de *P. generosa* en el Pacífico Nororiental con marcadores moleculares mitocondriales (secuencias de los genes Citocromo Oxidasa I y III) y nucleares (microsatélites, 3), encontrando una alta diversidad genética dentro y un elevado flujo genético entre las poblaciones.

Además, Suárez-Moo *et al.* (2013) realizaron un estudio integrando la ecología, la genética y la morfología de *Panopea globosa*. Se describen diferencias morfológicas entre las almejas de sifón que habitan en Bahía Magdalena y las poblaciones de la misma que se distribuyen en el Golfo de California, pero los análisis genéticos, mediante la PCR- RFLP de ADN ribosómico y la secuenciación del gen Citocromo Oxidasa I de ADNmt, confirman que pertenecen a la misma especie, *P. globosa*.

### **1.2.2. Desarrollo de marcadores microsatelitales mediante plataformas de secuenciación de siguiente generación.**

La identificación y caracterización de microsatélites a partir de secuencias generadas por las nuevas plataformas de secuenciación masiva se ha llevado a cabo en una gran variedad de organismos como plantas (Zalapa *et al.*, 2012), insectos (Perry & Rowe, 2011), aves (Abdelkrim *et al.*, 2009), reptiles (Castoe *et al.*, 2010a) y más recientemente se están aplicando en moluscos.

Greenley *et al.* (2011) usando la plataforma de pirosecuenciación 454 para la secuenciación del abulón amarillo (*Haliotis corrugata*) aislaron y caracterizaron 18 marcadores microsatelitales polimórficos que servirán para realizar análisis parentales y el estudio más al detalle de la estructura genética.

An & Lee (2012) generaron 22 marcadores microsatelitales polimórficos mediante la plataforma de secuenciación de siguiente generación 454 GS-FLX en la especie de mejillón, *Mytilus coruscus*, de gran importancia comercial en Korea, con el propósito de que se empleen en estudios genéticos que favorezcan la conservación y la gestión adecuada de la especie.

Inoue *et al.* (2012) identificaron 20 marcadores microsatelitales polimórficos en una especie de bivalvo de agua dulce, *Popenaias popeii*, cuyas poblaciones han declinado. Se espera que la aplicación de estos marcadores a estudios de genética poblacional dé resultados lo suficientemente robustos para que se apliquen las medidas de conservación necesarias.

O'Bryhim *et al.* (2012) desarrollaron 16 marcadores microsatelitales polimórficos usando la plataforma de secuenciación Illumina en la especie de bivalvo *Leptodea leptodon* que es endémica de EEUU y que está en peligro de extinción. Los nuevos marcadores se pretenden usar para estudios de conectividad genética que procuren información útil para las estrategias de conservación.

Slabbert *et al.* (2012) secuenciaron la especie *Haliotis midae*, cultivada y exportada en Sudáfrica, por pirosecuenciación en el sistema 454 GS-FLX. A partir de las secuencias generadas aislaron y caracterizaron 82 microsatélites polimórficos que pueden ser útiles para estudios parentales y demográficos.

Recientemente, Cruz-Hernández *et al.* (2014) desarrollaron 24 marcadores microsatelitales para *Panopea globosa* usando la plataforma de pirosecuenciación 454. Éstos servirán para estudios de genética poblacional y conectividad.

### 1.3. Justificación

El género *Panopea* sustenta una de las pesquerías más importantes del Pacífico Nororiental por la alta demanda asiática y el gran capital que se genera en los países donde se obtiene que son EEUU, Canadá y México. Las ganancias en el Estado de Washington para el periodo 2006-2008 se valoraron en más de 33 millones de dólares (<http://www.st.nmfs.noaa.gov>) , siendo en la Columbia Británica aproximadamente 35 millones de dólares anuales (Khan, 2006). En México, la pesquería de *Panopea* spp. ha llegado a generar 30 millones de dólares anuales (SAGARPA, 2007). Actualmente, las autoridades de pesca mexicanas (SAGARPA-CONAPESCA) proponen un plan de manejo para esta actividad, en donde la cuota actual de captura permite la extracción de 1% de la biomasa virgen, con el objetivo a largo plazo de extraer el 50% de ésta (SAGARPA, 2012). Esto refleja el gran potencial que tiene la pesquería de la almeja de sifón en México, pero siempre y cuando se desarrolle una captura sustentable sin que corra peligro la viabilidad del recurso.

En primer lugar, la garantía de una explotación sostenible es conocer el recurso. En México la pesquería de la almeja de sifón es relativamente reciente, por lo que los estudios de biología y ecología son aún escasos, siendo una evidencia clara de esto la inexistencia de un plan de gestión diferenciado para cada especie. Urge, por tanto, rellenar los huecos del conocimiento de cada especie de *Panopea* para asegurar una explotación específica sostenible (Aragón-Noriega *et al.*, 2012).

En segundo lugar, las especies de *Panopea* tienen aspectos biológicos y evolutivos (extensa longevidad, madurez tardía, lento crecimiento y dispersión larval) que las hacen especialmente susceptibles a la sobreexplotación. Por lo que se busca una extracción de individuos con un entendimiento preciso del stock poblacional así como apostar por técnicas alternativas de producción como la acuicultura.

Por último, se apuesta por la ecología molecular como un medio actual y eficaz para obtener datos relevantes que mejoren el manejo de las pesquerías y las técnicas de acuicultura con el fin de maximizar la producción mientras se asegura la conservación de cada especie.

## 1.4. Hipótesis

Con el análisis de las secuencias obtenidas por NGS se podrán encontrar loci microsatelitales candidatos a marcadores moleculares y la secuencia que codifica el ADN mitocondrial para cada una de las especies.

## 1.5. Objetivos

### 1.5.1. Objetivo general

- Analizar las secuencias provenientes de la NGS de *Panopea generosa* y *Panopea globosa* para el desarrollo de marcadores moleculares.

### 1.5.2. Objetivos específicos

- Analizar las secuencias de los genomas de ambas especies.
- Identificar loci microsatelitales y seleccionar los candidatos a marcadores moleculares.
- Optimizar las condiciones de amplificación por PCR de los loci microsatelitales candidatos y evaluar su nivel de polimorfismo.
- Realizar la anotación genética preliminar de los contigs obtenidos.
- Identificar y anotar la secuencia de ADN mitocondrial.

## 2. Materiales y métodos

### 2.1. Extracción de ADN

Se extrajo ADN de dos individuos frescos de *Panopea generosa* del Rincón de Ballenas, Ensenada y de varios individuos conservados en etanol al 96% de *Panopea globosa* de las localidades de San Felipe y Bahía Magdalena mediante el DNeasy Kit (Qiagen, Inc.). Para la secuenciación de siguiente generación se requieren una cantidad ( $> 50\text{ng}/\mu\text{l}$ ) y calidad ( $A_{260}/A_{280} > 1.80$ ) de ADN que se comprobaron por espectrofotometría, con un equipo Nanodrop Lite spectrophotometer (Thermo science, Madison, WI, USA) y Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA), y por electroforesis mediante geles de agarosa.

### 2.2. Secuenciación masiva

Se envió a secuenciar el ADN que se extrajo de los dos individuos de *P. generosa* y de la otra especie se mandó ADN extraído de dos individuos de San Felipe y uno de Bahía Magdalena. La plataforma que se eligió para la secuenciación masiva fue Illumina por el método de paired-end reads. Se prepararon bibliotecas independientes tipo TruSeq nano para cada individuo y se añadieron a los fragmentos obtenidos adaptadores de identificación. La corrida de secuenciación se llevó a cabo para cada individuo usando 1/12 parte de una línea del equipo HiSeq 2500 en formato “pair-end” 2x100 (Illumina, San Diego, CA) en el Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO, Irapuato, México). En el tipo de secuenciación “pair-end” 2x100 se secuencian 100 pb de ambos extremos del fragmento de ADN obtenido en la librería, a estas dos secuencias que provienen de un mismo fragmento de ADN genómico inicial se les denomina secuencias pareadas.

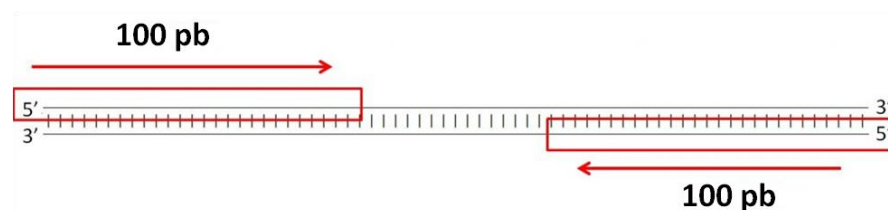


Figura 7. Representación gráfica de la secuenciación pair-end 2x100

## 2.3. Bioinformática

### 2.3.1. Análisis de secuencias

Lo primero que se hizo fue evaluar la calidad de las secuencias en crudo mediante dos programas distintos, FastQC v 0.10.1 (Babraham Institute, Cambridge, UK) (Andrews, 2010) y el CLC Genomic Workbench 7.0.3 (CLC Bio, Boston, MA) (Knudsen & Flensburg, 2008). Estos programas dan información estadística sobre parámetros como los valores de calidad de las secuencias, contenido en G+C, la sobrerrepresentación de secuencias (indicando que la librería genómica está sesgada o contaminada) y otras variables. Cabe destacar que el formato en el que envían las secuencias incluye un valor de calidad para cada base secuenciada, lo que se conoce como valor Q. Este valor indica la probabilidad de que una base determinada sea nombrada incorrectamente por el secuenciador (Tabla 1).

Tabla 1. Valor de calidad Phred y precisión de la secuenciación.

Phred Quality Score	Probability of Incorrect Base Call	Base Call Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

A continuación en el mismo programa CLC Genomics Workbench se hizo una limpieza de las secuencias (trimming) para eliminar o recortar los extremos con bases de baja calidad. En este mismo programa se realizó el ensamble *de novo* en el cuál, como si de un rompecabezas se tratara, las secuencias que solapan formarán “contigs”. Contig viene de la palabra contiguo y se define según el National Human Genome Research Institute como la cadena de superposiciones de secuencias de ADN que reconstruye la secuencia de ADN original de un cromosoma o una región del mismo (<http://www.genome.gov/Glossary/index.cfm?id=39>). Además otro dato importante que

se obtiene es la cobertura promedio para cada contig que depende del número de secuencias iniciales que coincidan en una misma posición del contig. No todas las secuencias pueden ensamblarse por lo que a aquellas secuencias únicas que permanecen sin ensamblar se les denomina “singletons”.

### **2.3.2. Identificación y selección de loci microsatelitales candidatos a marcadores moleculares**

La búsqueda de motivos repetitivos microsatelitales se hizo en el conjunto de Contigs obtenidos del ensamble *de novo* para cada especie por separado usando el programa Msatcommander (Faircloth, 2008). Este programa permite la identificación de microsatélites en los contigs de acuerdo a criterios tales como: perfectos y/o compuestos, tipo de motivo repetitivo, mínimo nº de repeticiones, etc. En este proyecto se dio prioridad a los microsatélites perfectos (no compuestos) formados por motivos repetitivos de di, tri o tetranucleótidos y con un número de unidades repetidas mínimo de 8, 8 y 6 veces respectivamente. Además con el mismo programa se diseñaron cebadores para la amplificación por PCR de los loci detectados. Asimismo algunos cebadores fueron mejorados con el programa Primer-BLAST. Este paso es muy importante ya que sólo aquellos loci con cebadores diseñados pueden ser candidatos a marcadores moleculares. El tamaño de los cebadores, el porcentaje de GC en la secuencia de los mismos, la temperatura de anillamiento y otras características determinarán la calidad de éstos y serán determinantes para su elección. Cabe destacar que es preferible seleccionar cebadores en los tramos del contig con una profundidad de secuencias mínimo 5X.

En este punto se seleccionaron 30 loci candidatos a marcadores moleculares para cada especie mediante análisis visuales en donde se tenían en cuenta diferentes factores: alta cobertura del contig en la posición del locus y los cebadores diseñados, mínima presencia de posibles bases variables sobretodo en la secuencia de los cebadores porque serían degenerados y son menos específicos, ausencia de otros loci entre los cebadores y el locus objetivo, temperaturas de anillamiento similares entre el par de cebadores para un mismo locus, porcentaje de GC adecuado, etc.

Finalmente se mandaron a sintetizar los 30 pares de cebadores para cada especie de *Panopea* a la empresa IDT.

## **2.4. Pruebas experimentales**

### **2.4.1. Optimización de las condiciones de amplificación de los loci microsatelitales candidatos a marcadores moleculares**

Se realizaron gradientes de temperatura de anillamiento para cada locus. De esta manera se busca encontrar la temperatura de anillamiento óptima así como las mejores condiciones químicas para la amplificación de la banda de ADN del tamaño esperado. Los resultados se revelaron en geles de agarosa al 1.5%. Posteriormente los cebadores optimizados fueron probados en 10 organismos para confirmar dicha optimización y para identificar los loci microsatelitales polimórficos. Estos resultados se visualizaron en geles de agarosa al 2%, al 2.5% y en geles de acrilamida.

El paso siguiente consistió en seleccionar aquellos loci más polimórficos que presentaron una amplificación positiva en la mayor parte de los 10 organismos probados. De los loci seleccionados se mandó a sintetizar el mismo cebador Forward con el que se amplificó, pero añadiéndole una extensión consistente en el cebador universal M13. Esta extensión o cola universal M13 consiste en 18 bases que permiten la incorporación de un fluoróforo al fragmento de ADN de interés mediante dos pasos de amplificación en los que intervienen tres cebadores distintos y dos temperaturas de anillamiento diferentes (la óptima para cada locus y la temperatura de 53°C que es la óptima para la unión de la secuencia de la cola M13) de acuerdo a lo propuesto por Schuelke (2000). Esta incorporación permite diferenciar los alelos de cada organismo para cada locus.

En este punto, los loci seleccionados pasaron a una segunda fase de optimización con el nuevo forward M13. De nuevo se realizaron gradientes de temperatura para verificar la temperatura de anillamiento óptima de cada locus y se amplificaron hasta 36 organismos de diferentes localidades para cada especie y locus. Los loci de *P. generosa* se probaron en 14 organismos de dos localidades cercanas de Washington,



EEUU (7 de Hood Canal y 7 de Alden Bank), junto con 21 de tres localidades del estado de Baja California, México (7 de Islas Coronado, 7 de San Quintín y 7 de Santa Rosalita). En el caso de *P. globosa* se experimentaron en organismos de tres localidades (12 de Bahía Magdalena, 12 de San Felipe y 12 de Puerto Peñasco).

## **2.4.2. Caracterización**

Los productos amplificados de cada locus seleccionado se mandaron a genotipificar (análisis de fragmentos) en un secuenciador de ADN automático ABI 3130xl. Posteriormente los resultados de la genotipificación se visualizaron en el software GeneMarker 2.4.0 para determinar los alelos de cada genotipo y la conversión de estos datos en tamaños alélicos discretos se realizó usando un macro de Excel, FLEXIBIN (Amos *et al.*, 2007). La heterocigosidad observada ( $H_o$ ) y esperada ( $H_E$ ) y el índice de contenido polimórfico (PIC) se estimaron usando otro macro de Excel, Mstools (Park, 2001). Finalmente se comprobó el equilibrio de Hardy-Weinberg para cada locus y de desequilibrio en el ligamiento entre cada par de loci con el programa ARLEQUIN ver. 3.1 (Excoffier *et al.*, 2005). Se hizo a posteriori una corrección de Dunn-Sidák para comparaciones múltiples que estrecha los resultados significativos en estos dos últimos análisis.

## **2.5. Anotación**

### **2.5.1. Anotación general de los contigs**

Los contigs se introdujeron el programa FASTANNOTATOR para obtener la anotación de forma automatizada (Chen *et al.*, 2012). Este programa revela la cantidad de contigs que fueron anotados por homologías dentro de la base de datos de Ontología genética, de enzimas y/o dominios proteicos. Se examinó la anotación específica para los contigs de los loci que fueron seleccionados.

### **2.5.2. Anotación mitocondrial**

La búsqueda de la secuencia del genoma mitocondrial se realizó mediante una búsqueda BLAST (Altschup *et al.*, 1990) de los contigs de mayor longitud y seleccionó aquel en el que se encontraron porcentajes de similitud altos con genomas mitocondriales.

A continuación se introdujo el contig seleccionado en programas específicos que anotan el ADN mitocondrial, el DOGMA (Wyman *et al.*, 2004) y el MITOS (Bernt *et al.*, 2013). Además se hicieron búsquedas en Blastx para la anotación de determinados genes así como se usó el programa específico “t-RNA scan-SE Search Server” para verificar y completar la anotación de los ARN de transferencia. Cuando fue posible se comparó con secuencias mitocondriales de otros bivalvos para establecer los codones de inicio y parada de cada gen.

### 3. Resultados

#### 3.1. Bioinformática

##### 3.1.1. Análisis de secuencias

Se obtuvieron entre 36 y 40 millones de secuencias totales de ADN para cada uno de los organismos de *Panopea* secuenciados en la plataforma Illumina con el formato 2x100. Excepto para uno de ellos donde no se llegaron a obtener 100 mil secuencias totales. En total para cada una de las especies se obtuvieron poco más de 77 millones de secuencias (Tabla 2).

Tabla 2. Cantidad de secuencias que se obtuvieron en la secuenciación masiva de cada organismo y el total de secuencias que se obtuvo para cada especie. Pgen= *Panopea generosa*, Pglo = *P. globosa*, RB =Rincón de Ballenas (Ensenada), SF = San Felipe, BM = Bahía Magdalena.

Muestra	Secuencias	Total
Pgen RB1	40,122,148	77,475,634
Pgen RB2	37,353,486	
Pglo SF02	93,852	77,381,780
Pglo SF03	36,884,220	
Pglo BM01	40,403,708	

La calidad de las secuencias fue muy buena con un valor de  $Q_{promedio} = 38$ , un valor de contenido en G+C de un 34-37%, poco contenido de bases ambiguas y menos de un 5% de secuencias sobrerrepresentadas. En el proceso de limpieza de las secuencias se modificaron menos del 10% y se descartó menos del 1% de las mismas en ambas especies (Tabla 3).

Tabla 3. Porcentaje de secuencias no modificadas, modificadas y descartadas en el proceso de limpieza y recorte (control de calidad).

Especie	Secuencias No modificadas	Secuencias Modificadas	Secuencias Descartadas
<i>P. generosa</i>	91.72%	7.52%	0.76%
<i>P. globosa</i>	91.23%	8.23%	0.54%

En el ensamblaje realizado se obtuvieron 868,521 mil y 736,851 mil contigs para *P. generosa* y *P. globosa* respectivamente. En el ensamblaje de la especie *P. globosa* el contig de mayor longitud es de 23,743pb y el de menor longitud de 80pb, en cambio en *P. generosa* el contig de mayor longitud es de 15,585pb y el de menor de 104pb. Sin embargo se obtuvo una longitud promedio de contigs alrededor de 400pb en ambas especies. Se ensamblaron casi 3/4 de las secuencias en cada una de las especies de *Panopea*, pero la cantidad de secuencias que se ensamblaron pareadas fue notablemente menor que las no pareadas (Tabla 4). Cuando ambas secuencias del mismo fragmento de ADN que se secuenció mediante paired-end coinciden en el mismo contig tras el proceso de ensamblaje, se habla de secuencias pareadas.

Tabla 4. Porcentaje total de secuencias ensambladas y porcentaje específico de las secuencias que se ensamblaron pareadas y las no pareadas.

Secuencias ensambladas	<i>P.generosa</i>		
	67.55%	pareadas	27.47%
		no pareadas	72.53%
	<i>P.globosa</i>		
	65.31%	pareadas	17.5%
no pareadas		82.5%	

### 3.1.2 Identificación y selección de loci microsatelitales candidatos a marcadores moleculares

Se obtuvieron loci de los tres tipos deseados: di-, tri- y tetranucleótidos. Los loci dinucleótidos fueron los más abundantes en ambas especies, seguidos de los trinucleótidos en *P. generosa* y de los tetranucleótidos en *P. globosa*. El porcentaje de loci identificados a los que se les pudo diseñar cebadores fue menor al 26% en todos los casos y éste es menor en *P. globosa* que en *P. generosa* en todos los tipos de loci identificados (Tabla 5).

Tabla 5. Número de loci di-, tri- y tetranucleótidos que se obtuvieron para cada especie de *Panopea* y el porcentaje de éstos a los que se les pudo diseñar cebadores.

Especie	Loci identificados		Loci con cebadores diseñados
<i>P. generosa</i>	Di-	8,060	24%
	Tri-	3,146	14%
	Tetra-	2,830	25%
<i>P. globosa</i>	Di-	4,956	18%
	Tri-	1,647	12%
	Tetra-	1,983	23%

Los 30 loci seleccionados como candidatos a marcadores moleculares de cada especie fueron 10 di-, 10 tri- y 10 tetra para *P. globosa* y 9 di-, 10 tri- y 11 tetra- para *P. generosa*.

## 3.2. Pruebas experimentales

### 3.2.1. Optimización de las condiciones de amplificación de los loci microsatelitales candidatos a marcadores moleculares

Las condiciones químicas que mostraron amplificaciones positivas son reacciones de PCR de 12.5  $\mu$ l con 200  $\mu$ M de dNTPs, 1 X PCR Perkin Elmer Buffer (10mM Tris HCl, 50mM KCl y 2mM MgCl), 0.5  $\mu$ M de cada cebador, 0.04 U Taq DNA pol (BioLabs o Kappa) y 20 ng de ADN genómico. Además con estas condiciones químicas se ajustó el perfil de termociclado de amplificación por PCR (Tabla 6).

**Tabla 6. Perfil de termociclado de PCR usado para la amplificación de los loci ensayados.**

	Tiempo	Temperatura
Hot Start	10 min	95 °C
Desnaturalización	60 seg	94 °C
Anillamiento	30 seg	X °C
Polimerización	120 seg	72 °C
Incubación final	30 min	72 °C

**X: la temperatura varió dependiendo de la composición de los primers**

En este paso se descartaron 10 loci de los 30 candidatos en cada una de las especies. Los loci que fueron desechados debido varios motivos: amplificación negativa, bandas inespecíficas, polimorfismo dudoso (Tabla 7 y 8).

Tabla 7. Descripción de los cebadores de los loci descartados y causa del descarte en *P. generosa* (Pgen) y *P. globosa* (Pglo).

Cebadores	Secuencia	Contig	Resultado experimental
Pgen2F_2	GAAACGGACTTTTCCACCACG	2071	Sin amplificación
Pgen2R_2	TGACGCTGGTTTATAGGCATG		
Pgen2F_4	GGCTTGGGAAATGTGTGGAG	5783	Sin amplificación
Pgen2R_4	GGCACTTTGCGTATACAGGG		
Pgen2F_5	ACAAACCATCAAGGCATGTTGA	6192	Polimorfismo dudoso
Pgen2R_5	GATGATGTACAAGGCGGCTG		
Pgen2F_8	TGTGTAGCTCCTTCCACTGG	9270	Polimorfismo dudoso
Pgen2R_8	TCCTTGAAGCACGCTACTTG		
Pgen2F_9	TGTGTGGAAGACAGATGGTTT	10208	Polimorfismo dudoso
Pgen2R_9	ACGGGAACATCACAGAGTTC		
Pgen3F_3	GGGCTGTTGTTAACCGATATTC	16367	Sin amplificación
Pgen3R_3	GGTCAGCAAGTTTGATCACG		
Pgen3F_5	TCCCTGAGAATGCACCATTACA	39469	Inespecífico
Pgen3R_5	ATTTACAGGGTCAATGGCAG		
Pgen4F_5	GAACTTGGAGTAGTCGTGGC	10351	Polimorfismo dudoso
Pgen4R_5	ACTCCTGGTGCGGTATTGAG		
Pgen4F_6	TTTCTGAACTGCCACTGACG	10559	Inespecífico
Pgen4R_6	TGGGTAACCGTCATGACTGAAG		
Pgen4F_8	AGTTCCAATTCCCGACCGAC	11409	Inespecífico
Pgen4R_8	TGGGTTTACAGGACGGTAC		
<b>Cebadores</b>			
Pglo2F_1	ARAGGGAGGCATGAAGGAGG	369	Polimorfismo dudoso
Pglo2R_1	TAATGCAAAGTGCGAGTGCG		
Pglo2F_4	GCACTGAATCTGCCAATGACC	5708	Inespecífico
Pglo2R_4	GAGGTGGGTATCGGGTATGG		
Pglo2F_7	ACTAGCGGTGTGTTGAACTG	10411	Inespecífico
Pglo2R_7	AGGTGAGCGCTGATAATATTCC		
Pglo2F_8	ACACTCTCTGGCAAGTCTGG	13588	Polimorfismo dudoso
Pglo2R_8	CTTGTGAGCCCTACTCCTAAAC		
Pglo2F_9	TTCTCCAGGTCCCACAACCTC	21972	Polimorfismo dudoso
Pglo2R_9	TGCTGTTGATATGGTCTGATGC		
Pglo2F_10	AAATACATGTTTCATGCCACCAC	24325	Inespecífico
Pglo2R_10	TCTCATCGGTACAATGCTGC		
Pglo3F_1	TTAGATGTGCTGTCCGGCTC	14573	Polimorfismo dudoso
Pglo3R_1	TGCATGCGTGGTATCCTTCA		
Pglo3F_6	CCGTTCAAAGTTGTCAACC	48210	Inespecífico
Pglo3R_6	TGTGGTGATCAGGTGAGCTG		
Pglo3F_7	ATATTCCGGCTSCTGCCGTT	77523	Polimorfismo dudoso
Pglo3R_7	CAAAGTCTGACGAGTGTTGA		
Pglo4F_7	GGGATTGGCATTTCGTCCG	208280	Inespecífico
Pglo4R_7	ATAACGACCAACCGCACATC		

Por tanto 20 loci por especie pasaron a la 2ª fase de laboratorio en la que se trato de realizar la unión del fluoróforo al producto amplificado. Se llevo a cabo el perfil de termociclado usado por de Arruda *et al.* (2010) en hasta 35 organismos para cada locus.

Finalmente 8 loci de *Panopea generosa* y 9 loci de *Panopea globosa* mostraron resultados de genotipado existososexitosos . La genotipificación del resto fracasó por diferentes motivos: amplificaciones inespecíficas, intensidades muy débiles, ausencia total de fluorescencia (Tabla 9 y 10).

De los 8 loci de *P. generosa* cinco son microsatélites tetra-, dos tri- y 1 dinucleotídicos. En el caso de *P. globosa* también 5 de ellos son tetra- y 1 di-, pero tres son trinucleotídicos.



**Tabla 8. Descripción de los cebadores de los loci que pasaron a la 2ª fase experimental y el resultado de la misma en *P. generosa*.**

Cebadores	Secuencia	Contig	Ta (°C)	Tamaño producto	Resultado experimental
Pgen2F_1	*GGCTGCCTAAAGTGAATGCC	239	55.6	270	Amplificación débil
Pgen2R_1	AGAAGGGTCATGGTTTCAGC				
Pgen2F_3	*GCGTTTGATTGCRGGTGTAT	2600	55.6	147	34 individuos genotipados
Pgen2R_3	CAGGCATCGTCGTGTAATGG				
Pgen2F_6	*CGAGTTACGCACAAAAGCGT	7376	62	225	Amplificación débil
Pgen2R_6	AGGTAAGACACTGCAAACACAA				
Pgen2F_7	*AAACCAGTCTCGCTCTTCCG	8230	55.6	222	Inespecífico
Pgen2R_7	ACGTCCAGACCTTCAAGCATC				
Pgen3F_1	*AGTGCATCAAGACAACCTGCC	1429	55.6	334	Inespecífico
Pgen3R_1	CGCTCAAGCTAAGCACTTGAT				
Pgen3F_2	*CGTCTTGCGGTTACATTCC	2687	66	292	Inespecífico
Pgen3R_2	GTTGGTTTGACGTCATGTGC				
Pgen3F_4	*ACGGCGAAAGAACGAAATGG	33800	55.6	318	17 individuos genotipados
Pgen3R_4	TTGGTGAGAGTTGTTGCAG				
Pgen3F_6	*CAGAAGTCYTTTTGTTGCATGCTT	52178	66	256	Amplificación débil
Pgen3R_6	ACCCATTGTCAGTGATAACGC				
Pgen3F_7	*GACAAACACCGCCTACACTG	54945	66	357	29 individuos genotipados
Pgen3R_7	TACGAATGCAGTCACCAAGC				
Pgen3F_8	*ACGTTTCATAGGAAGGCCGAC	59441	55.6	199	Inespecífico
Pgen3R_8	ACGTTGAGGACAAGCTTTGTTA				
Pgen3F_9	*ATATCGTCCGGGTTGCAGAC	65389	55.6	331	Inespecífico
Pgen3R_9	TGTGTGTCGGCTAACCTTCC				
Pgen3F_10	*TGCTACTACCCTGCATCACC	72902	55.6	220	Sin amplificación
Pgen3R_10	TTAGAAGACGAGGACCCAGG				
Pgen4F_1	*GSGTGGAATCCATTGGGGTA	4478	62	330	24 individuos genotipados
Pgen4R_1	ACCACCCTGGACACTCCTTA				
Pgen4F_2	*CATCTCCACAATGACGGGTC	7557	55.6	350	Amplificación negativa
Pgen4R_2	AAGCGCGTGTTAGTAGTACC				
Pgen4F_3	*GTTTGCCTTGTGCTCTGCAG	8886	62	220	35 individuos genotipados
Pgen4R_3	GGATCCCTGGAAAGTGTGGT				
Pgen4F_4	*TCTTCCTTTCTGTTGGCGTTG	9631	55.6	264	No incorporación
Pgen4R_4	CACACTGGTTGAGATCACTGG				
Pgen4F_7	*AACCATGACCTGAAGAACACA	11080	55.6	183	Inespecífico
Pgen4R_7	ACTTGTTAAGAGGGTTTTCCCA				
Pgen4F_9	*GTCAATCCAGCCAAGCACAG	14808	55.6	318	33 individuos genotipados
Pgen4R_9	GCGTGTTAGCCCTCAATAGC				
Pgen4F_10	*AACCGCAGCAGAACAAAGTC	17275	56.6	330	28 individuos genotipados
Pgen4R_10	ATCTTCGCTTAGGAGGCGG				
Pgen4F_11	*AAGTCAACCAGGATGTGCAC	21240	66	226	32 individuos genotipados
Pgen4R_11	CCATTAAAGGGTCACACGGC				

**Tabla 9. Descripción de los cebadores de los loci que pasaron a la 2ª fase experimental y el resultado de la misma en *P. globosa*.**

Cebadores	Secuencia	Contig	Ta (°C)	Tamaño producto	Resultado experimental
Pglo2F_2	*ACATCTGCTTCGCCAAACAC	596	62	199	Inespecífico
Pglo2R_2	AGAGCAGAGACGTTGCAGTC				
Pglo2F_3	*CATTACCCTACCCACATAACCC	3944	62	162	Inespecífico
Pglo2R_3	TGTTCAAGATATCGGAACTTGACG				
Pglo2F_5	*ATGTGGTATCTTTGCACGGG	7468	55.6	211	Amplificación débil
Pglo2R_5	AGTTTACCGATTTTCACGGCT				
Pglo2F_6	*GCCACATCAACTGAACCC	9716	66	269	22 individuos genotipados
Pglo2R_6	ACGAAACAACCAACATCGTCA				
Pglo3F_2	*ACGGAATTAGGAATATGCGCAC	17557	63	234	29 individuos genotipados
Pglo3R_2	ACCGAATATCACACACCATGC				
Pglo3F_3	*GGCCATCTCCCTTCTCTGAG	18265	60	256	Amplificación débil
Pglo3R_3	TGATCCTGGAGTTCTCGCTG				
Pglo3F_4	*GGTTTCATCGGCCACTGAATG	21038	62	244	21 individuos genotipados
Pglo3R_4	AACTCAAACACTGTCCTGCC				
Pglo3F_5	*TTCCCTGTATCGCGTTATGG	21175	62	328	21 individuos genotipados
Pglo3R_5	ACATTGCAGTTTGATTGGGC				
Pglo3F_8	*TCCAAAGGTCATGTGATTTCCG	93377	63	262	Interpretación dudosa
Pglo3R_8	TGAGATCACACAGTTCAAGCAGA				
Pglo3F_9	*TCGTTGGTAAATGGCGTTCC	282886	55.6	330	Amplificación débil
Pglo3R_9	ACAACAGCAGATCTGGACC				
Pglo3F_10	*TCTGTAGCTGATGACTGGGC	333069	55.6	187	Amplificación débil
Pglo3R_10	CGCATTAAATAGCGTCGAAACTC				
Pglo4F_1	*CTCTCGCCGGAATTGGTATG	1966	55.6	250	20 individuos genotipados
Pglo4R_1	TCACAACCCCACTAAACCGC				
Pglo4F_2	*AACAGCGTTAATTGAGGGCG	2149	68	376	28 individuos genotipados
Pglo4R_2	ATGATCGTGAAATCGCCAGC				
Pglo4F_3	*TGCCATTGCTATCACGCAT	43405	55.6	263	28 individuos genotipados
Pglo4R_3	AAGCACGATCATTACTCCGG				
Pglo4F_4	*ACTGGCCTTTCAAACCTTCC	44360	55.6	164	Inespecífico
Pglo4R_4	TGGCGACATTGACCATTCTG				
Pglo4F_5	*GTCTGTCAGCCCTCAATAGC	95135	66	246	15 individuos genotipados
Pglo4R_5	TCACTGCTCATGCATCCTTCA				
Pglo4F_6	*GTGAAAATTGCACAACGATAATGT	102948	66	346	23 individuos genotipados
Pglo4R_6	GCCACCACAATGGTAAGCTG				
Pglo4F_8	*ACCTGCCGTCATGTTAACTG	208457	62	393	Amplificación débil
Pglo4R_8	GCACACGTCCGAATCTCAAG				
Pglo4F_9	*ATAGATCGAGAAGCAGCCGG	241735	55.6	301	Amplificación débil
Pglo4R_9	TCCGCTCTCTTCTCAGTCA				
Pglo4F_10	*CCTTACAGGTGTTTGCCGT	252310	55.6	151	Sin amplificación
Pglo4R_10	CCGTCTGCAGCTGGTAATCC				

### 3.2.2. Caracterización

En los 8 loci de *P. generosa* encontramos de 5 a 22 alelos y en los 9 loci de *P. globosa* encontramos de 3 a 16 alelos. El locus Pgen4F\_9 presentó el mayor valor de heterocigosidad observada ( $H_o=0.818$ ) y el locus Pglo3F\_5 el de menor valor ( $H_o=0.286$ ). Los loci Pgen3F\_7, Pgen4F\_1 y Pgen4F\_11 no se encontraron en equilibrio de H-W debido a un déficit de heterocigotos tal y como muestran los valores de heterocigosidad observada y esperada para cada uno de estos locus. Lo mismo ocurrió en 5 loci de *P. globosa*: Pglo3F\_2, Pglo3F\_5, Pglo4F\_2, Pglo4F\_3 y Pglo4F\_5. El análisis de desequilibrio en el ligamiento indicó ligamiento entre los loci Pgen4F\_1 y Pgen4F\_10 así como entre los loci Pglo2F\_6 y Pglo4F\_2, entre Pglo3F\_2 y Pglo4F\_6 y entre Pglo4F\_5 y Pglo4F\_6. Los valores de PIC van de 0.391 a 0.939.

Tabla 10. Características de los 8 loci polimórficos de *P. generosa*.

Locus	Motivo	n	N <sub>a</sub>	Rango (pb)	H <sub>o</sub>	H <sub>E</sub>	PHWE	PIC
Pgen2F_3	(AT)8	34	5	159-167	0.559	0.673	0.311	0.6
Pgen3F_4	(ACG)10	17	6	316-337	0.529	0.617	0.223	0.559
Pgen3F_7	(AAC)17	29	15	345-393	0.552	0.912	0.000*	0.888
Pgen4F_1	(ACAG)14	24	22	314-474	0.667	0.962	0.000*	0.939
Pgen4F_3	(AAAC)6	35	7	236-320	0.429	0.548	0.141	0.488
Pgen4F_9	(AATC)9	33	13	281-381	0.818	0.891	0.045	0.866
Pgen4F_10	(ACGC)6	28	13	345-409	0.724	0.823	0.039	0.786
Pgen4F_11	(ATCC)8	32	10	242-278	0.625	0.848	0.002*	0.816

n: Tamaño de muestra; N<sub>a</sub>: n° de alelos; H<sub>o</sub>: Heterocigosidad observada; H<sub>E</sub>: Heterocigosidad esperada; P<sub>HWE</sub>: P valor de la prueba del equilibrio de Hardy-Weinberg, con asterisco los resultados significativos tras el ajuste de Bonferroni (p-valor < 0.006); PIC: contenido de información polimórfica.

Tabla 11. Características de los 9 loci polimórficos de *P. globosa*.

Locus	Motivo	n	Na	Rango	H <sub>O</sub>	H <sub>E</sub>	PHWE	PIC
Pglo2F_6	(AT)8	24	3	280-288	0.583	0.504	0.390	0.391
Pglo3F_2	(AAC)10	29	16	239-341	0.517	0.834	0.000*	0.808
Pglo3F_4	(AAT)16	21	5	236-248	0.429	0.587	0.010	0.538
Pglo3F_5	(AAC)9	21	8	336-378	0.286	0.827	0.000*	0.781
Pglo4F_1	(AATG)8	20	7	260-328	0.65	0.7	0.600	0.629
Pglo4F_2	(ACGC)8	28	14	317-445	0.607	0.906	0.000*	0.881
Pglo4F_3	(AAAT)7	28	6	261-285	0.464	0.747	0.003*	0.688
Pglo4F_5	(AATC)7	15	7	247-275	0.533	0.809	0.029	0.75
Pglo4F_6	(ACGC)8	23	10	349-385	0.565	0.886	0.000*	0.852

n: Tamaño de muestra; N<sub>a</sub>: nº de alelos; H<sub>O</sub>: Heterocigosidad observada; H<sub>E</sub>: Heterocigosidad esperada; P<sub>HWE</sub>: P valor de la prueba del equilibrio de Hardy-Weinberg, con asterisco los resultados significativos tras el ajuste de Bonferroni (p-valor < 0.005); PIC: contenido de información polimórfica.

### 3.3. Anotación

#### 3.3.1. Anotación general de los contigs

Los resultados arrojados por el programa FASTANNOTATOR muestran que alrededor del 5% de los contigs totales fueron anotados en ambas especies y también coinciden en la cantidad de homologías encontradas en cada una de las bases de datos (Fig.8).

Se anotaron 13 y 11 contigs en *P. generosa* y *P. globosa* respectivamente de los 30 contigs de cada especie donde se encuentran los loci seleccionados. En la mayoría de los casos son anotaciones que corresponden a proteínas hipotéticas.

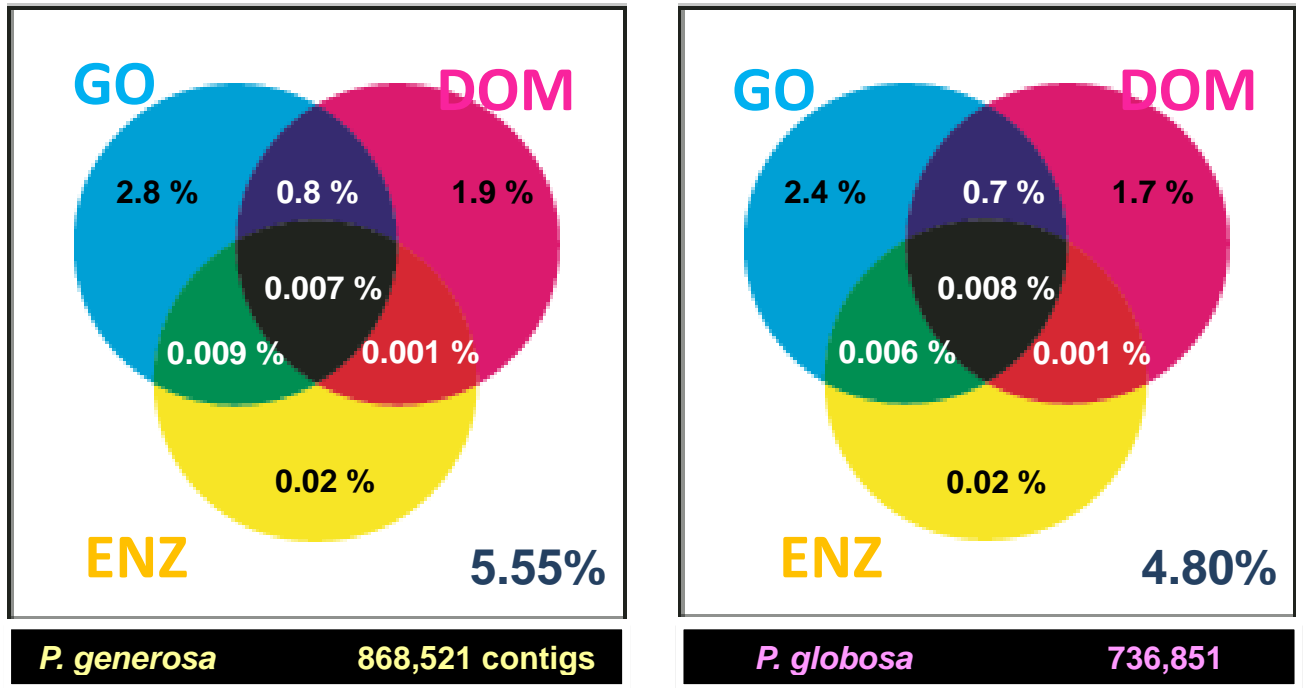


Figura 8. Anotación genética de los contigs totales de *P. generosa* y *P. globosa*. Bases de datos: GO: Ontología genética; DOM: Dominio y ENZ: Enzima.

### 3.3.2. Anotación mitocondrial

El mitogenoma de *P. generosa* se detectó en el contig 1715 y tiene una longitud de 15,585 pb con una composición de bases: 25.0% A, 38.7% T, 11.2% C y 25.0% G (Genbank accession KM580067) en cambio en *P. globosa*, detectado en el contig 899, la longitud es ligeramente menor con 15,469 pb y una composición de bases muy similar: 23.3% A, 40.4% T, 10.1% C and 26.1% G (Genbank accession KM580068). Ambos mitogenomas son muy similares en cuanto a la composición y localización de los genes y los codones de inicio y parada (Figs. 8 y 9). Los dos contienen 13 genes estructurales, 2 ARN ribosomales y 22 ARN de transferencia. Además en 11 de los 13 genes estructurales el codón de inicio es ATG o ATA y sólo en COX1 y Cytb se encontró un codón de inicio alternativo, GTG. Asimismo 8 de los genes estructurales terminan con TAG y 5 con TAA. La región no codificante de mayor longitud se encuentra entre los genes COX1 y COX2 en los dos genomas mitocondriales, siendo de 438 pb en *P. generosa* y 316 pb en *P. globosa*.

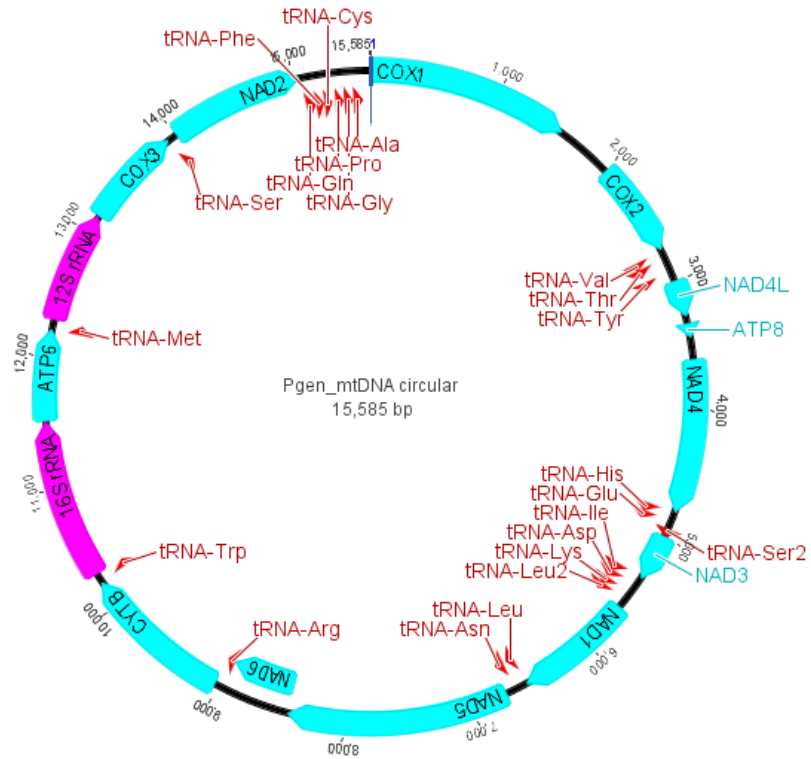


Figura 9. Mitogenoma de *Panopea generosa*.

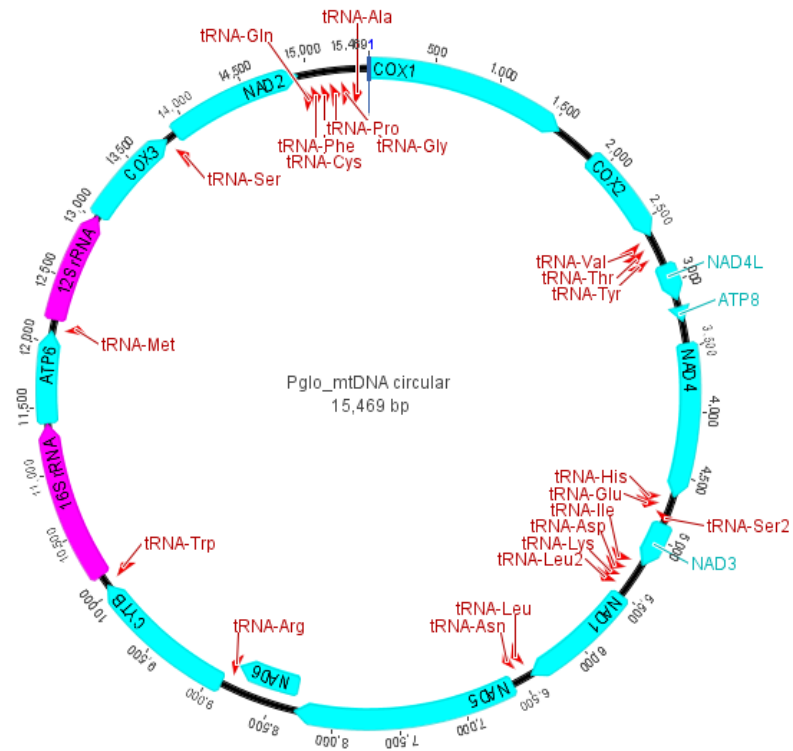


Figura 10. Mitogenoma de *Panopea globosa*.

## 4. Discusión

### 4.1. Análisis de las secuencias de ambas especies provenientes de la NGS

La cantidad de secuencias que se obtuvo para cada uno de los organismos secuenciados es similar, excepto para el organismo Pglo SF02 del cual se consiguió menos del 1% de las secuencias que se alcanzaron en cada uno de los otros organismos. Esto pudo ser resultado fortuito de alguna etapa del proceso de secuenciación, ya que la calidad y cantidad de ADN que se mandó fue igual en todos los organismos de *Panopea* y en el proceso de preparación de la librería no se reportó ninguna anomalía. Además, la secuenciación se realizó al mismo tiempo para todos los organismos por lo que tampoco puede atribuirse a un fallo en este proceso.

Las secuencias obtenidas tuvieron valores de calidad dentro de los márgenes de “buena calidad” que los dos programas utilizados definen. Esta buena calidad queda reflejada en el proceso de trimming, ya que aproximadamente el 90% de las secuencias de la misma especie no sufrieron ninguna alteración (ni modificadas, ni descartadas).

En el proceso de ensamblaje se observa que poco más de un cuarto de las secuencias no son ensambladas por lo que en este paso se pierde una parte sustancial de información. En otros estudios quedaron sin ensamblar cantidades mayores de secuencias, como un 40% (Parchman *et al.*, 2010); y hasta 74% (An & Lee, 2012). El proceso de ensamblar es una manera de eliminar secuencias repetidas y para el fin de este trabajo aumenta las posibilidades de encontrar loci microsatelitales con regiones flanqueantes de buena calidad para el diseño de cebadores (An & Lee, 2012; Silva *et al.*, 2013). Otro dato importante es la cantidad de secuencias pareadas que se encuentran en el mismo contig, esto hace más confiable el ensamblaje de dicho contig, ya que las secuencias pareadas son cercanas en el genoma y deberían situarse cerca una de la otra en la reconstrucción.

## **4.2. Identificación de loci microsatelitales y selección de los loci candidatos a marcadores moleculares.**

### **4.2.1. Elección de la plataforma de NGS**

La elección de la plataforma de NGS es un detalle importante según el objetivo de la investigación ya que no todas ofrecen los mismos resultados. En un principio, la plataforma 454 al ofrecer secuencias de mayor longitud que las otras pareció la idónea para el desarrollo de marcadores microsatelitales, ya que es más probable que exista suficiente región flanqueadora para la síntesis de los cebadores.

Sin embargo, algunos investigadores probaron con la plataforma Illumina con resultados igual de buenos que con 454 y con un costo significativamente menor por base secuenciada. Nunziata *et al.* (2012) y O'Bryhim *et al.* (2012), entre otros, usaron la opción paired-end de la plataforma Illumina y obtuvieron 42 y 16 marcadores microsatélites polimórficos para sus especies de estudio, respectivamente.

Rápidamente se despierta la duda de cuál plataforma es más eficaz en términos de resultados obtenidos, costo y tiempo para el aislamiento de microsatélites que puedan amplificarse. Castoe *et al.* (2012) comparó la eficacia para encontrar loci microsatelitales y cebadores flanqueadores entre ambas plataformas partiendo del mismo número de secuencias y un número similar de megabases en una especie de serpiente, *Python molurus bivittatus*. Como se puede observar en la tabla 12 los resultados fueron muy similares en ambas plataformas.



**Tabla 12. Comparación de los resultados obtenidos en dos plataformas diferentes de NGS, 454 e Illumina (IPE-119K), para la especie *Python molurus bivittatus*. \*PALs: potentially amplifiable microsatellite loci.**

<b>Sample set</b>	<b>Burmese Python</b>	
	<b>454</b>	<b>IPE-119K</b>
Millions of reads	0.119	0.119
Megabases of sequence	28.5	27.1
Reads containing one or more microsatellites	11,027	11,073
Total individual microsatellite loci	13,142	12,833
Compound loci	1,314	973
Mirosatellite reads per megabase of sequence	386.9	408.6
Discrete PALs	5,474	4,129
Discrete PAL rate	0.496	0.373

Lance *et al.* (2013) realizaron un estudio en el que analizaron la validez de la plataforma Illumina para obtener marcadores microsatelitales en 32 especies de diferentes taxones. Ellos además de presentar la cantidad de los loci microsatelitales potencialmente amplificables probaron la eficacia de éstos en 13 de estas especies. Los resultados positivos de las amplificaciones varían entre taxones, pero van desde un 49% a un 67%, y con valores posiblemente más altos si se empleará más esfuerzo en la optimización de algunos.

Recientemente se han valorado otras plataformas para la obtención de microsatélites. Elliott *et al.* (2014) hizo un estudio con plantas para comparar la plataforma 454 con la plataforma Ion Torrent (PMG). Esta última difiere de 454 e Illumina, ya que ambas detectan los nucleótidos mediante imágenes de fluorescencia, y la Ion Torrent lo hace mediante cambios de pH y por eso se habla de tecnología semiconductor. Con PMG se obtienen 10 veces más Gigabases por corrida que con 454, pero ésta ofrece secuencias de mayor longitud. Concordando con esto, en este estudio se obtuvieron más secuencias con PMG que con 454, pero de menor longitud. En cuanto a los microsatélites que se identificaron en cada caso, existen algunas diferencias en cuanto al número y el tipo de ellos, pero en definitiva en ambas plataformas se obtuvo un gran número de secuencias con microsatélites donde se pueden diseñar primers. Por otro

lado, al comparar los costos y el tiempo de corrida, ambos son menores en PMG que en 454. Por lo que parece que la plataforma PMG también puede ser una elección efectiva para el desarrollo de marcadores microsatelitales.

#### **4.2.2. Selección de los loci candidatos a marcadores moleculares**

En este estudio la cantidad de loci de interés identificados para las tres categorías (di-, tri- y tetranucleótidos) fue alta en ambas especies. Aunque los candidatos a marcadores moleculares quedan reducidos considerablemente, debido a que a menos de una cuarta parte de los loci totales se les pudo diseñar cebadores, y son estos los realmente útiles como marcadores moleculares, denominados por Castoe *et al.* (2010) como PAL 'potentially amplifiable loci'. En estudios similares también se reporta esta diferencia entre los loci identificados y los loci con primers diseñados (Castoe *et al.*, 2010; Csencsics *et al.*, 2010). Como ya se comentó en la sección anterior, la posibilidad de obtener más PAL será mayor cuanto más larga sea la secuencia en las que se realiza la búsqueda de loci, por lo que en este caso dependerá de la longitud de los contigs ensamblados.

Sin embargo, es importante tener en cuenta que estos resultados dependen en gran medida de los parámetros de búsqueda de los loci microsatelitales así como de los parámetros para el diseño de cebadores que se le determina a los programas informáticos (Abdelkrim *et al.*, 2009). Además, hay que tener en mente que más crítico que la cantidad es la calidad de estos PAL, ya que no debe subestimarse el esfuerzo requerido para el desarrollo de los marcadores moleculares una vez ya se han identificado. Squirrell *et al.* (2003) reportaron que del total de PAL sólo se obtiene el 50% de loci polimórficos con productos confiables de PCR. Por lo tanto, unos criterios de selección estrictos pueden asegurar mejores resultados en la fase de optimización de la condiciones de amplificación por PCR (Lepais & Bacles, 2011). En este último punto una ventaja sería seleccionar PAL que ya se supiera que son polimórficos a priori. Se ha visto que cuanto mayor es el número de repeticiones de un locus mayor es la tasa de mutación del mismo (Ellegren, 2004; Kelkar *et al.*, 2008). Aunque loci con un gran número de repeticiones también pueden presentar una mayor probabilidad de alelos nulos y/o stutters (Guichoux *et al.*, 2011). En este estudio se observa que los loci

seleccionados con mayor número de repeticiones sí mostraron en la mayoría de los casos una mayor cantidad de alelos (Tablas 11 y 12). Por otro lado, si la NGS se realiza con más de un organismo y con una gran cobertura podría observarse el polimorfismo en las mismas secuencias y realizar así bajo este criterio la selección de los loci candidatos (Guichoux *et al.*, 2011).

### **4.3. Optimización de las condiciones de amplificación por PCR de los loci microsatelitales candidatos y evaluación de su nivel de polimorfismo**

De los 30 loci que se probaron experimentalmente sólo 8 para *P. generosa* y 9 para *P. globosa* mostraron una amplificación confiable así como polimorfismo. Por lo que el rendimiento obtenido fue de un 27% y un 30% respectivamente para cada especie. En el estudio de Cruz-Hernández *et al.* (2014) en el que se desarrollaron 24 marcadores microsatelitales para *P. globosa* siendo 68 loci los que se probaron experimentalmente se consigue un rendimiento del 35%. En otros bivalvos se obtienen resultados con un éxito del 22% (Greenley *et al.*, 2011), 33% (O'Bryhim *et al.*, 2012), 43% (An & Lee, 2012) y 71% (Inoue *et al.*, 2012). Como se ve en la mayoría de los estudios una parte de los loci serán descartados por problemas de amplificación, por lo que es importante tener esto en cuenta a la hora de iniciar una investigación. Además, ciertos grupos taxonómicos entre los que se incluyen los bivalvos son más propensos a problemas de amplificación (Selkoe & Toonen, 2006).

Existen varias complicaciones que pueden impedir la correcta amplificación. Para empezar algunos loci con primers diseñados aparentemente apropiados no amplifican o lo hacen levemente (Squirrell *et al.*, 2003). El otro inconveniente suele ser la presencia de bandas que no corresponden, lo que impide la correcta interpretación de los alelos del locus en cuestión. Esto puede ocurrir porque los primers se anillan en sitios alternos además de en el lugar para el que se diseñaron, lo que se conoce como "Mispriming".

En este trabajo 10 loci de cada especie no pasaron a la 2ª fase experimental por diferentes razones: nunca se obtuvo un producto amplificado o la amplificación fue muy

pobre como para considerarse convincente, en otros casos la amplificación fue inespecífica y por último, varios de estos loci no fueron seleccionados porque no se detectó polimorfismo. De hecho en el conjunto final de microsatélites genotipados predominan los del tipo tetranucleótido y los dinucleótidos están en menor cantidad. Esto puede deberse a que fue más fácil detectar polimorfismo cuanto mayor era el motivo repetitivo del microsatélite. Por lo que es posible que algunos de estos microsatélites dinucleótidos sí sean polimórficos y pudieran en un futuro acabar de evaluarse para llegar a ser marcadores moleculares.

En la 2ª fase experimental el cambio de perfil de termociclado y técnica de PCR para la incorporación del fluoróforo obstaculizó la obtención de las amplificaciones de los organismos en cada loci. Realmente en varios loci no se consiguieron suficientes amplificaciones positivas para mandar a genotipificar. La cola M13 parece decrecer la eficiencia del PCR (Guichoux *et al.*, 2011); y es por eso que se usan perfiles de termociclado con un mayor número de ciclos (Schuelke, 2000; de Arruda *et al.*, 2010) .

Otra criba se realizó con el análisis de los genotipos ya que se detectaron varios fallos: loci que no incorporaron el fluoróforo, otros con bandas artefacto así como patrones tri o tetra alélicos en un mismo organismo, lo cual es imposible ya que estas especies son diploides. Estos últimos podrían deberse, de nuevo, al mispriming de los primers, pero otra posible causa es la asociación de estos microsatélites a elementos móviles del genoma como transposones que generarán la duplicación de regiones (Fernandez-Silva *et al.*, 2013). Por lo que en este caso si los cebadores han sido diseñados en regiones duplicadas por el genoma se amplificarán al mismo tiempo varios loci generando patrones multialélicos. Esto ya se ha reportado en diferentes taxones incluyendo especies de moluscos (McInerney *et al.*, 2011).

Finalmente, es importante tener en cuenta, antes de cualquier comparación entre los loci genotipados con éxito, que los análisis de cada locus tienen un tamaño de muestra diferente. Sin embargo, en cuanto al número de alelos encontramos el valor más bajo (3) y más alto (22) en dos loci que presenta el mismo número de organismos analizado (24). Lo cual indica la variabilidad encontrada en este carácter independientemente del

tamaño de la muestra. La relación entre el número de repeticiones del motivo microsatelital y el número de alelos sí se cumplió en su mayoría, por ejemplo el locus Pgen3F\_7 con 17 veces repetido el motivo reveló 15 alelos. En cambio hay casos que no siguen esta pauta como el locus Pglo3F\_4 con un motivo repetido 16 veces y sólo 5 alelos.

La calidad de un marcador genético se puede medir por su heterocigosidad y el valor de PIC (Shete *et al.*, 2000). Una diferencia significativa entre la heterocigosidad observada y la esperada para un marcador molecular nos está indicando un déficit de organismos heterocigotos. Este déficit en estudios de genética poblacional puede deberse a fenómenos como endogamia, cuello de botella o efecto Wahlund (Selkoe & Toonen, 2006). Este último ocurre cuando se tratan dos grupos genéticamente distintos como una misma población, lo que podría estar resultando en este estudio, ya que en los análisis de los loci se incluyeron organismos de diferentes localidades, algunas muy alejadas entre sí, que es posible que se diferencien genéticamente. Sin embargo, también es posible que la causa provenga de caracteres del mismo marcador. Existen varias causas que conllevan al déficit de heterocigotos, como puede ser la presencia de alelos nulos, deserción del alelo mayor o la homoplasia. En esta tesis los valores de heterocigosidad observada presentan un amplio rango (0.286- 0.818), pero en todos los loci menos uno, son menores a los valores de heterocigosidad esperada calculados. En los casos en que esta diferencia fue significativa, los loci se desviaron del equilibrio de Hardy-Weinberg, sin embargo aunque un locus no se ajuste al equilibrio de Hardy-Weinberg esto no es motivo suficiente para desecharlo (Selkoe & Toonen, 2006). En estudios anteriores donde se aislaron loci microsatelitales en estas especies se obtuvieron también marcadores que no lo cumplían (Vadopalas & Bentzen, 2000; Kaukinen, 2004; Cruz-Hernández *et al.*, 2014). En bivalvos estos resultados se atribuyen a mutaciones en la regiones de unión para el primer ya que se ha reportado una incidencia alta de alelos nulos en estos taxones (Hedgecock *et al.*, 2004; Becquet *et al.*, 2009).

El valor de PIC depende tanto del número de alelos así como de la frecuencia de los mismos en la población (Botstein *et al.*, 1980). Los valores de PIC obtenido nos indican

que la mayoría de los loci, son altamente informativos, es decir que son eficaces para detectar polimorfismo en una población.

Los resultados del test de desequilibrio de ligamiento indican que dos loci de *P. generosa* se segregan conjuntamente y en *P. globosa* se encontraron 3 parejas de loci que no eran independientes. Es muy importante que los marcadores moleculares que se usen en un estudio sean independientes, ya que de lo contrario se crea pseudo-replicación de loci en los análisis. Por lo que si el test de desequilibrio de ligamiento es significativo entre dos loci, uno de ellos debería ser descartado (Selkoe & Toonen, 2006a).

En definitiva, este estudio suma nuevos marcadores microsatelitales polimórficos así como reporta otros que podrían llegar a serlo, además de la gran cantidad de loci con primers diseñados que han sido identificados. Por lo que se cuenta con un mayor número de marcadores moleculares para cada una de las especies que servirán para estudiar los patrones de conectividad entre poblaciones, identificar “stocks” pesqueros, estimar el tamaño poblacional efectivo y emplear en programas de trazabilidad genética, información que debe incorporarse en los planes de manejo del recurso para definir el rendimiento máximo sostenible que garantice la conservación de ambas especies.

#### **4.4. Anotación general de los contigs**

La cantidad de contigs anotados fue de sólo un 5% del total, seguramente es consecuencia de las bases de datos que usa el programa informático. Existen menos genomas de invertebrados no artrópodos anotados completamente que de otros grupos como vertebrados e invertebrados artrópodos

([http://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/all/](http://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/)).

Cabe destacar que la información reunida en el aproximadamente 3% de los contigs anotados en la “Gene Ontology” va más allá que una anotación básica, ya que determina la distribución de los genes encontrados en tres categorías funcionales, de tal

manera que controla el vocabulario y estandariza las funciones de los genes en dichas categorías facilitando la comparación de anotaciones genómicas entre especies.

Sin embargo, hay que vigilar la propagación de anotaciones erradas, fruto de genes mal categorizados en las bases de datos que servirán como molde para otras anotaciones propagando en ese caso el error.

La importancia de la anotación de los contigs que contienen los loci seleccionados se debe a la posible relación entre la presencia de locus no polimórficos en contigs anotados, ya que aunque en su mayoría los loci microsatelites se encuentran en regiones neutras, puede ser que en ciertos casos se encuentren en regiones codificantes o adyacentes a éstas por lo que la tasa de mutación será inferior (Zane *et al.*, 2002). En este trabajo no se reportan indicios de esta relación, de hecho varios de los loci polimórficos (Pgen2F\_3, Pgen3F\_4, Pgen3F\_7, Pgen4F\_1, Pglo3F\_2, Pglo3F\_4, Pglo4F\_1) calificados como posibles marcadores moleculares y con un número de alelos considerable se encuentran en contigs anotados. Hay que tener en cuenta que la longitud de estos 7 contigs anotados es en la mayoría cercana o superior a 1000pb, por lo que es posible que la región que anotó el software quede fuera del rango del loci microsatelital.

#### **4.5. Anotación mitocondrial**

La anotación mitocondrial de ambas especies es una fuente más de marcadores moleculares. Los marcadores mitocondriales se aplican generalmente en estudios de filogeografía ya que el ADN mitocondrial presenta las características idóneas para este tipo de estudios: no recombina, tiene una tasa alta de mutación y es de herencia matrilineal (Gillespie, 1986; Harrison, 1989). Además permiten deducir cambios demográficos y de dispersión entre especies (Di Rienzo & Wilson, 1991; Godoy *et al.*, 2004).

En este estudio se obtuvo la anotación completa del ADN mitocondrial de ambas especies de *Panopea* y se publicaron dos notas científicas: "The complete mitochondrial DNA of the Pacific Geoduck clam (*Panopea generosa*)" (Bisbal-Pardo *et al.*, 2014) y

“Novel gene arrangement in the mitochondrial genome of the Corté’s Geoduck clam (*Panopea globosa*)” (Bisbal-Pardo *et al.*, 2014) en la revista *Mitochondrial DNA* con el fin de contribuir a la divulgación científica y añadir nuevos mitogenomas a los ya publicados para la Subclase Heterodonta de bivalvos (Dreyer & Steiner, 2006; Ren *et al.*, 2009; Meng *et al.*, 2012; Yuan *et al.*, 2012).

Los mitogenomas de ambas especies son muy parecidos, en cambio difieren significativamente de los mitogenomas de otros bivalvos que pertenecen a la Subclase Heterodonta, en concreto en el orden de los genes (Bisbal-Pardo *et al.*, 2014). Se ha observado que el orden de los genes es muy variable en los bivalvos heterodontos por lo que esta información es muy útil para aclarar la relaciones filogenéticas entre estos (Dreyer & Steiner, 2006; Yuan *et al.*, 2012).



## 5. Resultados significativos

Se obtuvieron aproximadamente 77 millones de secuencias de 101pb de buena calidad tanto para *Panopea generosa* como *Panopea globosa* a partir de secuenciación de siguiente generación, usando la plataforma Illumina.

Se ensamblaron alrededor del 66% de las secuencias de cada una de las especies por separado formando 868,521 mil y 736,851 mil contigs para *P. generosa* y *P.globosa* respectivamente con una longitud media de 400pb.

En los contigs de cada especie se identificaron tres tipos de microsatélites: 8,060 di-, 3,146 tri- y 2,830 tetranucleótidos en *P. generosa*; 4,956 di-, 1,647 tri- y 1,983 tetranucleótidos en *P. globosa*. Sólo fue posible diseñar cebadores a menos del 26% de los microsatélites detectados en todos los casos.

Se optimizaron 8 y 9 loci polimórficos en *P. generosa* y *P. globosa* respectivamente, que complementan a los marcadores moleculares existentes para realizar estudios de genética poblacional, entre otros posibles.

Se anotó alrededor del 5% de los contigs de cada especie. No se pudo encontrar una relación entre los loci seleccionados que se ubicaban en un contig anotado con su resultado experimental.

Se anotaron los genomas mitocondriales completos de cada una de las especies, mostrando un orden de los genes similar entre sí pero diferente al de otras especies pertenecientes a la subclase Heterodonta.

Por último, destacar que las secuencias en bruto, los contigs formados, los microsatelites identificados y los contigs anotados son un recurso de información genética que puede ser aprovechado para incrementar el conocimiento sobre las poblaciones de estas especies. Toda esta información esta almacenada en un disco duro externo en el Laboratorio de Ecología Molecular del CICESE bajo la tutela del Dr. Axayácatl Rocha-Olivares.

## Lista de Referencias

- Abdelkrim, J., Robertson, B., Stanton, J. A. and Gemmell, N. (2009). Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, 46(3), 185–92.
- Altschup, S. F., Gish, W., Pennsylvania, T. and Park, U. (1990). Basic Local Alignment Search Tool. *Journal of molecular biology*, 215(3), 403-410.
- Amos, W., Hoffman, J. I., Frodsham, A., Zhang, L., Best, S. and Hill, A. V. S. (2007). Automated binning of microsatellite alleles: Problems and solutions. *Molecular Ecology Notes*, 7(1), 10–14.
- An, H. S. and Lee, J. W. (2012). Development of Microsatellite Markers for the Korean Mussel, *Mytilus coruscus* (Mytilidae) Using Next-Generation Sequencing. *International Journal of Molecular Sciences*, 13(8), 10583–93.
- Andrews, S. (2011). FastQC High Throughput sequence QC report v.0.10.1. Babraham Bioinformatics. Available at: <http://www.bioinformatics.babraham.ac.uk>.
- Aragón-Noriega, E. A., Alcántara-Razo, E., Calderon-Aguilera, L. E. and Sánchez-Fourcade, R. (2012). Status of Geoduck Clam Fisheries in Mexico. *Journal of Shellfish Research*, 31(3), 733–738.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M.A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25-29,
- Bahassi, E. M. and Stambrook, P. J. (2014). Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis*, 29(5), 303–310.
- Becquet, V., Lanneluc, I., Simon-Bouhet, B. and Garcia, P. (2009). Microsatellite markers for the Baltic clam, *Macoma balthica* (Linné, 1758), a key species concerned by changing southern limit, in exploited littoral ecosystems. *Conservation Genetics Resources*, 1(1), 265–267.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., Pütz, J., Middendorf, M. and Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2), 313–9.
- Bisbal-Pardo, C. I., Del Río-Portilla, M. A., and Rocha-Olivares, A. (2014). The complete mitochondrial DNA of the Pacific Geoduck clam (*Panopea generosa*). *Mitochondrial DNA*, 1736(646), 1–2.

- Bisbal-Pardo, C. I., Del Río-Portilla, M. A. and Rocha-Olivares, A. (2014). Novel gene arrangement in the mitochondrial genome of the Cortés Geoduck clam (*Panopea globosa*). *Mitochondrial DNA*, 1736(646), 1–2.
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3), 314–331.
- Bureau, D., Hajas, W., Surry, N. W., Hand, C. M., Dovey, G. and Campell, A. (2002). Age, Size Structure and Growth Parameters of Geoducks (*Panopea abrupta*, Conrad 1849) from 34 Locations in British Columbia Sampled Between 1993 and 2000. Canadian Technical Report of Fisheries and Aquatic Sciences, 2413.
- Burton, R. S. (1996). Molecular tools in marine ecology. *Journal of Experimental Marine Biology and Ecology*, 200 (1-2), 85–101.
- Campbell, A., Yeung, G. and Zhang, Z. (2004). Population biology of the pacific Geoduck clam, *Panopea abrupta*, In experimental plots, Southern British Columbia, Canada, *Journal of Shellfish Research*, 23(3), 661-673.
- Castoe, T. a, Poole, A. W., de Koning, a P. J., Jones, K. L., Tomback, D. F., Oyler-McCance, S. J., Fike, J. A., Lance, S. L., Streicher, J. W., Smith, E. N. and Pollock, D. D. (2012). Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS One*, 7(2), e30953.
- Castoe, T. A., Poole, A. W., Gu, W., Jason de Koning, A. P., Daza, J. M., Smith, E. N. and Pollock, D. D. (2010). Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resources*, 10(2), 341–7.
- Chen, T.W., Gan, R.C. R., Wu, T. H., Huang, P.J., Lee, C.Y., Chen, Y.Y. M., Chen, C.C. and Tang, P. (2012). FastAnnotator- an efficient transcript annotation web tool. *BMC Genomics*, 13(7), S9.
- Cruz-Hernández, P., Munguia-Vega, A., Leyva-Valencia, I., Lucero-Burquez, F. and Lluch-Cota, D. B. (2014). Development of 24 tetra-nucleotide microsatellite markers in Cortes Geoduck *Panopea globosa* by next-generation sequencing. *Conservation Genetics Resources*, 6(3), 531–533.
- Csencsics, D., Brodbeck, S., and Holderegger, R. (2010). Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. *Journal of Heredity*, 101(6), 789–793.
- de Arruda, M. P., Gonçalves, E. C., Schneider, M. P. C., da Silva, A. L. D. C. and Morielle-Versute, E. (2010). An alternative genotyping method using dye-labeled universal primer to reduce unspecific amplifications. *Molecular Biology Reports*, 37(4), 2031–6.

- Di Rienzo, A. and Wilson, A. (1991). Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proceedings of the National Academy of Science*, 88(5), 1597–1601.
- Dreyer, H. and Steiner, G. (2006). The complete sequences and gene organisation of the mitochondrial genomes of the heterodont bivalves *Acanthocardia tuberculata* and *Hiatella arctica*--and the first record for a putative Atpase subunit 8 gene in marine bivalves. *Frontiers in Zoology*, 3(13),1-14.
- Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews. Genetics*, 5(June), 435–445.
- Elliott, C. P., Enright, N. J., Allcock, R. J., Gardner, M. G., Megléc, E., Anthony, J. and Krauss, S. L. (2014). Microsatellite markers from the Ion Torrent: a multi-species contrast to 454 shotgun sequencing. *Molecular ecology resources*, 14(3), 554-568.
- Environmental Defense Fund, Mexico. (2013). A value chain analysis of Mexico ' s emerging commercial geoduck trade. CapLog Reports, 1(February), 0–18.
- Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47–50.
- Faircloth, B. C. (2008). Msatcommander: Detection of Microsatellite Repeat Arrays and Automated, Locus-Specific Primer Design. *Molecular Ecology Resources*, 8(1), 92–94.
- Feldman, K., Vadopalas, B., Armstrong, D., Friedman, C., Hilborn, R., Naish, K., Orensanz, J., Valero, J., Ruesink, J., Suhrbier, A., Christy, A., Cheney, D. and Davis, J. P. (2004). *Olympia*, WA: Washington State Department of Natural Resources.
- Fernandez-Silva, I., Whitney, J., Wainwright, B., Andrews, K. R., Ylitalo-Ward, H., Bowen, B. W., Toonen, J. T., Goetze, E. and Karl, S. A. (2013). Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. *PloS One*, 8(2), e55990.
- Gan, H. M., Schultz, M. B. and Austin, C. M. (2014). Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evolutionary Biology*, 14(1), 19.
- Gillespie, J. H. (1986). Variability of evolutionary rates of Dna. *Genetics*, 113(4), 1077-1091
- Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11(5), 759–69.

- Godoy, J. A., Negro, J. J., Hiraldo, F. and Donázar, J. A. (2004). Phylogeography, genetic structure and diversity in the endangered bearded vulture (*Gypaetus barbatus*, L.) as revealed by mitochondrial DNA. *Molecular Ecology*, 13(2), 371–390.
- Greenley, A. P., Muguia-Vega, A., Saenz-Arroyo, A. and Micheli, F. (2011). New tetranucleotide microsatellite loci in pink abalone (*Haliotis corrugata*) isolated via 454 pyrosequencing. *Conservation Genetics Resources*, 4(2), 265–268.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F. and Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources*, 11(4), 591–611.
- Harrison, R. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution*, 4(1), 6-11.
- Hedgecock D., Li G., Hubert, S., Bucklin, K. and Ribes, V.(2004). Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *Journal Shellfish Research*, 23(2),379– 385.
- Inoue, K., Lang, B. K., and Berg, D. J. (2012). Development and characterization of 20 polymorphic microsatellite markers for the Texas hornshell, *Popenaias popeii* (*Bivalvia: Unionidae*), through next-generation sequencing. *Conservation Genetics Resources*, 5(1), 195–198.
- Kaukinen, K. H., Supernault, K. J. and Miller, K. M. (2004). Enrichment of tetranucleotide microsatellite loci from invertebrate species. *Journal of Shellfish Research*, 23(2), 621-626.
- Kelkar, Y. D., Tyekucheva, S., Chiaromonte, F., and Makova, K. D. (2008). The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Research*, 18(1), 30–38.
- Khan, A. (2006). Sustainability Challenges in the Geoduck Clam Fishery of British Columbia: Policy Perspectives. *Coastal Management*, 34(4), 443–453.
- Knudsen, T. B. and Flensburg, M. (2008). CLC Genomics Workbench, 4.0.
- Lance, S. L., Love, C. N., Nunziata, S. O., O'Bryhim, J. R., Scott, D. E., Flynn, R. W. and Jones, K. L. (2013). 32 species validation of a new Illumina paired-end approach for the development of microsatellites. *PloS One*, 8(11), e81853.
- Lepais, O. and Bacles, C. F. E. (2011). Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla* F. Muell. Ex Benth. *Molecular Ecology Resources*, 11(4), 711–724.

- Leyva-Valencia, I. (2012). Diferencias morfométricas en dos especies de la almeja generosa: *Panopea generosa* (gould 1850) y *P. globosa* (dall 1898) y filogenia molecular de cinco especies del género *Panopea*. Tesis de Doctorado en ciencias Uso, Manejo y Preservación de los Recursos Naturales. Centro de investigaciones Biológicas del Noroeste, S. C. 86p
- Loman, N. and Constantinidou, C. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*, 10 (9), 599–606.
- Mardis, E. R. (2006). Anticipating the 1,000 dollar genome. *Genome Biology*, 7(7), 112.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376–80.
- McInerney, C. E., Allcock, A. L., Johnson, M. P., Bailie, D. A. and Prodöhl, P. A. (2011). Comparative genomic analysis reveals species-dependent complexities that explain difficulties with microsatellite marker development in molluscs. *Heredity*, 106(1), 78–87.
- Meng, X., Zhao, N., Shen, X., Hao, J., Liang, M., Zhu, X., Cheng, H., Yan, B. and Liu, Z. (2012). Complete mitochondrial genome of *Coelomactra antiquata* (Mollusca: Bivalvia): The first representative from the family Mactridae with novel gene order and unusual tandem repeats. *Comparative Biochemistry and Physiology. Part D, Genomics & Proteomics*, 7(2), 175–9.
- Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research*, 15(12), 1767–76.
- Miller, K. M., Supernault, K. J., Li, S. and Withler, R. E. (2006). Population Structure in Two Marine Invertebrate Species (*Panopea Abrupta* and *Strongylocentrotus Franciscanus*) Targeted for Aquaculture and Enhancement in British Columbia. *Journal of Shellfish Research*, 25(1), 33–42.
- Nunome, T., Negoro, S., Miyatake, K., Yamaguchi, H. and Fukuoka, H. (2006). A protocol for the construction of microsatellite enriched genomic library. *Plant Molecular Biology Reporter*, 24(3-4), 305-312
- Nunziata, S. O., Karron, J. D., Mitchell, R. J., Lance, S. L., Jones, K. L. and Trapnell, D. W. (2012). Characterization of 42 polymorphic microsatellite loci in *Mimulus ringens* (Phrymaceae) using Illumina sequencing. *American Journal of Botany*, 99(12), e477–80.

- O'Bryhim, J., Chong, J. P., Lance, S. L., Jones, K. L. and Roe, K. J. (2012). Development and characterization of sixteen microsatellite markers for the federally endangered species: *Leptodea leptodon* (*Bivalvia*: Unionidae) using paired-end Illumina shotgun sequencing. *Conservation Genetics Resources*, 4(3), 787–789.
- Ostrander, E. a, Jong, P. M., Rine, J. and Duyk, G. (1992). Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences*, 89(8), 3419–23.
- Paetkau, D. (1999). Microsatellites obtained using strand extension: an enrichment protocol. *Biotechniques*, 26(4), 690-2.
- Parchman, T. L., Geist, K. S., Grahnen, J. a, Benkman, C. W. and Buerkle, C. A. (2010). Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*, 11(1), 180.
- Park, S. (2001). MStools v 3 (Excel spreadsheet toolkit for data conversion) Smurfit Institute of Genetics. *Trinity College, Dublin*, 2.
- Perez-Valencia, L. I. (2011). Variabilidad genética de la almeja de sifón *Panopea globosa* (dall, 1898) en el noroeste de México. Tesis de Maestría en Uso, Manejo y Preservación de los Recursos Naturales. Centro de Investigaciones Biológicas del Noroeste, S.C. 62p
- Perry, J. C. and Rowe, L. (2011). Rapid microsatellite development for water striders by next-generation sequencing. *The Journal of Heredity*, 102(1), 125–9.
- Rassmann, K., Schlötterer, C. and Tautz, D. (1991). Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. *Electrophoresis*, 12(2-3), 113-118.
- Ren, J., Shen, X., Sun, M., Jiang, F., Yu, Y., Chi, Z., and Liu, B. (2009). The complete mitochondrial genome of the clam *Meretrix petechialis* (*Mollusca*: *Bivalvia*: *Veneridae*). *Mitochondrial DNA*, 20(4), 78–87.
- Rocha-Olivares, A., Calderon-Aguilera, L. E., Aragón-Noriega, E. A., Saavedra-Sotelo, N. C. and Moreno-Rivera, V. M. (2010). Genetic and morphological variation of northeast Pacific *Panopea* clams: evolutionary implications. *Journal of Shellfish Research*, 29(2), 327–335.
- SAGARPA. (2007). Programa de investigación para el seguimiento de la pesquería de almeja generosa (*Panopea* spp) en las Costas de Baja California, México. Prospección y evaluación de nuevas áreas de aprovechamiento. Mazatlán, Sinaloa: CONAPESCA.
- SAGARPA.(2012). Plan de manejo para la pesquería de almeja generosa (*Panopea* spp.) en las costas de Baja California. Mazatlán, Sinaloa: CONAPESCA. 57 pp.

- Sanger, F. and Nicklen, S. (1977). DNA sequencing with chain-terminating. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467.
- Schlötterer, C. (2004). The evolution of molecular markers-just a matter of fashion? *Nature Reviews. Genetics*, 5(1), 63–9.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology*, 18(2), 233-234 .
- Schuster, S. C. (2008). Next-generation sequencing transforms today ' s biology. *Nature methods*, 5(1), 16–18.
- Selkoe, K. A. and Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, 9(5), 615–29.
- Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135–45.
- Shete, S., Tiwari, H. and Elston, R. C. (2000). On estimating the heterozygosity and polymorphism information content value. *Theoretical Population Biology*, 57(3), 265–271
- Silva, P. I. T., Martins, A. M., Gouvea, E. G., Pessoa-Filho, M. and Ferreira, M. E. (2013). Development and validation of microsatellite markers for *Brachiaria ruziziensis* obtained by partial genome assembly of Illumina single-end reads. *BMC Genomics*, 14(1), 17.
- Slabbert, R., Hepple, J., Rhode, C., Bester-Van der Merwe, a E., and Roodt-Wilding, R. (2012). New microsatellite markers for the abalone *Haliotis midae* developed by 454 pyrosequencing and in silico analyses. *Genetics and Molecular Research : GMR*, 11(3), 2769–79.
- Sloan, N.A. and Robinson, S.M. (1984). Age and gonad development in the Geoduck clam *Panopea abrupta* (Conrad) from southern British Columbia, Canada. *Journal Shellfish Research*, 4, 131–137
- Srikwan, S., Field, D. and Woodruff, D. S. (1996). Genotyping free-ranging rodents with heterologous pcr primer pairs for hypervariable nuclear. *Journal of the Science Society of Thailand*, 22, 267–274.
- Squirrell, J., Hollingsworth, P. M., Woodhead, M., Russell, J., Lowe, a. J., Gibby, M., and Powell, W. (2003). How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology*, 12, 1339–1348.
- Straus, K., Crosson, L. and Vadopalas, B. (2008). *Effects of geoduck aquaculture on the environment: a synthesis of current knowledge*. Washington Sea Grant Program, University of Washington.



- Suárez-Moo, P.J., Calderon-Aguilera, L. E., Reyes-Bonilla, H., Díaz-Erales, G., Castañeda-Fernandez-de-Lara, V., Aragón-Noriega, E. A. and Rocha-Olivares, A. (2013). Integrating genetic, phenotypic and ecological analyses to assess the variation and clarify the distribution of the Cortes geoduck (*Panopea globosa*). *Journal of the Marine Biological Association of the United Kingdom*, 93(03), 809–816.
- Tang, Z., Kuo, T., Shen, J. and Lin, R.J. (2000). Biochemical and Genetic Conservation of Fission Yeast Dsk1 and Human SR Protein-Specific Kinase 1. *Molecular and Cellular Biology*, 20(3), 816–824.
- Turcatti, G., Romieu, A., Fedurco, M. and Tairi, A.P. (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Research*, 36(4), e25.
- Vadopalas, B. and Bentzen, P. (2001). Isolation and characterization of di- and tetranucleotide microsatellite loci in geoduck clams. *Panopea abrupta*. *Molecular ecology*, 9(9), 1435-1436
- Vadopalas, B., Leclair, L. L. and Bentzen, P. (2004) Microsatellite and allozyme analyses reveal few genetic differences among spatially distinct aggregations of geoduck clams (*Panopea abrupta*, Conrad 1849). *Journal of Shellfish Research*, 23, 693–706.
- Vadopalas, B., Pietsch, T. W. and Friedman, C. S. (2010). The proper name for the geoduck: resurrection of *Panopea generosa* Gould, 1850, from the synonymy of *Panopea abrupta* (Conrad, 1849)(Bivalvia: Myoida: Hiatellidae). *Malacologia*, 52(1), 169-173.
- Vadopalas, B., Leclair, L. L. and Bentzen, P. (2012). Temporal Genetic Similarity Among Year-Classes of the Pacific Geoduck Clam (*Panopea generosa* Gould 1850): A Species Exhibiting Spatial Genetic Patchiness. *Journal of Shellfish Research*, 31(3), 697–709.
- VanKoeveeringe, M. (1998). *Molecular population genetics of British Columbia geoduck clams, Panope abrupta, based on mitochondrial DNA sequences*. Tesis de Mestría en Ciencias Biológicas. Simon Fraser University.60p
- Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), 1–9.
- Wyman, S. K., Jansen, R. K. and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics (Oxford, England)*, 20(17), 3252–5.
- Yuan, Y., Li, Q., Yu, H. and Kong, L. (2012). The complete mitochondrial genomes of six heterodont bivalves (*Tellinoidea* and *Solenoidea*): variable gene arrangements and phylogenetic implications. *PLoS One*, 7(2), e32353.

- Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R. and Simon, P. (2012). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany*, 99(2), 193–208.
- Zane, L., Bargelloni, L. and Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, 11(1), 1–16.
- Zhang, J., Chiodini, R., Badr, A. and Zhang, G. (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38(3), 95–109.