

**Centro de Investigación Científica y de Educación Superior de
Ensenada, Baja California**



**Maestría en Ciencias
en Ecología Marina**

**Variabilidad genética y filogeografía del pez arrecifal arcoíris de
Cortés, *Thalassoma lucasanum*, en el Pacífico Oriental Tropical**

Tesis

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

Presenta

Pablo Eladio Guevara Reyna

Ensenada, Baja California, México
2017

Tesis defendida por
Pablo Eladio Guevara Reyna

y aprobado por el siguiente Comité:

Dr. Axayácatl Rocha Olivares
Director de tesis

Dra. María Clara Arteaga Uribe
Dr. Luis Eduardo Calderón Aguilera
Dr. Omar Domínguez Domínguez
Dra. Fabiola Lafarga De la Cruz



Dra. María Lucila del Carmen Lares Reyes
Coordinador del Posgrado en Ecología Marina

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Resumen de la tesis que presenta **Pablo Eladio Guevara Reyna** como requisito parcial para la obtención del grado de Maestro en Ciencias en Ecología Marina

Variabilidad genética y filogeografía del pez arrecifal arcoiris de Cortés, *Thalassoma lucasanum*, en el Pacífico Oriental Tropical

Resumen aprobado por:

Dr. Axayácatl Rocha Olivares
Director de Tesis

En este proyecto se analizaron los patrones de variabilidad genética del arcoiris de Cortés, *Thalassoma lucasanum*, mediante un acercamiento filogeográfico y de genómica de poblaciones. El modelo de estudio es un pez arrecifal abundante en el Pacífico Oriental Tropical (POT). Para el análisis se emplearon polimorfismos de nucleótidos únicos (SNPs por sus siglas en inglés), mediante secuenciación genómica reducida (3RAD), además de un análisis de tres marcadores de ADN mitocondrial, para caracterizar patrones filogeográficos y probar hipótesis sobre su estructura poblacional. Se probaron hipótesis sobre la heterogeneidad genética espacial asociada a la presencia de barreras a la dispersión y otras fuerzas evolutivas. Se encontró que existe un bajo polimorfismo en los tres genes mitocondriales, de los cuales únicamente el citocromo oxidasa subunidad I (COI) fue polimórfico ($h= 0.105$; $\pi= 0.0002$), estos resultados sugieren que el bajo polimorfismo se extiende a todo el mitogenoma de la especie, por lo que no es apto para estudiar su genética poblacional. Esta baja diversidad estaría asociada con la biología de la especie, como resultado de una expansión demográfica, o por la naturaleza de la molécula de ADN mitocondrial. Por su parte, los SNPs evaluados revelaron una señal significativa de diferenciación genética en el POT así como un claro patrón filogeográfico. Se encontró que la brecha Centroamericana constituye una barrera efectiva que genera diferenciación genética en individuos del sur del Pacífico Oriental Tropical. Sin embargo, no se descarta la posibilidad de un flujo genético restringido de individuos del sur del Pacífico mexicano hacia esa zona. Por último, se encontró que existe una concordancia entre el origen geográfico de los individuos y la distribución de linajes en este sistema, debido quizá a una separación débil de linajes que podría ser efecto de una divergencia reciente entre clados del norte y sur de la brecha centroamericana.

Palabras clave: *Thalassoma lucasanum*, Pacífico Oriental Tropical, 3RAD, Barreras a la dispersión

Abstract of the thesis presented by **Pablo Eladio Guevara Reyna** as a partial requirement to obtain the Master of Science degree in Marine Ecology.

Genetic variability and Phylogeography of the rainbow wrasse, *Thalassoma lucasanum* from the Tropical Eastern Pacific.

Abstract approved by:

Axayácatl Rocha Olivares, PhD.
Thesis Director

This project analyzed the genetic variability patterns of the Cortez rainbow wrasse, *Thalassoma lucasanum*, using phylogeographical and population genomics approaches. Our study model is a coral reef species dominant across the Tropical Eastern Pacific (TEP) reef communities. We employed SNPs originated from a Reduced Representation sequencing platform (3RAD), and mitochondrial DNA sequences from three markers, in order to characterize both population structure and phylogeographic patterns. Our results allowed us to test hypotheses about the spatial genetic heterogeneity determined by barriers to dispersal as well as other evolutionary forces. We found low polymorphism in the three mitochondrial markers, of which only the cytochrome oxidase subunit I (COI) was polymorphic ($h=0.105$; $\pi=0.0002$), concluding that low polymorphism may be pervasive across the mitochondrial genome of this species, hence it is not suitable to analyze its population genetics. This pattern may result from the species' biology, a recent demographic expansion pattern or by the nature of the mitochondrion itself. On the other hand, the analyzed SNPs revealed significant genetic heterogeneity in the TEP as well as a clear phylogeographic pattern. We found that the Central American gap acts as an effective barrier to dispersal, causing genetic differentiation in individuals located at the southern-most part of the TEP system. Nevertheless, we do not discard the possibility of a restricted gene flow from individuals located at the south of the Mexican Pacific zone. In addition, we found a concordance between the lineage distribution of samples within the TEP system, which might be explained by a recent divergence between clades from the southern-most and northern parts of the Central American gap.

Keywords: *Thalassoma lucasanum*, Tropical Eastern Pacific, 3RAD, Dispersal barriers

Dedicatorias

A mi mamá, pues dicen por ahí que detrás de un gran pseudohombre siempre habrá una gran mujer. Soy quién soy y seguiré siendo gracias a ella. Gracias a Dios por permitirme haberla tenido como madre.

A mi familia, porque con ellos todo lo puedo y sin ellos nos soy nada. A todos ellos puesto que mis sueños no serían realizables si no me dieran la fuerza, cariño y comprensión que me han dado durante estos ya 26 años prácticamente.

A Cristina, pues si bien empezamos esta aventura y no la terminamos juntos, eres y seguirás siendo alguien muy importante en mi vida. Gracias por todo y sobre todo por la compañía, cariño y respeto que nunca había encontrado en alguien...

A mi grupo Scout y a mis hermanos scouts en todo el mundo, pues a la distancia han seguido presentes para mí y espero retribuirles todos y cada uno de los momentos maravillosos que me han dado

A mis amigos de Salamanca y Guanajuato. No soy de hacer muchas amistades pero ellos saben que son muy importantes en mi vida aunque rara vez hablemos. En general gracias a los barrenos y a los malignos jaja. Gracias a la banda de la secundaria y a mis compas de la prepa y la Carrera con quienes sigo teniendo contacto.

Esta tesis la dedico también a quienes ya no están conmigo físicamente, pero que continúan estando dentro de mí. Gracias por su cariño y ayuda, por ayudarme a crecer con y sin ustedes y por no dejarme abajo en ningún momento.

Te dedico esta tesis a tí papa, puesto que me has apoyado de alguna u otra manera y es muy cierto que esta tesis no hubiera sido posible sin tu ayuda. Gracias y ojalá algún día mejoremos nuestra relación

Y pues de una vez te la dedico a T1, si estás leyendo esto y te tomaste la molestia de llegar al fin jajaja...

Agradecimientos

Al CONACYT, quién me brindó la oportunidad de financiar mis estudios de posgrado durante poco más de dos años. Sin esta institución la ciencia y tecnología en México estaría en otro lugar muy distinto.

Al CICESE, por haberme aceptado y darme la oportunidad de desenvolverme en esta institución de la cual aprendí y me llevo muchas cosas...

Al posgrado de Ecología Marina, igualmente por haberme aceptado y darme la oportunidad de conocer una nueva familia de distintos estados de la República así como países distantes.

Al laboratorio de Ecología Molecular del CICESE, gracias a todos y cada uno de los chicos. Gracias a Ivonne y a Celia, a Viri, Ivette y Samantha, a Anita, a Francesco, al Chileno, a Miguelón, Edson y por supuesto a los que se han ido a cumplir nuevos retos (Pepe, Ale, Guamerús jaja, y la lista sigue!!). Y sobre todo gracias por aguantar todas, pero TODAS las sandeces que decía y dije durante estos dos años.

Al Dr. Axa, por haberme dado la oportunidad de trabajar en su laboratorio, y sobre todo por ayudarme a darme cuenta de que este negocio es de dedicación, esfuerzo y sobre todo de nunca estar conforme con lo que se tiene y a dónde se ha llegado.

A mis sinodales, al Dr. Luis Calderón, pues gracias a él mantuve mis deseos de terminar la maestría. A la Dra. Clara por haberme dado un excelente curso, así como hacerme preguntas geniales sin la hostilidad con la que lo hacía el jefe (perdón pero es la verdad). También al Dr. Omar por ayudarme a darme cuenta de mis errores de interpretación en biogeografía, y sobre todo por financiar este proyecto y darnos su apoyo en todo momento y con complete disposición para que dicho proyecto se llevara a cabo efectivamente.

A la División de Innovación Biomédica de CICESE, al Dr. Alexei, a Samanta, Hanna, a mi carnal Lalo, Vane, Vivi, Ricardo, los Fernandos, los Dres. Pierrick y Paty Juarez y a todo el crew de ese laboratorio al cual llegué siendo un sucio estudiante de licenciatura y salí prácticamente como licenciado en Biología Experimental.

A mi pandilla de Ensenada, no digo nombres porque la clicka crece cada vez más jaja. A todos y cada uno de ustedes por estar en las buenas, las malas y las peores. Siempre tendrán su casa en la tierra de este perro, y siempre tendrán un amigo, un hermano y un confidente en mi persona.

A la raza de la natación, del fut Shaolin, Cristiano, y a toda la demás bola del CICESE, UABC y adjuntos que conocí y de alguna u otra manera cambiaron mi vida para bien.

A todos y cada uno de ustedes, gracias por todo y a darle átomos!!!

Pablo...

Tabla de contenido

	Página
Resumen en español	iii
Resumen en inglés	iv
Dedicatorias	v
Agradecimientos	vi
Lista de figuras	ix
Lista de tablas	x
Capítulo 1. Introducción	
1.1 Efecto de las fuerzas evolutivas en la diversidad genética	1
1.2 RADseq: Una herramienta innovadora y costeable para la secuenciación en organismos no modelo	2
1.3 Biogeografía y filogeografía del Pacífico oriental tropical.....	3
1.4 Especie de estudio.....	7
1.4.1 Antecedentes genéticos de <i>Thalassoma lucasanum</i>	9
1.5 Justificación.....	10
1.6 Hipótesis.....	11
1.7 Objetivos.....	12
1.7.1 Objetivo general	12
1.7.2 Objetivos específicos.....	12
Capítulo 2. Metodología	
2.1 Muestreo y obtención de organismos.....	13
2.2 Extracción de ADN.....	13
2.3 Amplificación y análisis de fragmentos mitocondriales.....	13
2.4 Bibliotecas genómicas 3RAD	15
2.5 Diversidad genética, neutralidad y estructura poblacional.....	16
2.6 Reconstrucción filogenética.....	17
Capítulo 3. Resultados	
3.1 Extracciones de ADN en individuos de <i>T. lucasanum</i>	18
3.2 Diversidad mitocondriales	19
3.2.1 Identificación taxonómica de muestras de tejido.....	19

3.2.2 16s ARNr.....	19
3.2.3 Citocromo B (CYTB).....	20
3.2.4 Citocromo Oxidasa subunidad I (COI).....	20
3.3 Diversidad nuclear (SNPs).....	21
3.3.1 Secuenciación 3RAD.....	21
3.3.2 Diversidad.....	22
3.4 Patrones de estructura genética.....	23
3.4.1 Concordancia ecológica: efecto de las discontinuidades en el hábitat rocoso.....	23
3.4.2 Concordancia biogeográfica.....	24
3.4.3 Patrones entre localidades.....	24
3.5 Patrones filogeográficos	28
Capítulo 4 . Discusión	
4.1 Variabilidad genética de <i>Thalassoma lucasanum</i>	31
4.1.1. Variabilidad mitocondrial.....	31
4.1.2. Utilidad de datos mitocondriales para corroborar la identidad taxonómica de muestras de tejido	32
4.1.3 Variabilidad genómica	33
4.2 Estructura poblacional de <i>T. Lucasanum</i> en el POT.....	33
4.2.1. Efecto de barreras a la dispersión en la estructura poblacional de <i>T. lucasanum</i>	34
4.3 Patrones filogeográficos y concordancias génicas.....	35
Capítulo 5 . Conclusiones	
5.1 Conclusiones	37
Literatura citada	38

Lista de figuras

Figura	Página
1 Mapa del Pacífico Oriental Tropical. Se indican las localidades de muestreo (n= tamaño de muestra), las barreras a la dispersión (brechas) y las provincias biogeográficas demarcadas en esta zona. Tomado de Robertson y Cramer (2009) y modificado según (Briggs y Bowen, 2012).....	5
2 Filogeografía de la anémona <i>Nematostella vectensis</i> en cuatro linajes evolutivos (Nueva Escocia (NS); Massachussets (MA); Maryland (MD) y Carolina del Sur (SC)) obtenida a partir de SNPs de alta fidelidad en múltiples marcadores RAD resultado de la genotipificación de loci de 30 individuos. Tomada de (Reitzel <i>et al.</i> , 2013).....	7
3 Etapas del desarrollo de <i>Thalassoma lucasanum</i> en A) Fase juvenil (machos y hembras) y B) Macho en fase adulta. Tomado de (Hernández Olalde, 2008).....	8
4 Sitios variables en el gen COI. Los sitios se encuentran resaltados en amarillo. Los números representan la posición en el alineamiento. Los signos de interrogación corresponden a bases no identificadas (datos faltantes).....	21
5 Distribución de la variación genética en función de 3133 SNPs independientes del arcoíris de Cortés, <i>T. lucasanum</i> . En la imagen los dos factores principales explican cerca del 40 % de la variación genética total, teniendo los valores de eigenvalores (discriminantes) más robustos en el modelo (barras inferior derecha).....	26
6 Figura 6. Agrupamiento Bayesiano de la estructuración poblacional en <i>T. lucasanum</i> con el valor calculado de más probable de agrupamiento (K=3). La gama de colores representa el coeficiente de ancestría (q) de cada agrupamiento en cada individuo en función de : A) Localidad; B) Barreras a la dispersión; C) Biogeografía.....	28
7 Reconstrucción filogenética del arcoíris de Cortés <i>Thalassoma lucasanum</i> en el POT, en función del mejor modelo de sustitución molecular mediante máxima verosimilitud. La gama de colores en los agrupamientos corresponde a Sonora (rojo), Mulegé (naranja), La Paz (dorado), Revillagigedo (lima), Colima (verde), Guerrero (aguamarina), Oaxaca (azul marino), Costa Rica (magenta), Galápagos (púrpura). Las barras de color negro, azul y verde corresponden a la demarcación de localidades respecto al norte de la brecha de Sinaloa (norte de brechas), entre y al sur de la brecha Centroamericana (sur de brechas). Se indica también la demarcación por provincias biogeográficas. Se muestran solamente valores con alto soporte estadístico (bootstrap >99).....	30

Lista de tablas

Tabla		Página
1	Lista de cebadores (primers) para amplificación de marcadores mitocondriales de interés. Se ubican las secuencias y referencias bibliográficas por marcador.....	14
2	Localidades del muestreo, número de muestras analizadas por localidad (n) y coordenadas geográficas de cada sitio según Google Earth (https://www.google.com/earth/).....	18
3	Relación de muestras analizadas correspondientes a otras especies de teleósteos.....	19
4	Resumen de resultados para cada marcador mitocondrial utilizado en este estudio.....	21
5	Estadísticos descriptivos de diversidad genética nuclear de los 3133 SNPs independientes en <i>T. lucasanum</i> . Se observan los valores por localidad. π : Diversidad nucleotídica; H_o : Heterocigosidad observada; H_E . Heterocigosidad esperada; Fis: Coeficiente de endogamia.....	23
6	Valores de F_{st} pareado entre las localidades ubicadas al norte de la brecha de Sinaloa, entre brechas y al sur de la brecha centroamericana.....	23
7	Estructuración poblacional a nivel biogeográfico en <i>T. lucasanum</i> . Los valores de F_{st} pareados entre provincias se encuentran marcados en color negro; El valor de significancia a un nivel de $\alpha= 0.05$ se encuentra identificado por asteriscos. GRI corresponde a la provincia de las Grandes Islas (Galápagos).....	24
8	Valores de F_{st} pareado entre localidades (Parte superior derecha). Valores de significancia (parte inferior izquierda).....	25

Capítulo 1. Introducción.

1.1. Efecto de las fuerzas evolutivas en la diversidad genética

La diversidad genética se encuentra sujeta a la acción de distintas fuerzas evolutivas que moldean la capacidad de los organismos de responder a los cambios ambientales. Mientras que algunas de estas fuerzas aportan variabilidad genética (flujo genético y mutación), otras reducen esta variabilidad (selección natural y deriva genética). Los patrones resultantes pueden evaluarse mediante estudios de genética y genómica poblacional.

El flujo genético se encuentra asociado a una homogenización de la variabilidad genética en las poblaciones, disminuyendo la habilidad de los organismos de adaptarse a las condiciones locales (Limborg *et al.*, 2012). Por su parte, factores como la heterogeneidad ambiental ante el balance entre migración-selección-deriva podría tener un efecto selectivo traducido en patrones de adaptación local en especies marinas con amplia distribución (Yeaman y Whitlock, 2011; Limborg *et al.*, 2012).

Por otra parte, la mutación es el cambio en la información genética resultante en diversidad, la cual además puede tener efecto en el fenotipo de los individuos. Las tasas de mutación en organismos son altamente variables (Bromham, 2009), lo cual hace complejo el estudio de la evolución molecular.

Deriva genética se refiere a un cambio en las frecuencias alélicas por procesos azarosos debidos a fallas en el muestreo de poblaciones finitas (Templeton, 2006). Estas pérdidas aleatorias de diversidad son una consecuencia directa de los cambios en el tamaño efectivo poblacional (Ackermann y Cheverud, 2004).

La selección natural, por su parte, es el proceso mediante el cual una presión de selección afecta a los organismos de manera diferencial, los cuales pueden o no adaptarse a esta condición, actuando como modulador de la evolución. Algunos ejemplos de los tipos de selección natural (diferencial, disruptiva, balanceadora) se encuentran en (Ackermann y Cheverud, 2004; Haney, Silliman y Rand, 2007; Limborg *et al.*, 2012; Reitzel *et al.*, 2013).

Los patrones y efectos de estas fuerzas evolutivas pueden estudiarse mediante el uso de marcadores moleculares, los cuales son secuencias de distintos tamaños con mayores tasas mutacionales que regiones conservadas, las cuales guardan información sobre la historia evolutiva de los organismos (Avice, 2012).

Por una parte, el ADN mitocondrial (ADNmt) ha sido ampliamente utilizado debido a que presenta características importantes como una baja tasa de recombinación y una alta tasa de mutación en comparación con el genoma nuclear (Avice, 2000). Sin embargo, la interpretación de los datos generados a partir de marcadores mitocondriales tanto en filogeografía como en filogenia y demografía histórica se restringe a un solo locus, ya que todos los genes del genoma mitocondrial se encuentran físicamente ligados.

Por otra parte, los marcadores llamados “polimorfismos de nucleótido único (SNPs)”, son marcadores codominantes que han permitido caracterizar varios procesos poblacionales (Vignal *et al.*, 2002; Hodel *et al.*, 2016), tanto en regiones no codificantes (neutrales) como en regiones que codifican para distintos procesos biológicos (Hohenlohe *et al.*, 2010; Reitzel *et al.*, 2013). Mediante este tipo de marcadores se han realizado diversos estudios sobre la demografía histórica, distribución de linajes y estructura poblacional de distintas especies no modelo (Davey *et al.*, 2011, 2013). Las librerías de representación reducida (ej. RRL) han permitido evaluar el efecto de barreras históricas al flujo genético, tanto en ambientes terrestres (Emerson *et al.*, 2010) como en ambientes marinos (Tariel, Longo y Bernardi, 2016). En algunos casos se han calculado con exactitud el tiempo de divergencia entre dos o más linajes (Tariel, Longo y Bernardi, 2016; Younger *et al.*, 2017), lo cual permite interpretar objetivamente las relaciones evolutivas que guardan estos taxones mediante la evaluación de múltiples loci.

1.2. RADseq: Una herramienta innovadora y costeable para la secuenciación en organismos no modelo.

Las técnicas de secuenciación masiva han revolucionado la forma en la que los organismos pueden ser genotipificados mediante el uso de miles de marcadores distribuidos aleatoriamente en el genoma de la especie (Davey *et al.*, 2011). Por otro lado, se han descrito metodologías para la construcción de bibliotecas genómicas mediante procesos variados reduciendo costos y maximizando la cobertura por individuo (Davey *et al.*, 2013). Una de estas metodologías se denomina secuenciación de ADN asociado a sitios de restricción (RAD-seq por sus siglas en inglés), la cual fue descrita por Miller y colaboradores (2007) y posteriormente por Baird y colaboradores (2008) (Miller *et al.*, 2007; Baird *et al.*, 2008). Esta técnica consiste en la digestión del ADN genómico con enzimas de restricción con sitios de corte conocidos. Posteriormente, los fragmentos cortados son acoplados a secuencias de ADN para la identificación de las

muestras individuales (códigos de barras genéticos). Dichos fragmentos son amplificados por PCR y secuenciados mediante la plataforma Illumina (Baird *et al.*, 2008; Davey *et al.*, 2013; Wagner *et al.*, 2013). Mediante esta secuenciación genómica reducida es posible encontrar múltiples SNPs, ensamblar genomas de organelos y comparar secuencias desconocidas con secuencias de referencia; siendo un método práctico para el estudio genómico de organismos no modelo con pocos recursos genómicos disponibles (Reitzel *et al.*, 2013).

1.3. Biogeografía y filogeografía del Pacífico oriental tropical.

El Pacífico Oriental Tropical (POT) es una zona oceanográfica que comprende desde el Noroeste de México, hasta el archipiélago de las Galápagos en la zona oceánica, recorriendo además Centroamérica y el Caribe, hasta el norte de Perú en la plataforma continental. La diversidad biológica en esta zona se encuentra sujeta a distintos procesos oceanográficos como choque de corrientes y masas de agua, así como discontinuidades en el hábitat, modulando la capacidad de dispersión de dichas especies así como sus rangos de distribución en los distintos estadios del ciclo de vida (Craig *et al.*, 2006). El chorro de Tehuantepec, chorro de Papagayo y el chorro de Panamá son masas de aire con influencia en la dinámica oceanográfica que generan mezcla de masas de agua, enfriamiento superficial y aumento en concentración de fitoplancton y nutrientes orientados en dirección al oeste (Liang, McWilliams y Gruber, 2009). Además, asociados a esta dinámica se encuentran eventos de mesoescala, los cuales explican la variabilidad estacional en la circulación superficial (Manzello, Eakin y Glynn, 2017).

Biogeográficamente, el Pacífico Oriental Tropical se ha subdividido en varias provincias que han cambiado su configuración conforme se conoce más sobre los niveles de endemismo de la zona. Principalmente destacan cuatro provincias: California, Cortés, Panámica y Perú-Chile (Briggs y Bowen, 2012). Anteriormente se consideraba al Pacífico Mexicano Tropical como una provincia separada, previo a encontrar nueva evidencia la cual sugería esta zona no contaba con el nivel de endemismo requerido (ej. 10%) para ser considerada como tal (Robertson y Cramer, 2009; Briggs y Bowen, 2012). Con el tiempo se demostró además que el archipiélago de Galápagos era la única isla oceánica con el suficiente endemismo (ej. 13% peces costeros) para ser considerada una provincia separada, por lo que el resto de las islas oceánicas del POT (ej. Revillagigedo; 8% de endemismo) deberían considerarse como puntos extremos de la provincia panámica (Briggs y Bowen, 2012).

Además de las provincias, se han identificado dos discontinuidades en el hábitat costero del POT, que pueden constituir barreras físicas para la dispersión de las especies costeras que habitan fondos rocosos o sistemas arrecifales, puesto que son costas con fondos blandos sedimentarios arenosos o fangosos (Hastings, 2000). La primera es la llamada “Brecha de Sinaloa”, que se extiende 370 km desde Topolobampo hasta Mazatlán, Sinaloa; y la segunda es la “Brecha Centroamericana”, que se extiende por más de 1000 km desde el Istmo de Tehuantepec hasta el Golfo de Fonseca (Hastings, 2000; Craig *et al.*, 2006) (Fig. 1). Estas brechas podrían constituir barreras a la dispersión de las especies de fondos rocosos del hábitat circundante. Considerando este factor, es probable que las barreras a la dispersión pudieran generar rupturas en la conectividad genética de dichas especies, reflejado en una diferenciación genética cuantificable para cada especie.

En un estudio realizado por Lessios y Baums (2017) se evaluaron los niveles de flujo genético entre distintas poblaciones de corales, peces, invertebrados marinos, entre otros, los cuales componen parte de la fauna del Pacífico Oriental Tropical (Lessios y Baums, 2017). Tomando en cuenta estudios con marcadores mitocondriales y nucleares en todo el POT, los autores concluyen que pese a que existen barreras físicas a la dispersión, el nivel de flujo genético es variable según la especie de estudio, por lo que hablar de una alta diferenciación genética como un patrón generalizado en las poblaciones del POT es cuestionable (Lessios y Baums, 2017; Manzello, Eakin y Glynn, 2017). Por último, se menciona que existe un patrón de aislamiento por distancia en los patrones de diferenciación genética poblacional de las especies de corales y peces residentes del POT. Este patrón se presenta principalmente en las especies cercanas a la costa y específicamente en aquellas separadas entre las zonas más al norte y sur del sistema; principalmente debido a barreras a la dispersión, así como a las condiciones ambientales que difieren entre el golfo de California y el centro y sur del Pacífico Oriental.



Figura 1. Mapa del Pacífico Oriental Tropical. Se indican las localidades de muestreo (n= tamaño de muestra), las barreras a la dispersión (brechas) y las provincias biogeográficas demarcadas en esta zona. Tomado de Robertson y Cramer (2009) y modificado según (Briggs y Bowen, 2012).

La filogeografía es una rama de la biogeografía encargada de estudiar la distribución geográfica de los linajes evolutivos de las especies, relacionándolos con distintos procesos y eventos en escalas de tiempo geológicas (Avice, 2000). La mayor parte de los estudios filogeográficos se han centrado en secuencias mitocondriales debido a sus propiedades evolutivas tales como: 1) transferencia uniparental; 2) múltiples copias por célula somática; 3) evolución rápida en su secuencia nucleotídica, 4) baja o nula tasa de recombinación genética y 5) extenso rango de polimorfismos intraespecíficos (Avice, 2000).

Craig y colaboradores (2006) utilizaron el marcador mitocondrial Citocromo B (CYTB) para caracterizar la distribución de linajes de la cabrilla *Epinephelus labriformis*. Los autores encuentran que no existe una divergencia significativa entre los organismos costeros respecto de los ubicados en la zona oceánica,

sugiriendo panmixia en estas poblaciones (Craig *et al.*, 2006). Además, los autores consideraron que una posible ruptura biogeográfica no sería un factor que conllevara a un evento de especiación en la cabrilla.

En un trabajo realizado por Ludt y colaboradores (2012) se emplearon los marcadores citocromo oxidasa subunidad I y región control (ej. RC) para determinar las diferencias en la estructura genética asociadas a las variaciones en el nivel del mar debidas a los periodos glaciales, así como el tiempo de divergencia entre las especies de lábridos (género *Halichoeres*) ubicuas del Indo-Pacífico (Ludt *et al.*, 2012). Se encontró que la divergencia entre estos linajes es débil y reciente, debido a que existe conectividad entre las poblaciones costeras y las pertenecientes a la zona del arrecife externo. Este patrón tuvo como efecto una alta diversidad haplotípica, además de una baja diversidad nucleotídica, la cual sería resultado de una expansión poblacional atribuida a los periodos interglaciales del Pleistoceno (Ludt *et al.*, 2012).

Reitzel y colaboradores (2013) analizaron a la anémona *Nematostella vectensis* mediante el uso de SNPs (Reitzel *et al.*, 2013). Los autores desarrollaron un análisis filogeográfico de alta resolución (Fig. 2.0), así como un análisis de ontología genética (GO por sus siglas en inglés), encontrando regiones del genoma de la especie sujetas a selección disruptiva. Entre otros atributos, resaltan las ventajas del uso de los marcadores RAD en organismos no modelo y con pocos recursos genómicos para la obtención de datos genéticos robustos, así como en la identificación de regiones genómicas candidatas a estar sujetas a procesos de selección natural (Reitzel *et al.*, 2013). Esta alta resolución de los marcadores permitió encontrar un claro patrón de ruptura genética entre los clados ubicados al norte de la costa este de Estados Unidos y Canadá con la zona sur aledaña al golfo de México. Este patrón sería resultado de un evento de recolonización del hábitat costero posterior a los periodos glaciales de esa zona (Reitzel *et al.*, 2013).

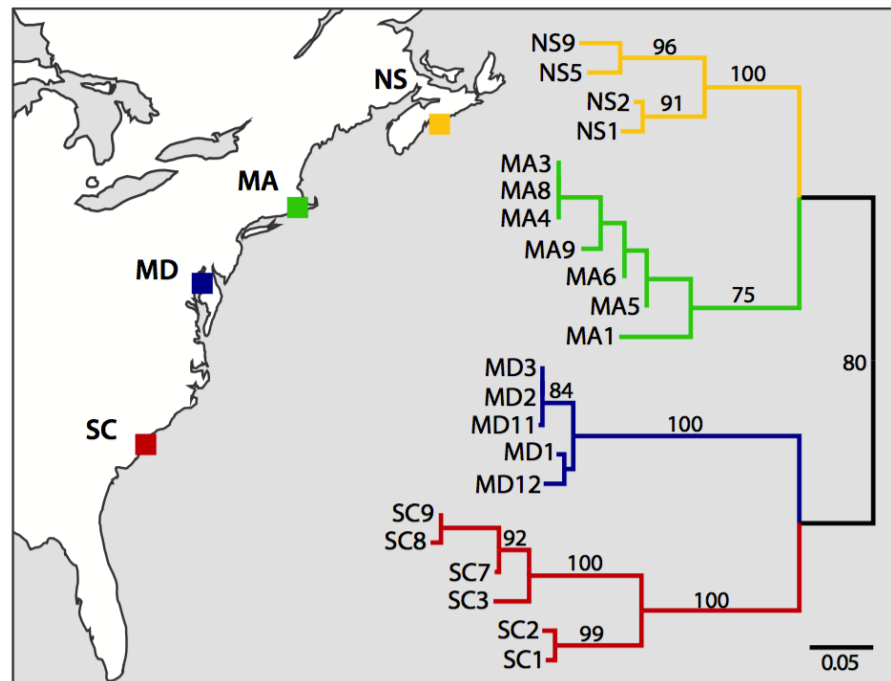


Figura 2. Filogeografía de la anémona *Nematostella vectensis* en cuatro linajes evolutivos (Nueva Escocia (NS); Massachussets (MA); Maryland (MD) y Carolina del Sur (SC)) obtenida a partir de SNP's de alta fidelidad en múltiples marcadores RAD resultado de la genotipificación de loci de 30 individuos. Tomada de Reitzel y colaboradores (2013).

Posteriormente, Tariel y colaboradores (2016) caracterizaron el tiempo de especiación en peces del género *Holacanthus*, los cuales resultaron fragmentados como resultado del surgimiento del Istmo de Panamá (Tariel, Longo y Bernardi, 2016). Mediante el uso de 1189 SNPs se caracterizó el tiempo de especiación en las especies del Pacífico Oriental Tropical y del Atlántico Occidental Tropical estimado en 1.5 M.A., consistente con el cierre del Istmo de Panamá. Además, estos autores hicieron inferencia sobre una posible divergencia debida a simpatria en los clados del Atlántico, así como de un caso de peripatria en los clados del Pacífico (Tariel, Longo y Bernardi, 2016).

1.4. Especie de estudio.

Thalassoma lucasanum (Gill, 1863) es una especie de pez arrecifal dominante en el Pacífico Oriental Tropical (Giraldo, Gomez y Ospina, 2001; Alvarez-Filip, Reyes-Bonilla y Calderon-Aguilera, 2006). Una de

las características principales de esta especie es ser hermafrodita protogínico, es decir, que algunas hembras pueden cambiar de sexo a machos. Esta inversión de género puede deberse a una respuesta de comportamiento social y a cambios en la densidad de machos y hembras (Warner, 1982; Giraldo, Gomez y Ospina, 2001). Si el número de machos dominantes desciende, algunas hembras juveniles cambian de sexo, de modo que los machos dominantes son reestablecidos en la población (Warner, 1982). *Thalassoma lucasanum* también presenta cambios en coloración según la fase inicial y final del desarrollo (Warner, 1982; Hernández Olalde, 2008) (Figura 3).

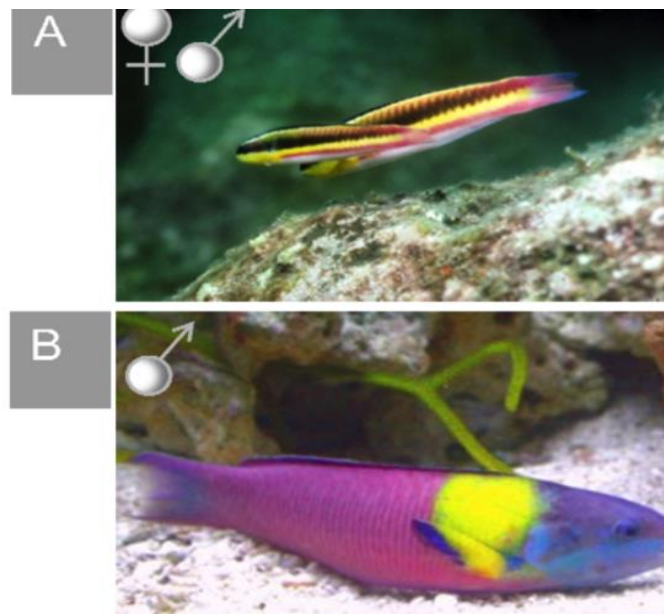


Figura 3. Etapas del desarrollo de *Thalassoma lucasanum* en A) Fase juvenil (machos y hembras) y B) Macho en fase adulta. Tomado de (Hernández Olalde, 2008).

La dispersión en esta especie se ve influenciada principalmente por dos factores: 1) la dispersión de huevos en la columna de agua, facilitada por las corrientes marinas (Hernández Olalde, 2008); y 2) El periodo de duración larvaria cuyo efecto en el crecimiento se ejerce una vez eclosionados los huevos (Giraldo, Gomez y Ospina, 2001). Debido a la variación en la densidad de machos juveniles aquellos que llegan a una etapa adulta presentan territorialismo como respuesta a una mayor competencia por recursos, así como el acceso a hembras. Este comportamiento se refleja en una baja capacidad de dispersión efectiva, además

de una mayor adecuación respecto a otros machos en estadio juvenil al tener varias hembras disponibles (Warner, 1982).

1.4.1. Antecedentes genéticos de *Thalassoma lucasanum*

En un esfuerzo por estudiar la variabilidad, estructura y flujo genéticos de *T. lucasanum* en el sistema del Pacífico Oriental Mexicano (PMT), Domínguez Rivera (2011) estudió 110 secuencias de la RC del ADNmt en individuos localizados desde La Paz hasta Huatulco. Encontró valores de 0.124 y 0.000059 de diversidad haplotípica y nucleotídica, respectivamente, en individuos muestreados a lo largo de este sistema. Esta baja diversidad genética sería resultado principalmente de las características intrínsecas y la demografía de la especie (ej. Ubicuidad, protoginia), más que un efecto debido por artefactos del marcador tales como la amplificación de pseudogenes (numts) (Dominguez-Rivera, 2011).

Una pregunta pendiente sobre el nivel de estructuración genética poblacional de *Thalassoma lucasanum* consiste en establecer si la baja diversidad observada en estudios anteriores se extiende a todo el mitogenoma y al genoma nuclear (Dominguez-Rivera, 2011). De no ser así, la ausencia de estructura genética podría deberse al bajo polimorfismo más que a un estado de panmixia. Además, se desconoce si la estructura poblacional del arcoíris de Cortés ha sido modificada y/ o moldeada por un aislamiento debido a barreras oceanográficas que pudieran impedir el flujo genético entre poblaciones (Hastings, 2000; Craig *et al.*, 2006). Estos resultados serían lo suficientemente robustos para caracterizar patrones de procesos ecológicos y evolutivos representativos en la cobertura del genoma de la especie.

En este trabajo se buscó caracterizar los patrones de variabilidad y filogeografía del arcoíris de Cortés, *Thalassoma lucasanum*, utilizando una aproximación de secuenciación genómica reducida (ej. RADseq), así como el uso de tres marcadores mitocondriales con distintos niveles de conservación en su secuencia, buscando patrones filogeográficos a lo largo del Pacífico Oriental Tropical, evaluando el efecto que las barreras a la dispersión para la especie en su estructuración genética y filogenia.

1.5 Justificación

Conocer los niveles de variabilidad y estructuración genética de las especies permite establecer las relaciones evolutivas y ecológicas que dichas especies guardan entre y dentro de sus poblaciones. Se propone realizar una evaluación multi-locus a lo largo del genoma, con mayor resolución que los análisis locus-específicos previamente realizados y que han revelado bajos niveles de variación en el mitogenoma de *Thalassoma lucasanum*. Los resultados derivados de este análisis permitirán hacer inferencias sobre los patrones filogeográficos de la especie en aras de establecer las bases para un mejor entendimiento de su biología aplicable en la conservación, tanto del arcoíris de Cortés como de los sistemas arrecifales que habita.

1.6 Hipótesis.

H1:

Las discontinuidades en el hábitat rocoso y coralino representan una barrera a la dispersión de *Thalassoma lucasanum*

- Predicción: Las poblaciones separadas por las Brechas de Sinaloa y Centroamericana se encuentran genéticamente diferenciadas

H2:

Las barreras a la dispersión identificadas mediante análisis de estructura poblacional producirán patrones filogeográficos concordantes en los linajes evolutivos de la especie

- Predicción: Existen concordancias genealógicas con las regiones biogeográficas en los árboles génicos de *Thalassoma lucasanum*

1.7 Objetivos

1.7.1 Objetivo General:

- Estimar los patrones de variabilidad y estructura genéticas de *Thalassoma lucasanum* en el Pacífico Oriental Tropical usando una perspectiva genómica.

1.7.2 Objetivos particulares:

- Estimar la influencia de barreras físicas, como las brechas costeras arenosas, en el nivel de diferenciación genética poblacional de *T. lucasanum* en Pacífico Oriental Tropical (POT) mediante el uso de SNPs
- Utilizar marcadores nucleares bialélicos para evaluar los patrones filogeográficos en *T. lucasanum*
- Utilizar tres genes mitocondriales para evaluar si existe un patrón generalizado de baja diversidad genética en el mitogenoma de la especie
- Relacionar los patrones filogeográficos de la especie de estudio con los eventos históricos en el POT.

Capítulo 2. Metodología.

2.1 Muestreo y obtención de organismos.

Las muestras de tejido provienen de 34 localidades del POT y se han colectado en el contexto del proyecto “**Conectividad y fragmentación en diferentes especies de peces del Pacífico Oriental Tropical**” (Fig. 1). Las muestras de aleta dorsal fueron guardadas en etanol sin desnaturalizar al 90% a -70°C.

2.2 Extracción de ADN.

La extracción de ADN de las muestras de tejido se realizó usando el kit de extracción DNeasy® Blood & Tissue kit de QIAGEN (Sample & Assay technologies™), cuya metodología se describe en Davey y colaboradores (2013). La calidad del ADN fue verificada mediante geles de agarosa 1.5 % y fluorométricamente con el fluorómetro Qubit® 2.0 (Thermo Fischer Scientific™). Cada muestra se estandarizó aproximadamente a 20 ng/μl de ADN por individuo. Las muestras de ADN de cada individuo (50μl) fueron enviadas a una concentración aproximada de 20 ng/ μl de ADN.

2.3 Amplificación y análisis de fragmentos mitocondriales.

Se utilizaron cebadores (primers) para la amplificación por PCR de los marcadores mitocondriales listados en la Tabla I. Los ciclos de amplificación para cada marcador fueron los reportados en la bibliografía y consisten en los siguientes pasos:

Para el citocromo c oxidasa subunidad I: 120 s a 95° C, 30 s a 94° C, 30 s a 54° C y 60 s a 72° C por 35 ciclos. Por último, se dejaron 10 minutos a 72 ° C para la extensión final. Para el citocromo b: 60 s a 93° C, 60 s a 50° C y 180 s a 72° C durante 35 ciclos, con 10 minutos de extensión final.

Para la subunidad ribosomal 16s: 45 s a 94° C, 45 s a 53° C y 60 s a 72° C durante 35 ciclos.

Tabla 1. Lista de cebadores (primers) para amplificación de marcadores mitocondriales de interés. Se ubican las secuencias y referencias bibliográficas por marcador.

Marcador	Primer L	Primer H	Referencia.
Citocromo b (CYTB)	L14841 (5'-AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA-3')	H15149 (5'-AAACTGCAGCCCCTCAGAA TG- ATATTTGTCCTCA-3')	(Kocher <i>et al.</i> , 1989; Bernardi <i>et al.</i> , 2004)
Citocromo c oxidasa subunidad I (COI)	FishF1: TCAACCAACCACAAAGACATTGGCAC	FishR1: TAGACTTCTGGGTGGCCAA AGAATCA	(Ward <i>et al.</i> , 2005)
Subunidad ribosomal 16s	16Sar-5'-CGCCTGTTTATCAAAAACAT - 3'	16Sbr-3'-CCGGTCTGAACTCAGATCA CGT-5'	(Ivanova <i>et al.</i> , 2007)

Las secuencias de marcadores mitocondriales fueron revisadas y editadas mediante CodonCode Aligner Versión 6.0.2. Posteriormente fueron alineadas con el programa MEGA 7.0.21 (Kumar, Stecher y Tamura, 2016). Con el fin de optimizar el tamaño de muestra de individuos con marcadores polimórficos se seleccionó un sub conjunto de muestras de ADN de *T. lucasanum* como base para realizar las amplificaciones de los distintos marcadores mitocondriales empleados en el estudio. Se secuenciaron 30, 63 y 63 amplicones de organismos para los genes 16s ARN, CYTB y COI, respectivamente. Después, se obtuvo el número de haplotipos por gen con el programa DnaSP (Librado y Rozas, 2009). Por último, se realizaron los cálculos estadísticos de índices de diversidad con Arlequin 3.5 (Excoffier y Lischer, 2010).

2.4 Bibliotecas genómicas 3RAD

Para la construcción de bibliotecas genómicas se utilizó el protocolo 3RAD (Graham *et al.*, 2015) y fueron realizadas y secuenciadas en el departamento de Ciencias Ambientales y de la Salud de la Universidad de Georgia, en Athens. Este protocolo utiliza una enzima de restricción adicional a la metodología de doble digestión (ddRADseq; Peterson *et al.*, 2012), con un sitio de corte similar a una de las enzimas del par principal, evitando así la aparición de dímeros de adaptadores y quimeras de secuencias en las bibliotecas. Para la digestión se utilizaron las enzimas *MspI*, *BamHI* y *Clal*, además de 3 distintos adaptadores iTru7 (ej., iTru7_110_01, iTru7_110_02, iTru7_110_03; (Hoffberg *et al.*, 2016)). Se utilizaron también distintos códigos de barras (ej. secuencias cortas) como adaptadores internos, con la finalidad de realizar el filtrado o “demultiplex” de las secuencias individuales. Las muestras de ADN de cada individuo seleccionado fueron secuenciadas (150 pb paired-end) con la plataforma Illumina Hi-seq 4000.

El análisis de los datos genómicos procedentes del RADseq se realizó mediante el programa Stacks[®], el cual ejecuta un modelo de máxima verosimilitud estadístico para el tratamiento de múltiples loci. Para ello realiza un ensamble de las lecturas individuales de ADN *de novo*, es decir, sin un genoma de referencia (Catchen *et al.*, 2011; Catchen, Hohenlohe, *et al.*, 2013). Primeramente, las secuencias pasaron por una limpieza y control de calidad mediante el comando *process_radtags*. Este comando permite discriminar entre los distintos locus que componen el genoma y aquellas lecturas que pudieran ser secuencias repetidas, lo que provocaría la sobreestimación de la profundidad de secuenciación. Adicionalmente, se empleó el comando *clone_filter*, el cual permite eliminar duplicados de PCR resultantes de la amplificación de dímeros de adaptadores y quimeras de secuencias.

Una vez realizado el filtrado y revisión de secuencias, se utilizó el comando de control *denovo.map.pl*, el cual permite procesar los datos generados por secuenciación reducida mediante cada uno de los pasos del pipeline. Este consiste primeramente en la utilización del comando *ustacks* para el ensamble de loci realizando un llamado de SNPs *de novo*. Posterior al ensamblaje, se ejecuta el comando *cstacks*, el cual permite formar un catálogo de todos los loci encontrados en los individuos de las distintas poblaciones. Por último, se ejecuta el comando *sstacks*, cuya función permite comparar cada una de los ensamblajes obtenidos con respecto al catálogo, de modo que se encuentren coincidencias determinando así loci polimórficos reales (Catchen *et al.*, 2011).

2.5 Diversidad genética y estructura poblacional

Para los marcadores mitocondriales se calcularon los índices de diversidad nucleotídica y haplotípica (π y h respectivamente) para los datos de cada marcador con el programa ARLEQUIN 3.5 (Excoffier, Hofer y Foll, 2009).

Para los datos generados por secuenciación genómica reducida; primeramente, se calcularon los F_{st} pareados entre las localidades que se encuentran distribuidas al norte de la brecha de Sinaloa, entre brechas (aquellas localidades situadas entre la brecha de Sinaloa y Centroamericana) y al sur de la brecha Centroamericana, para cuantificar el efecto de estas discontinuidades como barreras a la dispersión. Posteriormente se calcularon los F_{st} entre pares de localidades así como entre provincias biogeográficas, todo ello mediante el software GENODIVE (Meirmans, 2012), el cual utiliza un agrupamiento Bayesiano en función de un AMOVA, utilizando 1000 iteraciones con un nivel de significancia de $\alpha= 0.05$.

Además, dentro del pipeline del programa Stacks se empleó el comando *populations*, el cual permite calcular estadísticos generales de diversidad y estructura genética (heterocigosidad esperada, heterocigosidad observada, estadísticos F, diversidad nucleotídica) en función de un número de poblaciones putativo el cual puede ser editado (parámetro “p”) (Catchen *et al.*, 2011; Catchen, Hohenlohe, *et al.*, 2013). Se seleccionó el valor de $p=3$ debido a que maximizaba el número de sitios polimórficos a extraer para los análisis posteriores, además de tener sentido biológico con las zonas geográficas separadas por las Brechas de Sinaloa y Centroamericana.

Se utilizó un análisis discriminante de componentes principales (DAPC; (Jombart, Devillard y Balloux, 2010) para evaluar gráficamente el grado de diferenciación entre muestras; dicho análisis busca minimizar la varianza dentro de grupos, y maximizar la varianza dentro de ellos, de modo que se evalúa la distribución de la variación genética de los individuos. Este análisis se realizó con el paquete Adegenet en R (Team, 2013). Análogo a este análisis, se probó la posibilidad de encontrar un patrón de aislamiento por distancia debido a una estructuración genética críptica (Catchen, Bassham, *et al.*, 2013). Se realizó una prueba mantel (Mantel, 1967) correlacionando una matriz de distancia geográfica entre localidades de muestreo con una matriz de distancia euclidiana en función de las frecuencias alélicas (Edwards, 1970), también mediante el paquete Adegenet.

Finalmente, la estructura genética poblacional se evaluó mediante un método de agrupación bayesiano con el programa STRUCTURE (Pritchard, Stephens y Donnelly, 2000). Los resultados de STRUCTURE fueron visualizados en STRUCTURE HARVESTER (Earl, 2012) para obtener un resumen de las probabilidades y del ΔK descritos previamente en Evanno y colaboradores (2005).

2.6. Reconstrucción filogenética

La reconstrucción de árboles filogenéticos fue realizada con el programa RAxML (Stamatakis, 2014), el cual utiliza el método de máxima verosimilitud y puede ajustarse a distintos modelos de sustitución molecular. Se utilizó el soporte “bootstrap” con 10000 aleatorizaciones (Stamatakis, 2014).

Para la selección del mejor modelo de sustitución se utilizó jmodeltest-2.1.10 (Guindon y Gascuel, 2003; Darriba *et al.*, 2012), el cual realiza un análisis de probabilidad para arrojar el modelo de sustitución molecular de mayor ajuste. Se utilizaron los criterios de Akaike (AIC; (Akaike, 1974; Sugiura, 1978; Trust, 2016)), Bayesiano (BIC; (Schwarz, 1978)), así como el de teoría de decisión (DT; (Minin *et al.*, 2003)) para los sets de datos de las librerías 3RAD.

Capítulo 3. Resultados

3.1. Extracciones de ADN

Se extrajo ADN de 153 organismos distribuidos a lo largo del POT (ej. Fig. 1.0). No se observó una diferencia en la concentración y calidad del ADN a lo largo de todas las localidades, salvo en las muestras correspondientes a la localidad de Sonora, en las cuales la concentración superaba los 30 ng/l.

Tabla 2. Localidades del muestreo, número de muestras analizadas por localidad (n) y coordenadas geográficas de cada sitio según Google Earth (<https://www.google.com/earth/>).

Localidad	n	Coordenadas geográficas
Guaymas, Sonora	25	(28° 4'59.00"N, 111°16'32.30"O)
Mulegé, Baja California Sur	10	(26°25'52.33"N, 111°19'38.58"O)
La Paz, Baja California Sur	10	(23°22'30.00"N, 109°28'3.00"O)
Archipiélago de Revillagigedo, Colima	25	(18°43'31.78"N, 110°56'57.56"O)
Manzanillo, Colima	8	(18°41'48.46"N, 104° 2'28.08"O)
Ixtapa-Zihuatanejo, Guerrero	18	(17°39'3.58"N, 101°36'45.46"O)
Bahías de Huatulco, Oaxaca	12	(15°39'12.00"N, 96° 6'30.00"O)
Puntarenas, Costa Rica	17	(9°24'13.80"N, 84°10'8.80"O)
Panamá, Panamá	2	(7°28'40.70"N, 79°59'34.40"O)
Ecuador, Costa insular	2	(1°37'13.60"N, 80°49'35.50"O)
Archipiélago de Galápagos, Ecuador	24	(0°25'27.70"N, 90°21'30.40"O)

3.2 Diversidad mitocondrial

3.2.1 Identificación taxonómica de muestras de tejido

En el transcurso de la investigación se identificaron varias muestras que por sus niveles de divergencia mitocondrial y nuclear (SNPs) procedían de otras especies y no se incluyeron en análisis subsiguientes (Tabla III).

Tabla 3. Relación de muestras analizadas correspondientes a otras especies de teleósteos.

Muestra	Divergencia en genes analizados (p-value)			Identificación molecular
	CYTB	COI	SNPs * (x 10 ⁻³)	
RVG7	0.082	0.025	N.A	<i>Thalassoma sp.</i>
RVG9	0.066	0.075	0.494	<i>T. virens</i>
RVG10	N.A	0.104	0.422	<i>T. grammaticum</i>
RVG11	N.A	0.102	0.394	<i>T. grammaticum</i>
RVG22	N.A	0.104	0.404	<i>T. grammaticum</i>
GAL17	N.A	0.208	0.008	<i>Mycteroperca sp.</i>

*Las distancias por ML fueron calculadas mediante un estimado de máxima verosimilitud siguiendo Stamatakis (2014)

3.2.2 16s ARNr

Se amplificó un fragmento de 550 pb aproximadamente del gen 16S ribosomal, lo cual es comparable con el tamaño reportado en la bibliografía (Ivanova *et al.*, 2007). Tras haber analizado y editado las secuencias

se obtuvo un solo haplotipo en los 27 organismos provenientes de las 9 localidades. Debido a la nula variación genética del gen no se realizaron análisis subsiguientes.

3.2.3 Citocromo b (CYTB)

El producto de amplificación del CYTB es un fragmento de aproximadamente 450 pb, el cual es de menor tamaño al fragmento esperado (560 pb) reportado para peces óseos en la bibliografía (Kocher *et al.*, 1989; Bernardi *et al.*, 2004). Después de recortar los extremos ruidosos manualmente, el alineamiento resultante se redujo a 305 pb, en el que se identificó un solo haplotipo en las 41 secuencias analizadas.

3.2.4 Citocromo Oxidasa Subunidad I (COI)

El fragmento amplificado del locus COI, por su parte, tuvo una longitud de 655 pb, lo cual coincide con lo reportado en la bibliografía (Ward *et al.*, 2005). Tras la edición manual de las secuencias, el alineamiento de las 38 secuencias se redujo a 645 pb en el que se identificaron seis haplotipos. Dos de ellos se identificaron en 34 de las secuencias (ej. COL3, SON6), mientras que el resto se identificó en un solo organismo siendo tres el número máximo de sustituciones (Fig. 5.0). Estos seis haplotipos fueron ingresados en Arlequin (Excoffier y Lischer, 2010) para el cálculo de estadísticos de diversidad.

Los índices de diversidad haplotípica resultó ser de $h=0.105$, mientras que la diversidad nucleotídica presentó un valor de $\pi= 0.00020$. Dichos valores indican una baja diversidad haplotípica y nucleotídica, aunque mayor que respecto a los otros marcadores empleados en este estudio (Tabla IV).

	1	1	3	3	3	4	4	5	6
	0	9	0	7	8	0	1	7	1
	4	7	2	7	1	9	9	5	0
TLCOL3 new	A	A	G	G	G	G	T	G	
SON6	G
TLGAL2 R	A	A	C	?	?
TLGR7	?	G	A
TLMG6	?	?	C
TLSON15	?	?	?	A

Figura 4. Sitios variables en el gen COI. Los sitios se encuentran resaltados en amarillo. Los números representan la posición en el alineamiento. Los signos de interrogación corresponden a bases no identificadas (datos faltantes).

Tabla 4. Resumen de resultados para cada marcador mitocondrial utilizado en este estudio.

Marcador	Nº de individuos con secuencia	Nº de secuencias editadas	Tamaño de fragmentos obtenidos (pb)	Nº de haplotipos	Diversidad haplotípica (h)	Diversidad nucleotídica (π)
16S rARN	30	29	450	1	0	0
CYTB	63	41	305	1	0	0
COI	63	38	645	6	0.105	0.00020

3.3 Diversidad nuclear (SNPs)

3.3.1. Secuenciación 3RAD.

De las muestras disponibles, se enviaron a secuenciar 96 muestras de ADN cuya cantidad y calidad de ADN fue suficiente (>20 ng/ μ l; $A_{260/280}>1.8$), las muestras de las localidades de Panamá y Ecuador-Costa Insular no produjeron cantidad ni calidad de ADN adecuados para su análisis.

Tras la secuenciación masiva, se obtuvieron tres pares de archivos en formato FASTQ con diferentes identificadores o “tags” moleculares iTru7 empleados por Hoffberg y colaboradores (2016). Estos archivos fueron concatenados para generar dos archivos “Read 1” y “Read 2” resultado de la secuenciación pareada. Los archivos contuvieron un total de 192’610,804 lecturas. Al realizar el demultiplex para la obtención de

lecturas individuales y cobertura del genoma por individuo, se obtuvo un total de 185'129,486 lecturas, correspondiente al 95.4 % de lecturas crudas, con una cobertura promedio de 20X por individuo. Las lecturas descartadas son producto de una baja calidad en la secuenciación, por lo que la secuencia de esas lecturas es ambigua y podría proporcionar una fuente de error en el llamado de SNPs. De este 95.4%, en promedio 2% de las lecturas resultaron ser clones de PCR identificados mediante `clone_filter`, por lo que el número de secuencias viables se redujo a un total de 180'077,936 lecturas correspondientes al 93.4% de las lecturas crudas

Al ejecutar el pipeline de Stacks, se encontraron 6,289 sitios polimórficos con el set de parámetros cuyos resultados fueron más robustos. Se empleó un número mínimo de 5 lecturas idénticas para la formación de un stack ($m=5$), mientras que para la construcción del catálogo se optimizó el número de lecturas requeridas ($n=3$), así como agrupar al menos tres lecturas por locus para cada individuo ($M=3$). Este set optimizado resultó el más astringente al mismo tiempo que permitió maximizar el número de SNPs útiles para los análisis subsecuentes. Dichos análisis fueron realizados con un set de SNPs no ligados (mediante el comando `write_single_snp` en `populations`), evitando así sesgos por efecto de desequilibrio de ligamiento (Slatkin, 2008). Este parámetro `write_single_snp` permite obtener un SNP de cada locus RAD introduciendo así únicamente sitios polimórficos independientes. Por último, se homogenizaron las frecuencias alélicas mediante el parámetro `min_maf` ajustándolo a un valor de 0.25, resultando en un set final de 3,133 SNPs, los cuales se encontraron en una frecuencia alélica estandarizada evitando así sitios con una con baja representación en el genoma.

3.3.2. Diversidad.

Al analizar la diversidad genética de los SNPs se encontraron valores de heterocigosidad observada y esperada bajos, pero congruentes con los reportados para otras especies de peces (Catchen, Bassham, *et al.*, 2013; Jeffries *et al.*, 2016). Por otra parte, la diversidad nucleotídica presentó valores altos en todas las localidades con respecto a los genes mitocondriales analizados en este estudio (Tabla IV). Además, los valores del coeficiente de endogamia (ej. F_{is}) oscilaron entre 0.02-0.05. Este rango de valores concuerda con datos de otras especies con las cuales se ha trabajado secuenciación genómica reducida (Catchen, Bassham, *et al.*, 2013; Jeffries *et al.*, 2016).

Tabla 5. Estadísticos descriptivos de diversidad genética nuclear de los 3133 SNPs independientes en *T. lucasanum*. Se observan los valores por localidad. π : Diversidad nucleotídica; H_o : Heterocigosidad observada; H_e : Heterocigosidad esperada; Fis: Coeficiente de endogamia.

Localidad	% sitios polimórficos	H_o	H_e	π	Fis
Sonora	0.6929	0.0025	0.0031	0.0031	0.0014
La Paz	0.6901	0.0025	0.0030	0.0032	0.0015
Mulegé	0.6921	0.0024	0.0030	0.0031	0.0015
Revillagigedo	0.6891	0.0024	0.0029	0.0031	0.0015
Colima	0.6795	0.0020	0.0028	0.0032	0.0025
Guerrero	0.6742	0.0023	0.0028	0.0030	0.0017
Oaxaca	0.6887	0.0025	0.0029	0.0031	0.0014
Costa Rica	0.6718	0.0020	0.0028	0.0031	0.0024
Galápagos	0.6729	0.0024	0.0029	0.0030	0.0014

3.4. Patrones de estructura genética.

3.4.1. Concordancia ecológica: efecto de las discontinuidades en el hábitat rocoso

El set de 3,133 SNPs no ligados se usó para calcular los valores de F_{st} entre las localidades ubicadas al norte, entre, y al sur de las brechas de Sinaloa y Centroamericana; encontramos que sólo existe diferenciación genética en las localidades al sur de la Brecha Centroamericana respecto al resto (Tabla V).

Tabla 6. Valores de F_{st} pareado entre las localidades ubicadas al norte de la brecha de Sinaloa, entre brechas y al sur de la brecha centroamericana.

* Estadísticamente significativo; ** Altamente significativo; *** Muy significativo: $\alpha = 0.05$

Regiones geográficas	Norte de la Brecha de Sinaloa	Entre brechas	Sur de la brecha centroamericana
Norte de la brecha de Sinaloa	--	0.004	0.056
Entre brechas	NS	--	0.036

Sur de la brecha centroamericana	***	***	--
----------------------------------	-----	-----	----

3.4.2. Concordancia biogeográfica

Al evaluar mediante F_{st} el grado de diferenciación entre poblaciones de las distintas provincias biogeográficas se encontró diferenciación genética significativa entre las provincias Panámica, Cortés y de Grandes Islas (ej. Galápagos). Como era de esperarse, la diferenciación más importante se encontró asociada a las poblaciones de las Grandes Islas (Tabla VII).

Tabla 7. Estructuración poblacional a nivel biogeográfico en *T. lucasanum*. Los valores de F_{st} pareados entre provincias se encuentran marcados en color negro; El valor de significancia a un nivel de $\alpha= 0.05$ se encuentra identificado por asteriscos. GRI corresponde a la provincia de las Grandes Islas (Galápagos).

* Estadísticamente significativo; ** Altamente significativo; *** Muy significativo: $\alpha= 0.05$

Statistic	Cortés	Panámica	GRI
Cortés	--	0.01	0.057
Panámica	*	--	0.029
Grandes Islas	***	***	--

3.4.3. Patrones entre localidades.

Para identificar cuáles muestras fueron responsables de dicha diferenciación, se evaluaron los índices F_{st} entre pares de localidades. Se encontraron valores entre 0.003 y 0.066, de los cuales fueron estadísticamente significativos ($p < 0.05$) los de las localidades de Costa Rica y Galápagos respecto a la mayoría de las localidades (Tabla VI). Por su parte, la localidad de Guerrero, dentro de la clasificación como entre brechas, presentaba valores de F_{st} estadísticamente significativos con otras dos localidades (Mulegé, Revillagigedo), cuyo nivel de significancia desaparece al realizar una corrección de Bonferroni secuencial (Rice, 1989).

Tabla 8. Valores de Fst pareado entre localidades (Parte superior derecha). Valores de significancia (parte inferior izquierda) .

* Estadísticamente significativo; ** Altamente significativo; *** Muy significativo: $\alpha= 0.05$

	Sonora	Mulegé	La Paz	Revillagigedo	Colima	Guerrero	Oaxaca	Costa Rica	Galápagos
Sonora	--	0	0	0	0.005	0.003	0	0.066	0.062
Mulegé	NS	--	0.004	0	0	0.021	0.005	0.054	0.063
La Paz	NS	NS	--	0	0.003	0.009	0.003	0.06	0.058
Revillagigedo	NS	NS	NS	--	0	0.014	0	0.047	0.051
Colima	NS	NS	NS	NS	--	0.008	0.01	0.035	0.049
Guerrero	NS	NS	NS	NS	NS	--	0.012	0.052	0.042
Oaxaca	NS	NS	NS	NS	NS	NS	--	0.038	0.04
Costa Rica	***	**	**	**	NS	**	**	--	0
Galápagos	***	***	***	***	***	***	***	NS	--

Por su parte, el análisis exploratorio de DAPC reveló que las muestras de cada localidad son diferenciables unas de otras, puesto que se observa una clara agrupación de organismos de cada localidad, reflejando estructuración geográfica (Fig. 5).

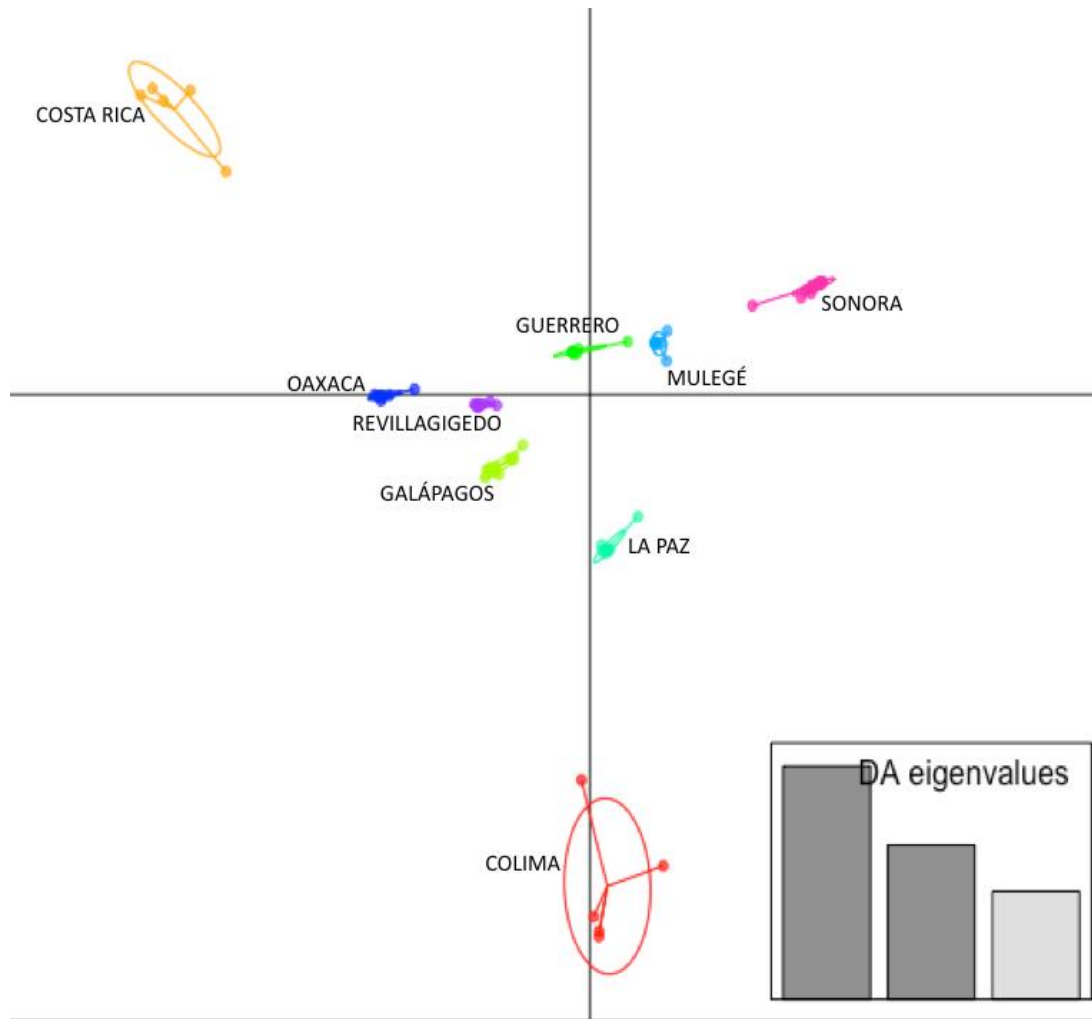


Figura 5. Distribución de la variación genética en función de 3133 SNPs independientes del arcoíris de Cortés, *T. lucasanum*. En la imagen los dos factores principales explican cerca del 40 % de la variación genética total, teniendo los valores de eigenvalores (discriminantes) más robustos en el modelo (diagrama de barras, recuadro inferior derecha).

Se observa que las localidades de Costa Rica y Colima presentan una composición genética distinta del resto de las localidades. Mientras que Colima se ve reflejada por una función discriminante distinta a la mayoría de las localidades (salvo La Paz), los individuos de Costa Rica tienen un nivel de variabilidad genética diferente al resto. Los eigenvectores (ej. 2 ejes) correspondientes a las funciones discriminantes de dicho análisis, las cuales a su vez representan casi el 40% de la variación total del modelo.

Por otro lado, al calcular el número de grupos más probables (denotado por “k”) mediante K medias en Adegenet, aquellos con los valores más bajos de BIC correspondían a las agrupaciones de k=9, lo cual

concuerta con el número de localidades. Por último, se encontró que no existe correlación entre la distancia genética entre las poblaciones y su distancia geográfica (prueba mantel, $r = -0.089$, $p > 0.05$).

La distribución de probabilidades posteriores obtenidas en STRUCTURE fue consistente con los resultados del DAPC, al encontrar cada grupo resuelto y sin traslapes en función de las localidades de muestreo. Sin embargo, el número de agrupamientos más probable fue de $K=3$ calculado por el método de Evanno. Al analizar los valores del coeficiente de ancestría (q) observamos que los organismos de Colima y La Paz presentan un componente genético (rojo) casi exclusivo que se encuentra también presente, aunque en menor proporción, en Costa Rica.

El segundo componente genético característico del golfo de California (verde) es compartido por Sonora, en donde es prácticamente exclusivo, y se encuentra en proporciones decrecientes hacia el sur, en Mulegé y en Guerrero. Por su parte, el tercer componente genético (azul) que se puede considerar característico del sur se encuentra al 100% en casi todos los organismos de Revillagigedo, Oaxaca, y Galápagos. Aunque también en Guerrero y hasta Mulegé. (Fig. 6).

Al analizar los gráficos de distribución posterior mediante un reagrupamiento de localidades en función: 1) de las barreras a la dispersión y 2) de la Biogeografía, no se encontraron cambios en el número de componentes genéticos ni tampoco en el valor de $\Delta K=3$ como el número de grupos más probable en función del número de Evanno. El único cambio presente es el cambio en la proporción de ancestría en Costa Rica, la cual se reduce a un único componente genético el cual comparte con localidades como Oaxaca y Revillagigedo.

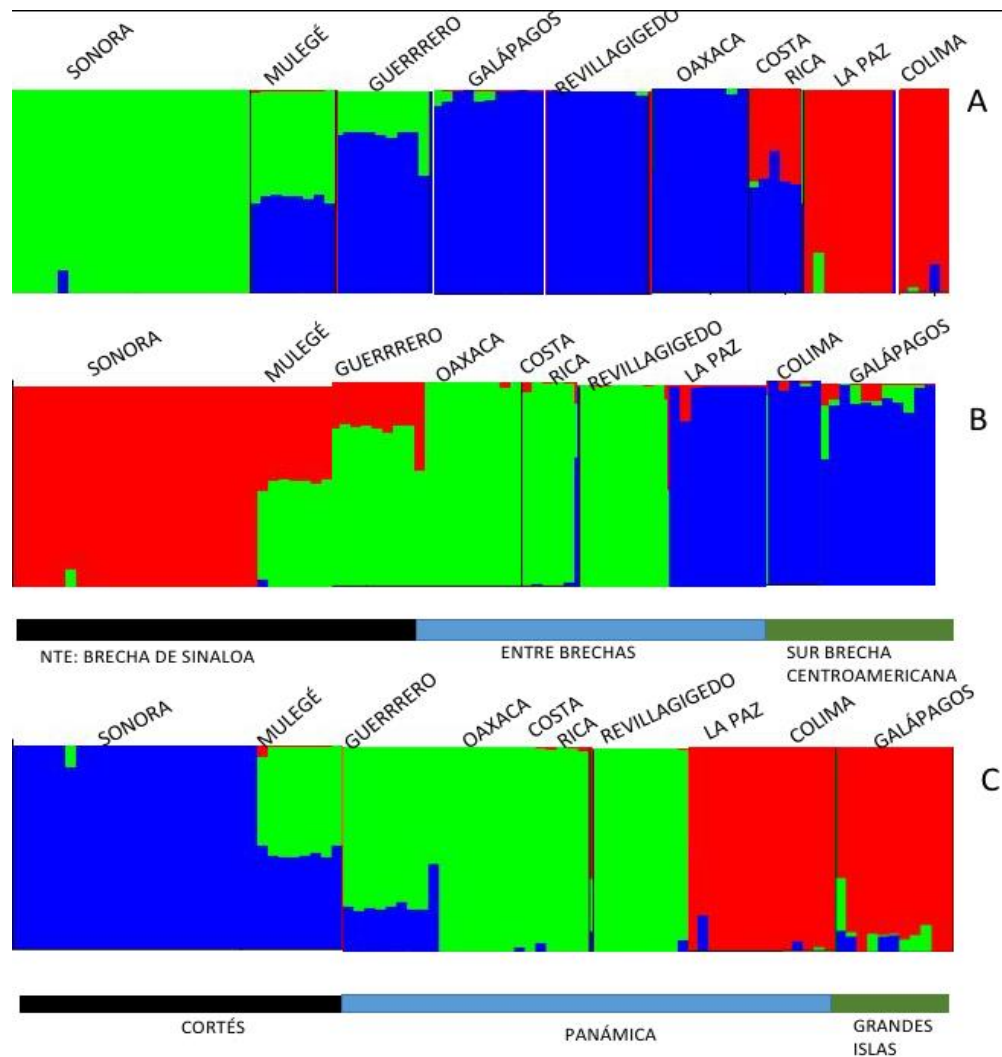


Figura 6. Agrupamiento Bayesiano de la estructuración poblacional en *T. lucasanum* con el valor calculado de más probable de agrupamiento ($K=3$). La gama de colores representa el coeficiente de ancestría (q) de cada agrupamiento en cada individuo en función de: A) Localidad; B) Barreras a la dispersión; C) Biogeografía.

3.5. Patrones filogeográficos

Para la reconstrucción filogenética basada en SNPs se utilizaron los archivos con los sitios variables y fijados aunados al resto de la secuencia en el alineamiento (opción `--phylip_var_all`). Se calcularon las probabilidades para cada modelo de sustitución y como resultado del cálculo de probabilidades el modelo más robusto para nuestro set de datos fue el “GTR” acoplado a heterogeneidad espacial en las tasas de sustitución de bases (Tavaré, 1986).

En general existe una concordancia entre el arreglo de cada uno de los clados formados respecto al origen geográfico de cada individuo. Así mismo, se observa que la mayor proporción de individuos pertenecientes a cada clado (coloración por rama) presenta una transición acorde a la geografía conforme se recorre el Pacífico Oriental Tropical de norte a sur. Se encontró también que los individuos pertenecientes a las localidades al sur de la brecha centroamericana (linajes magenta y púrpura en Figura 7) se agrupan como linajes hermanos a todos los demás distribuidos al norte. El único agrupamiento con un valor de bootstrap de 100 se encontró entre un individuo de “Guerrero” y uno de “Mulegé”.

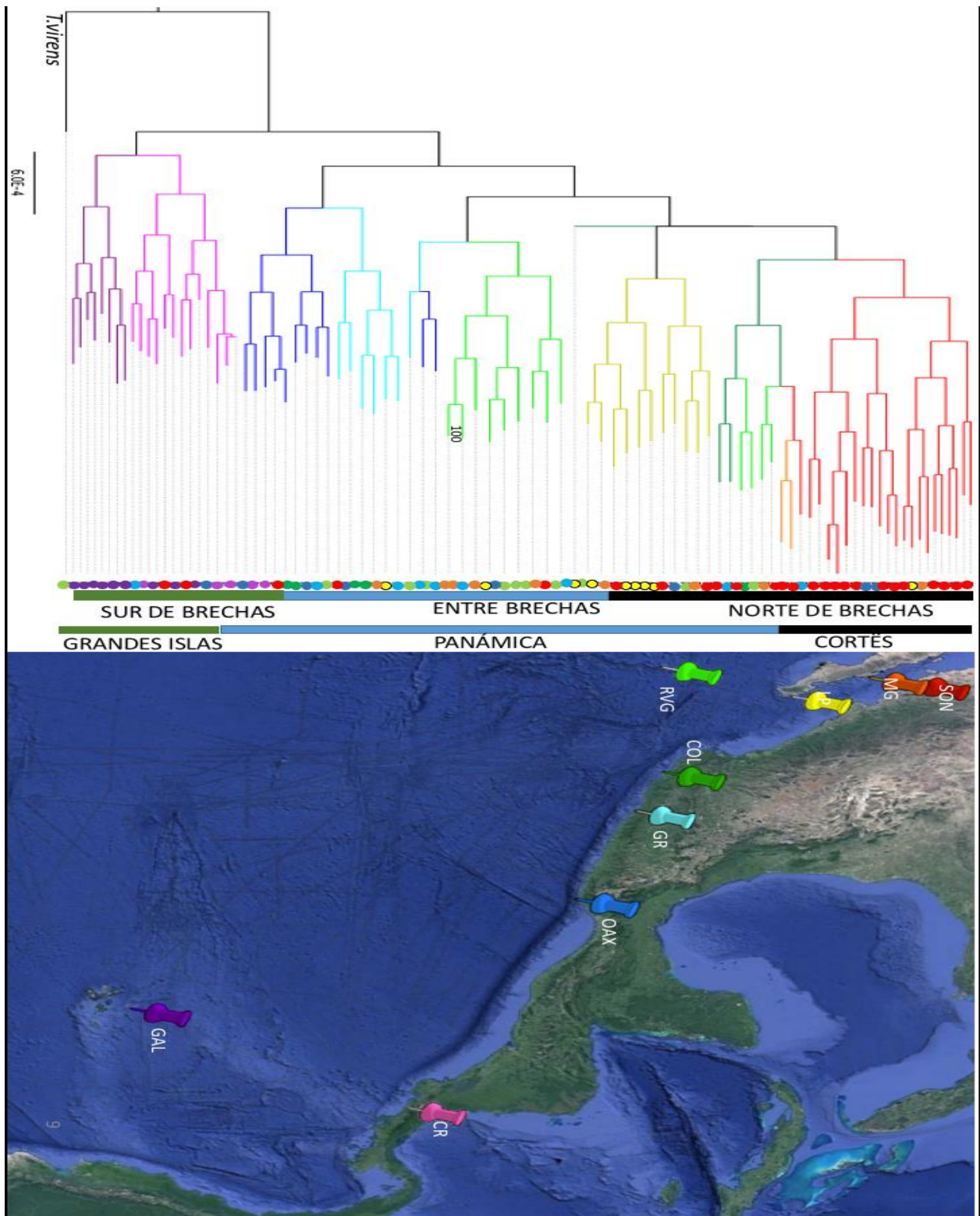


Figura 7. Reconstrucción filogenética del arcoíris de Cortés *Thalassoma lucasanum* en el POT, en función del mejor modelo de sustitución molecular mediante máxima verosimilitud. La gama de colores en los agrupamientos corresponde a Sonora (rojo), Mulegé (naranja), La Paz (dorado), Revillagigedo (lima), Colima (verde), Guerrero (aguamarina), Oaxaca (azul marino), Costa Rica (magenta), Galápagos (púrpura). Las barras de color negro, azul y verde corresponden a la demarcación de localidades respecto al norte de la brecha de Sinaloa (norte de brechas), entre y al sur de la brecha Centroamericana (sur de brechas). Se indica también la demarcación por provincias biogeográficas. Se muestran solamente valores con alto soporte estadístico (bootstrap >99).

Capítulo 4. Discusión

En este trabajo se usaron dos acercamientos con distintos marcadores moleculares para estimar la diversidad genética y estructura poblacional del pez arcoíris de Cortés, *Thalassoma lucasanum*. Por una parte se analizaron genes mitocondriales para evaluar si la baja variabilidad genética previamente encontrada en la región control (Dominguez-Rivera, 2011) se extendía a otros genes del genoma mitocondrial de la especie. Adicionalmente, se adoptó una perspectiva genómica (análisis de SNPs anónimos distribuidos en todo el genoma), para extender el número de loci analizados y evaluar la diversidad y estructura genéticas de la especie con un mayor poder estadístico. Usando estos SNPs polimórficos, se analizaron también los patrones de distribución geográfica de linajes (patrones filogeográficos) de la especie, evaluando así si existe concordancia entre la distribución de la especie en el POT respecto a el arreglo de linajes evolutivos de la misma.

4.1 Variabilidad genética de *Thalassoma lucasanum*

4.1.1. Variabilidad mitocondrial.

La variabilidad genética a nivel mitocondrial para el arcoíris de Cortés, *Thalassoma lucasanum*, parece ser consistentemente baja tanto a nivel nucleotídico como haplotípico. En este estudio se encontraron valores menores de diversidad haplotípica, así como un valor ligeramente mayor de diversidad nucleotídica, el cual pudiera estar enmascarado por la diferencia en el número de secuencias con los cuales se reportaron los valores para la Región Control de la especie (Dominguez-Rivera, 2011). Se observó un número menor de haplotipos que en los identificados por Domínguez Rivera (2011), así como una disminución en niveles de divergencia intraespecífico. El resultado es esperado, considerando que la región control no codificante generalmente es la que presenta la mayor tasa evolutiva del genoma mitocondrial. Debido a ello, las preguntas científicas a responder con este marcador difieren en general de aquellas que pueden ser esclarecidas con los marcadores empleados en este estudio. Mientras que la “CR” se ha empleado para determinar la estructura poblacional y demografía histórica de especies (Chen *et al.*, 2004; Haney, Silliman y Rand, 2007), se han reportado los genes mitocondriales empleados en este estudio (ej. 16s ARNr, CYTB y COI) en estudios para la reconstrucción de filogenias y filogeografía de especies en sistemas donde existen múltiples factores que influyen en la distribución de linajes (Bernardi *et al.*, 2004; Ward *et al.*, 2005;

Craig *et al.*, 2006). Debido a estos factores, nuestros resultados sugieren que el bajo polimorfismo originalmente en la región control por Dominguez-Rivera (2011) se extiende al resto del mitogenoma de *T. lucasanum* (Dominguez-Rivera, 2011). Anteriormente Rodriguez-Rivera (2011) había propuesto que el bajo polimorfismo podría deberse principalmente a una expansión demográfica reciente antecedida de un cuello de botella (Dominguez-Rivera, 2011). Otras posibles explicaciones de estos resultados podrían ser eventos de especiación espontánea ocurridos hace 5-10 millones de años (M.A) en especies del género *Thalassoma* (Bernardi *et al.*, 2004), con lo cual el tiempo de divergencia entre especies sería reciente, y por ende los marcadores de ADNmt empleados no tendrían la suficiente resolución. Este patrón de baja variabilidad también podría presentarse como resultado de la inversión de sexo de machos juveniles en hembras, resultado de la demografía sexual característica en el arcoiris de Cortés (Warner, 1982).

4.1.2. Utilidad de datos mitocondriales para corroborar la identidad taxonómica de muestras de tejido.

En las localidades de Revillagigedo y Galápagos se encontró al menos un organismo que fue identificado morfológicamente como *T. lucasanum*, cuyas secuencias de ADNmt y 3RAD fueron altamente divergentes del resto. La comparación de las secuencias de ADNmt permitió su identificación como de otras especies congénéricas (ej. *T. virens*, *T. grammaticum*), y un caso extremo en el cual uno de los individuos que no pudieron ser identificados a nivel de especie tenía los valores de identidad más altos con especies de la familia *Serranidae*. *T. virens* es una especie simpátrica del arcoiris de Cortés (Victor y Wellington, 2000), además de distribuirse en la zona del archipiélago de Revillagigedo, pudiendo migrar a zonas de Baja California durante periodos “El Niño” (<http://www.fishbase.org/summary/13821>;(Harrison, 1995)). Por otra parte, *T. grammaticum* presenta un rango de distribución geográfico amplio en el POT, pudiendo encontrarse además en las islas oceánicas (<http://www.fishbase.org/summary/Thalassoma-grammaticum.html>); (Bernardi *et al.*, 2004)). Esto denota la importancia de contar tanto con datos genéticos como morfológicos, en específico para especies con una alta variación en la coloración durante los estados de fase inicial y terminal (Warner, 1982).

Los niveles de divergencia inter e intraespecífico pueden ser otro factor discriminante para distinguir entre especies del género *Thalassoma*, como se maneja en el proyecto Código de Barras de ADN la Vida (Hebert, Ratnasingham y de Waard, 2003). Para el gen COI se calculó un porcentaje de divergencia intraespecífica

promedio de 0.3% entre los 38 organismos de *T. lucasanum* para los cuales se obtuvieron secuencias de buena calidad. Por otra parte, el porcentaje de divergencia interespecífico de especies del género *Thalassoma spp.* superó el 1.5 %. Dicho esto, nuestro estudio resalta la utilidad del marcado COI como un gen adecuado para la identificación taxonómica en taxones cuyos caracteres morfológicos resultan altamente similares, como en el caso del género *Thalassoma*.

4.1.3. Variabilidad genómica.

Los patrones de variabilidad genética nuclear presentaron mucho mayor resolución con respecto a los genes mitocondriales. Se encontró que el nivel de diversidad nuclear es mayor tanto al compararlos con los datos mitocondriales de la misma especie, como al compararlos con datos genómicos (p.ej., RADseq) de otras especies de teleósteos (Hohenlohe *et al.*, 2010; Catchen, Bassham, *et al.*, 2013; Jeffries *et al.*, 2016). Una excepción es el trabajo de Jansson y colaboradores (2016), quienes caracterizaron 149 SNPs del lábrido *Ctenolabrus rupestris*. De este total se validaron 48 SNPs aleatoriamente seleccionados y analizados genotípicamente como más polimórficos, los cuales presentaron un valor promedio de $H_E = 0.334$ (Jansson *et al.*, 2016), valor dos órdenes de magnitud mayor al promedio calculado para nuestros datos ($H_E = 0.0029$). Sin embargo, tanto el tamaño de muestra (ej. $N=150$) como el protocolo de obtención de SNPs en el pipeline y la exclusión de loci con valores de frecuencia alélica inferior (ej. MAF) a 0.3 pudieran ser parámetros en cuenta para dimensionar tanto las diferencias entre las preguntas de investigación como la metodología que se abordó en ambos casos.

4.2. Estructura poblacional de *T. lucasanum* en el POT

Los valores de F_{st} entre localidades separadas por las brechas de Sinaloa y Centroamericana, así como entre las provincias biogeográficas reflejan que existe estructuración genética en las localidades del sur separadas por la brecha Centroamericana, así como entre las tres provincias que conforman el Pacífico Oriental Tropical. Si bien existe diferenciación genética tanto a nivel de localidad como a nivel de biogeografía, los bajos valores de F_{st} indican una estructuración genética sutil, la cual ha sido descrita

tanto en peces (Catchen, Bassham, *et al.*, 2013) como en otros taxones distantes (ej., (Younger *et al.*, 2017)).

Los análisis de agrupamiento bayesiano coinciden en el número óptimo de “clusters” $k=3$ por el método de Evanno y colaboradores (2005), contrastando con el método multivariado empleado en Jombart y colaboradores (2010) el cual considera $k=9$, el número de localidades analizadas, como valor óptimo. Sin embargo, el gráfico de distribuciones posteriores (Fig. 6) permite observar cómo la separación entre cada agrupamiento coincide con las localidades de muestreo.

Por otra parte, la localidad de Colima presenta un componente genético que comparte mayoritariamente con Costa Rica y La Paz. Esto, aunque improbable, pudiera ser evidencia de una posible transición en la composición genética de los organismos debida a las discontinuidades del hábitat. No se descarta la posibilidad de un sutil flujo genético debido a gradientes ambientales que modulan la estructuración poblacional en especies aledañas a la costa (Saenz-Agudelo *et al.*, 2015). Además, dada la resolución de los grupos por localidad en el DAPC y el valor de $\Delta K=9$ mediante K-medias, es probable que los individuos de cada localidad sean subconjuntos con un componente genético intrínseco en cada localidad. El buscar darle sentido biológico a más de un nivel agrupamiento tiene validez en ciertos casos (Meirmans, 2015), tales como al encontrar una estructuración poblacional jerárquica (la cual pudiera ser el caso en el arcoíris de Cortés; revisar (Jombart, Devillard y Balloux, 2010)). Aunado a ello, ingresar los datos de la locación de cada individuo como un nivel de agrupamiento a priori (opción “LOCPRIO” en Structure; (Pritchard, Stephens y Donnelly, 2000)) pudiera ser un factor que modificara el número de grupos óptimos obtenidos, aunque en otros trabajos se ha sugerido que los cambios en el agrupamiento bayesiano no son significativos (Younger *et al.*, 2017).

Por último, para evaluar de manera integrativa los patrones de estructura poblacional de *T. lucasanum* queda pendiente el acoplar al componente genético el gradiente ambiental que, como es de esperarse, tendría efecto en la divergencia genética de distintas especies distribuidas en amplios sistemas geográficos (Saenz-Agudelo *et al.*, 2015).

4.2.1. Efecto de barreras a la dispersión en la estructura poblacional de *T. lucasanum*.

Se ha considerado que las barreras a la dispersión tienen efecto en la demografía histórica de especies con rangos de dispersión geográficamente amplios, reflejado en un patrón generalizado de aislamiento por distancia complementario a una diferenciación genética debida a una restricción en el flujo genético (Riginos, 2005; Lessios y Baums, 2017).

Debido al bajo polimorfismo encontrado en los distintos marcadores mitocondriales no fue posible caracterizar la estructura poblacional de *T. lucasanum* a nivel de mitogenoma. En otras especies de lábridos se han encontrado niveles bajos de diferenciación genética en el mitogenoma, los cuales se han atribuido a características intrínsecas de las especies más que a factores extrínsecos como el ambiente. Además, se ha hipotetizado la influencia de ciertos eventos históricos, tales como el cierre de las vías marítimas en el norte y sur de la península de Baja California, en la filogeografía de especies en el Pacífico norte (Bernardi, Findley y Rocha-Olivares, 2003).

Por otro lado, mediante los análisis realizados con SNPs no se descarta que, contrario a otras especies de peces en el POT, *Thalassoma lucasanum* presente una ligera ruptura genética traducida en diferenciación entre algunas de sus localidades, por lo que esta ruptura genética pudiera asociarse principalmente a las discontinuidades en el hábitat. Este patrón denota la utilidad de una aproximación multi-locus para especies con pocos recursos genómicos, debido a que anteriormente no se habían caracterizado patrones de estructuración genética debidos a las brechas de Sinaloa y Centroamericana, considerando datos de múltiples especies tanto con marcadores mitocondriales como nucleares (Lessios y Baums, 2017; Manzello, Eakin y Glynn, 2017).

4.3. Patrones filogeográficos y concordancias génicas

En general la filogenia de *T. lucasanum* presenta una concordancia filogeográfica entre los linajes representados en cada clado y su distribución geográfica.

La concordancia en los agrupamientos coincide en general con la delimitación de provincias biogeográficas propuesta por Briggs y Bowen (2012), quienes agruparon la provincia del Pacífico Mexicano dentro de la provincia Panámica. Esto debido a que el clado que agrupa los linajes encontrados en Costa Rica y Galápagos se encuentra únicamente conectado con individuos de Guerrero y Oaxaca. Sorprendentemente,

existen un par de puntos correspondientes a la localidad de “Sonora” que yace dentro de este clado. Esto se explica debido a que en el DAPC se observan los subconjuntos de Mulegé, Guerrero y Sonora como más cercanos entre sí, y a su vez la mayor parte de los grupos se encuentran estrechamente relacionados (salvo el caso de Colima y Costa Rica). Debido a la similitud en la composición genética de los individuos de Guerrero con los de Sonora es probable que la reconstrucción filogenética hubiese agrupado algunos individuos de estas localidades dentro de la rama en la cual Costa Rica y Galápagos se agrupan. Todo esto correspondería a una débil separación de los linajes, característico de una posible divergencia reciente entre los linajes del sur separados por la brecha Centroamericana respecto a los ubicados al norte de esta barrera.

Por último, se observó en la reconstrucción filogenética que los individuos correspondientes a Galápagos en general presentan un componente genético similar únicamente al componente correspondiente a Costa Rica. Esto podría ser indicio de una estructuración genética en una provincia aislada (ej. Grandes islas), lo cual concuerda con la demarcación de la provincia de las Grandes Islas (GRI) en función de los niveles de endemismo mencionados en Briggs y Bowen (2012).

La restricción del flujo genético entre las localidades al sur de la brecha Centroamericana con el resto del POT reflejada en diferenciación genética, podría deberse al aislamiento producido por la brecha centroamericana como barrera a la dispersión. Esta zona podría mostrar conectividad con las localidades de Oaxaca y Guerrero, ubicadas al sur de lo que anteriormente se consideraba a nivel biogeográfico como el Pacífico Mexicano (Briggs, 1974; Hastings, 2000).

Capítulo 5. Conclusiones.

- Existe una estructuración genética sutil entre las localidades al sur y al norte de la brecha Centroamericana debida, probablemente, a la biología y demografía de la especie más que por un patrón de aislamiento por distancia.
- Existen concordancias filogeográficas entre la distribución de los linajes con la geografía del Pacífico Oriental Tropical. Los organismos al sur y al norte de la brecha Centroamericana representan linajes independientes.
- La baja variabilidad genética mitocondrial se extiende a todos los genes analizados en este estudio. Parece improbable que exista mayor diversidad genética cuantificable en alguna otra región del mitogenoma de *Thalassoma lucasanum*. Esta baja diversidad genética estaría relacionada con la biología de la especie, algún proceso de expansión demográfica, y en última instancia debido a la naturaleza misma de la molécula de ADNmt.
- Este trabajo denota el mayor poder y resolución de los análisis genómicos para detectar niveles de estructuración genética poblacional y filogeografía sutiles como los de esta especie.
- Optimizar el tamaño muestral, el rango de frecuencias alélicas y la restricción a ciertos valores de heterocigosidad en los marcadores bialélicos podrían ser factores clave para caracterizar de mejor manera la variación genómica a través de todo el genoma del arcoíris de Cortés.

Literatura citada.

- Ackermann, R. R. y Cheverud, J. M. (2004) "Detecting genetic drift versus selection in human evolution.", *Proceedings of the National Academy of Sciences of the United States of America*, 101(52), pp. 17946–51. doi: 10.1073/pnas.0405919102.
- Akaike, H. (1974) "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, 19(6), pp. 716–723. doi: 10.1109/TAC.1974.1100705.
- Alvarez-Filip, L., Reyes-Bonilla, H. y Calderon-Aguilera, L. E. (2006) "Community structure of fishes in Cabo Pulmo Reef, Gulf of California", *Marine Ecology*, 27(3), pp. 253–262. doi: 10.1111/j.1439-0485.2006.00108.x.
- Avise, J. C. (2000) *Phylogeography: the history and formation of species*. Harvard university press.
- Avise, J. C. (2012) *Molecular markers, natural history and evolution*. Springer Science & Business Media.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. y Johnson, E. A. (2008) "Rapid SNP discovery and genetic mapping using sequenced RAD markers", *PLoS ONE*, 3(10), pp. 1–7. doi: 10.1371/journal.pone.0003376.
- Bernardi, G., Bucciarelli, G., Costagliola, D., Robertson, D. R. y Heiser, J. B. (2004) "Evolution of coral reef fish *Thalassoma* spp. (Labridae). 1. Molecular phylogeny and biogeography", *Marine Biology*, 144(2), pp. 369–375. doi: 10.1007/s00227-003-1199-0.
- Bernardi, G., Findley, L. y Rocha-Olivares, A. (2003) "Vicariance and Dispersal Across Baja California in Disjunct Marine Fish Populations", *Evolution*, 57(7), p. 1599. doi: 10.1554/02-669.
- Briggs, J. (1974) "Marine zoogeography", *New York: McGraw-Hill*, 1, p. 475.
- Briggs, J. C. y Bowen, B. W. (2012) "A realignment of marine biogeographic provinces with particular reference to fish distributions", *Journal of Biogeography*, 39(1), pp. 12–30. doi: 10.1111/j.1365-2699.2011.02613.x.
- Bromham, L. (2009) "Why do species vary in their rate of molecular evolution?", *Biology letters*, 5(3), pp. 401–404. doi: 10.1098/rsbl.2009.0136.
- Catchen, J., Bassham, S., Wilson, T., Currey, M. y Brien, C. O. (2013) "The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing", pp. 2864–2883. doi: 10.1111/mec.12330.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. y Cresko, W. A. (2013) "Stacks: An analysis tool set for population genomics", *Molecular Ecology*, 22(11), pp. 3124–3140. doi: 10.1111/mec.12354.
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H. y De Koning, D.-J. (2011) "Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences", *G3 & Genes/Genomes/Genetics*, 1(3), pp. 171–182. doi: 10.1534/g3.111.000240.
- Chen, C. A., Ablan, M. C. A., McManus, J. W., Bell, J. D., Tuan, V. S., Cabanban, A. S. y Shao, K. T. (2004) "Population structure and genetic variability of six bar wrasse (*Thalassoma hardwicki*) in northern South

China Sea revealed by mitochondrial control region sequences”, *Marine Biotechnology*, 6(4), pp. 312–326. doi: 10.1007/s10126-003-0028-2.

Craig, M. T., Hastings, P. A., Pondella, D. J., Ross Robertson, D. y Rosales-Casián, J. A. (2006) “Phylogeography of the flag cabrilla *Epinephelus labriformis* (Serranidae): Implications for the biogeography of the Tropical Eastern Pacific and the early stages of speciation in a marine shore fish”, *Journal of Biogeography*, 33(6), pp. 969–979. doi: 10.1111/j.1365-2699.2006.01467.x.

Darriba, D., Taboada, G. L., Doallo, R. y Posada, D. (2012) “jModelTest 2: more models, new heuristics and parallel computing”, *Nat Meth.* Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 9(8), p. 772. Disponible en: <http://dx.doi.org/10.1038/nmeth.2109>.

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K. y Blaxter, M. L. (2013) “Special features of RAD Sequencing data: Implications for genotyping”, *Molecular Ecology*, 22(11), pp. 3151–3164. doi: 10.1111/mec.12084.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M. y Blaxter, M. L. (2011) “Genome-wide genetic marker discovery and genotyping using next-generation sequencing”, *Nat Rev Genet.* Nature Publishing Group, 12(7), pp. 499–510. doi: 10.1038/nrg3012.

Dominguez-Rivera, A. (2011) *Variabilidad genética de la región control mitocondrial del pez arrecifal *Thalassoma lucasanum* en el Pacífico Tropical Mexicano*. CICESE.

Earl, D. A. (2012) “STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method”, *Conservation genetics resources*. Springer, 4(2), pp. 359–361.

Edwards, A. W. F. (1970) “Distances between populations on the basis of gene frequencies.”, *Heredity*, 25(Pt. 1).

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. a, Cresko, W. a, Bradshaw, W. E. y Holzapfel, C. M. (2010) “Resolving postglacial phylogeography using high-throughput sequencing.”, *Proceedings of the National Academy of Sciences of the United States of America*, 107, pp. 16196–16200. doi: 10.1073/pnas.1006538107.

Evanno, G., Regnaut, S. y Goudet, J. (2005) “Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study”, *Molecular ecology*. Wiley Online Library, 14(8), pp. 2611–2620.

Excoffier, L., Hofer, T. y Foll, M. (2009) “Detecting loci under selection in a hierarchically structured population.”, *Heredity*. Nature Publishing Group, 103(4), pp. 285–298. doi: 10.1038/hdy.2009.74.

Excoffier, L. y Lischer, H. E. L. (2010) “Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows”, *Molecular Ecology Resources*, 10(3), pp. 564–567. doi: 10.1111/j.1755-0998.2010.02847.x.

Giraldo, A., Gómez, C. y Ospina, F. (2001) “Abundance and spatial distribution of *Thalassoma lucasanum* (Gill, 1863) in a coral reef of the eastern tropical pacific”, *Bulletin of Marine Science*, 68(1), pp. 147–152.

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R. G., Martino, J. A., Pierson, T., Rogers, S. M., Wilson, J. Y. y Somers, C. M. (2015) “Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq)”, *Molecular Ecology Resources*, 15(6), pp.

1304–1315. doi: 10.1111/1755-0998.12404.

Guindon, S. y Gascuel, O. (2003) “A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood”, *Systematic Biology*, 52(5), pp. 696–704. doi: 10.1080/10635150390235520.

Haney, R. A., Silliman, B. R. y Rand, D. M. (2007) “A multi-locus assessment of connectivity and historical demography in the bluehead wrasse (*Thalassoma bifasciatum*)”, *Heredity*, 98(5), pp. 294–302. doi: 10.1038/sj.hdy.6800940.

Harrison, I. J. (1995) “Mugilidae: Lisas. In: Guia FAO para Identificación de Especies para lo Fines de la Pesca. Pacífico Centro-Oriental”, *Vol. III Part 2*, pp. 1293–1298.

Hastings, P. A. (2000) “Biogeography of the Tropical Eastern Pacific: distribution and phylogeny of chaenopsid fishes”, *Zoological Journal of the Linnean Society*, pp. 319–335. doi: 10.1006/zjls.1998.0196.

Hebert, P. D. N., Ratnasingham, S. y de Waard, J. R. (2003) “Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species”, *Proceedings of the Royal Society of London B: Biological Sciences*. The Royal Society, 270(Suppl 1), pp. S96–S99.

Hernández Olalde, L. (2008) “Ecología reproductiva de peces de arrecife rocoso en el Suroeste del Golfo de California”, p. 184.

Hodel, R. G. J., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., Gitzendanner, M. A., Douglas, N. A., Germain-Aubrey, C. C., Chen, S., Soltis, D. E. y Soltis, P. S. (2016) “The Report of My Death was an Exaggeration: A Review for Researchers Using Microsatellites in the 21st Century”, *Applications in Plant Sciences*, 4(6), p. 1600025. doi: 10.3732/apps.1600025.

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R. y Glenn, T. C. (2016) “RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data”, *Molecular Ecology Resources*, 16(5), pp. 1264–1278. doi: 10.1111/1755-0998.12566.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. y Cresko, W. A. (2010) “Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags”, *PLoS Genetics*, 6(2). doi: 10.1371/journal.pgen.1000862.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H. y Hebert, P. D. N. (2007) “Universal primer cocktails for fish DNA barcoding”, *Molecular Ecology Notes*, 7(4), pp. 544–548. doi: 10.1111/j.1471-8286.2007.01748.x.

Jansson, E., Taggart, J. B., Wehner, S., Dahle, G., Quintela, M., Mortensen, S., Kvamme, B. O. y Glover, K. A. (2016) “Development of SNP and microsatellite markers for goldsinny wrasse (*Ctenolabrus rupestris*) from ddRAD sequencing data”, *Conservation Genetics Resources*, 8(3), pp. 201–206. doi: 10.1007/s12686-016-0532-0.

Jeffries, D. L., Copp, G. H., Lawson Handley, L., Olsén, K. H., Sayer, C. D. y Hänfling, B. (2016) “Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L.”, *Molecular Ecology*. doi: 10.1111/mec.13613.

Jombart, T., Devillard, S. y Balloux, F. (2010) “Discriminant analysis of principal components: a new method for the analysis of genetically structured populations”, *BMC genetics*. BioMed Central, 11(1), p. 94.

- Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X., Wilson, A. C., Pääbo, S., Villablanca, F. X. y Wilson, A. C. (1989) "Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers.", *Proceedings of the National Academy of Sciences of the United States of America*, 86(16), pp. 6196–6200. doi: 10.1073/pnas.86.16.6196.
- Kumar, S., Stecher, G. y Tamura, K. (2016) "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets", *Molecular Biology and Evolution*, 33(7), pp. 1870–1874. Disponible en: <http://dx.doi.org/10.1093/molbev/msw054>.
- Lessios, H. A. y Baums, I. B. (2017) "Gene Flow in Coral Reef Organisms of the Tropical Eastern Pacific", en Glynn, P. W., Manzello, D. P., y Enochs, I. C. (eds.) *Coral Reefs of the Eastern Tropical Pacific: Persistence and Loss in a Dynamic Environment*. Dordrecht: Springer Netherlands, pp. 477–499. doi: 10.1007/978-94-017-7499-4_16.
- Liang, J. H., McWilliams, J. C. y Gruber, N. (2009) "High-frequency response of the ocean to mountain gap winds in the northeastern tropical Pacific", *Journal of Geophysical Research: Oceans*, 114(12), pp. 1–12. doi: 10.1029/2009JC005370.
- Librado, P. y Rozas, J. (2009) "DnaSP v5: A software for comprehensive analysis of DNA polymorphism data", *Bioinformatics*, 25(11), pp. 1451–1452. doi: 10.1093/bioinformatics/btp187.
- Limborg, M. T., Helyar, S. J., De Bruyn, M., Taylor, M. I., Nielsen, E. E., Ogden, R., Carvalho, G. R. y Bekkevold, D. (2012) "Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*)", *Molecular Ecology*, 21(15), pp. 3686–3703. doi: 10.1111/j.1365-294X.2012.05639.x.
- Ludt, W. B., Bernal, M. A., Bowen, B. W. y Rocha, L. A. (2012) "Living in the past: Phylogeography and population histories of Indo-Pacific wrasses (genus *Halichoeres*) in shallow lagoons versus outer reef slopes", *PLoS ONE*, 7(6), pp. 1–11. doi: 10.1371/journal.pone.0038042.
- Mantel, N. (1967) "The detection of disease clustering and a generalized regression approach", *Cancer research*. AACR, 27(2 Part 1), pp. 209–220.
- Manzello, D. P., Eakin, C. M. y Glynn, P. W. (2017) *Coral Reefs of the Eastern Tropical Pacific*. doi: 10.1007/978-94-017-7499-4.
- Meirmans, P. G. (2012) "AMOVA-based clustering of population genetic data", *Journal of Heredity*. Oxford University Press US, 103(5), pp. 744–750.
- Meirmans, P. G. (2015) "Seven common mistakes in population genetics and how to avoid them", *Molecular Ecology*, 24(13), pp. 3223–3231. doi: 10.1111/mec.13243.
- Miller, M., Dunham, J., Amores, a, Cresko, W. y Johnson, E. (2007) "genotyping using restriction site associated DNA (RAD) markers", *Genome Research*, 17, pp. 240–248. doi: 10.1101/gr.5681207.
- Minin, V., Abdo, Z., Joyce, P. y Sullivan, J. (2003) "Performance-Based Selection of Likelihood Models for Phylogeny Estimation", *Systematic Biology*, 52(5), pp. 674–683. doi: 10.1080/10635150390235494.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. y Hoekstra, H. E. (2012) "Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species", *PLoS ONE*, 7(5). doi: 10.1371/journal.pone.0037135.

- Pritchard, J. K., Stephens, M. y Donnelly, P. (2000) "Inference of population structure using multilocus genotype data", *Genetics*, 155(2), pp. 945–959. doi: 10.1111/j.1471-8286.2007.01758.x.
- Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q. y Shank, T. M. (2013) "Going where traditional markers have not gone before: Utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics", *Molecular Ecology*, 22(11), pp. 2953–2970. doi: 10.1111/mec.12228.
- Rice, W. R. (1989) "Analyzing tables of statistical tests", *Evolution*. JSTOR, 43(1), pp. 223–225.
- Riginos, C. (2005) "Cryptic vicariance in Gulf of California fishes parallels vicariant patterns found in Baja California mammals and reptiles.", *Evolution; international journal of organic evolution*, 59(12), pp. 2678–2690. doi: 10.1554/05-257.1.
- Robertson, D. R. y Cramer, K. L. (2009) "Shore fishes and biogeographic subdivisions of the Tropical Eastern Pacific", *Marine Ecology Progress Series*, 380(Ekman 1953), pp. 1–17. doi: 10.3354/meps07925.
- Saenz-Agudelo, P., DiBattista, J. D., Piatek, M. J., Gaither, M. R., Harrison, H. B., Nanninga, G. B. y Berumen, M. L. (2015) "Seascape genetics along environmental gradients in the Arabian Peninsula: Insights from ddRAD sequencing of anemonefishes", *Molecular Ecology*, 24(24), pp. 6241–6255. doi: 10.1111/mec.13471.
- Schwarz, G. (1978) "Estimating the Dimension of a Model", *The Annals of Statistics*, 6(2), pp. 461–464. doi: 10.1214/aos/1176344136.
- Slatkin, M. (2008) "Linkage disequilibrium--understanding the evolutionary past and mapping the medical future.", *Nature reviews. Genetics*, 9(6), pp. 477–85. doi: 10.1038/nrg2361.
- Stamatakis, A. (2014) "RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies", *Bioinformatics*, 30(9), pp. 1312–1313. doi: 10.1093/bioinformatics/btu033.
- Sugiura, N. (1978) "Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's", *Communications in Statistics-Theory and Methods*. Taylor & Francis, 7(1), pp. 13–26.
- Tariel, J., Longo, G. C. y Bernardi, G. (2016) "Molecular Phylogenetics and Evolution Tempo and mode of speciation in *Holacanthus* angelfishes based on RADseq markers", *Molecular Phylogenetics and Evolution*. Elsevier Inc., 98, pp. 84–88. doi: 10.1016/j.ympev.2016.01.010.
- Tavaré, S. (1986) "Some probabilistic and statistical problems in the analysis of DNA sequences", *Lectures on mathematics in the life sciences*, 17(2), pp. 57–86.
- Team, R. (2013) "R development core team", *RA Lang Environ Stat Comput*, 55, pp. 275–286.
- Templeton, A. R. (2006) *Population genetics and microevolutionary theory*. John Wiley & Sons.
- Trust, B. (2016) "Biometrika Trust Regression and Time Series Model Selection in Small Samples Author (s): Clifford M . Hurvich and Chih-Ling Tsai Published by : Oxford University Press on behalf of Biometrika Trust Stable URL : <http://www.jstor.org/stable/2336663> REFERE", 76(2), pp. 297–307.
- Victor, B. y Wellington, G. M. (2000) "Endemism and the pelagic larval duration of reef fishes in the

eastern Pacific Ocean”, *Marine Ecology Progress Series*, 205, pp. 241–248. doi: 10.3354/meps205241.

Vignal, A., Milan, D., SanCristobal, M. y Eggen, A. (2002) “A review on SNP and other types of molecular markers and their use in animal genetics”, *Genetics Selection Evolution*, 34(3), p. 275. doi: 10.1186/1297-9686-34-3-275.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A. y Seehausen, O. (2013) “Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation”, *Molecular Ecology*, 22(3), pp. 787–798. doi: 10.1111/mec.12023.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. y Hebert, P. D. N. (2005) “DNA barcoding Australia’s fish species.”, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(September), pp. 1847–1857. doi: 10.1098/rstb.2005.1716.

Warner, R. R. (1982) “Mating systems, sex change and sexual demography in the rainbow wrasse, *Thalassoma lucasanum*”, *Copeia*. JSTOR, pp. 653–661.

Yeaman, S. y Whitlock, M. C. (2011) “The genetic architecture of adaptation under migration–selection balance”, *Evolution*. Wiley Online Library, 65(7), pp. 1897–1911.

Younger, J. L., Clucas, G. V, Kao, D., Rogers, A. D., Gharbi, K., Hart, T. y Miller, K. J. (2017) “The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins”, (December 2016), pp. 3883–3897. doi: 10.1111/mec.14172.