



CENTRO DE INVESTIGACIONES BIOLÓGICAS
DEL NOROESTE, S.C.

Programa de Estudios de Posgrado

“PAISAJE MARINO GENÓMICO DEL GOBIO
Lythrypnus dalli (GILBERT 1890) EN EL GOLFO DE
CALIFORNIA”

T E S I S

Que para obtener el grado de

Maestro en Ciencias

Uso, Manejo y Preservación de los Recursos Naturales

Orientación en Biología Marina

P r e s e n t a

Camila Mac Loughlin Aleman

La Paz, Baja California Sur, junio de 2019

ACTA DE LIBERACIÓN DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 15 horas del día 06 del Mes de Junio del 2019, se procedió por los abajo firmantes, miembros de la Comisión Revisora de Tesis avalada por la Dirección de Estudios de Posgrado y Formación de Recursos Humanos del Centro de Investigaciones Biológicas del Noroeste, S. C., a liberar la Tesis de Grado titulada:

"Paisaje marino genómico del gobio *Lythrypnus dalli* (Gilbert 1890) en el Golfo de California"

Presentada por la alumna:

Camila Mac Loughlin Aleman

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

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

LA COMISIÓN REVISORA



Dr. Fausto Valenzuela Quiñones
Co-Director de Tesis



Dr. Eduardo Francisco Balart Páez
Co-Director de Tesis



Dr. Raul Octavio Martínez Rincón
Co-Tutor



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

Conformación de Comités

Comité Tutorial

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

Dr. Eduardo Francisco Balart Páez
Centro de Investigaciones Biológicas del Noroeste, S.C.
Co-Director de tesis

Dr. Raúl Octavio Martínez Rincón
CONACyT - Centro de Investigaciones Biológicas del Noroeste, S.C.
Co-tutor de tesis

Comité Revisor de Tesis

Dr. Fausto Valenzuela Quiñonez
Dr. Eduardo Francisco Balart Páez
Dr. Raúl Octavio Martínez Rincón

Jurado de Examen

Dr. Fausto Valenzuela Quiñonez
Dr. Eduardo Francisco Balart Páez
Dr. Raúl Octavio Martínez Rincón

Suplente

Dr. Juan Antonio de Anda Montañez

Resumen

La genómica del paisaje marino se enfoca en relacionar la estructura genética espaciotemporal a características del paisaje, como patrones de conectividad y gradientes ambientales. Un creciente número de evidencias ha puesto en duda el paradigma de homogeneidad genética en especies marinas, con fuertes implicaciones sobre las estrategias de manejo y conservación actuales. El Golfo de California (GC) posee complejos patrones oceanográficos y gran variabilidad ambiental, por lo que ofrece oportunidades únicas para probar hipótesis acerca del efecto de los factores que moldean la diversidad y estructura genética. El modelo de estudio es el gobio *Lythrypnus dalli*, pez criptobéntico de hábitats rocosos y coralinos. Posterior a su fase larval móvil presentan gran fidelidad al sitio, y su distribución geográfica atraviesa condiciones ambientales muy heterogéneas, por lo que sería esperable encontrar poblaciones con adaptaciones locales. El objetivo del presente trabajo es determinar la influencia de la distancia geográfica u oceanográfica sobre la estructura genómica neutral, y la influencia de la heterogeneidad ambiental sobre los patrones de diversidad adaptativa de *L. dalli* en el GC. Para ello, se colectaron individuos mediante buceo en 7 localidades de la costa occidental del GC, se obtuvieron patrones de estructura neutral (1,261 SNPs) y adaptativa (13 SNPs) a partir de librerías genómicas (ddRAD), y se compararon contra patrones de conectividad (derivados de simulaciones basadas en el flujo de corrientes del GC), temperatura superficial y clorofila-a (obtenidos de imágenes satelitales). La estructura genómica neutral indicó panmixia. No obstante, se detectó variación en los niveles de parentesco entre sitios y un patrón de aislamiento por distancia oceanográfica. En conjunto, los resultados sugieren una estructura metapoblacional con niveles sustanciales de conectividad, en la cual localidades norteñas fungen como fuentes de migrantes hacia el centro y sur del GC, de acuerdo a las corrientes predominantes durante la primavera. La búsqueda de patrones de diversidad adaptativa, en cambio, no fue posible debido a la cantidad y calidad de marcadores relacionados a procesos selectivos detectados. Esto se debe posiblemente a la baja cobertura genómica y espacial obtenida. Estos resultados proveen un aporte para refinar las estrategias de manejo y conservación de la biodiversidad en el GC.

Palabras Clave: *genómica del paisaje, conectividad, adaptación, ddRAD, Lythrypnus dalli.*

Vo. Bo.

Dr. Fausto Valenzuela Quiñonez
Co-Director de Tesis

Vo. Bo.

Dr. Eduardo Balart Páez
Co-Director de Tesis

Summary

Seascape genomics focuses on the relationship between spatiotemporal patterns of genetic structure and seascape features, like ocean currents and environmental gradients. An increasing number of evidences have recently challenged the genetic homogeneity paradigm in marine species, with strong implications to current management and conservation policies. The Gulf of California (GC) encompasses complex oceanographic currents and strong asymmetrical flow, in addition to high environmental heterogeneity, resulting in unique opportunities to test hypotheses about determinants of neutral and adaptive genetic structure patterns. Our model species was the small blue-banded goby *Lythrypnus dalli*, a small crypto-benthic reef fish. After a mobile larval phase, they have great site fidelity and a wide distribution throughout heterogeneous environmental conditions, which might allow local adaptations to occur. Our main objective is to determine the relative contribution of geographic or oceanographic distance over the neutral genetic structure, and the contribution of environmental heterogeneity over the adaptive genetic structure of *L. dalli* in the GC. Samples were collected from 7 sites in the GC, neutral (1,261 SNPs) and adaptive (13 SNPs) spatial structures were obtained from genomic libraries (ddRAD). Population structure was correlated to oceanographic connectivity hypotheses based on current flow, and environmental factors (temperature and chlorophyll) as predictors. Neutral genomic pattern indicates panmixia. However, variable relatedness levels and isolation by oceanographic distance were also detected. Overall, results point out to a metapopulation genetic structure with substantial connectivity, in which northernmost sites act as sources of migrants towards central and southern locations, according to predominant current flow during spring in the peninsular coast of the GC. The search for adaptive genomic patterns was limited given the amount and quality of detected candidates for selection, probably due to the low genomic and spatial coverage. This results help refine management and conservation strategies to maintain biodiversity in the GC.

Key words: *seascape genomics, connectivity, adaptation, ddRAD, Lythrypnus dalli.*

Vo. Bo.

Dr. Fausto Valenzuela Quiñonez
Co-Director de Tesis

Vo. Bo.

Dr. Eduardo Balart Páez
Co-Director de Tesis

Dedicatoria

“Nothing in biology makes sense except in the light of evolution”

Theodosius Dobzhansky

Agradecimientos

Al Centro de Investigaciones Biológicas del Noroeste, S.C., por aceptarme en el programa de estudios de Posgrado y por el apoyo académico durante el transcurso de estos casi tres años; también por la beca adicional otorgada para finalizar la tesis.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada de nivel maestría (CVU/Becario: 610080).

A mis directores, el Dr. Eduardo F. Balart Páez y el Dr. Fausto Valenzuela Quiñonez, por asesorarme, guiarme y por todo su apoyo incondicional, enseñanzas, confianza y sobre todo paciencia en la realización de éste, mi proyecto de tesis.

A los miembros de mi comité, Dr. Raúl Martínez Rincón y Dr. Adrián Munguía Vega por sus valiosos aportes en los campos de genética, modelación oceanográfica y conectividad, análisis de imágenes satelitales y sistemas de información geográfica, y análisis de datos en R.

Al proyecto SEP-CONACYT 2015-01-257524 "PAISAJE GENÓMICO EN EL MEDIO MARINO: PATRONES DE DIVERSIDAD ADAPTATIVA" por el financiamiento recibido para llevar a cabo la tesis.

A Alfredo Sayaz, Horacio Bervera y Salwa El Khattabi por su apoyo en el trabajo de campo.

Al personal del laboratorio de Genética Molecular, a M. C. Delia Rojas Posadas, Dra. Crisalejandra Rivera y Dra. Norma Hernández Saavedra, por todo su apoyo y el espacio de laboratorio brindado.

A M. C. Gabriela Mendoza Carrión, M. C. Victor López Flores y M. C. Karina Alejandra Vázquez por toda su ayuda con la elaboración de librerías y análisis bioinformáticos.

Al personal del departamento de posgrado, Ing. Horacio Sandoval, Lic. Osvelia Ibarra Morales, Lic. Leticia Gonzalez Rubio y Tania Núñez Valdez, por su apoyo y asesoramiento desde un principio.

La modelación de dispersión larvaria con el modelo oceanográfico tridimensional HAMSOM fue realizada por el Dr. Guido Marinone y M.C. Carolina Montaña-Cortés dentro del proyecto institucional CONACyT Fronteras de la Ciencia 292/2016 "ESTUDIO INTEGRATIVO DE LA BIODIVERSIDAD Y LA CONSERVACIÓN DEL GOLFO DE CALIFORNIA, BAJO UN ENFOQUE DE PAISAJE GENÉTICO MARINO Y CONECTIVIDAD", a cargo del Dr. Héctor Reyes-Bonilla.

Al excelente equipo de trabajo que estuvo allí en el laboratorio y eternas horas de computadora, Jorge Mares, Paulina Mejía Ruiz y Paulo Mora.

A mis amigos de aquí y allá, a mi familia y a mis perros, por ser mi sostén.

Contenido

Resumen	i
Summary	ii
Dedicatoria	iii
Agradecimientos	iv
Lista de figuras	viii
Lista de tablas	xi
Abreviaturas	xii
1. INTRODUCCIÓN	1
2. ANTECEDENTES	3
2.1 Genética poblacional y fuerzas evolutivas	3
2.2 Genómica del paisaje	6
2.3 Paisaje marino genético del Golfo de California	9
2.4 Biología de <i>Lythrypnus dalli</i>	13
3. JUSTIFICACIÓN	16
4. HIPÓTESIS	18
4.1 Preguntas científicas	18
4.2 Hipótesis	18
5. OBJETIVOS	19
5.1 Objetivo general	19
5.2 Objetivos particulares	19
6. MATERIAL Y MÉTODOS	20
6.1 Área de estudio y diseño de muestreo	20
6.2 Elaboración de librerías genómicas y secuenciación	22
6.2.1 Extracción de ADN	22
6.2.2 Librerías genómicas	23
6.3 Demultiplexado de librerías y obtención de SNPs	25
6.3.1 Optimización de parametrización de STACKS	26
6.3.2 Filtrado de individuos y secuencias	29
6.4 Identificación de candidatos a selección y SNPs neutrales	30
6.4.1 Detección de outliers	30
6.4.2 Obtención de SNPs neutrales	31
6.5 Diversidad genética y parentesco	33

6.6 Estructura genética poblacional neutral y adaptativa	33
6.7 Aislamiento por distancia	36
6.8 Aislamiento por ambiente	40
7. RESULTADOS	43
7.1 Muestreo	43
7.2 Elaboración de librerías genómicas y secuenciación.....	44
7.3 Demultiplexado de librerías y obtención de SNPs	46
7.3.1 Parametrización de STACKS	48
7.3.2 Filtrado de individuos y SNPs	52
7.4 Identificación de candidatos a selección y SNPs neutrales	53
7.4.1 Detección de outliers	53
7.4.2 Obtención de SNPs neutrales	55
7.5 Diversidad genética y parentesco	56
7.6 Estructura genómica poblacional neutral y adaptativa.....	57
7.7 Aislamiento por distancia – SNPs neutrales	68
7.8 Aislamiento por ambiente	72
8. DISCUSIÓN	81
8.1 Muestreo y librerías genómicas	82
8.2 Estructura genómica neutral	83
8.3 Aislamiento por distancia	85
8.4 Detección de <i>outliers</i> y aislamiento por ambiente.....	89
9. CONCLUSIONES	92
10. LITERATURA CITADA	93

Lista de figuras

Figura 1. Regiones geográficas del GC (A) de acuerdo con la comunidad de peces de plataforma continental descrita por Walker (1960) y (B) de acuerdo a comunidades de peces arrecifales (Thomson <i>et al.</i> , 2000). Imagen modificada de Petatán Ramírez (2015).....	11
Figura 2. <i>Lythrypnus dalli</i> (Imagen tomada de Robertson y Allen, 2015).....	13
Figura 3. Mapa de localidades de muestreo de <i>Lythrypnus dalli</i> . En orden latitudinal, BA: Bahía de los Ángeles, PSF: Puerto San Francisquito, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.	21
Figura 4. Imágenes de colecta de <i>Lythrypnus dalli</i> mediante el uso de una bolsa de 1 m de diámetro (A) y redes comunes de acuario (B).....	22
Figura 5. Representación esquemática de filtrado y detección de SNPs <i>neutrales</i> , <i>outliers</i> y <i>asociados a ambiente</i> ; <i>min_maf</i> : frecuencia del alelo menor, NAs: datos faltantes, LD: desequilibrio de ligamiento, HWE: equilibrio de Hardy-Weinberg, RDA: análisis de redundancia (ver sección 6.8).....	32
Figura 6. 59 sitios de liberación de partículas del modelado oceanográfico (puntos rojos) y sus respectivos polígonos (imagen provista por el Dr. Adrián Munguía Vega).....	39
Figura 7. Temperatura superficial del mar (SST) correspondiente a Julio de 2017 en el GC. Imagen generada a partir de la información obtenida del satélite AQUA MODIS.	41
Figura 8. Extracción y digestión de ADN para la elaboración de librerías genómicas. (A) Ejemplo de verificación de pureza y homogeneidad de diluciones de ADN de <i>Lythrypnus dalli</i> ; (B) Ejemplo de verificación de doble digestión y ligación de adaptadores. k: 1000 pb; E: marcador de peso molecular 1 kb Plus (Thermo Fisher Scientific).	45
Figura 9. Ejemplo de verificación de selección de fragmentos y amplificación de ADN vía PCR de los pools de <i>Lythrypnus dalli</i> (k: 1000 pb; E: marcador de peso molecular 1 kb Plus (Thermo Fisher Scientific)).	46
Figura 10. Diagrama de caja de calidad de secuencias crudas (<i>phred score</i> , eje y). Lecturas forward y reverse del pool Ld2. Los reportes son muy similares para los tres pools restantes. La línea azul representa las medianas y los colores verde, amarillo y rojo indican la calidad promedio buena (<i>phred</i> >30), razonable (<i>phred</i> >20) o mala (<i>phred</i> <20), respectivamente.	47
Figura 11. Número de RAD <i>loci</i> para <i>Lythrypnus dalli</i> (A) ensamblados y número de SNPs (B) obtenidos <i>de novo</i> iterando valores de <i>M</i> y <i>n</i> en STACKS.....	49
Figura 12. Promedio de profundidades de lectura por individuo (X_{ind}) de <i>Lythrypnus dalli</i> obtenidas iterando valores de <i>M</i> y <i>n</i> en STACKS.	50
Figura 13. Número de RAD <i>loci</i> de <i>Lythrypnus dalli</i> (A) y número de SNPs (B) obtenidos luego del filtrado de los genotipos obtenidos para cada combinación de <i>M</i> y <i>n</i>	51

- Figura 14.** Resultados de análisis de detección de *outliers* de *Lythrypnus dalli* a partir de 1711 SNPs del grupo 1 (en color rojo). (A) Prueba de Bayescan; (B) Prueba Arlequin. 55
- Figura 15.** Parentesco promedio (r , en %) de *Lythrypnus dalli* entre sitios muestreados para un subconjunto de 721 SNPs neutrales y polimórficos dentro de todas las localidades. Las barras de error indican los intervalos de confianza al 95% para los valores esperados y las líneas continuas indican los intervalos de confianza al 95% esperado bajo la hipótesis nula de apareamiento aleatorio. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo..... 57
- Figura 16.** Estimadores de estructura genética para *Lythrypnus dalli*. Índices de Fijación (A), diferenciación alélica (B), y distancia genética condicional (C) obtenido para los SNPs neutrales y *outliers*. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo. 59
- Figura 17.** Estructura genómica neutral mediante distancia genética condicional de *L. dalli*. (A) En el espacio geográfico, y (B) visualizado mediante el algoritmo de Fruchterman Reingold, que busca el equilibrio entre las fuerzas de atracción y repulsión entre nodos. El tamaño de nodos (localidades) representa la variación de la diversidad genética dentro de cada sitio y el ancho de los vínculos entre sitios es proporcional a la covarianza genética significativa. Bahía de los Ángeles (1), Isla San Marcos (2), San Cosme y Agua Verde (3), Isla El Pardito (4), Isla San Francisquito (5) y Punta Diablo (6)..... 61
- Figura 18.** dAPCs por localidad de *Lythrypnus dalli*. (A) Para 1,261 marcadores neutrales y (B) 13 *outliers*. PCs: componentes principales. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo..... 63
- Figura 19.** dAPCs de 1,261 SNPs neutrales de *Lythrypnus dalli*. (A) Por localidad y (B) por agrupamiento de acuerdo a *find_clusters*. Densidad de asignación de individuos a sus grupos sobre el primer eje discriminante de acuerdo a la retención de 45 componentes principales en ambos casos. Bahía de los Ángeles (1), Isla San Marcos (2), San Cosme y Agua Verde (3), Isla El Pardito (4), Isla San Francisquito (5) y Punta Diablo (6)..... 65
- Figura 20.** Número más probable de poblaciones de *Lythrypnus dalli* de acuerdo al método de Evanno. Para (A) 1,261 SNPs neutrales y (B) 13 *outliers*. Gráfica de probabilidad de pertenencia de individuos a las poblaciones en base a (C) los SNPs neutrales y (D) *outliers*. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo. 66
- Figura 21.** Prueba de Mantel ($r = 0.497$, $p=0.042$) entre el índice de diferenciación alélica (D_{est}) y distancia oceanográfica (Ha_9) estimada para *Lythrypnus dalli* a partir de 1,261 SNPs neutrales. La línea negra representa la asociación de Mantel, y la línea roja muestra la curva suavizada de variación de los datos..... 69
- Figura 22.** Hipótesis oceanográfica Ha_9 de dispersión larvaria. En versión fuente-destino (A), el color de las celdas indica la probabilidad de movimiento desde cada

sitio de liberación (fuente) hacia su destino, ambos en orden latitudinal. Retención local en la diagonal y las regiones del GC están marcadas en negrita y señaladas en ambos ejes (N: norte, C: centro, S: sur). (B) Distancias oceanográficas de los sitios de muestreo estandarizadas mediante la aplicación de la teoría de *population graphs*..... 70

Figura 23. Relación entre variables genéticas e hipótesis de conectividad Ha9. (A) Regresión lineal ($r^2_{aj} = 0.954$, $p=0.000$) de *Betweenness centrality* (BC) en base a 1,261 SNPs neutrales de *Lythrypnus dalli* contra el número de sitios hacia los que cada localidad exporta (n.exporta) derivado de Ha9. (B) Regresión lineal ($r^2_{aj} = -0.067$, $p=0.454$) entre parentesco promedio (proporción) en base a 721 SNPs neutrales y polimórficos dentro de todas las localidades y porcentaje de retención local derivado de Ha9..... 72

Figura 24. Frecuencias alélicas del único SNP identificado para *L. dalli* por Bayescan (A) y de uno de los alelos de cada uno de los 13 SNPs identificados por Arlequin (B). BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo..... 73

Figura 25. Variables climatológicas derivadas de imágenes del satélite aqua MODIS. (A) Temperatura superficial del mar y (B) clorofila-a. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo..... 75

Figura 26. Análisis de componentes principales de indicadores de variación de temperatura superficial (sst) y Clorofila a (Chla). BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo. 77

Figura 27. Gráfica triplot del modelo $pred_1$ con escalamiento de 1,711 SNPs (grupo 1, color gris) por la raíz cuadrada de los eigenvalores (*scaling* 1). Los predictores se muestran en azul y las localidades en colores. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo..... 79

Figura 28. Gráfica triplot del modelo $pred_2$ con escalamiento de 1,711 SNPs (grupo 1, color gris) por la raíz cuadrada de los eigenvalores (*scaling* 1). Los predictores se muestran en azul y las localidades en colores. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo..... 80

Lista de tablas

Tabla I. Parámetros involucrados en la obtención de genotipos con el programa STACKS V2.0. Los parámetros M , n y p fueron seleccionados y evaluados en cuanto a su efecto en la obtención de genotipos.	28
Tabla II. Hipótesis de conectividad oceanográfica de <i>Lythrypnus dalli</i> de acuerdo a su temporada reproductiva y duración de larva pelágica (PLD) en semanas.	38
Tabla III. Resumen del esfuerzo (buceos y cuadrantes) de muestreo en número (n .°); profundidad promedio (z); individuos colectados y tallas promedio por localidad (ordenados de norte a sur).....	43
Tabla IV. Filtrado de secuencias crudas en <i>process_radtags</i>	48
Tabla V. Filtrado de individuos de acuerdo a la profundidad de lecturas por individuo y el límite de datos faltantes por individuo (NAs). BA: Bahía de los Ángeles, PSF: Puerto San Francisquito, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.....	52
Tabla VI. Filtrado de grupo 1 de SNPs de <i>Lythrypnus dalli</i> para análisis posteriores. El porcentaje de SNPs retenidos es con respecto al número inicial..	53
Tabla VII. Indicadores de diversidad genética por localidad de <i>L. dalli</i> . Tamaño muestral (n), heterocigosis observada (H_O) y esperada (H_E), y riqueza alélica rarefaccionada (Ar). BA: Bahía de los Ángeles, PSF: Puerto San Francisquito, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.	56
Tabla VIII. Número más probable de grupos (k) de acuerdo al número de componentes principales (PCs) retenidos y porcentaje de varianza retenida.	64
Tabla IX. Pruebas de Mantel de aislamiento por distancia geográfica euclidiana (Dist. Geo.) u oceanográfica (probabilidad de conexión larvaria, Ha_i) en relación a la distancia genética neutral de <i>Lythrypnus dalli</i> . Grado de asociación (Mantel r), valor de p , temporada de movimiento larval y duración (PLD) de las hipótesis. ...	68
Tabla X. Características topológicas de Ha_9 por localidad; retención local, número de sitios (de 59 totales) hacia los que se exporta (n .exporta), número de sitios desde los cuales se importa (n .importa) y balance entre ellos (Rol.Fue.Sum.). BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.....	71
Tabla XI. Variables resumen de climatologías históricas utilizadas en la detección de <i>outliers</i> mediante análisis de redundancia, y en las pruebas de hipótesis de aislamiento por ambiente. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.....	76
Tabla XII. Aislamiento por ambiente mediante pruebas de Mantel en base a 13 SNPs <i>outliers</i> de <i>Lythrypnus dalli</i>	78

Abreviaturas

ADN	Ácido desoxirribonucleico.
Ar	Riqueza alélica.
BA	Islas de Bahía de los Ángeles.
cG	Distancia genética condicional (Dyer y Nason, 2004).
D_{est}	Índice de diferenciación alélica (Jost, 2008).
dAPC	Análisis discriminante de componentes principales.
F_{st}	Índice de fijación (Wright, 1943).
GC	Golfo de California.
G_{st}	Índice de fijación (Nei, 1987).
H_E	Heterocigosidad esperada.
H_O	Heterocigosidad observada.
HWE	Equilibrio de Hardy-Weinberg.
LD	Desequilibrio de ligamiento.
<i>maf</i>	Frecuencia del alelo menor (minor allele frequency).
N_e	Tamaño efectivo poblacional.
PDI	Punta Diablo.
PSF	Puerto San Francisquito.
RDA	Análisis de redundancia.
SCAV	San Cosme y Agua Verde.
SEISF	Isla San Francisquito.
SEPA	Isla El Pardito.
SM	Isla San Marcos.
X_{ind}	Profundidad de lecturas por individuo.
X_{locus}	Profundidad de lecturas promedio por locus.

1. INTRODUCCIÓN

La genómica del paisaje marino surge como una ramificación de la genética de poblaciones, y se enfoca en la influencia de la ecología, oceanografía y geografía sobre los patrones de diversidad neutral y adaptativa encontrados en poblaciones marinas (Riginos *et al.*, 2016). El medio marino se caracteriza por la ausencia de barreras obvias al flujo génico y las especies marinas han sido históricamente consideradas espacialmente homogéneas, es decir, unidades conectadas mediante movimiento larval, con apareamiento aleatorio y características demográficas como tasas de reclutamiento, crecimiento y mortalidad iguales (Bernatchez *et al.*, 2017; Ovenden *et al.*, 2015). Sin embargo, el paradigma de la uniformidad marina comienza a cambiar a medida que nuevas herramientas de análisis con poder estadístico sin precedente revelan patrones espaciales previamente no detectados, cuyos determinantes se relacionan en mayor medida con los sistemas de corrientes y gradientes ambientales de temperatura y salinidad (Selkoe *et al.*, 2016).

Las especies marinas ocupan rangos geográficos heterogéneos en términos de ambiente (temperatura, salinidad, tipo de fondo, etc.) que pueden promover efectos *locus* específicos o adaptaciones locales, a pesar del elevado flujo genético (Bernatchez, 2016; Xu *et al.*, 2017; Zhang *et al.*, 2016). Elucidar las dinámicas de conectividad mediante transporte larvario y los mecanismos de adaptación local de las poblaciones a su ambiente es de suma relevancia ya que, además de mejorar el entendimiento básico de la evolución por selección natural, resulta crítico para identificar los factores ambientales responsables de la diferenciación poblacional, para predecir cambios en la distribución futura de las especies, y elaborar estrategias de manejo y conservación adecuadas (Nielsen *et al.*, 2009; Riginos *et al.*, 2016; Selkoe *et al.*, 2016). Las estrategias de manejo y conservación en el medio marino se han incrementado en el mundo en las últimas décadas. Sin embargo, para que los esfuerzos sean efectivos ante los cambios

que se pronostican, es necesario comprender como los procesos regionales moldean la composición y capacidad adaptativa de las comunidades (González-Cabello y Bellwood, 2009).

El Golfo de California (GC) es un mar interior mexicano con elevados niveles de biodiversidad, endemismo, y productividad. Los actuales criterios utilizados para el manejo y conservación raramente incorporan información genética o de estructura poblacional. El objetivo del presente trabajo es determinar como la distancia geográfica u oceanográfica afectan al patrón de diversidad genética neutral, y cómo la heterogeneidad ambiental incide sobre el patrón de diversidad genética adaptativa, utilizando como modelo al gobio criptobéntico *Lythrypnus dalli*. Su cambio generacional anual, movilidad limitada, abundancia y amplia distribución dentro del GC lo convierten en un modelo ideal para el estudio de la influencia del paisaje en el genoma.

2. ANTECEDENTES

2.1 Genética poblacional y fuerzas evolutivas

La genética poblacional estudia la variación espaciotemporal de las frecuencias alélicas a distintos niveles de organización. Dicha variación resulta de la influencia de las fuerzas evolutivas: mutación, deriva génica, flujo génico y selección natural (Hedgecock *et al.*, 2003; Lowe *et al.*, 2017). Las mutaciones que ocurren durante la meiosis incorporan variabilidad en las poblaciones a escalas temporales relativamente extensas (en el orden de 10^{-3} a 10^{-6} por generación, dependiendo del marcador molecular y de la especie en cuestión). Por otro lado, la deriva génica es el cambio estocástico de frecuencias alélicas que resulta de la combinación de gametos durante la reproducción sexual de poblaciones finitas. Sus efectos son apreciables a escalas temporales cortas (inter-generacional) y es particularmente influyente en poblaciones reducidas. El flujo génico también aporta variabilidad a las poblaciones. Este se da cuando existen migrantes entre poblaciones que contribuyen con sus gametos a la población receptora y tiene un efecto homogeneizante sobre las frecuencias alélicas. La selección natural, a diferencia de las otras tres fuerzas que poseen un efecto global, actúa sobre aquellas regiones del genoma que confieren valor adaptativo o detrimental a los individuos (Black IV *et al.*, 2001).

El concepto de población es un tema central en biología. Waples y Gaggiotti (2006) revisaron numerosas definiciones y destacaron dos grandes grupos: el paradigma ecológico, que define una población como un grupo de individuos de la misma especie que co-ocurren e interactúan en un espacio y tiempo, y el paradigma evolutivo, que considera un grupo de individuos de la misma especie que vive en proximidad suficiente para que se dé el apareamiento. La mayoría de las especies se agrupan en poblaciones, y comúnmente presentan cierto grado de diferenciación genética dado por el apareamiento no aleatorio de individuos a través del rango de distribución. En la realidad es posible encontrar un continuo de

niveles de diferenciación genética, desde completo aislamiento, hasta completa homogeneidad genética y demográfica, lo que dificulta el establecimiento de un criterio objetivo y cuantitativo de población (Waples y Gaggiotti, 2006).

El reclutamiento en general es cualquier adición de individuos a una población, ya sea por auto-reclutamiento o conectividad. El auto-reclutamiento se entiende como la proporción de reclutas que se establecen en su población natal. La conectividad se refiere al movimiento de individuos entre poblaciones mediante dispersión o migración (Carr y Syms, 2006). Existe una distinción entre conceptos de conectividad: a nivel demográfico se evalúa la influencia de la migración sobre las tasas de crecimiento entre poblaciones de una especie, mientras que la conectividad genética es la influencia del flujo génico resultante de la migración efectiva (la cual incluye supervivencia de reclutas y aporte al acervo genético local) sobre las frecuencias alélicas y diversidad genética dentro de las poblaciones locales (Lowe y Allendorf, 2010).

Los patrones de diversidad o estructura genética son consecuencia de la interacción entre la historia evolutiva de las especies, factores demográficos, geográficos y ambientales (Selkoe *et al.*, 2016). Los patrones de estructura genética más comunes en la naturaleza se han clasificado de acuerdo con el predominio de las fuerzas evolutivas que los moldean.

El paradigma de las especies marinas es la *panmixia*, en donde no hay aislamiento reproductivo entre individuos de una especie y las frecuencias alélicas son homogéneas en el espacio y tiempo. En este caso las características demográficas, como tasas de crecimiento y mortalidad poblacional también son homogéneas en el espacio y tiempo.

Otro patrón espacial común ocurre cuando disminuye el flujo génico con el aumento de la distancia geográfica, un proceso no adaptativo que depende de las limitaciones a la dispersión y se conoce como *aislamiento por distancia* (Wright, 1943). Por otro lado, la variación en el flujo génico en función de factores

selectivos como la heterogeneidad ambiental a lo largo del rango de distribución de las especies puede generar patrones de diversidad adaptativa conocidos como *aislamiento por ambiente* (Wang y Bradburd, 2014). Los aislamientos por distancia y por ambiente no son mutuamente exclusivos, y en caso de correlación entre la distancia geográfica y heterogeneidad ambiental, se pueden producir patrones genéticos similares.

Un patrón de estructura genética intermedio entre panmixia y aislamiento es la *metapoblación*, un sistema en el que las sub-poblaciones o poblaciones locales habitan parches discretos, la dispersión entre parches no es tan pobre como para negar la conectividad demográfica, ni tan abundante como para poseer características genéticas y demográficas indiferenciables, incluyendo cierto grado de asincronía en el movimiento de individuos, en donde ciertos sitios actúan como fuentes y otros como sumideros o receptores de migrantes (Sale *et al.*, 2006). La conectividad es raramente simétrica en el medio marino y los flujos de larvas casi siempre serán más fuertes en una dirección, en donde los modelos fuente-sumidero describen el contexto en el que sub-poblaciones fuente exportan larvas o adultos hacia sumideros de menor productividad, cuya persistencia depende de este flujo de migrantes (Marine Stewardship Council, 2014).

Por último, el *parcheo genético caótico* es un patrón comúnmente relacionado a la varianza en el éxito reproductivo y se relaciona con la presencia de parches localizados con distinto grado de reclutamiento en el ambiente marino, en donde existe una gran variación espacio-temporal en las condiciones ambientales (Hedgecock y Pudovkin, 2011). El parcheo genético caótico también ha sido relacionado a las condiciones ambientales locales (Selkoe *et al.*, 2010).

Los estudios de estructuración genética neutral enfocados al balance de mutación, migración y deriva génica han sido ampliamente publicados (Hedgecock *et al.*, 2003; Selkoe *et al.*, 2016; Waples, 1987). Desde otra perspectiva, muchos estudios han demostrado casos de adaptación en plantas y animales como

resultado de presiones selectivas naturales y antropogénicas (Ficetola *et al.*, 2007; Jorde *et al.*, 2015; Murphy *et al.*, 2010). Estos estudios han resultado de gran utilidad para elucidar la conectividad entre poblaciones y su capacidad adaptativa, para el desarrollo de estrategias de manejo y conservación. Sin embargo, en el medio marino décadas de estudios moleculares han tenido un éxito limitado en la inferencia de conectividad genética, debido a la incapacidad de los marcadores moleculares clásicos de detectar niveles sutiles de diferenciación genética espacial presentes en muchas especies abundantes, con alta fecundidad y capacidad dispersiva (Riginos *et al.*, 2016; Selkoe *et al.*, 2016).

2.2 Genómica del paisaje

El desarrollo de técnicas de secuenciación masiva y el empleo de librerías de baja representación han permitido la generación e identificación de miles a decenas de miles de marcadores moleculares tipo polimorfismos de nucleótido simple (SNPs, por sus siglas en inglés) para el desarrollo de estudios genómico-poblacionales. La genómica poblacional posibilita el estudio en simultáneo de factores con efectos *locus*-específico, mismos que permiten identificar genes posiblemente sujetos a selección, y factores de efecto genómico global que nos informan acerca de la demografía e historia evolutiva de las especies (Allendorf *et al.*, 2010; Ferchaud y Hansen, 2016; Luikart *et al.*, 2003; Nielsen *et al.*, 2009). Las diferencias entre la estructura genética neutral y adaptativa permiten identificar distintos regímenes de selección (balanceadora, direccional, divergente) a partir de la diferenciación global resultante de los procesos neutrales. De esta manera, los *loci* bajo selección natural presentarán un patrón de variación entre poblaciones que difiere de lo esperado bajo neutralidad, por lo que se los suele llamar *outliers* (atípicos en inglés). Más aun, los estudios a nivel genómico en combinación con variables del paisaje permiten poner a prueba hipótesis acerca de los factores determinantes de la estructura genética, para una mejor comprensión de los procesos neutrales y adaptativos que delinear los distintos

linajes evolutivos (Gagnaire *et al.*, 2015). A su vez, proveen un medio para detectar regiones del genoma sujetas a selección sin necesidad de formular hipótesis *a priori* acerca de la ventaja (o desventaja) selectiva asociada (Reid *et al.*, 2016; Xu *et al.*, 2017).

La genómica del paisaje marino estudia la influencia de las variables ambientales sobre los patrones de diversidad, en el marco genético dado por la historia evolutiva y características demográficas de las especies que habitan el mar. Allí, los grandes tamaños poblacionales efectivos (N_e) combinados con elevado flujo génico, proveen un escenario en el que la influencia de la deriva génica en la diferenciación poblacional es relativamente baja, y la acción de la selección natural como motor de adaptación local podría ser detectado (Riginos *et al.*, 2016). El estudio de adaptaciones locales en especies marinas ha experimentado un auge reciente, y muchos trabajos se han enfocado particularmente en encontrar marcadores moleculares bajo selección natural, con la expectativa de contribuir al entendimiento de los mecanismos y genes involucrados en dichos procesos, y acerca de los factores naturales y antropogénicos causales (Benestan *et al.*, 2016; Bernatchez, 2016; Ferchaud y Hansen, 2016; Reid *et al.*, 2016).

Habiéndose cumplido 10 años desde el surgimiento de la genética del paisaje marino, Selkoe *et al.* (2016) realizaron una revisión bibliográfica y retuvieron 100 artículos académicos, principalmente basados en microsatélites, pero con una tendencia creciente en el uso de SNPs, entre los cuales evaluaron la cobertura taxonómica y geográfica generada. Según los autores, la mayoría de los estudios fueron realizados en cordados, principalmente peces (48%) e invertebrados (38%). En cuanto a geografía, se han publicado más trabajos en especies de hábitat templado (68%), y los trabajos de aguas tropicales (26%) se han concentrado en arrecifes coralinos. Casi todos los trabajos evaluaron la distancia geográfica como predictor, y aproximadamente la mitad de ellos reportó un aislamiento por distancia significativo. Más aun, de los trabajos que no reportaron aislamiento por distancia, la mitad encontró la influencia de algún otro factor del paisaje. Los

factores más comunes puestos a prueba en estos estudios fueron distintas variantes de temperatura y corrientes oceánicas, de los cuales un 43% y 31% respectivamente resultaron significativos (Selkoe *et al.*, 2016).

Dado que la disciplina es relativamente joven, existe una variedad de desafíos involucrados en la obtención de datos genómicos robustos, en su análisis en conjunto con datos del ambiente para la descripción de dinámicas meta-poblacionales, y en la detección de marcadores relacionados a adaptaciones locales (Catchen *et al.*, 2017; Lotterhos y Whitlock, 2015; Lowry *et al.*, 2016; Narum y Hess, 2011; Rellstab *et al.*, 2015). Los primeros estudios publicados se han concentrado en especies de interés comercial o especies cuyo genoma ha sido secuenciado (Gonçalves Da Silva *et al.*, 2015; Guo *et al.*, 2016; Lal *et al.*, 2016a; Therkildsen *et al.*, 2013), aunque trabajos basados en especies no-modelo progresan con el desarrollo de herramientas de secuenciación masiva y análisis bioinformáticos (Flanagan *et al.*, 2016; Hess *et al.*, 2013; Manthey y Moyle, 2015; Reid *et al.*, 2016; Saenz-Agudelo *et al.*, 2015).

Recientemente, Stanley *et al.* (2018) reportaron un quiebre filogeográfico multiespecie asociado a un gradiente climático, cuyo principal determinante es la temperatura mínima estacional. Este trabajo combina distintos marcadores moleculares (microsatélites y SNPs) y especies de características demográficas diversas y linajes evolutivos independientes, como el bacalao del Atlántico (*Gadus morhua*), la langosta americana (*Homarus americanus*), la vieira atlántica (*Placopecten magellanicus*), el camarón del norte (*Pandalus borealis*) y el cangrejo invasivo europeo (*Carcinus maenas*). Diversos análisis de estructura genética (agrupamiento bayesiano, análisis discriminante de componentes principales y análisis espacial de componentes principales) identifican el quiebre filogeográfico en una latitud cercana a un gradiente ambiental agudo, cuya influencia proponen explica el patrón observado en mayor medida que los patrones de conectividad y retención local dados por el sistema de corrientes en el Noroeste del Océano Atlántico, aunque no descartan su influencia, ni la posible influencia de eventos

vicariantes pasados cuya huella siguiera presente en la estructura genética contemporánea. Dado que los gradientes de temperatura mínima estacional tuvieron el mejor ajuste con la estructura genética observada, los autores proponen que el proceso de mortalidad diferencial post-reclutamiento asociado a las condiciones locales podría ser el mecanismo determinante (Stanley *et al.*, 2018).

En otro ejemplo del alcance de la genómica del paisaje, Matz *et al.* (2018) combinaron el análisis de 11,500 SNPs del coral *Acropora millepora* con modelos biofísicos y evolutivos para estimar el potencial adaptativo en la Gran Barrera de Coral australiana. Los autores desarrollaron un modelo de adaptación poligénica en un sistema de múltiples poblaciones interconectadas, y encontraron evidencias de adaptaciones locales relacionadas a la temperatura máxima de verano, incluso en presencia de elevadas tasas de migración, que a su vez concuerdan con un flujo predominantemente asimétrico en dirección al sur. Este flujo génico sería responsable de transmitir alelos tolerantes a altas temperaturas, favoreciendo la tolerancia al calentamiento global de la especie en la región durante los próximos 100 a 200 años. El modelo predice también un incremento en la susceptibilidad de la especie a fluctuaciones aleatorias de temperatura, como los eventos climáticos de El Niño u olas de calor, lo que concuerda con el incremento reciente de eventos catastróficos de blanqueamiento de corales.

2.3 Paisaje marino genético del Golfo de California

El GC (Fig. 1), también conocido como Mar de Cortez, posee una amplia heterogeneidad de paisaje que influye en la composición y distribución de las comunidades de peces e invertebrados. Muchos autores concuerdan en la división en 3 grandes biorregiones: Golfo norte, central y sur (Petatán Ramírez, 2015; Ulate *et al.*, 2016; Walker, 1960; Fig. 1). El Golfo norte presenta menor profundidad, menor salinidad y oscilaciones mareales extremas de hasta 10 m de

amplitud, además de tener la mayor variabilidad de temperatura superficial y menor cobertura de arrecifes rocosos que el resto de la costa peninsular (Thomson *et al.*, 2000). Tanto el Golfo norte como el central presentan giros ciclónicos en verano y anticiclónicos en invierno que generan retención de partículas, mientras que en la costa peninsular del Golfo sur la retención de partículas se da porque predominan corrientes estacionales débiles. En ambos casos existe potencial de auto-reclutamiento. En la costa continental, en cambio, existe gran potencial de dispersión dado por las fuertes corrientes estacionales (Lavin y Marinone, 2003; Marinone, 2012). El patrón de circulación de corrientes podría restringir el intercambio de migrantes entre las distintas regiones biogeográficas, aunque los modelos de conectividad oceanográfica predicen niveles sustanciales de conectividad y muchas especies son compartidas. Entre el Golfo norte y central no se conocen especies hermanas de peces, y la composición y distribución de las comunidades sugiere que los factores que moldean los rangos de distribución contemporáneos son ambientales (Riginos, 2005).

El Mar de Cortez tiene descritas 361 especies de peces arrecifales que habitan ambientes tanto coralinos como rocosos (Robertson y Allen, 2015). Los peces criptobénticos son un subgrupo de peces arrecifales muy abundante y diverso que generalmente posee una fase dispersiva larval y movilidad limitada luego de su asentamiento, por lo que pueden ser sensibles a cambios en las condiciones ambientales. La fauna criptobéntica del GC es especialmente vulnerable a disturbios debido a la alta especificidad en el uso de hábitat de las especies dominantes y su relativamente baja biodiversidad en comparación con otros sistemas arrecifales del Pacífico (González-Cabello y Bellwood, 2009).

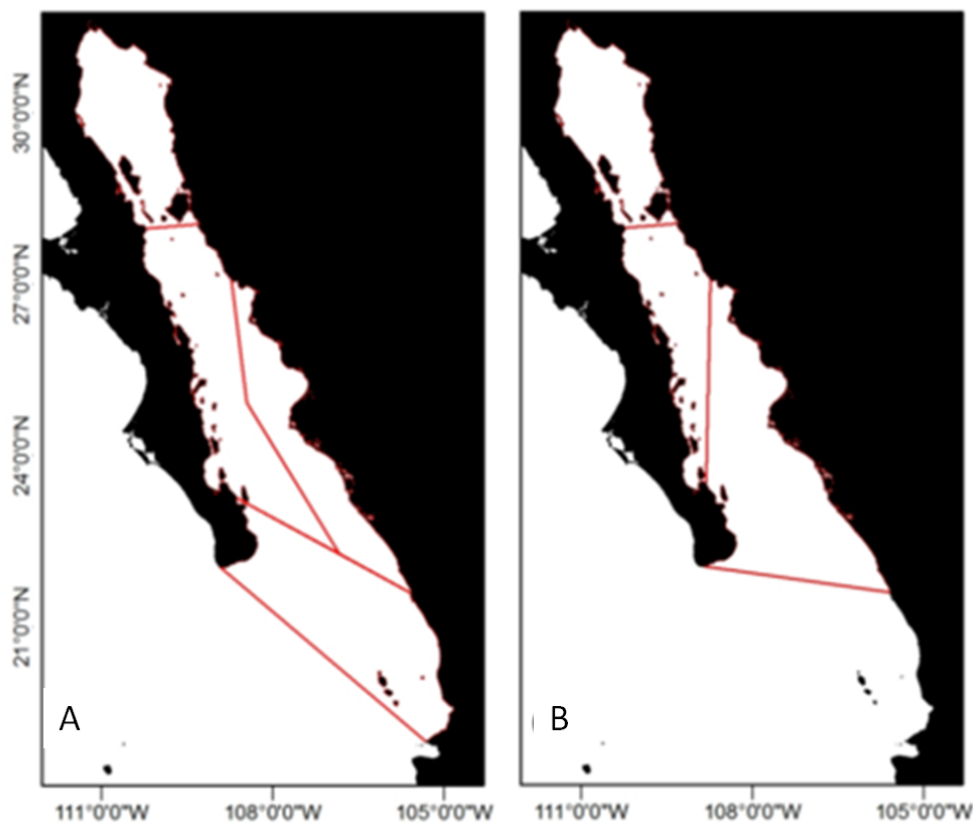


Figura 1. Regiones geográficas del GC (A) de acuerdo con la comunidad de peces de plataforma continental descrita por Walker (1960) y (B) de acuerdo a comunidades de peces arrecifales (Thomson *et al.*, 2000). Imagen modificada de Petatán Ramírez (2015).

Se han publicado numerosos trabajos que buscan la relación de la estructura genética con eventos históricos vicariantes (Bernardi *et al.*, 2003; Hurtado *et al.*, 2010; Lin *et al.*, 2009; Riginos, 2005). Por ejemplo, Bernardi *et al.* (2003) reportan dos patrones bien diferenciados para 12 especies disyuntas, es decir, ausentes o raras en el extremo sur del GC, pero presentes en ambas costas de la península. De las 12 especies, 8 presentan una clara diferenciación y señales de bajo flujo génico, mientras que las 4 restantes indican panmixia. Al interior del Golfo, Riginos (2005) encuentra una estructura espacial en 5 especies de peces costeros, que además coincide con patrones reportados para pequeños mamíferos y reptiles, por lo que posiblemente estén relacionadas a eventos históricos vicariantes.

Otra serie de trabajos se han enfocado en la relación de la estructura genética a eventos climáticos, procesos oceanográficos y dispersión larvaria (Beldade *et al.*, 2014; Hurtado *et al.*, 2010; Riginos y Nachman, 2001; Soria *et al.*, 2012). Por ejemplo, en el trabajo de Riginos y Victor (2001), se observa un claro aumento en la diferenciación de variantes haplotípicas ante la disminución de la dispersión larval. La mayoría de las especies marinas se dispersan durante un estadio larval planctónico, que presenta un amplio rango de variación temporal. La trayectoria de las larvas y los patrones resultantes de dispersión a largo plazo son difíciles de predecir, ya que la dispersión ocurre sobre un ambiente heterogéneo difícil para el humano de cuantificar, por lo que se sabe poco acerca del comportamiento e interacciones interespecíficas que ocurren durante esta etapa (Riginos *et al.*, 2016).

Varios estudios genéticos y ecológicos han comenzado a incorporar una adaptación del modelo tridimensional baroclínico de conectividad oceanográfica HAMSOM (*Hamburg Shelf Ocean Model*, Backhaus, 1985) para el sistema de corrientes del GC (Lavin y Marinone, 2003; Marinone, 2012). Este modelo se ha utilizado para evaluar la correlación entre estructura genómica y dispersión larval (Cisneros-Mata *et al.*, 2019; Munguia-Vega *et al.*, 2014, 2018a; Soria *et al.*, 2012). Los autores han encontrado una relación entre patrones de diversidad y estructura genética de diversas especies de peces e invertebrados con los flujos de corrientes predominantemente asimétricos y estacionales del GC durante la temporada y duración de movimiento larval de cada especie.

Es posible que la circulación oceanográfica y/o la selección natural en contra de individuos migrantes limiten el flujo génico entre las distintas regiones del Golfo. Sin embargo, la contribución relativa de migración y selección en la estructuración poblacional de la gran mayoría de las especies que habitan el GC aún no han podido ser elucidadas.

2.4 Biología de *Lythrypnus dalli*

El gobio bonito, gobio catalina azul o gobio de bandas azules *Lythrypnus dalli* (Gilbert, 1890; Fig. 2) es un pequeño pez criptobéntico de color naranja-rojizo, con una serie de bandas verticales de color azul (2-6 bandas) y una longitud estándar máxima de 50 mm (Behrents, 1983). Su distribución abarca desde las costas de California, en Estados Unidos, hasta el centro de Perú, incluyendo las Islas Mapelo, Cocos y Galápagos (Robertson y Allen, 2015). *L. dalli* habita ambientes rocosos sub-mareales hasta una profundidad de 75m, orientado en refugios horizontales o superiores y su alimentación se basa principalmente de zooplancton presente en la columna de agua (Behrents, 1989). La especie posee un estadio móvil de larva planctónica que eventualmente se asienta en los mismos microhábitats compartidos por juveniles y adultos (Behrents, 1983; Steele, 1997, 1999). El cambio generacional promedio ocurre aproximadamente cada 12 meses (Behrents, 1983; Calderón Parra, 2009), los adultos presentan muy baja movilidad y alcanzan la madurez sexual al mes de asentamiento de los juveniles, con una longitud estándar de 9-11 mm (Behrents, 1983). A su vez, presentan hermafroditismo simultáneo, cambio de sexo bi-direccional (Lorenzi *et al.* 2009; St. Mary, 2000) y cuidado parental de huevos demersales por parte de los machos (Behrents, 1983).



Figura 2. *Lythrypnus dalli* (Imagen tomada de Robertson y Allen, 2015).

Se han publicado trabajos de conectividad y flujo génico para *L. dalli* con alozimas en la costa de California (Waples, 1987), y con alozimas y ADN mitocondrial, comparando el norte del GC y la Costa Californiana de EEUU (Bernardi *et al.*, 2003). Bernardi *et al.* (2003) comparan varias especies de peces (*L. dalli* incluido) y adjudican las diferencias encontradas a procesos históricos climáticos y geológicos, como la separación de la Península de Baja California del continente y posibles migraciones durante el último período glacial, sin descartar la influencia de características demográficas sobre la estructura genética, como por ejemplo la amplitud de la temporada reproductiva y duración de la larva planctónica (PLD por sus siglas en inglés).

La duración de la temporada reproductiva es un factor influyente sobre los patrones espaciales de estructura genética neutral, ya que durante esta etapa ocurrirá el desove y subsecuente migración de las larvas plantónicas con las corrientes marinas hacia los sitios donde serán reclutadas. Calderón Parra (2009) realizó colectas mensuales de *L. dalli* en un mismo sitio ubicado en la Bahía de La Paz (Baja California Sur, México) desde junio de 2005 a noviembre de 2006. Según análisis histológicos, el autor reporta la presencia de hembras maduras y con desoves parciales de marzo a noviembre, y el reclutamiento de juveniles es registrado desde febrero a octubre, con un pico en julio. Algo similar se reporta para la especie en las costas de California (Estados Unidos), en donde se registra un período de ovoposición ligeramente más corto, de abril a septiembre, y el reclutamiento de juveniles comienza a fines de mayo, alcanza su pico en julio y agosto, y termina en enero (Behrents, 1983; Steele, 1997; Wiley, 1976). La duración del estadio de larva planctónica también es un factor muy influyente dada su relación con la capacidad dispersiva, especialmente en especies cuya fase adulta presenta baja o nula movilidad. Según datos de fecundidad y reclutamiento, para *L. dalli* se ha reportado una duración de larva pelágica de 40 días, 100 días o más (Archambeault *et al.*, 2016; Behrents, 1983).

El gobio bonito ha sido también modelo de estudios ecológicos relacionados a la disponibilidad de refugios, competencia por recursos y vulnerabilidad por predación (Hartney y Grorud, 2002; Gregor y Anderson, 2016; Steele, 1999), factores potencialmente influyentes en la capacidad de supervivencia y reproducción de los organismos.

3. JUSTIFICACIÓN

El GC es considerado un sitio de prioridad para la conservación por sus elevados niveles de biodiversidad y endemismo. Los peces arrecifales tienen un rol clave en el ecosistema dado que la mayoría de la biomasa fluye a través de ellos. La fauna críptica del GC puede ser particularmente vulnerable ante disturbios naturales y antropogénicos debido a la elevada dominancia en abundancia de especies especialistas en cuanto a hábitat y también al potencial limitado de redundancia funcional dentro del sistema (González-Cabello y Bellwood, 2009).

Los estudios previos que incluyen a *Lythrypnus dalli* han sido realizados en escalas espaciotemporales relativamente pequeñas o no incluyen análisis genéticos. El presente estudio posee un enfoque diferente, investigando los patrones espaciales a mayores escalas y a través de ambientes heterogéneos, en busca de determinar el papel que tienen la distancia geográfica (euclidiana), la distancia derivada de la circulación oceanográfica y el ambiente sobre los patrones de estructuración neutral y adaptativa del gobio bonito en el GC.

A medida que el paradigma de homogeneidad genética en especies marinas va siendo desplazado por nuevas evidencias, y en vista de los cambios que se pronostican a causa del calentamiento global, elucidar las dinámicas metapoblacionales determinadas por el movimiento larval, y la influencia de las condiciones locales sobre la diversidad y estructura genética resulta crítico para la implementación de estrategias de manejo y conservación adecuadas. La revolución genómica ha traído consigo la posibilidad de poner a prueba la influencia de factores bióticos y abióticos sobre los patrones de diversidad genética, permitiendo inferir la contribución relativa del flujo génico, deriva y selección natural en poblaciones naturales (Hand *et al.*, 2015), ya que permite detectar diferencias sutiles entre poblaciones que otros marcadores moleculares han sido incapaces de detectar y permite identificar procesos selectivos mediante los cuales se generan adaptaciones locales. Las bases moleculares de la

adaptación local de especies marinas no-modelo cuyo genoma no ha sido secuenciado, aún no han sido elucidadas. Comprender la habilidad de adaptación de los organismos a corto plazo, más allá de contribuir al conocimiento general de la evolución en el medio marino, es crítico para la predicción de la respuesta de las especies al cambio global (Nielsen *et al.*, 2009; Riginos *et al.*, 2016; Selkoe *et al.*, 2016).

El presente estudio busca aportar al entendimiento de la contribución relativa de la distancia geográfica, oceanográfica y la heterogeneidad ambiental sobre los patrones de estructura genética de peces criptobénticos arrecifales del GC, usando como modelo al gobio *Lythrypnus dalli*.

4. HIPÓTESIS

4.1 Preguntas científicas

¿Qué tipo de estructura genética neutral y adaptativa presenta *Lythrypnus dalli* en el Golfo de California?

¿Cuál será la contribución de la distancia geográfica, lineal u oceanográfica, sobre la estructura genética neutral de *Lythrypnus dalli* en el Golfo de California?

¿Cuál será la contribución de la heterogeneidad ambiental sobre la estructura genética adaptativa de *Lythrypnus dalli* en el Golfo de California?

4.2 Hipótesis

Dado que *Lythrypnus dalli* se encuentra distribuido a lo largo del Golfo de California y posee muy baja movilidad luego de su reclutamiento, se espera encontrar un patrón de aislamiento por distancia geográfica u oceanográfica en la estructura genética neutral, y un patrón de aislamiento por ambiente derivado de adaptaciones locales generadas mediante procesos selectivos.

5. OBJETIVOS

5.1 Objetivo general

Determinar la influencia de factores geográficos, oceanográficos y ambientales sobre la estructura genética poblacional de *Lythrypnus dalli* en el Golfo de California.

5.2 Objetivos particulares

1. Determinar la estructura genética neutral y adaptativa de *Lythrypnus dalli* en el Golfo de California.
2. Determinar la contribución de la distancia geográfica u oceanográfica sobre la estructura genética neutral de *Lythrypnus dalli* en el Golfo de California.
3. Determinar la contribución de la heterogeneidad ambiental sobre la estructura genética adaptativa de *Lythrypnus dalli* en el Golfo de California.

6. MATERIAL Y MÉTODOS

6.1 Área de estudio y diseño de muestreo

El GC consiste en un área semicerrada de más de 1500 km de largo, entre 100 y 200 km de ancho, y hasta 4 km de profundidad orientada en dirección noroeste-sureste. Se encuentra conectado al Océano Pacífico en su extremo sur, y al Delta del Río Colorado en su extremo norte. El Golfo se creó por la influencia de la falla de San Andrés, por donde se separó la Península de Baja California del resto del continente, durante un período de entre 3.5 a 12 millones de años atrás (Lavin y Marinone, 2003). Por un lado, presenta surgencias en ambas costas y un sistema de corrientes gobernado por un flujo saliente del Golfo hasta los 200 m de profundidad, entrante por debajo de los 200 m, y una serie de giros geostroficados a escala cuenca que alcanzan una profundidad mayor a 1000 metros y consistentemente cambian de dirección dos veces al año (Lavin y Marinone, 2003; Marinone, 2012). Por otro lado, el GC presenta una gran variabilidad espacio-temporal en cuanto a las condiciones ambientales, como temperaturas medias y rangos, salinidad, vientos y productividad, principalmente en el eje noroeste-sureste (Lavin *et al.* 2003).

Durante la temporada verano-otoño de 2017 se realizó la colecta de *Lythrypnus dalli* a lo largo de la costa occidental del GC, desde Bahía de los Ángeles (BA) hasta Punta Diablo (PDI) en la Bahía de la Paz (Fig. 3).

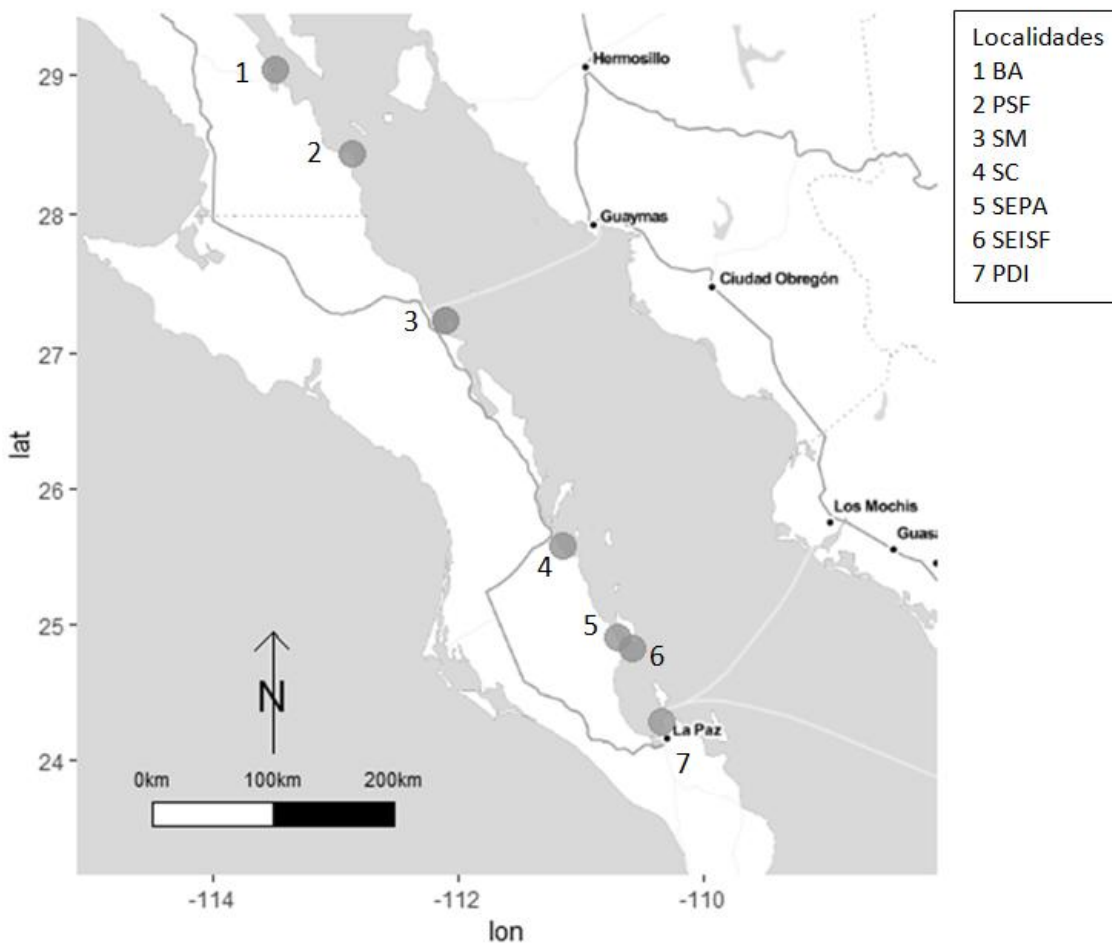


Figura 3. Mapa de localidades de muestreo de *Lythrypnus dalli*. En orden latitudinal, BA: Bahía de los Ángeles, PSF: Puerto San Francisquito, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

Los individuos fueron colectados mediante buceo SCUBA sobre tres tipos de sustrato: arrecife rocoso, pared rocosa y arrecife coralino. Durante cada inmersión se seleccionaron al azar 3 o más cuadrantes de 1x1 m, en donde se colectaron peces utilizando una mezcla de aceite de clavo:etanol (3:1) como anestésico. Para evitar que los peces se escaparan y minimizar el uso del anestésico, se colocó una bolsa de un metro de diámetro por encima del cuadrante seleccionado, para aplicar la mezcla en su interior y posteriormente capturar los organismos con cucharas de red de acuario (Fig. 4). Se registraron las coordenadas geográficas,

profundidad y longitud total los individuos. Éstos fueron preservados de manera íntegra en etanol al 96% para su traslado al laboratorio.



Figura 4. Imágenes de colecta de *Lythrypnus dalli* mediante el uso de una bolsa de 1 m de diámetro (A) y redes comunes de acuario (B).

6.2 Elaboración de librerías genómicas y secuenciación

6.2.1 Extracción de ADN

La extracción de ADN de los organismos se realizó mediante el protocolo de sales y cloroformo:alcohol-Isoamilico (Sambrook *et al.*, 1989). El ADN extraído fue cuantificado mediante Qubit® (Thermo Fisher Scientific), la integridad del ADN se

evaluó mediante un gel de Agarosa-Synergel al 1% y revelado en un fotodocumentador ChemiDoc™ (Bio-Rad). Todas las muestras de ADN fueron llevadas a una concentración de 20 ng/μL.

6.2.2 Librerías genómicas

Las librerías genómicas de baja representación se elaboraron siguiendo el método de secuenciación de doble digestión de ADN asociado a enzimas de restricción (ddRAD, por sus siglas en inglés) descrito por Peterson *et al.* (2012). Este método permite muestrear una fracción del genoma en múltiples individuos y obtener miles de SNPs a menores costos y tiempo que otras técnicas de tipo RAD (Peterson *et al.*, 2012). Todo el proceso se llevó a cabo en microplacas de 96 pozos y los pasos a seguir se presentan a continuación.

6.2.2.1 Doble digestión

Para la digestión se utilizaron 400 ng totales de ADN, el cual fue digerido con las enzimas de restricción EcoR1-HF (5' GAATTC, NEB) y MspI (5' CCGG, NEB). La reacción de digestión se llevó a cabo en un volumen de 30 μL con 400 ng de ADN, 1X Buffer CutSmart® (NEB), 10 U de EcoRI-HF (NEB) y 10 U de MspI (NEB). La reacción fue incubada a temperatura ambiente durante toda la noche. Los productos de la digestión fueron purificados con 1.5X de AMPure® XP beads (Beckman Coulter) de acuerdo con el protocolo del fabricante y el ADN recuperado fue eluido en 30 μL de TE 0.1X.

6.2.2.2 Ligación de adaptadores

Posteriormente, se procedió a la ligación de los adaptadores P1 (*barcode*), de los cuales se utilizaron con 48 *barcodes* diferentes, y P2 (*common*), que contienen la secuencia complementaria al extremo cohesivo de las enzimas de restricción EcoRI y MspI, respectivamente. Cada reacción de ligación se realizó en un volumen de 40 μL con 30 μL de ADN digerido purificado, 1X Buffer T4 Ligasa (NEB), 100 U de T4 ADN Ligasa (NEB), 0.1 μM de adaptador P2 y 0.1 μM de adaptador P1 con su respectivo *barcode*. Cada individuo contó con un *barcode*

específico. La reacción se incubó a temperatura ambiente durante toda la noche y fue posteriormente inactivada mediante la incubación a 65 °C por 10 minutos. El producto ligado fue nuevamente purificado con 1.5X AMPure® XP beads (Beckman Coulter) y eluído en 30 µL de TE 0.1X.

Para verificar la correcta ligación de los adaptadores se realizó una amplificación de los productos ligados mediante PCR. Las reacciones de PCR se llevaron a cabo en un volumen de 12.5 µL, con 1 µL de ADN ligado, 0.8 µM de cada primer Illumina Forward (3' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG), Reverse (3' CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTAAG) y 1X GoTaq® (Promega). Esto se llevó a cabo en un termociclador T100™ (Bio-Rad) iniciando con la desnaturalización del ADN a 94°C por 2 min, seguido por 20 ciclos de desnaturalización a 94°C por 30 s, alineamiento a 60°C por 30 s y extensión a 72°C por 45 s. La temperatura de reposo fue de 10°C. Los productos de PCR fueron visualizados en un gel de agarosa-Synergel al 2% y visualizado en un fotodocumentador ChemiDoc™ (Bio-Rad).

6.2.2.3 Combinación equimolar de muestras

Una vez verificada la ligación, se cuantificó la concentración de ADN ligado en Qubit® (Thermo Fisher Scientific), para poder combinar las muestras en concentraciones equimolares. Cada grupo de muestras combinadas o *pool* consta de 48 individuos, ya que corresponden a los 48 adaptadores P1 únicos. Dependiendo de la masa de ADN disponible, se elaboraron pools de 10 a 30 ng totales por individuo. Cada pool generado de volumen variable fue concentrado con 1.5X AMPure® XP beads (Beckman Coulter) y eluído en 30 µL de TE 0.1X.

6.2.2.4 Selección de fragmentos en Pippin prep

Se llevó a cabo una selección de fragmentos de cada pool mediante Pippin-prep (Sage Science), de 376 pb ± 50 pb de acuerdo con las instrucciones del

fabricante. Se recuperaron 40 μ L del ADN eluido luego de la selección de fragmentos.

6.2.2.5 Enriquecimiento e indexado de pools

A continuación, se realizó una segunda PCR con la finalidad de enriquecer los pools e incorporar los *index* específicos para cada pool, así como los adaptadores complementarios a la celda de flujo del secuenciador Illumina. A partir de cada pool se llevaron a cabo 8 reacciones de PCR. Cada reacción se llevó a cabo en un volumen de 20 μ L con 3 μ L de ADN, 1X Buffer Phusion-HF (NEB), 0.8uM dNTPs, 0.5 μ M de cada primer Illumina Forward y Reverse (de *index* variable), 0.5 U de TAQ Phusion ADN polimerasa (NEB). Las PCR se llevaron a cabo en un termociclador T100TM (Bio-Rad) iniciando con la desnaturalización del ADN a 94°C por 2 min, seguido por 10 ciclos de desnaturalización a 94°C por 30 s, alineamiento a 60°C por 30 s y extensión a 72°C por 45 s. La temperatura de reposo fue de 10°C. Los productos de PCR de cada pool fueron mezclados y concentrados con 1.5X AMPure® XP beads (Beckman Coulther). Cada pool fue visualizado en un gel de agarosa-synergel al 2% para verificar la correcta selección de fragmentos y posteriormente los pools fueron cuantificados en Qubit® (Thermo Fisher Scientific). Finalmente, las dos librerías finales se generaron a partir de la mezcla equimolar de dos pools cada una (4 pools en total). En total fueron incluidos 11 individuos por duplicado con distinta combinación de *barcode* e *index* para estimar el error de genotipificación. Las librerías se enviaron a secuenciar en dos líneas del secuenciador Illumina HiSeqX (Novogene, California) con lecturas pareadas de 150 pb.

6.3 Demultiplexado de librerías y obtención de SNPs

La calidad de las secuencias obtenidas fue verificada inicialmente con el programa Fastqc (Andrews, 2010). Las secuencias de las librerías fueron demultiplexadas, alineadas y filtradas con el programa STACKS V2.0 (Catchen *et al.*, 2013). Este

programa incorpora un modelo estadístico de máxima verosimilitud para la identificación de polimorfismos, genera un catálogo en donde se registran todos los *loci* de la población, y compara cada individuo con el catálogo para determinar que haplotipos están presentes en cada *locus* de cada individuo (Catchen, 2016). Las secuencias crudas fueron demultiplexadas y filtradas con el módulo *process_radtags* (Tabla I). Las lecturas individuales con una probabilidad de asignación menor al 90% (*phred score* < 10, ventana móvil del 15% de la longitud total), con errores en la secuencia del *barcode* (2 o más diferencias) o en los sitios de corte (EcoRI, MspI) fueron descartadas. La calidad de las secuencias obtenidas en los últimos 10 pb de los fragmentos no resultó igual que el resto (ver resultados), por lo que se decidió cortar las lecturas a 140 pb. Dado que no se cuenta con un genoma de referencia de la especie, o cercano a ella, los fragmentos o RAD *loci* fueron ensamblados *de novo* con el *pipeline denovo_map.pl* y posteriormente filtrados con el módulo *populations*.

6.3.1 Optimización de parametrización de STACKS

La descripción de parámetros involucrados en el filtrado y obtención de genotipos y su función se resume en la tabla I. Los parámetros recomendados por Paris *et al.* (2017) para este tipo de estudios son una profundidad de lecturas mínima para crear una pila de fragmentos de 3 ($m \geq 3$), la retención de SNPs presentes en el 80% de los individuos como mínimo ($r \geq 0.8$), y mantener una relación $n = M \pm 1$ en cuanto a los parámetros que controlan el grado de variación permitida dentro y entre pilas de fragmentos. El valor óptimo depende del objetivo del trabajo y del grado de polimorfismo de la especie, entre otros factores (Paris *et al.*, 2017). Específicamente, M controla el número de diferencias permitidas entre los dos alelos de una muestra heterocigota, mientras que n controla el número de diferencias entre pares de alelos de la población (Rochette y Catchen, 2017). Al respecto, se realizó una serie de análisis iterando valores de M y n (M entre 1 y 7, $n = M \pm 1$) para encontrar la combinación que permita obtener suficiente cantidad y calidad de marcadores moleculares. Otro parámetro influyente en la obtención

de genotipos es la frecuencia del alelo menor (*maf*, por sus siglas en inglés), comúnmente fijado a 0.05, aunque valores por encima del error de genotipificación estimado pueden ser utilizados si es conveniente.

Tabla I. Parámetros involucrados en la obtención de genotipos con el programa STACKS V2.0. Los parámetros M , n y p fueron seleccionados y evaluados en cuanto a su efecto en la obtención de genotipos.

Módulo	Parámetro	Descripción
<i>process_radtags</i>	-c	<i>Clean data.</i> Remueve fragmentos con bases no identificadas por el secuenciador.
	-q	Elimina secuencias de calidad menor al umbral deseado. Se usó el umbral de <i>phred score</i> de 10 (<i>default</i>).
	-r	Recupera fragmentos cuyo <i>barcode</i> o sitio de corte posee 1 diferencia como máximo de lo esperado (<i>default</i>).
	-t	Corta o trunca las secuencias para evitar errores de secuenciación comúnmente producidos por el secuenciador hacia el fin de los fragmentos. Se usó $t = 140$ pares de bases.
<i>denovo_map.pl</i>	m	Profundidad mínima (X) para crear una pila. Se usó $m = 4$, ya que se ha probado que valores mayores a 3 son suficientes para obtener datos robustos en distintas especies (Paris <i>et al.</i> 2017).
	M	Número de diferencias permitidas entre fragmentos dentro de cada individuo para formar una pila (posible alelo). De acuerdo con la evaluación de parametrización, se usó $M = 3$.
	n	Número de diferencias permitidas entre alelos de individuos para ser considerados como un RAD <i>locus</i> (conjunto de fragmentos homólogos). De acuerdo con la evaluación de parametrización, se usó $n = 4$.
<i>populations</i>	min_maf	Frecuencia del alelo menor. Proporción mínima de representación del alelo en la población para incluir el SNP que lo contiene. Se utilizó un valor mínimo de 0.03 para la frecuencia del alelo menor.
	-r	Frecuencia mínima de representación de los SNPs en los individuos. Se seleccionaron SNPs presentes en al menos el 80% de los individuos, ya que este parámetro ha sido considerado óptimo en la literatura (Paris <i>et al.</i> 2017).
	-p	Número mínimo de localidades en que los SNPs deben estar presentes. Para identificar candidatos a selección natural se seleccionaron SNPs presentes en 6 de las 7 localidades, mientras que para determinar la estructura genómica neutral se usó el subconjunto de SNPs presentes en todas las localidades.
	write_random_snp	Selecciona un único SNP por RAD <i>locus</i> , de modo de minimizar SNPs en LD.

6.3.2 Filtrado de individuos y secuencias

Con base en la combinación óptima de parámetros y habiendo eliminado del análisis aquellos individuos con una profundidad de lecturas promedio menor a 40X, se utilizó el módulo *populations* para filtrar los SNPs. El efecto del número mínimo de localidades en que cada SNP debe estar presente se seleccionó considerando el compromiso entre los SNPs que se retienen y el grado de datos faltantes por individuo asociado, estimado con VCF tools (Danecek *et al.*, 2011).

Se elaboraron dos grupos de datos con diferente filtrado para probar las potenciales hipótesis de relaciones de distancia y ambiente. Para la detección de candidatos a selección y probar las Hipótesis de aislamiento por ambiente se elaboró una matriz de genotipos donde se retuvieron aquellos SNPs presentes en **-p 6** de las 7 localidades, con la finalidad de obtener un mayor número de SNPs y con ello una incrementar la probabilidad de detectar *loci outliers* (de aquí en adelante a este grupo de datos se hará referencia como “grupo 1”). Un segundo grupo de datos para poner a prueba las hipótesis de aislamiento por distancia se elaboró mediante la selección del subconjunto de SNPs que estuvieran presentes en todas las localidades **-p 7** (alias “grupo 2”). Esta matriz de genotipos con un menor número de SNPs reduce la probabilidad de presencia de *outliers* y posee una menor proporción de datos faltantes. Una representación gráfica de los grupos de SNPs 1 y 2 se presenta en la figura 5 de la siguiente sección. Este filtrado diferencial se relaciona a la naturaleza de las preguntas a responder. En ambos casos es conveniente minimizar el grado de datos faltantes. Sin embargo, para la búsqueda de SNPs bajo posible selección natural, realizar filtrados muy estrictos podría dejar fuera los marcadores de interés, aunque también se puede inducir un sesgo de falsos positivos al relajar los parámetros (Hendricks *et al.*, 2018). En ambos casos se eliminaron individuos con más del 30% de datos faltantes, se seleccionaron SNPs presentes en al menos el 80% de los individuos, con un ***min_maf*** de 0.03 y únicamente un SNP al azar por RAD *locus*.

6.4 Identificación de candidatos a selección y SNPs neutrales

6.4.1 Detección de outliers

Para la identificación de *outliers* (candidatos a estar sujetos a selección natural) a partir del conjunto de SNPs del grupo 1 ($n=6$), se usaron dos programas basados en comparaciones de valores de diferenciación genética poblacional (F_{st}). Particularmente, se buscan valores de F_{st} que presenten desviaciones de lo esperado bajo Equilibrio de Hardy-Weinberg. El enfoque de diferenciación poblacional (DP) no involucra factores ambientales que podrían estar aplicando una presión selectiva para la detección de *outliers* y ha sido el mayormente usado. Se usó el programa BAYESCAN V2.01 (Foll y Gaggiotti, 2008) que determina la probabilidad posterior de que un *locus* esté sujeto a selección mediante un análisis bayesiano, en el cual contrasta dos modelos alternativos, un modelo bajo el efecto de la selección y otro sin selección, estimando la probabilidad posterior de cada modelo para cada *locus*, mediante una aproximación de salto reversible de cadenas de Markov de Monte Carlo (MCMC). El análisis se realizó con una probabilidad previa para el modelo neutral de 1:10, 20 corridas piloto de 5,000 iteraciones cada una, seguido de 100,000 iteraciones con 50,000 réplicas de calentamiento (*burn-in*). Por otro lado, se usó el programa ARLEQUIN v.3.5 (Excoffier y Lischer, 2010), que incluye una extensión del programa FDIST (Beaumont y Nichols, 1996), en donde se calculan valores de G_{st} para cada *locus* y se contrastan contra una distribución neutral resultante de una simulación coalescente, según un modelo finito de islas. Adicionalmente, se realizó una prueba de asociación genotipo-ambiente (AGA) que identifica aquellos SNPs que covarían con variables ambientales que pudieran estar ejerciendo presión selectiva (ver sección 6.8).

6.4.2 Obtención de SNPs neutrales

Para llevar a cabo las estimaciones de diversidad y estructura genética, estimar el grado de parentesco promedio entre localidades y finalmente poner a prueba la hipótesis de aislamiento por distancia, se realizaron para el grupo 2 de SNPs una serie de filtros adicionales.

Una de las características emergentes de trabajar con miles de marcadores moleculares, es que la probabilidad de encontrar desequilibrio de ligamiento (LD, por sus siglas en inglés) aumenta y puede contradecir el supuesto de independencia de observaciones de la mayoría de los análisis estadísticos. Además de seleccionar un único SNP por RAD *locus*, por cuestiones metodológicas relacionadas al tiempo necesario para llevar a cabo la prueba, se decidió realizar una prueba pareada de LD sólo para el grupo 2 mediante la estimación del grado de correlación entre pares de SNPs (r_{barD} ; Agapow y Burt, 2001) con el paquete *poppr* v2.1.8.1 (Kamvar *et al.*, 2015) en R v3.5.1 (R Core team, 2018). La significancia de las aproximadamente 600 mil comparaciones se evaluó mediante pruebas de permutación con 199 iteraciones cada una. El nivel de significancia estadístico fue corregido por múltiples pruebas mediante el método de *false discovery rate* (FDR) de la función *p.adjust* del paquete *stats* v.3.5.1 de R (R Core team, 2018).

El modelo de equilibrio de Hardy-Weinberg (HWE, por sus siglas en inglés), bajo los supuestos de tamaños efectivos poblacionales grandes, apareamiento aleatorio y ausencia de mutación, deriva génica y selección natural, permite predecir las frecuencias genotípicas de la progenie a partir de los genotipos parentales. Bajo estos supuestos (y el modelo de herencia mendeliano), las frecuencias alélicas permanecerían constantes a través de las generaciones, permitiéndonos hacer inferencia acerca de procesos evolutivos bajo un equilibrio mutación-deriva génica. En el presente estudio se realizó una prueba exacta de HWE por localidad, con 10,000 iteraciones de cadena de Markov y un nivel de

significancia de 0.05. El método de FDR de la función $p.adjust$ del paquete *stats* v.3.5.1 de R (R Core team, 2018) se usó para corregir el nivel de significancia por múltiples pruebas. Posteriormente se eliminaron aquellos SNPs en desequilibrio de HW en el 50% de las localidades o más.

Un resumen de los filtros realizados sobre los marcadores se puede consultar en la figura 5. Habiendo concluido los análisis y filtros previamente mencionados, se hará referencia a los SNPs candidatos a selección como *outliers* o *asociados a ambiente*, y a los marcadores relacionados a procesos neutrales como “*neutrales*”.

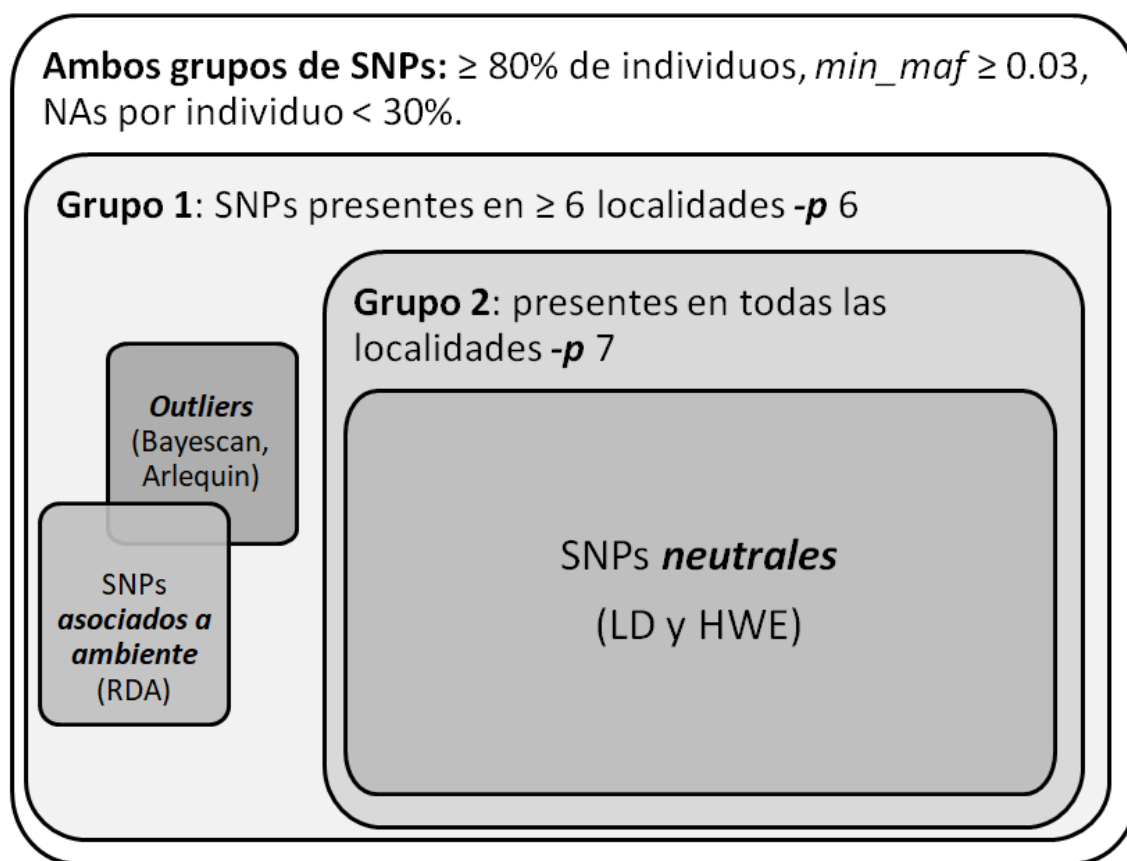


Figura 5. Representación esquemática de filtrado y detección de SNPs *neutrales*, *outliers* y *asociados a ambiente*; min_maf : frecuencia del alelo menor, NAs: datos faltantes, LD: desequilibrio de ligamiento, HWE: equilibrio de Hardy-Weinberg, RDA: análisis de redundancia (ver sección 6.8).

6.5 Diversidad genética y parentesco

La diversidad genética se relaciona con la capacidad adaptativa de las especies a su ambiente, y también es un indicador de potencial evolutivo. Al interior de cada localidad, se estimaron diversos indicadores de diversidad genética para los SNPs neutrales: heterocigosidad observada y esperada con el paquete *adegenet* v2.1.1 (Jombart, 2008; Jombart y Ahmed, 2011), y riqueza alélica rarefaccionada (ponderada por tamaño muestral) con el paquete *PopGenReport* v3.0.4 (Adamack y Gruber, 2014) en R v3.5.1 (R Core team, 2018).

Adicionalmente, como medida indirecta del grado de auto-reclutamiento, se realizó una prueba de parentesco promedio entre localidades mediante el grado de correlación (r) de frecuencias alélicas (Queller y Goodnight, 1989) en GenAIEx v6.5 (Peakall y Smouse, 2012). La prueba se realizó con 999 permutaciones para evaluar la significancia de r de acuerdo a un modelo nulo de apareamiento aleatorio y 1000 réplicas *bootstrap* para calcular los intervalos de confianza al 95%. Dado que la prueba no admite *loci* monomórficos al interior de cada población, los SNPs neutrales fueron filtrados exclusivamente para este análisis para retener solo aquellos polimórficos dentro de todas las localidades.

6.6 Estructura genética poblacional neutral y adaptativa

Para cubrir el primer objetivo en cuanto a la estructura genética neutral y adaptativa se realizó una serie de análisis que se diferencian en sus metodologías y supuestos, con la intención de contrastar los resultados obtenidos y evaluar su consistencia. Estas pruebas se realizaron para los SNPs neutrales y adaptativos.

Uno de los indicadores más usados para cuantificar la diferenciación genética entre localidades o poblaciones es el G_{st} , la adaptación de Nei (1987) del índice de fijación (F_{st}) descrito por Wright (1943), basado en la diferencia relativa entre la heterocigosidad esperada global, y la heterocigosidad esperada promedio de las

localidades muestreadas. El G_{st} puede variar entre cero y uno (los valores negativos son considerados como cero), y es una medida de la proporción de alelos fijos al interior de cada localidad, siendo máximo cuando en todas las localidades se ha dado la fijación de un alelo, sin importar cual sea el alelo (Jost *et al.*, 2018). Se realizó una prueba de G_{st} pareada entre localidades y se evaluó la significancia del G_{st} mediante una prueba de permutación con 10,000 réplicas utilizando el programa ARLEQUIN V.3.5 (Foll y Gaggiotti, 2008). Para evitar errores de falsos positivos el nivel de significancia de la prueba se corrigió mediante la corrección de FDR de la función *p.adjust* del paquete *stats* v. 3.5.1 de R v3.5.1 (R Core team, 2018). Para complementar la información provista por el G_{st} , se estimó también un indicador de diferenciación alélica, el D_{est} (Jost, 2008), que también oscila entre cero y uno; es cero cuando todas las localidades comparten los mismos alelos en la misma frecuencia, y tiene su valor máximo cuando las distintas localidades no comparten ningún alelo. El D_{est} se estimó con el paquete *mmod* (Winter, 2012) en R v3.5.1 (R Core team, 2018).

Tanto G_{st} como D_{est} son estimadores de fijación o diferenciación alélica derivados de los niveles de heterocigosidad local y global, bajo el modelo de EHW. Al ser una comparación pareada, el valor obtenido sólo depende del par de localidades que están siendo comparadas. En los últimos años ha crecido en popularidad un estimador que describe la distribución de la covarianza genética independientemente de modelos evolutivos y utiliza toda la información disponible (a diferencia de las medidas pareadas mencionadas anteriormente): la distancia genética condicional (cGD; Dyer y Nason, 2004). La cGD deriva de la aplicación de los *population graphs*, en donde la variación genética primero se descompone en dos componentes geométricos, *entre* y *dentro* de las localidades (de modo análogo al análisis AMOVA), y luego la estructura gráfico-teórica generada se modifica usando la covarianza genética condicional para determinar la configuración topológica mínima (en inglés, *shortest path distance*) necesaria para

describir la totalidad de la variación genética (Dyer, 2015). Así, es posible identificar vínculos redundantes en la red.

La cGD se puede visualizar en forma de matriz o como una red topográfica de nodos y vínculos. El diámetro de cada nodo representa la contribución al componente *dentro* de la varianza (análogo a la diversidad genética), y el grosor de cada vínculo representa la covarianza *entre* cada par de sitios. Para estimar la cGD y generar las redes topográficas se utilizaron los paquetes *gstudio* v1.5.0 y *popgraph* v1.5.0 (Dyer, 2016) en R v 3.5.1 (R Core team, 2018). La importancia de cada sitio dentro de la red se estimó mediante el índice de *betweenness centrality*, un indicador de su rol en la mantención de la conectividad. Las topologías fueron visualizadas geográficamente y mediante la aplicación del algoritmo de Fruchterman Reingold, que busca el equilibrio entre las fuerzas de atracción y repulsión entre nodos en un plano.

Adicionalmente, para identificar el número más probable de poblaciones se aplicó un algoritmo de agrupamiento de individuos mediante la función *find_clusters* con el paquete *adegenet* v1.3.1 (Jombart, 2008; Jombart y Ahmed, 2011) en R v3.5.1 (R Core team, 2018), evaluado mediante el criterio bayesiano de información (BIC). A partir del agrupamiento recuperado y para comparar con el agrupamiento *a priori* de muestras de acuerdo a las localidades, se realizaron tres análisis discriminantes de componentes principales (dAPC, por sus siglas en inglés) con *adegenet* v1.3.1 (Jombart, 2008; Jombart y Ahmed, 2011) en R v3.5.1 (R Core team, 2018). En estos análisis (SNPs neutrales por grupo; neutrales y *outliers* por localidad) se evalúa la agrupación de las localidades o grupos de individuos muestreados en el espacio multivariado de acuerdo con los genotipos individuales. Al estar basado en análisis de componentes principales, el método tampoco depende de supuestos evolutivos (EHW), ni supone la ausencia de LD (Jombart, 2008). El número de componentes principales a retener fue estimado mediante el método de validación cruzada (*xvalDapc*) con 100 réplicas, en donde el set de datos se divide en 90% de generación y 10% de validación, estimándose la

proporción de asignación correcta del set de validación con el aumento de componentes retenidos en el set de generación. El número óptimo de componentes a utilizar es aquel que presenta el mayor poder predictivo, y el número más probable de poblaciones se identifica de acuerdo con la distribución de las localidades o grupos en la gráfica resultante del análisis.

Finalmente, se determinó la estructura espacial y el número más probable de poblaciones (K) mediante un análisis Bayesiano con el programa STRUCTURE v.2.2 (Pritchard *et al.*, 2000). STRUCTURE se basa en un algoritmo de MCMC para asignar la probabilidad de pertenencia de los individuos de acuerdo a sus genotipos *multilocus* a una o varias poblaciones, si los genotipos demuestran mezcla (Falush *et al.*, 2007; Pritchard *et al.*, 2000). El análisis se ejecutó con 10,000 iteraciones de *burn in* y 100,000 iteraciones de MCMC, utilizando el modelo de mezcla, sin incluir el origen de las muestras. Se evaluaron valores de K de 1 a 7, con 10 réplicas para cada valor de K. El número más probable de poblaciones se identificó con el método de ΔK (Evanno *et al.*, 2005) implementado en el sitio web Structure Harvester (<http://taylor0.biology.ucla.edu/-structureHarvester/>). Los resultados fueron analizados con el programa CLUMPP v1.1.2 (Jakobsson y Rosenberg, 2007) y las figuras resultantes se generaron con el programa Distruct v1.1 (Rosenberg, 2004).

6.7 Aislamiento por distancia

Para llevar a cabo la comparación de estructura genómica neutral y distancia geográfica de acuerdo al objetivo 2, se calculó la distancia euclidiana entre sitios muestreados con el paquete *raster* v2.8-4 (Hijmans, 2018) en R v3.5.1 (R Core team, 2018).

La conectividad oceanográfica potencial se generó a partir de una simulación que predice patrones de dispersión larvaria de acuerdo al modelo HAMSOM de corrientes adaptado al GC descrito en detalle por Lavin y Marinone (2003), y

Marinone (2003, 2012). Este modelo originalmente descrito por Backhaus (1985) tiene una resolución geográfica de aproximadamente 1.3 x 1.5 km y 12 capas en el eje vertical, de las cuales solo se utilizó la capa superficial (0 a 10 m de profundidad) para la simulación. El modelo incluye campos de temperatura y salinidad completamente pronósticos, lo que permite un movimiento baroclínico temporalmente dependiente. En la capa superficial también se incluyeron forzamientos derivados de estaciones meteorológicas, como régimen de vientos y flujos de agua dulce.

A partir de 59 polígonos generados en las costas e islas del GC -en donde cada polígono contiene un sitio de liberación de partículas y un área determinada (Fig. 5)- se liberaron 4000 partículas por sitio cuya trayectoria fue registrada. Se estimó, para pleamar y bajamar de cada mes del año, la proporción de partículas cuyo destino final fuera cualquiera de los 59 polígonos mencionados. Así, a partir de las 2 matrices mensuales (12 meses del año), para valores de PLD de 4, 6 y 8 semanas (2 mareas x 12 meses x 3 PLD = 72 matrices de 59x59 polígonos) se seleccionaron y promediaron aquellas que pudiera presentar *Lythrypnus dalli*, resultando en 13 hipótesis oceanográficas derivadas de la combinación de distinta duración de larva pelágica y temporada reproductiva (uno a 10 meses de duración) de acuerdo a lo reportado en literatura para la especie (Tabla II). Las matrices de conectividad fueron realizadas por el Dr. Guido Marinone y la M.C. Carolina Montaña-Cortés (CICESE).

Tabla II. Hipótesis de conectividad oceanográfica de *Lythrypnus dalli* de acuerdo a su temporada reproductiva y duración de larva pelágica (PLD) en semanas.

Predictor	Temporada reproductiva	PLD (sem.)	Justificación
Ha1	may:jun	8	El pico de reclutamiento (jul:ago) ocurre luego de uno o pocos eventos dispersivos determinantes. Desoves parciales no significativos en cuanto a conectividad genética ^{1,2} .
Ha2	jun:jul	8	
Ha3	may:jun	6	
Ha4	jun:jul	6	
Ha5	may	4	
Ha6	jun	4	
Ha7	jul	4	
Ha8	abr:jun	8	Se contempla la llegada de reclutas durante el verano, luego de varios desoves parciales que ocurren en otoño ^{1,2} .
Ha9	abr:jun	6	
Ha10	abr:jun	4	
Ha11	dic:sep	8	Abarcan el mayor rango de reclutamiento reportado para la especie en Bahía de la Paz ² .
Ha12	dic:sep	6	
Ha13	dic:sep	4	

Referencias: ¹ Behrents (1983); ² Calderón Parra (2009).

Para poder llevar a cabo las comparaciones con los indicadores bidireccionales y simétricos de distancia genética entre cada par de sitios muestreados, las 13 matrices de conectividad oceanográfica (direccionales y asimétricas) que solo reflejan la conexión directa entre algunos pares de sitios de un evento dispersivo individual (lo que representaría una sola generación) de acuerdo a los meses promediados, fueron estandarizadas mediante la aplicación de la teoría de *population graphs* (Dyer y Nason, 2004; Dyer, 2015). Para ello se calculó la distancia sobre la gráfica (*graph distance*) de cada una de las 13 redes de conectividad con el paquete *popgraph* para obtener un estimador bidireccional de conectividad entre cada par de sitios que represente las conexiones indirectas (que integran múltiples eventos dispersivos en múltiples generaciones), de manera similar a estudios previos en el GC (Munguia-Vega *et al.*, 2014, 2018a).

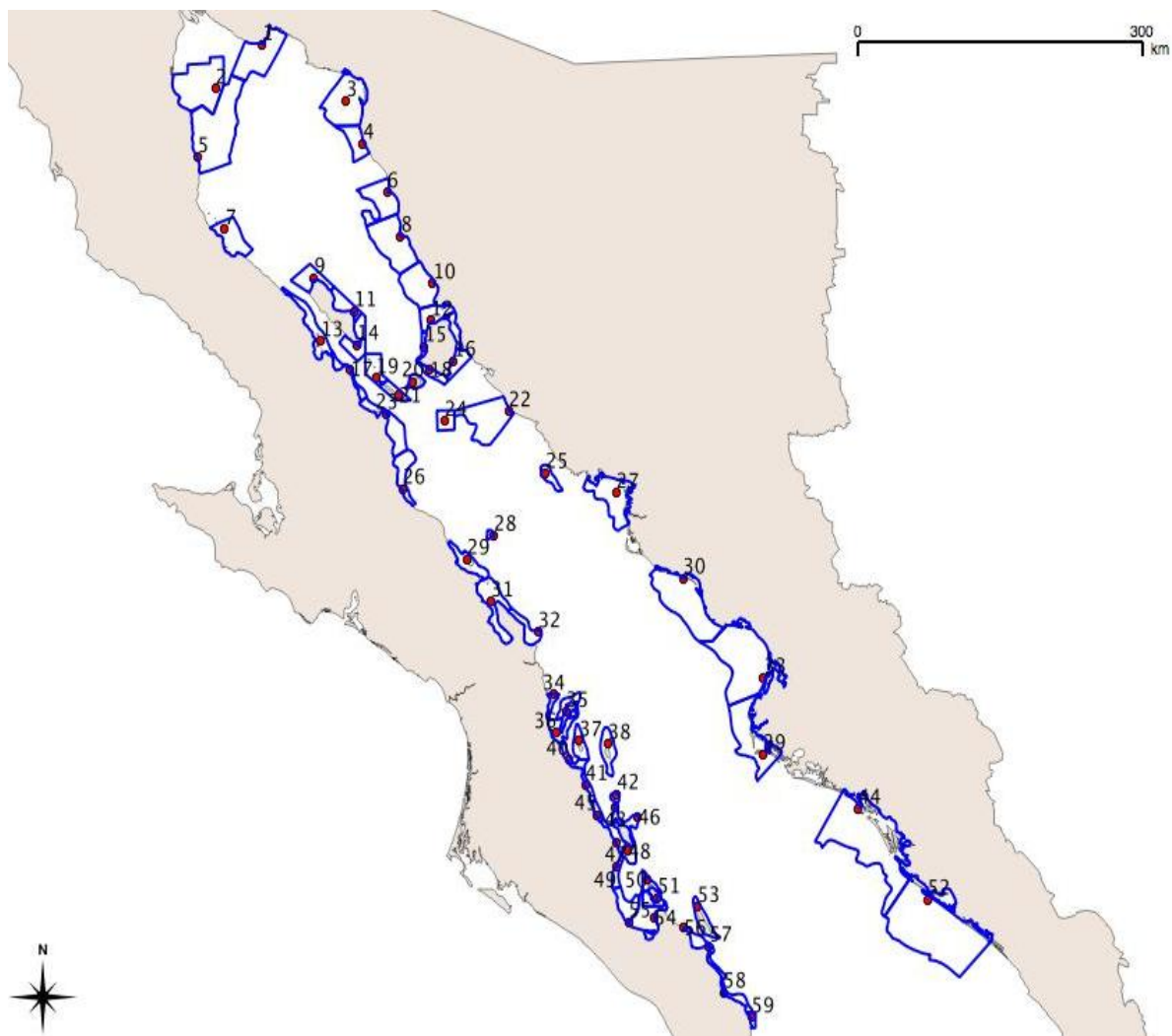


Figura 6. 59 sitios de liberación de partículas del modelado oceanográfico (puntos rojos) y sus respectivos polígonos (imagen provista por el Dr. Adrián Munguía Vega).

Una vez obtenidas las matrices de distancia geográfica euclidiana y oceanográficas, se compararon contra las matrices de G_{st} , D_{est} y cGD mediante pruebas de Mantel con la función *mantel.randtest* del paquete *ade4* (Dray y Dufour, 2007) en R v3.5.1 (R Core team, 2018) para poner a prueba la hipótesis de aislamiento por distancia. En base al modelo de conectividad oceanográfico más probable, se estimó para cada sitio el número de sitios de donde recibe (*n.importa*) y hacia donde exporta (*n.exporta*) migrantes, indicando así su rol en la

red como fuente (aquellos sitios que netamente exportan a más sitios que lo que importan) o sumidero (sitios que reciben migrantes de más sitios que de los que exportan) con el programa Gephi v0.9.2 (Bastian y Heymann, 2009). Adicionalmente, se relacionaron los niveles de diversidad genética y parentesco contra el grado de retención local y el rol como fuente o sumidero predicho mediante regresión lineal con la función *lm* del paquete *stats* de R v3.5.1 (R Core team, 2018). Los supuestos de la regresión fueron verificados gráficamente.

6.8 Aislamiento por ambiente

En cuanto al objetivo 3, que involucra la comparación de la estructura genómica adaptativa y la heterogeneidad ambiental, se obtuvieron imágenes de temperatura superficial del mar (SST) y clorofila-a (Chla) del satélite AQUA MODIS, que comprende el período de julio 2002 a julio 2017, para el GC (Fig. 7). A partir de las imágenes, se extrajeron datos de SST y Chla correspondientes a los 59 polígonos mencionados en la sección anterior, a partir de estos se calcularon promedios para los 12 meses del año mediante el uso del paquete *raster* v2.8-4 (Hijmans, 2018) de R v3.5.1 (R Core team, 2018). Usando los promedios mensuales se calculó el promedio anual, desviación estándar y rango (máximos y mínimos). Posterior a esto se realizó un análisis de componentes principales para evaluar su estructura y covariación. Adicionalmente, se calculó la distancia euclidiana ambiental entre pares de sitios mediante la función *vegdist* del paquete *vegan* v2.5-2 (Oksanen *et al.*, 2018) en R v3.5.1 (R Core team, 2018).

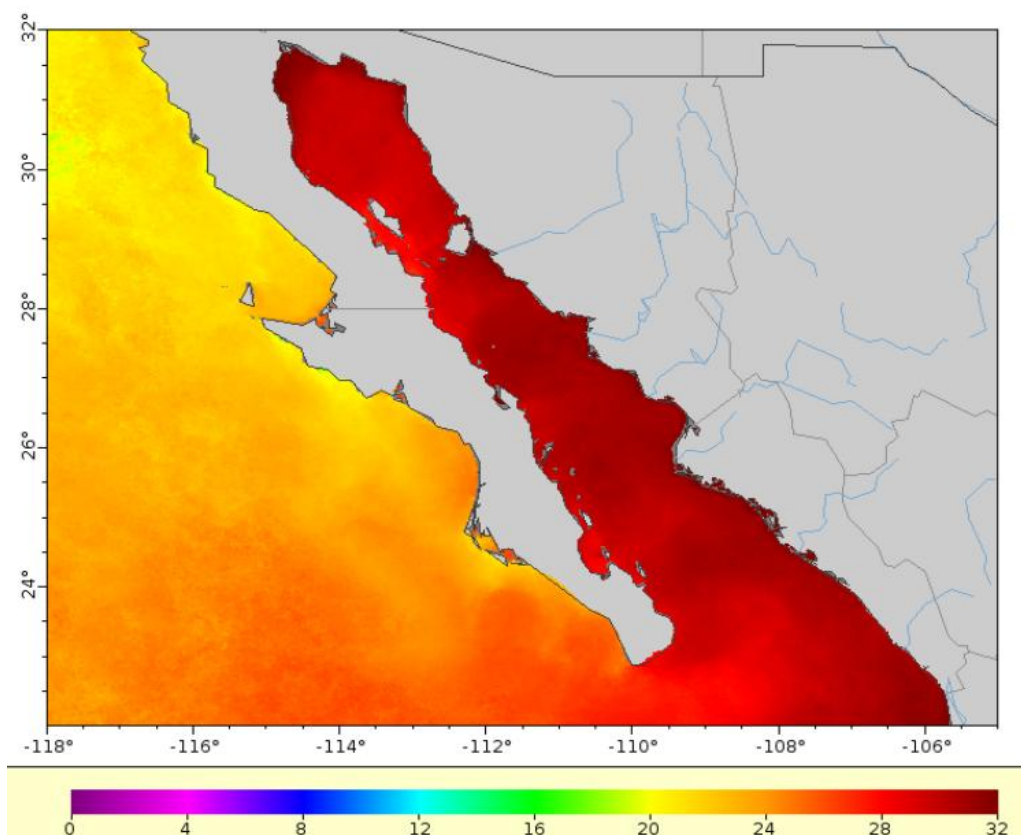


Figura 7. Temperatura superficial del mar (SST) correspondiente a Julio de 2017 en el GC. Imagen generada a partir de la información obtenida del satélite AQUA MODIS.

Para poner a prueba la hipótesis de aislamiento por ambiente se realizó por un lado la comparación entre las matrices de G_{st} , D_{est} y cGD de los SNPs *outliers* y la distancia ambiental, mediante pruebas de Mantel con la función *mantel.randtest* del paquete *ade4* (Dray y Dufour, 2007) en R v3.5.1 (R Core team, 2018).

Por otro lado, se usó uno de los análisis de asociación genético-ambiental recientemente descritos (AGA; Rellstab *et al.*, 2015), que se basa en la correlación entre las frecuencias alélicas y variables ambientales. Existe una amplia gama de análisis disponibles para este fin, de los cuales se seleccionó el análisis de redundancia (RDA, por sus siglas en inglés) por su buen desempeño en estudios de simulaciones bajo distintos escenarios genéticos, como aislamiento por distancia y el modelo de islas (Forester *et al.*, 2018; Lotterhos y Whitlock, 2015). El

RDA es un método multivariado de ordenamiento restringido que busca ordenar la variación genética de modo que maximice su relación con la variación ambiental, en este caso las variables resumen de temperatura y clorofila-a. El análisis fue llevado a cabo con el paquete *vegan* v2.5-2 (Oksanen *et al.*, 2018) en R v3.5.1 (R Core team, 2018). Dado que las variables ambientales presentaron un elevado nivel de multicolinealidad, se utilizaron como predictores un subgrupo de variables cuyo factor de inflación de la varianza (VIF) no fuera mayor a 4, y las cargas de los dos primeros componentes principales, por su cualidad de independencia, y porque en conjunto explican gran parte de la varianza total.

7. RESULTADOS

7.1 Muestreo

El muestreo se llevó a cabo entre junio y noviembre de 2017. Se realizó un total de 28 buceos, a razón de 1 a 3 buceos por localidad, y 3 a 6 cuadrantes por inmersión. *Lythrypus dalli* fue colectado en 14 sitios, mayoritariamente en buceos cercanos a islas, con excepción de dos sitios en la costa de la Península: Puerto San Francisquito (PSF, Baja California), y PDI (Bahía de La Paz, Baja California Sur). Los individuos colectados fueron agrupados en 7 localidades en función de la resolución espacial de las variables explicativas (modelo de dispersión larvaria y variables ambientales). La tabla III presenta un resumen de número de organismos colectados por localidad, en relación al número de buceos y cuadrantes, talla promedio y profundidad promedio de colección.

Tabla III. Resumen del esfuerzo (buceos y cuadrantes) de muestreo en número (n.º); profundidad promedio (z); individuos colectados y tallas promedio por localidad (ordenados de norte a sur).

Localidad	Código	n.º buceos	n.º cuadrantes	z (m)	n.º <i>L. dalli</i>	Talla (mm)
Islas Bahía de los Ángeles	BA	3	13	6.32	36	24.31
Puerto San Francisquito	PSF	1	4	1.91	5	32.20
Isla San Marcos	SM	2	8	6.56	27	19.33
San Cosme y Agua Verde	SCAV	5	21	6.48	51	16.54
Isla El Pardito	SEPA	1	4	7.32	15	18.07
Isla San Francisquito	SEISF	1	5	7.01	35	18.31
Punta Diablo	PDI	1	6	9.08	17	22.18
Total		14	61		186	

Con la excepción de PSF, en donde se colectaron solo 5 individuos, se cuenta con 15 a 51 individuos por localidad. No existe una relación evidente entre el esfuerzo de muestreo y el número de individuos capturados. No se observa una tendencia latitudinal en el número de colectas ni en la talla de la especie. La variación en el esfuerzo de muestreo tiene que ver con la ubicación de los sitios de colecta y la

resolución espacial de los predictores. Por ejemplo, en SCAV se realizaron 2 y 3 buceos que quedaron agrupados en un mismo conjunto. En cambio, Isla El Pardito (SEPA) e Isla San Francisquito (SEISF) fueron dos sitios de buceo muy cercanos entre sí, y los buceos se realizaron en un mismo evento de colecta desde la localidad de San Evaristo. Sin embargo, estos sitios pertenecen a polígonos diferentes en los predictores, por lo que fueron separados en dos localidades.

7.2 Elaboración de librerías genómicas y secuenciación

Se realizó la extracción de ADN de los 186 individuos colectados y se obtuvo ADN de alto peso molecular (Fig. 8A). Sin embargo, nueve individuos de SCAV, dos individuos de SEISF y un individuo de BA fueron excluidos del análisis debido a la baja masa inicial de tejido disponible para la obtención de ADN o debido a la baja eficiencia de la extracción en estos casos, restando 174 individuos. La doble digestión y posterior ligación de adaptadores fue verificada mediante un gel de agarosa-synergel al 2% (Fig. 8B). Las muestras fueron posteriormente combinadas equimolarmente, incluyéndose 10 ng de ADN por individuo para 3 pools, y 30 ng de ADN por individuo para el cuarto pool.

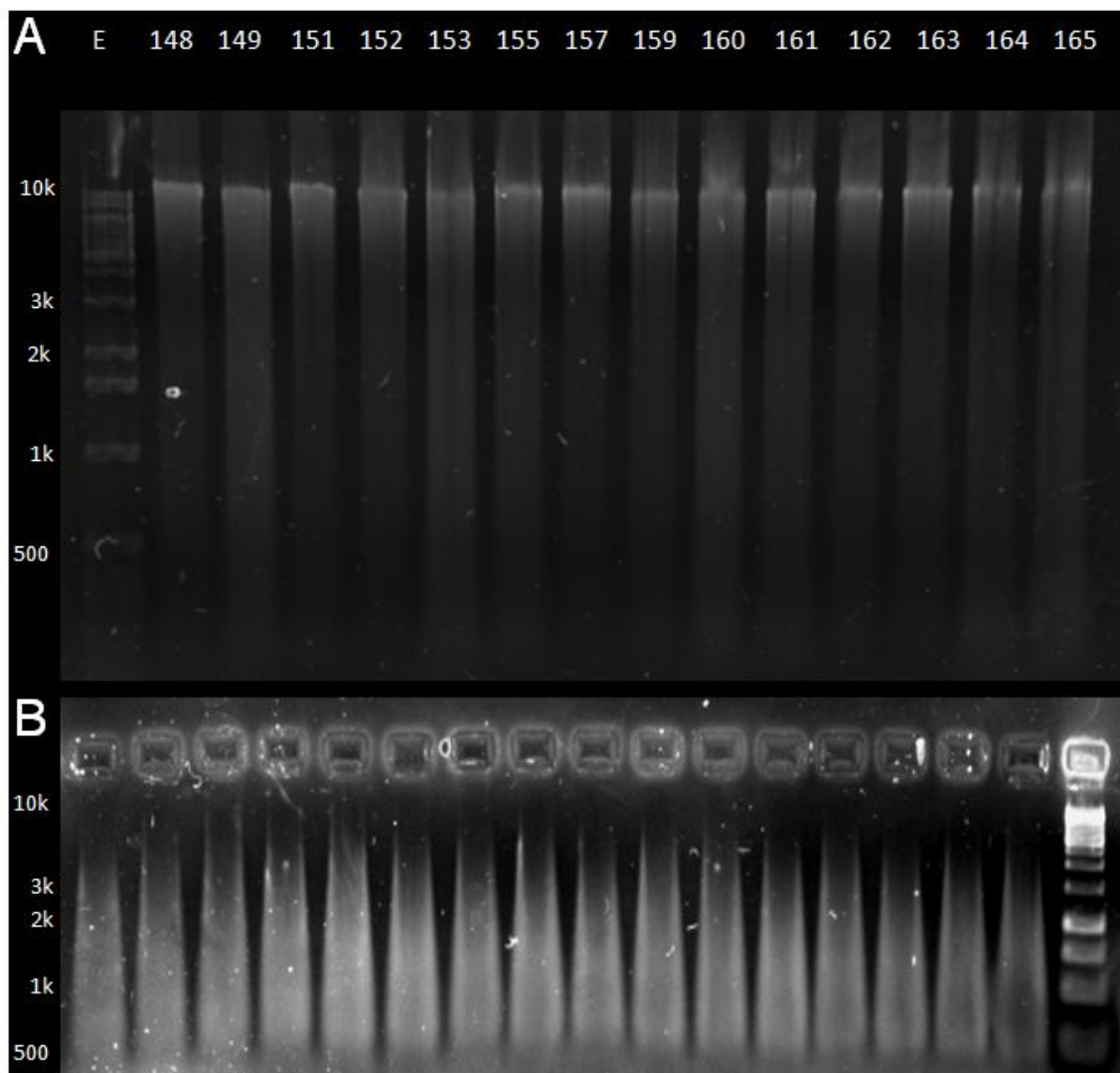


Figura 8. Extracción y digestión de ADN para la elaboración de librerías genómicas. (A) Ejemplo de verificación de pureza y homogeneidad de diluciones de ADN de *Lythrypnus dalli*; (B) Ejemplo de verificación de doble digestión y ligación de adaptadores. k: 1000 pb; E: marcador de peso molecular 1 kb Plus (Thermo Fisher Scientific).

Posteriormente, se realizó la selección de fragmentos y amplificación mediante PCR de los 4 pools, y se verificó mediante un gel de agarosa-synergel al 2% (Fig. 9).

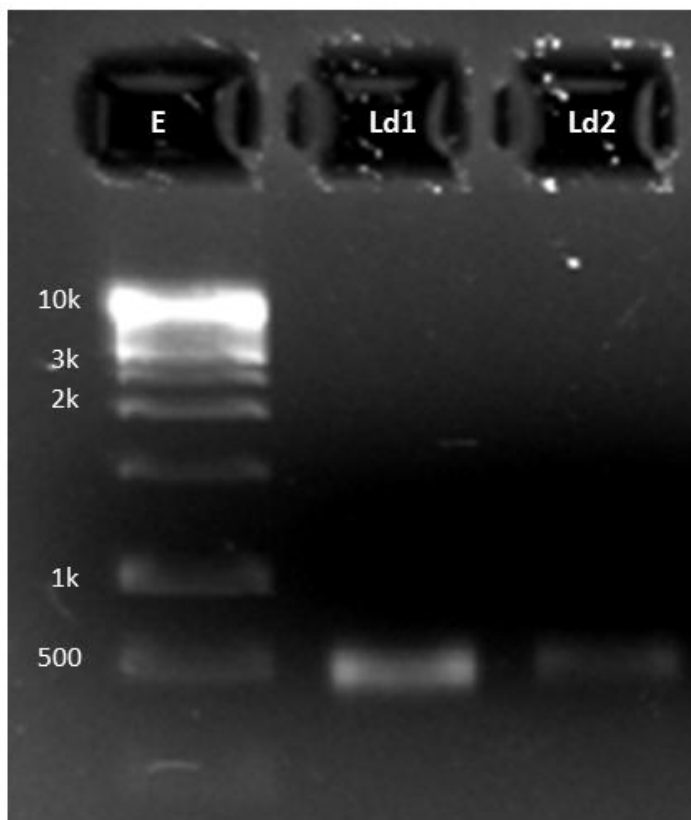


Figura 9. Ejemplo de verificación de selección de fragmentos y amplificación de ADN vía PCR de los pools de *Lythrypnus dalli* (k: 1000 pb; E: marcador de peso molecular 1 kb Plus (Thermo Fisher Scientific)).

Los pools fueron cuantificados y combinados equimolarmente en dos librerías finales compuestas por un par de pools cada una, con un contenido total de ADN de 70.2 ng y 126.4 ng.

7.3 Demultiplexado de librerías y obtención de SNPs

La evaluación de la calidad de las secuencias obtenidas para cada librería indicó que los datos de secuenciación obtuvieron valores de phred score > 20 a lo largo de todo el rango de la longitud de las secuencias de 150 pb, lo cual indica una elevada confianza en la correcta asignación de las bases (Fig. 10).

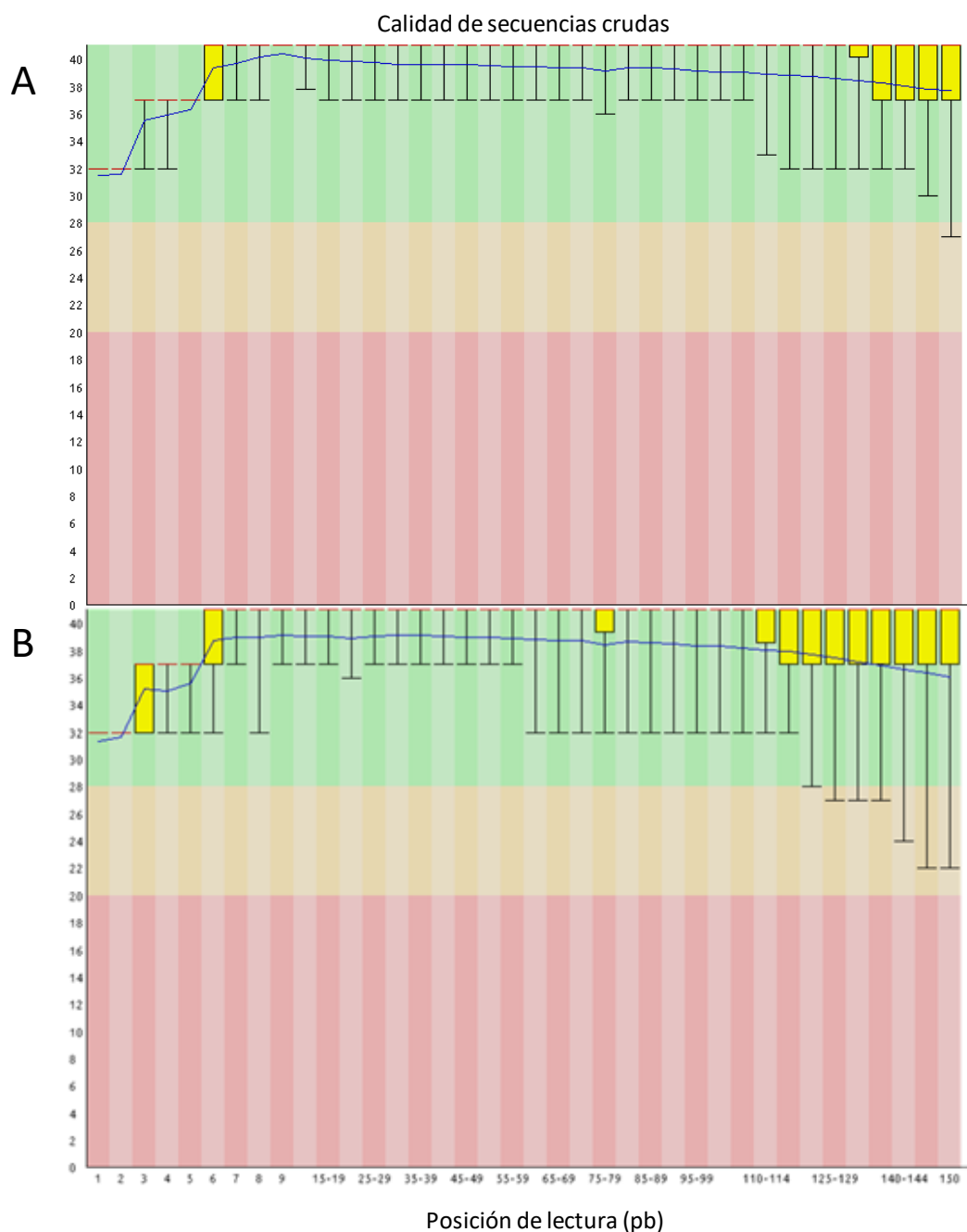


Figura 10. Diagrama de caja de calidad de secuencias crudas (*phred score*, eje y). Lecturas forward y reverse del pool Ld2. Los reportes son muy similares para los tres pools restantes. La línea azul representa las medianas y los colores verde, amarillo y rojo indican la calidad promedio buena (*phred* >30), razonable (*phred* >20) o mala (*phred* <20), respectivamente.

De un total de 1,507,953,258 secuencias crudas, el 97.79% pasó el filtro de calidad, la verificación del *barcode* y de las secuencias de corte de MspI y EcoRI durante el demultiplexado y filtrado inicial con *process_radtags* (Tabla IV).

Tabla IV. Filtrado de secuencias crudas en *process_radtags*.

	número	%
Secuencias crudas	1,507,953,258	100.00
Barcodes ambiguos	11,455,328	0.76
Phred score < 10	1,206,125	0.08
Corte de MspI ambiguo	20,670,750	1.37
Secuencias retenidas	1,474,621,055	97.79

7.3.1 Parametrización de STACKS

Para la identificación de parámetros óptimos en la obtención de los genotipos, dados los objetivos planteados, se realizó una serie de análisis *de novo* en STACKS iterando valores de M entre 1 y 6 y valores de n entre 1 y 7 para un subgrupo de 4 localidades con 26 a 42 individuos cada una. El aumento de diferencias permitidas entre fragmentos para generar una pila (M) no tuvo efecto en el número de RAD *loci* obtenidos (Fig. 11A), mientras que sí generó una disminución en el número de SNPs resultantes (Fig. 11B). En cambio, con el aumento de diferencias permitidas entre pilas para formar un RAD *loci* (n) se da una disminución en el número de RAD *loci* y un aumento en el número de SNPs ensamblados (Fig. 11). Esto se debe a que, al aumentar el número de diferencias que se permiten entre pilas para considerarlos parte de un mismo RAD *locus*, una mayor proporción de pilas resultan combinados en menos RAD *loci*, conteniendo estos ensamblados más sitios variables al interior.

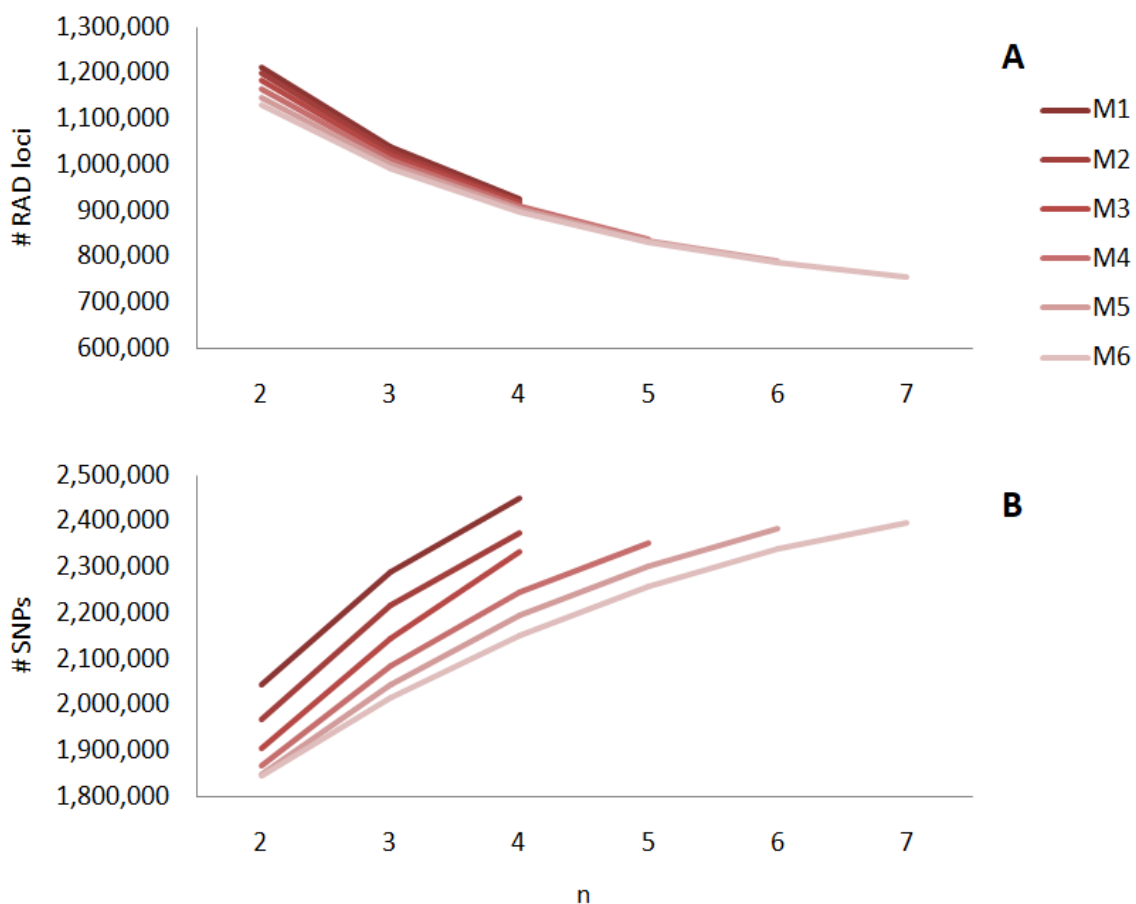


Figura 11. Número de RAD *loci* para *Lythrypnus dalli* (A) ensamblados y número de SNPs (B) obtenidos *de novo* iterando valores de **M** y **n** en STACKS.

Otro criterio de selección considerado fue la profundidad de lecturas por *locus* y por individuo. En este caso, no se observa un efecto de los parámetros **M** ni **n** en la profundidad efectiva por *locus* ($X_{locus} = 114 \pm 34$), mientras que la profundidad promedio por individuo aumenta con el aumento de **M** y no es afectada por el aumento de **n** (Fig. 12). Esto se debe a que, al permitir más diferencias, fragmentos de mayor variabilidad están siendo apilados en menos pilas, resultando en profundidades mayores.

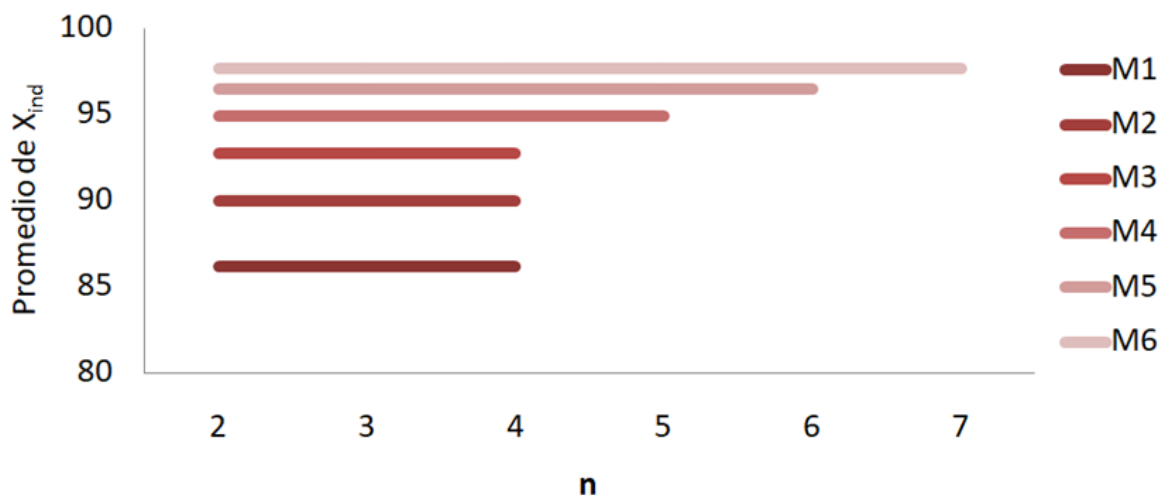


Figura 12. Promedio de profundidades de lectura por individuo (X_{ind}) de *Lythrypnus dalli* obtenidas iterando valores de M y n en STACKS.

Los SNPs obtenidos con las distintas combinaciones de M y n fueron filtrados con el módulo *populations* de STACKS para retener solo aquellos SNPs presentes en el 80% de los individuos o más (r 80%), con una *min_maf* de 0.05 y preliminarmente retener aquellos SNPs presentes en 3 de las 4 localidades incluidas en la parametrización (p 3). El número de RAD *loci* y SNPs retenidos disminuyen de manera similar con el aumento de M (Fig. 13). Cabe destacar que el número de SNPs obtenidos (manteniendo M constante) tiende a estabilizarse a valores de n mayores a 4, por lo que se seleccionó a n 4 como valor óptimo para el presente análisis. De este modo, el número máximo de diferencias entre fragmentos para generar una pila (M) entre 3 y 5 serían adecuados de acuerdo a las recomendaciones de Paris *et al.* (2017) sobre la relación de M y n , entre los cuales se seleccionó M 3, ya que se obtiene un mayor número de SNPs para análisis posteriores (Fig. 13), a pesar de resultar en una profundidad de lecturas por individuo ligeramente menor (Fig.12).

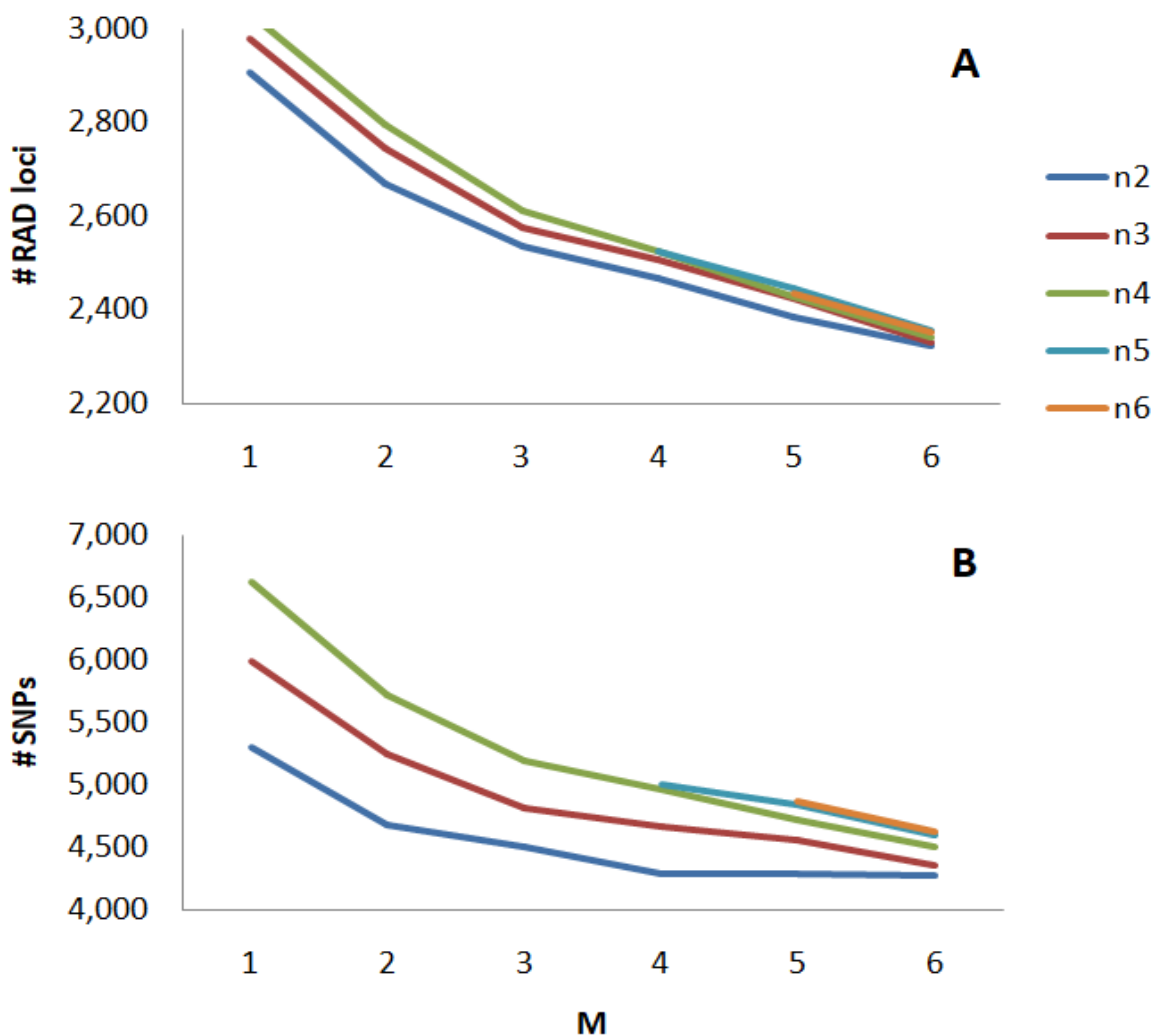


Figura 13. Número de RAD *loci* de *Lythrypnus dalli* (A) y número de SNPs (B) obtenidos luego del filtrado de los genotipos obtenidos para cada combinación de *M* y *n*.

Finalmente, se analizó el grado de datos faltantes resultantes para un subconjunto de combinaciones de *M* y *n* (ambos entre 2 y 4) con VCFtools, no encontrándose variación sustancial debida a estos parámetros ($18 \pm 1\%$ de datos faltantes por individuo en promedio \pm desviación estándar (DE)).

7.3.2 Filtrado de individuos y SNPs

Habiendo identificado los parámetros a utilizar (m 4, M 3, n 4), cinco individuos presentaron una profundidad de lecturas promedio menor a 40X y fueron eliminados de análisis posteriores (Tabla V). En el ensamblado *de novo* con los 161 individuos retenidos se obtuvieron 1,094,917 RAD *loci*, con 2,886,185 sitios variantes totales. La profundidad de lecturas efectiva por *locus* fue en promedio 115.6X, con un mínimo de 51.6X y un máximo de 190X; por individuo se obtuvo una profundidad entre 43.6X y 167.7X, con un promedio de 102X.

El filtrado en el módulo *populations* se realizó con los parámetros r 0.8, *min_maf* 0.03 y *write_random_snp*. En el caso del “grupo 1” de SNPs, en donde se retuvieron *loci* presentes en 6 de las 7 localidades como mínimo (p 6), se obtuvieron 1,711 SNPs con un $13 \pm 6\%$ (media \pm DE) de datos faltantes por individuo. Para el subconjunto de SNPs del “grupo 2”, en donde se retuvieron aquellos presentes en todas las localidades (p 7), quedaron 1270 SNPs con un $7 \pm 5\%$ (media \pm DE) de datos faltantes por individuo. En ambos casos fueron eliminados del análisis 8 individuos con más del 30% de datos faltantes (Tabla V).

Tabla V. Filtrado de individuos de acuerdo a la profundidad de lecturas por individuo y el límite de datos faltantes por individuo (NAs). BA: Bahía de los Ángeles, PSF: Puerto San Francisquito, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

	BA	PSF	SM	SCAV	SEPA	SEISF	PDI	Total general
individuos secuenciados	35	5	27	42	15	33	17	174
X < 40	1	0	0	1	1	2	0	5
subtotal	34	5	27	41	14	31	17	169
NAs > 30%	7	0	1	0	0	0	0	8
subtotal	27	5	26	41	14	31	17	161

A modo de ejemplo, del total de sitios variantes (2,886,185) del grupo 1, 96,566 SNPs (3.35% del total) están presentes en el 80% de los individuos y 38,922

SNPs (1.35% del total) en 6 de las 7 localidades como mínimo. De ellos, 5,859 (0.20% del total) marcadores presentan un *maf* mayor a 0.03 y finalmente 1,711 SNPs (0.06% del total) se retuvieron al seleccionar un único SNP por RAD *locus* (Tabla VI). El error de genotipificación se estimó en $0.99 \pm 0.59\%$, de acuerdo a la comparación de los genotipos de 10 individuos incluidos por duplicado. El onceavo individuo duplicado presentó valores anormales en la comparación, posiblemente debido a errores durante la elaboración de librerías genómicas, por lo que no se consideró para la estimación del error de genotipificación.

Tabla VI. Filtrado de grupo 1 de SNPs de *Lythrypnus dalli* para análisis posteriores. El porcentaje de SNPs retenidos es con respecto al número inicial.

Filtrado secuencial	número de SNPs	% retenidos
Inicial	2,886,185	-
80% de individuos	96,566	3.35
6 de 7 localidades	38,922	1.35
<i>maf</i> ¹ 0.03	5,859	0.20
1 SNP por RAD <i>locus</i>	1,711	0.06

Notas: ¹ frecuencia del alelo menor.

Por último, la localidad de Puerto San Francisquito con un tamaño muestral de 5 individuos fue eliminada de análisis posteriores, ya que la diferencia en tamaños muestrales con respecto al resto de las localidades generó en análisis preliminares una interferencia tanto en estimaciones de diversidad como de estructura genética. Por ende, ambos sets de SNPs: 1,711 del “grupo 1” y 1,270 del “grupo 2” quedaron compuestos por 156 individuos en total.

7.4 Identificación de candidatos a selección y SNPs neutrales

7.4.1 Detección de outliers

Para la identificación de *loci outliers* (candidatos a estar sujetos a presión selectiva) mediante análisis de diferenciación poblacional a partir del grupo 1 de

SNPs se realizaron dos pruebas de metodología y supuestos diferentes, con la intención de identificar aquellos candidatos de mayor robustez detectados por ambas pruebas. El programa BAYESCAN v2.01 (Foll y Gaggiotti, 2008) identificó un único outlier habiendo establecido el FDR al 5% (Fig. 14A) mientras que el programa ARLEQUIN v.3.5 (Excoffier y Lischer, 2010), identificó a 27 SNPs *outliers* ($p < 0.01$, Fig. 14B), 14 de los cuales fueron considerados falsos positivos y desestimados por encontrarse en HWE en todas las localidades, restando 13 SNPs *outliers* que fueron utilizados para análisis posteriores.

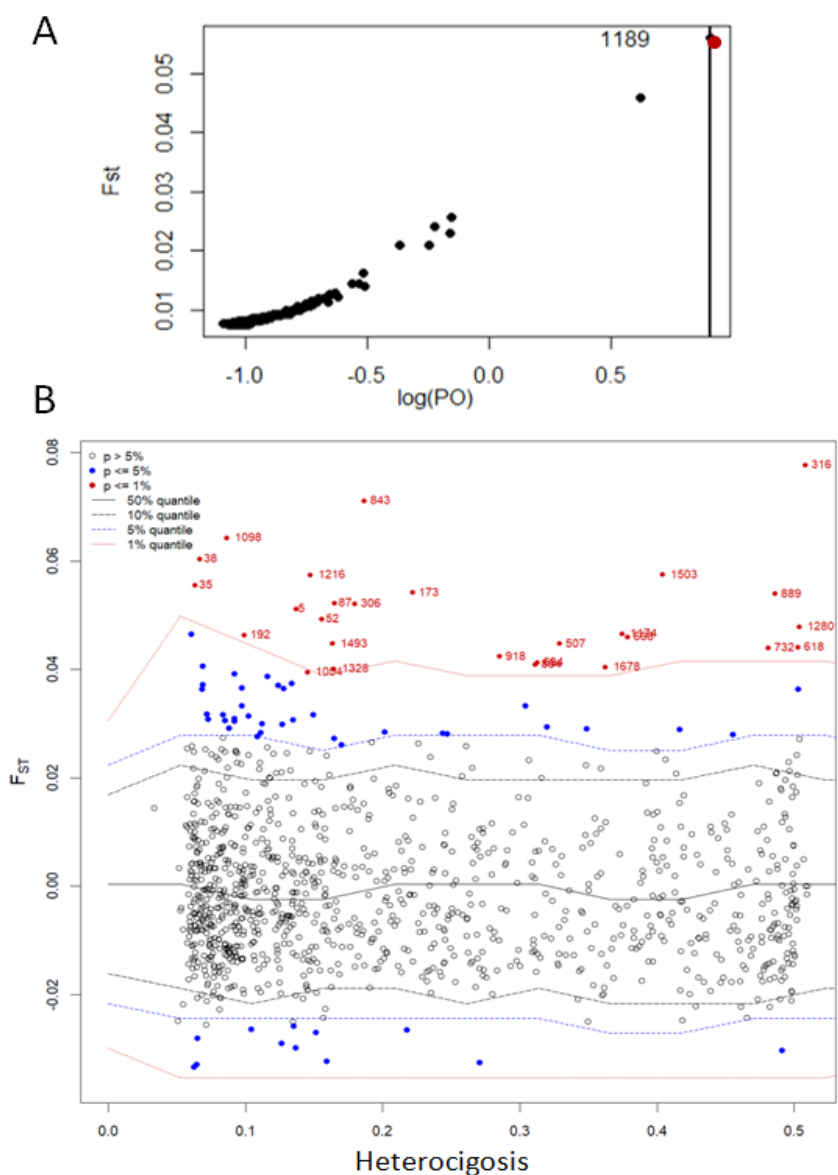


Figura 14. Resultados de análisis de detección de *outliers* de *Lythrypnus dalli* a partir de 1711 SNPs del grupo 1 (en color rojo). (A) Prueba de Bayescan; (B) Prueba Arlequin.

7.4.2 Obtención de SNPs neutrales

El subconjunto de 1270 SNPs del grupo 2 fueron sometidos a pruebas de LD mediante la estimación pareada del grado de correlación r_{barD} (Agapow y Burt, 2001), no encontrándose ningún par de SNPs significativamente en LD ($p > 0.5$).

para todas las comparaciones). En el último filtro aplicado, la prueba exacta de HWE por localidad (corregida mediante FDR), se eliminaron 9 SNPs en desequilibrio ($p < 0.05$) en el 50% de las localidades o más, restando un set de 1,261 SNPs *neutrales* para análisis posteriores.

7.5 Diversidad genética y parentesco

Los niveles de heterocigosidad observada, esperada y riqueza alélica rarefaccionada se presentan en la tabla VII. Se observa poca variación en los niveles de diversidad genética, siendo las localidades centro-sureñas (y muy cercanas entre sí) de SEISF (H_O) y SEPA (H_E , Ar) quienes presentan los menores niveles de diversidad, mientras que SCAV presentaron los valores máximos de heterocigosidad esperada y riqueza alélica. En cuanto a la heterocigosidad observada, el máximo valor corresponde a PDI, superando a SCAV por una centésima. A pesar de haber utilizado el método de rarefacción para estimar la riqueza alélica, aún se detecta un nivel de correlación elevado ($r^2 = 0.91$) entre Ar y tamaño muestral, por lo que deben ser interpretados con cautela.

Tabla VII. Indicadores de diversidad genética por localidad de *L. dalli*. Tamaño muestral (n), heterocigosis observada (H_O) y esperada (H_E), y riqueza alélica rarefaccionada (Ar). BA: Bahía de los Ángeles, PSF: Puerto San Francisquito, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

Localidad	n	H_O	H_E	Ar
BA	27	0.1897	0.2101	1.8164
SM	26	0.1904	0.2131	1.8087
SCAV	41	0.1926	0.2149	1.8282
SEPA	14	0.1890	0.2080	1.7814
SEISF	31	0.1868	0.2091	1.8109
PDI	17	0.1936	0.2084	1.7914

El análisis de parentesco revela un grado significativamente menor a lo esperado bajo apareamiento aleatorio en el caso de los individuos provenientes de BA ($r = -0.022$; $p = 0.001$), y un elevado y significativo grado de parentesco dentro de

las localidades de SCAV ($r = 0.026$; $p=0.001$; Fig. 15). Estos resultados indican que el apareamiento entre individuos de las localidades muestreadas no es aleatorio.

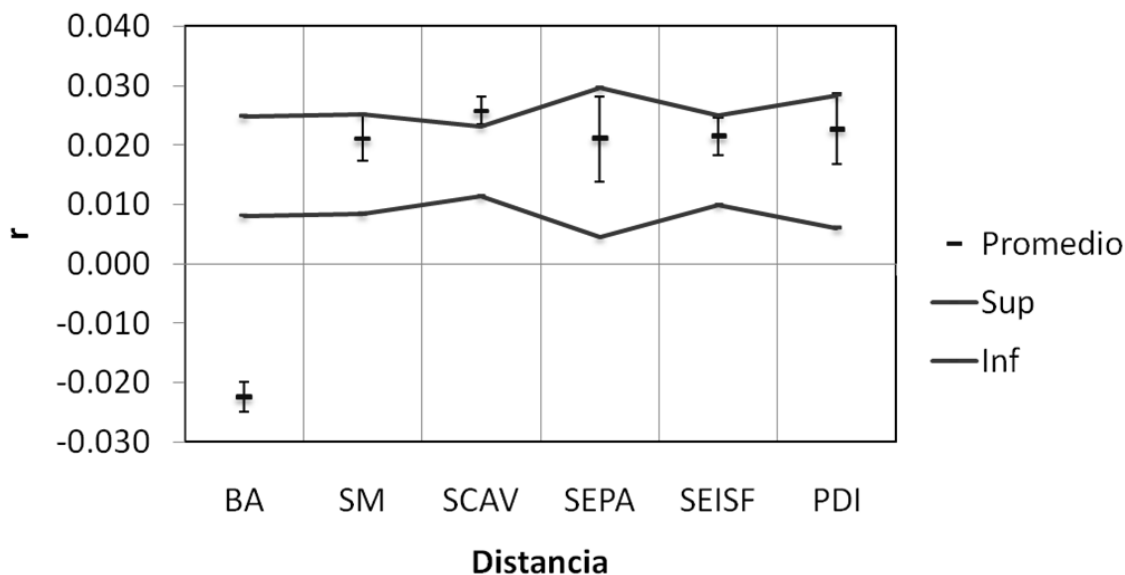


Figura 15. Parentesco promedio (r , en %) de *Lythrypnus dalli* entre sitios muestreados para un subconjunto de 721 SNPs neutrales y polimórficos dentro de todas las localidades. Las barras de error indican los intervalos de confianza al 95% para los valores esperados y las líneas continuas indican los intervalos de confianza al 95% esperado bajo la hipótesis nula de apareamiento aleatorio. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

7.6 Estructura genómica poblacional neutral y adaptativa

Para identificar qué tipo de estructura genómica neutral y adaptativa presenta el góbido *Lythrypnus dalli* en la costa occidental del GC, se estimaron en primera instancia tres indicadores de estructura genética: el índice de fijación (G_{st}), el índice de diferenciación alélica (D_{est}) y la distancia genética condicional (cGD) (Fig. 16). Los niveles de diferenciación resultaron mayores para los SNPs *outliers* que para los SNPs neutrales en todos los casos, siendo esto esperado dado que los análisis de detección de *outliers* se enfocan precisamente a identificar aquellos SNPs que se desvían de lo esperado bajo un escenario de evolución neutral. Para

evaluar si las diferencias observadas entre localidades son significativas, el G_{st} se estimó mediante una prueba de permutación con 10 mil réplicas, no encontrándose diferencias significativas entre localidades para los marcadores neutrales, mientras que todas las comparaciones de los SNPs *outliers*, con la excepción de SEPA-PDI al Sur del GC, resultaron significativas ($p < 0.05$). Estos resultados no nos permiten rechazar la hipótesis nula de estructura genómica neutral panmíctica, mientras que sí muestran evidencias de variación adaptativa que pudiera estar relacionada a adaptaciones locales mediante procesos selectivos.

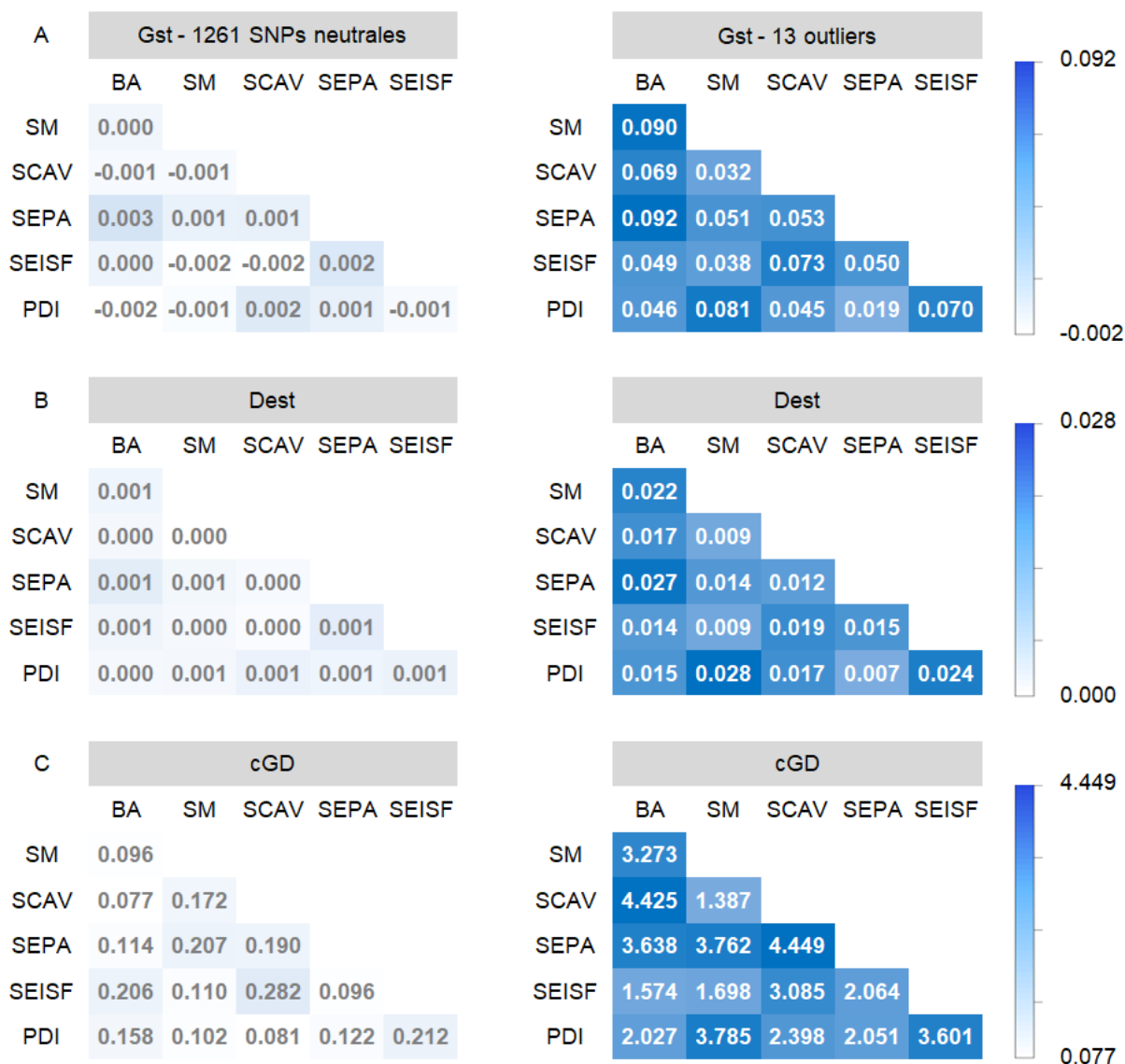


Figura 16. Estimadores de estructura genética para *Lythrypnus dalli*. Índices de Fijación (A), diferenciación alélica (B), y distancia genética condicional (C) obtenido para los SNPs neutrales y *outliers*. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

Cabe destacar que, en el caso de G_{st} y D_{est} la información que aportan es similar, dado que poseen una correlación de Pearson de 0.74 en los SNPs neutrales, y de 0.9 entre indicadores derivados de *outliers*. En cambio, los niveles de cGD sí difieren bastante en términos de estructura, obteniéndose niveles de correlación

de 0.31 con G_{st} y 0.17 con D_{est} para SNPs neutrales, y correlaciones de 0.50 y 0.64 con G_{st} y D_{est} adaptativos, respectivamente. Al calcular la cGD en su versión topográfica, se retienen 8 de los 15 vínculos posibles entre pares de sitios, descartándose 7 conexiones que no aportan información a la explicación de la varianza del sistema (Fig. 17). San Marcos (SM) y SEISF presentaron el mayor nivel de diversidad genética, seguidos por BA y SCAV, siendo SEPA y PDI las localidades de menor diversidad genética, en consistencia con lo reportado en términos de riqueza alélica (Tabla VII).

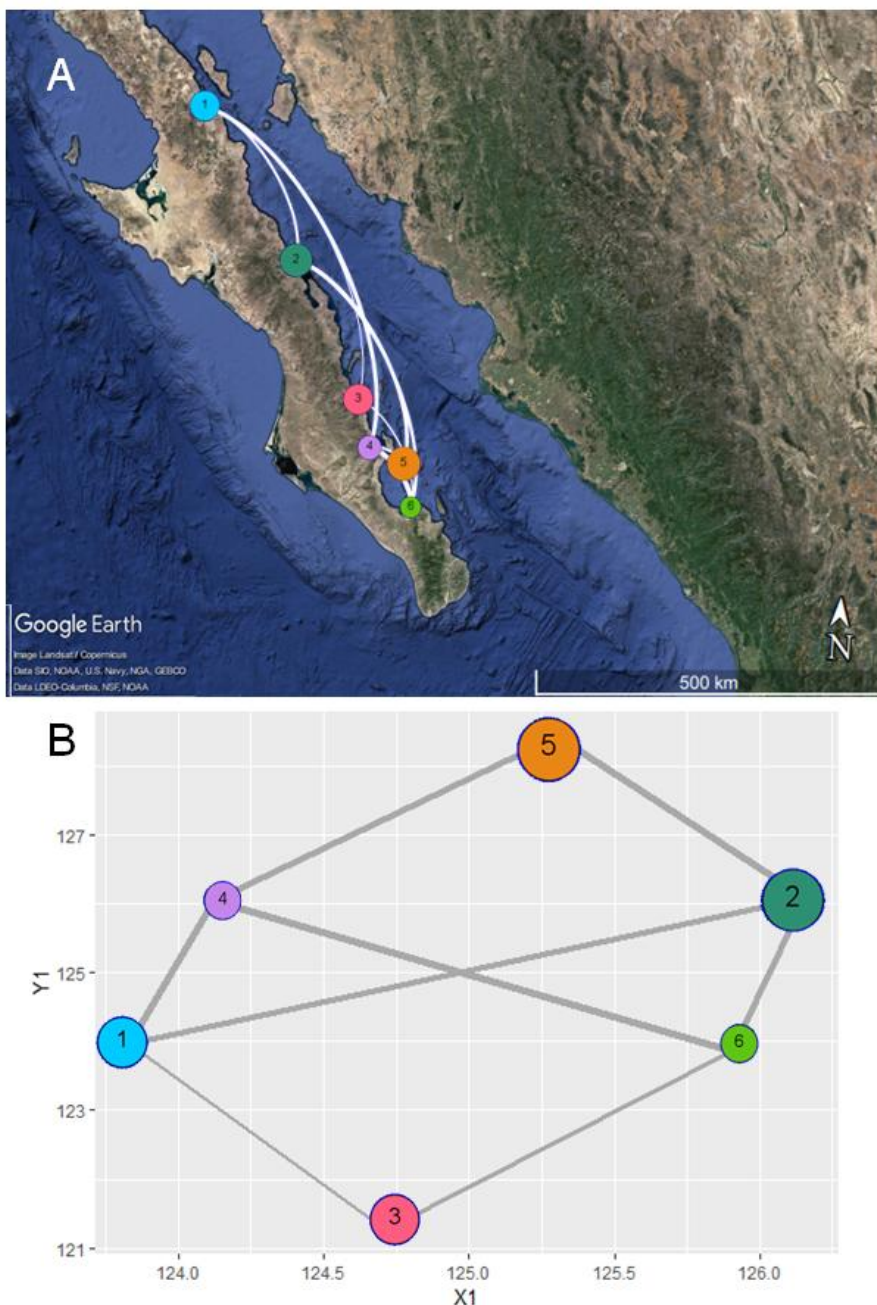


Figura 17. Estructura genómica neutral mediante distancia genética condicional de *L. dalli*. (A) En el espacio geográfico, y (B) visualizado mediante el algoritmo de Fruchterman Reingold, que busca el equilibrio entre las fuerzas de atracción y repulsión entre nodos. El tamaño de nodos (localidades) representa la variación de la diversidad genética dentro de cada sitio y el ancho de los vínculos entre sitios es proporcional a la covarianza genética significativa. Bahía de los Ángeles (1), Isla San Marcos (2), San Cosme y Agua Verde (3), Isla El Pardito (4), Isla San Francisquito (5) y Punta Diablo (6).

Analizando los vínculos entre sitios, es interesante destacar que tanto SEISF como las localidades de SCAV poseen niveles elevados de diversidad genética manteniendo sólo dos conexiones significativas cada una, de distinta magnitud. Más aún, los vínculos que enlazan a SCAV con el resto del sistema son relativamente débiles y no incluyen a las localidades más cercanas, por lo que es posible que la influencia del flujo génico en la mantención de los niveles de diversidad genética en esta localidad no sean iguales que en el resto del Golfo. La estimación del *betweenness centrality* (BC) revela un patrón claramente latitudinal, en donde los sitios de mayor relevancia en la conectividad de la red son BA (BC = 3) y SM (BC = 3), seguidos por SCAV (BC = 1) y SEISF (BC = 1). Las localidades de SEPA y PDI no participan en ninguna *shortest path distance* entre sitios (BC = 0). Estos resultados sugieren una estructura genética metapoblacional.

Una estrategia complementaria para identificar el número más probable de poblaciones fue la aplicación de análisis discriminantes de componentes principales (dAPCs, por sus siglas en inglés), para comparar entre marcadores neutrales y *outliers* agrupados por localidad (Fig. 18), o para comparar entre marcadores neutrales agrupados por localidad o a partir del número más probable de grupos según la función *find_clusters* y posterior aplicación del BIC para los SNPs neutrales (Fig. 19).

Los dAPC realizados por localidad se realizaron a partir de la retención de 80 componentes principales (72% de la varianza retenida) en el caso de los SNPs neutrales (Fig. 18A), y 6 componentes principales (70% de la varianza retenida) para resumir la variación de los 13 *outliers* (Fig. 18B), de acuerdo al método de validación cruzada. Las figuras muestran un agrupamiento de los sitios en un único grupo en ambos casos, siendo este resultado consistente con el G_{st} para los SNPs neutrales, pero contradictorio en cuanto a la diferenciación de *outliers* (Fig. 16).

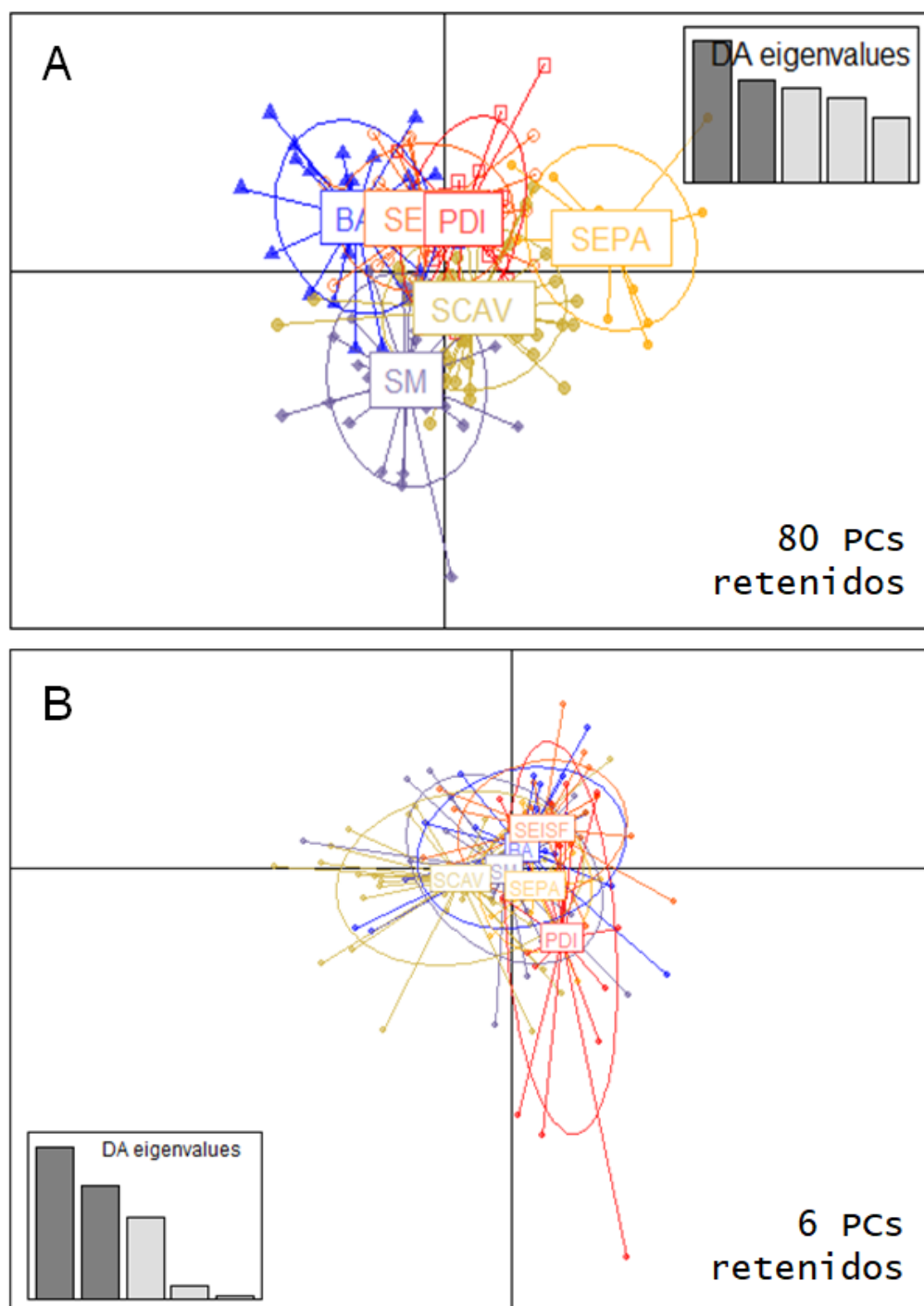


Figura 18. dAPCs por localidad de *Lythrypnus dalli*. (A) Para 1,261 marcadores neutrales y (B) 13 *outliers*. PCs: componentes principales. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

El número más probable de poblaciones identificado por la función *find_clusters* de acuerdo al criterio bayesiano de información depende del porcentaje de varianza retenido (Tabla VIII).

Tabla VIII. Número más probable de grupos (k) de acuerdo al número de componentes principales (PCs) retenidos y porcentaje de varianza retenida.

PCs retenidos	Var. explicada %	k más probable
10	18	6
20	28	4
30	38	3
40	45	2
50	53	2
60	62	1
70	68	1
80	72	1
90	78	1
100	82	1

A partir de la retención de 60 componentes principales (PCs) en adelante, el análisis devuelve la presencia de un único grupo, en consistencia con el dAPC realizado por localidad (Fig. 18A). Sin embargo, la retención de entre 40 y 50 PCs, representando aproximadamente un 50% de la varianza, revela la presencia de dos grupos (Tabla VIII, Fig. 19B). Esta agrupación no puede ser explicada por el sitio de origen de las muestras, ya que todas las localidades poseen individuos asignados a ambos clusters. Tampoco se detectaron diferencias entre tallas (19.87 ± 4.80 vs. 19.42 ± 4.56 , media \pm DE), por lo que es poco probable que el agrupamiento resulte de distintas cohortes de individuos. Por último, se comparó el porcentaje de datos faltantes promedio de los individuos en cada grupo y no se encontraron diferencias (6.40 ± 3.31 vs. 6.07 ± 2.50 , media \pm DE).

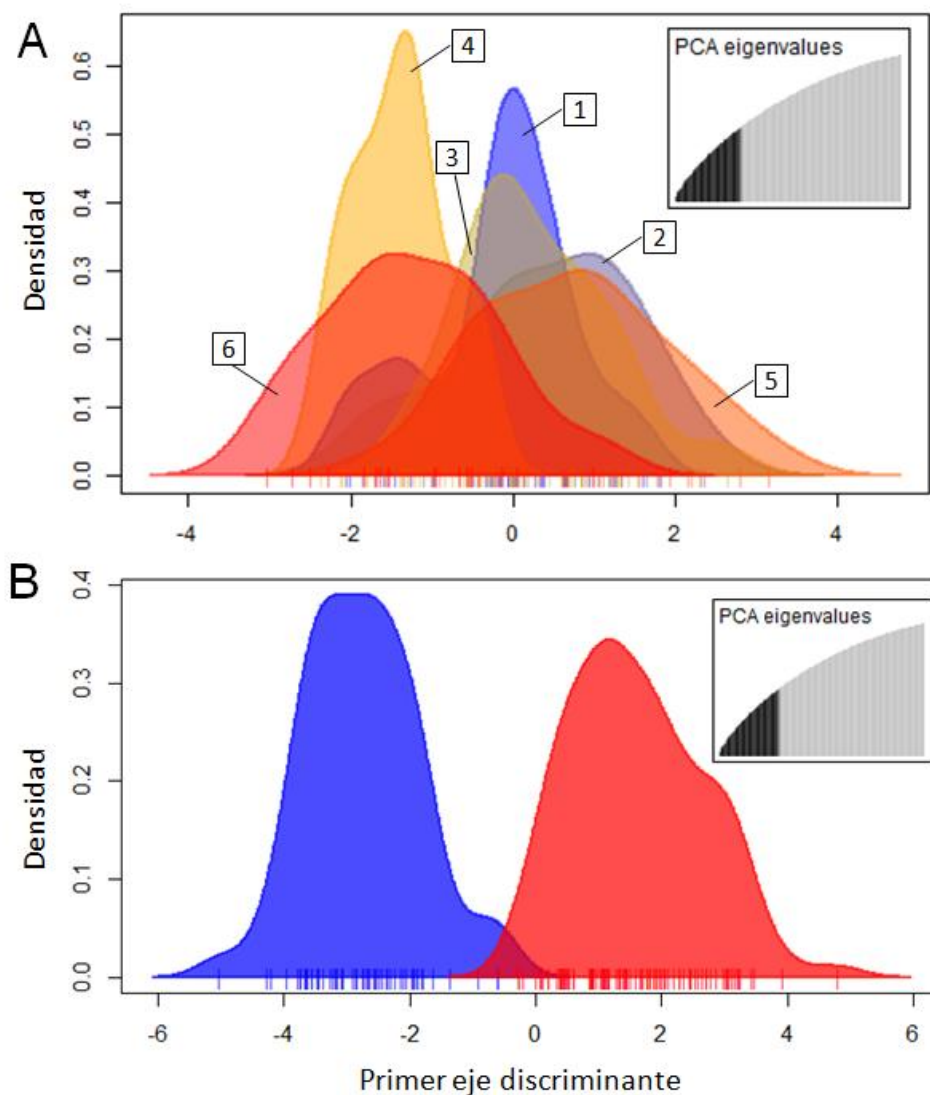


Figura 19. dAPCs de 1,261 SNPs neutrales de *Lythrypnus dalli*. (A) Por localidad y (B) por agrupamiento de acuerdo a *find_clusters*. Densidad de asignación de individuos a sus grupos sobre el primer eje discriminante de acuerdo a la retención de 45 componentes principales en ambos casos. Bahía de los Ángeles (1), Isla San Marcos (2), San Cosme y Agua Verde (3), Isla El Pardito (4), Isla San Francisquito (5) y Punta Diablo (6).

Finalmente, se determinó la estructura espacial y el número más probable de poblaciones mediante un análisis Bayesiano de clusters con el programa STRUCTURE para comparar resultados (Fig. 20).

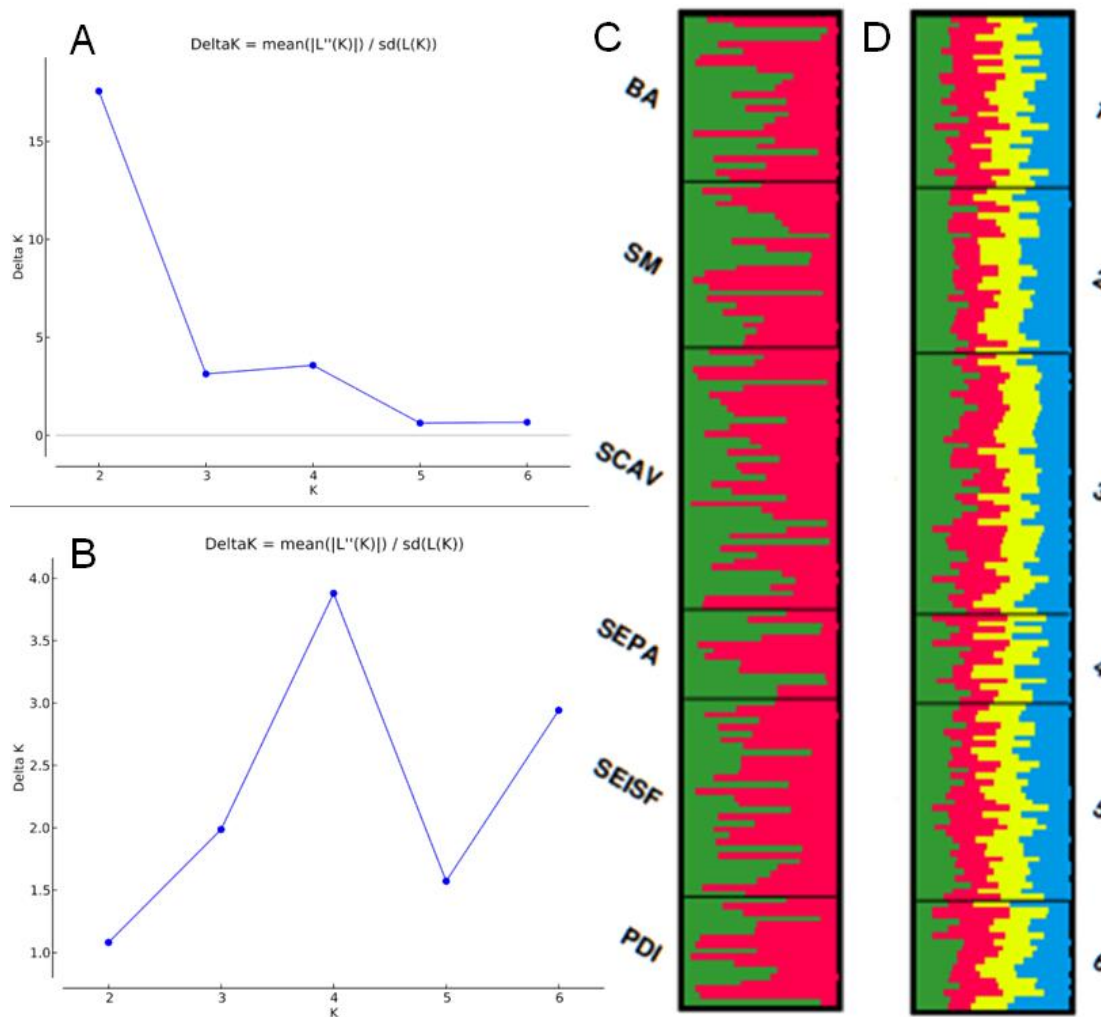


Figura 20. Número más probable de poblaciones de *Lythrypnus dalli* de acuerdo al método de Evanno. Para (A) 1,261 SNPs neutrales y (B) 13 *outliers*. Gráfica de probabilidad de pertenencia de individuos a las poblaciones en base a (C) los SNPs neutrales y (D) *outliers*. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

Los resultados del STRUCTURE son aparentemente consistentes con lo encontrado en los dAPCs. El número más probable de poblaciones según el método de Evanno *et al.* (2005) devuelve dos agrupaciones (Fig. 20A). Este método ha sido criticado por ser propenso a la identificación de $k = 2$ y es técnicamente incapaz de comparar el cambio de verosimilitud entre $k = 0$ y $k = 1$ para permitir la comparación (Janes *et al.*, 2017). La probabilidad posterior sobre

la presencia de un único grupo en este análisis tiene mayor soporte que el $k = 2$ sugerido por Evanno. Sin embargo, la proporción de pertenencia de los individuos a uno u otro grupo es elevada en la mayoría de los casos (Fig. 20C) y la asignación de individuos a uno u otro grupo, cuando STRUCTURE supone $k = 2$, coincide con el dAPC por cluster (Fig. 19b) sobre 151 individuos a uno u otro grupo, mientras que solo 5 individuos fueron discordantes entre métodos.

Tanto STRUCTURE como ambos dAPCs por localidad (Fig. 18A y 19A) resultaron incapaces de discernir entre una y dos poblaciones, como es esperable dados los bajos niveles de diferenciación que se detectaron entre marcadores neutrales (Fig. 16). No obstante, la proporción de varianza explicada por la agrupación en el dAPC (~50%) y la coincidencia de dos análisis de agrupamiento basados en metodologías diferentes disminuye la probabilidad de que el resultado fuera un artefacto en ambos casos, y pudiera en realidad deberse a algún factor causal desconocido.

El análisis a partir de los 13 *outliers* sugiere la presencia de 4 agrupamientos (Fig. 20B), aunque la pertenencia de los individuos a cada grupo no presenta ningún patrón espacial, y tampoco coincide con los resultados de los índices (Fig. 16), ni del dAPC (Fig. 18B). Posiblemente, la cantidad de SNPs utilizados para este análisis y su calidad (ver sección 7.8.) no son suficientes para arribar a un resultado robusto.

7.7 Aislamiento por distancia – SNPs neutrales

No se detectó una relación entre la distancia genética neutral (G_{st} , D_{est} , cGD) y la distancia geográfica euclidiana de acuerdo a las pruebas de Mantel (Tabla IX).

Tabla IX. Pruebas de Mantel de aislamiento por distancia geográfica euclidiana (Dist. Geo.) u oceanográfica (probabilidad de conexión larvaria, Ha_i) en relación a la distancia genética neutral de *Lythrypnus dalli*. Grado de asociación (Mantel r), valor de p, temporada de movimiento larval y duración (PLD) de las hipótesis.

Predictor	Temporada	PLD (sem.)	G_{st}		D_{est}		cGD	
			Mantel r	p	Mantel r	p	Mantel r	p
Dist. Geo.	-	-	-0.308	0.911	-0.023	0.586	-0.171	0.707
Ha1	may:jun	8	0.428	°0.032	0.026	0.464	0.359	0.130
Ha2	jun:jul	8	-0.339	0.897	0.044	0.407	-0.099	0.593
Ha3	may:jun	6	0.428	°0.030	0.026	0.462	0.359	0.127
Ha4	jun:jul	6	-0.212	0.827	-0.137	0.661	-0.045	0.431
Ha5	may	4	-0.348	0.889	-0.155	0.711	-0.521	0.956
Ha6	jun	4	-0.262	0.873	0.209	0.342	-0.123	0.480
Ha7	jul	4	-0.383	0.962	-0.235	0.814	-0.287	0.733
Ha8	abr:jun	8	-0.065	0.553	-0.044	0.534	-0.011	0.471
Ha9	abr:jun	6	0.197	0.213	0.497	*0.042	0.545	0.099
Ha10	abr:jun	4	0.080	0.367	0.320	0.106	0.020	0.434
Ha11	dic:sep	8	0.172	0.246	-0.210	0.757	-0.050	0.575
Ha12	dic:sep	6	-0.120	0.676	0.117	0.410	-0.226	0.762
Ha13	dic:sep	4	0.014	0.420	-0.400	0.956	-0.339	0.891

Notas: * $p < 0.05$; ° No cumple el supuesto de homocedasticidad, resultado inválido.

En cuanto a las 13 hipótesis de conectividad oceanográfica, se detectó una única correlación significativa ($r = 0.497$, $p = 0.042$, Tabla IX y Fig. 21) entre el D_{est} y la hipótesis de conectividad número 9 (Ha_9), que comprende una temporada de migración larvaria durante los meses de abril, mayo y junio, y una PLD de 6 semanas. La misma hipótesis resultó marginalmente significativa también en su relación con la distancia genética condicional ($r = 0.545$, $p = 0.099$).

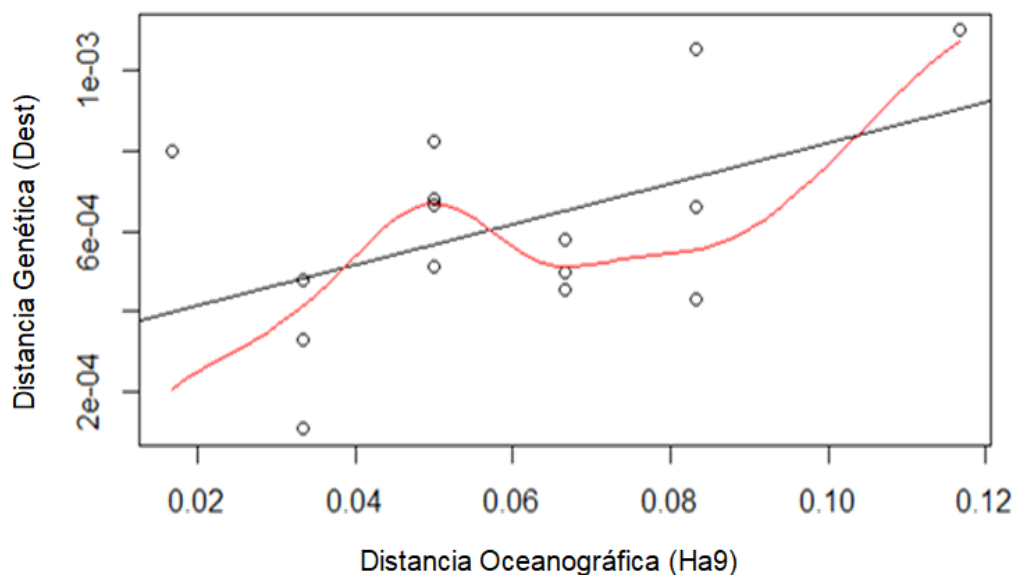


Figura 21. Prueba de Mantel ($r = 0.497$, $p=0.042$) entre el índice de diferenciación alélica (D_{est}) y distancia oceanográfica (Ha9) estimada para *Lythrypnus dalli* a partir de 1,261 SNPs neutrales. La línea negra representa la asociación de Mantel, y la línea roja muestra la curva suavizada de variación de los datos.

En la versión bidireccional de Ha9 (Fig. 22), en donde cada celda representa la probabilidad de conexión entre los 59 sitios de liberación de partículas (Fig. 6) durante la simulación de dispersión por flujo de corrientes, la diagonal representa la probabilidad de retención local y las líneas negras separan las regiones Norte, Central y Sur del Golfo en orden latitudinal. Allí predomina un flujo asimétrico desde el norte hacia el sur (mayor proporción de celdas amarillas por arriba de la diagonal, Fig. 22), con probabilidades de conexión directa entre pares de sitios de hasta un 33%. El nivel de conectividad potencial en el Golfo Norte es mayor que el resto, y hay muy poca exportación de localidades sureñas hacia el norte. El Golfo central resulta intermedio, mostrando conexiones hacia ambas regiones.

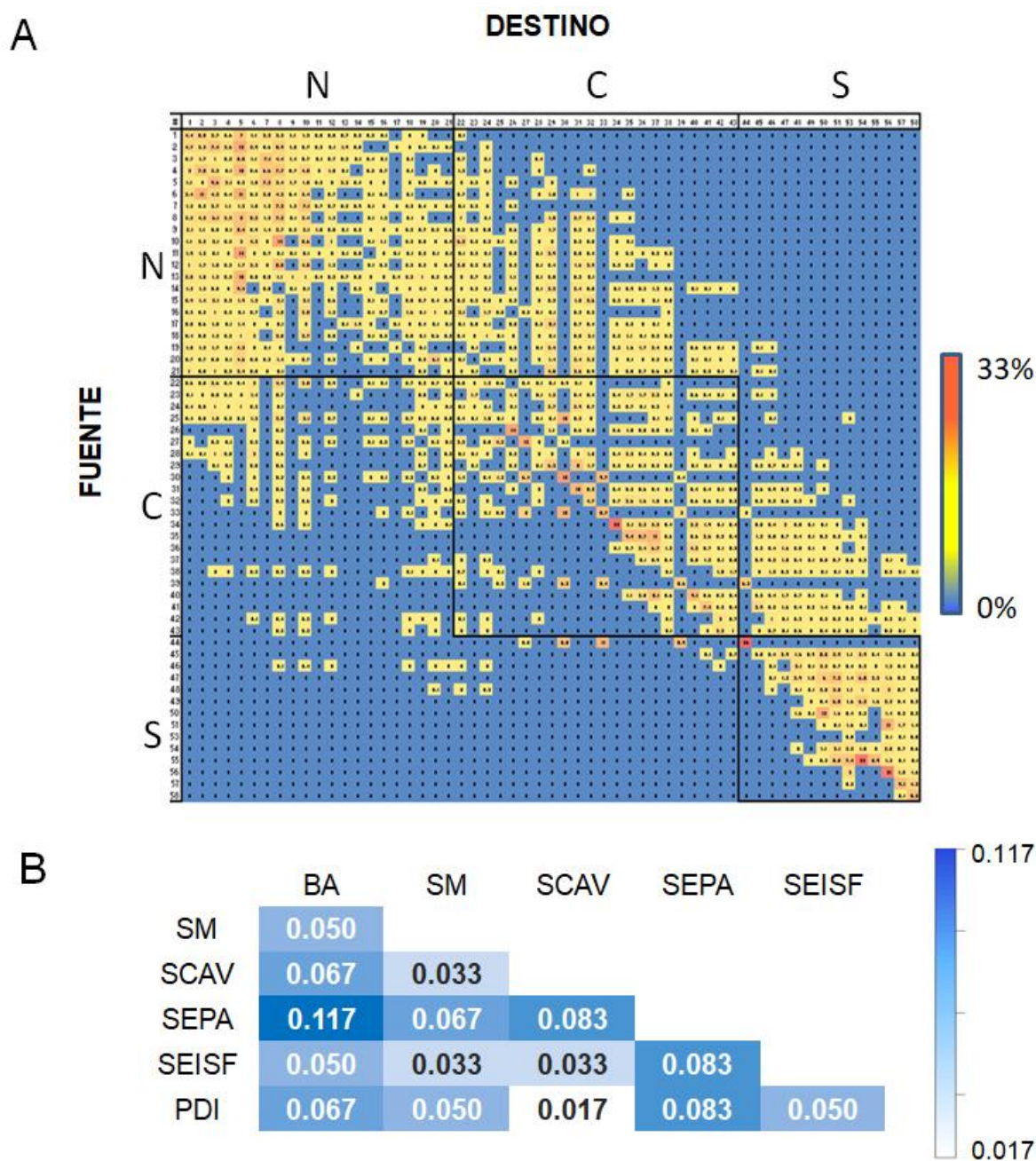


Figura 22. Hipótesis oceanográfica Ha9 de dispersión larvaria. En versión fuente-destino (A), el color de las celdas indica la probabilidad de movimiento desde cada sitio de liberación (fuente) hacia su destino, ambos en orden latitudinal. Retención local en la diagonal y las regiones del GC están marcadas en negrita y señaladas en ambos ejes (N: norte, C: centro, S: sur). (B) Distancias oceanográficas de los sitios de muestreo estandarizadas mediante la aplicación de la teoría de *population graphs*.

Dentro de la red, las localidades norteñas de BA y SM estarían funcionando como fuentes netas de larvas, mientras que el resto de las localidades hacia el sur presentan características de sumidero, observándose un claro gradiente norte-sur en el balance de importación y exportación entre sitios (Tabla X).

Tabla X. Características topológicas de Ha9 por localidad; retención local, número de sitios (de 59 totales) hacia los que se exporta (n.exporta), número de sitios desde los cuales se importa (n.importa) y balance entre ellos (Rol.Fue.Sum.). BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

Localidad	Retención local (%)	n.exporta	n.importa	Rol.Fue.Sum.
BA	0.70	28	9	19
SM	6.30	25	22	3
SCAV	9.07	15	16	-1
SEPA	1.20	11	13	-2
SEISF	0.08	13	18	-5
PDI	1.48	8	17	-9

Posteriormente, se investigó la relación entre variables derivadas de análisis genéticos con predictores derivados de Ha9 mediante regresiones lineales. Se detectó una relación lineal fuerte entre el *betweenness centrality* genético y el número de sitios de exportación ($r^2_{aj} = 0.954$, $p=0.000$, Fig. 23A), y en menor medida una relación con el rol fuente-sumidero, debido a que los dos predictores no son independientes entre sí ($r^2 = 0.89$ entre n.exporta y Rol.Fue.Sum.).

Los valores de retención local, en cambio, no presentan una relación lineal con los niveles de retención local derivados de Ha9 ($r^2_{aj} = -0.067$, $p=0.454$, Fig. 23B). No obstante, en BA coincide un nivel muy bajo de parentesco promedio con un bajo nivel de retención local, y en SCAV coinciden el máximo nivel de parentesco promedio con el máximo grado de retención local.

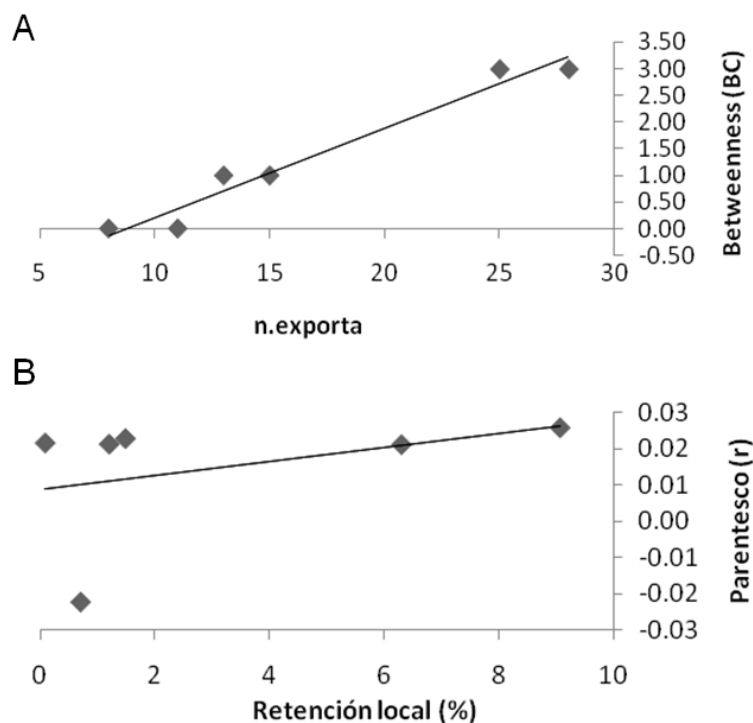


Figura 23. Relación entre variables genéticas e hipótesis de conectividad Ha9. (A) Regresión lineal ($r^2_{aj} = 0.954$, $p=0.000$) de *Betweenness centrality* (BC) en base a 1,261 SNPs neutrales de *Lythrypnus dalli* contra el número de sitios hacia los que cada localidad exporta (n.exporta) derivado de Ha9. (B) Regresión lineal ($r^2_{aj} = -0.067$, $p=0.454$) entre parentesco promedio (proporción) en base a 721 SNPs neutrales y polimórficos dentro de todas las localidades y porcentaje de retención local derivado de Ha9.

Tampoco se detectó una relación significativa entre la variación de niveles de diversidad genética (H_{obs}) y el grado de retención local ($r^2_{aj} = 0.026$, $p=0.347$) o rol como fuente o sumidero de los sitios ($r^2_{aj} = -0.067$, $p=0.713$).

7.8 Aislamiento por ambiente

Los análisis de Bayescan y Arlequin detectaron 1 y 13 SNPs, respectivamente, siendo las diferencias entre sitios significativas en cuanto al G_{st} . Sin embargo, al analizar las frecuencias alélicas se observa que, con excepción de unos pocos casos (por ejemplo: S0618, S0889), en la mayoría de los SNPs no hay variación

sustancial entre sitios, dado que es el mismo alelo a través de las localidades quien se encuentra cercano a la fijación (Fig. 24).

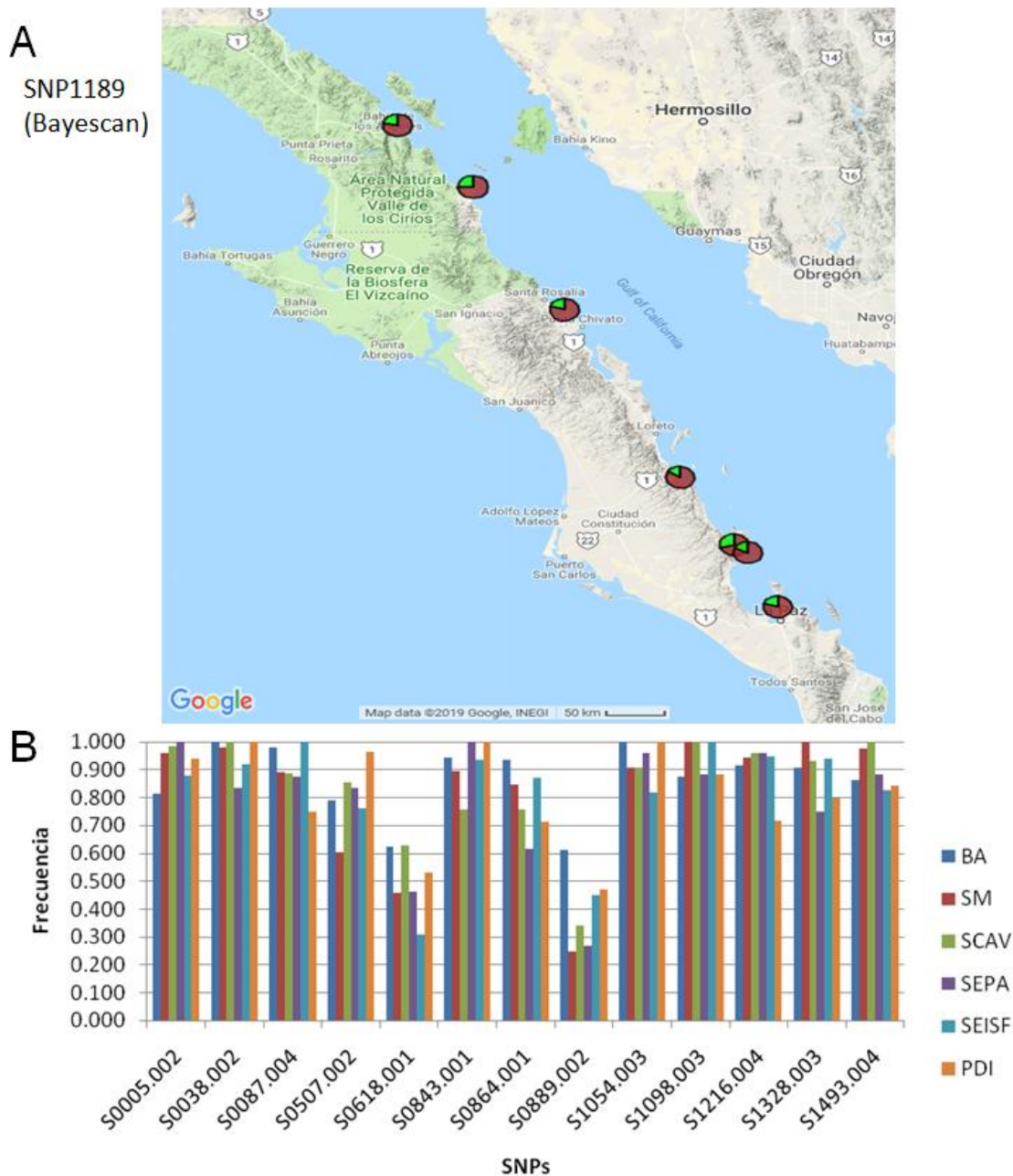


Figura 24. Frecuencias alélicas del único SNP identificado para *L. dalli* por Bayescan (A) y de uno de los alelos de cada uno de los 13 SNPs identificados por Arlequin (B). BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

Si mediante efectos selectivos se estuvieran favoreciendo distintas variantes alélicas asociadas a las condiciones locales, es poco probable que aquellas regiones del genoma afectadas por la presión selectiva hayan sido capturadas en el presente estudio.

Con respecto a los predictores, las variables climatológicas derivadas del análisis de imágenes satelitales se caracterizan, en cuanto a temperatura superficial del mar (SST), por presentar un gradiente latitudinal tanto en términos de promedios como en varianza. La temperatura media mensual aumenta de norte a sur, registrándose los valores más altos en los meses de julio a septiembre; la variabilidad disminuye de norte a sur (Fig. 25A, Tabla XI). En cuanto a los niveles de clorofila-a (Chla), también se observan diferencias espacio-temporales entre los sitios muestreados, con una tendencia general decreciente de norte a sur (Fig. 25B, Tabla XI).

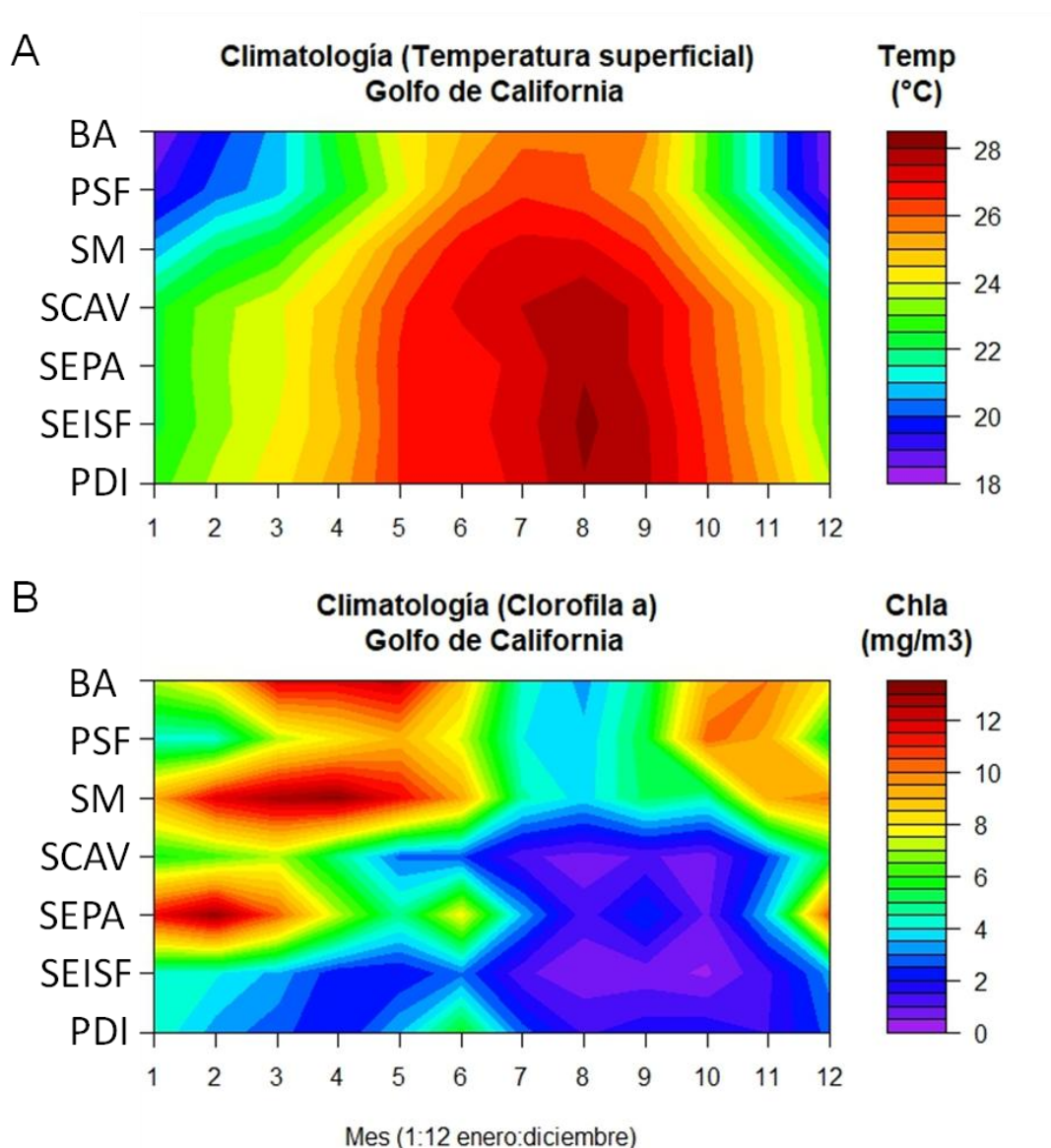


Figura 25. Variables climatológicas derivadas de imágenes del satélite aqua MODIS. (A) Temperatura superficial del mar y (B) clorofila-a. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

A partir de las variables climatológicas (promedios mensuales por sitio) se estimaron las variables resumen a utilizar como predictores en la prueba de hipótesis: promedio, máximo promedio, mínimo promedio y rango de sst y Chla (Tabla XI).

Tabla XI. Variables resumen de climatologías históricas utilizadas en la detección de *outliers* mediante análisis de redundancia, y en las pruebas de hipótesis de aislamiento por ambiente. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

	Localidad	BA	SM	SCAV	SEPA	SEISF	PDI
SST (°C)	media	22.51	24.24	25.34	25.37	25.43	25.60
	DE	2.82	2.47	2.00	1.90	1.95	1.72
	rango	7.44	6.59	5.56	5.58	5.78	5.21
	min	18.50	20.62	22.34	22.37	22.35	22.77
	max	25.94	27.21	27.90	27.95	28.13	27.97
Chla (mg/m3)	media	8.34	8.80	3.42	6.44	2.18	2.80
	DE	2.98	3.43	2.34	4.27	1.36	1.29
	Rango	9.02	9.42	6.37	12.26	3.83	4.35
	min	3.30	3.78	0.74	0.89	0.42	1.44
	max	12.32	13.20	7.10	13.15	4.25	5.79
ACP(SST-Chla)	PC1	3.47	1.90	-1.97	-0.41	-2.86	-2.83
	PC2	0.55	-1.00	-0.07	-2.98	1.20	0.75

Debido a los elevados niveles de multicolinealidad encontrados, se realizó una transformación de variables mediante análisis de componentes principales (Fig. 26). Las variables utilizadas en este análisis fueron: promedio, máximo, mínimo y rango de sst y Chla, respectivamente.

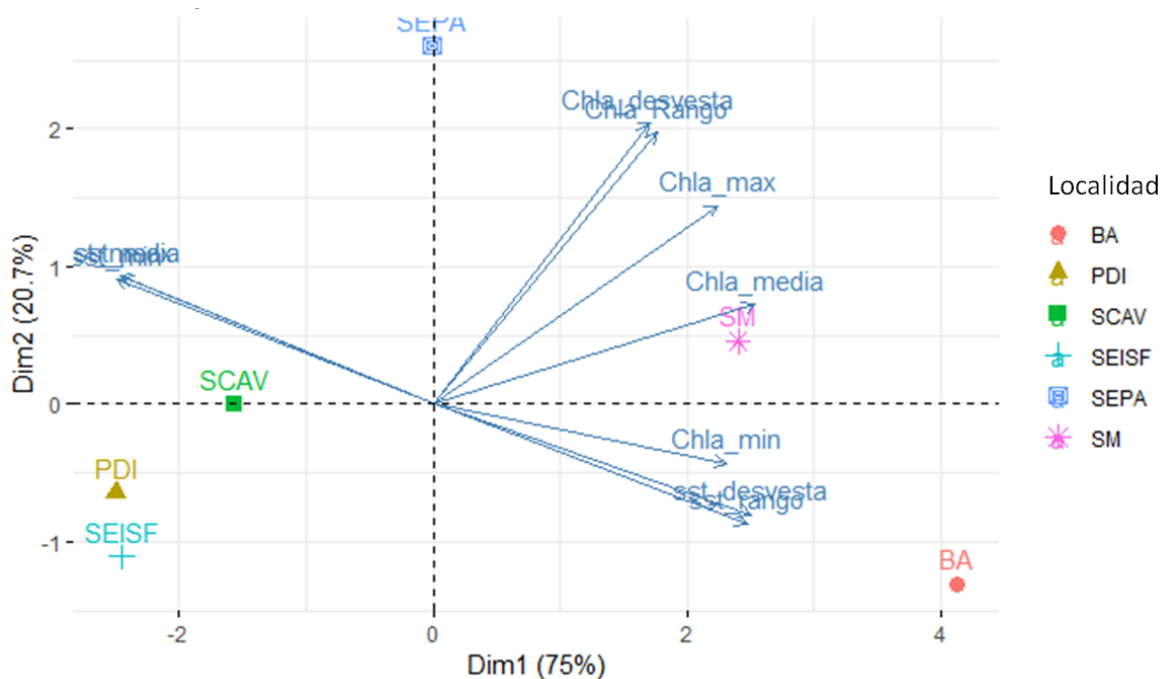


Figura 26. Análisis de componentes principales de indicadores de variación de temperatura superficial (sst) y Clorofila a (Chla). BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

El primer componente principal (75% de la varianza) separa a las localidades del norte de las localidades del sur, aunque no fue posible identificar una variable con mayor peso en la clasificación, ya que los valores de los eigenvectores del primer componente son similares entre sitios. El segundo componente, en cambio, tiene una mayor influencia de la varianza de clorofila-a. Las puntuaciones de las localidades en los dos primeros ejes principales fueron incluidos como predictores en las pruebas de Mantel por su característica de independencia, y porque en conjunto explican el 95.7% de la varianza.

La comparación entre las matrices de G_{st} , D_{est} y cGD de los SNPs *outliers* y la distancia ambiental mediante pruebas de Mantel no presentaron ninguna relación significativa (Tabla XII).

Tabla XII. Aislamiento por ambiente mediante pruebas de Mantel en base a 13 SNPs *outliers* de *Lythrypnus dalli*.

Predictor	G_{st}		D_{est}		cGD	
	Mantel r	p	Mantel r	p	Mantel r	p
PC1	-0.13	0.67	-0.12	0.66	-0.44	0.98
PC2	-0.08	0.62	-0.12	0.69	0.00	0.48
Chla.media	-0.41	0.89	-0.38	0.86	-0.47	0.98
Chla.max	-0.51	0.97	-0.47	0.95	-0.44	0.96
Chla.min	-0.10	0.67	-0.10	0.59	-0.35	0.97
Chla.rango	-0.53	0.98	-0.47	0.97	-0.29	0.88
sst.media	0.27	0.12	0.21	0.16	-0.28	0.83
sst.max	0.30	0.11	0.20	0.16	-0.32	0.90
sst.min	0.25	0.12	0.21	0.15	-0.26	0.82
sst.rango	0.24	0.16	0.27	0.17	-0.20	0.76

Por otro lado, a partir del conjunto de SNPs obtenidos para identificar de candidatos a selección natural (1,711 SNPs del grupo 2) se realizaron dos análisis de redundancia (RDA) utilizando como predictores un subconjunto de variables ambientales (alias “pred₁”, Fig. 27) y los primeros 2 componentes principales de la transformación (alias “pred₂”, Fig. 28).

En el primer caso se seleccionó la temperatura máxima (sst_max), rango de clorofila-a (Chla_Rango) y clorofila-a mínima (Chla_min) de acuerdo al criterio de inflación de la varianza (VIF 2.83, 1.16 y 2.95 respectivamente). Ambos modelos resultaron globalmente significativos (pred₁, $r^2_{aj} = 0.179$, $p < 0.05$; pred₂, $r^2_{aj} = 0.348$, $p < 0.05$). La prueba de permutación para identificar qué ejes canónicos del modelo son significativos identificó únicamente al primer eje (RDA1) en ambos casos (pred₁ y pred₂, $p < 0.05$), por lo que se utilizaron para identificar aquellos SNPs cuya variación estuviera fuertemente relacionada a ellos.

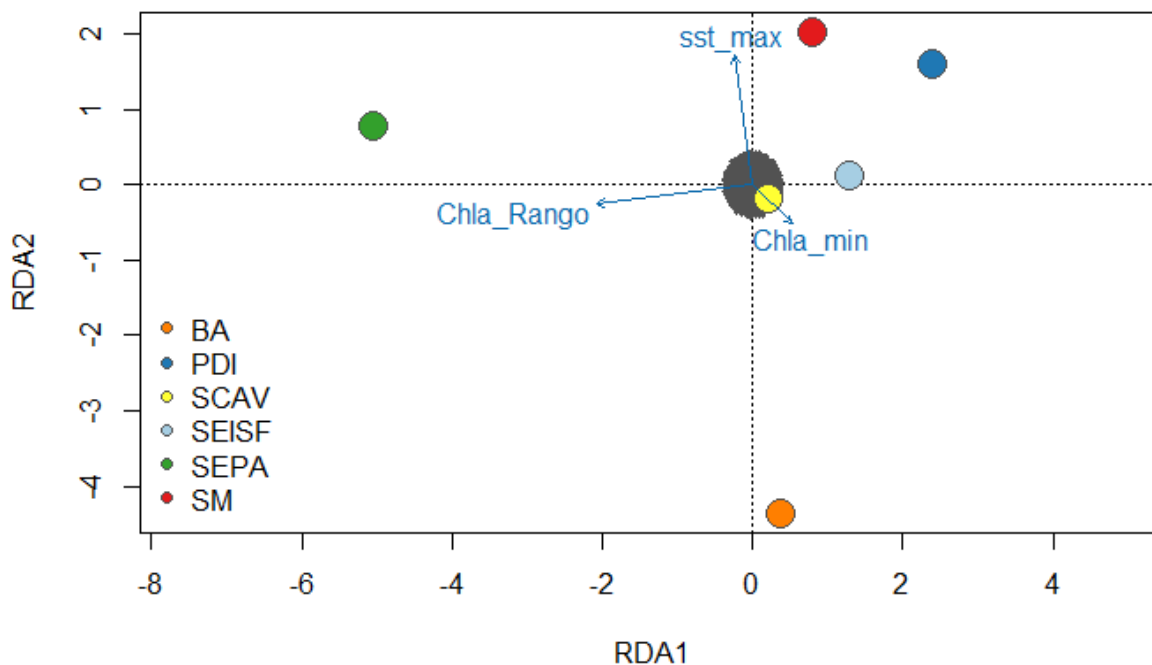


Figura 27. Gráfica triplot del modelo pred₁ con escalamiento de 1,711 SNPs (grupo 1, color gris) por la raíz cuadrada de los eigenvalores (*scaling 1*). Los predictores se muestran en azul y las localidades en colores. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

En ambas gráficas, la posición de las localidades representa su relación con respecto al modelo. Las variables ambientales de cada modelo son representadas mediante vectores azules, cuya longitud es proporcional a la magnitud de su contribución al modelo (el ángulo entre ellas no representa su correlación en este escalamiento), y cuya dirección apunta hacia valores altos (por ejemplo, SEPA posee elevado rango de Chla y PDI tiene bajo rango de Chla, Fig. 27). La posición de los SNPs representa la dirección de variación en las frecuencias alélicas en relación al modelo. Al ser un número elevado de puntos (1,711 SNPs), se forman nubes de puntos alrededor del origen y no se distingue en la gráfica separación o gradiente alguno, tal como ha ocurrido en otros trabajos que utilizan este método (Capblancq *et al.*, 2018; Forester *et al.*, 2018).

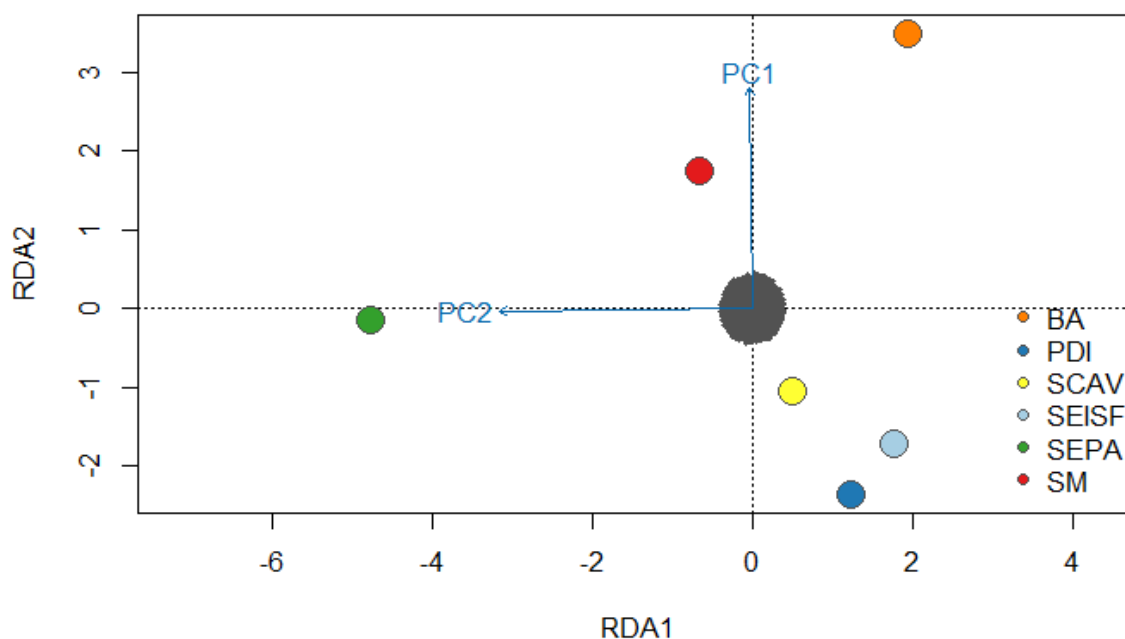


Figura 28. Gráfica triplot del modelo $pred_2$ con escalamiento de 1,711 SNPs (grupo 1, color gris) por la raíz cuadrada de los eigenvalores (*scaling 1*). Los predictores se muestran en azul y las localidades en colores. BA: Bahía de los Ángeles, SM: Isla San Marcos, SCAV: San Cosme y Agua Verde, SEPA: Isla El Pardito, SEISF: Isla San Francisquito y PDI: Punta Diablo.

Sin embargo, en ambos modelos la aplicación de la función para la detección de *outliers* presentes en las colas de la distribución de *scores* en los ejes canónicos, no identificó ningún marcador fuertemente relacionado aplicando el criterio laxo de corte de 2.0 DE de la media (equivalente a un $p=0.05$ de dos colas). Estos resultados van acorde con los resultados derivados de la detección de *outliers*, ya que los SNPs identificados no presentaron en su mayoría variación espacial, ni se detectaron asociaciones mediante las pruebas de Mantel (Tabla XII).

8. DISCUSIÓN

Los niveles sutiles de estructura genética neutral y adaptativa detectados para *Lytrhypnus dalli* en el GC (GC) coinciden con lo esperado para muchas especies marinas. Con respecto a la estructura neutral, los indicadores clásicos de estructura genética (G_{st} , dAPC por localidad y STRUCTURE) indican panmixia. No obstante, la distancia genética condicional reveló los primeros indicios de variación espacial, reteniendo una topología en donde la presencia e intensidad de vínculos entre localidades (proporcional a su variación conjunta) no es homogénea. La incorporación de hipótesis de conectividad oceanográfica que modelan el flujo de partículas con las corrientes del GC, más allá de las localidades muestreadas, ha aportado información a partir de la cual patrones genéticos espaciales aparentemente caóticos adquieren sentido biológico.

En primer lugar, la única relación significativa se da entre el D_{est} y una hipótesis de conectividad (Ha9) cuyos parámetros biológicos coinciden con literatura previa en cuanto al pico de reclutamiento y duración de movimiento larval (Fig. 21). En segundo lugar, se detectó una relación lineal fuerte entre el *betweenness centrality*, un indicador de la relevancia de cada localidad en la conectividad (covariación) genética, y el número de sitios de exportación o el rol como fuente o sumidero de cada localidad de acuerdo con Ha9 (Fig. 23A). En tercer lugar, el análisis de parentesco reveló diferencias significativas en las localidades de BA y SCAV, y dichos niveles tienen sentido a la luz del porcentaje de retención local, aunque la relación no sea linealmente significativa (Fig. 23B).

En conjunto, los resultados sugieren la presencia de una estructura metapoblacional con niveles muy bajos de diferenciación genética, aunque las poblaciones locales no han sido identificadas. El modelo de conectividad oceanográfica sugiere que el flujo predominante de corrientes asociado a los giros ciclónicos de verano transporta migrantes desde las localidades norteñas hacia el centro y sur del GC en la costa peninsular. Sin embargo, la estructuración metapoblacional sugerida no explica la variación en diversidad genética, ni el

agrupamiento de individuos en dos clusters presentes en todas las localidades, responsables de una proporción sustancial de la varianza genética (Fig. 19).

Con respecto a los patrones y determinantes de la variación adaptativa, los pocos candidatos a selección identificados en el presente estudio no presentaron una relación significativa con las variables del paisaje utilizadas, posiblemente debido a la baja cobertura genómica obtenida con la combinación de enzimas utilizadas para la digestión del ADN durante la elaboración de librerías, o a la baja cobertura geográfica con respecto al rango de distribución de *L. dalli*. Los principales hallazgos y limitaciones se discuten en mayor detalle a continuación.

8.1 Muestreo y librerías genómicas

En el presente estudio, mediante la obtención de librerías genómicas de baja representación, se pudieron comparar hasta 1,711 marcadores moleculares obtenidos de 156 individuos en 6 localidades de la costa occidental del GC. Una de las limitantes del estudio se relaciona con la amplitud de muestreo ya que, no pudieron colectarse organismos de la especie en los sitios visitados al norte de BA, ni al sur de PDI, teniendo como consecuencia una pobre representación de las regiones norte y sur del Golfo. Un mayor esfuerzo de muestreo en estas regiones, en conjunto con la incorporación de individuos provenientes de las costas continentales del Golfo, serían de gran utilidad para elucidar los patrones de diversidad y estructura genómica en la región. A su vez, el aumento en la cobertura geográfica permitiría abarcar una mayor variabilidad ambiental y complejidad oceanográfica al estudio.

La calidad de las secuencias obtenidas mediante la técnica de ddRAD (Peterson *et al.*, 2012) resultó muy elevada, siendo este un indicador de la confiabilidad de los genotipos obtenidos para análisis posteriores. La parametrización en STACKS (Catchen *et al.*, 2013) para la obtención de RAD *loci* y SNPs también permitió identificar la combinación óptima de parámetros de acuerdo al objetivo planteado

para *Lythrypnus dalli* en el GC, con profundidades de lectura por encima de 100X por *locus* y por individuo. Este nivel de profundidad de lectura, de hecho, está muy por encima del nivel habitual de otros trabajos de RADseq (10X, Lal *et al.*, 2016; 5X, Manthey y Moyle, 2015; 5X, Zhang *et al.*, 2016) y por encima del nivel mínimo de 25X recomendado por Paris *et al.* (2017). Por ende, la inclusión de al menos el doble de individuos con el mismo esfuerzo de secuenciación sería viable para *Lythrypnus dalli* en las condiciones que se realizó este trabajo.

Fountain *et al.* (2016) estimaron el error de genotipificación de acuerdo a distintos niveles de cobertura utilizando ddRAD y encontraron para ensamblajes *de novo* un error de genotipificación mayor al 11% y 4% para profundidades de al menos 5X y 30X, respectivamente. La estimación del error de genotipificación del 1% en promedio entonces resulta esperable para la profundidad de lecturas obtenida en el presente estudio. El establecimiento de un límite para la frecuencia del alelo menor (*maf*) permite descartar SNPs resultantes de errores durante la elaboración de librerías y genotipificación. Valores frecuentes de *maf* rondan entre 0.01 y 0.05 (Brauer *et al.*, 2016; Lal *et al.*, 2016b; Wang *et al.*, 2016). Al disminuir la frecuencia del alelo menor (*maf*) de 0.05 (seleccionado durante la parametrización de STACKS) a 0.03, con la confianza de estar por encima del error de genotipificación, resultó en entre 7% y 13% de datos faltantes para los SNPs neutrales y adaptativos, respectivamente, habiendo partido de 18% en promedio durante la parametrización.

El número de individuos secuenciados y de SNPs obtenidos se encuentran en el orden de otros estudios genómicos (Flanagan *et al.*, 2016; Lal *et al.*, 2016a; Sandoval-Castillo *et al.*, 2018; Van Wyngaarden *et al.*, 2016).

8.2 Estructura genómica neutral

Los bajos niveles de diferenciación obtenidos en el presente estudio coinciden con muchas otras especies marinas en el mundo (DiBattista *et al.*, 2017; Lah *et al.*,

2016; Therkildsen *et al.*, 2013) y en el GC (Cisneros-Mata *et al.*, 2019; Jackson *et al.*, 2015; Paz-García *et al.*, 2012; Soria *et al.*, 2012; Vázquez Rojas, 2018;), incluso en distintas escalas espacio-temporales y linajes evolutivos independientes, dando soporte a las hipótesis de fuerzas extrínsecas actuando como determinantes de la diversidad y estructura genética en los océanos. Los índices de fijación y diferenciación alélica por si solos indican panmixia. La distancia genética condicional tiene la ventaja de representar en el espacio geográfico o multivariado la partición de la varianza (o diversidad) genética entre estratos independientemente de su magnitud (Dyer, 2015), siendo de gran utilidad en especies marinas. Es notable mencionar que de 15 vínculos posibles en una red de 6 sitios, la mitad resultó redundante en términos de la varianza explicada. Más aún, el estadístico de *betweenness centrality* reveló diferencias entre la relevancia de las localidades en la covarianza del sistema.

La presencia de un único grupo panmítico también es soportada por el dAPC por localidad y el software STRUCTURE, en consistencia con el G_{st} . Panmixia ha sido el paradigma para las especies marinas por décadas, dados los elevados tamaños efectivos poblacionales y niveles de conectividad que ocurren comúnmente en este medio. No obstante, la función *find_clusters* y posterior dAPC en ausencia de información sobre la procedencia de las muestras identifica dos grupos, siendo esta separación soportada por cerca del 50% de la varianza. La coincidencia entre la asignación de individuos entre el dAPC por grupo y STRUCTURE para $k = 2$, y el porcentaje de varianza asociado sugiere que hay fuerzas actuando sobre la diversidad genética que no han podido ser identificadas en el presente estudio.

No se encontró una relación entre los niveles de diversidad genética y los predictores utilizados en el presente estudio. Sin embargo, los factores que inciden sobre la diversidad genética son numerosos, destacándose el tamaño efectivo poblacional, que no ha sido estimado. Adicionalmente, a pesar de haber sido calculada mediante el método de rarefacción, los niveles de riqueza alélica continuaron mostrando una relación fuerte con el tamaño muestral, por lo que su

interpretación debe ser tomada con cautela. El análisis de parentesco, en cambio, sí reveló diferencias significativas entre localidades. En BA, el nivel resultó significativamente inferior al resto de las localidades, incluso a lo esperado bajo apareamiento aleatorio, y también se detectaron niveles significativos de endogamia en las localidades de SCAV, restando soporte a la hipótesis de una única población panmíctica.

8.3 Aislamiento por distancia

A partir de la estimación de distancias euclidianas y de 13 hipótesis de conectividad oceanográfica generadas, la única relación significativa detectada mediante pruebas de Mantel se dio entre el D_{est} y el flujo pasivo de larvas estimado para un PLD y temporada reproductiva que coinciden con lo reportado para *L. dalli* en la bibliografía. Esta hipótesis presenta en promedio, para la estación primaveral (abril-junio) y un PLD de 6 semanas (42 días), un elevado nivel de conectividad entre los sitios del Golfo norte, debido a la presencia de un giro ciclónico y una corriente con dirección noroeste en la costa continental del Golfo norte, y un flujo predominante hacia el sur en la costa peninsular del Golfo central y sureño consistente con la presencia de un segundo giro ciclónico en el Golfo central.

Existe una creciente acumulación de evidencias del sistema de corrientes del GC como determinante de patrones de estructura y diversidad genética en especies de peces como la cabrilla sardinera *Mycteroperca rosacea* (Jackson *et al.*, 2015; Munguia-Vega *et al.*, 2014), la jaiba *Callinectes bellicosus* (Cisneros-Mata *et al.*, 2019) y la almeja burra *Spondylus calcifer* (Soria *et al.*, 2012), con temporada reproductiva veraniega, y también para la almeja de sifón *Panopea globosa* (Dall 1898) relacionada al giro anticiclónico invernal del Golfo Norte (Vázquez Rojas, 2018). Los estudios mencionados se concentran en las regiones norte y central del GC, con énfasis alrededor de las grandes Islas, mientras que los trabajos que

incluyen representantes de la región sur son escasos. Por ejemplo, Munguia-Vega *et al.* (2018a) recientemente explicaron niveles de diversidad y estructura genética con las dinámicas de dispersión larvaria del huachinango *Lutjanus peru*, con una temporada reproductiva durante el verano y un PLD de 28 días. Paz-García *et al.* (2012) también detectaron niveles de flujo génico moderado a elevado entre los corales *Pocillopora damicornis* (Linnaeus, 1758) y *Porites panamensis* (Verrill, 1866) dentro y fuera del GC, abarcando las tres regiones biogeográficas, y cuyo patrón espacial también adjudican al flujo de corrientes y la presencia de giros.

Recientemente, Cisneros-Mata *et al.* (2019) publicaron un trabajo de diversidad y estructura genética de la jaiba azul *Callinectes bellicosus* en las costas de Sonora del GC (Golfo Norte), con resultados muy similares a los del presente estudio. Los autores identifican una estructura metapoblacional del tipo fuente-sumidero siendo las localidades sureñas quienes actúan como fuentes, ya que en la costa occidental los giros del Golfo Norte y Central fluyen en dirección opuesta a la costa peninsular. El mismo análisis en *C. bellicosus* revela un grado menor de parentesco entre individuos de la localidad más norteña de su estudio, Puerto Peñasco, hecho que atribuyen a su comportamiento como sumidero de larvas de distintas localidades. En el presente estudio, en cambio, es BA la localidad identificada como fuente de migrantes, quien tiene un grado de parentesco menor al resto.

Dos posibles explicaciones surgen de esta contradicción. Por un lado, es posible que BA actúe como fuente de migrantes hacia el sur de la península, pero que ella misma sea sumidero de larvas provenientes de la costa continental, y de la región de las grandes Islas. Sánchez-Velasco *et al.* (2009) identificaron a *L. dalli* como una de las especies dominantes de ensambles de larvas de peces costeros demersales a principios del verano en el Golfo Norte.

Por otro lado, es posible que la región adyacente a BA, con distintos tamaños de Islas, Islotes, Cuencas, Bahías y demás variación de hábitats (conocida como

región de grandes Islas), sumado a la riqueza de especies y nutrientes que caracteriza a la región, provea un escenario favorable para el desarrollo de la especie a escala macro-evolutiva. De hecho, la región de las grandes islas se ha propuesto como un área estable durante el Pleistoceno tardío en donde poblaciones demográficamente estables de la cabrilla sardinera *M. rosacea* habrían actuado como fuentes de nuevas poblaciones durante períodos de variabilidad climática y grandes fluctuaciones en el nivel del mar (Jackson *et al.*, 2015).

En cuanto a los niveles significativos de endogamia reportados para SCAV, también es posible que la presencia y distribución de islas e islotes en la zona estén favoreciendo la retención local de larvas, generando a través de auto-reclutamiento el patrón observado. De hecho, aquí además se da una menor relación en cuanto a número e intensidad de vínculos con el resto de las localidades, de acuerdo a la distancia genética condicional (Fig. 17), por lo que podría suponerse que una mayor proporción de individuos locales estuviera contribuyendo al acervo genético de la localidad.

Considerando los principales hallazgos del presente estudio, se han encontrado evidencias que sugieren panmixia: la prueba de permutación de G_{st} , dAPC por localidad y la verosimilitud de STRUUCTURE para $k = 1$. No obstante, la presencia de apareamiento no aleatorio, la identificación de dos posibles grupos que explican la mitad de la varianza global (cuyos determinantes no han podido ser identificados) y las evidencias de aislamiento por distancia detectadas no dan soporte a una estructura panmíctica. De acuerdo al concepto de metapoblación de Sale *et al.* (2006) y al continuo de diferenciación poblacional posible entre aislamiento y panmixia descrito por Waples y Gaggiotti (2006), es probable para *Lythrypnus dalli* en la costa occidental del GC la presencia de poblaciones locales o sub-poblaciones con conectividad sustancial, suficiente para mantener flujo génico y un elevado grado de homogeneidad genética, en conjunto con dinámicas

de fuente-sumidero y un grado variable de auto-reclutamiento y aislamiento reproductivo.

Estas observaciones y el creciente caudal de evidencias acumuladas enfatizan la importancia de incorporar estudios de migración y flujo génico en el GC, un sistema en donde existe elevada asimetría y variabilidad espacio-temporal en términos de conectividad, y por ende el diseño de estrategias de manejo deben considerar sitios dentro del GC que actúen como exportadores de migrantes: para proteger la biodiversidad de un sitio en particular probablemente sea necesario identificar y aplicar estrategias en lugares “río arriba” del sitio de interés, entre otros factores (Munguia-Vega *et al.*, 2018b). A su vez, la posibilidad de predecir o validar PLD y temporadas reproductivas mediante el ajuste entre patrones de dispersión larvaria estacional y datos empíricos genéticos sobre un marco teórico metapoblacional podría resultar de suma utilidad en otras especies, en donde la información sobre temporada y duración reproductiva sea conflictiva o inexistente (Munguia-Vega *et al.*, 2018a).

Sin embargo, es necesario aclarar que factores como la distribución de hábitat, el comportamiento larval (migración vertical, escape a las corrientes), traslape generacional, eventos climáticos extremos como huracanes o “El Niño”, y otros factores biológicos y metodológicos pueden afectar la estructura genética espacial (Selkoe *et al.*, 2016) y no han sido considerados en el presente estudio. Para elucidar los mecanismos causales del patrón observado en el gobio *L. dalli*, futuros estudios se beneficiarían de la incorporación de muestras provenientes de la costa continental, de una mejor representación de la región de las grandes Islas que separan el norte y centro del GC, e idealmente de la incorporación de muestras a lo largo del rango de distribución de la especie y a través del tiempo. A su vez, incorporar una mejor representación del paisaje, en conjunto con un mayor número de marcadores moleculares con una mayor cobertura del genoma, podrían hacer un aporte al entendimiento de la evolución de *L. dalli* a escala

macro y micro-evolutiva, permitiendo así predecir futuras respuestas de la especie (y de especies similares) ante los pronósticos del cambio climático.

8.4 Detección de *outliers* y aislamiento por ambiente

La búsqueda de marcadores relacionados a selección natural direccional se ha establecido en los últimos años debido a su utilidad para entender mecanismos de adaptación, y por su relevancia para el manejo y conservación de recursos naturales (Schoville *et al.*, 2012). Para minimizar la identificación de falsos positivos e incorporar el efecto del método de identificación sobre los SNPs detectados, se realizaron tres análisis de detección de *outliers*, dos de ellos derivados de métodos de diferenciación poblacional (Bayescan y Arlequin), y un análisis de asociación genético-ambiental (RDA) (Ahrens *et al.*, 2018; Forester *et al.*, 2018; Narum y Hess, 2011; Rellstab *et al.*, 2015;). No hubo SNPs coincidentes entre métodos, y tanto Bayescan como RDA detectaron uno o ningún marcador candidato a selección. Arlequín, a pesar de ser uno de los métodos más populares de detección de *outliers* (Ahrens *et al.*, 2018), ha sido identificado mediante simulaciones como uno de los métodos con mayor incidencia de falsos positivos y negativos (Narum y Hess, 2011). Este método identificó 27 SNPs *outliers* de los cuales 14 se encuentran en HWE (por lo que fueron considerados falsos positivos), y de los 13 marcadores restantes, la mayoría posee frecuencias alélicas espacialmente similares (Fig. 24), siendo poco probable la influencia de procesos selectivos actuando sobre ellos. La ausencia de asociaciones significativas a partir de este conjunto de *outliers*, por lo tanto, no resulta sorprendente. Por otro lado, los análisis de redundancia sí encontraron una relación entre la variación genómica global estimada a partir de 1,711 SNPs filtrados (para maximizar la probabilidad de detectar marcadores candidatos a selección) y los predictores derivados de temperatura superficial y clorofila a, pero en esta relación no se identificaron marcadores candidatos a selección bajo el criterio más laxo usado en la literatura (Forester *et al.*, 2018).

Es posible que la cobertura genómica obtenida en este estudio no haya sido suficiente como para capturar SNPs bajo efectos selectivos, debido a su efecto *locus* específico sobre el ADN. El porcentaje genómico representado en este estudio es muy difícil de estimar, debido a que no se cuenta con un genoma de referencia filogenéticamente cercano para hacer la comparación. El genoma más cercano disponible se encuentra a un orden de distancia de los Perciformes, el de *Dichotomyctere nigroviridis*, una especie de pez globo del orden Tetraodontiformes con un genoma relativamente pequeño (aproximadamente 340 Mb). La digestión *in silico* del genoma de *D. nigroviridis* obtenido de la base de datos genómicos *ensemble* (ensamble a nivel cromosómico, www.ensemble.org) con las mismas enzimas de restricción usadas en el presente estudio (EcoRI y MspI) y posterior selección de fragmentos de 300±50 pb, devolvió un total de 5822 posibles RAD *loci*. El número obtenido aquí fue considerablemente menor, de 2571 RAD *loci*, a partir de los cuales se obtuvieron 1,711 SNPs para la detección de patrones de variación adaptativa. En general, la proporción de SNPs *outliers* o asociados a la variación ambiental detectada en organismos marinos para este tipo de estudios ronda el 0.1-3% (Manthey y Moyle, 2015; Van Wyngaarden *et al.*, 2016; Xuereb *et al.*, 2018; Zhang *et al.*, 2016), aunque existen casos extremos (9.4%, Guo *et al.*, 2016), y también casos en donde no se encuentran evidencias de adaptación (Lal *et al.*, 2016a). Cabe aclarar que la falta de evidencias de adaptación puede ser un reflejo de un patrón natural, o puede ser consecuencia de los métodos de secuenciación y análisis seleccionados (Flanagan y Jones, 2018; Hendricks *et al.*, 2018; O'Leary *et al.*, 2018). A partir de una muestra genómica de 1,711 SNPs, se podría esperar encontrar entre 51 (3%) y 17 (1%) candidatos a selección, ambos valores quizás muy pequeños como para haber sido detectados en el presente estudio.

Otra posible explicación a la ausencia de SNPs potencialmente adaptativos es la influencia del flujo génico entre sitios, ya que su efecto homogeneizante incide en la probabilidad de adaptación y su mantenimiento en la escala espacial del

estudio. Tan solo un migrante por generación es suficiente para homogeneizar frecuencias alélicas entre sitios (Mills y Allendorf, 1996), y el gobio *L. dalli* posee una fase larval móvil de 40 días o más, pudiendo recorrer cientos de km con las corrientes oceánicas. Recientemente, Munguía-Vega *et al.* (2018b) propusieron una serie de lineamientos para el diseño de una red de áreas naturales protegidas en el GC. Los autores encuentran que el mismo modelo oceanográfico usado en este estudio puede predecir la distancia recorrida de distintas especies por movimiento larval pasivo, de acuerdo a la duración larval (PLD). Del estudio mencionado, la especie con un PLD similar a *L. dalli* es la jaiba azul *C. bellicosus*, con un PLD de 60 días y una distancia de dispersión estimada entre 65 y 200 km (Tabla S6, material suplementario, Munguía-Vega *et al.*, 2018), suficiente para justificar las conexiones entre sitios del GC.

Para mejorar la probabilidad de detección de variación adaptativa, futuros estudios se enriquecerían en términos de cobertura genómica mediante la selección de una combinación de enzimas de restricción que generen un mayor número de fragmentos. Nuevamente, la ampliación de la cobertura geográfica, incluso por fuera del GC, sería ideal para cubrir una mayor proporción del rango de la especie, y tener así representantes de los extremos de la distribución, en donde posiblemente se encuentren los límites fisiológicos de tolerancia al ambiente, de interés en la búsqueda de selección natural como agente causal de diferenciación.

9. CONCLUSIONES

Los análisis de estructura genética de *Lythrypnus dalli* en el GC indican panmixia. En el presente estudio se han presentado y discutido las evidencias que dan soporte y contradicen la posibilidad de panmixia. La presencia de una estructura metapoblacional en donde sub-poblaciones presentan conectividad sustancial provee un sentido biológico al elevado nivel de homogeneidad genética, en conjunto con las dinámicas de fuente-sumidero y el grado variable de auto-reclutamiento y aislamiento reproductivo detectado, aunque no se han podido delimitar poblaciones locales.

La estructura metapoblacional permite explicar los resultados de los análisis genéticos que indican panmixia y también explica la detección de aislamiento por distancia con una hipótesis de conectividad oceanográfica derivada de una temporada y duración de movimiento larval que coincide con estudios pasados y recientes para la especie. El modelo de conectividad oceanográfica también pudo ser relacionado a la contribución de las localidades a la varianza global del sistema, estimada mediante el *betweenness centrality*. Por último, la estructura metapoblacional tiene cierto grado de coincidencia con la detección de diferencias significativas entre los niveles de parentesco promedio entre localidades.

Las sub-poblaciones mencionadas no han podido ser delimitadas en el espacio geográfico ya que, nuevamente, los niveles de diferenciación genética espacial son mínimos. No obstante, los resultados sugieren que las localidades del norte actúan como fuente de migrantes hacia el centro y sur de la costa occidental del GC.

El presente estudio, en cambio, no ha podido explicar los niveles de diversidad genética, ni la posible presencia de dos agrupamientos de individuos independientemente de su procedencia. Tampoco se ha tenido éxito en la detección de marcadores relacionados a procesos adaptativos, imposibilitando determinar la estructura genómica adaptativa y sus determinantes.

10. LITERATURA CITADA

- Adamack, A. T., B. Gruber. 2014. PopGenReport: Simplifying basic population genetic analyses in R. *Methods Ecol. Evol.* 5(4):384-387.
- Agapow, P. M., A. Burt. 2001. Indices of multilocus linkage disequilibrium. *Mol. Ecol. notes.* 1(1):101-102.
- Ahrens, C. W., P. D. Rymer, A. Stow, J. Bragg, S. Dillon, K.D.L. Umbers, R.Y. Dudaniec. 2018. The search for loci under selection: trends, biases and progress. *Mol. Ecol.* 27:1342-1356.
- Allendorf, F. W., P. A. Hohenlohe, G. Luikart. 2010. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* 11(10):697-709.
- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Archambeault, S., E. N. L. Rapp, D. Cerino, B. Bourque, T. Solomon-Lane, M. S. Grober, A. Rhyne, K. Crow. 2016. Reproduction, larviculture and early development of the Bluebanded goby, *Lythrypnus dalli*, an emerging model organism for studies in evolutionary developmental biology and sexual plasticity. *Aquac. Res.* 47(6):1899-1916.
- Backhaus, J. O. 1985. A three-dimensional model for the simulation of shelf sea dynamics. *Deutsche Hydrographische Zeitschrift.* 38(4):165-187.
- Bastian, M., S. Heymann. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Quality and Safety in Health Care.*
- Beaumont, M. A., R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *P. Roy. Soc. Lond. B. Bio.* 263(1377):1619-1626.
- Behrents, K. C. 1983. The comparative ecology and interactions between the sympatric gobies (*Lythrypnus dalli* and *Lythrypnus zebra*). Tesis (PhD in Biology) Los Angeles, California, USA. University of Southern California. 343 p.
- Behrents, K. C. 1987. The influence of shelter availability on recruitment and early juvenile survivorship of *Lythrypnus dalli* Gilbert (Pisces: Gobiidae). *J. Exp. Mar. Biol. Ecol.* 107(1):45-59.
- Behrents, K. C. 1989. The foraging ecology of two sympatric gobiid fishes: importance of behavior in prey type selection. *Environ. Biol. Fish.* 26(2):105-118.

Beldade, R., A. Jackson, R. Cudney-Bueno, P. Raimondi, G. Bernardi. 2014. Genetic structure among spawning aggregations of the gulf coney *Hyporthodus acanthistius*. Mar. Ecol. Prog. Ser. 499:193-201.

Benestan, L., B. K. Quinn, H. Maaroufi, M. Laporte, F. K. Clark, S. J. Greenwood, R. Rochette, L. Bernatchez. 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). Mol. Ecol. 25(20):5073-5092.

Bernardi, G., L. Findley, A. Rocha-Olivares. 2003. Vicariance and Dispersal across Baja California in Disjunct Marine Fish Populations. Evolution. 57(7):1599-1609.

Bernatchez, L. 2016. On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. J. Fish Biol.. 89(6):2519-2556.

Bernatchez, L., M. Wellenreuther, C. Araneda, D. T. Ashton, J. M. I. Barth, T. D. Beacham, G. E. Maes, J. T. Martinsohn, K. M. Miller, K. A. Naish, J. R. Ovenden, C. R. Primmer, H. Young Suk, N. O. Therkildsen, R. E. Withler. 2017. Harnessing the Power of Genomics to Secure the Future of Seafood. Trends Ecol. Evol.. 32(9):665-680.

Black IV, W. C., C. F. Baer, M. F. Antolin, N. M. DuTeau. 2001. Population Genomics: Genome-Wide Sampling of Insect Populations. Annu. Rev. Entomol. 46(1):441-469.

Brauer, C. J., M. P. Hammer, L.B. Beheregaray. 2016. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. Mol. Ecol. 25(20):5093-5113.

Calderón Parra, M. 2009. Análisis comparativo de la biología reproductiva y algunos aspectos de dinámica poblacional de *Lythrypnus dalli* (Gilbert, 1890) y *L. pulchellus* (Ginsburg, 1938) (Perciformes: Gobiidae) en la Bahía de La Paz, B.C.S. Tesis (Maestría en Ciencias en el Uso, Manejo y Preservación de los Recursos Naturales). Baja California Sur, México. Centro de Investigaciones Biológicas del Noroeste, S. C. 80 p.

Capblancq, T., K. Luu, M. G. B. Blum, E. Bazin. 2018. Evaluation of redundancy analysis to identify signatures of local adaptation. Mol. Ecol. Resour. 18(6):1223-1233.

Carr, M. H., C. Syms. 2006. Recruitment. En: Allen, L. G., D. J. Pondella, M. H. Horn (eds.). The Ecology of Marine Fishes: California and Adjacent Waters. University of California Press, Berkeley. First edition. California, USA. pp 411-427.

Catchen, J. M., P. A. Hohenlohe, S. Bassham, A. Amores, W. A. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* 22(11):3124-3140.

Catchen, J. M. 2016. Stacks Manual. Second edition. Illinois, USA. 37p.

Catchen, J. M., P. A. Hohenlohe, L. Bernatchez, W. C. Funk, K. R. Andrews, F. W. Allendorf. 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Mol. Ecol. Resour.* 17:362-365p.

Cisneros-Mata, M. A., A. Munguia-Vega, D. Rodríguez-Félix, E. A. Aragón-Noriega, J. M. Grijalva-Chon, J.A. Arreola-Lizárraga, L.A. Hurtado. 2019. Genetic diversity and metapopulation structure of the brown swimming crab (*Callinectes bellicosus*) along the coast of Sonora, Mexico: Implications for fisheries management. *Fish. Res.* 212:97-106.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin. 2011. The variant call format and VCFtools. *Bioinformatics.* 27(15):2156-2158.

DiBattista, J. D., M. J. Travers, G. I. Moore, R. D. Evans, S. J. Newman, M. Feng, S. D. Moyle, R. J. Gorton, T. Saunders, O. Berry. 2017. Seascape genomics reveals fine-scale patterns of dispersal for a reef fish along the ecologically divergent coast of Northwestern Australia. *Mol. Ecol.* 26(22):6206-6223.

Dray, S., A. B. Dufour. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* 22(4):1-20.

Dyer, R. J. 2015. Population Graphs and Landscape Genetics. *Annu. Rev. Ecol. Evol. S.* 46(1):327-342.

Dyer, R. J. 2016. gstudio: Tools Related to the Spatial Analysis of Genetic Marker Data. R package version 1.5.0.

Dyer, R. J., J. D. Nason. 2004. Population Graphs: The graph theoretic shape of genetic structure. *Mol. Ecol.* 13(7):1713-1727.

Evanno, G., S. Regnaut, J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14(8):2611-2620.

Excoffier, L., H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10(3):564-567.

Falush, D., M. Stephens, J. K. Pritchard. 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. notes.* 7(4):574-578.

Ferchaud, A. L., M. M. Hansen. 2016. The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: Three-spine sticklebacks in divergent environments. *Mol. Ecol.* 25(1):238-259.

Ficetola, G. F., T. W. J. Garner, F. De Bernardi. 2007. Genetic diversity, but not hatching success, is jointly affected by postglacial colonization and isolation in the threatened frog, *Rana latastei*. *Mol. Ecol.* 16(9):1787-1797.

Flanagan, S. P., E. Rose, A. G. Jones. 2016. Population genomics reveals multiple drivers of population differentiation in a sex-role-reversed pipefish. *Mol. Ecol.* 25(20):5043-5072.

Flanagan, S. P., A. G. Jones. 2018. Substantial differences in bias between single-digest and double-digest RAD-seq libraries: A case study. *Mol. Ecol. Resour.* 18(2):264-280.

Foll, M., O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics.* 180(2):977-993.

Forester, B. R., J. R. Lasky, H. Wagner, D. L. Urban. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol. Ecol.* 27(9):2215-2233.

Fountain, E. D., J. N. Pauli, B. N. Reid, P. J. Palsbøll, M. Z. Peery. 2016. Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Mol. Ecol. Resour.* 16(4):966-978.

Gagnaire, P. A., T. Broquet, D. Aurelle, F. Viard, A. Souissi, F. Bonhomme, S. Arnaud-Haond, N. Bierne. 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol. Appl.* 8(8):769-786.

Gilbert, C. H. 1890. Scientific results of explorations by the U.S. Fish Steamer Albatross No. 12. A preliminary report on the fishes collected by the Steamer Albatross on the Pacific coast of North America during the year 1889, with descriptions of twelve new genera. *P. US. Nat. Mus.* 13(12):49-126.

Gonçalves Da Silva, A., S. A. Appleyard, J. Upston. 2015. Establishing the evolutionary compatibility of potential sources of colonizers for overfished stocks: A population genomics approach. *Mol. Ecol.* 24:56-579.

- González-Cabello, A., D. R. Bellwood. 2009. Local ecological impacts of regional biodiversity on reef fish assemblages. *J. Biogeogr.* 36(6):1129-1137.
- Gregor, C. A., T. W. Anderson. 2016. Relative importance of habitat attributes to predation risk in a temperate reef fish. *Environ. Biol. Fish.* 99(6):539-556.
- Guo, B., Z. Li, J. Merila. 2016. Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Mol. Ecol.* 25(12):2833-2852.
- Hand, B. K., W. H. Lowe, R. P. Kovach, C. C. Muhlfeld, G. Luikart. 2015. Landscape community genomics: Understanding eco-evolutionary processes in complex environments. *Trends Ecol. Evol.*, 30:161-168p.
- Hartney, K. B. y K. A. Grorud. 2002. The effect of sea urchins as biogenic structures on the local abundance of a temperate reef fish. *Oecologia.* 131(4):506-513.
- Hedgecock, D., P. H. Barber, S. Edmands. 2003. Genetic Approaches to Measuring Connectivity. *Oceanography.* 20(3):70-79.
- Hedgecock, D., A. I. Pudovkin. 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *B. Mar. Sci.* 87(4):971-1002.
- Hendricks, S., E. C. Anderson, T. Antao, L. Bernatchez, B. R. Forester, B. Garner, B. K. Hand, P. A. Hohenlohe, M. Kardos, B. Koop, A. Sethuraman, R. S. Waples, G. Luikart. 2018. Recent advances in conservation and population genomics data analysis. *Evol. Appl.* 11:1197-1211.
- Hess, J. E., N. R. Campbell, D. A. Close, M. F. Docker, S. R. Narum. 2013. Population genomics of Pacific lamprey: Adaptive variation in a highly dispersive species. *Mol. Ecol.* 22(11):2898-2916.
- Hijmans, R. J. 2018. raster: Geographic Data Analysis and Modeling. R package version 2.8-4.
- Hurtado, L. A., M. Mateos, C. A. Santamaria. 2010. Phylogeography of supralittoral rocky intertidal *Ligia* isopods in the Pacific region from Central California to Central Mexico. *PLoS ONE.* 5(7):1-13.
- Jackson, A. M., A. Munguia-Vega, R. Beldade, B. E. Erisman, G. Bernardi. 2015. Incorporating historical and ecological genetic data for leopard grouper (*Mycteroperca rosacea*) into marine reserve design in the Gulf of California. *Conserv. Genet.* 16(4):811-822.

Jakobsson, M., N. A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 23(14):1801-1806.

Janes, J. K., J. M. Miller, J. R. Dupuis, R. M. Malenfant, J. C. Gorrell, C. I. Cullingham, R. L. Andrew. 2017. The K = 2 conundrum. *Mol. Ecol.* 26(14):3594-3602.

Jombart, T. 2008. Adegnet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*. 24(11):1403-1405.

Jombart, T., I. Ahmed. 2011. Adegnet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*. 27(21):3070-3071.

Jorde, P. E., G. Søvik, J. I. Westgaard, J. Albretsen, C. André, C. Hvingel, T. Johansen, A. D. Sandvik, M. Kingsley, K. E. Jørstad. 2015. Genetically distinct populations of northern shrimp, *Pandalus borealis*, in the North Atlantic: Adaptation to different temperatures as an isolation factor. *Mol. Ecol.* 24(8):1742-1757.

Jost, L. 2008. GST and its relatives do not measure differentiation. *Mol. Ecol.* 17(18):4015-4026.

Jost, L., F. Archer, S. Flanagan, O. Gaggiotti, S. Hoban, E. Latch. 2018. Differentiation measures for conservation genetics. *Evol. Appl.* 11(7):1139-1148.

Kamvar, Z. N., J. C. Brooks, N. J. Grünwald. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6:1-10.

Lah, L., D. Trense, H. Benke, P. Berggren, P. Gunnlaugsson, C. Lockyer, A. Öztürk, B. Öztürk, I. Pawliczka, A. Roos, U. Siebert, K. S. Ra, G. Víkingsson, R. Tiedemann. 2016. Spatially explicit analysis of genome-wide SNPs detects subtle population structure in a mobile marine mammal, the harbor porpoise. *PLoS ONE*. 11(10):1-23.

Lal, M. M., P. C. Southgate, D. R. Jerry, C. Bosserelle, y K. R. Zenger. 2016a. A Parallel Population Genomic and Hydrodynamic Approach to Fishery Management of Highly-Dispersive Marine Invertebrates: The Case of the Fijian Black-Lip Pearl Oyster *Pinctada margaritifera*. *PLoS ONE*. 11(8):1-26.

Lal, M. M., P. C. Southgate, D. R. Jerry, K. R. Zenger. 2016b. Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *Mar. Genom.* 25:57-68.

- Lavin, M. F. y S. G. Marinone. 2003. An Overview of the Physical Oceanography of the Gulf of California. *Nonlinear Processes in Geophysical Fluid Dynamics*. (3):173-204.
- Lin, H. C., C. Sánchez-Ortiz, P. A. Hastings. 2009. Colour variation is incongruent with mitochondrial lineages: Cryptic speciation and subsequent diversification in a Gulf of California reef fish (Teleostei: Blennioidei). *Mol. Ecol.* 18(11):2476-2488.
- Lotterhos, K. E., M.C. Whitlock. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* 24(5):1031-1046.
- Lowe, W. H., R. P. Kovach, F. W. Allendorf. 2017. Population Genetics and Demography Unite Ecology and Evolution. *Trends Ecol. Evol.* 32(2):141-152.
- Lowe, W. H., F. W. Allendorf. 2010. What can genetics tell us about population connectivity?. *Mol. Ecol.* 19(15):3038-3051.
- Lowry, D. B., S. Hoban, J. L. Kelley, K. E. Lotterhos, L. K. Reed, M. F. Antolin, A. Storfer. 2016. Breaking RAD: An evaluation of the utility of restriction site associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* 17:142-152.
- Luikart, G., P. R. England, D. Tallmon, S. Jordan, P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genetics.* 4(12):981-994.
- Manthey, J. D., R. G. Moyle. 2015. Isolation by environment in white-breasted nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: A landscape genomics approach. *Mol. Ecol.* 24(14):3628-3638.
- Marinone, S. G. 2003. A three-dimensional model of the mean and seasonal circulation of the Gulf of California. *J. Geophys. Res.* 108(10):3325.
- Marinone, S. G. 2012. Seasonal surface connectivity in the Gulf of California. *Estuar. Coast. Shelf S.* 100:133-141.
- Matz, M. V., E. A. Treml, G. V. Aglyamova, L. K. Bay. 2018. Potential and limits for rapid genetic adaptation to warming in a Great Barrier Reef coral. *PLoS Genet.* 14(4):1-19.
- Mills, L. S., F. W. Allendorf. 1996. The One-Migrant-per-Generation Rule in Conservation and Management. *Conservation Biology.* 10(6):1509-1518.

Marine Stewardship Council. 2014. Fisheries Standard and Guidance. Marine Stewardship Council Editorial. Version 2.0. London, United Kingdom. 540p.

Munguia-Vega, A., A. Jackson, S. G. Marinone, B. Erisman, M. Moreno-Baez, A. Girón-Nava, T. Pfister, O. Aburto-Oropeza, J. Torre. 2014. Asymmetric connectivity of spawning aggregations of a commercially important marine fish using a multidisciplinary approach. *PeerJ*. 2(511):1-33.

Munguia-Vega, A., S. G. Marinone, D. A. Paz-Garcia, A. Giron-Nava, T. Plomozo-Lugo, O. Gonzalez-Cuellar, A. H. Weaver, F. J. García-Rodríguez, H. Reyes-Bonilla. 2018a. Anisotropic larval connectivity and metapopulation structure driven by directional oceanic currents in a marine fish targeted by small-scale fisheries. *Mar. Biol.* 165(1):1-17.

Munguia-Vega, A., A. L. Green, A. N. Suarez-Castillo, M. J. Espinosa-Romero, O. Aburto-Oropeza, A. M. Cisneros-Montemayor, G. Cruz-Piñón, G. Danemann, A. Giron-Nava, O. Gonzalez-Cuellar, C. Lasch, M. M. Mancha-Cisneros, S. G. Marinone, M. Moreno-Báez, H. Morzaria-Luna, H. Reyes-Bonilla, J. Torre, P. Turk-Boyer, M. Walther, A. Hudson-Weaver. 2018b. Ecological guidelines for designing networks of marine reserves in the unique biophysical environment of the Gulf of California. *Rev. Fish Biol. Fisher.* 28:749-776.

Murphy, M. A., R. Dezzani, D. S. Pilliod, A. Storfer. 2010. Landscape genetics of high mountain frog metapopulations. *Mol. Ecol.* 19(17):3634-3649.

Narum, S. R., J. E. Hess. 2011. Comparison of FST outlier tests for SNP loci under selection. *Mol. Ecol. Resour.* 11(1):184-194.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press. First Edition. New York, USA. 512 p.

Nielsen, E. E., J. Hemmer-Hansen, P. F. Larsen, D. Bekkevold. 2009. Population genomics of marine fishes: Identifying adaptive variation in space and time. *Mol. Ecol.* 18(15):3128-3150.

O'Leary, S. J., J. B. Puritz, S. C. Willis, C. M. Hollenbeck, D. S. Portnoy. 2018. These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Mol. Ecol.* 27(16):1-26.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O. Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, H. Wagner. 2018. Package 'vegan' version 2.5-2.

Ovenden, J. R., O. Berry, D. J. Welch, R. C. Buckworth, C. M. Dichmont. 2015. Ocean's eleven: A critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish Fish.* 16(1):125-159.

Paris, J. R., J. R. Stevens, J. M. Catchen. 2017. Lost in parameter space: A road map for stacks. *Methods Ecol. Evol.* 8(10):1360-1373

Paz-García, D. A., H. E. Chávez-Romo, F. Correa-Sandoval, H. Reyes-Bonilla, A. López-Pérez, P. Medina-Rosas, M. P. Hernández-Cortés. 2012. Genetic Connectivity Patterns of Corals *Pocillopora damicornis* and *Porites panamensis* (Anthozoa: Scleractinia) Along the West Coast of Mexico. *Pac. Sci.* 66(1):43-61.

Peakall, R. P. E. Smouse. 2012. GenALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics.* 28(19):2537-2539.

Petatán Ramírez, D. 2015. Propuesta de zonación del Golfo de California con base en variables oceanográficas y distribución de macroinvertebrados. Tesis (Maestría en Ciencias Marinas y Costeras) La Paz, Baja California Sur, México. Universidad Autónoma de Baja California Sur. 106 p

Peterson, B. K., J.N. Weber, E. H. Kay, H. S. Fisher, H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE.* 7(5):1-11.

Pritchard, J. K., M. Stephens, P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics.* 155(2):945-959.

Queller, D. C., K. F. Goodnight. 1989. Estimating Relatedness Using Genetic Markers. *Evolution.* 43(2):258-275.

R Core team. 2018. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.

Reid, N. M., D. A. Proestou, B. W. Clark, W. C. Warren, J. K. Colbourne, J. R. Shaw, S. I. Karchner, M. E. Hahn, D. Nacci, M. F. Oleksiak, D. L. Crawford, A. Whitehead. 2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science.* 354(6317):1305-1308.

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* 24(17):4348-4370.

Riginos, C. 2005. Cryptic Vicariance in Gulf of California Fishes Parallels Vicariant Patterns Found in Baja California Mammals and Reptiles. *Evolution*. 59(12):2678.

Riginos, C., E. D. Crandall, L. Liggins, P. Bongaerts, E. A. Treml. 2016. Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. *Curr. Zool.* 62(6):581-601.

Riginos, C., M. W. Nachman. 2001. Population subdivision in marine environments: The contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, *Axoclinus nigricaudus*. *Mol. Ecol.* 10(6):1439-1453.

Riginos, C., B. C. Victor. 2001. Larval spatial distributions and other early life-history characteristics predict genetic differentiation in eastern Pacific blennioid fishes. *P. Roy. Soc. B.* 268(1479):1931-1936.

Robertson, D. R., G. R. Allen. 2015. Shorefishes of the Tropical Eastern Pacific: online information system. Version 2.0.

Rochette, N. C., J. M. Catchen. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* 12(12):2640-2659.

Rosenberg, N. A. 2004. DISTRUCT: A program for the graphical display of population structure. *Mol. Ecol. notes.* 4(1):137-138.

Saenz-Agudelo, P., J. D. DiBattista, M. J. Piatek, M. R. Gaither, H. B. Harrison, G. B. Nanninga, M. L. Berumen. 2015. Seascape genetics along environmental gradients in the Arabian Peninsula: Insights from ddRAD sequencing of anemonefishes. *Mol. Ecol.* 24(24):6241-6255.

Sale, P. F., I. Hanski, J. P. Kritzer. 2006. The Merging of Metapopulation Theory and Marine Ecology: Establishing the Historical Context. En: Kritzer, J. P., P. F. Sale (eds.). *Marine Metapopulations*. Elsevier Inc. First edition. Sydney, Australia. pp 3-28.

Sambrook, J., E. F. Fritsch, T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold spring harbor laboratory press,

Sánchez-Velasco, L., M. F. Lavín, M. Peguero-Icaza, C. A. León-Chávez, F. Contreras-Catala, S. G. Marinone, I. V. Gutiérrez-Palacios, V. M. Godínez. 2009. Seasonal changes in larval fish assemblages in a semi-enclosed sea (Gulf of California). *Cont. Shelf Res.* 29(14):1697-1710.

Sandoval-Castillo, J., N. A. Robinson, A. M. Hart, L. W. S. Strain, L. B. Beheregaray. 2018. Seascape genomics reveals adaptive divergence in a

connected and commercially important mollusc, the greenlip abalone (*Haliotis laevis*), along a longitudinal environmental gradient. *Mol. Ecol.* 27(7):1603-1620.

Schoville, S. D., A. Bonin, O. Francois, S. Lobreaux, C. Melodelima, S. Manel. 2012. Adaptive Genetic Variation on the Landscape: Methods and Cases. *Annu. Rev. Ecol. Evol. S.* 43:23-43.

Selkoe, K. A., J. R. Watson, C. White, T. Ben Horin, M. Iacchei, S. Mitarai, D. A. Siegel, S. D. Gaines, R. J. Toonen. 2010. Taking the chaos out of genetic patchiness: Seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Mol. Ecol.* 19(17):3708-3726.

Selkoe, K. A., C. C. D'Aloia, E. D. Crandall, M. Iacchei, L. Liggins, J. B. Puritz, S. Von Der Heyden, R. J. Toonen. 2016. A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Mar. Ecol. Prog. Ser.* 554:1-19.

Soria, G., A. Munguia-Vega, S. G. Marinone, M. Moreno-Báez, I. Martínez-Tovar, R. Cudney-Bueno. 2012. Linking bio-oceanography and population genetics to assess larval connectivity. *Mar. Ecol. Prog. Ser.* 463:159-175.

Stanley, R. R. E., C. DiBacco, B. Lowen, R. G. Beiko, M. Van Wyngaarden, N. W. Jeffery, P. Bentzen, L. Benestan, L. Bernatchez, C. Johnson, P. V. R. Snelgrove, B. F. Wringe, I. R. Bradbury. 2018. A climate-associated multi-species cryptic genetic cline in the northwest Atlantic.

Steele, M. A. 1997. The relative importance of processes affecting recruitment of two temperate reef fishes. *Ecology.* 78(1):129-145.

Steele, M. A. 1999. Effects of shelter and predators on reef fishes. *J. Exp. Mar. Biol. Ecol.* 233(1):65-79.

Therkildsen, N. O., J. Hemmer-Hansen, R. B. Hedeholm, M. S. Wisz, C. Pampoulie, D. Meldrup, S. Bonanomi, A. Retzel, S. M. Olsen, E. E. Nielsen. 2013. Spatiotemporal SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod *Gadus morhua*. *Evol. Appl.*

Thomson, D. A., L. T. Findley, A. N. Kerstitch. 2000. Reef Fishes of the Sea of Cortez: The Rocky-Shore Fishes of the Gulf of California. University of Texas Press First Edition. Austin, Texas, USA. 302 p.

Ulate, K., C. Sánchez, A. Sánchez-Rodríguez, D. Alonso, O. Aburto-Oropeza, L. Huato-Soberanis. 2016. Latitudinal regionalization of epibenthic macroinvertebrate communities on rocky reefs in the Gulf of California. *Mar. Biol. Res.* 1000:1-13.

- Vázquez Rojas, K. A. 2018. Estructura Genética mediante SNPs en bancos de almeja generosa (*Panopea globosa*; DALL 1898) del alto Golfo de California. Tesis (Maestría en Ciencias en el Uso, Manejo y Preservación de los Recursos Naturales) La Paz, Baja California Sur, México. Centro de Investigaciones Biológicas del Noroeste, S. C. 81 p.
- Walker, B. W. 1960. The Distribution and Affinities of the Marine Fish Fauna of the Gulf of California. *Syst. Zool.* 9(3):123-133.
- Wang, I. J., G. S. Bradburd. 2014. Isolation by environment. *Mol. Ecol.* 23(23):5649-5662.
- Wang, J., D. X. Xue, B. D. Zhang, Y. L. Li, B. J. Liu, J. X. Liu. 2016. Genome-wide SNP discovery, genotyping and their preliminary applications for population genetic inference in spotted sea bass (*Lateolabrax maculatus*). *PLoS ONE.* 11(6):1-19.
- Waples, R. S. 1987. A Multispecies Approach to the Analysis of Gene Flow in Marine Shore Fishes. *Evolution.* 41(2):385-400.
- Waples, R. S., O. Gaggiotti. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.* 15(6):1419-1439.
- Wiley, J. W. 1976. Life histories and systematics of the western North American gobies *Lythrypnus dalli* (Gilbert) and *Lythrypnus zebra* (Gilbert). *T. San Diego Soc. Nat. Hist.* 18:196-183.
- Winter, D. J. 2012. MMOD: An R library for the calculation of population differentiation statistics. *Mol. Ecol. Resour.* 12(6):1158-1160.
- Wright, S. 1943. Isolation by Distance. *Genetics.* 28(2):114-138.
- Van Wyngaarden, M., P. V. R. Snelgrove, C. DiBacco, L. C. Hamilton, N. Rodríguez-Ezpeleta, N. W. Jeffery, R. R. E. Stanley, I. R. Bradbury. 2016. Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus*, using RADseq derived SNPs. *Evol. Appl.* 10:1-16.
- Xu, S., N. Song, L. Zhao, S. Cai, Z. Han, T. Gao. 2017. Genomic evidence for local adaptation in the ovoviviparous marine fish *Sebastiscus marmoratus* with a background of population homogeneity. *Sci. Rep. UK.* 7(1):1562.
- Xuereb, A., L. Benestan, É. Normandeau, R. M. Daigle, J. M. R. Curtis, L. Bernatchez, M. J. Fortin. 2018. Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly

dispersive marine invertebrate (*Parastichopus californicus*). *Mol. Ecol.* 27(10):2347-2364.

Zhang, B. D., D. X. Xue, J. Wang, Y. L. Li, B. J. Liu, J. X. Liu. 2016. Development and preliminary evaluation of a genomewide single nucleotide polymorphisms resource generated by RAD-seq for the small yellow croaker (*Larimichthys polyactis*). *Mol. Ecol. Resour.* 16(3):755-768.