



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
IZTACALA**

**GENÓMICA DE POBLACIONES Y DIVERGENCIA
ECOLÓGICA: EL CASO DEL COPÉPODO
Leptodiptomus DE LOS LAGOS DE LA
CUENCA ORIENTAL**

T E S I S

**QUE PARA OBTENER EL TÍTULO
BIÓLOGO**

**P R E S E N T A:
JAVIER URBÁN OLIVARES**

**DIRECTOR DE TESIS
Dr. Jorge Ciros Pérez**



Los Reyes Iztacala, Tlalneantla de Baz, Estado de México, 2019



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

A la Universidad Nacional Autónoma de México y a la Facultad de Estudios Superiores Iztacala por la formación y conocimientos académicos y personales que me ha brindado.

Al proyecto PAPIIT IN222619 (DGAPA-UNAM) y al proyecto del CONACYT CB-255788, por todos los apoyos económicos brindados durante diferentes etapas, para la realización de esta tesis.

Un gran y sincero agradecimiento al Dr. Jorge Ciros Pérez, por todas sus enseñanzas, consejos, apoyo y confianza que me brindo durante toda su dirección, para concluir exitosamente esta tesis.

A los miembros del comité tutorial: Dr. Elizabeth Ortega Mayagoitia, Dr. Jorge Eduardo Campos Contreras, Dr. Elías Piedra Ibarra por sus valiosos consejos y acertados comentarios para mejorar este trabajo de investigación. Y un especial agradecimiento al Dr. José Arturo Alcántara Rodríguez por su gran apoyo en los análisis bio-informáticos.

A todos mis amigos con los que conviví durante toda mi formación universitaria.

Dedicatorias

A mi familia Genaro, María de Lourdes, Eduardo, Oscar y José Luis, por la educación, motivación, aprendizaje, confianza, consejos, ejemplo, amor, apoyo y amistad incondicional en todo momento...

A ti Italia por llegar y decidir estar siempre.

Índice

Resumen.....	5
Introducción.....	7
Antecedentes.....	15
<i>Leptodiptomus</i> grupo <i>sicilis</i> como modelo de estudio.....	16
Justificación.....	16
Hipótesis.....	17
Objetivos.....	18
Área de Estudio.....	19
Materiales y Métodos.....	21
Resultados.....	30
Discusión.....	41
Conclusión.....	48
Literatura Citada.....	49

Resumen

El copépodo *Leptodiptomus* grupo *sicilis* cuenta con cuatro poblaciones que habitan lagos ecológicamente contrastantes dentro de la Cuenta Oriental Mexicana que, en esta tesis, se denominan como *L.* “Atexcac”, *L.* “El Carmen”, *L.* “La Preciosa” y *L.* “Quechulac”. Estudios previos han mostrado la existencia de diferenciación ecológica y de aislamiento reproductivo precigótico por inviabilidad de los migrantes, además de aislamiento reproductivo postcigótico por inviabilidad del híbrido entre al menos dos de estas poblaciones, aunque, al utilizar marcadores moleculares estándar (citocromo oxidasa subunidad I; COI) la divergencia genética se describe como baja, pero con genotipos exclusivos entre lagos. Con el fin de describir y analizar mejor los patrones de diversificación de las poblaciones en tiempo ecológico, se realizó un estudio de genómica poblacional con un método de genotipado por secuenciación (*genotyping by sequencing*; GBS por sus siglas en inglés) que identificó 25,870 marcadores moleculares polimórficos (*Single Nucleotide Polymorphism*; SNPs por sus siglas en inglés). Dentro de este conjunto se identificaron 1,700 SNPs con frecuencias genotípicas que no se ajustan a la neutralidad para las cuatro poblaciones. Los resultados de diversidad genética mostraron diferencias importantes entre cada una de las poblaciones, mientras que los análisis jerárquicos de agrupamiento fueron convincentes definiendo siempre cuatro grupos independientes. El análisis de varianza molecular mostró que la mayor variación está dentro de las poblaciones cuando se analizan los 25,870 SNPs totales, por contrario el análisis basado en 1,700 SNPs atípicos muestra que la mayor variación molecular esta entre poblaciones. Los valores del índice F_{ST} mostraron el valor de diferenciación más alto entre las poblaciones *L.* “Atexcac” y *L.* “El Carmen” ($F_{ST} = 0.2692$), mientras que el valor más bajo se registró entre *L.* “La Preciosa” y *L.* “Quechulac” ($F_{ST} = 0.0857$), este patrón se repitió con ambos conjuntos de SNPs pero, cabe resaltar que con los 1,700 SNPs (entre *L.* “Atexcac” y *L.* “El Carmen” el $F_{ST} = 0.8063$ y entre *L.* “La Preciosa” y *L.* “Quechulac” el $F_{ST} = 0.3860$) los valores son considerablemente más altos en comparación a los que se obtiene con 25,870 SNPs. Las relaciones filogenéticas igualmente mostraron este patrón, definiendo que las poblaciones afines de copépodo *L.* grupo *sicilis* han divergido en 4 clados independientes (*L.* “Atexcac”, *L.* “El Carmen”, *L.* “La Preciosa” y *L.* “Quechulac”) y perfectamente definidos con historias evolutivas independientes. Dados todos estos resultados, donde los patrones de divergencia genética y las relaciones filogenéticas entre las poblaciones indican que las poblaciones dulceacuícolas (*L.* “La Preciosa” y *L.* “Quechulac”), son más cercanas entre sí y que la población más divergente es aquella proveniente del lago somero y altamente

variable (*L. "El Carmen"*), sugieren que las condiciones ecológicas diferenciales y bajo eventos recientes de deriva genética (persistencia de eventos fundadores), tasas constantes de mutación y reforzadas por selección divergente, resultado del aislamiento geográfico (resultado de una aparente tasa de dispersión baja), pero principalmente por los factores ecológicos contrastantes, han propiciado la divergencia interpoblacional, la adaptación local y un evento de especiación ecológica como se ha analizado en trabajos previos. Estos resultados contribuyen a sentar las bases moleculares para poder explorar a futuro las características genéticas y funcionales asociadas a la evolución adaptativa de este complejo de especies de copépodos afines a *L. grupo sicilis*.

Introducción

Diversificación Genética y Evolución Adaptativa

La genómica poblacional estudia la composición genética de las poblaciones, su distribución, los cambios genotípicos y las frecuencias alélicas de los individuos, a través de marcadores moleculares con cobertura a escala genómica dentro y entre poblaciones, desarrollando un marco matemático para describir las tasas de cambio evolutivo modeladas por mecanismos como la mutación, recombinación, deriva genética y selección natural, que operan directamente sobre la variación genética en escala ecológica. Así, esta disciplina ha demostrado ser una de las herramientas más eficiente para el estudio de la ecología evolutiva, dilucidando los eventos y procesos que originan y acumulan cambios progresivos en los organismos a través del tiempo, permitiéndoles sobrevivir y reproducirse dando forma a la variación y estructura genética de las poblaciones naturales, así como la importancia de los procesos adaptativos y no adaptativos, y si el ritmo predominante de cambio es gradual o episódico dentro de procesos de especiación posteriores (Barrick, y Lenski, 2013; Eguarte *et al.*, 2013; Ellegren y Galtier, 2016; Wolf y Ellegren, 2016).

Darwin (1859) concluyó que los organismos deben estar expuestos a condiciones nuevas durante varias generaciones para que se produzca un grado considerable de variación con base en su naturaleza y la naturaleza de las condiciones ambientales. Así, la diversidad genética o polimorfismo genético se define como las diferencias existentes en la secuencia de DNA entre individuos de una especie o población, la cual es un aspecto clave para que ocurra la evolución biológica, ya que la variación permite la resiliencia y supervivencia de una población frente a las condiciones ambientales dinámicas y contrastantes, considerándose una ventaja ante escenarios inesperados. Esta variabilidad entre los individuos es la responsable de que ocurran cambios adaptativos a lo largo del ciclo de vida de una especie, es decir, influye en los rasgos de la historia de vida de los genotipos y, en consecuencia, en la dinámica poblacional, impactando los sistemas de apareamiento, la prevalencia de los efectos de ligamiento y el tamaño efectivo de la población. Por ejemplo, se sabe que las especies con tasas altas de fecundación, que liberan un gran número de huevos en el ambiente (llamados estrategias *r*), son mucho más polimórficos en comparación con especies que producen un menor número de descendencia y proporcionan un mayor cuidado parental (llamados estrategias *K*; Ellegren y Galtier, 2016).

En una población ideal, panmíctica, se espera que se mantenga el tamaño efectivo de la población (N_e), que es el número de individuos reproductores en una población idealizada de tamaño constante con apareamiento aleatorio que mostrarían la misma cantidad de diversidad genética que una población real. Y si las poblaciones fueran infinitamente grandes y los procesos evolutivos como la selección, la mutación y la migración fueran ignorados, entonces la frecuencia de las variantes genéticas se mantendría igual durante generaciones (en Equilibrio de Hardy-Weinberg). Sin embargo, no ocurre así, en general, las poblaciones con un alto nivel de variación genética son el resultado de un elevado flujo génico entre poblaciones y de tamaños poblaciones grandes. Por el contrario, si las poblaciones de una especie son muy diferentes entre sí, es posible que se trate de poblaciones más pequeñas y/o que exista flujo génico bajo entre ellas (Eguiarte *et al.*, 2013; Ellegren y Galtier, 2016). Además, los patrones de apareamiento en las poblaciones naturales con flujo genético también juegan un papel importante en la diversidad genética, ya que suelen ser no aleatorios, es decir que ocurren bajo la selección de rasgos específicos, y estas elecciones pueden estructurar las frecuencias de genotipos que aparecerán en generaciones sucesivas. Como consecuencia, el flujo genético no aleatorio puede conducir a la endogamia (apareamiento entre individuos con genotipos muy locales), que tiende a homogenizar la variación genética o exogamia (apareamiento entre individuos con genotipos inmigrantes) que puede conducir a una variación genética más heterogénea a nivel interpoblacional (Klug y Cummings, 1996).

El flujo génico y la deriva génica pueden actuar en conjunto en la estructuración genética entre poblaciones. La deriva génica describe las fluctuaciones aleatorias (fijación y pérdida de alelos) en el número de variantes genéticas de una población, que ocurre cuando la aparición de formas variantes de un gen (alelos) aumenta o disminuye por casualidad a lo largo del tiempo. Por lo general, el efecto de la deriva genética se da en mayor magnitud en poblaciones pequeñas y ocurre después de que una población experimenta un cuello de botella, donde alelos raros (con poca frecuencia) se enfrentan a una mayor probabilidad de perderse o que se fije un alelo único en un locus particular dentro de una población, provocando la disminución de la diversidad genética, lo que puede causar que una nueva población sea genéticamente distinta a la población ancestral. Esto ha llevado a la hipótesis de que la deriva génica juega un papel importante en la diversificación de las poblaciones. Por su parte, si el movimiento de individuos entre poblaciones es exitoso y los migrantes tienen la capacidad de permanecer y tener apareamiento con individuos que se encuentran en el lugar de destino, contribuirán a la entrada de gametos portadores

de alelos que pueden alterar la proporción existente de alelos en la población de destino (Klug y Cummings, 1996; Tobler *et al.*, 2008). Así, en un modelo simple, donde los individuos tienen las mismas probabilidades de dispersarse entre poblaciones y mantienen una tasa constante de migración (m , que es la fracción de copias de genes dentro de una población introducida por inmigración por generación), un solo migrante por generación ($N_e m = 1$) podría constituir un flujo de genes suficiente para contrarrestar los efectos diversificadores de la deriva génica, pero, si la deriva o el efecto de fundación son persistentes y el flujo de migrantes no es constante, puede ocurrir la divergencia, y eventualmente la especiación (Andrews, 2010).

Las mutaciones, que son los cambios aleatorios y espontáneos dentro de las secuencias de DNA de un organismo, se dan por causas diversas, como son las fuentes de radiación, la exposición a productos químicos en el ambiente o por errores durante la replicación del DNA. Generalmente, se dividen en dos tipos: mutaciones puntuales y cromosómicas. Las mutaciones puntuales son aquellos cambios en uno o más nucleótidos, existen las sustituciones (una base se agrega incorrectamente durante la replicación), las inserciones (se adicionan uno o más nucleótidos en el DNA replicado) y las supresiones (uno o más nucleótidos se “saltan” cambiando el marco de lectura) de uno o varios pares de bases nuevos en las cadenas de DNA. Mientras que las mutaciones cromosómicas se dan a mayor escala, en el cual se alteran grandes segmentos de cromosomas. En este caso, los fragmentos de cromosomas pueden eliminarse, duplicarse, invertirse, trasladarse a cromosomas diferentes o reorganizarse de otro modo, lo que resulta en cambios como la modificación de la dosificación del gen, la ausencia completa de genes o la alteración de las secuencias y suelen ocurrir durante la meiosis por daño irreparable del DNA, errores de la polimerasa, eventos de recombinación intragenómica, movimientos de elementos transponibles o por deslizamientos durante la recombinación del DNA (Carlin, 2011). Las mutaciones pueden ser identificadas como benéficas, neutrales, deletéreas o letales con respecto a sus efectos sobre la eficacia biológica de los organismos o de las poblaciones en conjunto (Barrick y Lenski, 2013). Las mutaciones son esenciales para que ocurra la evolución ya que son la fuente de las nuevas características genéticas en los organismos, estas nuevas variantes (alelos) se propaga a través de la reproducción y, en caso de ser beneficiosas, permiten que un organismo se alimente, crezca, sobreviva y se reproduzca de manera más eficiente, por lo tanto, el alelo mutante, con el tiempo, tenderá a incrementar su frecuencia dentro de la población. En el caso de las mutaciones deletéreas también pueden generar cambios evolutivos importantes en las poblaciones, reduciendo las

frecuencias alélicas al eliminar individuos que podrían estar portando alelos adaptativos en otros genes, aunque este fenómeno se da principalmente en poblaciones pequeñas (Carlin, 2011). Las tasas de mutación cambian a lo largo del tiempo evolutivo, por lo tanto, es importante entender cómo los factores genéticos y ambientales afectan las tasas de mutación en poblaciones divergentes (Barrick y Lenski, 2013).

Por otro lado, la selección natural ocurre cuando los rasgos heredables de los organismos con una mayor supervivencia y éxito reproductivo incrementan su frecuencia dentro de la población, transmitiendo sus caracteres ventajosos a la siguiente generación. Y se da bajo las siguientes condiciones: la existencia de variabilidad genética (individuos con diversidad), que al menos una parte de dicha variabilidad debe ser heredable a la siguiente generación (es decir, que la variación tenga una base genética), que existan individuos más exitosos que otros y que los rasgos que mejoran la eficacia de los organismos no sean aleatorios y, por lo tanto, que sean seleccionados (Darwin, 1859). Cuando las poblaciones se encuentran sujetas a selección natural presentan tres escenarios evolutivos probables: 1) un único especialista adaptado de forma óptima a un hábitat (generalmente el más común o productivo), y poco adaptado a otros escenarios con pocas oportunidades de migrar y establecerse en nuevos hábitats. 2) Generalistas adaptados para tolerar tipos múltiples de hábitats, estos pueden ser generalistas intermedios, fenotípicamente uniformes o expresar fenotipos alternativos bajo diferentes condiciones ambientales, es decir, con alta plasticidad fenotípica. 3) Múltiples especialistas que pueden adaptarse localmente a los diferentes tipos de hábitats alternativos, por lo tanto, se esperarían fenotipos especializados divergentes que maximicen su eficacia biológica en un hábitat dado y no migren entre los diversos tipos de hábitats, lo que resulta en un flujo genético limitado entre las poblaciones (Tobler *et al.*, 2008). Este último dependerá de la velocidad a la que un alelo ventajoso se fije, y depende en parte de las relaciones de dominancia entre alelos en el locus (o los loci) en cuestión. Por lo tanto, *loci* bajo selección divergente deberían exhibir una mayor diferenciación que las regiones neutrales débilmente ligadas o desvinculadas y, si la selección actúa fuertemente, las poblaciones pueden llegar al aislamiento genético y reproductivo. Así, la selección natural es el mecanismo más importante del cambio de frecuencias alélicas y el único que genera la adaptación de organismos a sus entornos (Xu *et al.*, 2017).

Un reto importante ha sido el cómo visualizar y analizar adecuadamente la diversidad genética entre individuos y poblaciones, para ello, se han desarrollado múltiples

tecnologías que analizan la variación genética de las especies. Por ejemplo, recientemente, se han desarrollado técnicas para la genotipificación utilizando la secuenciación de próxima generación (NGS, por sus siglas en inglés *Next Generation Sequencing*) como un método útil para producir una gran cantidad de marcadores en genomas de alta densidad y a un bajo costo. Los polimorfismos de un solo nucleótido (SNP, por sus siglas en inglés *Single Nucleotide Polymorphism*) son los marcadores de DNA más comunes en la actualidad para el estudio genético de las poblaciones. La genotipificación por secuenciación (GBS, por sus siglas en inglés *Genotyping By Sequencing*) utiliza los beneficios de las enzimas de restricción para la reducción de la complejidad de los genomas y crear bibliotecas genómicas adecuadas para la identificación de SNPs y poder realizar estudios genómicos caracterizando poblaciones, patrones reproductivos y mapeo de rasgos genéticos de diversos organismos asociados a la adaptación local (Elshire *et al.*, 2011).

El grado de divergencia genética a menudo se mide utilizando los estadísticos F de Wright (1931) que son un conjunto de medidas jerárquicas de las correlaciones de alelos dentro de los individuos y dentro de las poblaciones. El estadístico F más relevante para el estudio del flujo de genes es el índice de fijación F_{ST} desarrollado por Wright (1996), este parámetro es especialmente útil para analizar los datos de SNPs, el índice tiene varias interpretaciones; el más famoso es la varianza en las frecuencias alélicas entre poblaciones, σ^2_p , estandarizada por la frecuencia media de alelos (p) para ese locus: $F_{ST} = \sigma_p^2 / [p(1-p)]$. Donde los valores más grandes representan una mayor diferenciación entre las poblaciones. Esta derivación mide el grado de diferenciación genética y el valor de F_{ST} puede oscilar entre 0 (sin diferenciación) y 1 (diferenciación completa, en la que las otras poblaciones se fijan para alelos exclusivos; Norgard y Schultz, 2008; Andrews, 2010; Nosil y Buerkle, 2010;). La diferenciación de las poblaciones depende estrictamente de la diversidad genética intraespecífica y de la diversidad fenotípica de cada población, por ello, es importante entender cómo los factores determinantes de la diversidad genética (véase, Fig.1) actúan sobre la historia de vida de los individuos que componen las poblaciones, y como varían sus estructuras genéticas bajo diferentes contextos ecológicos, y si estas variaciones son resultado de los factores que determinan la diversidad genética a nivel local y/o regional (Fig. 1).

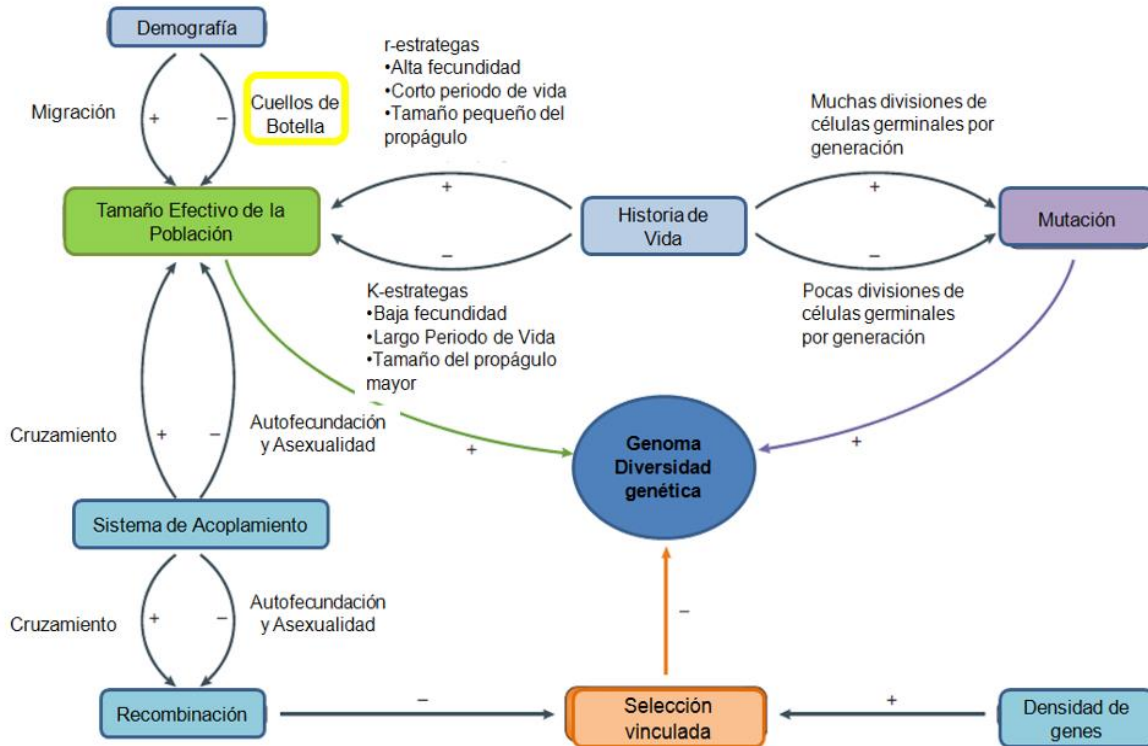


Figura 1. Visión general de los factores determinantes de la diversidad genética (tamaño efectivo de la población, deriva genética, mutación y selección) donde la dirección de correlación se indica con los símbolos – y +. Tomado y modificado de Ellegren y Galtier (2016).

La divergencia ecológica se puede definir como el proceso mediante el cual una especie puede formar dos o más linajes y es promovido principalmente por mecanismos evolutivos, barreras geográficas y/o ecológicas como respuesta a condiciones bióticas y/o abióticas contrastantes, la cual conduce a la divergencia de rasgos genéticos y aislamiento reproductivo entre las poblaciones. En modelos que conducen a la especiación ecológica, el aislamiento reproductivo evoluciona incidentalmente como un subproducto, pero cada vez que ocurre selección natural divergente entre las poblaciones, puede acompañarse de selección directa en el aislamiento pre reproductivo (Arnegard *et al.*, 2014). Se conoce que una variedad de agentes selectivos bióticos puede conducir al aislamiento reproductivo, como la interferencia reproductiva, el uso diferencial de recursos, la competencia interespecífica por recursos, la depredación y el parasitismo (Tobler *et al.*, 2008). Es por ello, que la selección divergente solamente puede actuar en escenarios con base en factores ecológicos contrastantes. Sin embargo, los eventos fortuitos también desempeñan un papel importante en la divergencia ecológica como la poliploidización, la hibridación, la

deriva genética, los cuellos de botella y los eventos de efecto fundador (Coyne y Orr, 2004). Aunque estos eventos suelen estar acompañados de algún tipo de selección (Schluter, 2000). Así mismo, la selección entre ambientes divergentes es consistente con el modelo clásico de especiación alopátrica, aunque la separación geográfica no es un requisito previo, pues también puede ocurrir entre las poblaciones simpátricas que ocupan nichos separados dentro de una misma área geográfica o un mismo sistema ecológico. Así, cuando la selección divergente persiste lleva a la especiación ecológica, teniendo como condición que las poblaciones presenten una base ecológica de selección divergente, una forma de aislamiento reproductivo y un mecanismo genético que los mantenga vinculados (Rundle y Nosil, 2005). Sin embargo, la prevalencia en la naturaleza de la selección natural divergente entre ambientes tampoco está clara, si existen entornos intermedios y los fenotipos intermedios funcionan bien en ellos, entonces, en teoría es posible que las poblaciones adaptadas a diferentes entornos se hayan separado unas de otras por deriva genética solamente (Gavrilets, 2004; Schluter, 2000). Por ello es importante analizar y describir la variación y estructura genética en las poblaciones para poder inferir y comprender mejor: (1) ¿Cómo los factores que determinan la variación genética conducen la evolución? (2) ¿Qué efecto tienen las fluctuaciones de la variación genética en conjunto con los mecanismos evolutivos en las poblaciones a lo largo del tiempo ecológico? Y, (3) ¿En qué grado la divergencia genética determina los procesos de adaptación, divergencia de linajes y eventualmente la especiación?

Copépodos de Aguas Continentales

Los sistemas lénticos (lagos y estanques) son un conjunto diverso de hábitats de agua dulce o salina que proporcionan recursos y hábitats esenciales para los organismos terrestres y acuáticos. Tienen una distribución amplia, discreta a manera de islas que representan poco más del 3% de la superficie de la Tierra (Downing *et al.*, 2006). La dinámica y estructuración de las comunidades biológicas de estos sistemas están determinadas por el conjunto de factores bióticos y abióticos que los caracterizan, de manera que los factores físicos y químicos del agua pueden llegar a limitar la amplitud de las distribuciones de especies debido a las capacidades ecofisiológicas que tienen para hacer frente a las condiciones locales. Por ejemplo, existen especies que pueden tolerar la temporada de sequía y por ello mantienen una distribución más amplia, colonizando una mayor diversidad hábitats, mientras que las especies que no toleran las sequías están restringidas a los lagos permanentes. Así, los lagos están conformados por elementos y factores muy particulares,

que, aunado a su aislamiento intrínseco dado su patrón de distribución, los hacen únicos, por lo que no es raro encontrar alta diversidad específica cuando se comparan, algunas de ellas endémicas que han evolucionado, principalmente, a través de la radiación adaptativa (Cristescu *et al.*, 2010). Por estas razones, los lagos son excelentes sistemas para el estudio de la divergencia ecológica, ya que sus características ambientes contrastantes generan tasas de evolución altas en cada cuerpo de agua, por lo que se han descrito altos grados de adaptaciones locales, es decir, el ajuste de los rasgos de los organismos a su ambiente en condiciones particulares, así como las diferencias genéticas interpoblacionales teniendo repercusiones directas en la distribución, dinámicas y estructuras genéticas poblacionales y comunitarias (De Meester *et al.*, 2002; Heino, 2011).

Los copépodos son uno de los metazoos más abundantes en la Tierra, estos pequeños crustáceos acuáticos muestran una amplia gama de adaptaciones morfológicas y fisiológicas a diferentes hábitats, colonizado casi todos los ecosistemas acuáticos bentónicos y planctónicos, en la mayoría de casos dominan sobre otros grupos de invertebrados crustáceos y no crustáceos en cuerpos de aguas continentales, donde están representados por ~3,000 especies (Boxshall y Defaye, 2008), pertenecientes a seis órdenes: Platycopioidea, Misophrioida, Calanoida, Cyclopoida, Harpacticoida y Gelyelloida.

Como se describió previamente, la distribución y colonización de los hábitats en los sistemas acuáticos continentales y subterráneos es, a menudo, el resultado de invasiones múltiples, frecuentemente seguidas de radiación adaptativa a través de la diversificación de nichos (Vaillant *et al.*, 2013). Los copépodos en pequeñas escalas se ven fuertemente afectados por la heterogeneidad en el tipo de hábitat, como la disponibilidad de recursos y la presencia de depredadores. Su distribución a escala biogeográfica se relaciona principalmente con procesos geológicos, climáticos y geográficos que ocurrieron en escalas de tiempo de mediano a largo plazo, afectando los patrones de colonización y la diversificación de los copépodos de aguas continentales, lo que ha dado lugar a una serie de reliquias filogenéticas con una distribución de alto grado de endemismo a diferentes niveles taxonómicos y se ve reflejado en la composición de comunidades de copépodos caracterizadas por relaciones filogenéticas distantes (Galassi *et al.*, 2009). Así, la amplia diversidad biológica y ecológica de la subclase Copepoda proporciona ventajas únicas para explicar problemas clave en los sistemas acuáticos desarrollando estrategias de análisis genómicos que ayuden a descifrar, por ejemplo, los patrones de evolución y la identificación de mecanismos genéticos y los cambios fisiológicos asociados con la adaptación a nuevos

entornos. En este sentido, la disponibilidad de recursos genómicos a gran escala con que se cuenta actualmente permite abordar cuestiones clave para entender procesos complejos en diversos ámbitos de la ecología y evolución (Bron *et al.*, 2011).

Antecedentes

En un estudio reciente, Barrera-Moreno *et al.* (2015) analizaron las poblaciones de estos copépodos, y demostraron que la selección divergente y los factores ecológicos pueden explicar los patrones de diversificación en estos organismos zooplanctónicos. Estos autores reportan una divergencia en al menos tres fenotipos diferentes: dos adaptados localmente, especializados en los lagos con salinidad constante (salina vs. agua dulce) y un fenotipo intermedio generalista que habita en el lago temporal con fluctuación de la salinidad, lo que ocasiona que el éxito de los inmigrantes entre lagos sea poco probable. La divergencia genética para el gen citocromo oxidasa subunidad I (COI) es baja ($>0.4\%$), pero existen la presencia de genotipos exclusivos en algunos lagos, que, en conjunto, permite inferir la existencia de restricciones en el flujo genético y que genera una fragmentación alopátrida de las poblaciones.

Recientemente, Barrera-Sánchez (2016) realizó un estudio con las poblaciones de Atexcac y La Preciosa de estos copépodos en el que documenta la existencia de aislamiento reproductivo postcigótico asimétrico, resultados que sugieren la divergencia del linaje en al menos dos especies biológicas diferentes, que habitan sistemas contrastantes, el lago salino Atexcac y otra en el lago dulceacuícola La Preciosa.

***Leptodiptomus* grupo *sicilis* como modelo de estudio**

La distribución geográfica del copépodo *L. sicilis* (Forbes, 1882) se encuentra restringida al norte de América, abarcando cuerpos de agua tanto dulces como salinos (Robertson y Gannon, 1981; Balcer *et al.*, 1984). En México, se registró por primera vez en el lago Atexcac (Macek *et al.*, 1994) y, posteriormente, en la misma cuenca hidrológica se adicionaron otras tres poblaciones afines a la especie en los lagos La Preciosa, El Carmen y Quechulac (Barrera-Moreno, 2010), los antecedentes presentados exponen diferenciación genética y reproductiva entre al menos dos poblaciones, lo que hace a *L. sicilis* de estas poblaciones un excelente modelo de estudio para el estudio de la divergencia ecológica.

Justificación

Basado en lo anterior, es necesario realizar análisis más finos de la diversidad y estructura genética/genómica dentro y entre poblaciones de los copépodos afines a *L. sicilis*, así, como análisis de inferencia bayesiana de las relaciones filogenéticas mediante el mapeo genómico de los individuos que componen cada una de las poblaciones, usando metodologías moleculares que revelen los procesos que han dado lugar a patrones de diversificación en tiempo reciente, como es el uso de genotipado por secuenciación (GBS) que permiten visualizar polimorfismos de un solo nucleótido (SNPs). Su uso representa una ventaja al momento de evaluar diversidad genética con una mayor cobertura a nivel genómico, y así, poder inferir los mecanismos evolutivos que han modelado y estructurado la variación genética dentro y entre las poblaciones en estos ambientes contrastantes. Para que eventualmente, permita la posibilidad de estudiar los patrones adaptativos e identificar mecanismos genéticos y cambios fisiológicos y/o morfológicos con un valor adaptativo bajo condiciones locales contrastantes en organismos que han sido bien caracterizados ecológicamente, como es el caso de los copépodos denominados hasta ahora como *L. grupo sicilis*.

Hipótesis

1. Las poblaciones del copépodo *L.* grupo *sicilis* (*L.* cf. *sicilis*; véase, Barrera-Moreno *et al.*, 2015) presentarán valores significativos de diferenciación genética interpoblacional, formando una estructura genética particular para cada lago y relaciones filogenéticas bien diferenciadas como resultado de habitar en sistemas ecológicos contrastantes.
 - 1.1 Las poblaciones que habita los lagos profundos y estables, Atexcac, La Preciosa y Quechulac presentarán una mayor proporción de homocigosis, resultado de un nicho ecológico poco variable, que han conducido a la adaptación a las condiciones locales y al poco éxito de la entrada de nuevos migrantes. Por el contrario, El Carmen presentará una frecuencia de mayor diversidad genética y heterocigosis debidas a la alta variabilidad ambiental estacional propia de este lago.
 - 1.2 Como resultado de una alta estructuración genética entre poblaciones y el bajo o nulo flujo genético en tiempos recientes, las relaciones filogenéticas mostrarán unidades evolutivas independientes por lago.

Objetivos

Objetivo General

Analizar la diversidad y estructura genética para entender las relaciones filogenéticas de las poblaciones de los copépodos *L.* grupo *sicilis* (i.e., *L.* "Atexcac", *L.* "El Carmen", *L.* "La Preciosa" y *L.* "Quechulac) que habitan lagos ecológicamente contrastantes en la Cuenca Oriental, México, por medio del análisis genómico de sitios altamente polimórficos, y entender el papel de los procesos de divergencia ecológica en poblaciones que divergieron en tiempo reciente.

Objetivos Particulares

- Identificar marcadores moleculares polimórficos y de calidad a lo largo de todo el genoma de individuos provenientes de los diferentes lagos, para describir la diversidad genética dentro y entre poblaciones.
- Analizar la variabilidad de las frecuencias alélicas y describir la estructura genética dentro y entre poblaciones.
- Inferir procesos evolutivos a través de la descripción de las relaciones filogenéticas que ha estructurado las poblaciones de los copépodos *L.* grupo *sicilis*.

Área de Estudio

La Cuenca Oriental Mexicana es una cuenca endorreica, que se localizan entre los estados de Puebla, Tlaxcala y Veracruz, en las coordenadas 18°48' - 19°43' N y 97°09' - 98°03' W (Figura 2), dentro del Distrito Oeste de la Faja Volcánica Transmexicana (Escalante *et al.*, 2007) a una latitud promedio de 2,300 m s.n.m., la cual abarca una superficie de 5,250 km², aproximadamente (Alcocer *et al.*, 1993), y comprende los llanos de San Juan y San Andrés. Predominan suelos con sustrato calizo tipo Rendzina, además de Regosol, Litosol, Feozem, Andosol y Cambisol. Tiene contacto con zonas tropicales húmedas al este y templadas al norte y al oeste. Sus recursos hídricos lénticos son los lagos: Alchichica, San Luis Atexcac, La Preciosa, Aljojuca, San Miguel Tecuitlapa, Quechulac y Totolcingo (i.e., El Carmen), los pantanos de Tepeyahualco y algunas presas. Las características básicas limnológicas generales son una salinidad de 1.2-14 g/l; pH = 8-9; temperatura anual promedio entre 12.8-14.4°C; precipitación anual promedio entre 425-656 mm. El análisis del agua de los lagos señala una elevada concentración de aniones y cationes, lo que permite dividirlos en dos grupos: los salinos formado por los lagos de Alchichica y Atexcac, y aquellos con menor carga iónica formado por los lagos restantes. El sedimento de todos los lagos es de textura arenosa y baja concentración de nutrientes. Los lagos cráter corresponden a lagos tipo *maar* y se les incluye dentro del tipo 11 de la clasificación de Hutchinson (1975). Tienen su origen en conos con explosión volcánica durante el Pleistoceno tardío (Carrasco-Nuñez *et al.*, 2007), y presentan altas concentraciones de cloruros, bicarbonatos de sodio, magnesio y una elevada conductividad eléctrica (Davies *et al.*, 2002). En la actualidad los lagos de Tepeyahualco y El Carmen son lagos someros, terminales y polimícticos. El agua que alimenta los lagos proviene de los mantos freáticos principalmente y, en menor medida, de la precipitación pluvial directa (Álvarez, 1950). Quechulac, La Preciosa y Atexcac son lagos monomícticos cálidos, que presentan generalmente una estratificación durante los meses de abril a diciembre y una mezcla invernal de enero a marzo. Sus características limnológicas son contrastantes, van desde el agua dulce hasta la salina, con un promedio de salinidad desde 0.42 a 6.5 g L⁻¹ (Tabla 1) y son oligotróficos. El lago El Carmen es un lago tipo playa, somero, estacional con una profundidad máxima ~30 cm y presenta el gradiente de salinidad más amplio (véase Tabla 1).

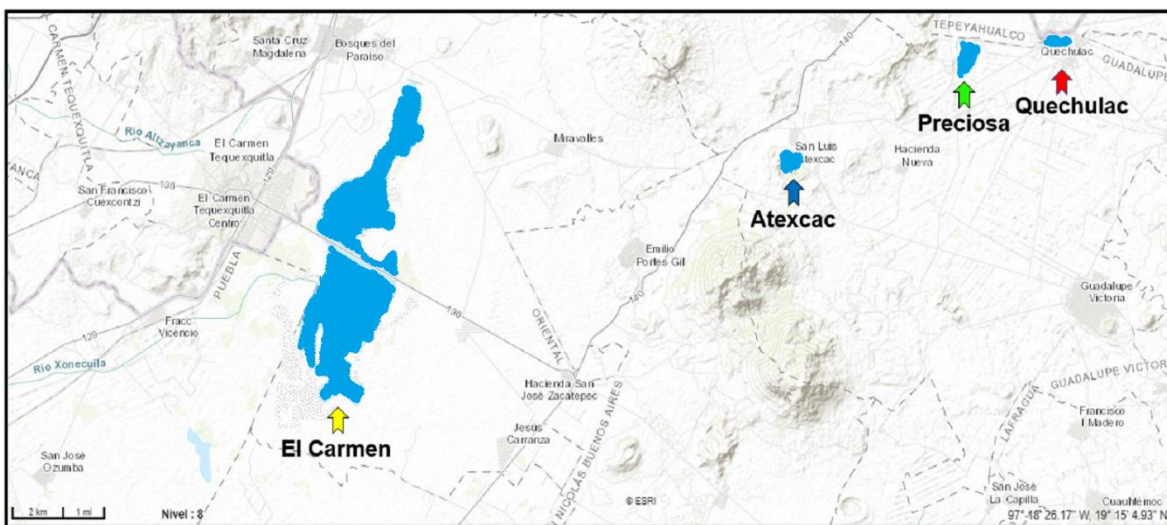


Figura 2. Mapa de la localización de los lagos El Carmen, San Juan Atexcac, La Preciosa y Quechulac en la Cuenca Oriental Mexicana (tomado del Instituto Nacional de Estadística y Geografía (INEGI) y modificado para esta tesis).

Tabla 1. Comparación de algunas características limnológicas importantes y contrastantes de los cuatro lagos donde habita los copépodos *L. grupo sicilis* ubicados en la Cuenca Oriental, México (modificado de Caballero *et al.*, 2001).

Nombre del Lago	Ubicación	Salinidad (g L ⁻¹)	Profundidad (m)	Presencia de Peces
Quechulac (Perenne)	19° 22' N, 97° 21' W	0.42 ± 0.05	40	Si
La Preciosa (Perenne)	19° 22' N, 97° 23' W	1.18 ± 0.09	45.5	Si
Atexcac (Perenne)	19° 20' N, 97° 27' W	6.54 ± 0.29	39	No
El Carmen (Temporal)	19° 09' a 19° 26' N, 97° 33' a 97° 47' W	1.4 a 10	<0.30	No

Barrera-Moreno (2010), registró copépodos con características morfológicas que corresponden a la especie *L. sicilis* (Forbes, 1882), habitando en los Lagos La Preciosa, El Carmen, Atexcac y Quechulac. Dados los resultados recientes que revelan que estos copépodos presentan patrones de aislamiento precigótico y postcigótico entre sus poblaciones (véase, Barrera-Moreno *et al.*, 2015; Barrera-Sánchez, 2016), y para facilitar la comunicación en esta tesis, los hemos denominado adicionando el topónimo del lago que habitan, por lo que los llamaremos *L. "Atexcac"*, *L. "El Carmen"*, *L. "La Preciosa"* y *L. "Quechulac"*.

Materiales y Métodos

Trabajo en Campo

Se tomaron muestras de zooplancton en cada uno de los lagos (Quechulac, la Preciosa, Atexcac y El Carmen; ver Tabla 2) con un arrastre vertical en la columna de agua utilizando una red de cono con apertura de poro de 80 μm ; se concentraron en frascos de 250 ml, posteriormente se fijaron con etanol absoluto, se transportaron en frío y oscuridad al laboratorio para preservar mejor el ADN. En el laboratorio se etiquetaron y se almacenaron en un congelador de refrigerador doméstico a $-4\text{ }^{\circ}\text{C}$ aproximadamente para su posterior análisis.

Tabla 2. Muestra las fechas en las que se realizó la colecta de muestras de copépodos *L. "Atexcac"*, *L. "El Carmen"*, *L. "La Preciosa"* y *L. "Quechulac"*.

Sitio de Colecta	Fecha de Colecta
Atexcac	14 de junio del 2016
Quechulac	2 de septiembre del 2010
La Preciosa	3 de marzo del 2011
El Carmen	5 de septiembre del 2010 y 2 de diciembre del 2015

Trabajo de Laboratorio

Con ayuda de un microscopio estereoscópico y de claves especializadas (Thorp y Covich; 2009; Barrera-Moreno; 2010), se aislaron e identificaron 12 individuos hembras adultas de cada una de las poblaciones (i.e., 3 individuos × 4 poblaciones = 12 individuos totales) para realizar una prueba piloto y evaluar distintas enzimas de restricción. Adicionalmente, se aislaron 94 individuos de (i.e., 24 individuos de *L. "Quechulac"* y *L. "El Carmen"*, y 23 de *L. "Atexcac"* y *L. "La Preciosa"*, respectivamente = 94 individuos totales) para el genotipado basado en secuenciación masiva.

Extracción, amplificación y purificación del ADN

El ADN se extrajo de cada individuo mediante el método de HotSHOT descrito por Montero-Pau *et al.* (2008), ajustando el volumen de la solución total a 20 µl por individuo con 10 µl de solución de lisis y 10 µl de buffer neutralizador. Después de esto, el ADN se amplificó siguiendo el protocolo del kit REPLI-g single cell (*Qiagen*TM) partiendo en cuatro cada reacción para ajustar a un volumen total a 12.5 µl por reacción. Las reacciones se incubaron en un termociclador a 30°C por 10 h y, finalmente, se inactivó la REPLI-g ADN polimerasa a 65 °C por 3 min. La purificación de cada producto de PCR se realizó siguiendo un protocolo estándar para precipitación de ADN con acetato de sodio 3 M y etanol al 100% y 70%, retirando el sobrenadante y resuspendiendo el ADN en 50 µl de buffer EB (*Qiagen*TM). Las muestras de ADN se cuantificaron por espectrofotometría en un equipo Epoch 2. (*BioTek*TM) para determinar la concentración y verificar la calidad del ADN (≥ 50 ng/µL) en una relación de absorbancia 260/280 (≥ 1.8 QC), para cumplir con los requerimientos solicitados por el Centro Genómico de la Universidad de Minnesota (CGUM), proveedores de la construcción y secuenciación de las librerías genómicas.

Digestión enzimática

Con 12 muestras iniciales de ADN, se realizó en el CGUM un análisis piloto para evaluar la digestión de ADN con un panel de 8 diferentes enzimas de restricción candidatas como lo sugiere el protocolo de GBS descrito por Elshire *et al.* (2011), y así, elegir la(s) enzima(s) óptima(s) que permitía(n) obtener mejores lecturas de ADN y por lo tanto una mayor cobertura del genoma. Se determinó que las enzimas de restricción MSPI y Nsil-HF eran la

mejor opción para usar en la construcción de librerías genómicas de las muestras de copépodos.

Secuenciación y construcción de librerías genómicas

Posteriormente, las 94 muestras de ADN de alta calidad genómica por individuo (>50 ng/ μ l; ≥ 2.5 μ g totales) fueron enviadas al CGUM para su análisis y digestión con las enzimas de restricción elegidas. Los métodos de preparación de librerías, la secuenciación y el análisis bioinformático para identificar SNPs, se basan en la re-secuenciación del genoma en pequeños fragmentos de ADN por individuo. Para este proceso se evaluó la calidad de ADN en cantidad (≥ 50 ng/ μ L) y pureza (≥ 1.8 QC) de las cuales 93 muestras cumplieron con los requisitos y se descartó una muestra que pertenecía a la población de *L. "La Preciosa"* por no tener la cantidad suficiente de ADN. El trabajo de construcción de librerías por GBS se realizó por medio de secuenciación masiva utilizando la plataforma Illumina® en una ejecución NextSeq 550 1x150-bp. El ADN fue digerido con las enzimas de restricción MSPI y Nsil-HF, después se le ligaron adaptadores para los índices de multiplex para la plataforma de secuenciación Illumina®, se usaron secuencias de 1-150 pb para identificar y amplificar cada muestra por medio de PCR. Las regiones detectadas para cada uno de los códigos de barras generaron > 1 M lecturas aproximadamente por muestra. Las lecturas para cada muestra tuvieron en promedio puntuaciones de calidad phred $\geq Q30$. Todas las bibliotecas de las 93 muestras fueron separadas para obtener los fragmentos que se conectaron a las secuencias crudas entre líneas en un solo archivo fastaq; adicionalmente se realizó un filtro para eliminar errores de secuenciación y se almacenaron las 93 muestras con 109,299 SNPs, en un Formato de Llamado de Variantes (VCF) para su descarga desde el portal electrónico del CGUM.

Análisis Bioinformático

Filtrado de SNPs

El llamado de SNPs se realizó a partir de los datos crudos de la secuenciación (formato fastq) se filtraron con base en su calidad, utilizando solo aquellos con un valor phred superior a 30 (99.9% de confianza) y se realizó la identificación de SNPs informativos a lo largo del genoma. Las secuencias sin procesar se recortaron utilizando trimmomatic para

el control de calidad y se mapearon contra las bases de datos genómicas (humano, ratón, E. coli, levadura, Phix y cloroplasto) para detectar la contaminación del ADN. Las secuencias que pasaron los filtros de calidad se mapearon contra la referencia genómica ensamblada (muestra CAR-16) utilizando el software BWA (Li & Durbin, 2009). Del conjunto de datos de sitios polimórficos, se eliminaron los individuos con > 20% de genotipos faltantes y las variantes con llamadas faltantes en más del 5% de las muestras, así como las variantes con MAF < 0.01. Los datos de secuencia procesadas y obtenidas de las lecturas de código de barras de alta calidad por línea con un total de 109,299 SNPs fueron ordenadas y sometidas a un filtro por sitios implementado en el software Tassel v.5.0. bajo las siguientes condiciones: (1) conteo mínimo = 5; (2) frecuencia mínima = 0.2; (3) repetición máxima = 0.8. Este filtrado arrojó un total de 25,870 marcadores tipo SNP de alta calidad que se colocaron en un formato VCF para continuar con los análisis de diversidad y estructura genética (Elshire *et al.*, 2011; Glaubitz *et al.*, 2014).

Diversidad y diferenciación genética

En los análisis descriptivos de la diversidad y divergencia genética se utilizó el conjunto de SNPs (25,870 pb) rigurosamente filtrados para el análisis de la diversidad nucleotídica, la tasa de mutación y una prueba de neutralidad basado en la D de Tajima entre las poblaciones utilizando Tassel 5.0. y el programa MEGA 7.0. (Kumar *et al.*, 2016). Las frecuencias alélicas (heterocigocidad observadas, H_o , y esperadas H_e) se estimaron usando la librería popr para el software R v.3.5.1. (R Core Team, 2013). La proporción de sitios segregantes o distancias P_s , que se calcula como la proporción de sitios homólogos que difieren en las secuencias de una población o sitios segregantes S sobre el total de sitios de alineamiento N , de acuerdo con a Nei y Kumar (2000): ($P_s = S/N$). Esta medida está influenciada por la presencia de alelos deletéreos (Hedrick, 2004) por que se encuentran en baja frecuencia en la población, y ya que el valor S no toma en cuenta la frecuencia de mutantes, ni frecuencias alélicas en general. Por ello, una medida más adecuada para medir la diversidad es la diversidad nucleotídica π , y que se relaciona directamente con las frecuencias alélicas.

La distribución y media de las frecuencias alélicas se analizan para inferir las implicaciones evolutivas que se pueden deducir a partir del principio de Hardy-Weinberg que establece que la variación genética en una población se mantendrá constante de una

generación a otra en ausencia de factores perturbadores. Así, cuando el apareamiento es aleatorio en una población sin circunstancias disruptivas, se predice que tanto las frecuencias genotípicas como las alélicas permanecerán constantes porque se mantienen en equilibrio. El equilibrio de Hardy-Weinberg puede ser perturbado por los mecanismos evolutivos como las mutaciones, la selección natural, el apareamiento no aleatorio, la deriva genética y el flujo de genes. Y debido a que estos mecanismos ocurren comúnmente en la naturaleza, el equilibrio de Hardy-Weinberg rara vez se cumple en poblaciones naturales. Por lo tanto, el equilibrio de Hardy-Weinberg describe un estado idealizado, y las variaciones genéticas en la naturaleza se pueden medir como cambios a partir de este estado de equilibrio para inferir procesos evolutivos a través de la distribución y media de las frecuencias alélicas de las poblaciones para describir implicaciones evolutivas a partir del teorema de Hardy-Weinberg (Salanti *et al.*, 2005).

Estructura Genética

Red de expansión mínima

Se realizó con ayuda del paquete R v.3.5.1 un análisis independiente útil para visualizar la estructura de la población a través de una red de expansión mínima (MSN). La MSN agrupa genotipos multilocus (MLG), basado en las distancias genéticas existentes entre ellos. Cada MLG es un nodo, y estos están conectados por la distancia mínima entre muestras y la distancia genética está representada por los bordes. En los estudios de secuenciación de alto rendimiento, donde la densidad de marcadores es alta, cada muestra generalmente consta de un genotipo único y se realiza con una matriz de distancias basada en el número de diferencias alélicas entre muestras (Kamvar *et al.*, 2015).

Detección de *loci* sujetos a selección

Se utilizó la librería pcadapt v.4.0.3 (Luu *et al.*, 2017), dentro del paquete R v.3.5.1, para detectar marcadores genéticos involucrados en la adaptación biológica, pcadapt proporciona herramientas estadísticas para la detección de valores atípicos (*outliers*), basados en el análisis de componentes principales (PCA), aquí se realizó un escaneo del genoma para la selección de valores atípicos en función de los datos de genotipos de 93 individuos de nuestras cuatro poblaciones y los 25,870 SNPs.

El análisis de los componentes principales (PCA), se realizó con la matriz de genotipos centrada y escalada, para calcular el estadístico de prueba o test y los valores P , basados en las correlaciones entre los SNPs y los primeros K (K = número de componentes principales), evaluando el porcentaje de variación que es explicado por cada PC y para tener una indicación sobre la presión evolutiva. Las estadísticas de prueba son las cargas, que corresponden a las correlaciones entre cada PC y cada SNP. Los valores P se calculan haciendo una aproximación gaussiana para cada PC y estimando la desviación estándar de la distribución nula.

El patrón ideal para la mejor elección del PC es indicado en una gráfica con una curva pronunciada seguida de una curva y una línea recta. Los valores propios que corresponden a la variación aleatoria se encuentran en una línea recta, mientras que los que corresponden a la estructura de la población se encuentran en una curva pronunciada, también se eligió el número de CP basado en el 'gráfico de puntuación' que muestra la estructura de la población. La gráfica de puntuación muestra las proyecciones de los individuos en los componentes principales especificados, así, la elección de K puede limitarse a los valores K que corresponden a un nivel significativo de la estructura de la población (Luu *et al.*, 2017).

El estadístico de prueba se basa en las puntuaciones z obtenidas al hacer una regresión de SNP con los K componentes principales. El estadístico de prueba se utilizó para detectar SNP atípicos, éste se basa en la distancia de Mahalanobis, que es un enfoque multidimensional que mide qué tan lejos está un punto de la media, también se configuró el parámetro de Frecuencia de Alelos Menores (MAF) del 5%. Los valores P de los SNP's con una frecuencia de alelo menor más pequeña que el umbral no se calculan, para este caso, se utilizó un modelo de divergencia asumiendo que 1,700 SNPs confieren una ventaja selectiva, este conjunto de SNPs también fue utilizado en los análisis posteriores.

Análisis de Componentes Principales

Se realizó un análisis inicial de agrupación para comprender la estructura de la población por medio de un análisis de componentes principales (PCA) que permite agrupar las diferentes variantes genéticas y evaluar la estructura poblacional, con el paquete R v.3.5.1, basado en las características genéticas de los individuos provenientes de cada lago. El análisis de PC se basa en un algoritmo matemático que reduce la dimensionalidad de los

datos y conserva la mayor parte de su variación, esta reducción se logra mediante la identificación de direcciones, llamadas componentes principales, a lo largo de las cuales la variación en los datos es máxima. Los datos se pueden trazar y son utilizados para explorar si hay similitudes y diferencias significativas entre las muestras, así como visualizar patrones de agrupamiento, como conjuntos (*clusters*), gradientes (*clines*) y valores atípicos (*outliers*) entre el conjunto de datos (Ringnér, 2008).

Adicionalmente, para la determinación de la estructura genética de las poblaciones se utilizó el paquete ADMIXTURE v.1.3.0. (Alexander *et al.*, 2009), que realizó una estimación de máxima verosimilitud de ancestros individuales a partir de conjuntos de datos de genotipos y SNPs multiloci y compara las secuencias entre sí para analiza las diferencias en la distribución de las variantes genéticas entre poblaciones por medio de un algoritmo de inferencia bayesiana iterativo, asignando en poblaciones asumidas a individuos que comparten patrones de variación genética similares. ADMIXTURE utiliza un enfoque de relajación de bloques para actualizar alternativamente los parámetros de frecuencias de alelos y fracción de ascendencia (Zhou *et al.*, 2011). Este análisis se realizó con 93 individuos de las cuatro poblaciones con 25,870 SNPs genómicos, y se probaron 1-5 grupos (K) en función de su fracción de pertenencia estimada entre cada grupo.

Análisis de Varianza Molecular (AMOVA) y Divergencia Genética (F_{ST})

Los procesos evolutivos, como la selección, la adaptación local y la deriva genética dependen en gran medida de la estructura de la población presente y pasada de las poblaciones. Por lo tanto, nuestra comprensión de los procesos evolutivos a menudo depende mucho del conocimiento detallado sobre la estructura de la o las poblaciones de una especie, así, la determinación de la estructura de la población de una especie constituye una parte importante de la genética de poblaciones. La fortaleza de la estructura de la población genética de las especies generalmente se describe en términos de la distribución estadística F que relacionan la diversidad dentro de la población con la diversidad genética total (Meirmans, 2012). El estadístico de F_{ST} compara el nivel de variación genética que existe dentro de dos o más subpoblaciones. F_{ST} a menudo se estima como: $F_{ST} = (T-S) / T$, donde T es la diferencia promedio entre las frecuencias de alelos, tomadas al azar de la población total, y S es la diferencia promedio entre las frecuencias alélicas, ambas tomadas al azar de la misma subpoblación. Si el apareamiento en la

población total es aleatorio, entonces se espera que T y S tengan el mismo valor y F_{ST} será 0. Sin embargo, si el apareamiento entre poblaciones no es aleatorio, la variación genética en cada subpoblación será menor que la variación en la población total, y F_{ST} tenderá a un máximo de 1 con aislamiento reproductivo. Por lo tanto, F_{ST} es una medida de la desviación del apareamiento aleatorio causado por la estructura genética de la población. Aunque F_{ST} es informativo sobre las diferencias promedio dentro de las poblaciones y entre ellas, nos dice poco sobre la consistencia o varianza de estas diferencias (Bamshad *et al.*, 2004).

El AMOVA es un método estadístico para cuantificar la diversidad genética a través de las pruebas estadísticas F en diferentes niveles jerárquicos de manera que la diversidad dentro del grupo se minimiza, mientras que la diversidad entre grupos se maximiza, además, AMOVA puede usarse para detectar agrupamientos o población en un conjunto de datos genéticos, en este trabajo se analizó el estadístico F_{ST} y el AMOVA del conjunto de 25,870 SNPs totales y 1,700 SNPs sujetos a selección para cuantificar la variación genética en dos niveles jerárquicos diferentes: dentro y entre las cuatro poblaciones, con el paquete R V.3.5.1. La estructura genética se interpretó a partir de las estadísticas asociadas con diferentes niveles jerárquicos en los que se distribuye la variación molecular (Excoffier *et al.* 1992; Meirmans, 2012) La implementación de AMOVA en R requiere dos componentes muy básicos: (1) una matriz de distancia derivada de los datos y (2) una tabla separada utilizada para dividir los datos en diferentes estratificaciones. En poblaciones panmícticas esperaríamos ver que la mayor parte de la varianza surja dentro de las muestras. Si vemos que la mayor parte de la varianza ocurre entre poblaciones, entonces, hay evidencia de que tenemos algún tipo de estructura poblacional.

Reconstrucción de las Relaciones Filogenéticas

El análisis comparativo de las secuencias moleculares es esencial para reconstruir sus historias evolutivas e inferir los mecanismos evolutivos que las han configurado, entendiendo la evolución de los genes y sus relaciones filogenéticas entre ellas, para ello se utilizó el paquete MrBayes v.3.2. (Ronquist *et al.*, 2012), que es un programa de inferencia bayesiana y permite la elección de una gama amplia de modelos filogenéticos y evolutivos, utilizando el método de Monte Carlo basado en cadenas de Markov (MCMC) para estimar la probabilidad posterior de los parámetros del modelo, y que permite la inferencia y estimación de árboles filogenéticos, así como de los parámetros de sustitución

y las tasas de variación entre las poblaciones (Kumar *et al.*, 2016). En nuestro caso se utilizó el método “4by4”, que es el método más usado y sugerido para marcadores moleculares tipo SNPs dentro de MrBayes para estimar las distancias evolutivas e inferencia de grandes filogenias, incluso en presencia de una variación extensa en la tasa evolutiva entre poblaciones o sesgos significativos en la composición de nucleótidos y la relación de transición/transversión. El principio de este método es similar al de verosimilitud máxima, y su finalidad es encontrar pares de unidades taxonómicas operacionales (OTU) que minimicen la longitud total de la rama en cada etapa del agrupamiento de OTU, comenzando con un árbol no enraizado. El uso de este método permite obtener rápidamente las longitudes de las ramas y la topología de un árbol parsimonioso, usando una simulación para obtener una filogenia correcta no arraigada (Saitou y Nei, 1987). Logrando un árbol general que ilustra las relaciones ancestro-descendiente entre los 93 individuos analizados pertenecientes a las cuatro poblaciones y caracterizados con los 25,870 SNPs totales y los 1700 SNPs atípicos obtenidos. Para ampliar el análisis de las relaciones filogenéticas se crearon 50 subconjuntos de 200 variantes (SNPs) cada uno, tomados aleatoriamente de la muestra de 25,870 y 1700 SNPs, para reconstruir un árbol de distancia con el método UPGMA (método de grupo de pares no ponderados con media aritmética) para cada subconjunto y se superponen todos los árboles reconstruidos con el uso de la función DensiTree implementada en R v.3.5.1. Esta superposición de árboles se utiliza para determinar si los subconjuntos respaldan nuestra hipótesis de la diferenciación de las poblaciones, o si existen grupos de árboles únicos para diferentes subconjuntos.

Resultados

Los resultados obtenidos después del multiplex y el filtrado por calidad $\geq Q30$ y códigos de barras ambiguos de 1-150 pb con un $\geq 1M$ lecturas por individuo, arrojó un total de 93 secuencias con 109,299 SNPs de alta calidad.

Después del filtrado para retener los SNPs de alta calidad, se adicionó otro filtro por sitios con las siguientes condiciones: (1) conteo mínimo = 5; (2) frecuencia mínima = 0.2; (3) repetición máxima = 0.8 y la presencia en al menos el 80% de las muestras, a partir del cual se eliminaron 83,429 SNPs, quedando un matriz de secuencias alineadas compuestas de 25,870 SNPs polimórficos para los 93 individuos que fueron usados para los análisis posteriores.

Diversidad Genética

Los análisis de diversidad genética que se realizaron con el conjunto de secuencias conformadas por 25,780 SNPs pertenecientes a 93 individuos de cuatro poblaciones describió una diversidad nucleotídica $\pi = 0.4368$ y poblacional $\theta = 0.1976$, y una D de Tajima = 4.18. La descripción genética por población se resume en la Tabla 3, donde podemos observar que, en general, los datos entre poblaciones son muy similares. La tasa de mutación poblacional fue muy similar para cada una de las poblaciones. Los resultados de las pruebas de D de Tajima se alejaron en todos los casos del modelo de neutralidad con valores mayores 0, por lo que podría ser que las poblaciones están bajo algún nivel de selección positiva. La heterocigosidad observada (H_o) y los análisis de la diversidad media de las poblaciones (R) oscilan en valores muy similares (1.19-1.24) describiendo una diversidad intrapoblacional similar en cada población.

Tabla 3. Describe la diversidad genética de las cuatro poblaciones del copépodo. π = diversidad nucleotídica, θ = diversidad nucleotídica poblacional, D = test estadístico de Tajima, H_o = heterocigosidad observada, Std H_o = desviación estándar de la H_o y R = diversidad media.

Poblaciones	π	θ	Tajima D	H_o	Std H_o	R
<i>L.</i> "Atexcac"	0.3578	0.2528	1.74	0.3464	0.224	1.24
<i>L.</i> "El Carmen"	0.351	0.2485	1.68	0.3175	0.2123	1.19
<i>L.</i> "Quechulac"	0.3973	0.2644	2.04	0.3815	0.1931	1.22
<i>L.</i> "La Preciosa"	0.3827	0.2693	1.75	0.3818	0.194	1.21

Red de expansión mínima

El resultado del análisis (ver Figura 3) con los datos de las poblaciones de *L. grupo sicilis* definió cuatro grupos o nodos conformados cada uno por las poblaciones asumidas (Atexcac, El Carmen, La Preciosa y Quechulac). Como se observa, la red de expansión mínima optimiza la visualización de las relaciones existentes entre las secuencias de ADN de cada uno de los individuos que componen las poblaciones y, a partir de ello, se puede inferir la estructura que se presenta entre las poblaciones. Este análisis está basado en la distancia genética y sus conexiones, la distancia o el grado de relación está representado por las líneas que une la red, líneas anchas significan una distancia genética con valores cercanos a 0, es decir, mayor parentesco, mientras que líneas más angostas significan una distancia genética mayor y una relación o parentesco menor entre individuos. De acuerdo con estos resultados, se observa que existe un grupo conformado por los individuos de la población de *L. "La Preciosa"* y los individuos de la población de *L. "Quechulac"*, aunque, en esta última hay 6 individuos que se alejan ligeramente de los demás.

En los individuos de la población de *L. "El Carmen"* se puede observar que la distancia entre ellos es muy baja entre la mayoría de ellos, formando un grupo compacto, aunque ocho individuos se distancian de manera importante. Los individuos de la población de *L. "Atexcac"* son los más dispersos en la red, aunque se agrupan más cercanamente entre sí que respecto a los individuos de las demás poblaciones.

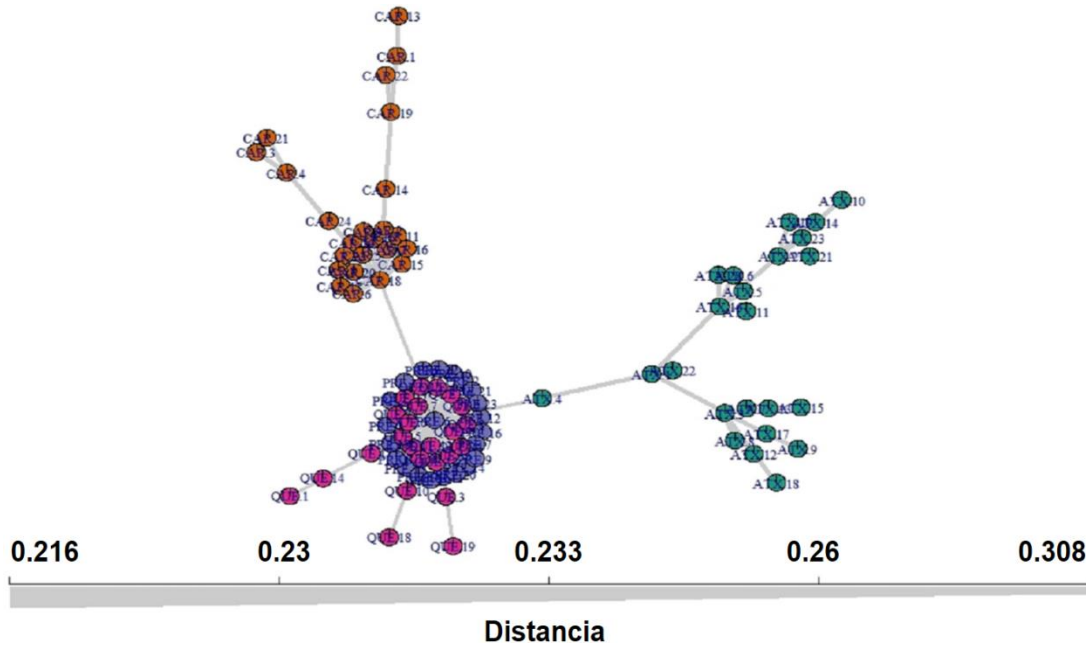


Figura 3. Red de expansión mínima de la relación genética entre los 93 individuos analizados: 24 individuos de *L. "Quechulac"* en color rosa (QUE), 24 individuos de *L. "El Carmen"* (CAR) en color naranja, 23 individuos de *L. "Atexcac"* (ATX) en color verde y 22 individuos del *L. "La Preciosa"* en color morado (PRE).

Detección de *loci* sujetos a selección

Las herramientas analíticas de todo el genoma detectan la selección al buscar patrones inusuales de variación genética que postulan que la historia demográfica de la población afecta la variación en todos los loci, mientras que la selección natural opera en loci específicos. Conocidas como pruebas de valores atípicos, las huellas de la selección se buscan al escanear el genoma, utilizando una medida de diferenciación basada en la población como F_{ST} o mediante un enfoque basado en el individuo, centrado en los modelos de inferencia bayesiana, como es el que se utiliza dentro del paquete *pcadapt* (Duforet *et al.*, 2015). En este trabajo el paquete *pcadapt* identificó 1,700 SNPs (Figura 4) dentro del conjunto de 25,870 SNPs de las cuatro poblaciones del copépodo con valores atípicos ($MAF > 0.05$), como posibles SNPs candidatos a estar bajo selección o conferir alguna ventaja bajo algún escenario selectivo particular.

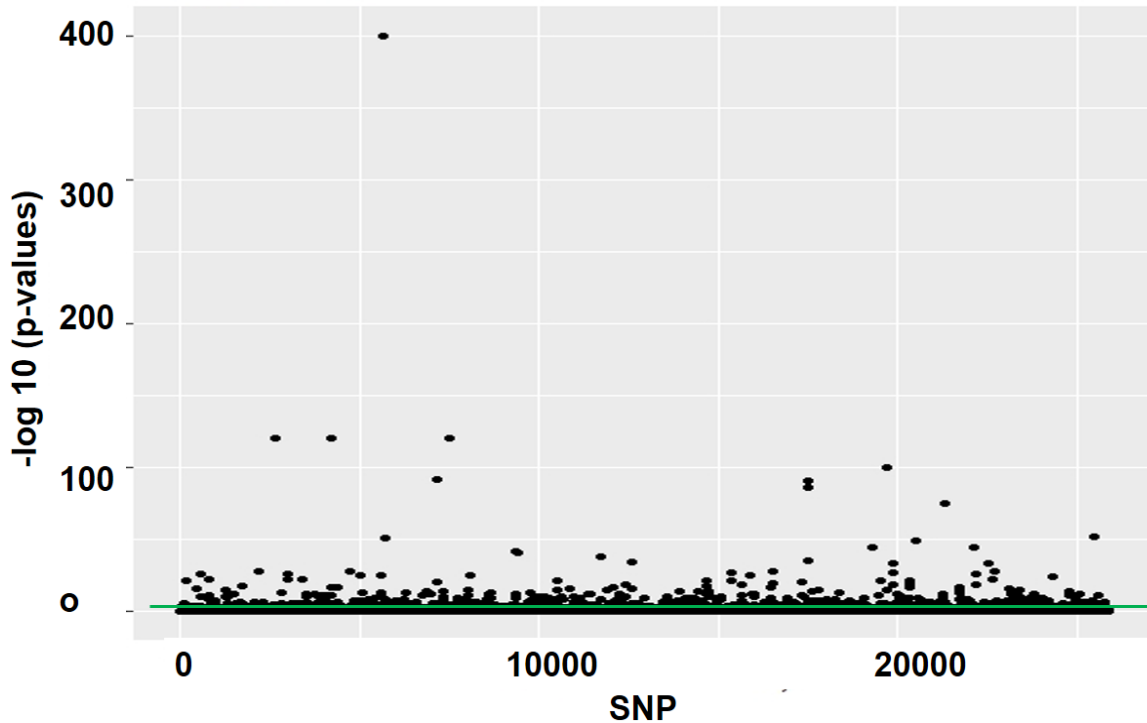


Figura 4. Diagrama de Manhattan que representa la distribución de los valores de significancia en todos los datos analizados (25,870 SNPs) $-\log_{10}(\text{p-value})$, y muestra a los 1,700 SNPs con valores MAF > 0.05, representada por la línea verde.

Análisis de Componentes Principales

En la Figura 5 se resumen los resultados del análisis de componentes principales, se grafican los primeros dos componentes (CP1 y CP2) que explican aproximadamente 19% y 28% de la variación genética, respectivamente. Con este análisis se revela el patrón de agrupación asumido en la hipótesis, con cuatro poblaciones (Atexcac, Carmen, Quechulac y Preciosa) claramente definidas por cuatro agrupamientos. Ambos ejes separan a la población de *L. "El Carmen"* respecto a las demás, y coloca a las poblaciones de *L. "Quechulac"* y *L. "La Preciosa"* cercanamente relacionadas entre sí. Estos resultados de agrupación que revelaron los CP son muy similares entre los análisis de los 25,870 SNPs (ver Figura 5A) y los 1,700 SNPs con valores atípicos (ver Figura 5B). Finalmente, se puede visualizar a *L. "Atexcac"* separada de las otras poblaciones a lo largo del CP 2 en ambos análisis.

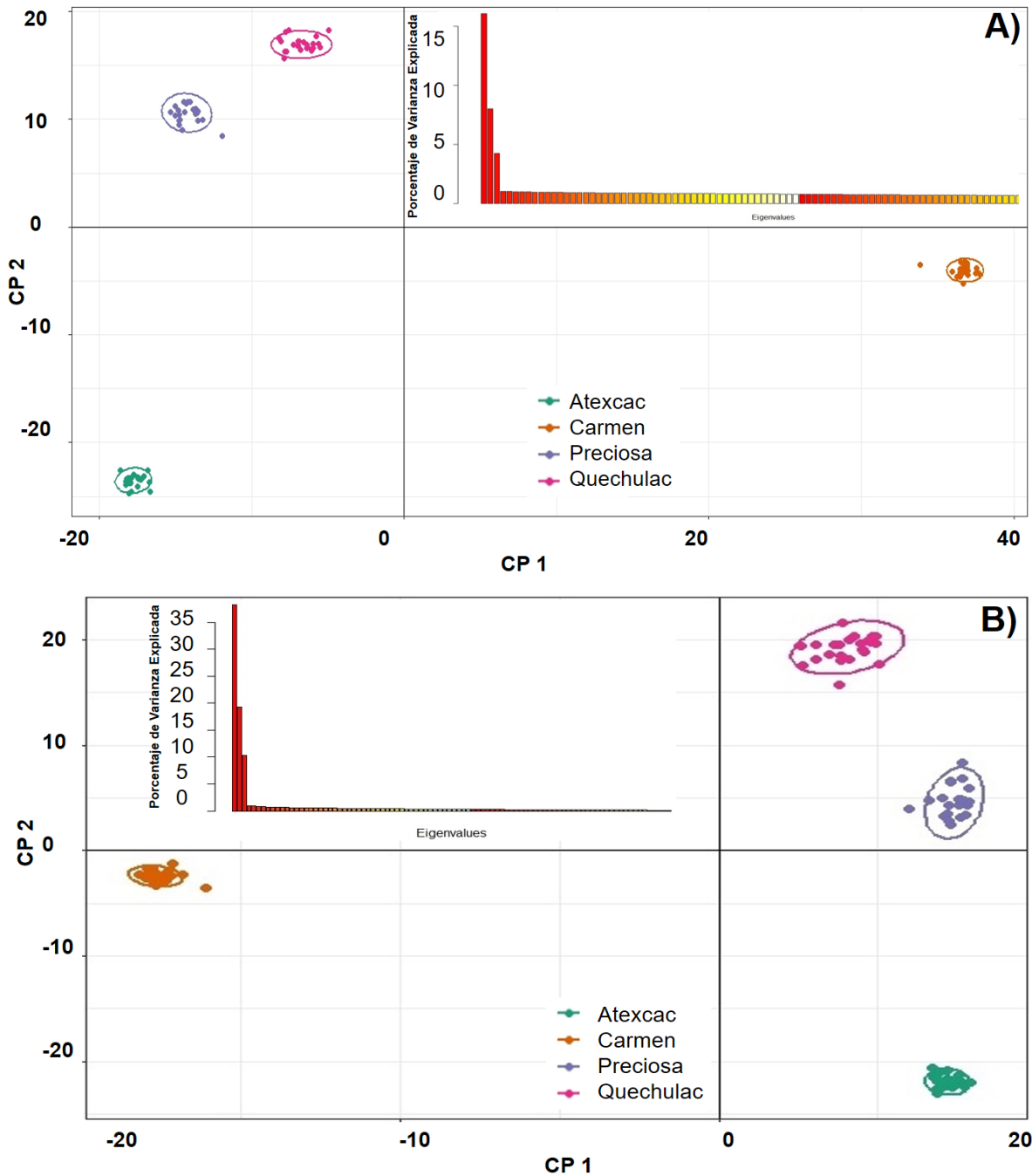


Figura 5. Resultado del análisis de componentes principales. Los 2 ejes corresponden a los dos primeros componentes (CP 1 y CP 2) que explican la mayor parte de la variación molecular (~30 %; véase la figura inserta análisis con los autovalores calculados para los componentes generados por el análisis) de los 93 individuos analizados, en la parte superior (A) con 25,870 SNPs y en la parte inferior (B) con los 1,700 SNPs atípicos. Los elipsoides corresponden al 95% de intervalo de confianza para cada población. Las poblaciones se representan con diferentes colores.

Análisis de Estructura genómica

La evaluación de asignación de grupos (Figura 6) entre los 93 individuos con 25,870 SNPs se realizó sobre la base de la validación del mínimo error de cruzamiento y asumiendo *a priori* la posibilidad de diferente número de poblaciones, $K = 1$ a $K = 5$. Los resultados muestran que la validación del mínimo error de cruzamiento (CV) se encuentra cuando $K = 4$, lo que sugiere que el número más probable de poblaciones que conforman la estructura genética del conjunto de datos del copépodo *Leptodiptomus* es de cuatro.

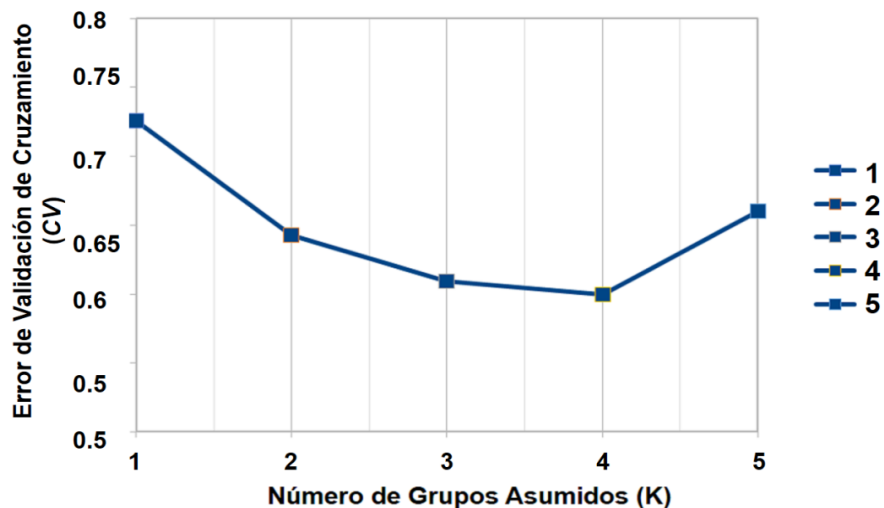


Figura 6. Validación de Cruzamiento para la estimación de grupos dentro del conjunto de datos de 25,870 SNPs para valores de K que van de 1 a 5.

Los resultados de la cuantificación de la máxima probabilidad de pertenencia de cada individuo en 4 poblaciones ancestrales (valor de K) asumidas por la validación del mínimo error de cruzamiento muestran una estructura genética definida (Fig. 6). La asignación de los individuos en el primer grupo (barras color azul; Fig. 7) lo componen 23 individuos de *L.* “Atexcac” con una máxima probabilidad de pertenencia de $Q \sim 1$ en promedio, en el segundo grupo (barras color amarillo) fueron asignados 24 individuos a *L.* “El Carmen”, de igual manera, con un valor de $Q \sim 1$ en promedio; el tercero (barras color verde) se compone de 22 individuos de *L.* “La Preciosa” y el cuarto grupo está conformado por 24 individuos de *L.* “Quechulac” (barras color rojo). En los dos últimos grupos se observa una ligera probabilidad de pertenencia a otra población ($> 5\%$), en el caso del grupo tres con 6 individuos afines a *L.* “Quechulac” y 5 individuos a *L.* “Atexcac”, mientras que el grupo cuatro tiene 5 individuos afines a *L.* “Atexcac”, 8 individuos a *L.* “El Carmen”, y 13 individuos a *L.* “La Preciosa”.

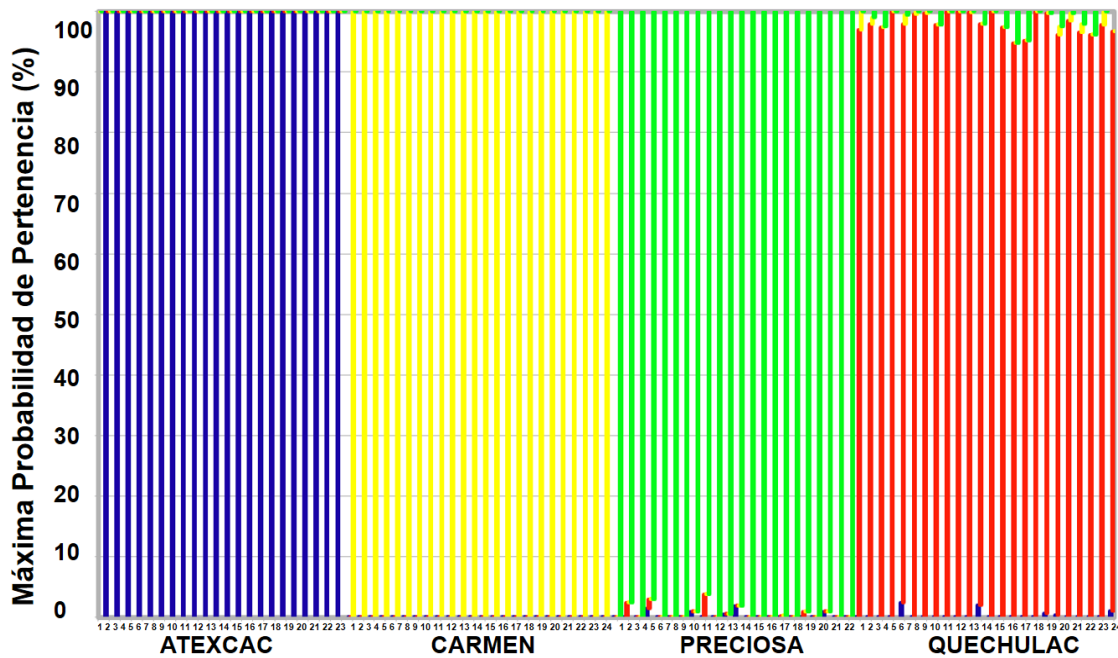


Figura 7. Estructura poblacional de los copépodos *L. grupo sicilis* de los lagos de Cuenca Oriental analizada para conjunto de datos de 25,870 SNPs con el paquete ADMIXTURE y la máxima probabilidad de pertenencia ($K = 4$); cada individuo analizado ($n = 93$) está representado por una barra vertical que muestra la probabilidad de pertenencia a los diferentes grupos; cada grupo o población está representado por un color diferente.

Análisis de Varianza Molecular (AMOVA) y Divergencia Genética (F_{ST})

La variación genética dentro de todo el conjunto de datos y entre pares de poblaciones se analizó con un AMOVA que compara la diversidad y variación genética del conjunto de los 25,870 SNPs (Figura 8A) y los 1,700 SNPs atípicos (Figura 8B) entre y dentro de las poblaciones, los resultados se resumen en la Figura 8. Se observa que el mayor porcentaje de variación de los SNPs totales se encuentra dentro de las poblaciones, es decir entre individuos de cada población, que entre poblaciones. Estos resultados son consistentes tanto para el conjunto total de poblaciones como en los análisis por pares, mostrando que no hay una estructura genética definida. Sin embargo, cuando se analizan los SNPs atípicos se observa que se invierten los resultados, es decir el mayor porcentaje de variación se encuentra entre las poblaciones, que, dentro de las poblaciones, y también son consistentes en el análisis con el conjunto total de poblaciones como en los análisis pareados, en este caso se muestra que existe una estructura genética definida. La estructura genética clara

también se observa en los análisis de diferenciación genética (F_{ST}) entre poblaciones en la Tabla 4.

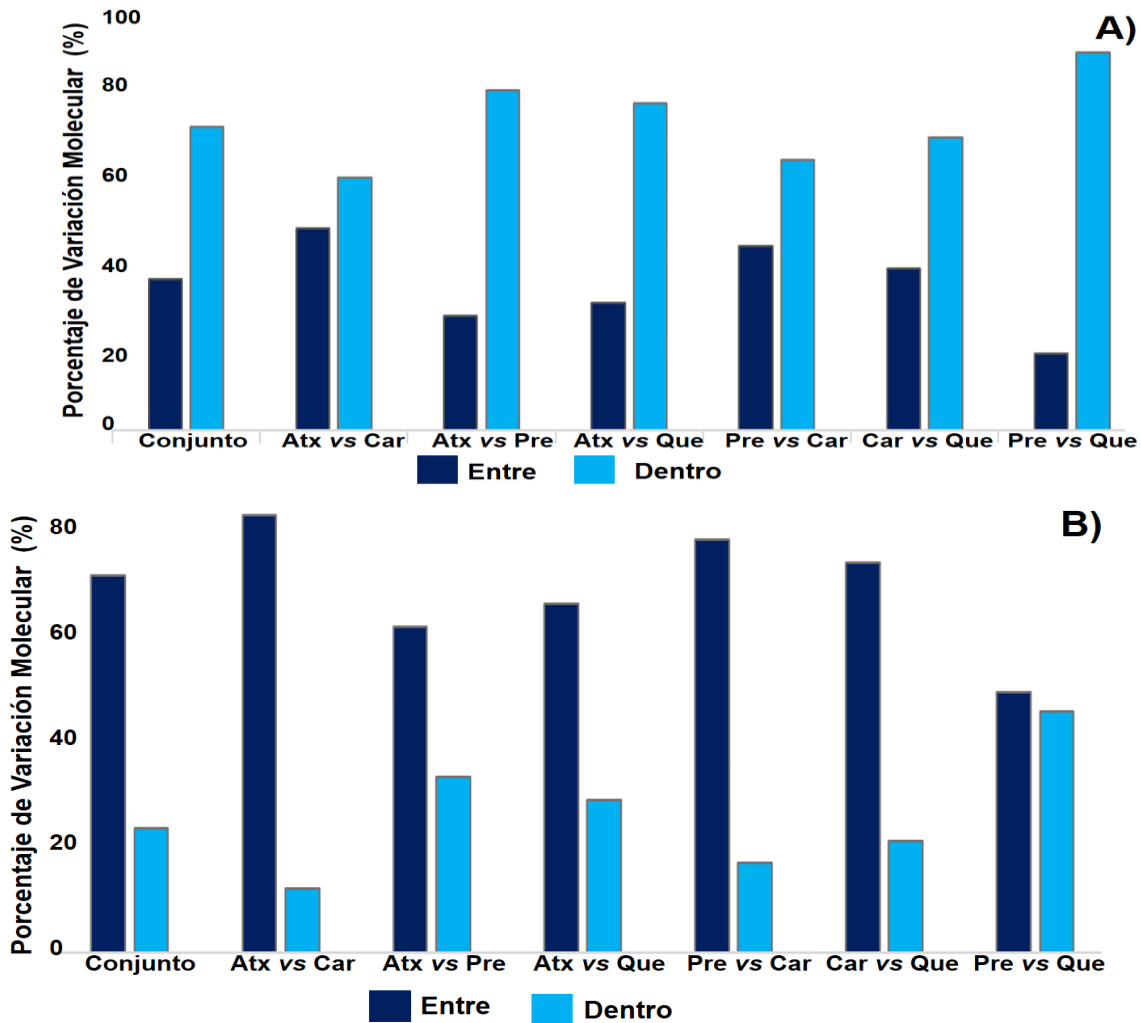


Figura 8. Resultados que exhiben los Porcentajes de la Varianza Molecular (AMOVA) entre y dentro de los grupos (poblaciones) constituidos por 93 individuos de los copépodos *L. grupo sicilis*, con los 25,870 SNPs en la parte superior (A) y en la parte inferior con 1,700 SNPs atípicos (B).

La Tabla 4A muestra que la mayor divergencia genética presente entre las poblaciones de *L. grupo sicilis* está entre las poblaciones de *L. "Atexcac"* y *L. "El Carmen"* ($F_{ST} = 0.2693$), y las que presentaron el valor mínimo de divergencia son las poblaciones de *L. "La Preciosa"* y *L. "Quechulac"* ($F_{ST} = 0.0858$), consistente con los datos previos. En el caso de la Tabla 4B que presenta los resultados de divergencia genética obtenidos a partir de los SNPs atípicos, mostraron un patrón similar entre poblaciones, aunque con valores más altos de divergencia que con el total, el mayor valor entre *L. "Atexcac"* y *L. "El*

Carmen” ($F_{ST} = 0.8063$) y el valor mínimo entre L. “La Preciosa” y L. “Quechulac” ($F_{ST} = 0.3861$).

A)	Atexcac	El Carmen	La Preciosa	Quechulac
Atexcac	-	-	-	-
El Carmen	0.2692	-	-	-
La Preciosa	0.1358	0.2448	-	-
Quechulac	0.1521	0.2084	0.0857	-

B)	Atexcac	El Carmen	La Preciosa	Quechulac
Atexcac	-	-	-	-
El Carmen	0.8063	-	-	-
La Preciosa	0.5303	0.7287	-	-
Quechulac	0.5664	0.6700	0.3860	-

Tabla 4. Estadísticos de divergencia genética (F_{ST}) pareados con 25,870 SNPs en la parte superior (A) y con 1,700 SNPs atípicos en la parte inferior (B).

Reconstrucción de las Relaciones Filogenéticas

Los análisis de la reconstrucción filogenética realizados bajo un modelo estándar de sustitución de nucleótidos “4by4” generaron un árbol filogenético de máxima credibilidad que se ilustra en la Figura 9, donde las líneas coloreadas relacionan a los individuos de cada grupo y los nodos principales con un soporte de probabilidad posterior alto (>0.9), donde también se muestra la comparación de las topologías de los árboles analizados con el conjunto de 25,870 SNPs (Figura 9A) y los 1,700 SNPs atípicos (Figura 9B).

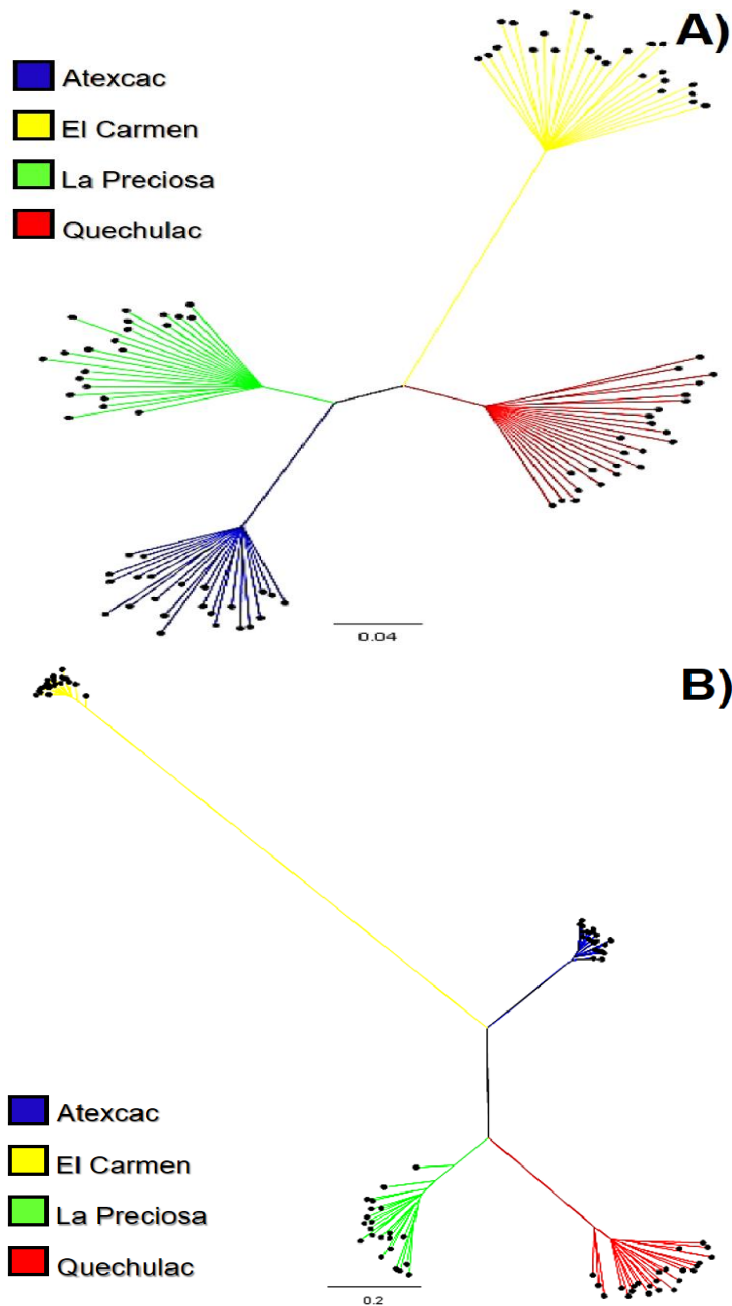


Figura 9. Reconstrucción de las relaciones filogenéticas de 93 secuencias compuestas de 25,870 SNPs (A) y 1,700 SNPs (B) de las cuatro poblaciones de los copépodos *L. grupo sicilis*. Las historias evolutivas se infirieron a partir del modelo evolutivo “4by4+GTR” sugerido para datos moleculares tipo SNPs.

También se muestra un resumen de los árboles filogenéticos muestreados a lo largo de los análisis de reconstruidos bajo el método agrupación UPGMA (Figura 10A, basado en 25,870 SNPs, y Figura 10B, basado en 1,700 SNPs atípicos) y que resalta la

incertidumbre dentro de los clados que se genera en busca del árbol más creíble y al mismo tiempo muestra la certeza de converger en una topología que delimita claramente entre grupos, generando 4 clados monofiléticos independientes, cada uno, correspondiente a cada una de las poblaciones asumidas *a priori*.

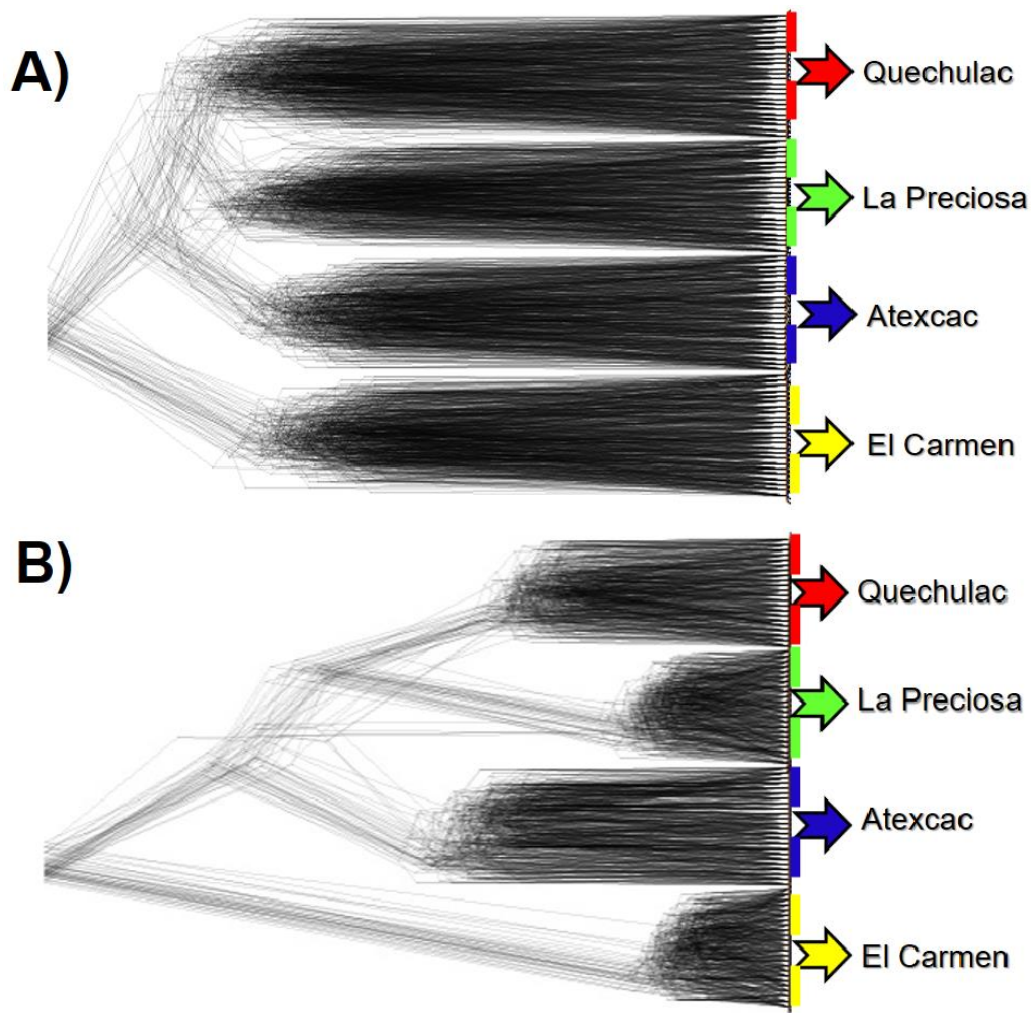


Figura 10. Reconstrucción de las relaciones filogenéticas de árboles consenso realizados con 93 secuencias compuestas de 25,870 SNPs (A) y 1,700 SNPs (B) de alta calidad que pertenecen a 4 poblaciones de los copépodos *L. grupo sicilis*.

El patrón de relaciones y divergencia genética es similar entre los análisis con los dos diferentes métodos de reconstrucción filogenética (Figura 9 y 10), ya que definen una divergencia genética clara entre las cuatro poblaciones, donde la distancia genética es más corta entre las poblaciones de *L. "La Preciosa"* y *L. "Quechulac"*, mientras que la población de *L. "Atexcac"* aparece un poco más distante, y, finalmente, la longitud de mayor distancia

pertenece a la población de *L. "El Carmen"*. Estos resultados se observan con mayor claridad en los análisis basados en los SNPs sujetos a selección, por lo cual se puede seguir afirmando que los loci incluidos dentro de este conjunto son fuertes candidatos para analizar los efectos de la selección, y que por lo tanto deben tener un papel importante en la estructuración de las poblaciones de los copépodos *L. grupo sicilis*, dentro de los cuales se deben encontrar regiones que han promovido la adaptación local y la evolución adaptativa.

Discusión

La alta diversidad biológica y ecológica de los copépodos ofrece ventajas para el desarrollo de análisis genómicos que ayudan a descifrar los mecanismos genéticos y fisiológicos asociados a su adaptación a ambientes divergentes (Bron *et al.*, 2011). Esta idea es fácil de entender ya que, los gradientes ambientales que se observan en los ambientes acuáticos continentales pueden actuar como factores de selección potenciales y generar como resultados variantes genéticas adaptadas localmente. Dadas estas circunstancias, el resultado se define como un cambio genético conducido por selección natural, ya que es producto de selección divergente, reforzado por aislamiento geográfico producto de la existencia de mecanismos de dispersión poco eficientes y por lo tanto baja posibilidad de colonización de los organismos que, finalmente, conduce a la explotación de nichos ecológicos alternativos (véase, p. e., Arnegard *et al.*, 2014). En el caso particular de los copépodos *L. grupo sicilis*, que se encuentra distribuido en cuatro poblaciones geográficamente aisladas (< 30 km de distancia entre poblaciones) y hábitats ecológicamente contrastantes, se ha reportado un aislamiento reproductivo entre al menos dos poblaciones (Atexcac vs. La Preciosa) y patrones de variación fisiológica y genética entre las cuatro poblaciones usando marcadores moleculares estructurales (véase, Barrera-Moreno *et al.*, 2015; Barrera-Sánchez, 2016).

Para reforzar la hipótesis que plantea una divergencia genética entre poblaciones como resultado de habitar sistemas ecológicos contrastantes, en esta tesis se realizó un análisis genómico, utilizando metodologías genómicas de última generación y los análisis clásicos utilizados en la genética de poblaciones para resolver las incertidumbres filogenéticas, y para caracterizar los patrones de diversidad genética que puedan proporcionar información valiosa sobre los procesos evolutivos que dan forma a la estructura genética de las poblaciones y, por lo tanto, que ayudan a mejorar nuestra

comprensión de como los organismos pueden responder en ambientes ecológicos contrastantes.

Diversidad genética

Este estudio de genómica poblacional de los copépodos *L.* grupo *sicilis* de la Cuenca Oriental, muestra que la diversidad genética (H_o , π , θ y R) intrapoblacional en general es relativamente alta, aunque *L.* “Quechulac” es la población que tiene la diversidad genética más alta, seguido de la población de *L.* “La Preciosa”. Además, en todos los casos, las poblaciones tienen valores de D de Tajima > 0 , lo que sugiere un incremento de la heterocigosidad promedio, y que las poblaciones han estado bajo presiones de selección positiva o negativa. La población que presentó el nivel de diversidad más bajo es la de *L.* “El Carmen” respecto a las demás poblaciones, esta posible reducción de la diversidad genética intrapoblacional podría ser el reflejo de efectos de flujo genético limitado y eventos de deriva genética recientes, promovidos por las condiciones ecológicas particulares de su nicho, recordando que habitan en lago estacional tipo playa, con desecación temporal, alta variabilidad salina, efectos antropogénicos y fragmentación de hábitat, lo que podría limitar la dispersión y reproducción entre individuos, aunque podría deberse a restricciones debidas al muestreo. Sin embargo, en general los niveles de diversidad genética fueron similares en todas las poblaciones, y altos respecto a otros estudios de poblaciones de copépodos (véase, p.e., Young *et al.*, 2013; Kochanova *et al.*, 2018; Wang *et al.*, 2018), así que estos parámetros pueden ser un punto de referencia para estudios más específicos de cada población.

Estructura Genética

Red de expansión mínima

Los resultados de distancia y estructura genética mostraron consistencia, definiendo cuatro poblaciones genómicamente bien diferenciadas. La composición de la red de distancia genética medida a través de la variabilidad genética entre poblaciones mostró una estrecha relación en la asociación genética entre las poblaciones de *L.* “La Preciosa” y *L.* “Quechulac”, estas poblaciones habitan a una menor distancia geográfica, pero más importante, habitan lagos perennes, con la presencia de peces y de baja salinidad (La

Preciosa = $1.18 \pm 0.9 \text{ g L}^{-1}$ y Quechulac = $0.42 \pm 0.05 \text{ g L}^{-1}$). La población de *L. "El Carmen"* se mostró como el agrupamiento que claramente está distante respecto de los demás, la cual puede entenderse dado que esta población se encuentra geográficamente más distante del resto, además de que, al habitar en el único lago efímero, anualmente presenta diferencias ecológicas considerables, especialmente en la permanencia de la columna de agua y en el gradiente de salinidad (1.4 a 14 g L^{-1}). La población *L. "Atexcac"* mostró tener una asociación media respecto a las *L. "La Preciosa"* y *L. "Quechulac"*, y más distante de *L. "El Carmen"*, cabe mencionar que esta población habita un lago perenne, con una mayor estabilidad ecológica, aunque tiene la salinidad más alta de todos los lagos ($6.54 \pm 0.29 \text{ g L}^{-1}$) y no se ve sujeta a la presión de depredación por peces como el caso de las otras poblaciones que habitan lagos perennes. La distancia genética y las condiciones de salinidad contrastante parecen estar asociados *a priori*, como lo proponen Barrera-Moreno *et al.* (2015) con patrones de divergencia asociada al uso diferencial de nichos ecológicamente contrastantes, lo cual coincide con la literatura especializada, donde se ha descrito una divergencia y expresión genética diferencial en respuesta al estrés en gradientes de salinidad y temperatura en copépodos (Pereira *et al.*, 2017; DeBiasse *et al.*, 2018).

Detección de *Loci* sujetos a selección

Los análisis de valores atípicos (*outliers*) pueden ser una herramienta poderosa que sirve para detectar loci que están bajo selección, particularmente cuando los niveles de diferenciación genómica de fondo son bajos o ambiguos (Tigano *et al.*, 2017) o bien, para poder aclarar incertidumbres y por tanto reforzar hipótesis de divergencia de las poblaciones producto de selección natural. La proporción de valores atípicos detectados a partir de un conjunto de 25,870 SNPs fue del 6.5% es decir, 1,700 SNPs que, en consecuencia, muy probablemente son reflejo de la selección diversificadora y no simplemente por la posibilidad de efectos estocásticos, lo que los convierte en candidatos para buscar marcadores asociados a factores adaptativos. Este conjunto de *loci* sujetos a selección permitió obtener una evidencia más sólida de la posible existencia de selección divergente operando sobre las poblaciones de los copépodos *L. grupo sicilis*, esto se pudo observar en los análisis comparados de diferenciación genética poblacional basados en los SNPs totales vs. los SNPs atípicos. Vale la pena mencionar que también puede haber falsos positivos dentro de nuestros loci candidatos y podrían representar

procesos estocásticos o vinculados con otros candidatos (Cullingham *et al.*, 2014). Por lo tanto, los valores atípicos en las poblaciones de este copépodo requieren de un mayor estudio y tratar de asociarlos a un genoma de referencia en busca de las rutas metabólicas asociadas.

Análisis de componentes principales

El análisis de componentes principales agrupó en solo dos componentes (CP1 y CP2) el ~30% de la variabilidad molecular total de los individuos analizados, tanto para los SNPs totales como con los atípicos, mostraron claramente cuatro grupos. El CP1 separó a todos los grupos, aunque los individuos del *L. "El Carmen"* se separaron más ampliamente del resto, mientras que el CP2 separó claramente al grupo de *L. "Atexcac"* con respecto a los grupos *L. "La Preciosa"* y *L. "Quechulac"*, estos últimos se encuentran más cercanos a lo largo de este eje. Sin embargo, en el caso del análisis con los datos atípicos, se observaron las agrupaciones más distantes que con el análisis de todos los SNPs, aunque, en conjunto, las poblaciones se comportan como unidades con cohesión genética independiente y una amplia distancia genética respecto a las demás poblaciones en este espacio bidimensional, siendo esto un indicativo de una fuerte división y estructura genética definida.

Análisis de estructura genómica

Este análisis se realizó únicamente con los 25,870 SNPs totales y estuvo basado en un método de inferencia bayesiana que resolvió el número óptimo de agrupación genética con una $K = 4$ soportada por la validación del mínimo error de cruzamiento (CV), este resultado dio un soporte mayor a la hipótesis de divergencia genética entre poblaciones por la diferenciación de nicho ecológico, modelado por la selección divergente. Estos resultados permitieron visualizar de manera más clara la estructura genética con valores de pertenencia altos (Fig. 6), por lo que se puede asumir que cada grupo se comporta como una unidad evolutiva, potencialmente independiente, pues la probabilidad de mezcla es casi nula con una pertenencia cercana al 100 % en las poblaciones de *L. "El Carmen"* y *L. "Atexcac"*, mientras que los individuos de *L. "La Preciosa"* y *L. "Quechulac"* mostraron una probabilidad de pertenencia ~ 95%. Así, a pesar de que geográficamente los lagos se encuentran a una distancia relativamente corta (< 30 km) y que su capacidad de dispersión de sus estructuras de resistencia podría ser suficiente, los resultados revelados por el

análisis de estructura genética indican que la probabilidad de que nuestras poblaciones mantengan eventos de migración, colonización y flujo genético reciente es casi nula y, por lo tanto, se encuentran en un proceso avanzado de divergencia.

Análisis de Varianza Molecular (AMOVA) y Divergencia Genética (F_{ST})

Cuando se utilizaron todos los SNPs en el análisis jerárquico de variación molecular se observa que la mayor variación genética se encuentra dentro de los grupos y la menor variación entre grupos, este pareciera romper con la hipótesis de divergencia ecológica y adaptación local en las poblaciones de los copépodos *L. grupo sicilis* que se ha demostrado en este trabajo. Sin embargo, cuando se utilizaron los SNPs atípicos se invirtieron los resultados, teniendo la mayor variación entre poblaciones y no dentro de las poblaciones. Además, de igual manera, la diferenciación genética (F_{ST}) es más fuerte cuando se utilizaron los marcadores atípicos, aunque ambos resultados muestran que las poblaciones han divergido y se ve reforzada la hipótesis de una relación más estrecha entre los copépodos que habitan los lagos más parecidos en salinidad. Así, a primera vista pareciera que las altas diferencias genéticas entre poblaciones están explicadas por los SNPs neutrales, y que las poblaciones han estado evolucionando de manera neutral. Sin embargo, en la búsqueda de marcadores atípicos, se detectó una diferenciación más convincente de las poblaciones sobre la base de los SNPs que podría estar asociados a función y eventualmente fijados por divergencia ecológica y adaptación local.

Reconstrucción de las Relaciones Filogenéticas

Los análisis filogenéticos de los 93 individuos con los 25,870 y 1,700 SNPs de individuos de *L. grupo sicilis* revelaron, de manera consistente, la divergencia de las cuatro poblaciones ya reconocidas con los análisis descritos previamente. Los árboles filogenéticos basados en dos diferentes métodos mostraron resultados similares, identificando cuatro linajes monofiléticos, cada uno correspondiente a nuestras cuatro poblaciones, manteniendo el mismo patrón de divergencia, es decir, a las poblaciones de *L. "La Preciosa"* y *L. "Quechulac"* con la relación más estrecha, a la población *L. "Atexcac"* como un linaje intermedio y, finalmente, a la población de *L. "El Carmen"* como el grupo más alejado. Esto nos permite inferir que las poblaciones han estado sujetas a diferentes

presiones ambientales, experimentando posiblemente diferentes eventos de deriva genética, a través, de la fijación y pérdida aleatoria de variantes genéticas especialmente neutrales y que la selección natural divergente es el mecanismo evolutivo que mejor explica la estructura genética de estos copépodos, a través de las variantes y/o regiones genéticas particulares, las cuales posiblemente se encuentran dentro del conjunto de *loci* sujetos a selección (1700 SNPs) y que podrían ser responsables de alterar la expresión genética, la estructura y función de las proteínas en diferentes entornos, afectando la eficacia individual por selección natural, conduciendo a adaptación local (Pereira *et al.*, 2017). Lo que es más evidente en las relaciones filogenéticas basadas en los SNPs atípicos, pues, aunque tienen un patrón similar de relaciones filogenéticas mostrado con los SNPs totales, estos árboles filogenéticos presentan ramas más prolongadas, además, la agrupación de los cuatro clados monofiléticos es más compacta, mostrando relaciones más estrechas dentro de grupo poblaciones y más distantes con respecto de los otros grupos.

Discusión general e implicaciones ecológicas

El conjunto de factores ambientales heterogéneos, la alta diversidad genética intrapoblacional, el tamaño efectivo de la población, tasas de mutación constantes, el aislamiento geográfico, la deriva genética y la acción de la selección natural divergente, son promotores importantes para la diferenciación genética entre poblaciones. En esta tesis se analizó la estructura y divergencia genética entre las poblaciones del copépodo *L. grupo sicilis*, donde los análisis de estructura genética a nivel genómico son el punto de partida para lograr una mejor comprensión del efecto potencial de los procesos evolutivos de la migración, el flujo de genes, el sistema de apareamiento, la selección, la adaptación y la deriva genética en las poblaciones (véase, p.e., Meimars, 2012).

El genotipado por secuenciación realizado aquí ha demostrado ser útil y confiable para la identificación de SNPs de alta calidad, así como para la identificación de ~6% de SNPs con valores atípicos que se utilizaron para investigar la diversidad genética y definir la estructura de las poblaciones de estos copépodos. La combinación de herramientas de agrupamiento bayesiano y jerárquico demostró ser eficaz para dilucidar la estructura genética de las poblaciones, los resultados mostraron una estructura genómica definida con niveles de diversidad altos en cada una de las poblaciones, sugiriendo un flujo genético restringido con historias evolutivas independientes, resultado de tasas de dispersión bajas

y condiciones ecológicas diferenciales que ejercen altas presiones de selección importantes, lo que mantiene aislando a los genotipos a sus ambientes particulares.

Las relaciones filogenéticas entre estas poblaciones sugieren una divergencia asociada con las condiciones ambientales de los lagos (particularmente con los gradientes de salinidad), sugiriendo adaptación local modelada en un inicio y reforzada por selección divergente, separando a las poblaciones del copépodo *L. grupo sicilis* en cuatro linajes, cada uno adaptado a su condición salina, *L. "El Carmen"* que se halla en un hábitat con salinidad variable, *L. "Atexcac"* que se encuentra en el ambiente más salino, y las dos restantes en ambientes dulceacuícolas *L. "La Preciosa"* y *L. "Quechulac"*, demostrando que el ambiente es un predictor importante de la diferenciación genética en poblaciones de copéodos. Además, existe la probabilidad que alguno de los *loci* sujetos a selección (1700 SNPs) pueda estar asociados a genotipos particulares, estos a fenotipos especializados y ambos a nichos ecológicos particulares, que se asocien a la adaptación local por variantes genéticas especializadas, por lo que podrían amortiguan la entrada de nuevos genotipos y reducir la probabilidad de establecimiento de migrantes en hábitats diferentes al suyo (De Mester *et al.*, 2002; Alcántara *et al.*, 2012; Matute, 2013). Al mismo tiempo, los resultados presentados en esta tesis, sumados a los de Barrera-Moreno *et al.* (2015) y Barrera-Sánchez (2016), refuerzan la interpretación de que ha sido principalmente la selección divergente la que ha modelado la estructura genómica de las poblaciones de estos copéodos, conduciendo al aislamiento reproductivo precigótico por inviabilidad de los migrantes, y post-cigótico por inviabilidad del híbrido en al menos dos linajes, y que en última instancia han conducido a la especiación ecológica.

Los recursos genómicos obtenidos aquí serán útiles para análisis posteriores que ayuden a clarificar mejor el papel de la selección en el proceso de divergencia genética que se ha demostrado aquí y, si está divergencia a lo largo de un gradiente ambiental puede ser indicativo de adaptación local (Kawecki y Ebert, 2004; Savolainen *et al.*, 2013). Bajo este enfoque, posteriormente se deberá continuar con la investigación que aclare los factores genéticos y ecológicos que han dado lugar a la adaptación local en estas poblaciones, utilizando métodos de exploración genómica más especializados que puedan identificar genotipos asociados a rasgos selectivos con función genética como base de la adaptación local en cada uno de los contextos ecológicos de las poblaciones de estos copéodos. Actualmente, se sabe que la exploración genómica ha permitido identificar las

huellas genéticas de la adaptación local en poblaciones naturales, incluso en ausencia de hipótesis planteadas *a priori* sobre los factores selectivos y los rasgos específicos implicados (Savolainen *et al.*, 2013; Tiffin y Ross-Ibarra, 2015), por lo que este modelo de estudio parece ser muy prometedor para entender los procesos que han dado lugar a estos patrones de divergencia biológica.

Conclusiones

Se logró analizar la diversidad y estructura genética de cuatro poblaciones del copépodo *L. grupo sicilis* a partir de un método de GBS y la identificación de sitios altamente polimórficos con marcadores moleculares tipo SNPs de alta calidad. Se describió la diversidad genética de cada población, que es relativamente alta en todas las poblaciones. Adicionalmente, se identificó un conjunto de marcadores atípicos (*loci* sujetos a selección) que ayudó a mejorar los análisis de estructura interpoblacional, revelando una estructura particular para cada población de marcadores potencialmente asociados a la divergencia ecológica y la adaptación local.

Las relaciones filogenéticas fueron consistentes con la hipótesis de trabajo, mostrando que las poblaciones afines del copépodo *L. grupo sicilis* han divergido en 4 clados independientes (*L. "Atexcac"*, *L. "El Carmen"*, *L. "La Preciosa"* y *L. "Quechulac"*) y perfectamente definidos y asociados a divergencia de hábitat, lo que sugiere que son resultado de historias evolutivas independientes promovidas principalmente por la selección natural divergente y nichos ecológicos contrastantes.

Literatura Citada

- Alexander, D. H., Novembre, J., y Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655-1664.
- Alcántara-Rodríguez, J. A., Ciroso-Pérez, J., Ortega-Mayagoitia, E., Serranía-Soto, C. y Piedra-Ibarra, E. 2012. Local adaptation in populations of a *Brachionus* group *plicatilis* cryptic species inhabiting three deep crater lakes in Central Mexico. *Freshwater Biology*, 57, 728-740.
- Álvarez, J. 1950. Contribución al conocimiento de los peces de la región de los Llanos, Estado de Puebla (México). *Anales de la Escuela Nacional de Ciencias Biológicas*, IPN, 6 (1-4), 81-107.
- Andrews, C. A. 2010. Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. *Nature Education Knowledge* 3(10), 5. Obtenido de <https://www.nature.com/scitable/knowledge/library/natural-selection-genetic-drift-and-gene-flow-15186648>
- Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., Kabir, S y Kingsley, D. M. 2014. Genetics of ecological divergence during speciation. *Nature*, 511(7509), 307.
- Arredondo, J.L. 1995. Los axalapasos de la Cuenca Oriental, Puebla. *En: De la Lanza-Espino, G. y J. L. García-Calderón (eds.). Lagos y Presas de México*. Centro de Ecología y Desarrollo, México, DF. pp. 65-87.
- Bamshad, M., Wooding, S., Salisbury, B. A., y Stephens, J. C. 2004. Deconstructing the relationship between genetics and race. *Nature Reviews Genetics*, 5(8), 598.
- Barrera-Moreno, O. 2010. Análisis de las poblaciones de los copépodos *Leptodiatomus* cf. *sicilis* (Copepoda: Calanoida) en los lagos de la Cuenca Oriental, México. Tesis de Maestría (Limnología). Universidad Nacional Autónoma de México, México.

- Barrera-Moreno, O.A., Ciro-Pérez, J., Ortega-Mayagoitia, E., Alcántara-Rodríguez, J.A. y Piedra-Ibarra, E. 2015. From local adaptation to ecological speciation in copepod populations from neighboring lakes. *PLoS ONE*, 10(4), e0125524.
- Barrick, J. E. y Lenski, R. E. 2013. Genome dynamics during experimental evolution. *Nature Reviews Genetics*, 14(12).
- Beaumont, M. A. y Nichols, R. A. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B*, 263, 1619–1626.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. y Buckler, E. S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633-2635.
- Bron, J. E., Frisch, D., Goetze, E., Johnson, S. C., Lee, C. E. y Wyngaard, G. A. 2011. Observing copepods through a genomic lens. *Frontiers in Zoology*, 8(1), p. 22.
- Boxshall, G.A., Defaye, D. 2008. Global diversity of copepods (Crustacea: Copepoda) in freshwater. *Hydrobiologia*, 595,195-207.
- Caballero, M., Vilaclara, G., Rodríguez, A. y Juárez, D. 2003. Short-term climatic change in lake sediments from lake Alchichica, Oriental, Mexico. *Geofísica Internacional*, 42(3), 529-537.
- Carlin, J. L. 2011. Mutations Are the Raw Materials of Evolution. *Nature Education Knowledge*, 3(10),10. Obtenido de <https://www.nature.com/scitable/knowledge/library/mutations-are-the-raw-materials-of-evolution-17395346>
- Carrasco, N. G., Ort M.H. y Romero C. 2007. Evolution and hydrological conditions of a maar volcano (Atexcac crater, Eastern México) *Journal of Volcanology and Geothermal Research*, 159, 179-197.
- Coyne, J.A. y Orr, H.A. 2004. *Speciation*. Sinauer Associates, Inc., Sunderland, MA.

- Cowardin, L. M., Carter, V., Golet, F. C. y LaRoe, ET. 1979. Classification of wetlands and deepwater habitats of the United States. *U.S. Fish and Wildlife Service*, 79(31), p. 131.
- Cristescu, M. E., Adamowicz, S. J., Vaillant, J. J. y Haffner, D. G. 2010. Ancient lakes revisited: from the ecology to the genetics of speciation. *Molecular Ecology*, 19(22), 4837-4851.
- Cullingham, C. I., Cooke, J. E. y Coltman, D. W. 2014. Cross-species outlier detection reveals different evolutionary pressures between sister species. *New Phytologist*, 204(1), 215-229.
- Darwin, C. (2009). *El origen de las especies por medio de la selección natural*. Editorial CSIC-CSIC Press.
- Davies, S.J., Metcalfe, S.E., Caballero, M. y Juggins, S. 2002. Developing diatom-based transfer functions for Central Mexican lakes. *Hydrobiologia*, 467, 199-213.
- DeBiasse, M. B., Kawji, Y. y Kelly, M. W. 2018. Phenotypic and transcriptomic responses to salinity stress across genetically and geographically divergent *Tigriopus californicus* populations. *Molecular Ecology*, 27(7), 1621-1632.
- De Meester, L., Gómez, A., Okamura, B. y Schwenk, K. 2002. La monopolización hipótesis y la paradoja flujo de dispersión de genes en organismos acuáticos. *Acta Oecologica*, 23, 121-35.
- Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G. y Middelburg, J. J. 2006. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography*, 51(5), 2388-2397.
- Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, E. y Blum, M. G. 2015. Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 Genomes data. *Molecular Biology and Evolution*, 33(4), 1082-1093.

- Ellegren, H. y Galtier, N. 2016. Determinants of genetic diversity. *Nature Reviews. Genetic*, 17(7), 22-433.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. y Mitchell, S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), e19379.
- Eguiarte, L. E., Aguirre-Liguori, J. A., Jardón-Barbolla, L., Aguirre-Planter, E. y Souza, V. 2013. Genómica de poblaciones: nada en evolución va a tener sentido si no es a la luz de la genómica, y nada en genómica tendrá sentido si no es a la luz de la evolución. *TIP Revista Especializada En Ciencias Químico-Biológicas*, 16(1), 42-56.
- Escalante, T., Rodríguez, G., Gámez, N., León-Paniagua, L., Barrera, O. y Sánchez-Cordero, V. 2007. Biogeografía y conservación de los mamíferos. *Biodiversidad de la faja volcánica transmexicana*, (ed. by I. Luna, J.J. Morrone and D. Espinosa), pp. 485–502. Las Prensas de Ciencias, UNAM, Mexico City.
- Galassi, D. M., Huys, R. y Reid, J. W. 2009. Diversity, ecology and evolution of groundwater copepods. *Freshwater Biology*, 54(4), 691-708.
- Gavrilets, S. 2004. *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, NJ.
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. y Buckler, E. S. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*, 9(2), e90346.
- Hastie, L. C., Rasmussen, J., Angel, M. V., Boxshall, G. A., Chambers, S. J., Conway, D. V. P. y Wootton, M. 2015. Zooplankton identification manual for north European seas (ZIMNES). In M. F. Watson, C. Lyal, y C. Pendry (Eds.), *Descriptive Taxonomy: The Foundation of Biodiversity Research*, (pp. 107-110). Cambridge: Cambridge University Press.

- Heino, J. 2011. A macroecological perspective of diversity patterns in the freshwater realm. *Freshwater Biology*, 56, 1703-1722.
- Hoverman, J. T. y Johnson, P. T.J. 2012. Ponds and Lakes: A Journey Through the Life Aquatic. *Nature Education Knowledge*, 3(6),17.
- Jones, F. C., Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J. y Birney, E. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484, 55-61.
- Kamvar, Z. N., Brooks, J. C. y Grünwald, N. J. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6, p. 208.
- Kawecki, T. J., y Ebert, D. 2004. Conceptual issues in local adaptation. *Ecology Letters*, 7(12), 1225-1241.
- Kumar, S., Stecher, G. y Tamura, K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870-1874.
- Luu, K., Bazin, E. y Blum, M. G. 2017. pcadapt: anR package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67-77.
- Marrone, F., Lo Brutto, S., Hundsdoerfer, A. K. y Arculeo, M. 2013. Overlooked cryptic endemism in copepods: Systematics and natural history of the calanoid subgenus *Occidodiaptomus* Borutzky 1991 (Copepoda, Calanoida, Diaptomidae). *Molecular Phylogenetic Evolution*, 66, 190-202.
- Meirmans, P. G. 2012. AMOVA-based clustering of population genetic data. *Journal of Heredity*, 103(5), 744-750.

- Montero-Pau, J., Gómez, A. y Muñoz, J. 2008. Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. *Limnology and Oceanography Methods*, 6(6), 218-222.
- Hastie, L. C., Rasmussen, J., Angel, M. V., Boxshall, G. A., Chambers, S. J., Conway, D. V. P., y Wootton, M. 2015. Zooplankton identification manual for north European seas (ZIMNES). *Descriptive Taxonomy: The Foundation of Biodiversity Research*, 107.
- Nei M. y Kumar S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Norrgard, K. y Schultz, J. 2008. Using SNP data to examine human phenotypic differences. *Nature Education*, 1(1), 85.
- Pereira, R. J., Sasaki, M. C. y Burton, R. S. 2017. Adaptation to a latitudinal thermal gradient within a widespread copepod species: the contributions of genetic divergence and phenotypic plasticity. [*Philosophical Transactions of the Royal Society B: Biological Sciences*](#), 284(1853), 20170236.
- Porras, H. L., Ruiz, L., y Santos, C., Phillips, C., Carracedo, Á. y Lauren, M. V. 2013 An overview of STRUCTURE: applications parameter settings and supporting software. *Frontiers in Genetics*, 4, p. 98.
- Pritchard, J. K., Stephens, M. y Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.
- Klug, W. S. y Cummings, M. R. 1996. *Essentials of genetics* (No. Ed. 2). Obtenido de <https://www.nature.com/scitable/ebooks/essentials-of-genetics-8/118523195>
- Kochanova, E., Sarvala, J. y Fefilova, E. 2018. Phylogenetic Study of Dioecious and Parthenogenetic Populations of *Canthocamptus staphylinus* (Crustacea, Copepoda, Harpacticoida). *Bulletin, Southern California Academy of Sciences*, 117(2), 138-150.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Rice, A. M., Rudh, A., Ellegren, H. y Qvarnström, A. 2011. A guide to the genomics of ecological speciation in natural animal populations. *Ecology letters*, 14(1), 9-18.

Ringnér, M. 2008. What is principal component analysis? *Nature Biotechnology*, 26(3), 303.

Rundle, H. D. y Nosil, P. 2005. Ecological speciation. *Ecology Letters*, 8(3), 336-352.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. y Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61, 539–542.

Saitou, N. y Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425.

Salanti, G., Amountza, G., Ntzani, E. E., y Ioannidis, J. P. 2005. Hardy–Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *European journal of human genetics*, 13(7), 840.

Savolainen, O., Lascoux, M. y Merilä, J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11).

Scheffer, M. y Van Nes, E. H. 2007. Shallow lakes theory revisited: various alternative regimes driven by climate, nutrients, depth and lake size. *Hydrobiologia*, 584(1), 455-466.

Schluter, D. 2000. *The ecology of adaptive radiation*. OUP Oxford.

Schnack-Schiel, S. B. 2001. Aspects of the study of the life cycles of Antarctic copepods. In *Copepoda: Developments in Ecology, Biology and Systematics*, (pp. 9-24). Springer, Dordrecht.

- Seebens, H., Einsle, U. y Straile, D. 2009. Copepod life cycle adaptations and success in response to phytoplankton spring bloom phenology. *Global Change Biology*, 15(6), 1394-1404.
- Steinberg, K. M., Antonacci, F., Sudmant, P. H., Kidd, J. M., Campbell, C. D., Vives, L. y Lema, G. 2012. Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nature Genetics*, 44(8), 872-80.
- SOKI Wiki 2014 Profile: Copepods - Biota - Confluence, SOKI, Antarctic Climate and Ecosystems Co-operative Research Centre. <http://soki.aq/x/dAO7>.
- Tajima F. 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.
- Tamura, K., Stecher, G. y Kumar, S. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870-1874.
- Thorp, J. H. y Covich, A. P. (Eds.) 2009. *Ecology and classification of North American freshwater invertebrates*. Academic Press.
- Tiffin, P. y Ross-Ibarra, J. 2014. Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, 29(12), 673-680.
- Tigano, A., Shultz, A. J., Edwards, S. V., Robertson, G. J. y Friesen, V. L. 2017. Outlier analyses to test for local adaptation to breeding grounds in a migratory arctic seabird. *Ecology and Evolution*, 7(7), 2370-2381.
- Tobler, M., DeWitt, T. J., Schlupp, I. y Garcia de León, F. J. 2008. Toxic hydrogen sulfide and dark caves: Phenotypic and genetic divergence across two environmental gradients *Poecilia mexicana*. *Evolution*, 62, 2643-2649.

- Travers, A., Hiriart, E., Churcher, M., Caserta, M. y Mauro, E. D. 2010. The DNA Sequence-dependence of Nucleosome Positioning in vivo and in vitro. *Journal of Biomolecular Structure and Dynamics*, 6, 713-724.
- Vaillant, J. J., Bock, D. G., Haffner, G. D. y Cristescu, M. E. 2013. Speciation patterns and processes in the zooplankton of the ancient lakes of Sulawesi Island, Indonesia. *Ecology and Evolution*, 3(9), 3083-3094.
- Wang, X., Ma, X., Hu, W. y Yin, M. 2018. Genetic diversity and population differentiation of the freshwater copepod *Sinocalanus tenellus* (Calanoida, Centropagidae) in China. *Journal of Limnology*. 77(2).
- Wetzel, R. G. 2001. *Limnology: lake and river ecosystems*. gulf professional publishing.
- Wolf, J. B. y Ellegren, H. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nature Genetics*, 18(2), p. 87.
- Xu, S., Song, N., Zhao, L., Cai, S., Han, Z. y Gao, T. 2017. Genomic evidence for local adaptation in the ovoviparous marine fish *Sebastes marmoratus* with a background of population homogeneity. *Scientific Reports*, 7(1), p. 1562.
- Young, S. S., Lin, S. C. y Liu, M. Y. 2013. Genetic diversity and population structure of two freshwater copepods (Copepoda: Diaptomidae), *Neodiaptomus schmackeri* (Poppe and Richard, 1892) and *Mongolodiaptomus birulai* (Rylov, 1922) from Taiwan. *Diversity*, 4, 796-810.
- Zhou, H., Alexander, D. y Lange, K. 2011. A quasi-Newton acceleration for high-dimensional optimization algorithms. *Statistics and Computing*, 2, 261-273.