



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FES IZTACALA

BIOLOGÍA EVOLUTIVA

**Ensamblado *de novo* y anotación del mitogenoma del copépodo lacustre
*Leptodiptomus garciai***

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

Biól. José Miguel Carreto Romero

TUTOR PRINCIPAL DE TESIS: Dra. Elizabeth Ortega Mayagoitia

Facultad de Estudios Superiores Iztacala, UNAM

COMITÉ TUTOR:

Dr. Juan Servando Núñez Farfán

Instituto de Ecología, UNAM

Dr. Elías Piedra Ibarra

Facultad de Estudios Superiores Iztacala, UNAM

Los Reyes Iztacala, Tlanepantla, Estado de México, Agosto, 2022.



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.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FES IZTACALA

BIOLOGÍA EVOLUTIVA

**Ensamblado *de novo* y anotación del mitogenoma del copépodo lacustre
*Leptodiptomus garciai***

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

Biól. José Miguel Carreto Romero

TUTOR PRINCIPAL DE TESIS: Dra. Elizabeth Ortega Mayagoitia

Facultad de Estudios Superiores Iztacala, UNAM

COMITÉ TUTOR:

Dr. Juan Servando Núñez Farfán

Instituto de Ecología, UNAM

Dr. Elías Piedra Ibarra

Facultad de Estudios Superiores Iztacala, UNAM

Los Reyes Iztacala, Tlanepantla, Estado de México, 2022.

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

OFICIO CPCB/633/2022

ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **04 de abril de 2022** se aprobó el siguiente jurado para el examen de grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **(Biología Evolutiva)** del estudiante **CARRETO ROMERO JOSÉ MIGUEL** con número de cuenta **311021064** con la tesis titulada **“Ensamblado de novo y anotación del mitogenoma del copépodo lacustre *Leptodiptomus garciai*.”**, realizada bajo la dirección de la **DRA. ELIZABETH ORTEGA MAYAGOITIA**, quedando integrado de la siguiente manera:

Presidente: DR. JORGE CIROS PÉREZ
Vocal: DRA. DULCE ALEJANDRA ROUGON CARDOSO
Vocal: DR. JUAN ALBERTO PONCIANO GÓMEZ
Vocal: DRA. CORINA DIANA CEAPA
Secretario: DR. JUAN SERVANDO NÚÑEZ FARFÁN

Sin otro particular, me es grato enviarle un cordial saludo.

ATENTAMENTE
“POR MI RAZA HABLARÁ EL ESPÍRITU”
Ciudad Universitaria, Cd. Mx., a 29 de junio de 2022

COORDINADOR DEL PROGRAMA



DR. ADOLFO GERARDO NAVARRO SIGÜENZA



Agradecimientos institucionales

Al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México, por la oportunidad brindada.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo otorgado para la realización de estudios de maestría.

Al proyecto CONACYT (CB2015-1) 255788 “Evolución adaptativa en ambientes divergentes: organismos zooplanctónicos lacustres como modelo ecogenómico” y al proyecto PAPIIT IN222916 “Evolución de la tolerancia a la salinidad en copépodos calanoides *Leptodiptomus*” que financiaron esta investigación.

A mi tutora principal, Dra. Elizabeth Ortega Mayagoitia por la oportunidad, la confianza y el apoyo brindado durante esta investigación.

A mi comité tutor, Dr. Juan Núñez Farfán y Dr. Elías Piedra Ibarra por los valiosos comentarios y observaciones que me proporcionaron a lo largo de este proyecto. Su ayuda enriqueció esta investigación.

Agradecimientos a título personal

A la Doctora Elizabeth Ortega Mayagoitia, tutora y mentora. Muchas gracias por ser parte de mi formación científica, siempre he apreciado todas sus observaciones y consejos a lo largo de estos años. Gracias por su dedicación y tiempo.

Al Doctor Jorge Ciros Perez, por todo el apoyo que me ha brindado desde que llegué al laboratorio. Siempre he encontrado sabiduría y guía en sus comentarios

A todas las personas que son y fueron parte del laboratorio, en especial a Ana, Javi, Arturo y Omar. Siempre he aprendido grandes cosas de todos ustedes.

A los miembros del jurado y a mi comité tutor. A la Dra. Diana Corina Ceapa, el Dr. Juan Alberto Ponciano Gomez y la Dra. Dulce Alejandra Rougón Cardoso por el valioso tiempo invertido en este trabajo.

A mi familia, en especial a Sebastián por todas las horas que invertiste escuchando y viendo mis ensayos de presentaciones o cursos, siempre me fue posible aclarar mis ideas con tus comentarios y sugerencias.

A Diana, mi compañera de vida. Has estado a mi lado desde siempre, me has escuchado, dado consejos y apoyado en todas mis decisiones. Eres el amor de mi vida y nunca te podré agradecer lo suficiente por tanto que me has dado.

1. Introducción	3
1.1 El sistema genético	3
1.2 Genómica comparada	4
1.3 Genoma mitocondrial	5
1.4 La Subclase Copepoda	7
1.5 El copépodo <i>Leptodiaptomus garciai</i>	8
2. Objetivos	10
2.1 Objetivo general	10
2.2 Objetivos particulares	10
3. Metodología	10
3.1 Zona de estudio	10
3.2 Características ecológicas de <i>Leptodiaptomus garciai</i>	10
3.3 Obtención de organismos, extracción de ADN y secuenciación	11
3.4 Análisis Bioinformático	12
3.4.1 Limpieza de secuencias	12
3.4.2 Armado y anotación del genoma mitocondrial	12
3.5 Análisis evolutivos	13
3.5.1 Divergencia nucleotídica	14
3.5.2 Arquitectura del ADN mitocondrial	15
3.5.3 Análisis filogenético	15
3.6 Genoma nuclear	16
4. Resultados	16
4.1 Descripción y limpieza de datos	16
4.2 Ensamblado y anotación	17
4.3 Análisis evolutivos	21
4.3.1 Características del mitogenoma en Copepoda	21
4.3.2 Distancia evolutiva y divergencia nucleotídica	23
4.3.3 Arquitectura del ADN mitocondrial	25
4.3.4 Análisis filogenómicos	29
5. Discusión	34
5.1 Características del genoma mitocondrial	34
5.1 Evolución nucleotídica	35
5.2 Evolución de la arquitectura del mitogenoma	36
5.3 Análisis filogenómicos	38
5.4 Patrones evolutivos del mitogenoma de Copepoda	42
6. Conclusión	44
7. Referencias bibliográficas	45
8. Anexo	58
8.1 Ensamble del genoma	58
8.2 Ensamble del mitogenoma	62

Resumen

En años recientes el estudio del genoma mitocondrial ha adquirido gran relevancia, siendo utilizado para comprender procesos a una escala tanto microevolutiva como macroevolutiva. Dentro de la subclase Copepoda, actualmente existen doce mitogenomas reportados de los cuales solo uno pertenece a un copépodo lacustre de vida libre. En este trabajo, mediante una aproximación a través de herramientas genómicas y bioinformáticas, se presenta el borrador del mitogenoma de *Leptodiptomus garciai*, un copépodo calanoide que habita el lago Alchichica en la Cuenca Oriental, Puebla. La longitud del ensamble fue de 14, 654 pb, recuperando un total de 37 genes. Adicionalmente, se realizó una comparación de este mitogenoma contra copéodos de distintos órdenes. En conjunto, se identificó que los genes NAD (1-3) y ATP6 presentan tasas de sustitución nucleotídica elevada. Así mismo, se observan eventos de reordenamiento masivo en genes codificantes de proteínas, destacando que el ordenamiento genético de *L. garciai* es único dentro de la subclase. Por último, se realizaron análisis filogenéticos utilizando distintos genes mitocondriales, los cuales confirman la afinidad taxonómica de *L. garciai* a copéodos del orden Calanoida. Sin embargo, los escasos recursos genómicos de Copepoda, dificultan el establecer una filogenia con soporte estadístico robusto. Los recursos moleculares obtenidos en esta tesis sentarán las bases genómicas del estudio de procesos ecológicos y evolutivos de este importante grupo de animales acuáticos.

Abstract

In recent years, mitochondrial genome research has acquired particular attention, especially in the field of evolutionary biology. Although the Copepoda subclass exhibits a vast species diversity, currently there are only twelve reported mitogenomes, of which only one belongs to a lacustrine free-living copepod. Here, through a genomic and bioinformatic approach, we report a mitochondrial draft genome assembly of *Leptodiaptomus garciai*, a calanoid copepod that inhabits lake Alchichica from the Cuenca Oriental (Puebla), México. The genome assembly length is of 14,645 bp, with an annotation of 37 genes. Furthermore, comparisons between different copepods belonging to different orders were made. Overall, the genes NAD (1-3) and ATP6 seem to have the highest nucleotide substitution rates. In addition, the mitochondrial DNA architecture of different copepod species reveal multiple rearrangements of protein coding genes, highlighting the uniqueness of *L. garciai* mitogenome. Finally, phylogenetic analyses were performed using different mitochondrial genes, supporting the taxonomic affinity of *L. garciai* to the order Calanoida. The molecular resources provided in this thesis will shed light on the ecological and evolutionary events that take place in this important group of aquatic animals.

Ensamblado *de novo* y anotación del mitogenoma del copépodo lacustre *Leptodiptomus garciai*

1. Introducción

1.1 El sistema genético

Uno de los principales objetivos de la teoría de evolución biológica es esclarecer la historia natural de las especies. Desde la publicación de *El Origen de las Especies* por Darwin en 1859, se generó un parteaguas en la investigación evolutiva, ya que a partir de ese momento se comprendió que el estudio de la variación en los caracteres, la herencia y el ambiente, son puntos clave para explicar la extensa diversidad que existe entre los seres vivos. Años más tarde, con el surgimiento de áreas como la genética de poblaciones y la genética cuantitativa, las investigaciones se comenzaron a realizar en el marco de trabajo de la síntesis evolutiva y con esto, se comenzaron a explicar los patrones de evolución con mayor precisión y robustez estadística. En los últimos años, además del paradigma de la síntesis, han surgido distintas áreas de investigación que han ampliado y desafiado la visión actual del modelo evolutivo, diversificando aún más este campo. Algunas de estas nuevas ramas son: la epigenética, la biología evolutiva del desarrollo (evo-devo), la evolucionabilidad, la plasticidad fenotípica y la acomodación genética (Pigliucci, 2007).

Un punto en común que tienen estas áreas de trabajo es el estudio de los sistemas de transmisión de información. Estos sistemas son realmente importantes en la teoría evolutiva, ya que son los responsables de la herencia, son versátiles y con ellos las especies pueden hacer frente a las condiciones ambientales, siendo así, la pieza clave con la cual la evolución genera cambios en los organismos a través del tiempo. Siguiendo esta idea, algunos autores (Smith & Szathmary, 1997) proponen que la complejidad de los sistemas vivos surgió a partir de cuatro transiciones (el origen de los cromosomas, de los eucariontes, del sexo y de los organismos pluricelulares) que modificaron directamente el proceso de transmisión de información en las especies. Estos eventos pudieron repercutir en la complejidad de los patrones evolutivos, así como en un aumento de la diversidad taxonómica. Es así que, para comprender la historia natural de las especies, es fundamental conocer y analizar estos sistemas, entre ellos, el más sencillo de identificar y examinar es el sistema genético (Jablonka & Lamb, 2014). En él, la información hereditaria se encuentra organizada en una secuencia de nucleótidos, siendo el gen una unidad compuesta de ácidos nucleicos, la cual es decodificada en ARN y proteínas (Jablonka, 1994). Teniendo un conocimiento

íntegro de la información genómica, se pueden analizar las frecuencias genéticas para conocer patrones de selección (Nielsen, 2005) y adaptación (Orr, 2005). Esto a su vez, posibilita la exploración de vestigios evolutivos a nivel molecular y, además, se pueden asociar estos datos con información obtenida a partir de características morfológicas, fisiológicas y ecológicas.

Por lo general, el parentesco entre linajes se exploraba a partir de enzimas, secuencias nucleares y mitocondriales (Taylor *et al.*, 1996). Sin embargo, analizar regiones genéticas cortas se ha vuelto una limitación cada vez más evidente para dar explicaciones contundentes sobre la morfología, la fisiología y la adaptación de los organismos. Así, el avance de las técnicas genómicas aplicadas en la medicina, como la secuenciación del genoma humano (Venter *et al.*, 2001), atrajo el interés de investigadores a explorar la variación de las especies a nivel genómico (Shubin, 2020). Derivado de esto, surgió una gran cantidad de tecnologías y metodologías para la obtención parcial (polimorfismos) o completa (ensamble de genoma) de la secuencia de ADN de las especies (Zhang *et al.*, 2011), lo cual impulsó el desarrollo de tecnologías de secuenciación masiva (NGS, *next-generation sequencing*)

El crecimiento de las NGS ha ocasionado el origen de distintas metodologías que se han utilizado en el estudio de la biología evolutiva. Dentro de ellas se encuentran: el ensamblado de *novo* del genoma, el estudio de la variación genómica y epigenética, y los análisis de expresión diferencial (Park & Kim, 2016). Debido a que los costos de secuenciación se han reducido en los últimos años y los programas bioinformáticos son cada vez más diversos y accesibles al usuario con conocimientos limitados de programación, estos análisis son cada vez más sencillos de implementar en proyectos de investigación (Paszkiwicz & Studholme, 2010). Esto ha ocasionado innumerables posibilidades en lo referente al estudio de la diversidad, ya que ahora es más simple realizar comparaciones entre regiones genómicas de especies, lo cual permite conocer distintas características como secuencias linaje-específicas, secuencias compartidas entre especies, patrones de selección, regiones funcionales, entre otras (Wray, 2010).

1.2 Genómica comparada

El estudio de la variabilidad en las especies es sumamente importante, ya que a partir de esto, podemos conocer cómo se encuentran relacionadas entre sí y a su vez, formular hipótesis sobre procesos que moldearon su historia evolutiva. Esta variabilidad se puede estudiar a nivel de genoma, realizando comparaciones en cuanto a la estructura y función de secuencias de ADN. Por lo general, estos análisis se enfocan en el estudio del genoma nuclear y mitocondrial de los organismos.

Para analizar el genoma de un organismo, es necesario efectuar un muestreo representativo del ADN a través de una técnica de secuenciación masiva. A partir de esto, se efectúa un ensamble, el cual es la representación del genoma a partir de la estructuración jerárquica de la información obtenida (lecturas, *contigs* y *scaffolds*) mediante programas bioinformáticos (Miller *et al.*, 2010). Aunque estas técnicas han

avanzado de manera importante en los últimos años, la mayoría de estos ensamblajes distan de ser trabajos definitivos y en muchos casos se encuentran a un nivel de borrador (Phillippy *et al.*, 2008). Sin embargo, estos proyectos se van completando a lo largo del tiempo, perfeccionando la estrategia de secuenciación, ensamblado y anotado, logrando así la mejor calidad e integridad de un genoma.

La implementación de técnicas de NGS en el campo de la genómica comparada ha incrementado la comprensión de las dinámicas del genoma. Por ejemplo, estas técnicas han permitido conocer la divergencia en regiones funcionales en Metazoa mediante la descripción de regiones de *splicing* alternativo, la detección de zonas de regulación (micro-RNAs) y también la identificación de eventos genómicos como duplicaciones y pérdida de genes (Copley, 2008). Del mismo modo, en Arthropoda se han podido explicar de manera detallada las diferencias en patrones de selección tanto en regiones codificantes (Wright & Andolfatto, 2008) como no codificantes (Haddrill *et al.*, 2008), esclarecer patrones de especiación mediante el análisis de genes ortólogos (Gabaldón & Koonin, 2013) y estudiar los cambios en la arquitectura genómica (Sheffer *et al.*, 2021). La exploración del funcionamiento de regiones genómicas ha ocasionado que las investigaciones evolutivas se realicen con una precisión más alta. Esto se ve reflejado en distintos campos de la biología, como en una mayor exactitud en la clasificación de especies y en un entendimiento más completo del funcionamiento del acervo genético de los organismos frente a distintas condiciones ambientales (Van Straalen y Roelofs, 2011).

Del mismo modo, el estudio del genoma es una herramienta invaluable para la comprensión de los procesos ecológicos de una especie. Por ejemplo, en los peces espinosos (*Gasterosteus aculeatus*), a partir de un genoma ensamblado, se identificaron regiones genéticas que se asocian a la adaptación ocurrida en la transición de aguas marinas a epicontinentales (Jones *et al.*, 2012). También, el genoma del cladóceros *Daphnia* ha sido utilizado para relacionar características ecológicas y fisiológicas con análisis de expresión genómica, con la finalidad de entender los mecanismos adaptativos que han moldeado su evolución (Ebert, 2011).

1.3 Genoma mitocondrial

La función llevada a cabo por la mitocondria no solo se limita a la capacidad de producir ATP, sino que abarca una intrincada dinámica de cascadas de señalización (McBride & Neuspiel, 2006). La complejidad de su sistema de membranas y la interacción de los elementos mitocondriales con la fisiología y el metabolismo, ha generado interés en el estudio de este organelo no solamente en las ciencias biológicas, sino que también dentro del campo de la medicina (Kalyanaraman *et al.*, 2018). Una de las características que más destacan de este organelo es que, al originarse de un endosimbionte, cuenta con su propio genoma, el cual consiste principalmente en genes involucrados en el proceso de fosforilación oxidativa (Chinnery & Hudson, 2013).

El genoma de la mitocondria es una molécula de ADN circular de doble cadena, generalmente contiene un total de 37 genes, entre ellos, 22 ARN de transferencia, 2 ARN ribosomales y 13 genes codificantes de proteínas (CDS). Estos CDS corresponden al citocromo B, tres subunidades del citocromo C oxidasa (COX 1-3), dos subunidades del complejo ATPasa (ATP6 y ATP8) y siete subunidades de NADH deshidrogenasa (NAD 1-6, 4L; Wolstenholme, 1992). Los ARN de transferencia se caracterizan por ser moléculas pequeñas y en algunos casos se ha reportado superposición de estos genes (Taanman, 1999). El mitogenoma no contiene intrones a excepción de la región control, la cual se compone de elementos repetidos y contiene el origen de la replicación del mitogenoma. También, existen cambios en el código genético, lo que significa que en algunas especies hay una reasignación o alteración de codones (Smith & Keeling, 2015).

La arquitectura del ADN mitocondrial se define como el orden del conjunto de regiones codificantes, tomando en cuenta el contenido y la posición de los genes. Aunque la cantidad de genes se mantiene constante en distintas especies, se ha observado que algunas copias de CDS y/o ARNt se pueden perder (Gissi *et al.*, 2008). También, se reconoce la existencia de cuatro eventos de reordenamiento comunes dentro del mitogenoma, entre ellos se encuentran: inversión, transposición, transposición inversa y TDRL (Figura 1). Este último evento (TDRL) es de los más considerables en la evolución de la arquitectura genómica, de éste se pueden derivar los otros eventos de reordenamiento e implica una duplicación de una región completa del genoma (Bernt *et al.*, 2011).



Figura 1. Eventos de reordenamiento, de izquierda a derecha: inversión, transposición, transposición inversa y TDRL (duplicación en tándem seguida de pérdida al azar (*tandem duplication and random loss*)). Imagen tomada de Bernt *et al.*, 2013)

El genoma mitocondrial presenta características únicas, como una gran cantidad de mitocondrias por célula, su naturaleza homoplasmática, la herencia por línea materna, un tamaño de genoma corto (~16 Kpb) y la ausencia de recombinación (Wilson *et al.*, 1985). Esto ha facilitado la investigación de características genómicas estructurales, como la variación en el tamaño del genoma, cambios en la arquitectura genética, tasas de sustitución, secuencias no codificantes y estructuras secundarias de ARN (Gissi *et al.*, 2008).

Es así, que el ADN de este organelo se ha utilizado de manera habitual para explorar procesos biológicos. Por ejemplo, la información obtenida del mitogenoma ha revelado

eventos evolutivos importantes en metazoos, como que la multicelularidad y bilateralidad han ido acompañados de transiciones en la arquitectura genómica, pérdida de genes mitocondriales (Lavrov, 2007) y pérdida parcial o total de ARNt (Salinas-Giegé *et al.*, 2015). Además de esto, se ha usado para esclarecer relaciones filogenéticas, como en el proyecto de código de barras de la vida, el cual emplea un fragmento del gen COX1 como herramienta sistemática (Hebert *et al.*, 2003).

Con el desarrollo acelerado de las tecnologías de NGS, se ha hecho cada vez más simple y rutinaria la obtención y el análisis de genomas mitocondriales completos (Tang *et al.*, 2014). Esto mismo ha facilitado el uso de distintos genes mitocondriales (además de nucleares) para realizar análisis filogenómicos. En algunas ocasiones, utilizar distintos genes es un procedimiento que puede generar una mayor confianza estadística en comparación con métodos tradicionales de un solo gen (Lambert *et al.*, 2015), reflejando con mayor precisión las relaciones ancestro-descendiente de las especies (Havird & Santos, 2014; Brito & Edwards, 2009). A su vez, establecer filogenias que presenten una robustez estadística más alta, puede ayudarnos a estimar el tiempo de divergencia entre linajes, efectuar la reconstrucción de estados ancestrales de rasgos particulares y conocer patrones de diversificación y especiación, en conjunto con la biogeografía de las especies (Young & Gillung, 2020).

Además de esto, se ha estudiado la asociación entre el genoma mitocondrial y nuclear, ya que muchos de los productos de la mitocondria se encuentran codificados en el ADN del núcleo (Hill, 2015). Esto podría esclarecer patrones de especiación, puesto que se ha observado que las disrupciones de las interacciones mitonucleares pueden tener un efecto significativo en la eficacia de los híbridos de especies con divergencia elevada en genes mitocondriales (Healy & Burton, 2020).

1.4 La subclase Copepoda

Los artrópodos son uno de los grupos más diversos del planeta y representan aproximadamente un 80% de las especies descritas de animales (Zhang, 2011). El interés sobre la evolución de este grupo se ha visto reflejado en iniciativas como la “Sequencing Five Thousand Arthropod Genomes” (I5K), la cual a través de las tecnologías NGS busca ensamblar y anotar genomas de artrópodos que abarquen la mayoría de la diversidad existente y que actualmente suma ~500 genomas publicados (I5K consortium, 2013; Poelchau *et al.*, 2015). Aunque, dentro del filo se han descrito familias genéticas con función biológica específica, la mayoría de ellas aún tienen una función desconocida, lo cual refleja vacíos de conocimiento que se tiene sobre la función a nivel molecular dentro del grupo (Thomas *et al.*, 2020).

Dentro de los ambientes lacustres, la subclase Copepoda (Arthropoda: Hexanauplia) es una de las más abundantes (Wetzel, 1975), constituyendo entre el 50% - 75% de la biomasa del zooplancton tanto en aguas marinas como continentales (Elías-Gutiérrez *et al.*, 2008). Actualmente, se tiene registro de un total de 11,139 especies de copépodos (Walter & Boxshall, 2021) y, aunque los estudios relacionados a la ecología

y fisiología de esta subclase son abundantes, los recursos genómicos son escasos. Ejemplo de ello es que al momento de escribir la presente tesis, la base de datos del *National Center for Biotechnology Information* (NCBI) cuenta únicamente con un total de 12 genomas completos representativos de 12 especies de copépodos.

En años recientes, el interés en esta subclase ha ido en aumento, lo cual se ve evidenciado en investigaciones de índole genómico, proteómico y metabolómico, las cuales pretenden dilucidar la respuesta fisiológica y ecológica a nivel molecular (Amato & Carotenuto, 2018). Se ha identificado que las técnicas genómicas en muchas ocasiones son complicadas de implementar en los copépodos debido a su variación en el tamaño del genoma. Sin embargo, en los últimos años se han producido esfuerzos para solventar estos problemas, estableciendo metodologías que faciliten el ensamble del genoma y el análisis de la subclase a través de la genómica de poblaciones (Choquet *et al.*, 2019). Esto ha ayudado a generar genomas con una calidad elevada como el del copépodo *Eurytemora affinis*, con el cual se han podido identificar loci específicos sobre los cuales la selección natural ha actuado durante los eventos de invasión de ambientes continentales (Stern & Lee, 2020).

Por otra parte, el avance en la generación de recursos genómicos abre la posibilidad de establecer relaciones filogenéticas a partir de distintos genes nucleares o mitocondriales. Generalmente, en la subclase se han utilizado marcadores moleculares tradicionales para establecer relaciones ancestro-descendiente (COX1 e ITS). No obstante, existen diferentes problemáticas asociadas a esto, por ejemplo, dificultades metodológicas de extracción de COX1 con respecto a la afinidad de oligonucleótidos (Rocha *et al.*, 2018), baja resolución taxonómica empleando marcadores nucleares (ITS; Caputa *et al.*, 2017) y discordancia entre filogenias moleculares y morfológicas (Braga *et al.*, 2017). Por lo cual, el emplear recursos genómicos puede ser una alternativa en la resolución de relaciones filogenéticas dentro del grupo.

La relevancia de expandir el repertorio genómico de esta subclase no solo recae en la importancia ecológica que tiene este linaje, sino que también permitiría comprender los procesos evolutivos a los que han estado sujetos. Por ejemplo, conocer las bases genéticas de la adaptación en los eventos repetidos de colonización de aguas continentales, el mecanismo genético que subyace en la diapausa, comprender la variación del tamaño del genoma y resolver las relaciones filogenéticas de esta subclase desde una perspectiva filogenómica (Bron *et al.*, 2011).

1.5 El copépodo *Leptodiptomus garciai*

Dentro de la Cuenca Oriental, Puebla, existe una serie de lagos en los cuales habitan copépodos del género *Leptodiptomus* y que en años recientes se han comenzado a estudiar mediante tecnologías de NGS. Estas poblaciones exhiben diferencias significativas en cuanto al tamaño del genoma (Carreto-Romero, 2019), así como una

fuerte estructura poblacional revelada por marcadores moleculares polimórficos (SNPs; Urban-Olivares, 2019, Ortega Mayagoitia *et al.*, 2022). Entre estos lagos, se encuentra Alchichica, un lago tropical cálido monomítico (Alcocer *et al.*, 2000) que cuenta con una gran cantidad de endemismos, entre ellos el copépodo calanoide *Leptodiptomus garciai* (Osorio-Tafall 1942). Las características limnológicas únicas del lago han ocasionado que *L. garciai* presente adaptaciones que le permiten ser la especie dominante de este sistema. Por ejemplo, la presencia de depredadores, la incidencia de luz ultravioleta y los bajos recursos alimenticios han ocasionado que se encuentre adaptado mediante rasgos como la capacidad de migración vertical y tener bajo requerimiento de alimento para mantener en equilibrio a su población (Co; Ciro-Pérez *et al.*, 2015). Además, esta especie se encuentra adaptada a las condiciones locales de salinidad del lago (9 g L^{-1}) (Montiel-Martínez *et al.*, 2008); también tiene plasticidad fenotípica en el número de huevos, el esfuerzo reproductivo y el número de nauplios debido a la disponibilidad de alimento y a la temperatura (Ortega-Mayagoitia *et al.*, 2018).

En conjunto, estas investigaciones han generado una descripción bastante completa de los rasgos con los que *L. garciai* enfrenta las condiciones particulares de su ambiente. Estas características ecológicas, en conjunto con la afinidad filogenética de la especie (analizada con COX1) (Montiel-Martínez, 2008) y el tamaño de su genoma (Carreto-Romero, 2019), la hacen candidata a una descripción detallada de su genoma, lo cual podría establecer a esta especie como un modelo ecológico-evolutivo. La propuesta de este proyecto es generar el primer ensamble y anotación del mitogenoma de *L. garciai*, el cual puede ser utilizado para explorar regiones genéticas de interés, patrones evolutivos y esclarecer relaciones filogenéticas. Cabe destacar que este sería uno de los primeros ensambles de un mitogenoma de copépodo lacustre y a su vez el segundo genoma de la familia *Diaptomidae*, que es de las más diversas en ambientes epicontinentales (Boxshall & Defaye, 2007).

2. Objetivos

2.1 Objetivo general

Ensamblar y anotar el genoma mitocondrial de *Leptodiptomus garciai* para inferir sus patrones de evolución dentro de la subclase Copepoda.

2.2 Objetivos particulares

1. Realizar una descripción del mitogenoma de *L. garciai*.
2. Comparar la evolución del mitogenoma de *L. garciai* a nivel nucleotídico y sinténico utilizando información genómica de bases de datos públicas.
3. Estimar las relaciones ancestro-descendiente de *L. garciai* utilizando información genómica.

3. Metodología

3.1 Zona de estudio

La Cuenca Oriental, en el estado de Puebla, se encuentra ubicada en la parte oriente del Eje Neovolcánico, es una estructura endorreica delimitada por las coordenadas 18° 48' a 19° 43' de latitud norte y 97° 09' a 97° 03' de longitud oeste, abarcando una superficie aproximada de 5,250 km² en los estados de Puebla, Tlaxcala y Veracruz (Can-Chulim *et al.*, 2011). La fisiografía volcánica dentro de la cuenca se produjo desde principios del Cenozoico hasta el Cuaternario, conformándose como malpaís (Figueroa *et al.*, 1983). Dentro de ella se encuentra una serie de lagos de origen volcánico tipo *maar*, entre ellos Alchichica. Alchichica es un lago cálido monomítico que presenta una temporada de mezcla en los meses de enero - mayo y estratificación en los meses de junio - diciembre (Lugo *et al.*, 1999). Este lago es oligotrófico, endorreico y con profundidad máxima de 62 m, salinidad de 9 g L⁻¹ y con una prevalencia de los iones Cl-Na (Armienta *et al.*, 2008).

3.2 Características ecológicas de *L. garciai*

El ciclo de vida de los copépodos comprende una etapa larval con seis estadios de nauplio, seguido de seis estadios juveniles denominados copepoditos; el último de ellos es la forma adulta. El dimorfismo sexual se hace evidente a partir de la forma adulta, siendo los machos generalmente más pequeños y delgados que las hembras. *L. garciai* es un copépodo calanoide perteneciente a la familia Diaptomidae y es la

especie de zooplancton dominante del lago Alchichica. La coloración de su cuerpo es blanquecino-transparente, la talla de los adultos macho es de $906 \pm 49 \mu\text{m}$ y de hembras de $1001 \pm 53 \mu\text{m}$. El tamaño de puesta es variable a lo largo del año, el promedio más bajo es de $2.1 \pm 0.5 \mu\text{m}$ (mayo) y el más alto de $9.3 \pm 2.0 \mu\text{m}$ (diciembre; Hernández-Martínez, 2014). La migración vertical de los copepoditos C4, C5 y adultos de *L. garciai* a través de la columna de agua podría evitar la depredación ante el pez planctívoro *Poblana alchichica* (Atheriniformes) (Ciros-Pérez *et al.*, 2015). Se ha observado una correlación positiva ($p < 0.05$) entre la cantidad de comida y el esfuerzo reproductivo ($R^2 = 0.57$), el número de huevos ($R^2 = 0.64$) y el número de nauplios por hembra ($R^2 = 0.66$). Y una correlación negativa ($p < 0.001$) entre la temperatura y la talla de las hembras ($R^2 = -0.94$; Ortega-Mayagoitia *et al.*, 2018).

3.3 Obtención de organismos, extracción de ADN y secuenciación

La colecta del copépodo *L. garciai* se realizó el 23 de febrero del 2020 en el lago Alchichica mediante una red de arrastre con una malla de $100 \mu\text{m}$. Una vez en el laboratorio, se aislaron aproximadamente 200 individuos adultos, machos y hembras, lo cual corresponde a $\sim 1,466 \mu\text{g}$ de biomasa seca calculado a partir de una regresión talla - peso (Dumont *et al.*, 1975). Con la finalidad de reducir la contribución de contenido intestinal y bacterias al extracto de ADN, se estableció un cultivo adicionando penicilina al 5%. El cultivo se efectuó en matraces Erlenmeyer de 1 L a la salinidad del lago (9 g L^{-1}) preparada mediante sales comerciales (Instant Ocean®, Aquarium Systems™) en agua electrodesionizada (Millipore®, Elix 5) y autoclavada. Los organismos se mantuvieron con un fotoperiodo de 12 h luz y 12 h oscuridad a temperatura constante (18°C), durante las 48 h previas a la extracción de ADN. Ésta se realizó mediante el kit DNeasy Blood and Tissue de Qiagen. Por último, se utilizó el fluorómetro Qubit para evaluar la concentración y calidad del ADN ($150 \text{ ng}/\mu\text{l}$; QC ≥ 1.8).

Posterior a esto se envió el ADN al Centro de Genómica en la Universidad de Minnesota (UMGC por sus siglas en inglés, <https://genomics.umn.edu>). En donde se realizó la preparación de bibliotecas únicas de doble índice (UDI) mediante el kit ADN TruSeq Nano de Illumina y la secuenciación de un solo carril en un equipo de la serie NovaSeq SP a $2 \times 150 \text{ bp}$. La cobertura de los datos genómicos se calculó a partir de la ecuación de Lander/Waterman (1988) (ecuación 1) a partir del tamaño de genoma conocido de *L. garciai* (Carreto-Romero, 2019).

$$C = \frac{LN}{G}$$

Ecuación 1. C = Cobertura; L = Longitud del read; N = Número de secuencias; G = Tamaño del genoma haploide; donde $L=151$, $N= 111,814,448$ y $G=0.4 \text{ pg} = 391 \text{ Mpb}$.

3.4 Análisis Bioinformático

El análisis de las secuencias comenzó con dos archivos pareados en formato FASTA, el procedimiento informático se realizó en un entorno UNIX, con software de distribución libre y recursos informáticos en línea.

3.4.1 Limpieza de secuencias

La evaluación de calidad de las secuencias se realizó mediante FastQC (Andrews, 2010), realizando por separado la limpieza por calidad por medio de Trimmomatic (Bolger *et al.*, 2014) y AdapterRemoval (Schubert *et al.*, 2016). Por medio de FastQ Screen (Wingett & Andrews, 2018) se identificaron genomas contaminantes mediante el mapeo de las secuencias de *L. garciai* contra el genoma de *Scenedesmus* sp. (GCA_014905635.1), *Tetraselmis striata* (GCA_006384855.1), *Escherichia coli* (GCA_000005845.2), *Cyclotella cryptica* (GCA_013187285.1), *Brachionus plicatilis* (GCA_010279815.1) y *Homo sapiens*. Se utilizaron los genomas de *Scenedesmus* y *Tetraselmis* debido a que son del mismo género que las especies de microalgas que se utilizan como alimento para el cultivo de copépodos en nuestro laboratorio, *Cyclotella* por ser del mismo género que la especie dominante del fitoplancton del lago Alchicica (*C. alchichicana*), *E. coli* y *H. sapiens* para descartar la contaminación producida en la manipulación de las muestras.

Para limpiar las secuencias de los contaminantes identificados en cada set de datos (Trimmomatic y AdapterRemoval) se empleó Bowtie2 (Langmead & Salzberg, 2012). Seguido de esto, se descargó la base de datos Kraken2 (Wood *et al.*, 2019) para realizar una limpieza general contra bibliotecas de bacteria y protista. Por último, para verificar que no existieran rastros de contaminación, se alineó el ensamblaje a través de la herramienta BLAST contra la base de datos del NCBI (Megablast),

3.4.2 Armado y anotación del genoma mitocondrial

El ensamblaje del mitogenoma se realizó a partir de las secuencias de *L. garciai* sin contaminantes obtenidas a partir de Trimmomatic. Para recuperar lecturas exclusivas de ADN mitocondrial se utilizó BWA para mapear las secuencias de 6 genomas mitocondriales de copépodos depositados en el NCBI contra el set completo de secuencias de *L. garciai*. Los copépodos utilizados en el mapeo fueron: *Lernaea cyprinacea* (KM235194.1), *Phyllodiaptomus tunguidus* (MN927223.1), *Calanus hyperboreus* (JX678968.1), *Sinergasilus polycolpus* (NC_028085.1), *Sinergasilus undulatus* (NC_054173.1) y *Amphiascoides atopus* (NC_023783.1). Para efectuar el ensamblaje del mitogenoma se usó el programa Unicycler (Wick *et al.*, 2017), generando la anotación genética mediante MitoZ (Meng *et al.*, 2019) y Mito Web Server (MWS;

Bernt *et al.*, 2013). De manera adicional, se realizó la validación de los ARNt mediante el servidor tRNAscan-SE (Chan *et al.*, 2021) y la reconstrucción de la estructura terciaria del ARN ribosomal mediante RNAfold-web server (Lorenz *et al.* 2011) (Figura 2).

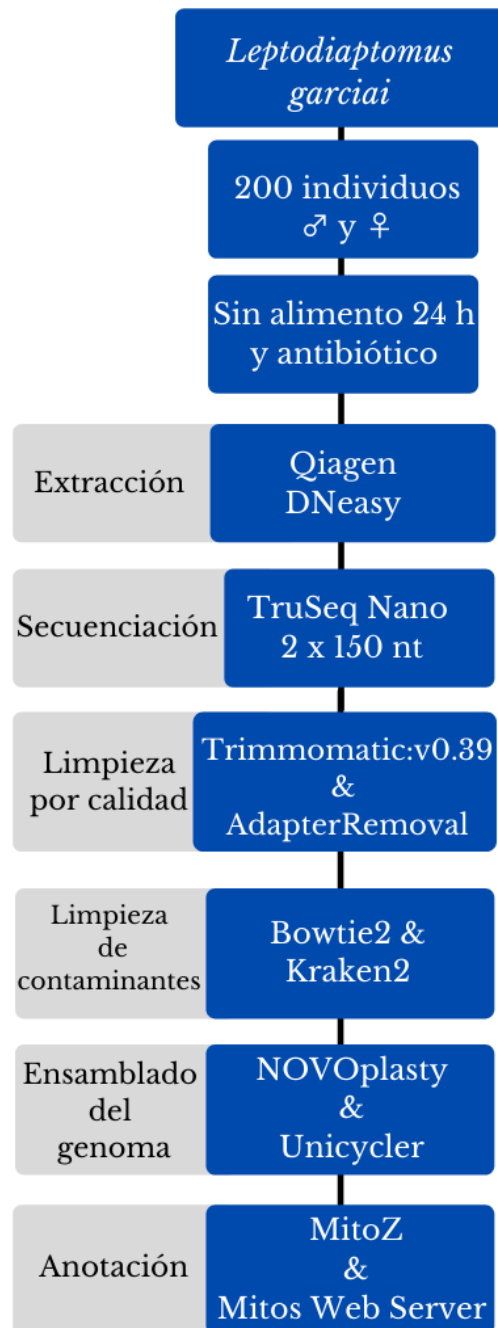


Figura 2. Flujo de trabajo del ensamble y anotación del mitogenoma de *L. garciai*

3.5 Análisis evolutivos

Los análisis evolutivos a nivel mitocondrial se llevaron a cabo utilizando la anotación de *L. garciai* proveniente de MitoZ (ver sección anterior), añadiendo los siguientes mitogenomas depositados en el NCBI: *C. glacialis* (MG001883.1) *Paracyclopsina nana* (EU877959.1), *Pandarus rhincodonicus* (HG942363.1), *Tigriopus kingsejongensis* (MK598762.1) y *Tigriopus californicus* (DQ917374.1). En conjunto, los 11 mitogenomas se utilizaron para vislumbrar patrones evolutivos en tres niveles de resolución: nucleotídico, sinténico y filogenómico.

Cabe resaltar que algunos análisis no fueron llevados a cabo con todos los genomas previamente mencionados, debido a que algunas secuencias depositadas dentro del NCBI se encontraban incompletas o presentaban algún error en la anotación genética.

3.5.1 Divergencia nucleotídica

Para conocer la divergencia a nivel nucleotídico entre el mitogenoma de *L. garciai* y las especies de copépodos mencionadas anteriormente, se realizó un análisis de distancia-p mediante MEGA X con la anotación de los 13 genes codificantes de proteínas (CDS) transformados en aminoácidos.

En seguida, procedió al alineamiento de las secuencias mitocondriales en codones mediante MEGA X por medio de ClustalW. Mediante comparaciones pareadas de *L. garciai* contra los copépodos mencionados anteriormente, se estimó la tasa de sustituciones sinónimas por sitio sinónimo (K_s), sustituciones no sinónimas por sitio no sinónimo (K_a) y la tasa de sustitución nucleotídica (K_a/K_s o ω ; Kryazhimskiy & Plotkin, 2008; Nei & Gojobori, 1986) a través de DNAsp (Rozas *et al.*, 2017). El análisis de sustitución nucleotídica (ω) se realizó utilizando un archivo concatenado de 11 CDS y, también, sobre cada gen de manera independiente, usando el código genético de ADN mitocondrial de *Drosophila*. De este análisis se excluyeron los genes NAD4L-*ubiquinone oxidoreductase chain 4L (NAD4L)* y *Mitochondially Encoded ATP Synthase Membrane Subunit 8 (ATP8)* debido a la variación en la longitud de las secuencias.

Un valor de $\omega > 1$ nos indica una mayor cantidad de mutaciones no sinónimas dentro de una región genética, lo cual puede tomarse como un sitio que putativamente está bajo selección natural divergente, $\omega = 1$ nos indica un escenario de evolución neutral y $\omega < 1$ un escenario de evolución balanceadora o purificadora.

3.5.2 Arquitectura del ADN mitocondrial

La arquitectura del genoma mitocondrial de *L. garciai* se analizó mediante la comparación de la sintenia utilizando los 13 CDS de todos los genomas mitocondriales de copépodos mencionados anteriormente. Mediante Simple Synteny (Veltri *et al.*, 2016) se realizó una comparación visual de los eventos de reordenamiento genético y seguido de esto, se usó el servidor CREx (Bernt *et al.*, 2007) para conocer las distancias de reordenamiento mitocondrial, así como los posibles escenarios evolutivos que tuvieron lugar para generar el ordenamiento actual de los genomas mitocondriales.

3.5.3 Análisis filogenético

Para inferir las relaciones ancestro-descendiente, se hizo un análisis filogenético mediante la concatenación de distintos CDS de *L. garciai* en conjunto con la información de los mitogenomas de copépodos mencionados en la sección 3.5. Como grupo externo se utilizó el cladóceros *Daphnia similis* (NC044415.1) por ser la especie filogenéticamente más cercana con recursos genéticos comparables disponibles. Para el análisis se utilizaron los siguientes genes: Nicotinamida adenina dinucleótido subunidades 1 a 6 (NAD1, NAD2, NAD3, NAD4, NAD4L, NAD5 y NAD6), Citocromo b (CYTB), Citocromo c oxidasa subunidades 1 a 3 (COX1, COX2, COX3), ATP6 y ATP8.

En primer lugar, se realizó el alineamiento de aminoácidos mediante MEGA X por medio de CLUSTALW. A continuación, se realizó la concatenación de genes siguiendo tres combinaciones: A) COX 1, 2 y 3 (1,196 posiciones de aminoácidos); B) COX 1, 2, 3 y CYTB (1,321 posiciones de aminoácidos); C) los 13 CDS (2,677 posiciones de aminoácidos). Las primeras dos combinaciones se seleccionaron a partir de la información obtenida en el análisis de sustitución nucleotídica (ω), en donde se identificaron genes bajo selección purificadora ($\omega < 1$). Para cada una de las combinaciones se obtuvo el mejor modelo evolutivo mediante MEGA X, eligiendo el modelo con el menor puntaje de BIC (*Bayesian Information Criterion*). Por último, se realizaron dos análisis filogenéticos con cada combinación de genes, utilizando máxima verosimilitud (MV) e inferencia bayesiana (IB).

El análisis de MV se efectuó en MEGA X empleando un valor de *bootstrap* de 5,000 repeticiones con el modelo de sustitución LG (Le & Gascuel, 2008) + F y una distribución gamma (+G) de cinco categorías. El análisis de IB se llevó a cabo con BEAST v1.10.4, en el que se usó el modelo de sustitución LG + G de 5 categorías y con un valor de MCMC de 3,000,000 generaciones. Mediante Tracer v1.7.2 se evaluó el tamaño efectivo de muestreo (ESS > 2,000) para cada conjunto de datos y se visualizó el árbol filogenético utilizando TreeAnnotator v1.10.4 y FigTree v1.4.4.

Finalmente, se estimó el tamaño ancestral del mitogenoma a partir del paquete de phytools (Revell, 2012) en R, con un intervalo de confianza del 95%. En este análisis se utilizó la topología de la filogenia con mayor soporte estadístico.

3.6 Genoma nuclear

Por último, se utilizaron los datos provenientes de la sección 3.4.1 para intentar hacer un ensamble del genoma nuclear de *L. garciai*. Se utilizaron los datos provenientes de Trimmomatic y AdapterRemoval. En cada caso, se probaron dos vías de armado, una por ensamble *de novo* y otra por ensamble con referencia. No fue posible la obtención de un ensamble con una calidad e integridad adecuada, sin embargo, se generó un set de datos limpio de contaminantes. El procedimiento de ensamblado y los resultados obtenidos se pueden encontrar en la sección de anexos (8.1). Estos recursos pueden ser empleados en el futuro para construir un borrador del genoma completo, en conjunto con otra tecnología de secuenciación.

4. Resultados

4.1 Descripción y limpieza de datos

La información genómica proveniente de la UMGC consistió en dos archivos de lecturas pareadas con una longitud de 151 pb. Cada archivo contiene 111,814,448 secuencias, con un porcentaje de GC del 45% y una cobertura de 43X. La mejor calidad de lecturas, evaluadas a través de la calidad phred > 33 y el %GC (Anexo A) se obtuvo con AdapterRemoval, sin embargo, de las secuencias iniciales para cada archivo solo quedaron 57,582,255 secuencias. En comparación, con Trimmomatic, se obtuvo una calidad más baja, pero con un total de 96,131,935 secuencias. En la Tabla 1 se muestran los resultados de la limpieza de las lecturas mediante estos dos programas.

Tabla 1. Comparación de secuencias resultantes con AdapterRemoval y Trimmomatic.

	Secuencias	Cobertura
Archivos crudos	111,814,448	43X
AdapterRemoval	57,582,255	22X
Trimmomatic	96,131,935	36X

Para cada lectura, el programa FastQ Screen detectó un porcentaje bajo de secuencias contaminantes (Anexo B). Del mismo modo, la limpieza por medio de Kraken2 indicó que el 99.15% de secuencias (95,188,418 pb) no fueron catalogadas dentro de la base de datos de este programa, lo que corresponden a 816,037 secuencias contaminantes; en el Anexo C se muestra el origen de la contaminación identificada.

4.2 Ensamblado y anotación

Mediante el ensamblador Unicycler se obtuvo el genoma mitocondrial de *novo* de *L. garciai*, el cual se encuentra contenido en un solo *contig* circularizado con longitud de 14,654 pb. En la Tabla 2 se muestra la comparación de la anotación llevada a cabo mediante MitoZ y Mitos Web Server (MWS); los dos programas recuperaron la misma cantidad de CDS, con diferencias en cuanto al número de ARNt y ARNr.

Tabla 2. Comparación de genes anotados

Programa	Proteínas	ARNt	ARNr
MitoZ	13	19	2
Mitos Web Server	13	22	1

Dado que los análisis evolutivos se realizaron a partir de los 13 CDS, y que no hubo diferencias en la anotación de estos genes a través de los dos programas, se eligió utilizar la anotación de MitoZ a lo largo de este trabajo. En la Figura 3 se muestra la representación gráfica circular del mitogenoma y en la Figura 4 la representación lineal. Entre los CDS anotados, se encuentran siete genes codificados en la cadena (+) y seis en la cadena (-), dos ARNr codificados en la cadena (-), siete ARNts codificados en la cadena (+) y doce en la cadena (-). En la Tabla 3 se muestran las características de los CDS anotados y en la Tabla 4 la de los ARNt y ARNr.

En comparación con MitoZ, el servidor MWS realizó la anotación completa de los ARNr, ARNt y la estructura secundaria de los mismos (Anexo D, E F y G); sin embargo, la validación por medio de tRNAscan-SE solo detectó 13 ARNt (Anexo H, I y J). Esto se puede deber a que algunos ARNt aún podrían estar fragmentados o no cuentan con el valor estadístico necesario para que se identifiquen como secuencias válidas a través de este servidor. Asimismo, dentro del mitogenoma, se identificó una región intergénica única, sin embargo esta región varía según el programa; MitoZ la identifica en la posición 10,051 - 11,911 pb mientras que MWS la muestra entre 9,962 - 11,043 pb. Posiblemente esta zona sea la región de control del mitogenoma, esto debido a que en la mayoría de las especies, esta región se encuentra ubicada en la sección única no-

codificante del mitogenoma y puede ser un área de oportunidad para futuros análisis evolutivos (Shadel & Clayton, 1997; Zhang & Hewitt, 1997).

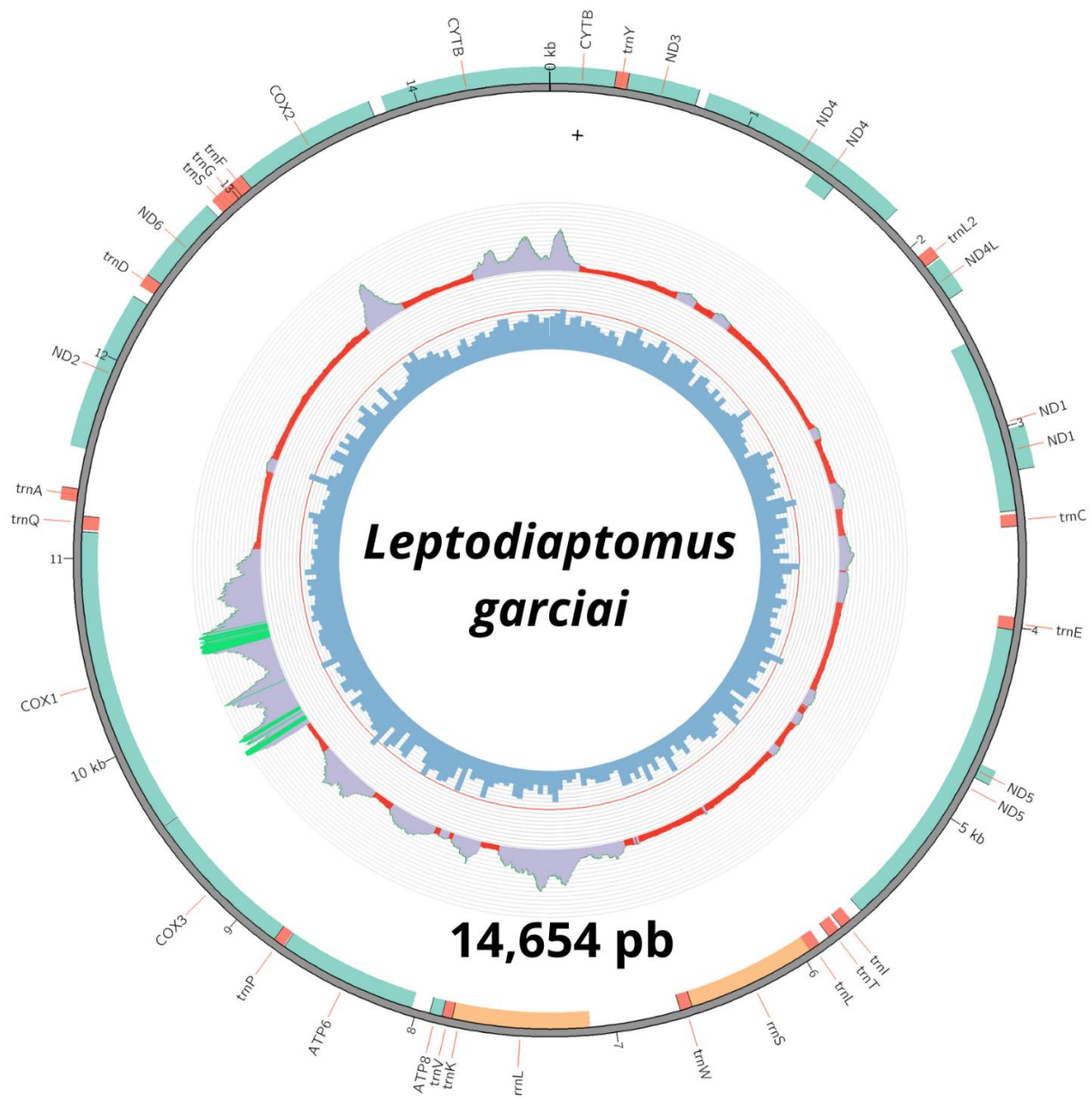


Figura 3. Representación circular del mitogenoma de *Leptodiptomus garciai*. Las barras interiores (azules) representan el contenido de GC, las exteriores (grises) muestran la cobertura de las lecturas (en verde fosforescente se muestran las regiones con mayor cobertura), los CDS se encuentran representados en color azul-verde, los ARN ribosomales en color naranja y los ARN de transferencia en color rojo. La figura fue generada mediante el software MitoZ.

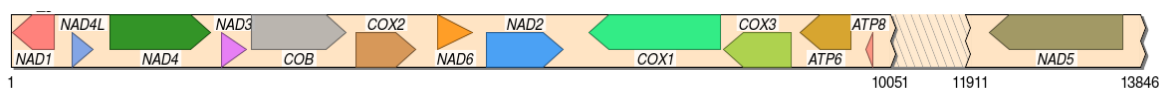


Figura 4. Representación lineal de los 13 genes mitocondriales codificantes de proteínas de *L. garciai*; se muestra la posible región intergénica ubicada entre las posiciones 10 051 y 11 911. La figura fue generada mediante el recurso en línea Simple Synteny.

Tabla 3. Descripción de la ubicación y tamaño de los CDS anotados en el mitogenoma de *L. garciai*.

Gen	Inicio de la secuencia	Fin de la secuencia	Longitud (pb)	Cadena
NAD1	2	540	539	-
NAD4L	728	944	217	+
NAD4	1,010	2,291	1,282	+
NAD3	2,330	2,681	352	+
CYTB	2,740	3,874	1,135	+
COX2	3,935	4,626	692	+
NAD6	4,816	5,284	469	+
NAD2	5,402	6,270	869	+
COX1	6,572	8,123	1,552	-
COX3	8,125	8,917	793	-
ATP6	8,977	9,685	709	-
ATP8	9,750	9,843	94	-
NAD5	12,077	13,698	1,622	-

Tabla 4. Descripción de la ubicación y tamaño de los ARN de transferencia y ribosomales anotados en el mitogenoma de *L. garciai*.

Gen	Inicio de la secuencia	Fin de la secuencia	Longitud (pb)	Cadena
ARN-Leu	939	1,001	63	+
ARN-Tyr	2,680	2,739	60	+
ARN-His	3,874	3,933	60	+
ARN-Phe	4,635	4,696	62	+
ARN-Ser	4,759	4,817	59	+
ARN-Asp	5,282	5,344	63	+
ARN-Ala	6,372	6,433	62	+
ARN-Gln	6,502	6,569	68	-
ARN-Val	9,843	9,906	63	-
ARN-Lys	9,903	9,963	61	-
ARN-Trp	11,043	11,105	63	-
ARN-Leu	11,752	11,815	64	-
ARN-Met	11,815	11,876	62	-
ARN-Thr	11,876	11,933	58	-
ARN-Ile	11,952	12,013	62	-
ARN-Glu	13,701	13,764	64	-
ARN-Arg	13,772	13,829	58	-
ARN-Asn	13,832	13,894	63	-
ARN-Cys	14,217	14,277	61	-
ARNr-16S	9,893	11,240	1348	-
ARNr-12S	11,105	11,836	732	-

4.3 Análisis evolutivos

4.3.1 Características del mitogenoma en Copepoda

Aunque el tamaño del mitogenoma de *L. garciai* (14,654 pb) es similar al de las demás especies de copéodos (14,000 - 18,000 pb), se puede observar una gran variación cuando se compara con el grupo de especies filogenéticamente más cercanas de Calanoida; en contraste con *C. glacialis* (27,342) y *C. finmarchicus* (>29,462) que tienen el mitogenoma más grande reportado dentro de la subclase (Weydmann *et al*, 2017; Figura 5). En la Tabla 5 se presenta una comparación de las características principales del mitogenoma de *L. garciai* y de los mitogenomas de copéodos disponibles en el NCBI. Por otra parte, se observa una variación en el contenido de GC, abarcando un rango de 30% a 46%. Cabe mencionar que el genoma de *C. finmarchicus* no se encuentra disponible dentro del NCBI, por lo tanto, no fue posible utilizarlo en los análisis evolutivos posteriores.

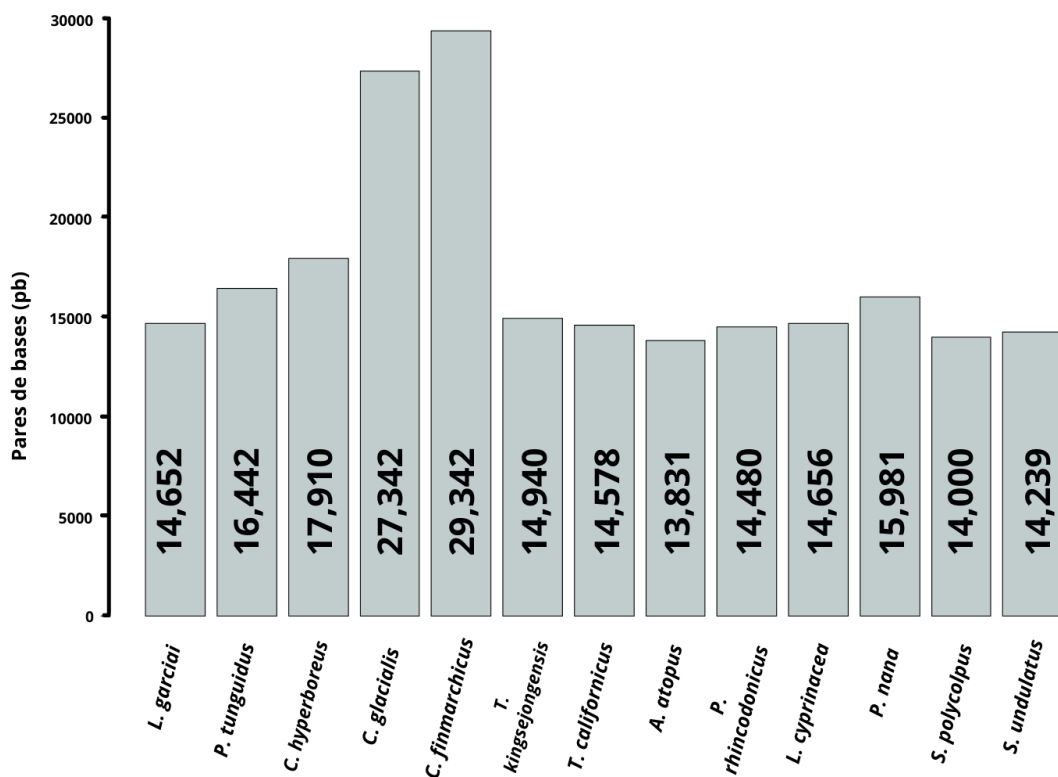


Figura 5. Comparación del tamaño de mitogenomas de algunas especies de Copepoda. La figura fue generada mediante el software R studio.

Tabla 5. Comparación de genes anotados (MitoZ) de *L. garciai* y otros mitogenomas de copépodos disponibles en el NCBI.

Especie	Familia/Orden	Tamaño del ensamblado	%GC	Longitud de la región intergénica	CDS	ARNr	ARNt	Total de genes
<i>L. garciai</i>	Diaptomidae/ Calanoida	14,654 pb	31.9	1860 pb	13	2	19	34
<i>P. tunguidus</i>	Diaptomidae/ Calanoida	16,442 pb	30.6	na	13	2	22	37
<i>C. hyperboreus</i>	Calanidae/ Calanoida	17,910 pb	33.7	3496 pb	13	2	22	37
<i>C. glacialis</i>	Calanidae/ Calanoida	27,342 pb	40	na	11	2	18	31
<i>C. finmarchicus</i>	Calanidae/ Calanoida	29,462	na	na	na	na	na	na
<i>T. californicus</i>	Harpacticidae/ Harpacticoida	14,578 pb	38.5	na	13	2	22	37
<i>T. kingsejongensis</i>	Harpacticidae/ Harpacticoida	14,940 pb	46.2	879 pb	13	2	22	37
<i>A. atopus</i>	Miraciidae/ Harpacticoida	13,831 pb	35.9	na	13	2	22	35
<i>P. rhincodonicus</i>	Pandaridae/ Siphonostomatoida	14,480 pb	38.6	na	12	2	22	36
<i>L. cyprinacea</i>	Lernaeidae/ Cyclopoida	14,656 pb	29.3	446 pb	13	2	22	37
<i>P. nana</i>	Cyclopoididae/ Cyclopoida	15,981 pb	29.2	2350 pb	12	2	23	37
<i>S. undulatus</i>	Ergasilidae/ Cyclopoida	14,239 pb	29.6	na	13	2	22	37
<i>S. polycolpus</i>	Ergasilidae/ Cyclopoida	14,000 pb	28.6	155 pb	13	2	22	37

4.3.2 Distancia evolutiva y divergencia nucleotídica

Mediante el análisis de distancia-p, considerando un concatenado de 11 CDS, se encontró que la divergencia más elevada (0.531) es entre *T. californicus* y *L. cyprinacea*, mientras que, la menor (0.145) fue entre *C. glacialis* y *C. hyperboreus* (Anexo K). En cuanto a la divergencia de *L. garciai*, la más elevada (0.517) fue con *P. rhincodonicus* y la menor (0.304) con *P. tunguidus*. En la Figura 6 se muestra la distancia evolutiva entre pares de especies, en donde se destaca la afinidad filogenética de las especies de copépodos. En la Figura 7 se contrasta la distancia evolutiva entre *L. garciai* y especies de los Órdenes cercanos (*P. tunguidus* y *C. glacialis*) y lejanos (*P. nana* y *P. rhincodonicus*).

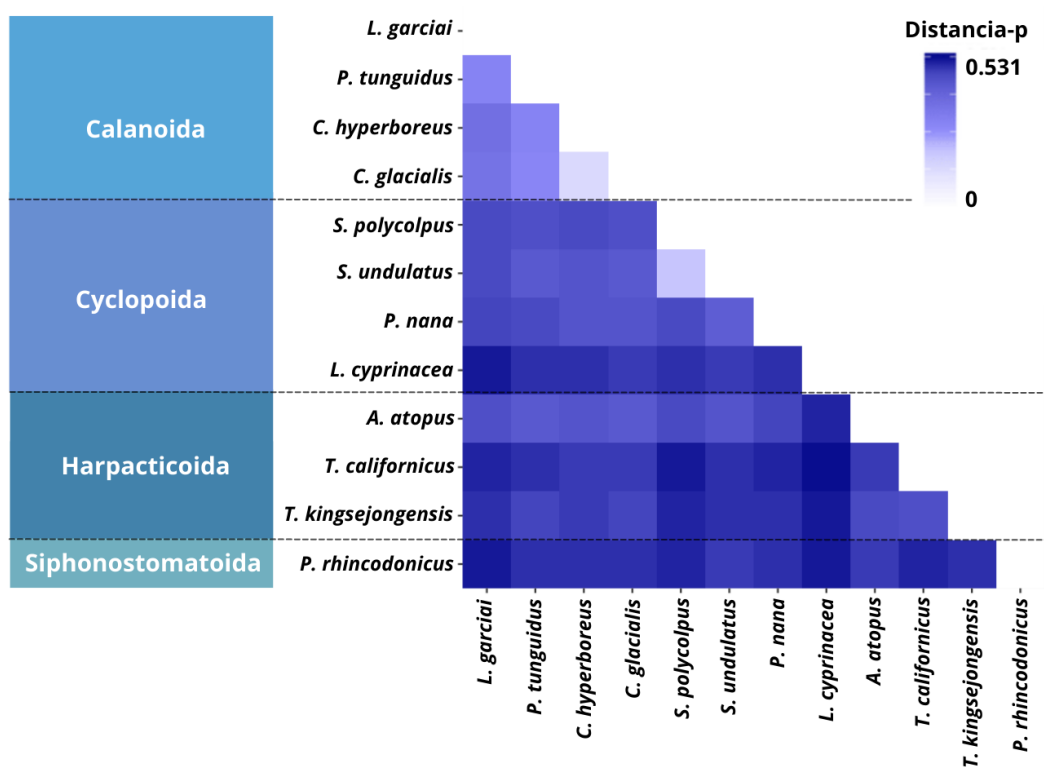


Figura 6. Matriz de distancia-p entre mitogenomas representada a través de un mapa de calor, en donde los valores más claros representan una distancia más cercana a 0 y colores fuertes una escala cercana al valor máximo de 0.531. Del lado izquierdo se muestran los Órdenes a los que pertenecen las especies. La figura fue generada mediante el recurso en línea Heatmapper.

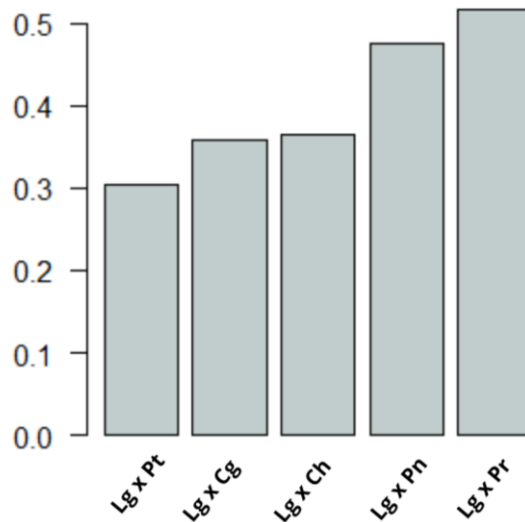


Figura 7. Comparación del número de polimorfismos intragénicos entre especies. *L. garciai* (Lg), *P. tunguidus* (Pt), *C. hyperboreus* (Ch), *P. nana* (Pn), *P. rhincodonicus* (Pr). La figura fue generada mediante el software R studio.

A continuación, se presenta la divergencia de acuerdo con la tasa Ka/Ks entre *L. garciai* y las especies de copépodos de comparación. El concatenado de los 11 CDS abarcó un total de 10,325 pb, lo cual corresponde al 70% del mitogenoma de *L. garciai*. El análisis individual de cada uno de los 11 CDS disponibles se muestra en la Figura 8 y un promedio de este valor se muestra en la Tabla 6. Los genes que presentaron una mayor tasa de sustitución fueron NAD3 y ATP6, mientras que los genes que presentan la tasa más baja son los del complejo de COX (Anexo L). Las especies que tienen la tasa más elevada de Ka/Ks (>1) comparado con *L. garciai* son *S. undulatus* (NAD1, NAD2, NAD3, ATP6), *S. polycolpus* (NAD1, NAD6, ATP6), *P. nana* (NAD2, NAD5, CYTB; Anexo L), *A. atopus* (NAD3, NAD4, ATP6) y *L. cyprinacea* (NAD3, NAD5, ATP6); todos ellos pertenecientes a los órdenes Cyclopoida y Harpacticoida.

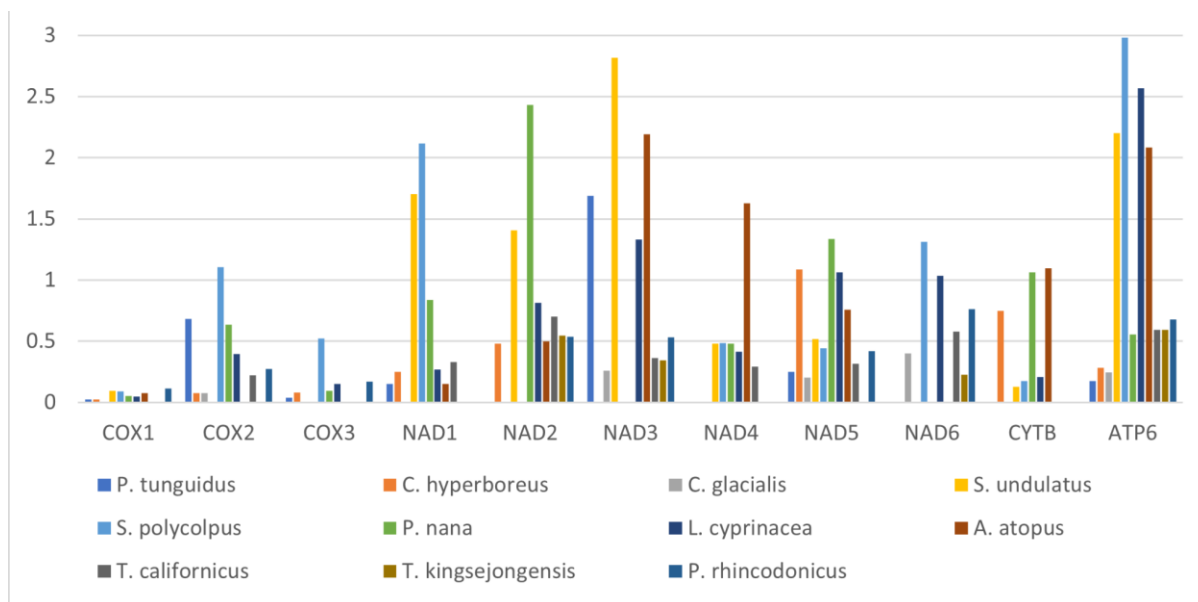


Figura 8. Comparación de la tasa de sustitución Ka/Ks de 11 genes entre *L. garciai* y otras especies de copéodos.

Tabla 6. Promedio de la tasa de sustitución Ka/Ks por gen.

COX1	COX2	COX3	NAD1	NAD2	NAD3	NAD4	NAD5	NAD6	CYTB	ATP6
0.065 ± 0.035	0.434 ± 0.356	0.177 ± 0.177	0.726 ± 0.770	0.927 ± 0.680	1.191 ± 0.973	0.631 ± 0.494	0.640 ± 0.398	0.719 ± 0.404	0.570 ± 0.455	1.178 ± 1.051

4.3.3 Arquitectura del ADN mitocondrial

La comparación entre la arquitectura mitocondrial de los CDS en copéodos mostró un alto grado de reordenamiento. En la Figura 9 se hace una representación comparada del mitogenoma de nueve especies, resaltando exclusivamente los CDS. Destacan los cambios de posición y de orientación de los genes, esto es, un cambio de posición de una cadena a otra (+ y -) y/o en el sentido de lectura. En este trabajo se excluyeron análisis involucrando regiones de ARN debido a que su anotación (ARNt y ARNr) no fue consistente (Tabla 2) en *L. garciai*.

En la Figura 10 se muestra una matriz de acuerdo con el reordenamiento calculado en CREx (Anexo M), donde el valor más alto (150) indica un ordenamiento de genes idéntico entre especies, mientras que los valores descendentes indican una menor similitud entre los mitogenomas. Las especies con un ordenamiento mitogenómico más parecido a *L. garciai* son *P. tunguidus* (94), *C. hyperboreus* (16) y *T. californicus* (16). Del mismo modo, se obtuvieron los posibles escenarios evolutivos que permiten explicar la sintenia de *L. garciai* a partir de las demás especies (Tabla 7), donde se observa que la especie que comparte una menor cantidad de eventos de

reordenamiento con *L. garciai* es *P. tunguidus*. Aparte de este copépodo, todas las comparaciones muestran escenarios de TDRL, los cuales sugieren múltiples eventos de duplicación respecto a CDS.

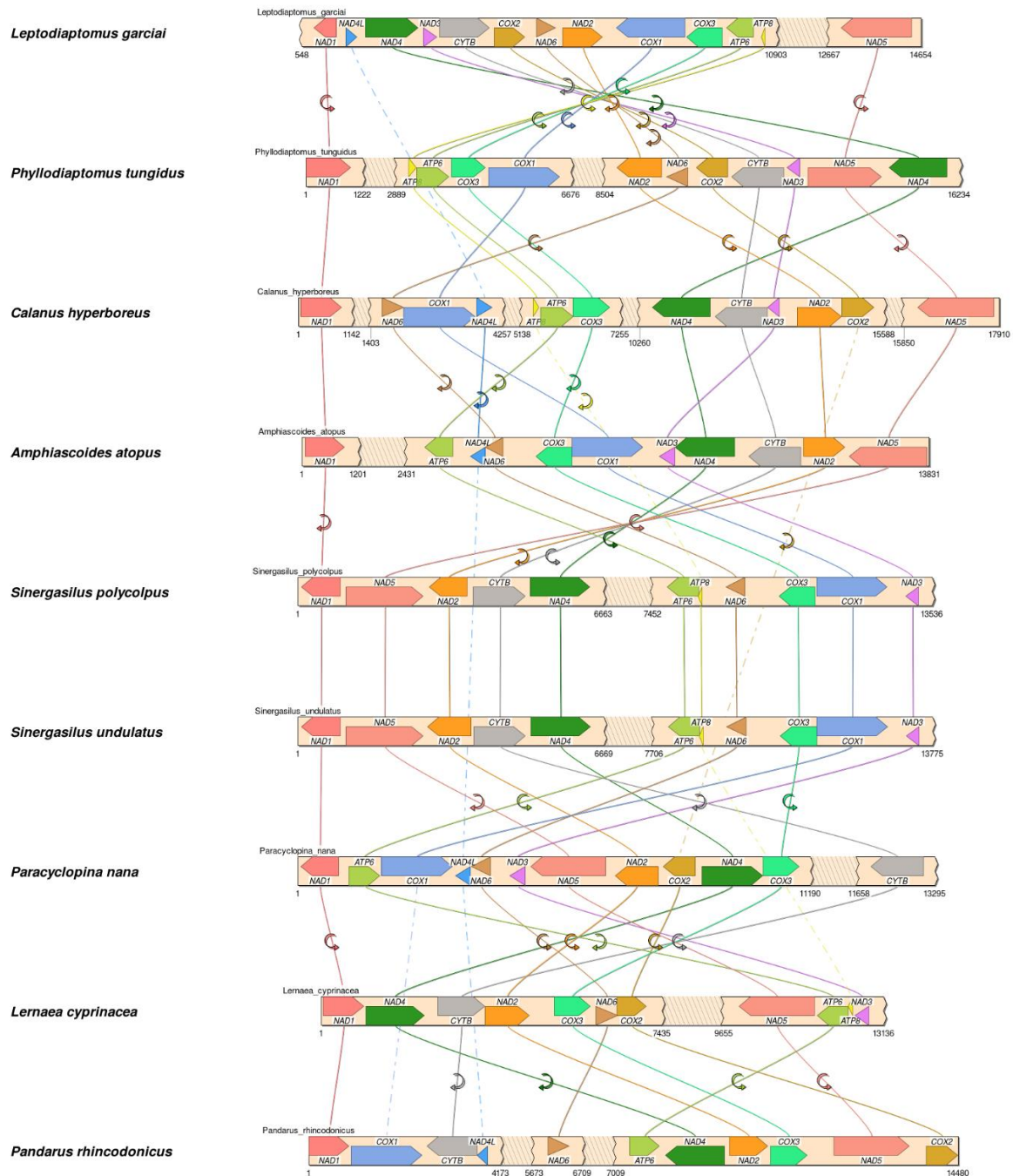


Figura 9. Comparación de la representación lineal de los CDS en el mitogenoma de algunos copépodos. El mitogenoma comienza arbitrariamente con el gen NAD1 y las flechas indican un cambio en la orientación (inversión o transversión) de genes. La figura fue generada mediante Simple Synteny.

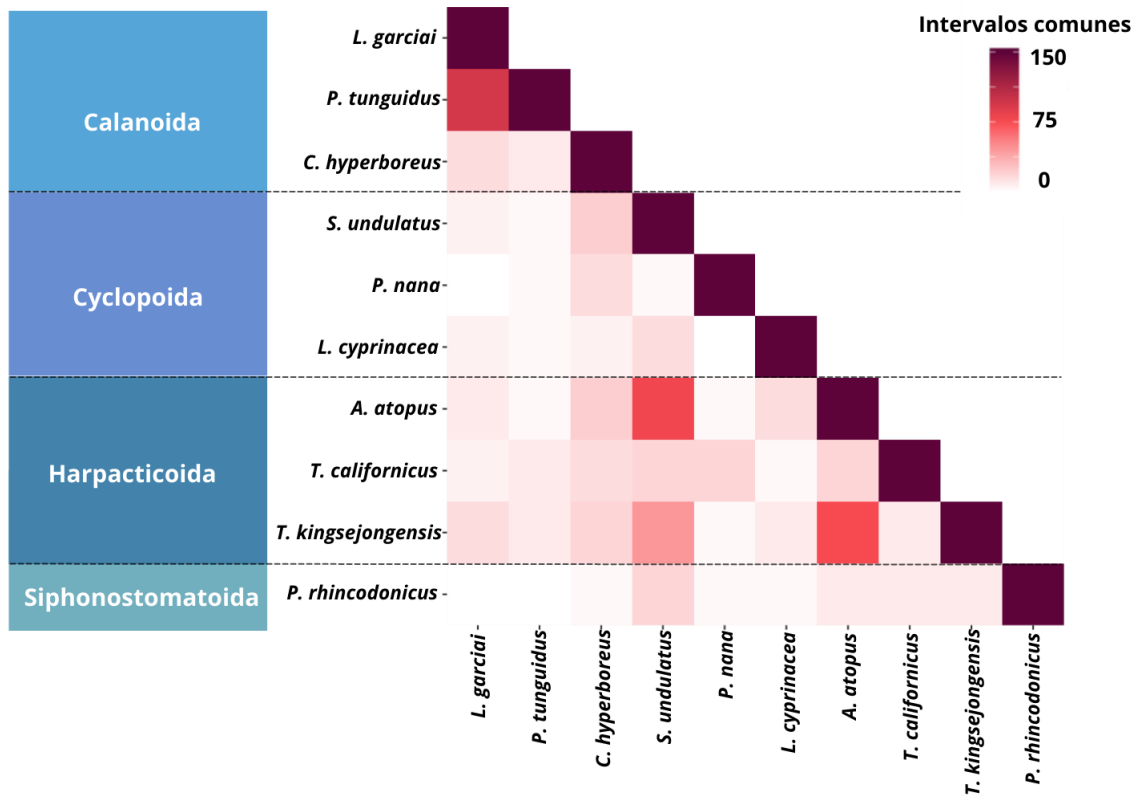


Figura 10. Matriz de intervalos comunes obtenidos mediante el servidor CREx. Mayor número de intervalos comunes representan un ordenamiento genético mitocondrial concordante mientras que valores bajos indican un ordenamiento divergente. La figura fue generada mediante el recurso en línea Heatmapper.

Tabla 7. Eventos de reordenamiento genético entre *L. garciai* y otras especies de copépodos, estimados mediante el servidor CREx. T: Transposición, I: Inversión, TI: Transposición inversa, TDRL: Duplicación en tándem seguido de pérdida al azar.

	<i>L. garciai</i>
<i>P. tunguidus</i>	T,R
<i>C. hyperboreus</i>	T, TI, I (2), TDR (2)
<i>L. cyprinacea</i>	T (2), I (3), TDRL (2)
<i>A. atopus</i>	T (2), TI, I (3), TDRL (2)
<i>P. nana</i>	I (2), TDRL (2)

Tabla 7. Continuación

	<i>L. garciai</i>
<i>P. rhincodonicus</i>	I (2), TDRL (3)
<i>S. undulatus</i>	I (5), T (2), TDRL
<i>T. californicus</i>	T, I (2), TDRL (2)
<i>T. kingsejongensis</i>	I (2), T, TDRL

Por otra parte, comparando el ordenamiento de CDS en los mitogenomas de Copepoda, se observó la presencia de una región genética compartida en calanoides que comprende los genes COX3/APT6/ATP8 (Figura 11).

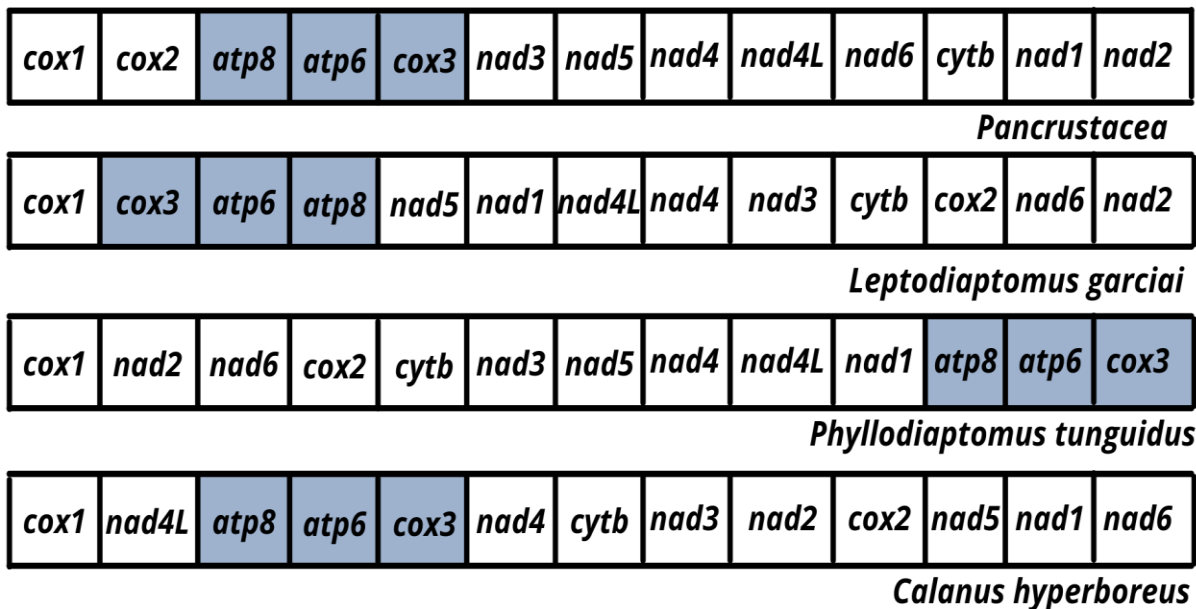


Figura 11. Ordenamiento mitocondrial de 13 CDS en Pancrustacea, *L. garciai* y 2 copépodos calanoides con información mitogenómica disponible, se resalta el patrón compartido de los genes ATP8-ATP6-COX3.

4.3.4 Análisis filogenómicos

Las relaciones filogenéticas de *L. garciai* se estimaron a partir de las tres combinaciones de genes mencionadas en la sección 3.5.3, utilizando un análisis de MV (Figura 12, 14 y 16) e IB (Figura 13, 15 y 17). La posición de *L. garciai* dentro de las filogenias obtenidas concuerda con lo esperado de acuerdo con la afinidad taxonómica de las especies. *L. garciai* se encuentra en un clado en conjunto con *P. tunguidus*, *C. hyperboreus* y *C. glacialis*, todos ellos pertenecientes al orden Calanoida. Sin embargo, fuera de esta similitud, los métodos de MV y el de IB mostraron diferencias en cuanto a las relaciones filogenéticas de las especies, donde los cambios principales en la topología de estos árboles involucran a las especies de los órdenes Harpacticoida y Siphonostomatoida.

El soporte de los nodos de las filogenias de MV fue muy bajo. En ellas, se observa que los árboles construidos mediante la combinación A y B (Figura 12 y 14) presentan una topología similar, en comparación con la combinación C (Figura 16). Por otro lado, los árboles de IB (Figura 13, 15 y 17) tienen los valores de probabilidad posterior más elevados, en especial la combinación C (Figura 17). Las tres combinaciones revelan una topología similar, en donde se observa a Calanoida y Cyclopoida como clados hermanos, con la excepción del copépodo ciclopoide *L. cyprinacea*. De igual modo, los copépodos harpacticoides se encuentran distribuidos en dos ramas distintas, parecido a las combinaciones A y B de las filogenias de MV.

En la Figura 18 se presenta la reconstrucción del estado ancestral del tamaño del mitogenoma en Copepoda. Se muestran los cambios ocurridos a través de la historia evolutiva de estos órdenes, denotando en colores cálidos las especies con un genoma más pequeño y en colores fríos las especies con un genoma más grande. Utilizando los datos disponibles hasta el momento, se estima que el estado ancestral del tamaño del genoma de la subclase Copepoda fue de 14,873 pb (nodo B). La mayoría de las especies exhiben un tamaño de genoma mitocondrial conservado (~13,800 a ~18,000), con la excepción del nodo I, correspondiente al ancestro común del orden Calanoida, cuyo tamaño de genoma experimentó una expansión con respecto al nodo anterior (G). Los linajes descendientes sufrieron procesos de contracción (familia Calanidae, nodo K) o expansión (Diptomidae, nodo L) en el tamaño del mitogenoma.

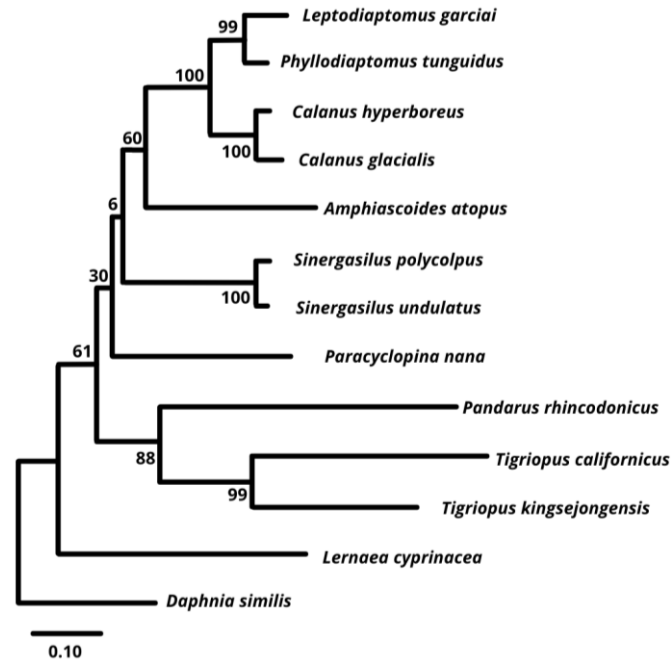


Figura 12. Filograma de MV utilizando la combinación de genes A (COX 1, 2 y 3). Se utilizó un modelo de sustitución LG + F + G, en los nodos se muestran los valores de *bootstrap*. Como grupo externo se utilizó a *D. similis*. La figura se realizó mediante el software Mega X.

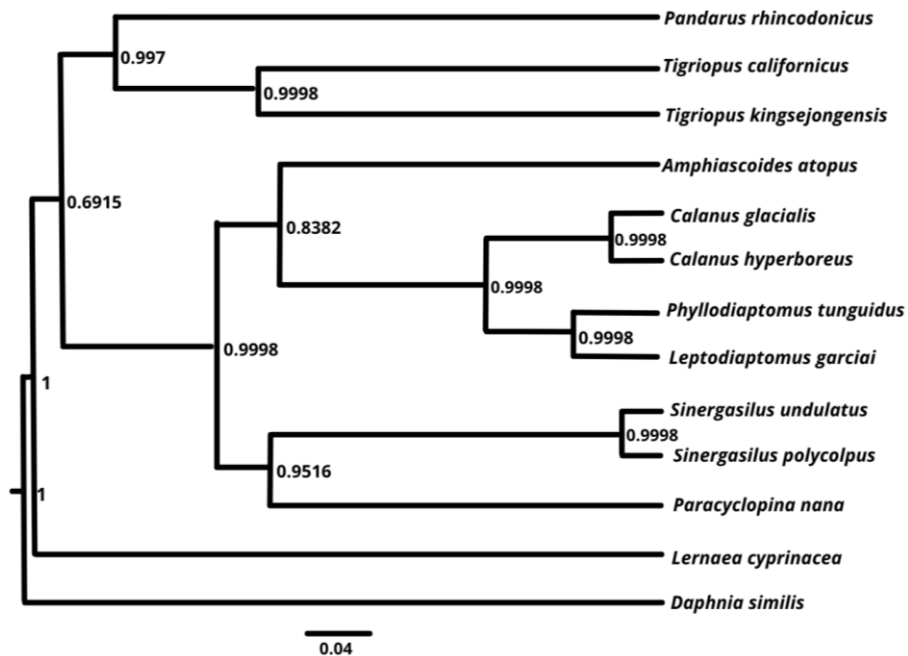


Figura 13. Filograma de IB utilizando la combinación de genes A (COX 1, 2 y 3). Se utilizó un modelo de sustitución LG + G de 5 categorías utilizando un valor de MCMC de 3,000,000 generaciones, en los nodos se muestran los valores de probabilidad posterior. Como grupo externo se utilizó a *D. similis*. La figura se realizó mediante el software TreeAnnotator y FigTree.

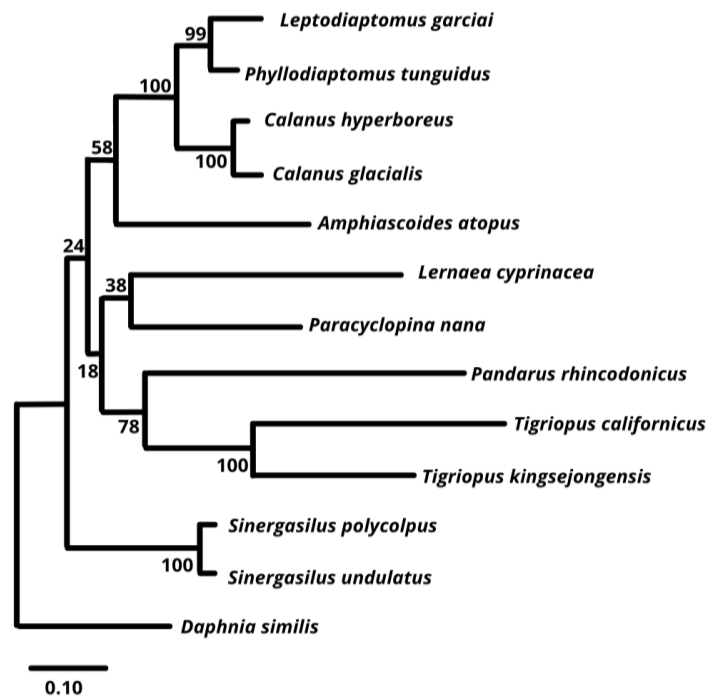


Figura 14. Filograma de MV utilizando la combinación de genes B (COX 1, 2, 3 y CYTB). Se utilizó un modelo de sustitución LG + F + G, en los nodos se muestran los valores de *bootstrap*. Como grupo externo se utilizó a *D. similis*. La figura se realizó mediante el software Mega X.

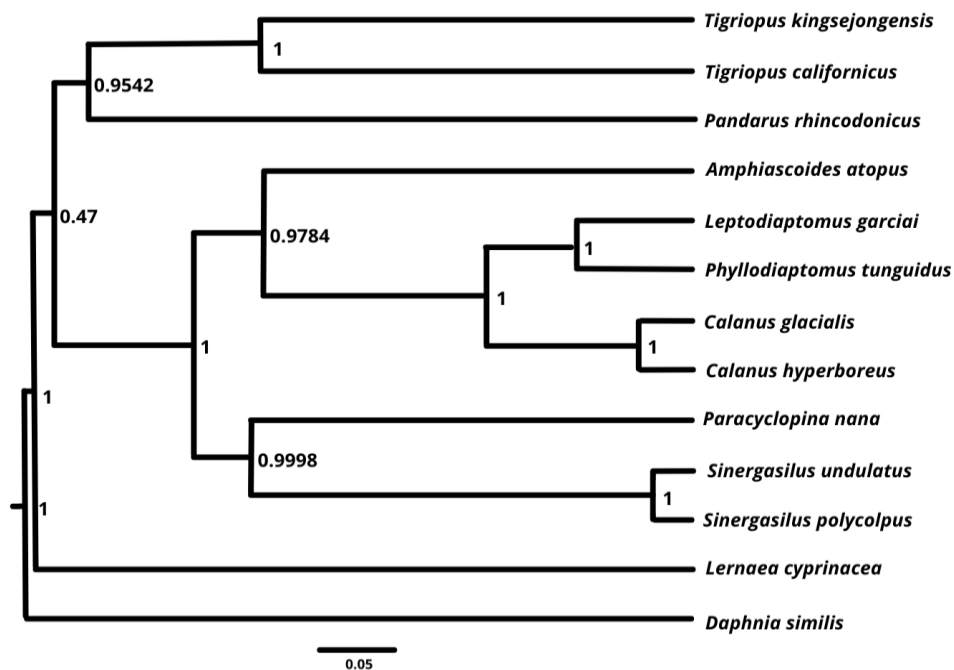


Figura 15. Filograma de IB utilizando la combinación de genes B (COX 1, 2, 3 y CYTB). Se utilizó un modelo de sustitución LG + G de 5 categorías utilizando un valor de MCMC de 3,000,000 generaciones, en los nodos se muestran los valores de probabilidad posterior. Como grupo externo se utilizó a *D. similis*. La figura se realizó mediante el software TreeAnnotator y FigTree.

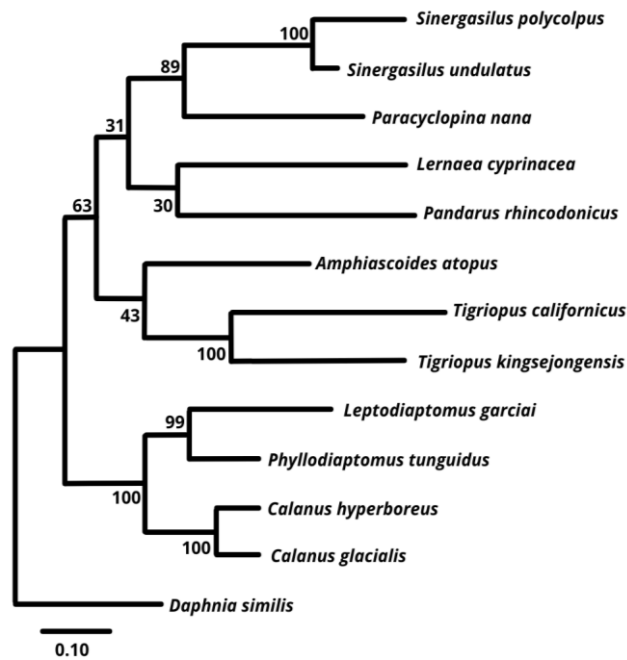


Figura 16. Filograma de MV utilizando la combinación de genes C (13 CDS). Se utilizó un modelo de sustitución LG + F + G, en los nodos se muestran los valores de *bootstrap*. Como grupo externo se utilizó a *D. similis*. La figura se realizó mediante el software Mega X.

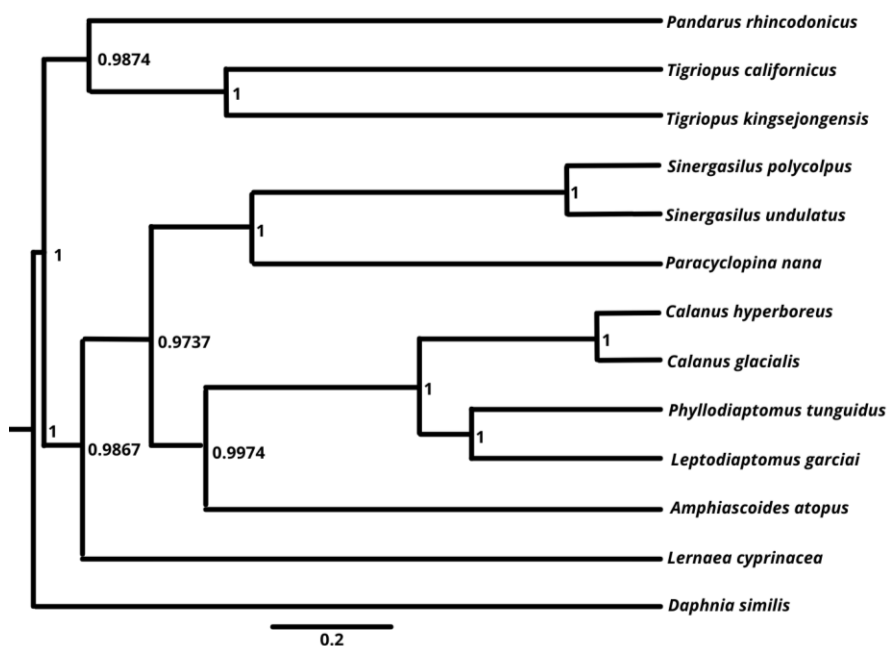


Figura 17. Filograma de IB utilizando la combinación de genes C (13 CDS). Se utilizó un modelo de sustitución LG + G de 5 categorías utilizando un valor de MCMC de 3,000,000 generaciones, en los nodos se muestran los valores de probabilidad posterior. Como grupo externo se utilizó a *D. similis*. La figura se realizó mediante el software TreeAnnotator y FigTree.

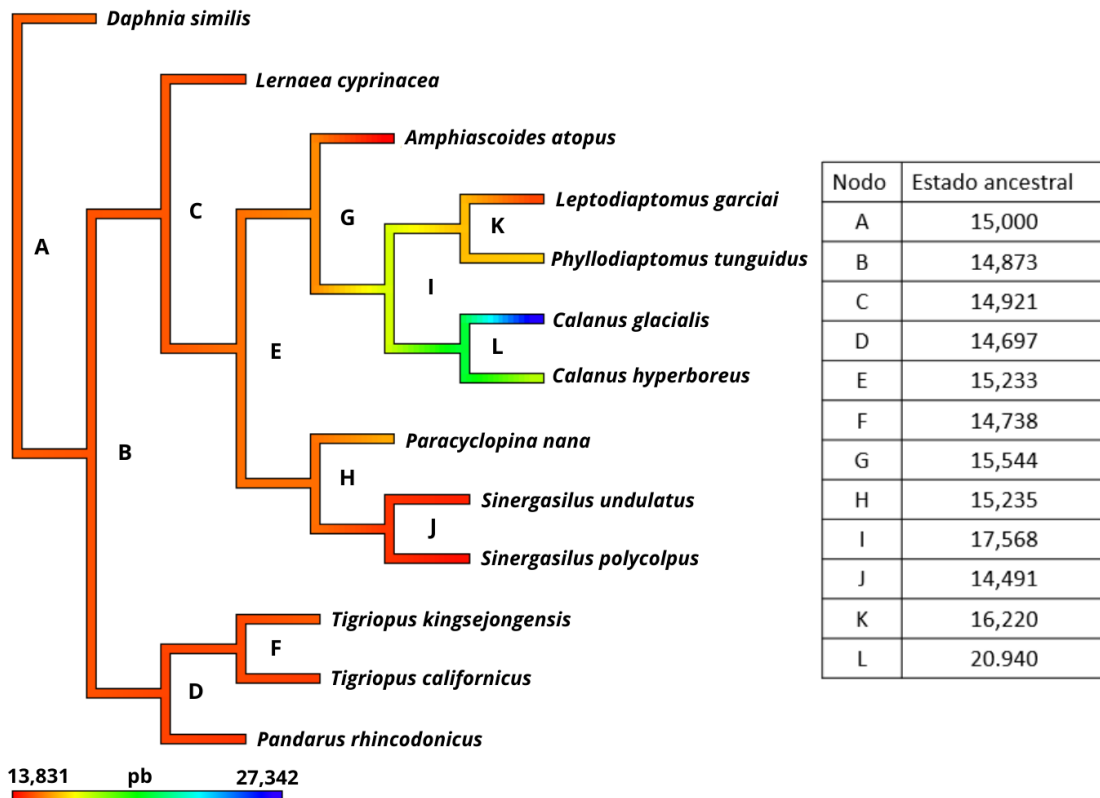


Figura 18. Reconstrucción del estado ancestral del tamaño del genoma mitocondrial (pb). El mapa de calor en el cladograma representa el cambio en el tamaño del mitogenoma. Los valores de la tabla representan la reconstrucción del estado ancestral de cada nodo (longitud en pb). La figura fue generada mediante el paquete de R *phytools*.

5. Discusión

5.1 Características del genoma mitocondrial

El ensamble realizado en este trabajo es el primer borrador del mitogenoma de la especie endémica *L. garciai* y a su vez, uno de los primeros mitogenomas de copépodos de ambientes continentales. Se recuperaron en su totalidad las copias completas de CDS, ARNt y ARNr (13, 22 y 2 respectivamente). Sin embargo, dependiendo del programa de anotación (MitoZ o MWS) se recuperó una cantidad distinta de genes de ARN. Aunque esto puede ser causado por errores en el ensamblado o en la anotación, se considera que el ensamble es de buena calidad debido a la circularización y a la ausencia de regiones intergénicas a excepción de la región control.

La disponibilidad de recursos genómicos de la subclase Copepoda es escasa y se aprecia un sesgo de información en los órdenes Calanoida y Cyclopoida (NCBI, 2022). Una de las características notorias del mitogenoma de esta subclase, es la diferencia que se presenta en su tamaño. La longitud del mitogenoma de *L. garciai* es semejante al de copépodos del orden Cyclopoida, Harpacticoida y Siphonostomatoida (~14,000 pb). Sin embargo, cuando se compara con copépodos calanoides, cercanos filogenéticamente, se puede observar una gran discordancia en tamaño; la variación que más destaca es la encontrada en *C. glacialis*, *C. finmarchicus* y *C. hyperboreus*. Esto mismo se puede apreciar en la Figura 18, en donde se muestra la reconstrucción del estado ancestral del tamaño de genoma. Se observa que el nodo ancestral del orden Calanoida presenta un tamaño elevado (nodo I), siendo este clado el que presenta la variación más grande dentro de la subclase. La diferencia en el tamaño del mitogenoma puede tener dos causas: la cantidad de genes y la longitud de la región intergénica (Wolstenholme, 1992).

Por lo general, la cantidad de genes del mitogenoma en copépodos es constante, sin embargo, se observa que algunos presentan una cantidad menor (Tabla 5). Entre estas diferencias, se destaca la pérdida de CDS de tres copépodos, *C. glacialis*, *P. rhincodonicus* y *P. nana*. En el caso de *P. nana*, Ki *et al.* (2009) reportaron la pérdida del gen ATP8 y la ganancia de un ARNt. Sin embargo, la variación de genes del copépodo *C. glacialis* es artificial, ya que el número reportado en el NCBI difiere de la publicación original (Weydmann *et al.*, 2017). Por lo tanto, debe de tomarse con precaución las afirmaciones de pérdida de genes dentro de la subclase y es poco probable que estas contribuyan de manera sustancial a la variación en el tamaño del genoma.

En cuanto a la región intergénica, se conoce que es la principal causa de la variación en el tamaño del mitogenoma de Copepoda (Weydmann *et al.*, 2017). Por ejemplo, el

copépodo *C. hyperboreus* presenta la zona intergénica más grande dentro de Crustacea (Kim *et al.*, 2013). Con los datos publicados hasta el momento, se observa que esta variación se encuentra únicamente en copépodos calanoides. Cabe resaltar que, de todos los mitogenomas de copépodos depositados en el NCBI, solo en algunos se tiene bien definida la zona de control (*L. cyprinacea*, *P. nana*, *P. rhincodonicus*, *T. kingsejongensis*, *C. glacialis* y *C. finmarchicus*). Por lo que respecta a este trabajo, no fue posible definir con exactitud el inicio y término de esta región y se necesita una validación con secuenciación dirigida para definir con certeza su posición y estructura. El analizar de manera más precisa esta zona sería valioso, ya que podría ser utilizada con el propósito de resolver relaciones filogenéticas problemáticas (Bronstein *et al.*, 2018) o, también, estudiar patrones de diversificación geográfica (Lin *et al.*, 2021).

Por último, se observa que el contenido de GC en los mitogenomas de Copepoda es variable y muy bajo, alcanzando un valor mínimo de 29.3%. Esto es similar al contenido de GC reportado en el genoma nuclear (30%), siendo una de las principales causas de la dificultad en el ensamblado de genomas nucleares en Copepoda (Jørgensen *et al.*, 2019).

5.1 Evolución nucleotídica

La tasa de sustitución sinónima y no sinónima de nucleótidos (ω) en el archivo concatenado de 11 CDS fue < 1 . Esto refleja que, en promedio, la región codificante de proteínas se encuentra en un escenario de selección purificadora. Este patrón es similar en distintas especies y se ha observado que la selección purificadora es una fuerza dominante en la evolución de la mitocondria (Meiklejohn *et al.*, 2007). Del mismo modo, se observa que las tasas Ka/Ks y de distancia- p entre especies filogenéticamente cercanas a *L. garciai* son más bajas en comparación con copépodos de distintos órdenes (Figura 6 y 8), lo cual se puede asociar a la afinidad filogenética de la especie.

Analizando los genes de manera independiente, se observa que el complejo COX, particularmente COX1 y COX3, presentan una menor tasa Ka/Ks entre todas las especies de copépodos. En contraste, se identificó que NAD1, NAD2, NAD3 y ATP6 podrían encontrarse en un escenario de selección positiva. Se puede observar que los genes que presentan las tasas más elevadas de Ka/Ks (Anexo L) corresponden a la comparación de *L. garciai* con copépodos del orden Cyclopoida. Aunque esto es una primera aproximación a las tasas evolutivas dentro del mitogenoma, estos resultados coinciden con un patrón similar reportado para el copépodo *T. californicus*, en donde las tasas más altas de divergencia se presentaban en los genes NAD3, NAD5, NAD6 y ATP6 (Barreto *et al.*, 2018). Aunque el patrón de sustitución en genes mitocondriales difiere en distintos grupos de metazoos, es común que los genes NAD se encuentren con una tasa de sustitución elevada. Esto se debe a la participación que tienen en el complejo 1 de la cadena de transporte de electrones, el cual es el de mayor tamaño en

el proceso de fosforilación oxidativa y es el que recibe una mayor cantidad de electrones en el proceso de oxidación y reducción (Fonseca *et al.*, 2008). Del mismo modo, se ha observado que los genes de ATP también presentan una alta tasa de divergencia y, en algunos casos, pueden ser eliminados del mitogenoma (Gissi *et al.*, 2008).

Por lo general, la selección mantiene las tasas de sustitución de los CDS conservadas, debido a la importancia que tienen en el proceso de fosforilación oxidativa (OXPHOS) y la comunicación nuclear-mitocondrial para llevar a cabo este proceso (Reinecke *et al.*, 2009). Sin embargo, algunos de estos genes, sí pueden encontrarse bajo un proceso de evolución adaptativa. Por ejemplo, se ha asociado la variación en estos genes a una reducción de energía generada por la mitocondria (ATP), lo cual repercute en el almacenamiento de lípidos y la velocidad a la cual estos son utilizados por el organismo, sirviendo como un mecanismo de respuesta a la inanición en algunos ambientes (Ballard & Melvin, 2010). Del mismo modo, es común que existan genes con tasas más elevadas de sustitución en comparación con otros dentro de un mismo mitogenoma, lo cual puede estar relacionado a la adaptación de una especie al estrés ambiental (Ramos *et al.*, 2020).

Aunque no se puede afirmar que la variación a nivel nucleotídico en copépodos responda a una demanda similar, se conoce que muchas de las especies se encuentran en ambientes ecológicamente estresantes. Especialmente se conoce que *L. garciai* se encuentra adaptado a las condiciones de alimento bajas en su ambiente (Ciros-Pérez *et al.*, 2015), y sería interesante investigar si las variaciones encontradas en su genoma se pueden asociar a respuestas fenotípicas (p. ej., la reserva de lípidos) que indiquen realmente evolución adaptativa.

5.2 Evolución de la arquitectura del mitogenoma

Es común que el orden genético se mantenga invariable y que los eventos de reordenamientos sean episodios extraordinarios (Boore, 1999). Comúnmente, dentro de Crustacea se utiliza como primera línea comparativa el patrón basal de pancrustacea (Boore *et al.*, 1998), ya que generalmente los cambios en Crustacea son en el reordenamiento de pocos ARNt (Shen *et al.*, 2010) hasta inclusive un solo ARNt (Ji *et al.*, 2014). En este trabajo, no se realizó la comparación con el modelo de Pancrustacea ya que se conoce que los copépodos tienen un arreglo sustancialmente diferente al patrón basal, tanto en ARNt como en CDS (Weydmann *et al.*, 2017), siendo estos eventos algo común dentro de la subclase (Hickerson & Cunningham, 2000; Zhang *et al.*, 2020).

Así, la arquitectura del mitogenoma de *L. garciai* no es semejante a la de ningún copépodo descrito hasta ahora. Las especies emparentadas más cercanamente a *L. garciai* son *P. tunguidus* y *C. hyperboreus*, siendo el genoma de la primera la que cuenta con un menor número de eventos de reordenamiento. A parte de esta especie,

cuando se compara el mitogenoma de *L. garciai* con los demás copépodos, se observa que la similitud en su ordenamiento genético es muy baja (Figura 10, Anexo K). A su vez, las especies emparentadas no siempre tienen un ordenamiento genético similar, por ejemplo, se presenta una mayor discordancia entre *T. californicus* y *T. kingsejongensis* que entre *T. kingsejongensis* y *A. atopus* (Figura 10, Anexo K). Hasta el momento, solo dos especies presentan una sintenia idéntica en cuanto a CDS, es decir, las dos del género *Sinergasilus*.

En los escenarios evolutivos obtenidos mediante CREx, se observa que muchos de los cambios en el ordenamiento del mitogenoma de Copepoda pueden explicarse con uno o más eventos de TDRL. Este resultado es aún preliminar y hace falta la comparación entre especies cercanas, así como una reconstrucción del ordenamiento mitocondrial de un ancestro común a nivel de familia y orden. No obstante, los eventos aquí encontrados nos indican que el mitogenoma en copépodos ha estado sujeto a grandes eventos de duplicación a lo largo de su historia evolutiva. Es posible, dada la plasticidad en la arquitectura del genoma de copépodos, que estos eventos, en principio, no sean adaptativos (Boussau *et al.*, 2011). Esto debido a que todos los genes mitocondriales se transcriben de manera policistrónica (Dowton *et al.*, 2002), por lo cual, el que se presente una secuencia diferente no afecta el resultado final de la transcripción del mitogenoma. Una posible explicación a la arquitectura divergente de los mitogenomas, es que éstas hayan sido fijado por deriva genética o cuellos de botella (Xia *et al.*, 2016).

Aunque es poco probable que estos eventos de reordenamiento tengan una repercusión directa en la eficacia de las especies, sí la pueden tener en especies que se encuentren en las primeras etapas de especiación, ya que se tiene evidencia que las arquitecturas mitocondriales divergentes pueden ocasionar una barrera precigótica (Ellison & Burton, 2008) o postcigótica (Healy & Burton, 2020) debido a incompatibilidades mito-nucleares. Este último caso podría explorarse con mayor detalle utilizando el complejo de especies de *Leptodiaptomus* grupo *sicilis*, filogenéticamente cercanas a *L. garciai*, las cuales se encuentran en un proceso avanzado de especiación ecológica (Barrera-Moreno *et al.*, 2015).

Por último, la región genética peculiar encontrada en copépodos (Figura 11) no contiene espacios intergénicos ni tampoco copias de ARN entre genes, encontrándose presente en todos los copépodos calanoides analizados en este trabajo. Anteriormente se había mencionado este patrón por parte de Kim *et al.* (2013), como algo característico de organismos pertenecientes al superorden Gymnoplea (Calanoida), y que posiblemente también se encontrara presente en Podoplea (Cyclopoida, Harpacticoida, Siphonostomatoida, entre otros) como un rasgo conservado de copépodos. De manera contrastante con la información recopilada en este trabajo, se observa que la región COX3/ATP6/ATP8 es conservada exclusivamente en copépodos calanoides y únicamente *T. kingsejongensis* (Harpacticoida, en Podoplea) presenta un patrón parecido (COX3/ATP8/ATP6). Aunque la información genómica aún es escasa, este patrón podría ser tomado como una sinapomorfía del orden Calanoida.

5.3 Análisis filogenómicos

En el pasado, la especie *L. garciai* había sido identificada como *L. novamexicanus* (Herrick, 1895) Años más tarde, mediante el análisis de diferencias morfológicas, fisiológicas y genéticas, Montiel-Martínez *et al.* (2008) concluyeron que *L. garciai* es una especie biológica independiente; reportando su afinidad filogenética con especies congéneres a través del marcador molecular COX1. En este trabajo, se utilizó la información genómica obtenida de *L. garciai* para inferir las relaciones cladísticas de este copépodo utilizando los recursos genómicos disponibles. Todas las filogenias realizadas, agruparon a *L. garciai* en el clado del orden Calanoida, compartiendo un ancestro en común con el copépodo *P. tunguidus*. Estos análisis confirman la identidad del taxón y su afinidad taxonómica a copépodos del orden Calanoida. Del mismo modo, se observa que el uso de las tres combinaciones de genes (A, B y C), agrupó a *L. garciai* con copépodos de la misma familia, por lo cual, la información genómica obtenida en este trabajo podría ser relevante para la resolución de relaciones filogenéticas dentro del orden Calanoida.

Las filogenias obtenidas mediante el método de MV presentaron un soporte muy bajo de nodos, sin embargo, se destaca la filogenia de la combinación C (Figura 16). Este árbol recupera a Calanoida como grupo monofilético, hermano de Harpacticoida + Siphonostomatoida + Cyclopoida. Lo cual es similar a las filogenias construidas mediante caracteres morfológicos (Ho, 1990; Boxshall, 1999) y moleculares (concatenado de 24 genes nucleares; Eyun, 2017; Figura 20-A), que sugieren a Calanoida como clado basal monofilético, hermano de los órdenes mencionados anteriormente. La topología de este árbol es la más parecida a las relaciones taxonómicas de Copepoda aceptadas actualmente (Eyun, 2017).

En cuanto a las filogenias obtenidas mediante IB, se observa que las tres combinaciones de genes generaron árboles con topologías similares, siendo la filogenia con la combinación C (Fig. 17) donde se observa el mayor soporte estadístico de nodos. En la figura 19 se muestra este árbol filogenético donde se denotan las relaciones a nivel de orden. Se destaca que el árbol divide a las especies de Cyclopoida y Harpacticoida en ramas distintas y agrupa a Calanoida + parte de Cyclopoida + parte de Harpacticoida en un solo clado. Cuando se compara la filogenia de la combinación C con otras filogenias construidas a partir de genes mitocondriales, se observa que es distinta a lo reportado cuando se utilizan 13 CDS + genes ribosomales (Hua *et al.*, 2021; Figura 20-B), pero idéntica cuando se utilizan 13 CDS mitocondriales (Jooste *et al.*, 2019; Zhang *et al.*, 2020; Figura 20-C). Aunque este árbol concuerda con datos publicados, las topologías de estas filogenias podrían estar influenciadas por cambios no neutrales en genes con tasas de evolución elevadas, como p. ej., en el complejo de genes NAD o ATP. Aunado a esto, en estos árboles se esperaría encontrar cohesión de

clados a nivel de orden, sin embargo, no es el caso y se considera que este análisis no refleja de manera satisfactoria su historia evolutiva.

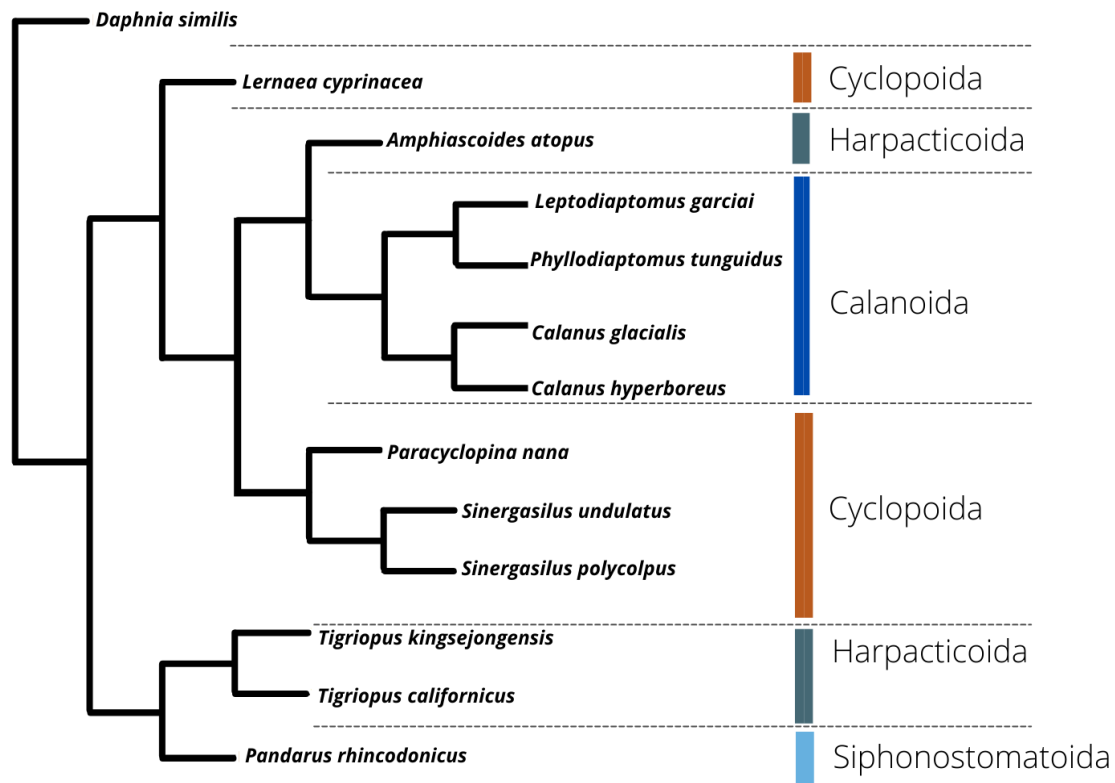


Figura 19. Cladograma obtenido mediante el análisis de IB (Fig. 18) mostrando la relación de las especies a nivel de orden.

Los análisis taxonómicos en la subclase Copepoda han sido controversiales en los últimos años, principalmente por las distintas propuestas filogenéticas a nivel de Orden. Una de las principales causas de ello es que, la mayoría de las investigaciones de taxonomía molecular de la subclase utiliza el marcador genético COX 1 propuesto en el proyecto de código de barras (600-700 pb; Hebert *et al.*, 2003). Esto ha generado que los recursos genéticos disponibles en Copepoda se limiten a utilizar este marcador, sin embargo, se sabe que presenta una baja eficiencia de resolver relaciones filogenéticas a una escala taxonómica mayor (Lefébure *et al.*, 2006). Del mismo modo, la mayoría de los trabajos filogenéticos se especializan en resolver relaciones taxonómicas a nivel de especie, género o familia, por lo cual existen pocos trabajos que establezcan inferencias evolutivas a nivel de Orden, dificultando una comparación más extensa.

Aunque los copépodos son uno de los grupos más abundantes de sistemas acuáticos y brindan servicios ecológicos en distintos ambientes, su relación taxonómica a nivel de orden aún es problemática y se han realizado pocos trabajos para tratar de resolver su estado actual. Aunque, utilizar el concatenado de 13 CDS mitocondriales es debatible debido a las distintas topologías que resultaron en este trabajo, es importante comenzar a explorar distintas metodologías para establecer relaciones evolutivas

confiables. Particularmente, porque la mayoría de las relaciones cladísticas de Copepoda se basan en filogenias de genes particulares. El avance de las técnicas de secuenciación masiva ha generado un nuevo campo de inferencia filogenómica (Edwards, 2009) y recientemente se ha visto que al utilizar una mayor cantidad de genes se pueden resolver problemas taxonómicos importantes, generando árboles filogenéticos de especies confiables (Jarvis *et al.*, 2014). Aunque los datos genómicos aún son escasos, es probable que la disponibilidad de información genómica vaya en aumento en los próximos años, conjuntando datos genómicos mitocondriales y nucleares, además de una mayor cantidad de especies, podría ayudar al establecimiento de relaciones evolutivas a un mayor nivel de resolución (Xi *et al.*, 2016).

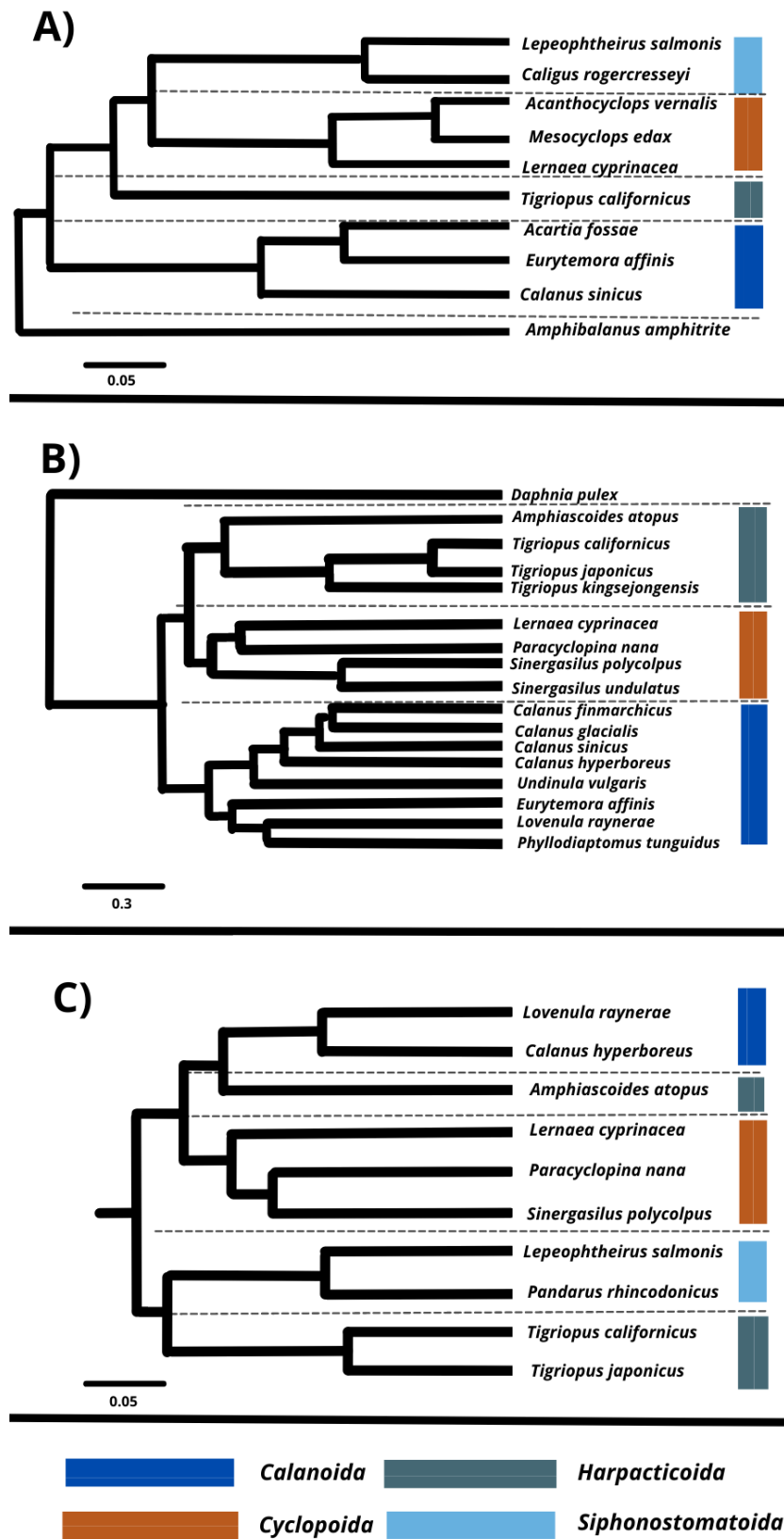


Figura 20. Comparación de las relaciones filogenéticas de Copepoda utilizando A) genes nucleares (Eyun, 2017), B) ARN ribosomales (Hua *et al.*, 2021) y C) 13 CDS mitocondriales (Jooste *et al.*, 2019).

5.4 Patrones evolutivos del mitogenoma de Copepoda

La característica más llamativa del genoma de las especies de copépodos analizadas en este trabajo son los cambios encontrados en su arquitectura genética. Aunque no se conoce con exactitud si la sintenia del mitogenoma puede tener repercusión en la eficacia de las especies, se ha reportado que cambios en tasas de sustitución nucleotídica se correlacionan con eventos de reordenamiento genómico, sin embargo, no se conoce con exactitud la causalidad entre ellos (Bernt *et al.*, 2013). Los eventos de reordenamiento mitocondrial no siguen una tendencia similar al reloj molecular (Podsiadlowski *et al.*, 2014), esto se ve reflejado en Copepoda, donde los cambios en la arquitectura del mitogenoma son extensos y al parecer no se encuentran relacionados con ancestría común entre especies. Sin embargo, sí existen patrones generales (COX3/ATP8/ATP6) que reflejan una historia compartida del clado, al menos en Calanoida.

Observando los resultados obtenidos en este trabajo, es probable que los cambios en el mitogenoma de los copépodos (sustitución nucleotídica y sintenia) se originaran a partir de distintas fuerzas evolutivas. Se ha considerado que los cambios en el mitogenoma de las especies pueden ser causados por deriva genética (Bazin *et al.*, 2006) o por evolución adaptativa (Xia *et al.*, 2016). En el caso de los copépodos, el efecto fundador se encuentra presente en los repetidos eventos de colonización de ambientes marinos hacia continentales (Boxshall & Jaume, 2000). Estos eventos pudieron haber propiciado la fijación de arquitecturas genéticas novedosas que, a su vez, podrían fungir como barreras reproductivas. En principio, estos procesos son evolutivamente neutrales, empero se conoce que cambios en regiones específicas, como en la región control o en el ordenamiento de genes, pueden incidir en la tasa de mutación (Minelli *et al.*, 2016). Esto se debe a que alteraciones en la replicación del genoma pueden aumentar procesos como la desaminación de nucleótidos (Marín & Xia, 2008; Zhang *et al.*, 2020).

Por otro lado, la divergencia nucleotídica pudo haber sido originada por evolución adaptativa, ya que cambios en el genoma mitocondrial pueden ser seleccionados de manera positiva por su incidencia en la eficacia biológica de las especies (Ueno & Takahashi, 2021). En el copépodo *Tigriopus californicus*, se ha encontrado que la variación en genes mitocondriales puede generar tasas de oxidación diferenciales en el sistema de transporte de electrones (Harrison & Burton, 2006), así como disminuir la eficacia de híbridos debido a incompatibilidades mito-nucleares (Healy & Burton, 2020). Es probable que situaciones similares a esta, se encuentren en distintas especies de copépodos habitando diversos ambientes, por lo cual, sería de interés conocer si cambios en genes particulares pueden asociarse a ambientes específicos.

Del mismo modo, el mantenimiento de variación genética en el mitogenoma, algunas veces catalogada como neutral, puede ocasionar cambios fenotípicos importantes que, a su vez, puede incrementar la eficacia de las especies (Dowling *et al.*, 2008). Un

ejemplo de esto es el aumento de la eficiencia del sistema de fosforilación oxidativa, lo cual reduce los requerimientos energéticos de una población (Ramos *et al.*, 2018).

Aunque los eventos de reordenamiento en Copepoda podrían tener una naturaleza evolutiva neutral, la interrupción en arquitecturas genómicas podría ocasionar tasas de evolución diferenciales a lo largo del mitogenoma. Esta variación a su vez podría estar moldeada por selección direccional —si se tiene un efecto positivo en la eficacia de los organismos— o por selección purificadora (Figura 21). Por otro lado, es probable que las filogenias realizadas únicamente a partir de genes mitocondriales no reconstruyan con precisión relaciones ancestro-descendiente, debido a que, en algunos casos, los polimorfismos existentes podrían estar reflejando adaptaciones al ambiente.

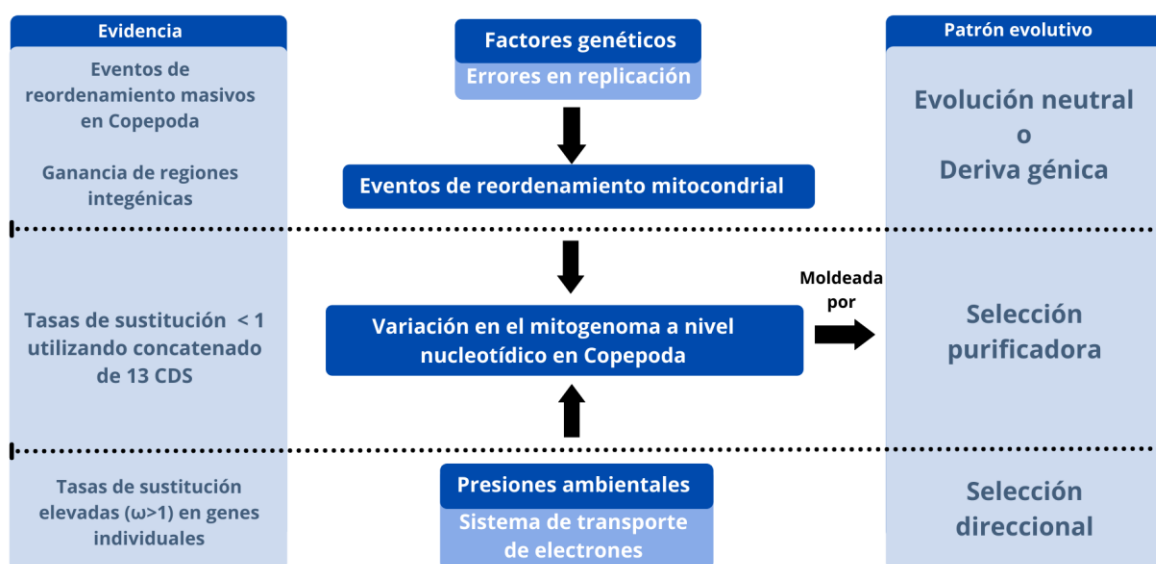


Figura 21. Relación de la variación del mitogenoma en Copepoda con factores intrínsecos (genéticos) y extrínsecos (ambientales), así como el patrón de evolución que mantienen.

Para terminar, me gustaría hacer un llamado a la investigación de manera extensiva de índole genómico en copépodos. Se ha observado que algunas especies de este grupo muestran patrones genómicos similares en respuesta a fluctuaciones ambientales, ocasionando una respuesta positiva a la colonización de nuevos nichos ecológicos (Stern & Lee, 2020). Cambios específicos a nivel de locus en respuesta a cambios ambientales, se han estudiado con anterioridad en otras especies de animales y se han vuelto realmente importantes para predecir la respuesta que tendrán algunos taxones frente al cambio climático (Fitzpatrick & Edelsparre, 2018). El generar conocimiento sobre la evolución de especies como copépodos, utilizando información genómica y ecológica, no solo ayudará a entender patrones complejos de evolución sino que además, puede ayudar a establecer estrategias claras de conservación, generando modelos ecológicos-evolutivos (Waldvogel *et al.*, 2020).

6. Conclusión

Se obtuvo el primer borrador del genoma mitocondrial de *L. garciai*, el cual es un ensamblaje circularizado con una longitud de 14,654 pb. El mitogenoma contiene un total de 34 genes, entre ellos, 13 CDS, 19 ARNt y 2 ARNr. Las características genéticas del mitogenoma son similares a las de otros copépodos, sin embargo, se presenta una gran diferencia en cuanto a su tamaño en comparación con especies del mismo orden. Comparando los CDS de diferentes copépodos, se identificó que los genes NAD (1-3) y ATP6 presentan las tasas más elevadas de divergencia nucleotídica. En cuanto a la arquitectura del ADN mitocondrial, el ordenamiento genético de *L. garciai* es único dentro de la subclase. Además, se observan múltiples eventos de reordenamiento en los mitogenomas de Copepoda. Utilizando la información mitogenómica de *L. garciai*, se obtuvieron distintas filogenias, las cuales demostraron las afinidades taxonómicas de *L. garciai* dentro del orden Calanoida y dentro de Copepoda, en general.

7. Referencias bibliográficas

- Alcocer, J., Lugo, A., Escobar, E., Sánchez, M. D. R., & Vilaclara, G. (2000). Water column stratification and its implications in the tropical warm monomictic Lake Alchichica, Puebla, Mexico. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen*, 27(5), 3166-3169.
- Amato, A., & Carotenuto, Y. (2018). Planktonic calanoids embark into the “omics” era. *Trends in Copepod Studies—Distribution, Biology and Ecology* Uttieri, M., Ed, 287-314.
- Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Armienta, M. A., Vilaclara, G., De la Cruz-Reyna, S., Ramos, S., Cenicerros, N., Cruz, O., & Arcega-Cabrera, F. (2008). Water chemistry of lakes related to active and inactive Mexican volcanoes. *Journal of Volcanology and Geothermal Research*, 178(2), 249-258.
- Ballard, J. W. O., & Melvin, R. G. (2010). Linking the mitochondrial genotype to the organismal phenotype. *Molecular Ecology*, 19(8), 1523-1539.
- Barrera-Moreno, O. A., Ciroso-Pérez J., Ortega-Mayagoitia E., Alcántara-Rodríguez J. A., & Piedra-Ibarra E. (2015). From local adaptation to ecological speciation in copepod populations from neighboring lakes. *PLoS One*, 10(4): e0125524.
- Barreto, F. S., Watson, E. T., Lima, T. G., Willett, C. S., Edmands, S., Li, W., & Burton, R. S. (2018). Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. *Nature Ecology & Evolution*, 2(8), 1250-1257.
- Bazin, E., Glémin, S., & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, 312(5773), 570-572.
- Bernt, M., Braband, A., Schierwater, B., & Stadler, P. F. (2013). Genetic aspects of mitochondrial genome evolution. *Molecular Phylogenetics and Evolution*, 69(2), 328-338.
- Bernt, M., Chen, K. Y., Chen, M. C., Chu, A. C., Merkle, D., Wang, H. L., Chao, K., & Middendorf, M. (2011). Finding all sorting tandem duplication random loss operations. *Journal of Discrete Algorithms*, 9(1), 32-48.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., & Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2), 313-319.

- Bernt, M., Merkle, D., Ramsch, K., Fritsch, G., Perseke, M., Bernhard, D., & Middendorf, M. (2007). CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics*, 23(21), 2957-2958.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, 170.
- Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, 27(8), 1767-1780.
- Boore, J. L., & Brown, W. M. (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics & Development*, 8(6), 668-674.
- Boore, J. L., Lavrov, D. V., & Brown, W. M. (1998). Gene translocation links insects and crustaceans. *Nature*, 392(6677), 667-668.
- Boussau, B., Brown, J. M., & Fujita, M. K. (2011). Nonadaptive evolution of mitochondrial genome size. *Evolution: International Journal of Organic Evolution*, 65(9), 2706-2711.
- Boxshall, G. A., & Defaye, D. (2007). Global diversity of copepods (Crustacea: Copepoda) in freshwater. *Freshwater Animal Diversity Assessment*, 195-207, Springer, Dordrecht.
- Boxshall, G. A., & Jaume, D. (2000). Making waves: the repeated colonization of fresh water by copepod crustaceans. *Advances in Ecological Research*, 31, 61-79.
- Braga, E., Zardoya, R., Meyer, A., & Yen, J. (1999). Mitochondrial and nuclear rRNA based copepod phylogeny with emphasis on the Euchaetidae (Calanoida). *Marine Biology*, 133(1), 79-90.
- Brito, P. H., & Edwards, S. V. (2009). Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, 135(3), 439-455.
- Bron, J. E., Frisch, D., Goetze, E., Johnson, S. C., Lee, C. E., & Wyngaard, G. A. (2011). Observing copepods through a genomic lens. *Frontiers in Zoology*, 8(1), 1-15.
- Bronstein, O., Kroh, A., & Haring, E. (2018). Mind the gap! The mitochondrial control region and its power as a phylogenetic marker in echinoids. *BMC Evolutionary Biology*, 18(1), 1-15.
- Can-Chulim, Á., Ortega-Escobar, H. M., García-Calderón, N. E., Reyes-Ortigoza, A. L., González-Hernández, V., & Flores-Román, D. (2011). Origen y calidad del agua subterránea en la cuenca oriental de México. *Terra Latinoamericana*, 29(2), 189-200.

- Carreto-Romero, J. M. (2019). Análisis del tamaño del genoma de copépodos calanoides en lagos vecinos en la Cuenca Oriental Puebla. Tesis de Licenciatura. Carrera de Biología. FES Iztacala. Universidad Nacional Autónoma de México.
- Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021) tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *BioRxiv*, 614032
- Chinnery, P. F., & Hudson, G. (2013). Mitochondrial genetics. *British Medical Bulletin*, 106(1), 135-159.
- Choquet, M., Smolina, I., Dhanasiri, A. K., Blanco-Bercial, L., Kopp, M., Jueterbock, A., & Hoarau, G. (2019). Towards population genomics in non-model species with large genomes: a case study of the marine zooplankton *Calanus finmarchicus*. *Royal Society Open Science*, 6(2), 180608.
- Ciros-Pérez, J., Ortega-Mayagoitia, E., & Alcocer, J. (2015). The role of ecophysiological and behavioral traits in structuring the zooplankton assemblage in a deep, oligotrophic, tropical lake. *Limnology and Oceanography*, 60(6), 2158-2172.
- Copley, R. R. (2008). The animal in the genome: comparative genomics and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1496), 1453-1461.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2).
- Darwin, C. (1859). *On the origins of species by means of natural selection*. London: Murray.
- Di Capua, I., Maffucci, F., Pannone, R., Mazzocchi, M. G., Biffali, E., & Amato, A. (2017). Molecular phylogeny of Oncaeidae (Copepoda) using nuclear ribosomal internal transcribed spacer (ITS rDNA). *PLoS One*, 12(4), e0175662.
- Dowling, D. K., Friberg, U., & Lindell, J. (2008). Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology & Evolution*, 23(10), 546-554.
- Dowton, M., Castro, L. R., & Austin, A. D. (2002). Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome morphology'. *Invertebrate Systematics*, 16(3), 345-356.
- Dumont, H. J., Van de Velde, I., & Dumont, S. (1975). The dry weight estimate of biomass in a selection of Cladocera, Copepoda and Rotifera from the plankton, periphyton and benthos of continental waters. *Oecologia*, 19(1), 75-97.

- Ebert, D. (2011). A genome for the environment. *Science*, 331(6017), 539-540.
- Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging?. *Evolution: International Journal of Organic Evolution*, 63(1), 1-19.
- Elías-Gutiérrez, M., Suárez-Morales, E., Gutiérrez-Aguirre, M. A., Silva-Briano, M., Granados-Ramírez, J., & Garfias-Espejo, T. (2008). *Cladocera y copepoda de las aguas continentales*. CONABIO, UNAM, México.
- Ellison, C. K., & Burton, R. S. (2008). Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution*, 62(3), 631-638.
- Eyun, S. I. (2017). Phylogenomic analysis of Copepoda (Arthropoda, Crustacea) reveals unexpected similarities with earlier proposed morphological phylogenies. *BMC Evolutionary Biology*, 17(1), 1-12.
- Figueroa, J. L., Borrego-Enriquez, L. E., Castillo-Dominguez, R. M., & Valladolid-Laredo, M. A. (1983). Batimetria y morfometria de los lagos "maars" de la Cuenca Oriental, Puebla, México. *Biotica*, 8, 37-47.
- Fitzpatrick, M. J., & Edelsparre, A. H. (2018). The genomics of climate change. *Science*, 359(6371), 29-30.
- Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J., & Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, 9(1), 1-22.
- Gabaldón, T., & Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, 14(5), 360-366.
- Gissi, C., Iannelli, F., & Pesole, G. (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, 101(4), 301-320.
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075.
- H. R., Boxshall G (1991) Copepod evolution. The Ray Society, London
- Haddrill, P. R., Bachtrog, D., & Andolfatto, P. (2008). Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Molecular Biology and Evolution*, 25(9), 1825-1834.
- Hall, M. B. (2019) Rasusa: Randomly subsample sequencing reads to a specified coverage.

- Harrison, J. S., & Burton, R. S. (2006). Tracing hybrid incompatibilities to single amino acid substitutions. *Molecular Biology and Evolution*, 23(3), 559-564.
- Havird, J. C., & Santos, S. R. (2014). Performance of single and concatenated sets of mitochondrial genes at inferring metazoan relationships relative to full mitogenome data. *PLoS One*, 9(1), e84080.
- Healy, T. M., & Burton, R. S. (2020). Strong selective effects of mitochondrial DNA on the nuclear genome. *Proceedings of the National Academy of Sciences*, 117(12), 6616-6621.
- Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512), 313-321.
- Herrick, C. L. (1895). *Microcrustacea from New Mexico*. VEB G. Fischer Verlag.
- Hernández-Martínez, O. (2014). Variación temporal de la biología reproductiva de *Leptodiaptomus garciai* (Copepoda: Calanoida) en el lago Alchichica (Puebla, México). Tesis de Maestría. Posgrado en ciencias del mar y limnología. Universidad Nacional Autónoma de México.
- Hickerson, M. J., & Cunningham, C. W. (2000). Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, Anomura). *Molecular Biology and Evolution*, 17(4), 639-644.
- Hill, G. E. (2015). Mitonuclear ecology. *Molecular Biology and Evolution*, 32(8), 1917-1927.
- Ho, J. S. (1990). Phylogenetic analysis of copepod orders. *Journal of Crustacean Biology*, 10(3), 528-536.
- Hua, C. J., Su, M. Y., Sun, Z. W., Lu, Y. H., & Feng, J. M. (2021). Complete mitochondrial genome of the copepod *Sinergasilus undulates* (Copepoda: Poecilostomatoida). *Mitochondrial DNA Part B*, 6(3), 1226-1228.
- i5K Consortium. (2013). The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, 104(5), 595-600.
- Jablonka, E. (1994). Inheritance systems and the evolution of new levels of individuality. *Journal of Theoretical Biology*, 170(3), 301-309.
- Jablonka, E., & Lamb, M. J. (2014). *Evolution in four dimensions, revised edition: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. MIT press.

- Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y., Faircloth, B. C., Nabholz, B., Howard, J. T., Suh, A., Weber, C. C., Fonseca, R. R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L.,... & Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346(6215), 1320-1331.
- Ji, Y. K., Wang, A., Lu, X. L., Song, D. H., Jin, Y. H., Lu, J. J., & Sun, H. Y. (2014). Mitochondrial genomes of two brachyuran crabs (Crustacea: Decapoda) and phylogenetic analysis. *Journal of Crustacean Biology*, 34(4), 494-503.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., & Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55-61.
- Jooste, C. M., Emami-Khoyi, A., Gan, H. M., Wasserman, R. J., Dalu, T., & Teske, P. R. (2019). The complete mitochondrial genome of Africa's largest freshwater copepod, *Lovenula raynerae*. *Mitochondrial DNA Part B*, 4(1), 725-727.
- Jørgensen, T. S., Petersen, B., Petersen, H. C. B., Browne, P. D., Prost, S., Stillman, J. H., & Hansen, B. W. (2019). The genome and mRNA transcriptome of the cosmopolitan calanoid copepod *Acartia tonsa dana* improve the understanding of copepod genome size evolution. *Genome Biology and Evolution*, 11(5), 1440-1450.
- Kalyanaraman, B., Cheng, G., Hardy, M., Ouari, O., Lopez, M., Joseph, J., & Dwinell, M. B. (2018). A review of the basics of mitochondrial bioenergetics, metabolism, and related signaling pathways in cancer cells: Therapeutic targeting of tumor mitochondria with lipophilic cationic compounds. *Redox Biology*, 14, 316-327.
- Ki, J. S., Park, H. G., & Lee, J. S. (2009). The complete mitochondrial genome of the cyclopoid copepod *Paracyclops nana*: a highly divergent genome with novel gene order and atypical gene numbers. *Gene*, 435(1-2), 13-22.
- Kim, S., Lim, B. J., Min, G. S., & Choi, H. G. (2013). The complete mitochondrial genome of Arctic *Calanus hyperboreus* (Copepoda, Calanoida) reveals characteristic patterns in calanoid mitochondrial genome. *Gene*, 520(1), 64-72.
- Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genetics*, 4(12), e1000304.
- Lambert, S. M., Reeder, T. W., & Wiens, J. J. (2015). When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. *Molecular Phylogenetics and Evolution*, 82, 146-155.

- Lander, E. S., & Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics*, 2(3), 231-239.
- Langmead, B., & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357-359.
- Lavrov, D. V. (2007). Key transitions in animal evolution: a mitochondrial DNA perspective. *Integrative and Comparative Biology*, 47(5), 734-743.
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7), 1307-1320.
- Lefébure, T., Douady, C. J., Gouy, M., & Gibert, J. (2006). Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, 40(2), 435-447.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094-3100.
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26(5), 589-595.
- Lin, Z. J., Wang, X., Wang, J., Tan, Y., Tang, X., Werren, J. H., & Wang, X. (2021). Comparative analysis reveals the expansion of mitochondrial DNA control region containing unusually high GC tandem repeat arrays in *Nasonia vitripennis*. *International Journal of Biological Macromolecules*, 166, 1246-1257.
- Lorenz, R., Bernhart, S. H., Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). "ViennaRNA Package 2.0", *Algorithms for Molecular Biology*, 6(1), 1-26.
- Lugo, A., González, M. E., del Rosario Sánchez, M., & Alcocer, J. (1999). Distribution of *Leptodiaptomus novamexicanus* (Copepoda: Calanoida) in a Mexican hyposaline lake. *Revista de Biología Tropical*, 141-148.
- Marín, A., & Xia, X. (2008). GC skew in protein-coding genes between the leading and lagging strands in bacterial genomes: new substitution models incorporating strand bias. *Journal of Theoretical Biology*, 253(3), 508-513.
- McBride, H. M., Neuspiel, M., & Wasiak, S. (2006). Mitochondria: more than just a powerhouse. *Current Biology*, 16(14), 551-560.
- Meiklejohn, C. D., Montooth, K. L., & Rand, D. M. (2007). Positive and negative selection on the mitochondrial genome. *Trends in Genetics*, 23(6), 259-263.

- Meng, G., Li, Y., Yang, C., & Liu, S. (2019). MitoZ: A toolkit for mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, 47(11), 63.
- Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95(6), 315-327.
- Minelli, A., Boxshall, G., & Fusco, G. (2016). *Arthropod Biology and Evolution*. Springer-Verlag Berlin An.
- Montiel-Martínez, A., Ciro-Pérez, J., Ortega-Mayagoitia, E., & Elías-Gutiérrez, M. (2008). Morphological, ecological, reproductive and molecular evidence for *Leptodiptomus garciai* (Osorio-Tafall 1942) as a valid endemic species. *Journal of Plankton Research*, 30(10), 1079-1093.
- National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [citada 2022 enero 18]. Available from: <https://www.ncbi.nlm.nih.gov/>
- Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3(5), 418-426.
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review Genetics*, 39, 197-218.
- Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, 6(2), 119-127.
- Ortega-Mayagoitia, E., Hernández-Martínez, O., & Ciro-Pérez, J. (2018). Phenotypic plasticity of life-history traits of a calanoid copepod in a tropical lake: Is the magnitude of thermal plasticity related to thermal variability? *PLoS One*, 13(4), e0196496.
- Ortega-Mayagoitia, E., J.A. Alcántara-Rodríguez, J. Urbán-Olivares, and J. Ciro-Pérez. 2022. Genomic signatures of divergent selection in lacustrine copepods. *Freshwater Biology* 67: 1045–1062.
- Osorio-Tafall, B. F. (1942). Un nuevo Diaptomus del México central (Copepoda, Diaptomidae). *Revista Brasileira de Zoologia*, 2(2), 147-154.
- Park, S. T., & Kim, J. (2016). Trends in next-generation sequencing and a new era for whole genome sequencing. *International Neurourology Journal*, 20(2), 76.
- Paszkiwicz, K., & Studholme, D. J. (2010). De novo assembly of short sequence reads. *Briefings in Bioinformatics*, 11(5), 457-472.

- Phillippy, A. M., Schatz, M. C., & Pop, M. (2008). Genome assembly forensics: finding the elusive mis-assembly. *Genome Biology*, 9(3), 1-13.
- Pigliucci, M. (2007). Do we need an extended evolutionary synthesis? *Evolution: International Journal of Organic Evolution*, 61(12), 2743-2749.
- Podsiadlowski, L., Mwinyi, A., Lesný, P., & Bartolomaeus, T. (2014). 19 Mitochondrial gene order in Metazoa—theme and variations. *Deep Metazoan Phylogeny: The Backbone of the Tree of Life*, 459-472.
- Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C. Y., & Hackett, K. (2015). The i5k Workspace@ NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research*, 43(D1), D714-D719.
- Ramos, B., González-Acuña, D., Loyola, D. E., Johnson, W. E., Parker, P. G., Massaro, M., Dantas, G., Miranda M. D., & Vianna, J. A. (2018). Landscape genomics: natural selection drives the evolution of mitogenome in penguins. *BMC Genomics*, 19(1), 1-17.
- Reinecke, F., Smeitink, J. A., & Van Der Westhuizen, F. H. (2009). OXPHOS gene expression and control in mitochondrial disorders. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1792(12), 1113-1121.
- Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217-223.
- Rocha, G. M., Martinelli, A. B., Utz, L. R. P., & Fernandes, L. F. L. (2019). Using internal transcribed spacers for Calanoida (Crustacea, Copepoda) species identification. *Marine Biodiversity*, 49(3), 1565-1571.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299-3302.
- Salinas-Giegé, T., Giegé, R., & Giegé, P. (2015). tRNA biology in mitochondria. *International Journal of Molecular Sciences*, 16(3), 4518-4559.
- Schubert, Lindgreen, & Orlando (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, 9(1), 88.
- Shadel, G. S., & Clayton, D. A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annual review of biochemistry*, 66(1), 409-435.

- Sheffer, M. M., Hoppe, A., Krehenwinkel, H., Uhl, G., Kuss, A. W., Jensen, L., & Prost, S. (2021). Chromosome-level reference genome of the European wasp spider *Argiope bruennichi*: a resource for studies on range expansion and evolutionary adaptation. *Giga Science*, 10(1).
- Shen, X., Wang, H., Ren, J., Tian, M., & Wang, M. (2010). The mitochondrial genome of *Euphausia superba* (Prydz Bay)(Crustacea: Malacostraca: Euphausiacea) reveals a novel gene arrangement and potential molecular markers. *Molecular Biology Reports*, 37(2), 771-784.
- Shubin, N. (2020). *Some assembly required: decoding four billion years of life, from ancient fossils to DNA*. Vintage.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212.
- Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-3.0*. 1996-2010
<http://www.repeatmasker.org>
- Smit, AFA, Hubley, R. *RepeatModeler Open-1.0*. 2008-2015
<http://www.repeatmasker.org>.
- Smith, D. R., & Keeling, P. J. (2015). Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences*, 112(33), 10177-10184.
- Smith, J. M., & Szathmary, E. (1997). *The major transitions in evolution*. Oxford University Press.
- Stanke, M., y Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33, 465-467.
- Stern, D. B., & Lee, C. E. (2020). Evolutionary origins of genomic adaptations in an invasive copepod. *Nature Ecology & Evolution*, 4(8), 1084-1094.
- Taanman, J. W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1410(2), 103-123.
- Tang, M., Tan, M., Meng, G., Yang, S., Su, X. U., Liu, S., & Zhou, X. (2014). Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Research*, 42(22), 166-166.

- Taylor, D. J., Hebert, P. D., & Colbourne, J. K. (1996). Phylogenetics and Evolution of the *Daphnia longispina* Group (Crustacea) Based on 12S rDNA Sequence and Allozyme Variation. *Molecular Phylogenetics and Evolution*, 5(3), 495-510.
- Thomas, G. W., Dohmen, E., Hughes, D. S., Murali, S. C., Poelchau, M., Glastad, K., Anstead, C. A., Ayoub, N. A., Batterham, P., Bellair, M., Binford, G. J., Chao, H., Chen, Y. H., Childers, C., Dinh, H., Doddapaneni, H. V., Duan J. J., Dugan, S.,... Richards, S. (2020). Gene content evolution in the arthropods. *Genome Biology*, 21(1), 1-14.
- Ueno, T., & Takahashi, Y. (2021). Mitochondrial polymorphism shapes intrapopulation behavioural variation in wild *Drosophila*. *Biology Letters*, 17(7), 20210194.
- Urban-Olivares, J. (2019). Genómica de poblaciones y divergencia ecológica: el caso del copépodo *Leptodiaptomus* de los Lagos de la Cuenca Oriental. Tesis de Licenciatura. Carrera de Biología. FES Iztacala. Universidad Nacional Autónoma de México.
- Van Straalen, N. M., & Roelofs, D. (2011). *An introduction to ecological genomics*. OUP Oxford.
- Veltri, D., Malapi-Wight, M. & Crouch, J. A. (2016). SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. *Nucleic Acids Research*, 44, 41-45.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Kalush, F. (2001). The sequence of the human genome. *Science*, 291(5507), 1304-1351.
- Waldvogel, A. M., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C., Kofler, R., Mock, T., Schmid, K., Schmitt, I., Bataillon, T., Savolainen, O., Bergland, A., Flatt, T., Guillaume, F., & Pfenninger, M. (2020). Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters*, 4(1), 4-18.
- Walter, T.C., Boxshall, G. (2021). World of Copepods database. Accessed at <http://www.marinespecies.org/copepoda> on 2021-01-22.
- Wetzel, R. G. (1975). Planktonic communities: Zooplankton and their interactions with fish. *Limnology Lake and River Ecosystems*, 395-488.
- Weydmann, A., Przyłucka, A., Lubośny, M., Walczyńska, K. S., Serrão, E. A., Pearson, G. A., & Burzyński, A. (2017). Mitochondrial genomes of the key zooplankton copepods Arctic *Calanus glacialis* and North Atlantic *Calanus finmarchicus* with the longest crustacean non-coding regions. *Scientific Reports*, 7(1), 1-11.

- Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, 13(6), e1005595.
- Wilson, A. C., Cann, R. L., Carr, S. M., George, M., Gyllensten, U. B., Helm-Bychowski, K. M., & Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, 26(4), 375-400.
- Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res*, 7, 1338.
- Wolstenholme, D. R. (1992). Animal mitochondrial DNA: structure and evolution. *International Review of Cytology*, 141, 173-216.
- Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20(1), 1-13.
- Wray, G. A. (2010). Integrating genomics into evolutionary theory. *Evolution: The extended synthesis*, 97-116.
- Wright, S. I., & Andolfatto, P. (2008). The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annual Review of Ecology, Evolution, and Systematics*, 39, 193-213.
- Xi, Z., Liu, L., & Davis, C. C. (2016). The impact of missing data on species tree estimation. *Molecular Biology and Evolution*, 33(3), 838-860.
- Xia, Y., Zheng, Y., Murphy, R. W., & Zeng, X. (2016). Intraspecific rearrangement of mitochondrial genome suggests the prevalence of the tandem duplication-random loss (TDLR) mechanism in *Quasipaa boulengeri*. *BMC Genomics*, 17(1), 1-9.
- Young, A. D., & Gillung, J. P. (2020). Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45(2), 225-247.
- Zhang, D. X., Szymura, J. M., & Hewitt, G. M. (1995). Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution*, 40(4), 382-391.
- Zhang, D. X., & Hewitt, G. M. (1997). Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*, 25(2), 99-120.

- Zhang, D., Zou, H., Zhang, J., Wang, G. T., & Jakovlić, I. (2020). Evolutionary history of inversions in the direction of architecture-driven mutational pressures in crustacean mitochondrial genomes. *bioRxiv*.
- Zhang, J., Chiodini, R., Badr, A., & Zhang, G. (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38(3), 95-109.
- Zhang, X., Xu, S., Liu, P., Huang, Q., Li, H., Dumont, H. J., & Han, B. P. (2020). The mitochondrial genome of *Phyllodiaptomus tunguidus* Shen & Tai, 1964 (Copepoda, Calanoida) and its phylogeny. *Mitochondrial DNA Part B*, 5(3), 2521-2522.
- Zhang, Z. Q. (2011). Animal biodiversity: an introduction to higher-level classification and taxonomic richness. *Zootaxa*, 3148(1), 7-12.

8. Anexo

8.1 Ensamble del genoma

El ensamble se llevó a cabo utilizando dos metodologías, por ensamble *de novo* y por ensamble con referencia. Para el genoma con referencia se realizó el mapeo de las lecturas de *L. garciai* contra genomas de copépodos mediante BWA (Li & Durbin, 2010) y Minimap2 (Li, 2018), llevando a cabo una manipulación post-alineamiento mediante la herramienta SAMtools (Danecek, 2021) para conocer el porcentaje de alineamiento. Se corrieron distintos mapeos utilizando las siguientes combinaciones de genomas disponibles:

- 1) *Eurytemora affinis*
- 2) *E. affinis* + *Acartia tonsa*
- 3) *Leptodiptomus* Atexcac
- 4) *L. Atexcac* + *L. El Carmen*
- 5) 11 genomas de copépodos disponibles en el NCBI (*Tigriopus kingsejongensis* (GCA_012959195.1), *Tisbe holothuriae* (GCA_900659605.1), *Apocyclops royi* (GCA_900607525.1), *Acartia tonsa* (GCA_900241095.1), *Calanus glacialis* (GCA_002740985.1), *Oithona nana* (GCA_900157175.1), *Eurytemora affinis* (GCA_000591075.2), *Tigriopus californicus* (GCA_007210705.1), *Calanus finmarchicus* (GCA_002740975.1), *Tigriopus japonicus* (GCA_010645155.1) y *Lepeophtheirus salmonis* (GCA_016086655.2)) y borradores no publicados de *Leptodiptomus* Atexcac (facilitado por Javier Urban) y *Leptodiptomus* El Carmen (facilitado por Arturo Alcántara).

Para el ensamble *de novo* se probaron los ensambladores SPAdes, SOAP *de novo*, Platanus, Megahit, ABYSS y Velvet. Para conocer las estadísticas básicas y N50 de cada ensamble se utilizó QUAST (Gurevich et al. 2013). Del mismo modo, se utilizó BUSCO (Simão *et al.*, 2015) para estimar el nivel de integridad del genoma. Esta herramienta realiza una comparación de grupos de genes ortólogos dentro de una base de datos predeterminada, para así conocer la cantidad de grupos genéticos que se tiene en un ensamble. El enmascaramiento del genoma se realizó por medio de RepeatModeler (Smit & Hubley, 2015) y RepeatMasker (Smit & Hubley, 2010). Finalmente, se utilizó AUGUSTUS (Stanke & Morgenstern, 2005) para realizar una anotación ab initio.

8.1.1 Ensamble con referencia

En la Tabla 2 se muestra el porcentaje de mapeo de las secuencias de *L. garciai* en los genomas de referencia con el set de datos de Trimmomatic. Los valores más elevados se obtuvieron utilizando un archivo concatenado de 13 genomas de copépodos

Tabla 2. Mapeo de lecturas de *L. garciai* contra genomas de referencia mediante BWA y Minimap 2.

Genoma de referencia	Porcentaje de mapeo
BWA (Burrows-Wheeler Aligner)	
<i>Eurytemora affinis</i>	23.93%
<i>E. affinis</i> + <i>A. tonsa</i>	29.95%
<i>Leptodiaptomus</i> Atexcac	23.93%
<i>L. Atexcac</i> + <i>L. El Carmen</i>	24.12%
11 genomas + <i>L. Atexcac</i> + <i>L. El Carmen</i>	37.34%
Minimap 2	
13 Genomas	36.57%

Se observa que las secuencias de *L. garciai* tienen baja representación en los genomas previamente publicados. Utilizando el archivo con el porcentaje más elevado de mapeo (37.34%) se realizaron distintas pruebas con diferentes ensambladores, sin embargo, SPAdes fue el único ensamblador que pudo completar el armado del genoma de *L. garciai*. En la tabla 3 se presentan las estadísticas de Quast y BUSCO de los ensamblados, siendo el ensamblado con mejor calidad (N50) e integridad (BUSCO) el llevado a cabo con un Kmero de 77.

Tabla 3. Ensamblados de referencia completados con SPAdes usando el set de datos de Trimmomatic.

kmero	N50	scaffolds	BUSCO
63	729 pb	38,223	--
71	710 pb	33,734	0.49%
77	694 pb	30,880	0.69%
81	683 pb	28,099	0.49%
91	659 pb	25,678	--

Para el set de datos de AdapterRemoval no fue posible completar ningún ensamble debido a la limitación de recursos computacionales, por lo cual no se muestra ningún resultado.

8.1.2 Ensamble *de novo*

En la Tabla 4 se presentan las características de los ensambles *de novo* con los estadísticos más altos de calidad e integridad. Para los datos provenientes de Trimmomatic, SOAP de novo ha sido el ensamblador que ha dado los mejores resultados. Algunos programas no pudieron completar el ensamble debido a la limitación de recursos computacionales. Para solucionar esto, en algunos casos fue posible utilizar el software Rasusa (Hall, 2019) el cual, a partir de las lecturas limpias, realiza un muestreo aleatorio con la finalidad de obtener un par de archivos con una cobertura más baja, *i. e.*, dos archivos con una menor cantidad de secuencias.

Tabla 4. Ensamblados *de novo* completados con set de datos de Trimmomatic.

Ensamblador	kmero	N50	scaffolds	BUSCO
SOAP de novo	61	281 pb	195,854	0.78%
SPAdes (Rasusa 20 X)	77	826 pb	67,796	0.49%
SPAdes (Rasusa 30 X)	77	991 pb	125,323	0.49%
Platanus	77	3,350 pb	27,548	0
Megahit	127	530 pb	2,551,315 contigs*	--
ABYSS	91	651 pb	2,792,055	--

En la tabla 5 se presentan los ensambles completados mediante el set de datos de AdapterRemoval, observándose una mejora en cuanto al N50 y al número de scaffolds respecto a los datos provenientes de Trimmomatic. SPAdes fue el único programa que completó el ensamblado utilizando estos datos, siendo el ensamble de SPAdes con kmero de 77 el que tiene una mayor calidad e integridad.

Tabla 5. Ensamblados de novo completados con set de datos de AdapterRemoval.

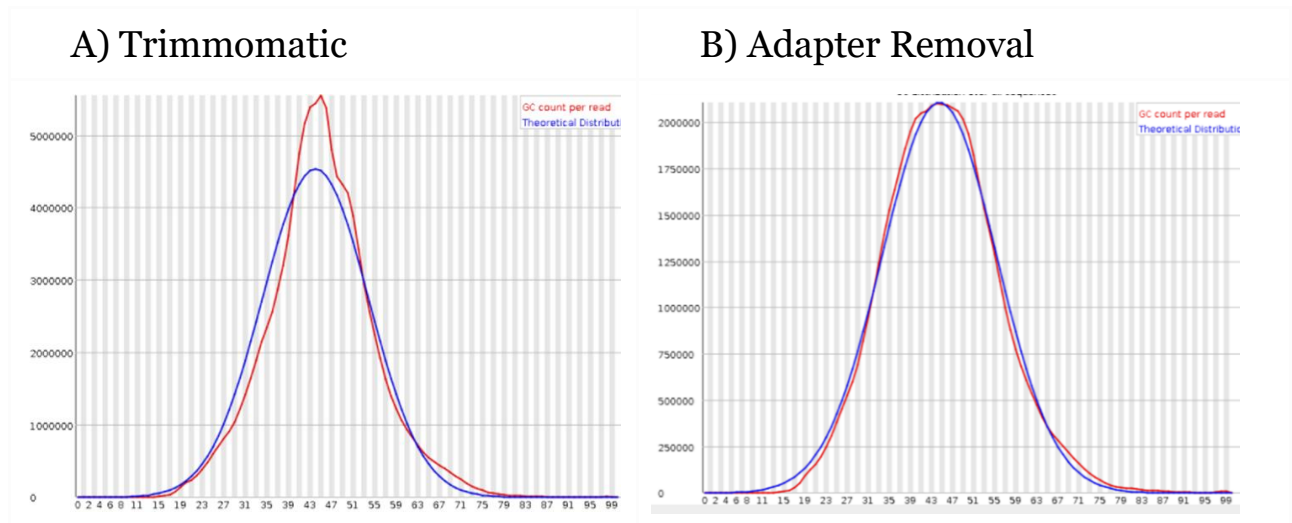
Ensamblador	kmero	N50	scaffolds	BUSCO
SPAdes	77	927 pb	153,806	0.59%
SPAdes	91	828 pb	151,752	0.29%
SPADES FastX collapser	77	921 pb	109.594	0.39%

Los ensamblados obtenidos hasta el momento presentan un alto grado de fragmentación que puede verse reflejado en el número de scaffolds. Una de las causas principales de esta fragmentación puede ser el número elevado de regiones de baja complejidad (secuencias repetidas) dentro del genoma de *L. garciai*. Se intentaron dos aproximaciones para tratar de resolver este problema, utilizando Rasusa (tabla 4) y FastX collapser el cual elimina secuencias repetidas en cada read. Aunque se puede observar una mejoría en el número de scaffolds, el valor de N50 no incrementa de manera sustancial.

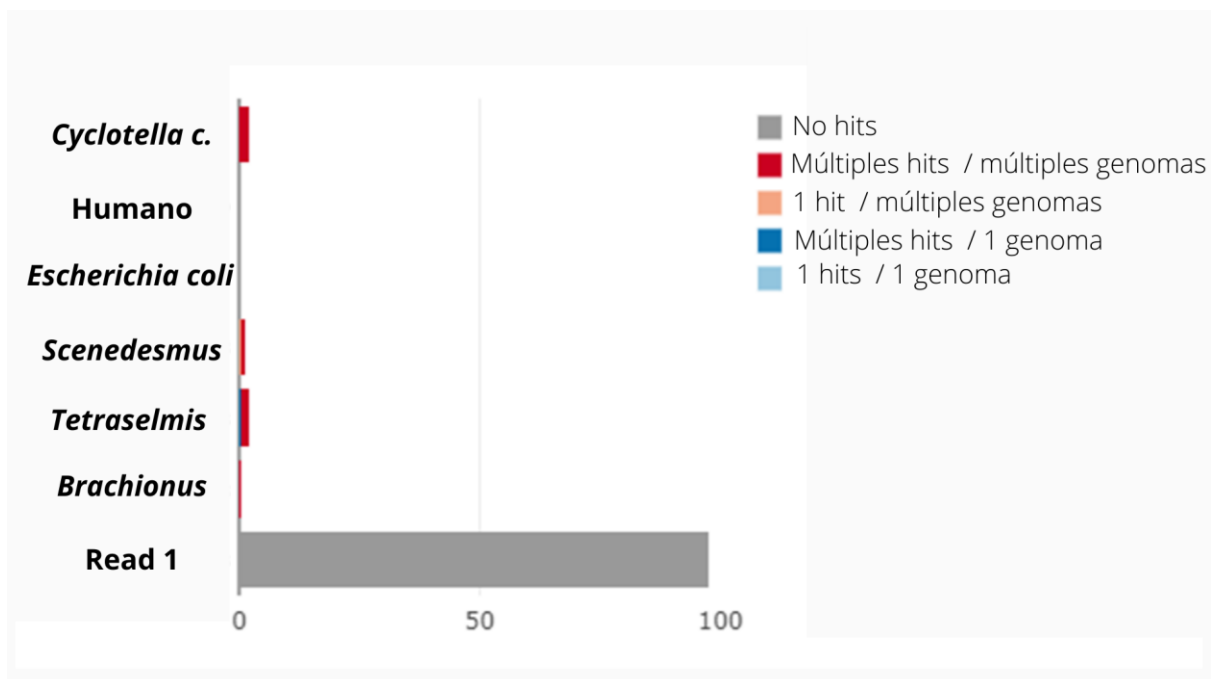
Se utilizó el ensamblado de Spades (kmero 77, %GC de 42.31) proveniente de los datos de AdapterRemoval (Tabla 5) para realizar el enmascaramiento del genoma. RepeatMasker enmascaró un 48.32% del ensamblado como elementos no conocidos dentro de su base de datos y un total de 1.33% como elementos repetidos simples, sin identificar elementos transponibles (SINEs, LINEs o LTRs). La predicción *ab initio* de AUGUSTUS identificó un total de 35,344 genes.

Los ensamblados de referencia y de novo realizados hasta el momento presentan una calidad e integridad baja, además de una elevada cantidad de genes anotados en comparación con otras especies de copépodos (~20,000). Aunque las secuencias presentan un grado de contaminación bajo, así como una buena calidad en general, estos resultados hacen evidente la necesidad de un esfuerzo mayor de secuenciación para obtener un genoma robusto de *L. garciai*.

8.2 Ensamble del mitogenoma



Anexo A. Distribución de GC en las secuencias después de la fase de limpieza. La curva azul muestra la distribución teórica y la curva roja el conteo real por lectura.

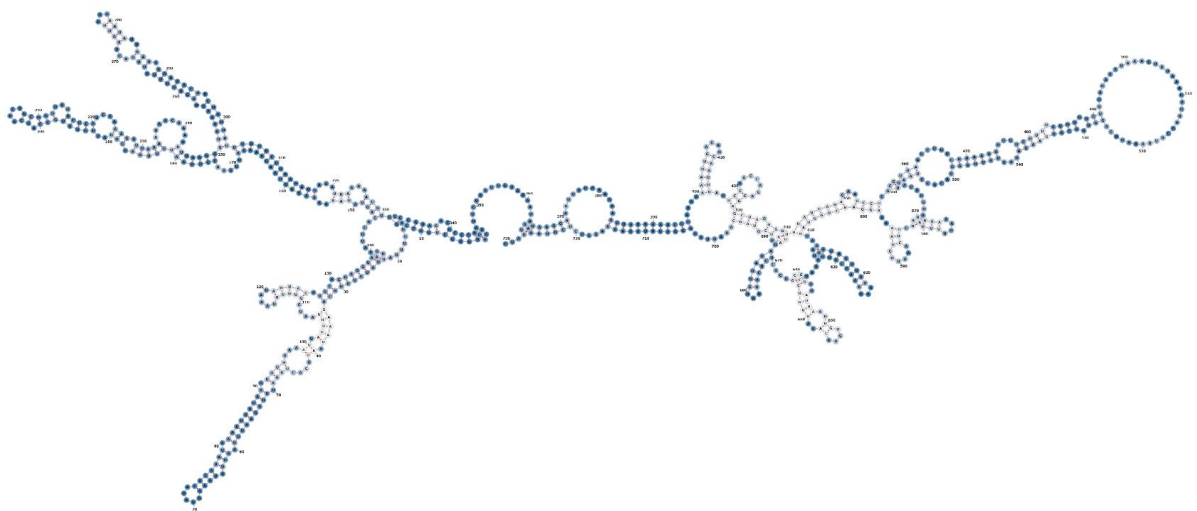


Anexo B. Porcentaje de mapeo de genomas contaminantes contra las secuencias genómicas provenientes de la secuenciación de *Leptodiptomus garciai* a través de

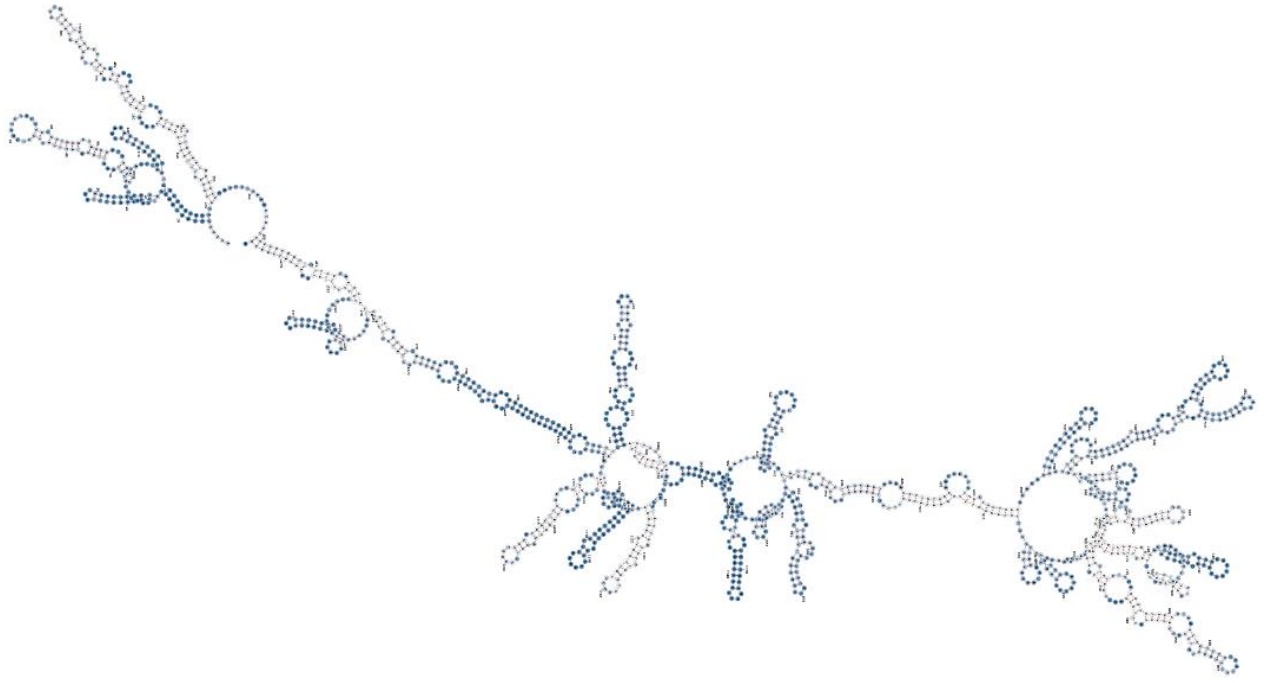
FastQ screen. La barra gris representa secuencias no pertenecientes a especies contaminantes.

Anexo C. Grupos identificados mediante el análisis de Kraken 2.

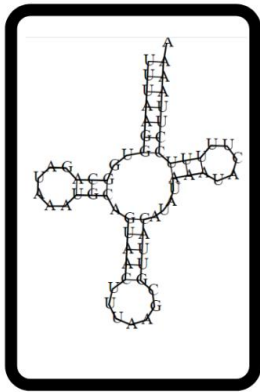
Grupo	Porcentaje de contamiación	Grupo	Porcentaje de contamiación
Enterobacterales	0.04%	Bacteroidetes/Chlorobi	0.03%
Enterobacteriaceae	0.02%	Bacteroidetes	0.03%
Erwiniaceae	0.01%	Flavobacteriia	0.02%
Buchnera	0.01%	Flavobacteriales	0.02%
Buchnera aphidicola	0.01%	Cyanobacteria	0.01%
Mixta	0.01%	Tenericutes	0.01%
Mixta theicola	0.01%	Mollicutes	0.01%
Yersiniaceae	0.01%	Mycoplasmatales	0.01%
Yersinia	0.01%	Betaproteobacteria	0.06%
Pseudomonadales	0.03%	Burkholderiales	0.06%
Moraxellaceae	0.02%	Burkholderiaceae	0.03%
Acinetobacter	0.02%	Burkholderia	0.02%
Acinetobacter shaoyimingii	0.02%	Alphaproteobacteria	0.02%
Pseudomonadaceae	0.01%	Rhodobacterales	0.01%
Pseudomonas	0.01%	Terrabacteria group	0.29%
Xanthomonadales	0.01%	Firmicutes	0.19%
Xanthomonadaceae	0.01%	Bacilli	0.17%
Oceanospirillales	0.01%	Bacillales	0.16%
Halomonadaceae	0.01%	Staphylococcaceae	0.11%
Archaea	0.01%	Straphylococcus aureus	0.10%
Euryarchaeota	0.01%		



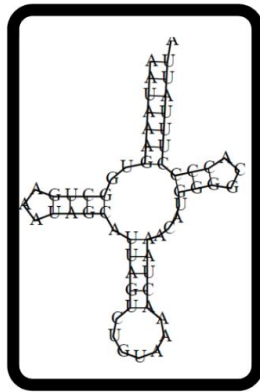
Anexo D. RNA ribosomal subunidad s (s-rRNA)



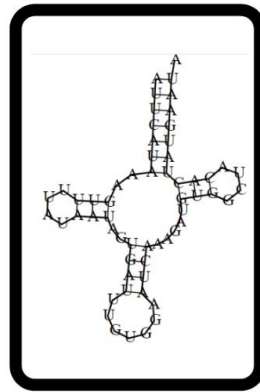
Anexo E. RNA ribosomal subunidad I (1-rRNA)



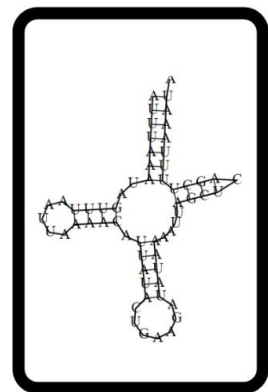
trnL2 (tta)



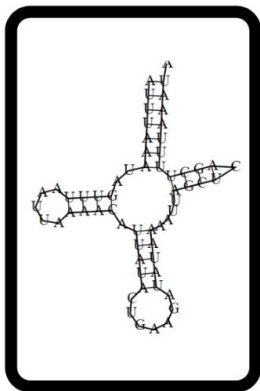
trnY (tac)



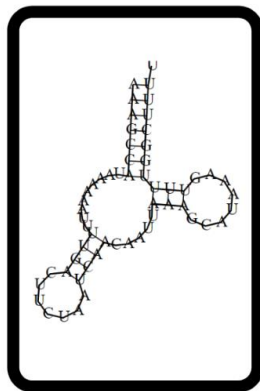
trnH (cac)



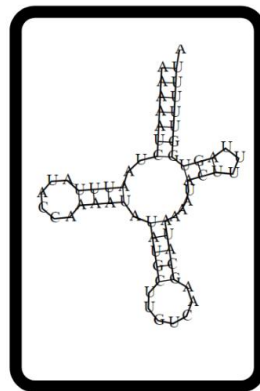
trnF (ttc)



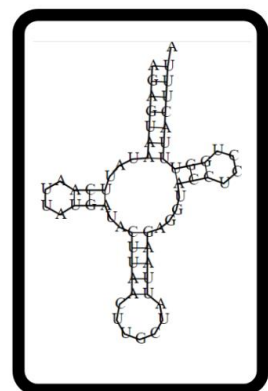
trnG (gga)



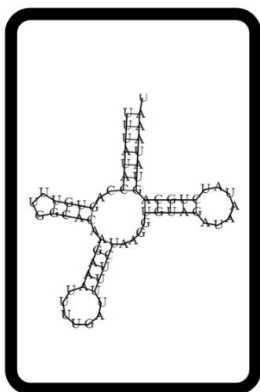
trnS1 (aga)



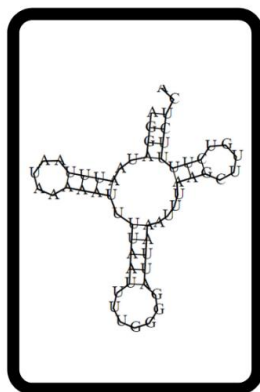
trnD (gac)



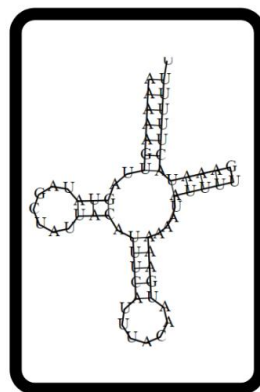
trnA (gca)



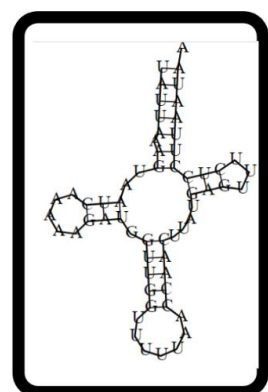
trnQ (caa)



trnP (cca)

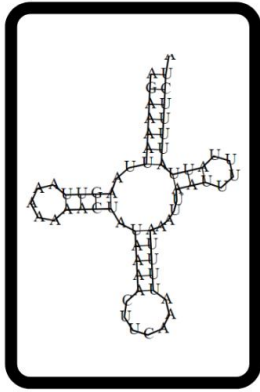


trnV (gta)

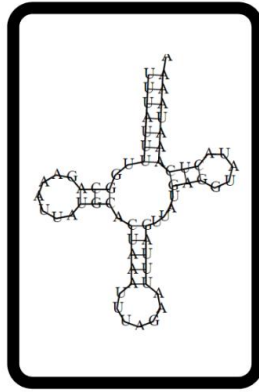


trnK (aaa)

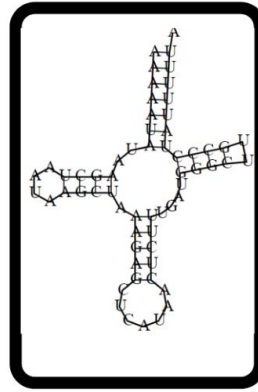
Anexo F. Estructura secundaria de RNAt obtenida a partir de MWS.



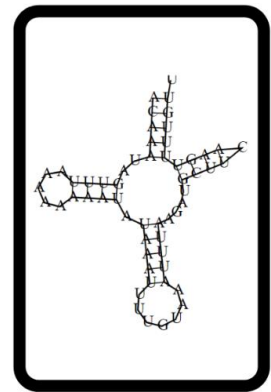
trnW (tga)



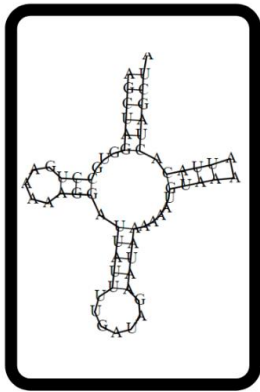
trnL1 (cta)



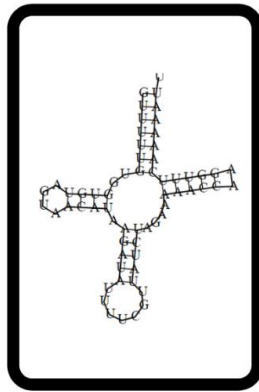
trnM (atg)



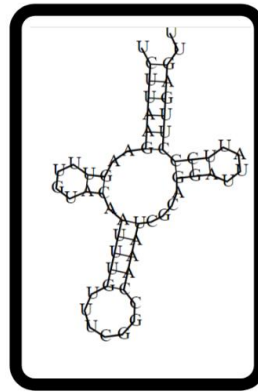
trnT (aca)



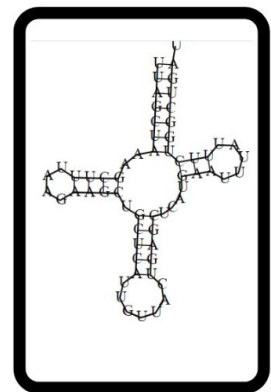
trnI (atc)



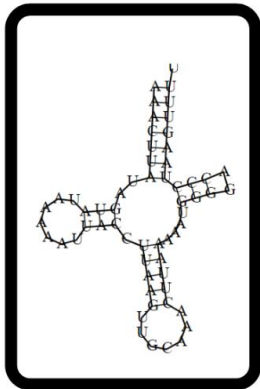
trnE (gaa)



trnR (cga)



trnN (aac)

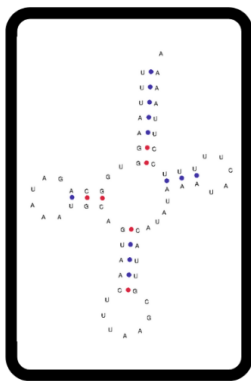


trnC (tgc)

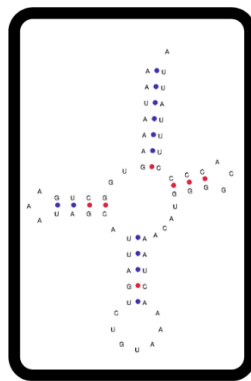
Anexo G. Estructura secundaria de RNAt obtenida a partir de MWS.

Anexo H. RNAt encontrados utilizando tRNAscan-SE.

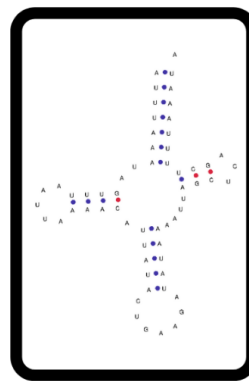
RNA t	Inicio de la secuencia	Fin de la secuencia	Longitud	Tipo de RNAt	Anticodón
1	939	1,000	61	Leu	TAA
2	2,680	2,738	58	Tyr	GTA
3	4,635	4,695	60	Phe	GAA
4	4,696	4,758	62	Gly	TCC
5	5,282	5,343	61	Asp	GTC
6	6,372	6,432	60	Ala	TGC
7	14,276	14,217	59	Cys	GCA
8	13,893	13,832	61	Asn	GTT
9	13,763	13,701	62	Gllu	TTC
10	12,012	11,952	60	Ile	GAT
11	11,875	11,815	61	Met	CAT
12	11,814	11,752	62	Leu	TAG
13	9,904	9,843	61	Val	TAC



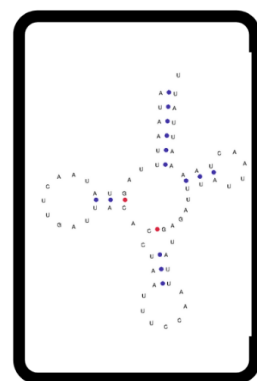
taa (Leu)



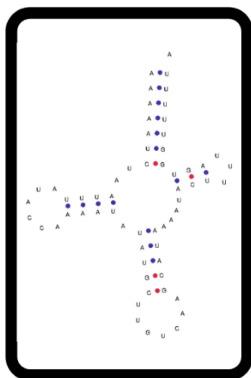
gra (Tyr)



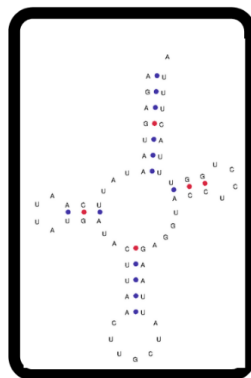
gaa (Phe)



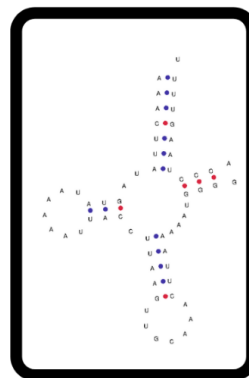
tcc (Gly)



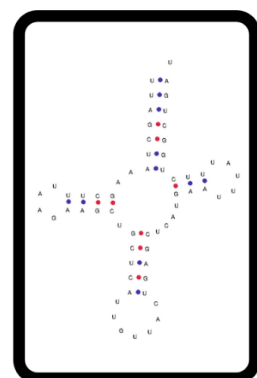
gtc (Asp)



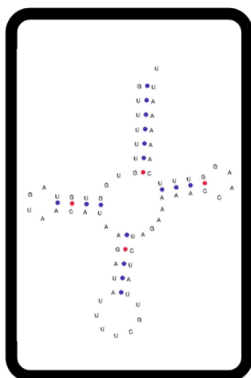
tgc (Ala)



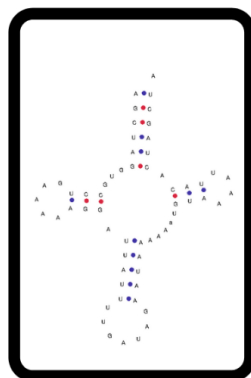
gca (Cys)



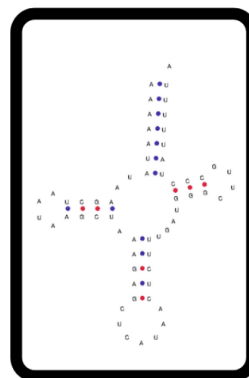
gtt (Asn)



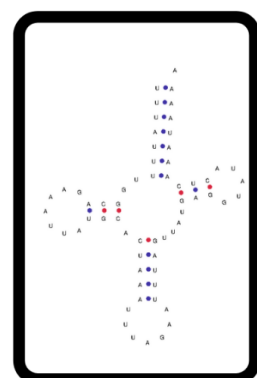
ttc (glu)



gat (Ile)

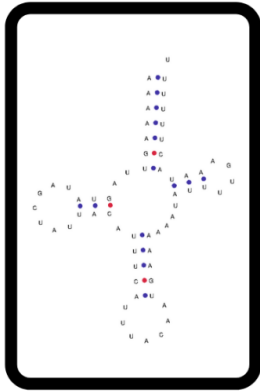


cat (Met)



tag (Leu)

Anexo I. Estructura secundaria de RNAt obtenida a partir de tRNAscan-SE



tac (Val)

Anexo J. Estructura secundaria de RNAt obtenida a partir de tRNAscan-SE

Anexo K. Matriz de distancia de distancias-p. A) *L. garciai*, B) *P. tunguidus*, C) *C. hyperboreus*, D) *C. glacialis*, E) *A. atopus*, F) *S. polycolpus*, G) *S. undulatus*, H) *P. nana*, I) *L. cyprinacea*, J) *T. californicus*, K) *T. kingsejongensis*, L) *P. rhincodonicus*.

	A	B	C	D	E	F	G	H	I	J	K	L
A	-	-	-	-	-	-	-	-	-	-	-	-
B	0.304	-	-	-	-	-	-	-	-	-	-	-
C	0.364	0.301	-	-	-	-	-	-	-	-	-	-
D	0.360	0.290	0.145	-	-	-	-	-	-	-	-	-
E	0.453	0.435	0.437	0.428	0	-	-	-	-	-	-	-
F	0.463	0.454	0.460	0.452	0.460	-	-	-	-	-	-	-
G	0.466	0.432	0.442	0.432	0.441	0.194	-	-	-	-	-	-
H	0.476	0.457	0.440	0.440	0.476	0.466	0.422	-	-	-	-	-
I	0.510	0.494	0.494	0.483	0.507	0.498	0.485	0.493	-	-	-	-
J	0.506	0.489	0.484	0.479	0.486	0.515	0.496	0.503	0.531	-	-	-
K	0.489	0.469	0.479	0.468	0.466	0.507	0.493	0.493	0.515	0.454	-	-
L	0.517	0.497	0.491	0.489	0.488	0.500	0.485	0.498	0.516	0.509	0.499	-

Anexo L. Tasas de sustitución Ka/Ks entre *L. garciai* y once especies de copépodos. A) Calanoida B) Cyclopoida C) Harpacticoida D) Siphonostomatoida. Los valores resaltados representan los genes que se encuentran bajo un escenario de selección putativa.

Orden	Especie	COX1	COX2	COX3	NAD1	NAD2	NAD3	NAD4	NAD5	NAD6	CYTB	ATP6
A	<i>P. tunguidus</i>	0.022	0.684	0.040	0.149	NA	1.691	NA	0.250	NA	NA	0.176
	<i>C. hyperboreus</i>	0.022	0.077	0.080	0.249	0.482	NA	NA	1.088	NA	0.749	0.282
	<i>C. glacialis</i>	na	0.077	NA	na	na	0.258	NA	0.203	0.399	NA	0.246
B	<i>S. undulatus</i>	0.095	na	na	1.704	1.406	2.816	0.482	0.517	NA	0.130	2.200
	<i>S. polycolpus</i>	0.092	1.108	0.524	2.115	na	na	0.486	0.444	1.312	0.175	2.980
	<i>P. nana</i>	0.050	0.637	0.094	0.836	2.434	na	0.481	1.338	NA	1.064	0.557
	<i>L. cyprinacea</i>	0.049	0.394	0.151	0.267	0.813	1.332	0.415	1.061	1.034	0.206	2.570
C	<i>A. atopus</i>	0.078	na	na	0.153	0.501	2.194	1.627	0.758	NA	1.096	2.083
	<i>T. californicus</i>	na	0.222	na	0.332	0.701	0.361	0.292	0.317	0.577	NA	0.593
	<i>T. kingsejongensis</i>	na	na	na	na	0.544	0.346	NA	na	0.228	NA	0.591
D	<i>P. rhincodonicus</i>	0.115	0.275	0.171	na	0.538	0.531	NA	0.421	0.763	NA	0.676

Anexo M. Matriz de reordenamiento de CREx. Los valores representan la cantidad de intervalos comunes entre genomas, valores altos indican un orden genético similar entre genomas., A) *L. garciai*, B) *P. tunguidus*, C) *C. hyperboreus*, D) *A. atopus*, E) *S. undulatus*, F) *P. nana*, G) *L. cyprinacea*, H) *T. californicus*, I) *T. kingsejongensis*, J) *P. rhincodonicus*

	A	B	C	D	E	F	G	H	I	J
A	150	-	-	-	-	-	-	-	-	-
B	94	150	-	-	-	-	-	-	-	-
C	16	12	150	-	-	-	-	-	-	-
D	12	6	22	150	-	-	-	-	-	-
E	8	6	22	80	150	-	-	-	-	-
F	0	6	16	4	4	150	-	-	-	-
G	8	6	8	16	16	0	150	-	-	-
H	8	10	18	20	20	20	6	150	-	-
I	16	10	20	78	48	4	12	12	150	-
J	0	2	4	10	20	6	4	10	10	150