

La investigación reportada en esta tesis es parte de los programas de investigación del CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California).

La investigación fue financiada por el CONAHCYT (Consejo Nacional de Humanidades Ciencias y Tecnología).

Todo el material contenido en esta tesis está 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 o titular de los Derechos de Autor.

CICESE@ 2023. Todos los derechos reservados

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



**Maestría en Ciencias
en Acuicultura**

***Análisis del genoma mitocondrial y contigs obtenidos
mediante secuenciación de segunda generación de los robalos
Centropomus viridis y Centropomus nigrescens***

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

Presenta:

Alejandra del Carmen Castillo Collado

Ensenada, Baja California, México
2023

Tesis defendida por
Alejandra del Carmen Castillo Collado

y aprobada por el siguiente Comité

Dr. Miguel Ángel del Río Portilla
Director de tesis

Dra. Carolina Álvarez Delgado

Dra. Fabiola Lafarga de la Cruz



Dr. Roberto Cruz Flores
Coordinador del Posgrado en Acuicultura

Dra. Ana Denise Re Araujo
Directora de Estudios de Posgrado

Resumen de la tesis que presenta **Alejandra del Carmen Castillo Collado** como requisito parcial para la obtención del grado de Maestro en Ciencias en Acuicultura.

Análisis del genoma mitocondrial y contigs obtenidos mediante secuenciación de segunda generación de los robalos *Centropomus viridis* y *Centropomus nigrescens*

Resumen aprobado por:

Dr. Miguel Ángel del Río Portilla
Director de tesis

En México, la acuicultura es una actividad muy importante debido a que contribuye a aportar una fuente de alimento rica en proteínas, vitaminas y lípidos esenciales, asimismo, ayuda a la economía del país. De la gran diversidad de peces en el país, los robalos *Centropomus viridis* y *Centropomus nigrescens* son las especies que presentan más popularidad en mercados locales. El objetivo del presente estudio fue obtener y analizar los genomas mitocondriales (mitogenomas) de las especies *Centropomus viridis* y *Centropomus nigrescens* y compararlos con otros mitogenomas de especies similares. Para conseguir los mitogenomas de estos organismos se extrajo el ADN total del músculo y de la aleta de dos peces obtenidos por donaciones siguiendo los protocolos del Laboratorio de Genética Acuícola del Departamento de Acuicultura del CICESE. Las muestras de mejor calidad se enviaron a secuenciar mediante el método de segunda generación con la plataforma Illumina. Las lecturas obtenidas de la secuenciación se evaluaron utilizando las regiones de mejor calidad, con estas lecturas se realizó un ensamblaje *de Novo*, un mapeo y extracción de contigs, y un análisis de Blastn y Blastx para identificar regiones de interés en el resto de las lecturas. Los mitogenomas completos obtenidos de *C. viridis* y *C. nigrescens* son similares a los genomas mitocondriales reportados en el GenBank que corresponden a especies del orden Carangaria *incertae sedis*. Con los análisis filogenéticos utilizando los genomas mitocondriales completos y la región 18S ARNr, se logró comprender mejor las relaciones filogenéticas de ambas especies, además se identificaron genes de interés e información de microsatélites lo que, se espera, facilitará investigaciones futuras en estos robalos. Se espera que la información generada en este trabajo pueda ayudar a seleccionar reproductores y contribuir a preservar el acervo genético.

Palabras clave: Genética, ADN, Secuenciación masiva, Filogenia, Ensamblaje *de Novo*.

Abstract of the thesis presented by **Alejandra del Carmen Castillo Collado** as a partial requirement to obtain the Master of Science degree in Aquaculture

Analysis of the mitochondrial genome and contigs obtained by second generation sequencing of snooks *Centropomus viridis* and *Centropomus nigrescens*

Abstract approved by:

Ph. D. Miguel Ángel del Río Portilla
Thesis Director

In Mexico, aquaculture is an important activity because it contributes to provide a food source rich in protein, vitamins and essential lipids, and it also helps the country's economy. There are various species of fish that are consumed throughout the country, but the snooks *Centropomus viridis* and *Centropomus nigrescens* are very popular species in local markets. The aim of this study was to obtain, analyze and compare the mitochondrial genomes (mitogenomes) of the *Centropomus viridis* and *Centropomus nigrescens* species with the mitogenomes of closely related species. To obtain the mitogenomes of these organisms, total DNA was extracted from the muscle and fin of two fish obtained by donations and following the protocols of the Aquaculture Genetics Laboratory of the CICESE Department of Aquaculture. The best quality samples were sent for sequencing using the second-generation method with the Illumina platform. The obtained reads from the sequencing were evaluated using the best quality regions, with these reads, a *de Novo* assembly. The complete mitogenomes obtained from *C. viridis* and *C. nigrescens* are similar to the mitochondrial genomes reported stored in the GenBank that correspond to species of the order Carangaria *incertae sedis*. With the phylogenetic analyzes using the complete mitochondrial genomes and the 18S rRNA region, it was possible to better understand the phylogenetic relationships of both species, in addition, genes of interest and microsatellite information were identified, which, it is considered, will facilitate further investigations. It is expected that the information generated in this work can help select broodstock and contribute to preserving gene stocks.

Keywords: Genetics, DNA, gun-shot Sequencing, Phylogeny, *De novo* assembly

Dedicatoria

A mi Padre por nunca dejarme sola en mi sueño de dedicarme a la ciencia.

A mis hermanos Michelle, Juan, Elizabeth, Teresa y Adolfo por ser lo máspreciado que tengo en la vida, los amo.

Agradecimientos

A Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por otorgarme la beca (1097885) para realizar mis estudios de posgrado. Al Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California (CICESE) por aceptarme como parte de su matrícula de estudiantes.

Al Departamento de Acuicultura, por brindarme la oportunidad de ampliar mis conocimientos en un área nueva y a cada uno de mis maestros que con paciencia y amabilidad me explicaron cada tema y me impulsaron a seguir adelante. A todo el personal administrativo de Acuicultura gracias por su amabilidad.

Al Dr. Miguel Ángel del Río Portilla por brindarme el honor de ser su estudiante y confiar en mí en todo momento, gracias por compartirme todo su conocimiento con mucha paciencia y por su apoyo incondicional. Gracias por el trato amable y respetuoso, por enseñarme a ser una mejor estudiante, pero sobre todo una mejor persona.

A la Dra. Fabiola Lafarga de la Cruz gracias por formar parte de mi comité por sus consejos, por sus observaciones en mi investigación y por las clases para ampliar mis conocimientos, sin duda disfruté mucho de ellas.

A la Dra. Carolina Álvarez Delgado gracias por formar parte de mi comité y por la buena retroalimentación en cada avance y por esa fascinante clase sobre mitocondrias.

Al Dr. Roberto Cruz Flores, por aconsejarme en muchas ocasiones y por apoyarnos a todos los estudiantes de acuicultura ante cualquier situación.

Al Dr. Antonio López Carvallo por resolver alguna de mis dudas y enseñarme algunos trucos bioinformáticos.

A la M. C. Carmen Elvira Vargas Peralta por los consejos para mejorar mis técnicas en el laboratorio de genética acuícola.

A la Dra. Edith Elizondo Reyna por las pláticas, los consejos y su amabilidad.

A la M. en Ed. Laura Rebeca Pineda Meléndez, muchas gracias por cuidar de cada estudiante foráneo y hacer más cómoda nuestra estancia en Ensenada. Gracias por la amistad, los abrazos, pláticas y atenciones.

A la M. en Ed Nadua Yamile Dajlala Spingler y a todo el personal de biblioteca por las reuniones de revisión de formato de tesis.

Al grupo de difusión científica “Ciencia y Serendipia” por ser unas grandes compañeras, inspirarme a ser mejor, por esas reuniones y salidas a divulgar la ciencia.

A mi familia por hacerme sentir una persona querida y apoyada siempre. A mi Padre el Dr. Rolando Castillo Santiago por apoyarme. A mis tías Karen Nadiut y Yeni Jazmín por ser como unas madres para mí, dándome consejos y apoyándome a la distancia.

A mi abuelito Arnulfo, mi tío José de Jesús y mi tío Carlos por el apoyo y las llamadas por las tardes para platicar. A mi tía Juana, mi tío Memo, mi tía Joana y a mi madrina Pili por acompañarme a distancia y quererme como una hija más, los quiero mucho.

A la Dra. Emma Estela Hernández Domínguez, tía Emmi muchas gracias por todo tu apoyo y tus consejos para mi vida personal y profesional.

A los primeros amigos que hice en la maestría David Mireles, Héctor Santos, Guillermo Chin, Naoki Kawamoto, Augusto Alvarado, Narda Pérez y Manuel Salud por las pláticas profundas, los consejos, las risas, las salidas juntos. Gracias por cuidar de mí y hacer linda mi estancia en Ensenada.

A la familia Barrera-Martínez por las tardes de juegos de mesa, su apoyo y amabilidad por abrirme las puertas de su casa y recibirme siempre con una sonrisa, apoyándome cuando me sentía sola.

A mis amigas Mriya López y Areli Castro gracias por esa amistad tan linda por cuidar de mí y abrirme las puertas de sus casas, por las salidas y los días de picnic. A mis amigos Karlita, Ámbar, Karen y Erick gracias por tanto los quiero mucho.

Al gatito prestado, Morrison “peluso” por acompañarme y hacerme sentir feliz, aunque extrañaba a mi familia.

A la señora Lolita García gracias por las tardes de pláticas, por permitirme degustar comida riquísima en su casa y por brindarme un trato hospitalario siempre.

Al Dr. Luis Daniel Jiménez por impulsarme a iniciar mi camino en la ciencia.

A todas las personas que conocí y me brindaron su amabilidad, me cuidaron e hicieron muy linda mi estancia en Ensenada, muchas gracias. Escoger CICESE como centro de investigación para mi formación y estudiar el Posgrado en Acuicultura ha sido de las mejores decisiones que he tomado.

Tabla de contenido

	Página
Resumen en español.....	ii
Resumen en inglés.....	iii
Dedicatoria	iv
Agradecimientos.....	v
Lista de figuras.....	xi
Lista de Tablas.....	xii
Capítulo 1. Introducción.....	1
1.1 Antecedentes	3
1.1.1 <i>La familia Centropomidae</i>	3
1.1.2 <i>Biología de Centropomus viridis y Centropomus nigrescens</i>	3
1.1.3 <i>Importancia del robalo en México</i>	5
1.2 Importancia de trabajos genéticos para la acuicultura.....	6
1.2.1. Trabajos genéticos en Centropomidae y especies de familias afines	7
1.3 Secuenciación masiva.....	9
1.3.1 <i>Mitogenomas de vertebrados mediante secuenciación masiva</i>	9
1.4 Microsatélites.....	10
1.5 Justificación	12
1.6 Hipótesis.....	13
1.7 Objetivos	13
1.7.1 <i>Objetivo general</i>	13
1.7.2 <i>Objetivos específicos</i>	13
Capítulo 2. Metodología.....	14
2.1 Origen del material biológico y extracción de ADN	14
2.1.1 <i>Extracción de ADN</i>	14
2.1.2 <i>Evaluación de la calidad y cantidad del ADN</i>	14

2.2	Protocolo de secuenciación.	15
2.3	Calidad de las lecturas y ensamblaje <i>de Novo</i>	15
2.3.1	<i>Corroboración de la calidad y eliminación de regiones de baja calidad</i>	16
2.3.2	<i>Ensamblaje de Novo usando TRINITY</i>	16
2.3.3	<i>Calidad de las lecturas y ensamblaje de Novo utilizando CLC</i>	16
2.3.3.1	Sequencing QC Report	16
2.3.3.2	Ensamblaje <i>de Novo</i>	17
2.3.4	<i>Ensamblaje de Novo utilizando NOVOPlasty</i>	17
2.4	Mapeo y extracción de “contigs”	17
2.4.1	<i>Mapeo y extracción de contigs con el servidor Omica</i>	17
2.4.2	<i>Mapeo y extracción de contigs con el servidor Omica</i>	18
2.4.3	<i>Mapeo y extracción de contigs con CLC Genomics Workbench</i>	18
2.5	Anotación del mitogenoma.....	18
2.5.1	<i>Registro de los genomas mitocondriales en el GenBank</i>	19
2.6	Análisis filogenético.....	19
2.6.1	<i>Análisis con clustal Omega</i>	19
2.6.2	<i>Selección del modelo mediante Jmodel</i>	19
2.6.3	<i>Árbol filogenético mediante inferencia Bayesiana</i>	20
2.7	Blastn y Blastx.....	20
2.8	Ontología genética	21
2.9	Evaluación del número de microsatélites	21
2.9.1	<i>Obtención y evaluación de los microsatélites</i>	21
2.9.2	<i>Búsqueda de “primers” con CLC Genomics Workbench</i>	22
Capítulo 3.	Resultados	23
3.1	Origen del material biológico y extracción de ADN	23
3.1.1	<i>Extracción, calidad y cuantificación del ADN</i>	23
3.2	Verificación de la calidad de las lecturas y ensamblaje <i>de Novo</i>	24
3.2.1	<i>Corroboración de la calidad y eliminación de regiones de baja calidad</i>	25
3.2.1.1	Sequencing QC Report	25

	x
3.2.1.2 Ensamblaje <i>de Novo</i> CLC Genomics Workbench.....	27
3.2.2 <i>Ensamblaje de Novo usando NOVOPlasty</i>	27
3.3 Mapeo y extracción de “contigs”	28
3.4 Anotación del genoma mitocondrial.....	28
3.4.1 <i>Registro de los genomas mitocondriales en GenBank</i>	31
3.6 Análisis filogenético	31
3.5 Blastn y Blastx.....	35
3.6 Ontología genética (GO).....	35
3.6.1 <i>Ontología genética (GO) para Centropomus viridis y Centropomus nigrescens</i>	35
3.7 Evaluación del número de microsatélites	37
Capítulo 4. Discusión.....	39
4.1 Extracción y calidad de las muestras de ADN	39
4.2 Análisis bioinformático.....	39
4.2.1 <i>Calidad de las secuencias y ensamblaje de Novo</i>	39
4.3 Anotación y registro del genoma mitocondrial.....	41
4.4 Análisis filogenéticos	42
4.5 Ontología genética	43
4.6 Microsatélites.....	44
Capítulo 5. Conclusiones	45
Literatura citada	46
Anexos.....	53

Lista de figuras

Figura	Página
1. <i>Centropomus viridis</i> ejemplar utilizado en el estudio de la revisión de caracteres morfológicos de robalos (Martínez-Brown <i>et al.</i> , 2021). (Foto Rodrigo Martínez Moreno).	1
2. Robalo negro <i>Centropomus nigrescens</i> [fotografía], por Chris Wheaton, 2019, Mexican fish (Róbalo Negro Fish.com mexicana (mexican-fish.com)).....	2
3. Mapa de distribución geográfica <i>C. viridis</i> marcado en verde y <i>C. nigrescens</i> marcado en azul.....	4
4. Genomas mitocondriales y filogenética [ilustración], por Elizabeth Collado, 2023.	7
5. Gel de electroforesis para la extracción de ADN. E marca la escalera, CVI (<i>Centropomus viridis</i>) CNI (<i>Centropomus nigrescens</i>).....	24
6. Datos generales de la secuenciación masiva para <i>Centropomus viridis</i>	26
7. Datos generales de la secuenciación masiva para <i>Centropomus nigrescens</i>	26
8. Comparación de los contigs obtenidos del ensamblaje de Novo para <i>Centropomus viridis</i> utilizando CLC (un contig) y NOVOPlasty (11 contigs), los contigs propuestos por NOVOPlasty se encuentran en verde y el del CLC fue el contig 284.	27
9. Anotación del genoma mitocondrial de <i>Centropomus viridis</i> creada con MitoFish	28
10. Anotación del genoma mitocondrial de <i>Centropomus nigrescens</i> creada con MitoFish.....	29
11. Alineamiento de los genomas mitocondriales de <i>Centropomus viridis</i> , <i>Centropomus nigrescens</i> , <i>Lates calcarifer</i> , <i>Sphyaena barracuda</i> y <i>Sphyaena pinguis</i> realizado con el programa CLC Genomics Workbench 10.1	31
12. Árbol filogenético realizado con información de genomas mitocondriales completos	33
13. Árbol filogenético realizado con secuencias del gen 18S ARNr	34
14. Diagrama de Venn para la ontología genética (GO) de <i>Centropomus viridis</i>	36
15. Diagrama de Venn para la ontología genética (GO) de <i>Centropomus nigrescens</i>	37
16. Estadísticas generadas por KRAIT para longitud de microsatélites para- <i>C. viridis</i> , <i>C. nigrescens</i> y <i>Lates calcarifer</i>	38
17. Estructura de los ARNt de las especies <i>Centropomus viridis</i> (a) y <i>Centropomus nigrescens</i> (b)...	56

Lista de Tablas

Tabla	Página
1. Clasificación taxonómica de la Familia Centropomidae (Ahyong <i>et al.</i> , 2023).....	3
2. Información sobre los genes mitocondriales de <i>Centropomus viridis</i> obtenidos de la base de datos NCBI, debido a que había secuencias repetidas solo se incluyó la información y el número de acceso de una secuencia por gen. (Revisado el 11/06/23).....	8
3. Listado de los genes mitocondriales en <i>Centropomus nigrescens</i> obtenidos de la base de datos NCBI (Revisado el 11/06/23).....	8
4. Secuencias utilizadas como "semilla" para la configuración de NOVOPlasty.....	17
5. Cuantificación por Nanodrop.....	23
6. Cuantificación por Qubit3 y dilución para enviar a secuenciar. Volumen final 20 µl.....	24
7. Estadísticas de Ensamblaje realizado con CLC.....	27
8. Comparativo de genes mitocondriales de las especies <i>Centropomus viridis</i> , <i>Centropomus nigrescens</i> , <i>Lates calcarifer</i> , <i>Sphyrana barracuda</i> , <i>Sphyrana pinguis</i> del orden Carangaria <i>incertae sedis</i>	30
9. Tabla de genes identificados para <i>C. viridis</i> utilizando Blastx, los números de acceso presentado son de secuencias de otras especies que previamente tenían anotación.....	63
10. Tabla de genes de interés identificados para <i>C. nigrescens</i> utilizando Blastx, los números de acceso presentado son de secuencias de otras especies que previamente tenían anotación.....	64
11. Comparación de costos entre protocolos de extracción de ADN.....	65

Capítulo 1. Introducción

Centropomus viridis, Lockington, 1877, y *Centropomus nigrescens*, Günther, 1864, son especies de la familia Centropomidae, morfológicamente se caracterizan por alcanzar grandes tallas llegando a medir hasta 2 metros de longitud, aunque la talla comercial es a partir de 60 cm. El cuerpo de *C. viridis* o robalo plateado como se le conoce comúnmente, presenta tonalidades plateadas o grises (Granados, 2018) (Fig.1). En cuanto al robalo prieto (*C. nigrescens*) generalmente es plateado y el rasgo característico que posee es una línea lateral negra en todo su cuerpo.

Es importante destacar que estas especies son muy populares en nuestro país debido al sabor de su carne y alto valor nutricional. De acuerdo con el anuario estadístico de acuicultura y pesca publicado en el año 2021, se obtuvieron 8,355 toneladas (t) de robalo obtenidas por pesca, mientras que en la acuicultura se produjo en el mismo año 303 t. Por lo tanto, debido a la producción pesquera el robalo se encuentra en la posición 23, pero, en cuanto a su valor comercial se encuentra en la posición 13.

A pesar de tener un valor comercial alto se estima que la tasa media de crecimiento anual en cuanto a producción acuícola es de -2.20 % de crecimiento anual en los últimos 10 años, mientras que el sector acuícola presenta un crecimiento sostenido en las últimas décadas. Esto sugiere que podría haber una problemática en la cual la genética o estudios genéticos pueden ayudar a incrementar los niveles de producción (CONAPESCA, 2021). Los principales avances tecnológicos en el cultivo de robalo plateado (*C. viridis*) se iniciaron en el 2015. Las primeras etapas consistieron en adaptar a los reproductores a vivir en cautiverio brindándoles un ambiente óptimo para su aclimatación y reproducción, este proceso fue realizado con éxito por personal del Centro de Investigación en Alimentación y Desarrollo (CIAD) en Mazatlán obteniendo 120,000 crías con una supervivencia mayor al 60 % en 2017 (INAPESCA, 2020).



Figura 1 *Centropomus viridis* ejemplar utilizado en el estudio de la revisión de caracteres morfológicos de robalos (Martínez-Brown *et al.*, 2021). (Foto Rodrigo Martínez Moreno).

Mientras que en el robalo prieto (*C. nigrescens*), en el año 2018, se realizaron las primeras pruebas de adaptación de la especie con el propósito de crear un sistema de cultivo extensivo, pero a pesar de que la especie presenta cualidades como una alta capacidad adaptativa, este proceso no obtuvo el éxito esperado, obteniendo como resultados un porcentaje de fertilización de óvulos del 0.7 % (Escárcega-Rodríguez, 2018).

En la actualidad existen avances en el estudio genético de estas especies, investigaciones previas demuestran que el genoma mitocondrial ha sido importante para conocer los avances evolutivos del organismo con relación a la filogenia (Tringali *et al.*, 1999). Conocer el genoma mitocondrial de esta especie ayudará a la problemática de diferenciación morfológica marcada en el artículo de (Martínez-Brown *et al.*, 2021)

Adicionalmente al conocimiento del mitogenoma, se podrían identificar otros genes de interés para la acuicultura, lo que contribuirá a la caracterización genética de las especies. Por mencionar unos ejemplos, una posible solución a los problemas de reproducción, podría ser la identificación de genes que ayuden a la reproducción y fisiología para la adaptación de la especie a cambios de temperatura y emplearlos en programas de mejoramiento genético con el objetivo de resolver las problemáticas presentes en los sistemas de cultivo, contribuyendo a la producción acuícola.

Además, la información obtenida puede aportar beneficios para la conservación de la especie, manteniendo el flujo genético a través de cruza planeadas. Por lo anterior, la información que se obtenga de la secuenciación del mitogenoma podrá ser de utilidad al sector gubernamental, cooperativas y particulares en el uso adecuado de este recurso con el fin de mejorar la calidad de vida de los organismos y la economía del país.



Figura 2 Robalo negro *Centropomus nigrescens* [fotografía], por Chris Wheaton,2019, Mexican fish (Róbalo Negro | Fish.com mexicana (mexican-fish.com))

1.1 Antecedentes

1.1.1 La familia Centropomidae

Esta familia se encuentra integrada por un total de 13 especies que pueden habitar en lugares tropicales tanto en aguas marinas como en agua dulce. Anteriormente se consideraba que en este grupo se encontraban dos subgéneros, *Centropomus* distribuidos desde el océano Pacífico al Atlántico; mientras que *Latidae* se distribuye en el Indo Pacífico principalmente en lugares como Australia y África (Greenwood, 1976). En la actualidad la taxonomía a nivel de Orden no está resuelta del todo por lo que se describe como *Carangaria incertae cedis* (Tabla 1).

Tabla 1. Clasificación taxonómica de la Familia Centropomidae (Ahyong *et al.*, 2023)

Reino	Animalia
Filo	Chordata
Clase	Actinopteri
Orden	<i>Carangaria incertae cedis</i>
Familia	Centropomidae
Género	<i>Centropomus</i>

1.1.2 Biología de *Centropomus viridis* y *Centropomus nigrescens*

El robalo blanco, *Centropomus viridis*, es originario del continente americano, se distribuye desde el Golfo de California, México hasta Perú y las Islas Galápagos. Su talla mínima comercial es de 60 cm, poseen un cuerpo alargado que puede alcanzar hasta los 2 metros de largo cubierto de escamas ásperas de tonalidades verdosas o café amarillento con vientre blanco (Granados, 2018).

Son peces carnívoros que se alimentan de una gran variedad de crustáceos y otros peces pequeños (Vergara-Chen, 2014). Los parámetros fisicoquímicos que necesita esta especie para vivir adecuadamente en cautiverio son los siguientes: temperatura (27.2-31.9 °C), oxígeno disuelto (5.38-14.17 mg l⁻¹), pH (7.4-8.0), salinidad (0.2-25.0), alcalinidad total (95-176 mg l⁻¹), dureza total (150-6000 mg l⁻¹), nitrógeno amoniacal total (0.031-0.051 mg l⁻¹) (Escárcega-Rodríguez, 2018). En cuanto a su ciclo reproductivo, esta

especie es protrándrica, es decir, madura primero como macho y luego experimenta una conversión sexual a hembra cuando alcanza aproximadamente los 15 cm de longitud (Alvarez-Lajonchère & Tsuzuki, 2008)

Centropomus nigrescens, conocido como robalo negro, se puede encontrar desde Mazatlán Sinaloa, Baja California Sur México hasta Buena Aventura, Colombia (Rivas, 1986). *C. nigrescens* alcanza tallas de 117 cm, sus escamas son plateadas con una línea lateral color negro, las aletas dorsales y caudales también presentan un color oscuro.

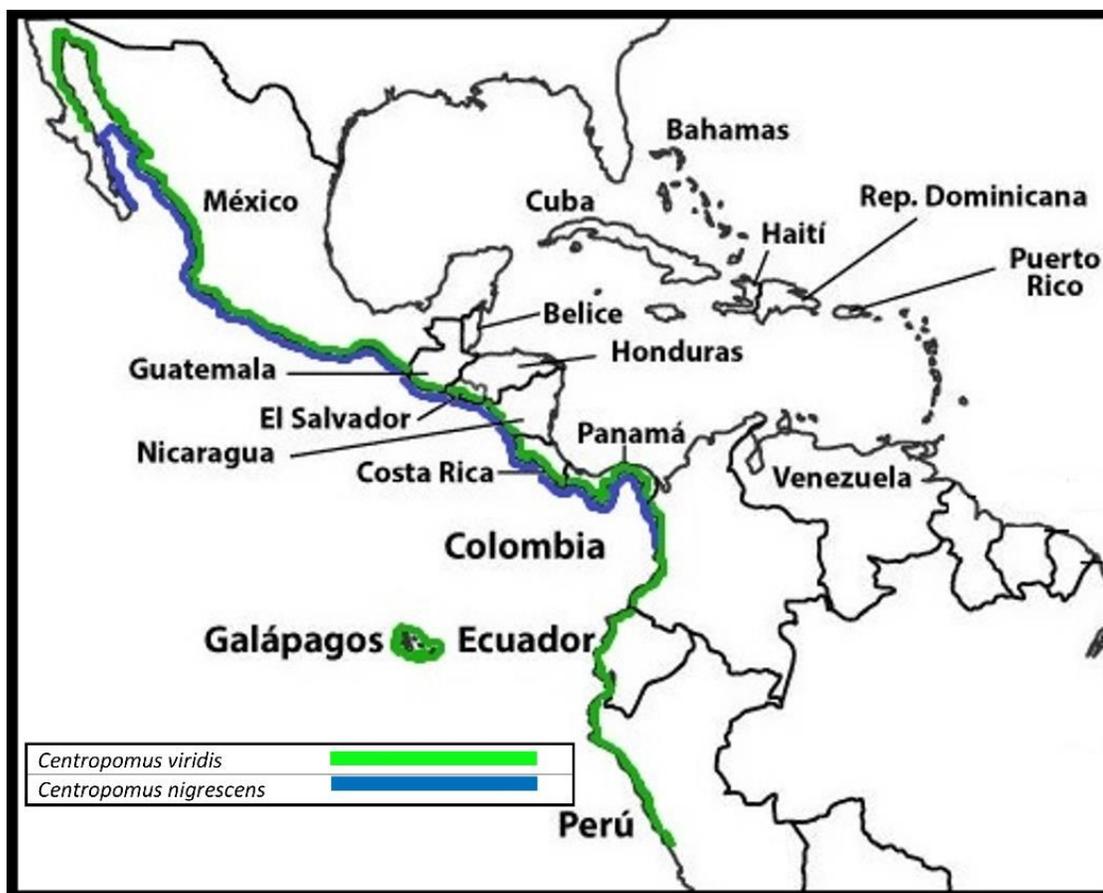


Figura 3 Mapa de distribución geográfica *C. viridis* marcado en verde y *C. nigrescens* marcado en azul

Los parámetros fisicoquímicos que necesita esta especie para llevar a cabo todos sus procesos fisiológicos son: temperatura (27 -31.9 °C), oxígeno disuelto (5.38-14.17 mg l⁻¹), pH (7.4-8.0), salinidad (0.2-25.0), alcalinidad total (95-176 mg l⁻¹), dureza total (150-6000 mg l⁻¹), nitrógeno amoniacal total (0.031-0.051 mg l⁻¹). A pesar de tener sus parámetros establecidos son resistentes a condiciones críticas, sobreviven en aguas de alta turbidez y baja calidad con niveles de oxígeno disuelto por debajo de 1.0 mg l⁻¹ (Escárcega-Rodríguez, 2010).

1.1.3 Importancia del robalo en México

Los robalos forman parte del grupo de peces que alcanzan mayor precio en el mercado con un aproximado de \$135 el kg , aunque estos datos pueden variar de acuerdo con la localidad. La mayoría de este producto se obtiene de la pesca y no del cultivo lo que ocasiona que no cubra la demanda del mercado nacional e internacional y que durante ciertos periodos el robalo sea escaso en mercados locales, aumentando su valor en ciertas temporadas del año(CIBNOR, 1994).

Por estas razones, no solo es valioso a nivel nutricional, su venta también ayuda a la economía del país. Datos oficiales del año 2017 reportaron un volumen comercial de 18,114 t con un valor estimado de \$ 755,000 miles de pesos, mientras que en el año 2018 estas cifras disminuyeron a 18,056 t con un valor de \$ 948,457 miles de pesos donde la mayoría del pescado obtenido es por parte de la pesca (CONAPESCA, 2018), mientras que para el 2020 solo se obtuvieron 9,391 t con un valor de \$ 814,374 miles de pesos.

Las estadísticas de ese año también reportaron que de la acuicultura solo se obtuvieron 101 t. Por lo tanto de acuerdo a este reporte, en los últimos 10 años la tasa media de producción anual es de 3.08% (CONAPESCA, 2020). Para apoyar al desarrollo en la acuicultura de este organismo el Centro de Investigación en Alimentación y Desarrollo Mazatlán (CIAD) está trabajando con el cultivo de robalo blanco.

De acuerdo con el Dr. Martínez-Brown (CIAD) se busca adaptar a reproductores de *C. viridis* a vivir en cautiverio y lograr un ambiente adecuado para su reproducción. En el año 2017 se obtuvieron alrededor de 1,630,000 juveniles. En 2019, fue la última producción reportada, esta actividad se suspendió durante el periodo 2020-2022 debido a la pandemia COVID-19. Se espera continuar con este proyecto en el año 2023 (Martínez-Brown, CIAD, comunicación personal, 19 de junio de 2023).

Existen investigaciones que proporcionan información valiosa sobre el cultivo de estas especies con el propósito de optimizar la acuicultura a escala comercial, proponiendo el uso de jaulas flotantes como una manera técnica y económicamente viable para el cultivo del robalo blanco (Giovanni *et al.*, 2022). Mientras que para el robalo prieto (*C. nigrescens*) la situación es distinta, en 2018 se realizó una primera prueba para el cultivo extensivo de esta especie en la costa de Michoacán, México, aunque estos organismos poseen grandes capacidades adaptativas, el proyecto no pudo implementarse con éxito, por lo se necesita seguir investigando con el fin de lograr producciones a gran escala (Escárcega-Rodríguez, 2018).

1.2 Importancia de trabajos genéticos para la acuicultura

Los objetivos de los trabajos en la acuicultura son muchos, pero entre los principales, se encuentran optimizar la tasa de conversión alimenticia, reducir el impacto ambiental y propiciar la resistencia a enfermedades, crear programas de repoblamiento y establecer las relaciones filogenéticas de cada población (Hahn-von Hessberg *et al.*, 2015).

Emplear trabajos genéticos en acuicultura nos ayuda a conocer y evaluar la filogenia de los organismos acuáticos, elaborar programas de mejoramiento genético para los cultivos con el objetivo de preservar el acervo genético, darles una mejor calidad de vida a las especies y mantener las características de interés, en la criopreservación de los organismos y en aspectos nutricionales evaluando genes del sistema digestivo. Conocer la información adecuada podría ayudar a hacer un uso oportuno de estos recursos por parte de pesquerías privadas, sector gubernamental y centros de investigación para estudios a futuro (Donaldson & Wilson, 1999).

Para lograr todos estos objetivos que se han planteado a lo largo de los años con el fin de proporcionar entre muchas cosas, alimento de calidad, es relevante conocer y comprender la estructura genómica y la organización de las especies acuícolas debido a que a partir de la información genética se pueden desarrollar tecnologías para la selección asistida a través de marcadores moleculares donde es importante el uso de un conjunto de herramientas y recursos genómicos, esto incluye las secuencias de referencia y sus anotaciones tomando en cuenta los elementos reguladores codificantes y no codificantes.

El análisis de las secuencias del genoma en varios estados de ensamblaje (parcial o completo) permite la identificación de variaciones genómicas, como inserciones o deleciones, polimorfismos de un solo nucleótido entre otras. A pesar de ello esta información solo es útil cuando se usa para predecir fenotipos que tengan un impacto positivo en la producción o calidad del producto (Abdelrahman *et al.*, 2017).

La secuenciación de los genomas y el análisis de los transcriptomas son las bases para generar recursos genómicos que posteriormente servirán para realizar estudios genéticos (Shen & Yue, 2019). Es importante indicar que el ADN se localiza tanto en el núcleo de las células eucariotas como en las mitocondrias, y en estas últimas su tamaño, por lo general, no excede los 20 000 pares de bases (pb). Estudios realizados sobre secuenciación completa de mitogenomas sugieren que la información presentada es una fuente valiosa para estudios de taxonomía molecular, diversidad genética y estructura de población además de contribuir a estrategias de conservación (Sharma *et al.*, 2020).

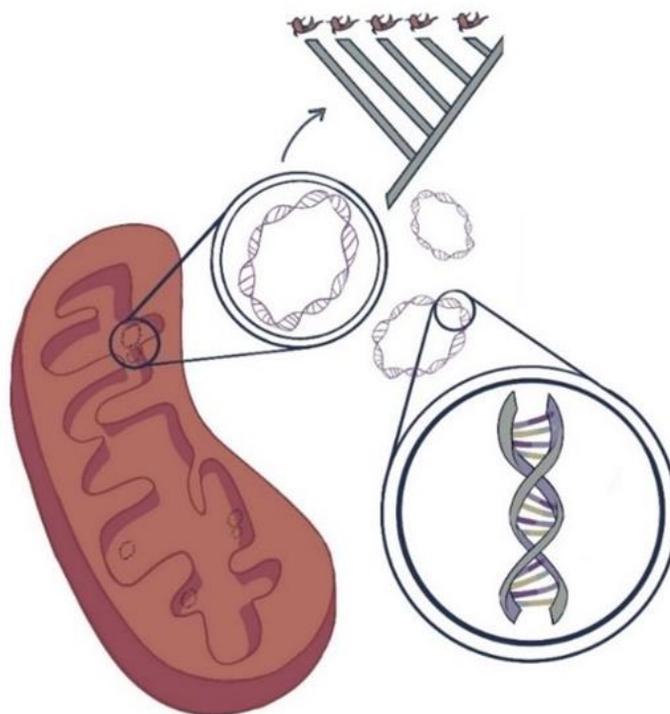


Figura 4. Genomas mitocondriales y filogenética [ilustración], por Elizabeth Collado, 2023.

1.2.1. Trabajos genéticos en Centropomidae y especies de familias afines

Se han realizado trabajos para conocer la taxonomía a nivel molecular, uno de ellos fue realizado por (Tringali *et al.*, 1999) donde se seleccionaron 12 especies de la familia Centropomidae y se analizaron mediante electroforesis utilizando como marcador el gen de la subunidad 16s del ribosoma (16S ARNr) de la mitocondria, dando como resultados de acuerdo a la filogenia obtenida que tenían gran similitud con los datos propuestos por (Rivas,1986).

En la actualidad existen 65 anotaciones de genes relacionados con *C. viridis* en el Centro Nacional de Información Biotecnológica (NCBI) de las cuales, 24 corresponden a secuencias de ADN mitocondrial del robalo plateado (Tabla 2).

El robalo prieto (*C. nigrescens*) solo cuenta con 6 secuencias registradas en la base de datos NCBI y estas anotaciones corresponden a genes mitocondriales (Tabla 3).

Tabla 2. Información sobre los genes mitocondriales de *Centropomus viridis* obtenidos de la base de datos NCBI, debido a que había secuencias repetidas solo se incluyó la información y el número de acceso de una secuencia por gen. (Revisado el 11/06/23)

Número de acceso	Secuencia	Referencias
MN196669	<i>Centropomus viridis</i> aislado ARN ribosómico de la subunidad grande del haplotipo 3.	(Martínez-Brown <i>et al.</i> , 2021)
MN196668	<i>Centropomus viridis</i> aislado gen de ARN ribosomal de la subunidad grande del haplotipo 2, secuencia parcial; mitocondrial.	(Martínez-Brown <i>et al.</i> , 2021)
DQ310724 DQ307689	Gen de la subunidad I (COI) del citocromo oxidasa. Gen del ARN ribosomal 16S de <i>Centropomus viridis</i> , secuencia parcial mitocondrial.	(Peregrino-Uriarte <i>et al.</i> , 2007) (Peregrino-Uriarte <i>et al.</i> , 2007)
DQ533169	<i>Centropomus viridis</i> 12s ribosomal secuencia parcial, ARNt-Val secuencia completa	(Smith & Wheeler, 2006)
MH397708	<i>Centropomus viridis</i> peroxisoma proliferador activado por el proliferador beta (ppar) ARNm, cds completos	(Parisi and Llera-Herrera, 2018) No publicado
MH397707	<i>Centropomus viridis</i> peroxisoma activado por el proliferador del receptor alfa (ppar) ARNm, cds completos	(Parisi and Llera-Herrera, 2018) No publicado
AF119325	<i>Centropomus viridis</i> NADH5 y genes de NADH6, genes mitocondriales	(Donaldson & Wilson, 1999)
MN756176	<i>Centropomus viridis</i> aislado MZ37 citocromo c oxidasa subunidad 1 (cox1) gen, cds parciales; mitocondrial	(Munguía-Vega <i>et al.</i> ,) No publicado
AF018603	<i>Centropomus viridis</i> aislado viri4 gen del citocromo b, gen mitocondrial que codifica la proteína mitocondrial, cds parciales	(Tringali <i>et al.</i> , 1999)
AF119322	<i>Centropomus viridis</i> , secuencias de ARNt-Pro, Región control y ARNt-Phe	(Donaldson & Wilson, 1999)

Tabla 3. Listado de los genes mitocondriales en *Centropomus nigrescens* obtenidos de la base de datos NCBI (Revisado el 11/06/23)

Número de Acceso	Secuencia	Referencias
MW201712	Gen de la subunidad I (CO1) del citocromo oxidasa de <i>Centropomus nigrescens</i> voucher LCL 1051 cd parciales; mitocondrial.	(Miranda <i>et al.</i> , 2021)
AF018604	Gen del citocromo b de <i>Centropomus nigrescens</i> , gen mitocondrial que codifica la proteína mitocondrial, cd parcial.	(Bert <i>et al.</i> , 2016) No publicado
U85015	Gen 16S ARN ribosomal de <i>Centropomus nigrescens</i> mitocondrial, secuencia parcial.	(Tringali <i>et al.</i> , 1999)

1.3 Secuenciación masiva

La secuenciación por el método de Sanger, conocida como de primera generación ha sido de las más utilizadas y una prueba estándar para detectar variables en el ADN, pero el alto costo que implicaba al principio, la cantidad de pares de bases que se podían obtener y el tiempo excesivo que se necesitaba para secuenciar un genoma completo abrieron paso a la creación de nuevas técnicas conocidas como la secuenciación masiva, secuenciación de siguiente generación o Next Generation Sequencing (NGS) (aunque actualmente se está utilizando el término de Segunda Generación) es un grupo de tecnologías diseñadas para secuenciar una gran cantidad de segmentos de ADN de forma masiva, en paralelo, en menor tiempo y a un menor costo.

Iniciando desde el 2005, la comercialización de pruebas NGS han permitido la obtención de lecturas de secuencias de forma masiva y paralela en menor tiempo y costo. Además, esta tecnología tiene como punto principal la profundidad (representando el número de veces que se está secuenciando un fragmento en promedio) y la cobertura (que indica el porcentaje de bases del genoma de referencia que están siendo secuenciados en una cantidad determinada de veces) (Rubio *et al.*, 2020).

Actualmente, aunque existen los métodos de tercera generación, la compañía Illumina (San Diego, C.A, USA) de segunda generación aún lidera el mercado debido a que ha desarrollado a lo largo del tiempo distintos modelos de secuenciadores, todos los equipos de Illumina realizan secuenciación por síntesis (Ju *et al.*, 2006).

Debido a la NGS, la biotecnología por sí sola ya no es tan adecuada para generar información detallada sobre los peces. Ahora se necesita complementar a la biotecnología con la bioinformática para proporcionar detalles moleculares cruciales que puedan ayudar a el cultivo y conservación de las especies y que sean previamente desconocidos (Rather *et al.*, 2023).

1.3.1 Mitogenomas de vertebrados mediante secuenciación masiva

El ADN mitocondrial es un tipo de genoma extra cromosómico que cuenta con características únicas como: herencia materna, baja tasa de recombinación, estructura simple. Estas características lo hacen útil en el estudio de identificación molecular y relaciones genéticas (Li *et al.*, 2022).

Generalmente el genoma mitocondrial tiene un tamaño aproximado de 16,000 a 17,000 pb y se encuentra conformado por 22 ARNt(moléculas que poseen un papel principal en la traducción debido a que forman nexos con ARNm que dan origen a la síntesis de proteínas (Kirchner & Ignatova, 2015)), 13 genes que codifican a proteínas y 2 ARNr(conocido también como ARN ribosomal) (16s y 12s) (Boore, 1999).

Debido a todas estas propiedades del ADN mitocondrial ha recibido mucha atención en el área de la acuicultura(Lin *et al.*, 2006). Los trabajos secuenciación masiva antes de la NGS se realizaban utilizando la técnica “shotgun sequencing” que se basa en amplificar regiones mediante PCR (Lin *et al.*, 2006).

Se han secuenciado los genomas mitocondriales de peces de potencial acuícola como el pejelagarto, *Atractosteus tropicus* (Gill, 1863) donde de acuerdo con la información obtenida y utilizando los fragmentos de los genes citocromo B y citocromo oxidasa II, se descubrió una baja diversidad genética en esta especie (Del Río-Portilla *et al.*, 2016).

Utilizando la secuenciación masiva de segunda generación se han logrado obtener los mitogenomas de otros organismos no modelo, pero de interés acuícola, que ayudan a las pesquerías a aportar recursos socioeconómicos como: la tenguayaca (*Petenia splendida*, Günter, 1862) (Del Río-Portilla *et al.*, 2016), el lenguado de California (*Paralichthys californicus*, Ayres, 1859) (Vargas-Peralta *et al.*, 2020). Asimismo, se ha adquirido información de peces marinos como el bacalao negro (*Anoplopoma fimbria*, Pallas,1814) (Galván-Tirado *et al.*, 2015) y especies de importancia ecológica como el beaubrummel (*Stegastes flavilatus*, Gill, 1872) (León-Pech *et al.*, 2016). Obtener el genoma mitocondrial completo para todas estas especies es el paso principal para establecer filogenias y comprender mejor a estos recursos poco estudiados.

1.4 Microsatélites

Una acuicultura sostenible depende de muchos factores, entre ellos, la disponibilidad y uso de recursos genéticos. Es importante generar información para la creación y uso de marcadores moleculares, esto ayuda a comprender los mecanismos que contribuyen a una variación fenotípica, estructura molecular y estudios de parentesco (Yue *et al.*, 2023). Los marcadores moleculares también han resultado beneficiosos para la selección de reproductores como para la planeación de estrategias basadas en la diversidad genética de varias especies de peces (Pandey & Parhi, 2022)

En los últimos años, se ha avanzado mucho en el desarrollo de marcadores de ADN en especies acuícolas como salmones, tilapia, bagre, ostras y camarones. Una de las preguntas más importante al momento de realizar investigaciones genómicas es qué tipo de marcador molecular es el más adecuado para nuestro trabajo (Liu & Cordes, 2004).

Los microsatélites, también conocidos como repeticiones de secuencia simple (SSR), representan secuencias repetidas en tándem (una serie de 2-6 nucleótidos, repetidos una detrás de la otra) que se encuentran distribuidos a lo largo de todo el genoma y presentan altos niveles de polimorfismo (presencia de dos o más variantes de una secuencia específica de ADN, estas pueden producirse en diferentes especies o entre diferentes poblaciones).

Son marcadores codominantes, por lo cual, permiten diferenciar a los organismos homocigotos de los heterocigotos, además de que pueden amplificarse fácilmente con la reacción en cadena de la polimerasa (PCR) (Chistiakov *et al.*, 2006).

En cuanto a su composición, se pueden clasificar de la siguiente manera: a) perfectos, si están compuestos en su totalidad por repeticiones iguales con un solo motivo, b) imperfecto se le conoce así si entre repeticiones aparece un par de bases que no pertenece al motivo, c) interrumpido, si se inserta una secuencia de una o más bases en el motivo; o d) compuesto se le conoce así cuando está formado por motivos múltiples, contiguos y repetitivos (Oliveira *et al.*, 2006).

Se puede afirmar que la presencia de SSR es menor en regiones génicas (regiones que codifican a proteínas), debido a que poseen una alta tasa de mutación que podría afectar la expresión génica. Sin embargo, algunos estudios indican que en las regiones codificantes se pueden encontrar microsatélites con motivo de repetición del tipo tri- y hexa-nucleótido resultado de la presión de selección frente a mutaciones que alteran el marco de lectura (Xu *et al.*, 2013).

Se han realizado estudios empleando estos marcadores moleculares en especies de peces con alto valor comercial como la Corvina del Golfo (*Cynoscion othonopterus*) (Jordan & Gilbert, 1882) y la cual se encuentra regulada mediante un plan de manejo oficial (NOM-063-PESC-2005). En esta investigación se analizaron 187 muestras de Corvina del Golfo, para once loci de microsatélites para poder estimar la diversidad genética sobre el estado de conservación de esta especie, además de si existía alguna evidencia de cuello de botella. Como resultado se obtuvo un promedio de 16.27 alelos por locus y de las tres pruebas realizadas no se obtuvo evidencia de un efecto de cuello de botella (Arce-Valdés *et al.*, 2023)

Existen diferentes programas de búsqueda de SSR como el msatcommander (Faircloth, 2008) y el KRAIT (Du *et al.*, 2018). Se ha utilizado la herramienta KRAIT con información obtenida mediante la secuenciación masiva para peces betta (*Betta splendens* Regan, 1910).

Con este programa se pudieron obtener microsatélites perfectos e imperfectos usando la configuración predeterminada. Para conocer el genotipado de estos peces se tomó la información de 12 loci de microsatélites para probar la efectividad de los microsatélites obtenidos se realizaron PCR validando con éxito dos loci (Wattanadilokchatkun *et al.*, 2022)

1.5 Justificación

Los robalos son especies de interés en las pesquerías y existen esfuerzos por iniciar su cultivo. Para el cultivo de especies acuáticas, todos los esfuerzos inician con la captura de reproductores del medio natural y posteriormente se van planteando programas para obtener reproductores de alto rendimiento a través de programas de mejoramiento genético que requiere de marcadores moleculares para lograrlo. Se considera importante que los programas de mejoramiento genético inicien junto con los esfuerzos de cultivo de las especies, con el fin de acelerar el proceso de adaptación al cautiverio de las especies en cultivo.

El aporte principal de este trabajo es que permitirá la caracterización de los genomas mitocondriales completos de las especies *Centropomus viridis* y *Centropomus nigrescens*, comparándolos con algunas secuencias registradas de miembros de la familia Centropomidae, lo cual permitirá conocer las relaciones filogenéticas de la especie a nivel mitogenoma. Asimismo, la información que se obtenga ayudará a investigaciones posteriores, programas de mejoramiento genético, conservación, a un uso adecuado de los recursos genéticos para un buen manejo en el cultivo de estos organismos y a la economía del país.

Por otro lado, en la mayoría de los artículos solo se han presentado los mitogenomas y no se han analizado otro tipo de secuencias obtenidas en la secuenciación masiva, a pesar de que se menciona un gran número de contigs formados. Estos contigs podrían proporcionar información importante sobre la genética de estas especies, analizar estos contigs podría ayudar obtener la información para elaborar mejores sistemas de cultivo empleando la información genética, seleccionando genes de interés para procesos fisiológicos, resistencia a enfermedades y resistencia a cambios de temperatura.

1.6 Hipótesis

Existirá una similitud en cuanto al orden de los genes y longitud entre los mitogenomas de las especies *Centropomus viridis* y *Centropomus nigrescens* con el de *Lates calcarifer* (barramundi).

1.7 Objetivos

1.7.1 Objetivo general

Comparar los genomas mitocondriales de las especies *Centropomus viridis* y *Centropomus nigrescens* con otros mitogenomas completos del orden Carangaria.

1.7.2 Objetivos específicos

Obtener el genoma mitocondrial de las especies *Centropomus viridis* y *Centropomus nigrescens*.

Establecer las relaciones filogenéticas de *Centropomus viridis* y *Centropomus nigrescens* con otras especies del orden Carangaria a nivel mitogenómico.

Identificar genes en el resto de los contigs obtenidos en las secuencias de *Centropomus viridis* y *Centropomus nigrescens*, y hacer énfasis en aquellos que codifican a proteínas de importancia fisiológica o acuicultural.

Evaluar el número de microsatélites identificados con la secuenciación parcial de *Centropomus viridis* y *Centropomus nigrescens*.

Comparar las relaciones filogenéticas de *Centropomus viridis* y *Centropomus nigrescens* con otras especies del orden Carangaria con genes nucleares de importancia taxonómica

Capítulo 2. Metodología

2.1 Origen del material biológico y extracción de ADN

Las muestras de *Centropomus viridis* provenían del noreste del estero del Yugo, Mazatlán Sinaloa (23° 18' 11" N, 106° 29' 00" W) y las de *Centropomus nigrescens* de un mercado local de pescados y mariscos en Tonalá Chiapas (Martínez-Brown *et al.*, 2021).

Las muestras que se utilizaron fueron donadas para el trabajo de (Martínez-Brown *et al.*, 2021), los tejidos se encontraban en buenas condiciones para ser usados en este proyecto y fueron conservados en tubos falcón con etanol al 70% a -20 °C en el Laboratorio de Genética Acuícola (LGA) del Departamento de Acuicultura del CICESE.

2.1.1 Extracción de ADN

Para realizar la extracción de ADN, se tomaron las muestras de branquias de *Centropomus viridis* y de las aletas de *Centropomus nigrescens* clasificados en 1, 2 y 3. Estas muestras provinieron de las utilizadas en el trabajo de Martínez-Brown *et al.* (2021). A estos tejidos se les extrajo el ADN mediante el protocolo de sales del LGA (el protocolo completo se encuentra en anexo A con el título: Protocolo de extracción de ADN del laboratorio de genética acuícola del CICESE).

2.1.2 Evaluación de la calidad y cantidad del ADN

Antes de enviar las muestras a secuenciar, se cuantificó y verificó la calidad e integridad del ADN mediante un Nanodrop, gel de electroforesis y cuantificación mediante flourometría (Du *et al.*, 2018), con el objetivo de evaluar que el ADN era de buena calidad y sin contaminación que pudiera interferir con la secuenciación masiva.

La cuantificación de ADN total extraído ($\text{ng } \mu\text{l}^{-1}$) se realizó mediante espectrofotometría en un espectrofotómetro marca NanoDrop™2000 (ThermoFisher Scientific) por medio de la lectura de absorbancias con una longitud de onda específica de 260 nm.

La pureza del ADN fue determinada mediante los parámetros 260/230 y 260/280, que son las razones de las lecturas obtenidas a longitudes de onda a 260,230 y 280 nm. Los rangos aceptables para el parámetro 260/230 son de entre 2.0 a 2.2, este parámetro estima la contaminación por fenoles. Para estimar la contaminación por proteínas se emplea el parámetro 260/280 donde los valores aceptables están entre 1.8 a 2.0 (Thermo Scientific, 2016).

Para verificar la calidad del ADN extraído también se realizó mediante un gel de electroforesis al 1% usando como amortiguador el Super Buffer 1x (Zhang *et al.*, 2011) a 90 Volts durante 60 min. El ADN se tiñó con GelRed™ (Biotium, 2009), una vez que concluyó el proceso en la cámara de electroforesis, los resultados se visualizaron en un fotodocumentador Gel Doc™ XR + Imager (Bio-Rad) (Protocolo completo en Anexo B : Protocolo para elaborar gel de electroforesis al 1%)

Por recomendaciones de la empresa Omega Bioservices, se realizó una cuantificación más a las muestras utilizando Qubit3 (ThermoFisher Scientific) para tener mejor precisión, debido a que se buscaba que los resultados de la secuenciación masiva fueran de buena calidad. Las muestras de mejor calidad fueron enviadas a la empresa Omega Bioservices (Norcross GA, Estados Unidos) de acuerdo con los parámetros sugeridos por la compañía, con una concentración de 8 ng μl^{-1} y un volumen total de 20 μl .

2.2 Protocolo de secuenciación.

De acuerdo con la información enviada por Omega Bioservices, para la secuenciación se utilizó un kit HyperPrep de Kapa Biosystems (Kapa Biosystems, Wilmington, MA, EUA) para la construcción de las bibliotecas, el ADN se fragmentó utilizando un sonicador biorruptor (Diagenode, Denville, NJ, EUA). Los extremos 3' del ADN fueron ligados a los adaptadores Illumina, los resultados obtenidos fueron empleados amplificados mediante PCR con cebadores indexados en un secuenciador Illumina HiSeq X10 (Illumina, San Diego, CA, EUA) utilizando un formato de ejecución de 150 pb (Baeza, 2018)

2.3 Calidad de las lecturas y ensamblaje *de Novo*

Una vez que se recibieron los datos de la empresa Omega Bioservices, se verificó la calidad de las lecturas para poder realizar el ensamblaje *de Novo*, este tipo de ensamblajes se utiliza cuando no hay información genómica de las especies de interés (Formenti *et al.*, 2021). Se utilizó el servidor Omica del CICESE

(<http://omica:8000/>), donde a través de bitácoras electrónicas de Jupyter (<https://jupyter.org>), se llevaron a cabo estos procesos.

2.3.1 Corroboración de la calidad y eliminación de regiones de baja calidad

En una bitácora de Jupyter se llevó a cabo todo el filtrado de los archivos fastq.gz generados por secuenciación masiva. Este proceso se hizo para las dos especies de robalo simultáneamente, mediante un código para poder correr el programa desde la bitácora (Anexo D: Código para realizar FastQC y trimmomatic).

Para eliminar los adaptadores, en las bitácoras se realizó un código para trimmomatic. Trimmomatic es un programa que nos permite filtrar secuencias de formato de lecturas individual o en pares (<http://www.usadellab.org/cms/?page=trimmomatic>). Siguiendo las instrucciones de la página oficial, con los archivos obtenidos del Trimmomatic se llevó a cabo un segundo análisis de FastQC, con el objetivo de corroborar la calidad de las secuencias después del proceso de recorte o “trimming”.

2.3.2 Ensamblaje *de Novo* usando TRINITY

Trinity es un ensamblador *de Novo* desarrollado principalmente para transcriptomas, aun así, se probó para ensamblar los mitogenomas de *C. viridis* y *C. nigrescens*. El proceso con Trinity se elaboró mediante el servidor Omica utilizando un código donde se colocaron las secuencias de los archivos pareados generadas por “trimmomatic”, utilizando la versión de Trinity 2.4.0. (código en Anexo E: Ensamblaje *de Novo* usando TRINITY) en donde se generó un archivo “.fasta” con la información de todos los “contigs” obtenidos, el cual descargó para análisis posteriores.

2.3.3 Calidad de las lecturas y ensamblaje *de Novo* utilizando CLC

2.3.3.1 Sequencing QC Report

Para evaluar la calidad de las secuencias CLC-Genomics Workbench V. 10.1 (QIAGEN) cuenta con la opción

“Sequencing QC report” para obtener las estadísticas generales de la secuenciación masiva. Las características evaluadas fueron las siguientes: porcentaje de Guanina y Citosina, número de errores en la secuenciación Illumina (representado con una N), porcentaje de distribución en cada par de bases.

2.3.3.2 Ensamblaje *de Novo*

La base del ensamblaje *de Novo* es unir las secuencias “limpias” que sean complementarias para armar un fragmento largo el cual se denomina “contig”, este proceso fue realizado con las secuencias que tenían un valor de Phred 20.

2.3.4 Ensamblaje *de Novo* utilizando NOVOPlasty

En NOVOPlasty (Dierckxsens *et al.*, 2016) es un programa que se utiliza para la obtención de plásmidos y otras secuencias cortas como los cloroplastos. Con este programa no fue necesario hacer el proceso de limpieza y filtrado debido a que su configuración no lo requiere. A diferencia de los otros ensambladores este necesitó de dos secuencias “semillas” presentadas en la tabla 4, los parámetros para el rango del genoma fueron de 16000-17000 pb. (Código utilizado se encuentra en anexo F: Código para realizar ensamblaje con NOVOPlasty)

Tabla 4. Secuencias utilizadas como "semilla" para la configuración de NOVOPlasty

Especie	Gen	Número de acceso	Autores
<i>Centropomus viridis</i>	Segmento del gen 16S rRNA, haplotipo 3	MN196669	(Martínez-Brown <i>et al.</i> , 2021)
<i>Centropomus nigrescens</i>	Segmento del gen 16S rRNA	U85015	(Tringali <i>et al.</i> , 1999)

2.4 Mapeo y extracción de “contigs”

2.4.1 Mapeo y extracción de contigs con el servidor Omica

Después de realizar el ensamblaje *de Novo* usando Trinity, el programa generó varios archivos, entre ellos

los que tienen el nombre de Trinity.fasta a estos archivos que contenían todos los “contigs” (estructuras conformadas por lecturas contiguas o superpuestas formando una secuencia de mayor longitud de nucleótidos), los cuales se descargaron a través una bitácora de Jupyter en el servidor y se les agregó un identificador para diferenciarlos y poder trabajar con los procesos de extracción, anotación y alineamiento que se realizaron después.

2.4.2 Mapeo y extracción de contigs con el servidor Omica

Después de realizar el ensamblaje *de Novo* usando Trinity, el programa generó varios archivos, entre ellos los que tienen el nombre de Trinity.fasta a estos archivos que contenían todos los “contigs” (estructuras conformadas por lecturas contiguas o superpuestas formando una secuencia de mayor longitud de nucleótidos), los cuales se descargaron a través una bitácora de Jupyter en el servidor y se les agregó un identificador para diferenciarlos y poder trabajar con los procesos de extracción, anotación y alineamiento que se realizaron después.

2.4.3 Mapeo y extracción de contigs con CLC Genomics Workbench

A las lecturas obtenidas de la secuenciación masiva, se obtuvo la estadística de calidad de la secuenciación con la opción Create Sequencing QC Report. Después se eliminaron regiones de baja calidad de las lecturas con la opción “Trim Sequences”, ambas opciones dentro del menú “NGS Core Tools”.

Posteriormente se realizó el ensamblaje “*de Novo*” y el mapeo de las lecturas a los contigs obtenidos con la opción “Map Reads to Contigs”.

Concluido el mapeo se hizo un proceso de filtrado para eliminar aquellos contigs que tenían menos de 3 lecturas y una cobertura menor a 1.5, ya que se concluyó todo el filtrado se seleccionaron y extrajeron todos los contigs en un solo archivo (.fasta) para cada especie y su análisis posterior.

2.5 Anotación del mitogenoma

Una vez que se generaron los archivos del ensamblaje, se identificaron los contigs, con longitudes

superiores de 1500 pb (pares de bases), en los que se encontraba el mitogenoma de *Centropomus viridis* y *Centropomus nigrescens* utilizando la herramienta Blast (Herramienta básica de búsqueda de alineación local). Posterior a este proceso se hizo una anotación mediante Mitofish (Iwasaki *et al.*, 2013) y Mitos (Bernt *et al.*, 2013) y para verificar el número de ARNt y su orden se pusieron los archivos “fasta” de los contigs que corresponden a los mitogenomas en la plataforma tRNA-scanSE (Chan and Lowe, 2019), para corroborar los tRNA.

2.5.1 Registro de los genomas mitocondriales en el GenBank

Se revisaron los requisitos para registrar los genomas mitocondriales en la base de datos NCBI y con base a esos lineamientos se registraron las secuencias para obtener un número de acceso y se utilizó la herramienta GenBank: Bankit (<https://www.ncbi.nlm.nih.gov/WebSub/>) en línea para someter los mitogenomas.

2.6 Análisis filogenético

2.6.1 Análisis con clustal Omega

Se realizó una búsqueda en el GenBank para obtener información sobre las secuencias a nivel genoma mitocondrial de especies del orden Carangaria. Con los mitogenomas obtenidos, se realizó un alineamiento con Clustal Omega, junto con las secuencias de *C. viridis* y *C. nigrescens*. De igual manera, se realizó un análisis con las secuencias del gen 18s para hacer una comparación y respaldo de los resultados obtenidos.

2.6.2 Selección del modelo mediante Jmodel

Este proceso se realizó mediante una serie de comandos en una bitácora electrónica de Jupyter utilizando la herramienta jModelTest2 (Darriba *et al.*, 2012) este proceso nos permitió seleccionar a través de la estadística el mejor modelo para el análisis filogenético, los modelos de sustitución de nucleótidos son fundamentales a la hora de realizar cualquier análisis filogenético ya que nos permite obtener

estimaciones de máxima verosimilitud de los parámetros del modelo, implementa distintos criterios para la selección de un modelo incluyendo los criterios de información bayesianos y de Akaike(AIC y BIC) (Posada & Buckley, 2004)

2.6.3 Árbol filogenético mediante inferencia Bayesiana

Como resultado del Blast obtuvimos una lista de 100 especies, de las cuales solo se seleccionaron las que tenían un porcentaje de identidad arriba del 96% y que contaran con genoma mitocondrial completo. Se utilizó el programa Mr. Bayes (Ronquist *et al.*, 2012) para realizar un árbol filogenético a nivel genoma mitocondrial y de las secuencias 18S del ARN ribosomal (18S ARNr) de las especies seleccionadas. Se utilizó como grupo externo (outgroup) a *Amblyraja radiata* del orden Rajiformes tomando en cuenta la referencia en la investigación de (Lin *et al.*, 2006).

El número de generaciones (ngen) que se utilizó para realizar este análisis fue de 30 000, frecuencia de muestra (samplefreq) 100, el parámetro “printfreq” controla la frecuencia con la cual la información se imprimirá en la pantalla el parámetro utilizado fue de 100, al final se coloca el comando “sump” el cual sirve para examinar la desviación estándar asociada a cada clado. El análisis se consideró significativo cuando se obtuvo un valor de $p < 0.05$ y se detuvo el proceso. El código utilizado se encuentra en el Anexo G (Código para realizar árboles filogenéticos con Mr. Bayes). El árbol obtenido se pasó a formato png con el FigTree (Rambaut,2018, [FigTree \(ed.ac.uk\)](http://FigTree.ed.ac.uk)) y se editó la imagen con PowerPoint.

2.7 Blastn y Blastx

Los procesos de Blastn y Blastx fueron utilizados para identificar a qué genes correspondían el resto de los contigs, estos se realizaron mediante servidor Omica utilizando las bases de datos nt y swissprot-unitprot, respectivamente.

El código para este proceso se encuentra en Anexos. Los archivos generados por los Blast se guardaron en formato de valores separados por tabulaciones y se importaron a Excel (.xlsx), donde se realizaron tablas dinámicas para filtrar los genes de interés mencionados en los objetivos.

2.8 Ontología genética

En la página de UniProt (<https://www.uniprot.org/>) se descargó la base de datos “uniprot-compressed_true_download_true_fields_accession_2C-reviewed_2C-2023.02.24-06.02.50.80.tsv.gz” en formato. tsv, después de que se descargó el archivo se descomprimió en la terminal usando el comando unzip. Se relacionó el identificador obtenido en el blastx para relacionarla con el código de ontología genética. Con ello se pudo obtener la función de la proteína, así como si se tuvo un valor de Proceso Biológico (P), Función Biológica (B) o Componente Celular(C) para crear los diagramas de Venn.

En una bitácora de jupyter se descargaron las paqueterías: matplotlib, pandas, numpy, gzip y matplotlib_Venn para la creación de los diagramas. En una celda se leyó el archivo descargado de la base de datos uniprot. Para el diagrama se crearon 3 columnas “P” para proceso biológico, “C” para componentes celulares y “F” para función biológica. También en el archivo se separó la columna sseqid y se eliminaron los puntos de las versiones. Al final a los resultados se les anexó un comando para colocar también los porcentajes dentro de cada conjunto.

2.9 Evaluación del número de microsatélites

2.9.1 Obtención y evaluación de los microsatélites

Los microsatélites se realizaron utilizando la herramienta bioinformática KRAIT (Du *et al.*, 2018) con la información extraída del mapeo se obtuvieron los microsatélites SSRs (simple sequence repeats) y los iSSRs (imperfect simple sequence repeats). Asimismo, con las bases de datos del NCBI se obtuvieron los microsatélites para el genoma completo del barramundi, *Lates calcarifer* (PRJNA294489) y se hizo una comparación entre los resultados de *C. viridis*, *C. nigrescens* y *L. calcarifer*. Después del diseño de los microsatélites, se seleccionaron todos los identificadores de los SSRs y iSSRs para cada especie para hacer más fácil el diseño de los iniciadores (“primers”) que se utilizaron para identificar las regiones de interés. Este diseño se realizó con la opción “Redo Design Primers” y se exportaron todos en formato “tsv” para hacer la identificación de cada contig más fácil en Excel.

2.9.2 Búsqueda de “primers” con CLC Genomics Workbench

Con los resultados que se obtuvieron se escogieron regiones que correspondían a los genes de interés mencionados en los objetivos específicos, con la herramienta “find” se buscaron las secuencias de los primers y se marcaron en cada “contig” seleccionado.

Capítulo 3. Resultados

3.1 Origen del material biológico y extracción de ADN

Las muestras donadas que fueron utilizadas se encontraban en óptimas condiciones para realizar los procesos de extracción de ácidos nucleicos. Se realizaron tres pruebas para verificar la calidad e integridad de las muestras de ADN con el objetivo de cumplir con los parámetros de calidad sugeridos por Omega Bioservices y obtener mejores resultados en la secuenciación masiva.

3.1.1 Extracción, calidad y cuantificación del ADN

El método de sales nos permitió obtener una mayor cantidad y mejor calidad de ADN para branquias que para aletas. La calidad del ADN se verificó mediante tres métodos debido a que los tejidos de *C. nigrescens* presentaban un precipitado color blanco producto del tipo de tejido.

La cuantificación por Nanodrop presentó valores adecuados con relación a los parámetros 260/280 260/230, ya que *C. viridis* tuvo valores de 2.02 y 2.07, respectivamente, mientras que *C. nigrescens* tuvo valores de 1.95 y 0.90 respectivamente (Tabla 5).

Sin embargo, fue necesario hacer un gel de electroforesis (Fig. 5) para visualizar el peso molecular del ADN y evaluar las muestras de *C. nigrescens*. Adicionalmente, por recomendación del proveedor del servicio de secuenciación, se cuantificó la cantidad del ADN mediante Qubit que dio valores adecuados para el envío a secuenciar (Tabla 6).

Tabla 5. Cuantificación por Nanodrop

Muestra	Ácidos nucleicos	A260	A280	260/280	260/230
CVI (<i>Centropomus viridis</i>)	469.2 ng μl^{-1}	9.384	4.654	2.02	2.07
CNI (<i>Centropomus nigrescens</i>)	116.8 ng μl^{-1}	2.336	1.195	1.95	0.90

Tabla 6. Cuantificación por Qubit3 y dilución para enviar a secuenciar. Volumen final 20 μ l.

ID	Concentración	ADN	H2O
CVI	185.5 ng μ l ⁻¹	0.9 μ l	19.1 μ l
CNI	66.4 ng μ l ⁻¹	2.4 μ l	17.6 μ l

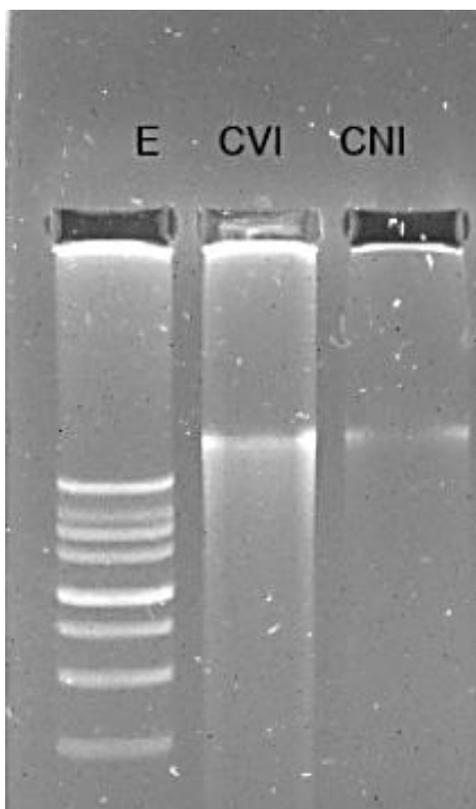


Figura 5. Gel de electroforesis para la extracción de ADN. E marca la escala, CVI (*Centropomus viridis*) CNI (*Centropomus nigrescens*)

3.2 Verificación de la calidad de las lecturas y ensamblaje *de Novo*

Los archivos fastq se descargaron comprimidos de la página de basespace (<http://basespace.illumina.com/>) y guardados en diferentes medios para su uso posterior. Los datos fueron manejados y ensamblados con distintos programas con el propósito de conocer cuál era el que proporcionaba la información más adecuada o si pudiera existir una diferencia entre los resultados.

3.2.1 Corroboración de la calidad y eliminación de regiones de baja calidad

A partir de la recepción de los resultados que envió la empresa Omega Bioservices, se evaluó la calidad de las secuencias obtenidas de *Centropomus viridis* y *Centropomus nigrescens*; en cuanto a la información de *C. viridis* las lecturas iniciales tuvieron un total de 1 511 583 084 bases, con un número total de lecturas de 10 010 484 y un contenido de guanina citosina (GC) de 42.43%.

Para *C. nigrescens* obtuvimos un total de 944 762 002 bases, con un total de lecturas de 6 256 702 y un contenido de GC 42.72%. Este proceso se realizó con distintos valores de calidad Phred debido a que en el servidor omicas no se podía colocar un valor menor a 33, mientras que en CLC Genomics Workbench 10.1 (CLC) se utilizó un valor de 20.

Debido a que la secuenciación masiva en términos generales dio buenos resultados, no era necesario utilizar un control de calidad tan alto, por lo que se pudo notar que se perdía información considerable al usar un valor de Phred alto, lo que pudo influir en los resultados de el ensamblaje con Trinity, ya que con este ensamblador el contig más largo fue de 10,000 pb

Con este ensamblador no se obtuvieron los mitogenomas completos para ninguna especie, esto puede ser debido a la configuración del programa ya que generó un archivo "fasta" con lecturas de una media de longitud de contig de 250 pb, con un total de bases 47,087,387 ensambladas y un porcentaje de GC del 41.78%. Debido a que Trinity es un programa utilizado especialmente para ensamblar transcriptomas, también en sus estadísticas nos arroja información de "transcritos", lo cual nos sirvió para hacer un Blastn o Blastx posteriormente en el cual se compararon los resultados con los de CLC.

3.2.1.1 Sequencing QC Report

Sequencing QC Report de CLC nos proporcionó datos generales de los archivos de la secuenciación masiva, los reportes generados fueron los mismos presentados por el programa FastQC y se encuentran en las figuras 6 y 7. Donde en el primer cuadro nos marca la información general del programa total de secuencias y total de nucleótidos obtenidos mediante la secuenciación masiva, también podemos observar los porcentajes de guanina-citosina y una longitud promedio de 151 pb por contig.

Reporte CLC para *Centropomus viridis*

Creation date:	Tue Oct 25 08:17:17 PDT 2022
Generated by:	migueldelrio
Software:	CLC Genomics Workbench 10.1.1
Based upon:	1 data set
1CVI_S1_L007_R1_001 (paired):	10,010,484 sequences in pairs
Total sequences in data set	10,010,484 sequences
Total nucleotides in data set	1,511,583,084 nucleotides

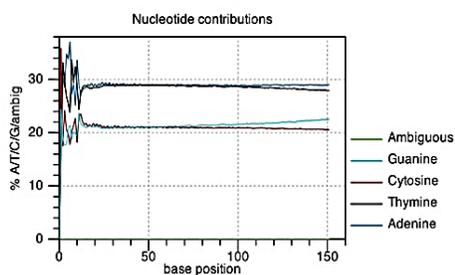
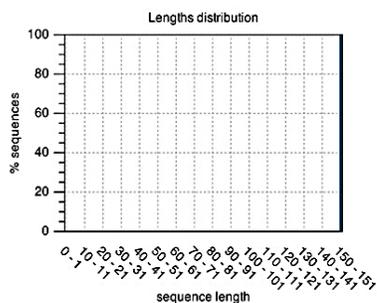
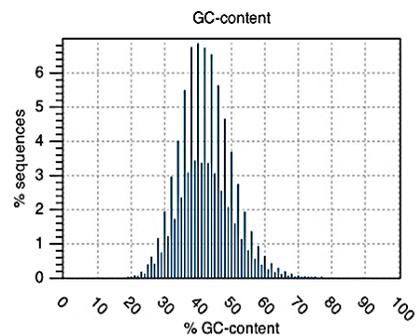


Figura 6 Datos generales de la secuenciación masiva para *Centropomus viridis*

Reporte CLC para *Centropomus nigrescens*

Creation date:	Tue Oct 25 08:16:26 PDT 2022
Generated by:	migueldelrio
Software:	CLC Genomics Workbench 10.1.1
Based upon:	1 data set
2CNI_S2_L007_R1_001 (paired):	6,256,702 sequences in pairs
Total sequences in data set	6,256,702 sequences
Total nucleotides in data set	944,762,002 nucleotides

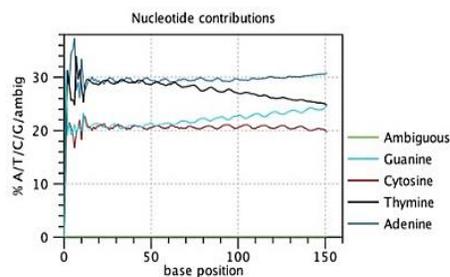
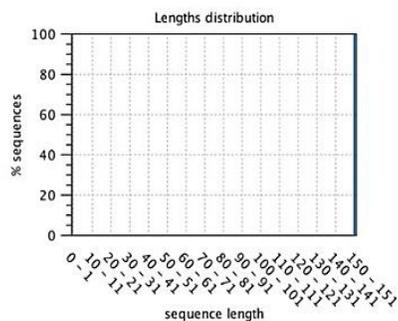
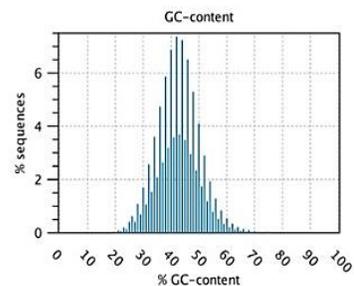


Figura 7 Datos generales de la secuenciación masiva para *Centropomus nigrescens*

3.2.1.2 Ensamblaje *de Novo* CLC Genomics Workbench

Este ensamblador ayudó a obtener el mitogenoma completo de ambas especies, además el hecho de que tenga un interfaz más gráfico y varias opciones hizo posible realizar más objetivos de este trabajo. El contig más largo identificado para *C. viridis* fue de 16,656pb mientras que para *C. nigrescens* fue de 16,665pb. Las estadísticas del ensamblaje se presentan en la tabla 7.

Tabla 7 Estadísticas de Ensamblaje realizado con CLC.

Especies	Lecturas ensambladas	N75	N50	N25	Min	Max	Promedio	Contigs
<i>C. viridis</i>	181,810,247	333	419	560	120	16,656	410	439,105
<i>C. nigrescens</i>	76,613,663	296	309	375	200	16,665	324	236,437

N: Valores Percentiles

3.2.2 Ensamblaje *de Novo* usando NOVOPlasty

El código utilizado con el NOVOPlasty (Anexos) permitió obtener tres propuestas de genomas mitocondriales para cada especie.

Sin embargo, cada una de ellas no representaban el genoma mitocondrial completo debido a que al momento de observar los resultados de un alineamiento con los fragmentos generados con el Novoplasty y con el del CLC Genomics Workbench 10.1 no se observaba un solo contig completo por lo cual se decidió quedarnos solo con el resultado de CLC y utilizarlo para hacer las anotaciones (Fig. 8).

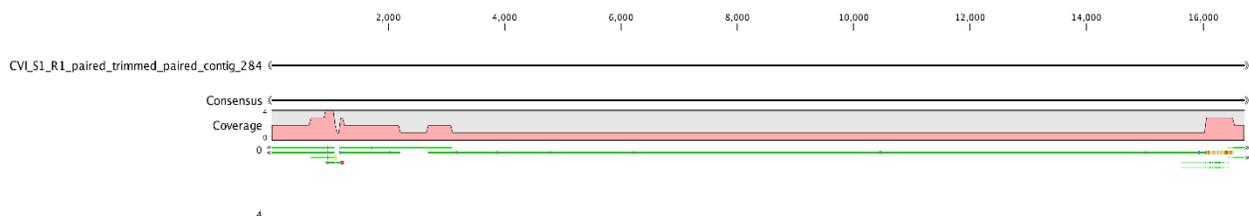


Figura 8 Comparación de los contigs obtenidos del ensamblaje de Novo para *Centropomus viridis* utilizando CLC (un contig) y NOVOPlasty (11 contigs), los contigs propuestos por NOVOPlasty se encuentran en verde y el del CLC fue el contig 284.

Tabla 8. Comparativo de genes mitocondriales de las especies *Centropomus viridis*, *Centropomus nigrescens*, *Lates calcarifer*, *Sphyræna barracuda*, *Sphyræna pinguis* del orden *Carangaria incertae sedis*

Nombre del Gen	Ubicación (tamaño en pb)					Sentido	Codón de inicio/anticodón	Codón de paro
	<i>Centropomus viridis</i> ^a	<i>Centropomus nigrescens</i> ^b	<i>Lates calcarifer</i> ^c	<i>Sphyræna barracuda</i> ^d	<i>Sphyræna pinguis</i> ^e			
Tamaño	16,656 pb	16,665 pb	16,526 pb	16,707 pb	16620 pb			
tRNA-F	1-69 (69)	1-70 (70)	1-68 (68)	1-69 (69)	1-68(68)	+	GAA	
12s ribosomal	70-166 (997)	71-1075 (1005)	69-1033 (965)	70-1022(952)	69-1022(955)	+		
tRNA-V	1067-1138 (72)	1076-1148 (73)	1034-1106 (73)	1023-1095(72)	1023-1094(72)	+	TAC	
16s ribosomal	1139-2881 (1701)	1213-2914 (1702)	1107-2860 (1754)	1096-2838 (1743)	1095-2839(1745)	+		
tRNA-L	2882-2954 (73)	2971-3043 (73)	2863-2935 (73)	2954-3027(73)	2888-2961(74)	+	TAA	
NADH deshidrogenasa subunidad 1	2957-3925 (969)	3052-4014 (963)	2936-3910 (975)	3028-4002(974)	2965-3936(972)	+	ATG ^(a-c) ATC ^(b, e) ATA ^(d)	CAA ^(a-d) CAG ^(b, e) TAA ^(d)
tRNA-I	3934-4003 (70)	4023-4092 (70)	3914-3984 (70)	4008-4078(70)	3941-4010(70)	+	GAT	
tRNA-Q	4003-4073 (71)	4092-4162 (71)	3985-4056 (72)	4078-4148(70)	4010-4080(71)	-	TTC	
tRNA-M	4086-4156 (71)	4173-4242 (71)	4056-4125 (70)	4157-4225(68)	4080-4150(71)	+	CAT	
NADH deshidrogenasa subunidad 2 (ND2)	4157-5191 (1035)	4243-5280 (1038)	4126-5172 (1047)	4226-5275(1049)	4152-5198(1047)	+	ATA ^(a) ATG ^(b, c, d, e)	CTC ^(a) ATG ^(b) TAA ^(c, d, e) CTT ^(e)
tRNA-W	5216-5285 (70)	5290-5361 (72)	5173-5244 (72)	5318-5388(41)	5206-5278(73)	+	TCA	
tRNA-A	5287-5355(69)	5363-5431 (69)	5246-5314 (69)	5392-5460(69)	5279-5347(69)	-	TGC	
tRNA-N	5356-5428 (73)	5434-5056 (73)	5316-5389 (74)	5462-5534(73)	5349-5421(73)	-	GTT	
tRNA-C	5452-5519(68)	5540-5606 (67)	5425-5489 (65)	5573-5638(66)	5458-5522(65)	-	GCA	
tRNA-Y	5521-5590(70)	5609-5678 (70)	5490-5559 (70)	5640-5709(70)	5592-5523(70)	-	GTA	
Citocromo oxidasa subunidad I (COI)	5595-7133 (1539)	5689-7221 (1533)	5561-7111 (1550)	5711-7273(1563)	5594-7156(1563)	+	ATG ^(a) ATC ^(b, d, e) GTG ^(c)	TAA ^(a, b, c) CAA ^(d, e)
tRNA-S	7159-7229 (71)	7247-7317 (71)	7112-7182 (71)	7262-7332(71)	7148-7218(71)	+	GCT	
tRNA-D	7233-7303 (71)	7321-7391 (71)	7187-7253 (67)	7336-7407(72)	7222-7290(69)	+	GTC	
Citocromo oxidasa subunidad II (COII)	7322-8005 (684)	7410-8093 (684)	7262-7956 (695)	7413-8103(691)	7298-7988(691)	+	ATG ^(a, b, c, d, e)	CTT ^(a, b) CCT ^(c) GAG ^(d) GCC ^(e)
tRNA-K	8013-8087 (75)	8101-8175 (75)	7957-8031 (74)	8104-8178(75)	7989-8063(75)	+	TTT	
ATP sintasa subunidad 8 (ATP8)	8089-8271 (183)	8177-8338 (162)	8034-8201 (164)	8184-8351(168)	8069-8236(168)	+	ATG ^(a, b, c, d, e)	TGG ^(a, b) TAA ^(c) TAC ^(d) CAC ^(e)
ATP sintasa subunidad 6 (ATP6)	8268-8945 (678)	8335-9012 (678)	8192-8875 (683)	8342-9024(683)	8227-8909(683)	+	ATG ^(a, b, c, d, e)	AAC ^(a) AAT ^(b, c, d, e) GTC ^(a) TCA ^(b)
Citocromo oxidasa subunidad III (COIII)	8985-9767 (783)	9062-9844 (783)	8875-9659 (784)	9025-9809(785)	8910-9695(786)	+	ATG ^(a, b, c, d, e)	TCC ^(a, d) TCT ^(b) ATA ^(c) GCC ^(e)
tRNA-G	9794-9864 (71)	9866-9936 (71)	9660-9731 (71)	9810-9880(71)	9702-9771(70)	+	TCC	
NADH deshidrogenasa subunidad 3 (ND3)	9865-10209 (345)	9937-10281 (375)	9732-10080 (348)	9881-10229(349)	9772-10120(349)	+	ATG ^(a, c, d, e) ATA ^(b)	GAA ^(a, b) AGT ^(c) GAA ^(d, e)
tRNA-R	10211-10279 (69)	10283-10351 (69)	10081-10149 (68)	10230-10298(69)	10121-10189(69)	+	TCG	
NADH deshidrogenasa subunidad 4L(ND4L)	10283-10576 (294)	10352-10645 (294)	10919-11215 (297)	10299-10595(367)	10190-10486(297)	+	ATA ^(a) ATG ^(b, d, e) GTG ^(c)	TGC ^(a, b, d, e) TAA ^(c)
NADH deshidrogenasa subunidad 4 (ND4)	10573-11937 (1365)	10642-12015 (1374)	10441-11821 (1380)	10589-11969(1381)	10480-11860(1381)	+	ATG ^(a, b, c, d, e)	GTT ^(a) GGG ^(b) TCT ^(c) TGA ^(d) GGA ^(e)
tRNA-H	11954-12023 (70)	12026-12095 (70)	11822-11890 (68)	11970-12038(69)	11861-11930(70)	+	CTG	
tRNA-S	12024-12098 (75)	12096-12167 (72)	11891-11958 (67)	12039-12106(68)	11931-11998(68)	+	TGA	
tRNA-L	12104-12176 (73)	12173-12245 (73)	11963-12035 (72)	12111-12183(73)	12003-12075(73)	+	TAG	
NADH deshidrogenasa subunidad 5 (ND5)	12195-14003 (1809)	12261-14072 (1812)	12036-13883 (1847)	12184-14022(1839)	12076-13914(1839)	+	ATA ^(a) ATC ^(b, e) ATG ^(c) ATT ^(d)	CTC ^(a, d) CTT ^(b) TAA ^(c) ACG ^(e)
NADH deshidrogenasa subunidad 6 (ND6)	14019-14537 (519)	14093-14611 (519)	13880-14412 (532)	14019-14540(522)	13911-14432(522)	-	ATG ^(a, b, c, d, e)	GTC ^(a) GTT ^(b, d, e) TAG ^(c)
tRNA-E	14538-14606 (69)	14612-14680 (69)	14411-14479 (68)	14541-14609(69)	14433-14501(69)	-	TTC	
Citocromo b (Cytb)	14609-15739 (1131)	14683-15813 (1131)	14484-15624 (1140)	14613-15753(1141)	14506-15646(1141)	+	ATG ^(a, b, c, d, e)	CTT ^(a, b) CTC ^(b) CAT ^(c) CTA ^(d)
tRNA-T	15753-15824 (72)	15827-15898 (72)	15628-15696 (68)	15754-15826(73)	15647-15717(71)	+	TGT	
tRNA-P	15824-15893 (70)	15898-15968 (71)	15701-15767 (66)	15826-15898(73)	15718-15788(71)	-	TGG	
D-Loop	15893-16656 (763)	15968-16665 (697)	15767-16635 (868)	15899-16707(809)	15789-16620(832)	+		

3.4.1 Registro de los genomas mitocondriales en GenBank

En este proceso se siguieron las instrucciones del GenBank para poder someter la información de los genomas mitocondriales obtenidos mediante el ensamblaje *de Novo*. Posteriormente, se necesitó la información de las anotaciones realizadas, usar distintos programas para este proceso nos ayudó a respaldar la información registrada en la base de datos.

Verificamos la longitud y anticodones de los ARNt obtenidos usando tRNAscan-SE, mientras que para los ARNr 12s y 16s se observó la longitud de estas regiones con el CLC Genomics Workbench 10.1 y se realizó un Blastn. Con el Blastn pudimos observar que estas regiones tenían un alto porcentaje de similitud con las secuencias previamente registradas de *C. viridis* y *C. nigrescens*, la información de los CDS obtenidas mediante MITOS y Mitofish se verificó en Expsy para observar los codones de paro. Una vez que todas las anotaciones fueron aceptadas por GenBank se les asignó los números de acceso a *C. viridis* (**OR068388**) y *C. nigrescens* (**OR068389**) Se realizó un alineamiento con las especies marcadas en la tabla 8 los resultados pueden observarse en la Fig.11.

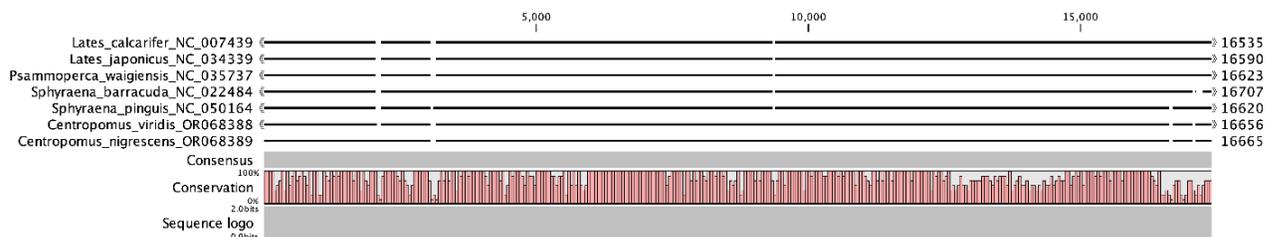


Figura 11 Alineamiento de los genomas mitocondriales de *Centropomus viridis*, *Centropomus nigrescens*, *Lates calcarifer*, *Sphyaena barracuda* y *Sphyaena pinguis* realizado con el programa CLC Genomics Workbench 10.1

3.6 Análisis filogenético

Los modelos evolutivos describen la probabilidad de una sustitución de nucleótidos homóloga a lo largo del tiempo. Mediante el programa Jmodel (Santorumet *etal.*, 2014) (<https://github.com/ddarriba/jmodeltest2>) se obtuvo el mejor modelo de análisis filogenético. Para seleccionar el modelo se utilizan los criterios de información de Akaike (AIC), dicho modelo nos proporciona dimensiones altas aportando una evaluación de un intervalo continuo correspondiente a la

probabilidad precisa y determinada de los hechos mientras que el criterio de información Bayesiana (BIC) aporta dimensiones finitas y consistentes. Para seleccionar el mejor modelo de AIC la probabilidad debe ser menor que 1 (Wang & Liu, 2006).

El mejor modelo seleccionado fue AIC = GTR+I+G y BIC= GTR+I+G, donde GTR se refiere al modelo de sustitución nucleotídica reversible en el tiempo, I se estima para la configuración de sitios variables y G es el uso de una función gamma. En los resultados de este análisis el modelo fue el mismo seleccionado por ambos métodos.

Con base a los resultados de Jmodel se obtuvo el árbol filogenético en el programa Mr. Bayes (Huelsenbeck & Ronquist, 2001) ([MrBayes | index \(nbisweden.github.io\)](https://github.com/nbisweden/mrbayes3)) , a partir del cual se obtuvo la imagen en formato png con el programa FigTree (Rambaut,2018)([FigTree \(ed.ac.uk\)](https://www.igb.berkeley.edu/~rambaut/figtree/)) y se editó con el programa PowerPoint (Fig. 12).

En la imagen se puede observar a *L. calcarifer* agrupado en un clado distinto a los robalos marcados con flechas rojas, lo que sugiere que pertenecen a una familia distinta, como se encuentra reportado en WoRMS (Ahyong *et al.*, 2023) mientras que la familia más cercana a *C. viridis* y *C. nigrescens* es la familia *Sphyraenidae* de acuerdo con los resultados obtenidos.

Con el propósito de dar más soporte a la información de la filogenia, se respaldó esta información realizando un árbol filogenético con los fragmentos del 18S ARNr, los fragmentos para *C. viridis* y *C. nigrescens* se obtuvieron debido a la secuenciación masiva realizada y se registraron en el Genbank con el siguiente número (2734175) Hacer este análisis solamente con las especies del orden Carangaria fue complicado debido a la falta de información en el Genbank de algunas de las especies seleccionadas.

Por lo tanto, con la secuenciación parcial de *C. viridis* y *C. nigrescens* identificamos los contigs que correspondían a esta región debido a que estas secuencias aún no se encontraban en el GenBank. Los contigs que fueron seleccionados (248 para *C. viridis*) y (el 15580 para *C. nigrescens*) se sometieron a un Blast para ver con que especies coincidían dichos fragmentos.

El árbol filogenético con la información de genomas mitocondriales completos se encuentra en la figura 12 y el árbol filogenético con las secuencias del gen 18S ARNr se encuentra en la figura 13 los datos taxonómicos se revisaron en WoRMS (Ahyong *et al.*, 2023)

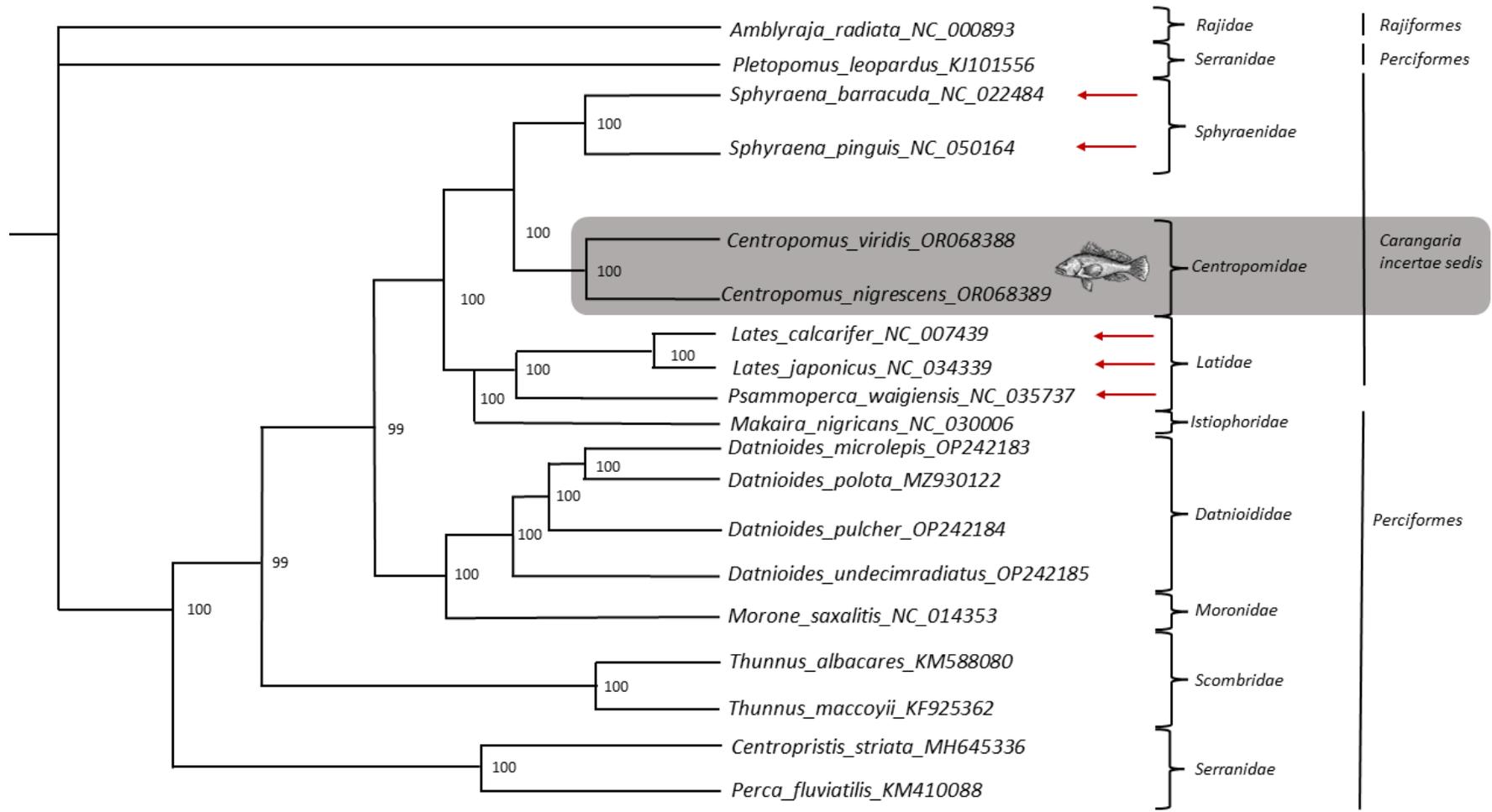


Figura 12 Árbol filogenético realizado con información de genomas mitocondriales completos

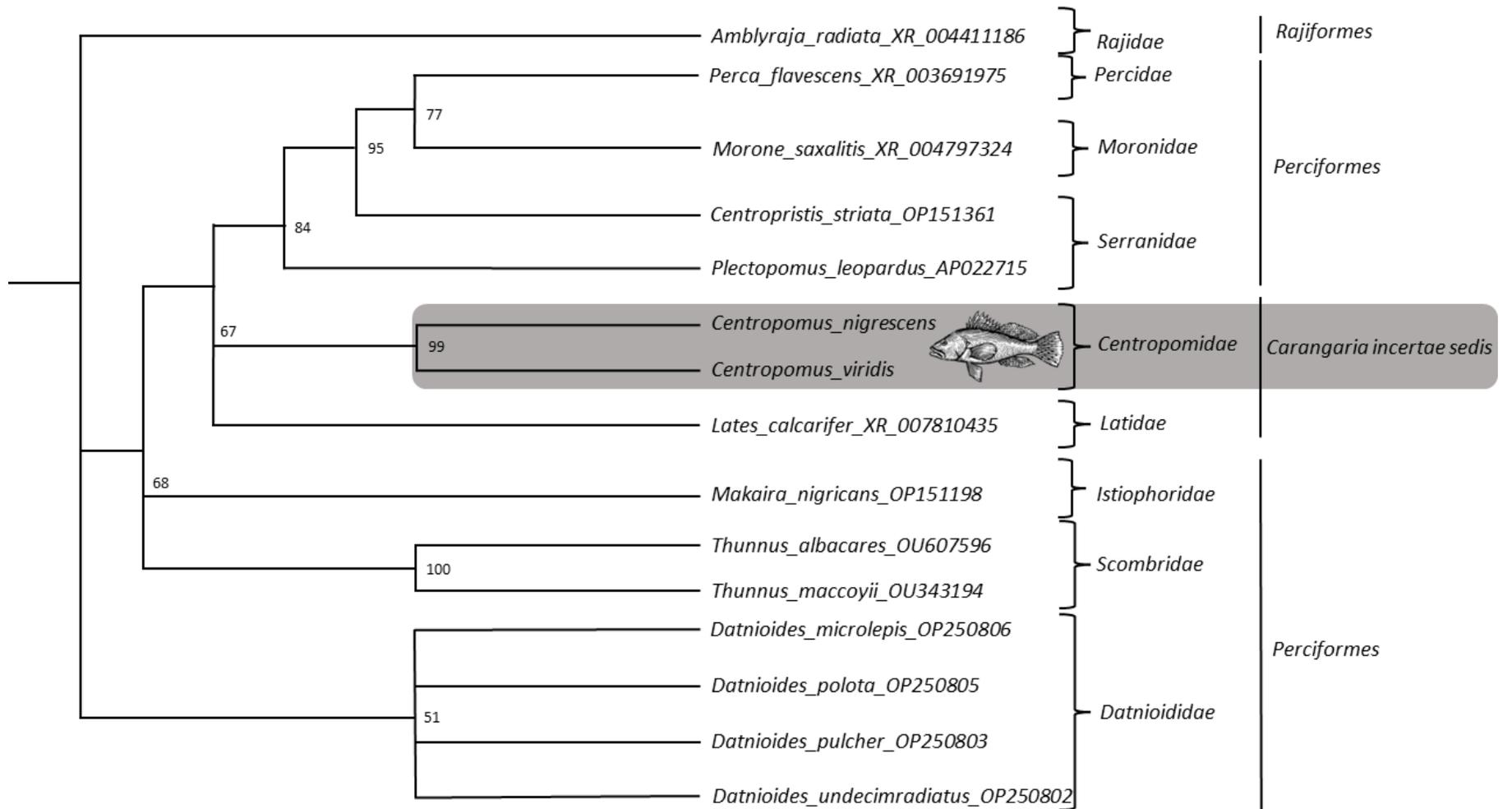


Figura 13 Árbol filogenético realizado con secuencias del gen 18S ARNr

3.5 Blastn y Blastx

De los 439,105 contigs obtenidos con *C. viridis* se realizó un Blastn y se obtuvo como resultado que 331,412 (75.47%) secuencias que tuvieron anotación a secuencias nucleotídicas, muchas de ellas pertenecían a especies de peces como la carpa común (*Cyprinus carpio*). El Blastn de *C. nigrescens* dio como resultados 354,148 anotaciones nucleotídicas y así como en *C. viridis* muchas de ellas pertenecían a especies como pez Cebra (*Danio rerio*) y carpa común.

El Blastn nos ayudó a identificar regiones como la secuencia del 18S ARNr para cada una de estas especies las cuales fueron utilizadas para realizar un árbol filogenético con especies del orden Carangaria. Los genes de interés como el 18S ARNr y genes de mantenimiento fueron identificados y obtenidos mediante el Blastx, esta herramienta nos permitió conocer las secuencias de nucleótidos que se traducían a proteína para *C. viridis* tuvimos un total de 1304 datos y para *C. nigrescens* un total de 47964 datos. Las tablas con los genes de interés se encuentran en el Anexo I con el título identificación de genes de interés mediante Blastx.

3.6 Ontología genética (GO)

Las ontologías genéticas (GO por sus siglas en inglés) proporcionan un lenguaje sistemático que describe productos génicos a través de tres dominios conocidos como proceso biológico, función biológica y componente celular. GO también facilita la relación de grandes datos, como en este caso los obtenidos de la secuenciación parcial de *C. viridis* y *C. nigrescens*. Tener este análisis nos permitió conocer las funciones de algunos genes en la secuenciación parcial.

3.6.1 Ontología genética (GO) para *Centropomus viridis* y *Centropomus nigrescens*

La anotación con la ontología genética (GO por sus siglas en inglés) permitió saber que genes tenían asociación a procesos biológicos "P", funciones biológicas "F" y componente celular "C". Los resultados obtenidos con el Blastx realizado con los datos del ensamblaje de CLC para la especie *Centropomus viridis* donde solo se tomaron para el análisis las secuencias que tenían ontología genética (GO por sus siglas en inglés) quedando solo 709 datos en todo el conjunto.

De estos 709 datos de anotación genética 575 (81.1%) genes pertenecían a la conjunción de proceso biológico, componente celular y función biológica, mientras que 21 (3.0%) datos de anotación genética pertenecían a la conjunción de proceso y función biológica y 29 (4.1%) datos a los procesos de componente celular y función biológica.

Las secuencias de genes que pertenecen a proceso biológico y componente celular son el 62(8.7%), de las anotaciones que solo corresponden a proceso biológico son 2 (0.3%), componente celular 9 (1.3%) y función biológica 11 (1.6%).

En la figura 14 se muestra en Diagramas de Venn los resultados de la anotación GO de la secuenciación parcial de *C. viridis*.

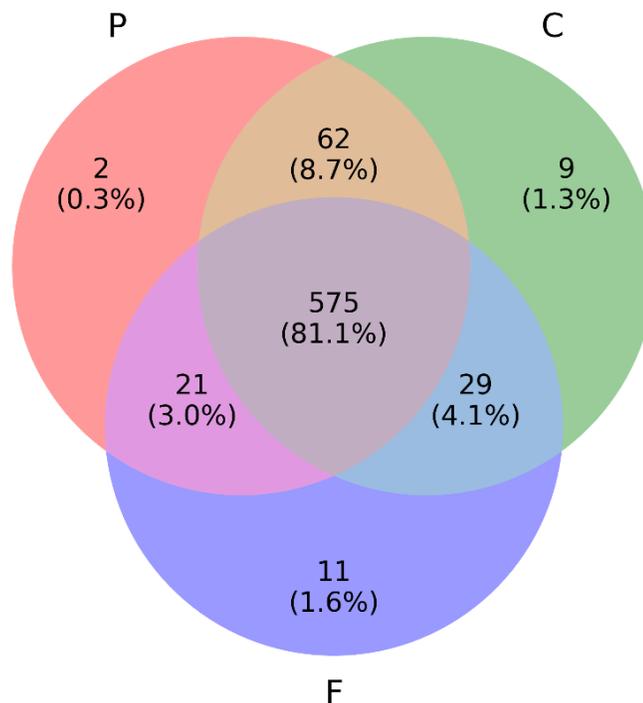


Figura 14 Diagrama de Venn para la ontología genética (GO) de *Centropomus viridis*.

Para *Centropomus nigrescens* los resultados de ontología genética nos dieron 13,822 datos. De los cuales, 10,991 (79.5%) estaban involucrados en la intersección de los procesos de función biológica, componente

celular y proceso biológico, mientras que 1,276 (9.2%) datos de anotación genética estaban involucrados en proceso biológico y componente celular.

En los procesos de función biológica y proceso celular se encontraban 299 (2.2%) anotaciones de genes y en función biológica y componente celular 661 (4.8%) anotaciones.

En el conjunto P que corresponde a proceso biológico solo se encontraron 61 (0.4%) anotaciones de genes, mientras que en el conjunto C que corresponde a componente celular 372 (2.7%) y en el conjunto F que es de función biológica 162 (1.2%) anotaciones de genes. En la figura 15 se muestra en Diagramas de Venn los resultados de la anotación GO de la secuenciación parcial de *C. nigrescens*.

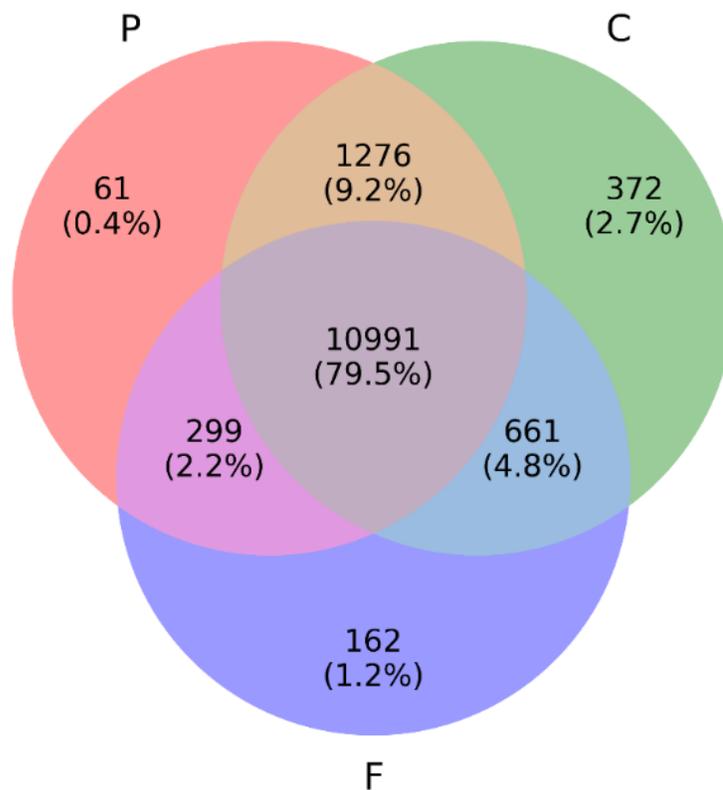


Figura 15 Diagrama de Venn para la ontología genética (GO) de *Centropomus nigrescens*

3.7 Evaluación del número de microsatélites

Se obtuvo una serie de datos de los microsatélites mediante KRAIT realizando una comparación con el genoma completo de *L. calcarifer* a pesar de que en los robalos se presentan datos de una secuenciación

parcial, no existe mucha diferencia en porcentajes de la longitud de los microsatélites.

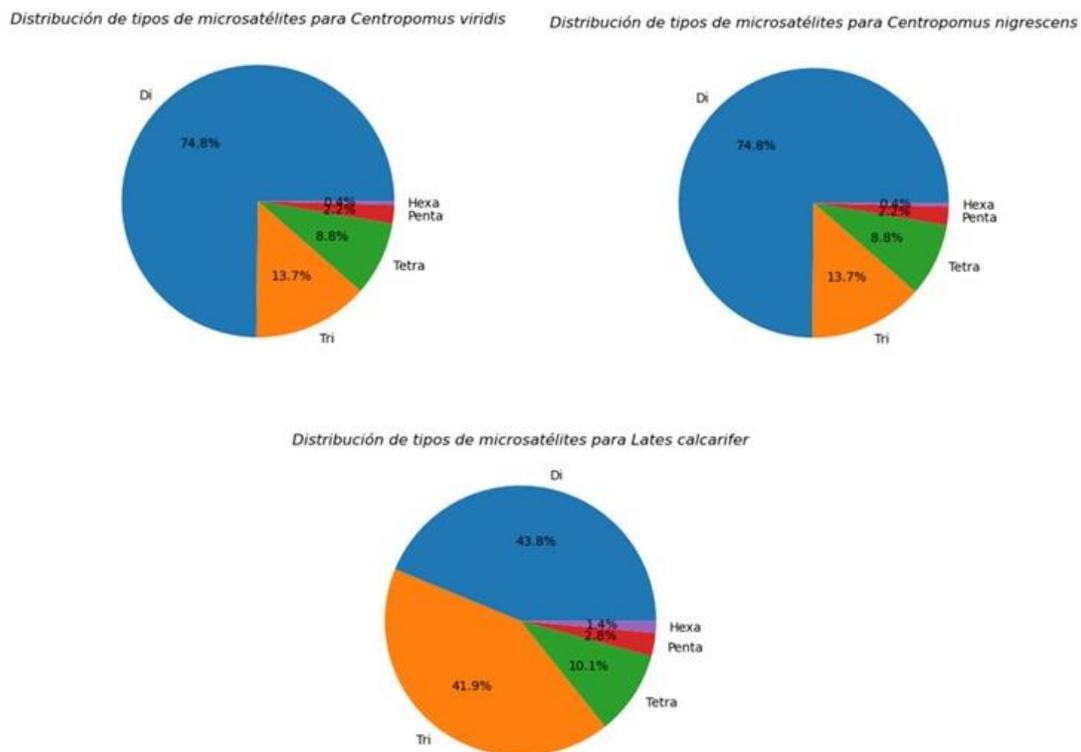


Figura 16 Estadísticas generadas por KRAIT para longitud de microsatélites para-*C. viridis*, *C. nigrescens* y *Lates calcarifer*.

Capítulo 4. Discusión

4.1 Extracción y calidad de las muestras de ADN

Existen estudios que demuestran la efectividad del protocolo de sales para realizar extracción de ácidos nucleicos en aletas de peces clasificándolo como un método económico, simple y reproducible (Lopera-Barrero *et al.*, 2008).

Este método nos permitió obtener altas concentraciones y calidad de ADN para cada uno de los organismos, dando resultados favorables para llevar a cabo el proceso de secuenciación de segunda generación. Esta técnica proporcionó muchas ventajas a la hora de realizar la extracción, como se sugirió en el artículo de (Wen *et al.*, 2000), no requirió de equipos y reactivos costosos y que afectaran al ambiente.

Se ha demostrado en otros estudios de peces que realizar extracciones de la aleta es un método ideal y no destructivo debido a que no se requiere del sacrificio del animal y por lo tanto no hay pérdidas económicas (Nunes-Pazdiora *et al.*, 2016). El estudio de (Martínez-Brown *et al.*, 2021) también mostró originalmente resultados favorables en la extracción de ADN de tejido de aletas para la amplificación de un solo gen 16S ARNr. Sin embargo, la cantidad de concentración de ADN obtenida depende del tipo de tejido y el tipo de método de extracción que se esté utilizando.

En este estudio no se realizó comparación con otros métodos debido a que es un procedimiento estándar del LAGA. se realizó el método de sales, lo cual permitió obtener mucha cantidad de ácidos nucleicos en branquias, pero no en aletas, aun así, los resultados fueron aceptables para la secuenciación masiva, lo cual nos indica que se podría utilizar este tejido para estudios posteriores.

4.2 Análisis bioinformático

4.2.1 Calidad de las secuencias y ensamblaje *de Novo*

La evaluación de la calidad del total de las lecturas obtenidas es muy importante al momento de realizar

un análisis bioinformático debido a que esto nos ayuda depurar las regiones de baja calidad de toda la información que proviene de la secuenciación masiva.

La principal herramienta para evaluar calidad de lecturas es programa FastQC en diversos estudios de análisis genómico y transcriptómico, por ejemplo, especies como el salmón (*Salmo salar*, Linnaeus,1758) (Boison *et al.*, 2019), el bagre amarillo (*Pelteobagrus fulvidraco*, Richardson,1846) (Luo *et al.*, 2022) entre otras.

La ventaja de FastQC es que es gratuito y además proporciona una serie de estadísticas y gráficos para un archivo fastq (Ward *et al.*, 2020). Sin embargo, no es la única herramienta que hace este tipo de análisis. El CLC Genomics Workbench 10.1 es un conjunto de procedimientos que incluye el proceso de filtrado, eliminación de adaptadores y ensamblaje en un solo programa, aunque su precio puede no ser accesible a muchos estudiantes o investigadores.

El propósito de hacer el ensamblaje *de Novo* con distintas herramientas fue hacer una comparación entre los distintos programas y evaluar la calidad de ensamblaje para obtener en un solo contig el genoma mitocondrial de cada organismo.

Aunque Trinity sea un ensamblador para transcriptomas existen trabajos realizados en peces como trucha arcoíris en los cuales contribuyó al descubrimiento de genes a través de la secuenciación del transcriptoma, al desarrollo de marcadores moleculares y a la identificación del polimorfismo de un solo nucleótido (SNPs)(Salem *et al.*, 2015).

El resultado obtenido mediante Trinity ayudó a hacer una comparación del resto de los genes identificados con CLC, en cuanto a NOVOPlasty se considera necesario que para obtener el genoma mitocondrial completo en algunas especies se ajuste la configuración del número de k-meros. Sin embargo, no fue posible obtener el mitogenoma de los robalos con este programa.

El CLC Genomics Workbench (con diferentes versiones) se ha utilizado exitosamente para la obtención de genomas mitocondriales de peces marinos (Del Río-Portilla *et al.*, 2016, Galván-Tirado *et al.*, 2016, Vargas-Peralta *et al.*, 2020) y peces de agua dulce (Barriga-Sosa *et al.*, 2016) (Del Río-Portilla *et al.*, 2016), elasmobranquios(Castillo-Páez *et al.*, 2014) y moluscos (Magallón-Gayón *et al.*, 2019)

4.3 Anotación y registro del genoma mitocondrial

La identificación y conformación de los genomas mitocondriales pueden ayudar a establecer relaciones filogenéticas precisas, entre otras cuestiones mencionadas a lo largo de este documento. (Prabhu *et al.*, 2020) sugieren que realizar análisis con los datos del genoma mitocondrial completo proporcionan una mejor comprensión para análisis filogenéticos que las investigaciones elaboradas con algunos genes. Haber obtenido la secuencia completa en un solo contig ayudará a realizar otras investigaciones y minimizar costos.

Estudios enfocados en mitogenomas sugieren que aunque el orden y estructura de los genes puede variar en algunos casos, en la mayoría de las especies esta información se encuentra muy bien conservada (Formenti *et al.*, 2021). El orden y estructura de los genes totales de los mitogenomas obtenidos en este proyecto coincidió con el de la mayoría de los trabajos realizados en vertebrados. El utilizar distintos anotadores es indispensable, ya que estos programas están optimizados para algunas especies y pueden presentar diferencias entre programas.

Por ejemplo: la plataforma MITOS no arrojó los CDS (coding sequencing por sus siglas en inglés y se refiere a la región codificante) con el codón de paro, mientras que Mitofish nos proporcionó anotaciones precisas. Por lo que es indispensable corroborar la información de programas de anotación automáticos.

Complementar esta tarea con tRNAscan-SE, nos permitió obtener la estructura secundaria y anticodones de los ARNt. Aunque, tRNAscan-SE no fue diseñado para identificar ARNt en orgánulos, se ha demostrado que da buenos resultados en la anotación de mitocondrias de mamíferos y vertebrados (Chan & Lowe, 2019), la anotación con MITOS nos permitió verificar que los datos presentados con este programa eran fiables.

La base de datos NCBI posee un intercambio diario de datos con el Archivo Europeo de Nucleótidos (ENA) y el Banco de datos de ADN de Japón (DDBJ) garantizando una cobertura mundial (Sayers *et al.*, 2023), subir la información generada a esta plataforma es de gran aporte debido a que esta plataforma es de acceso gratuito y cualquier persona puede acceder a ella. Por ello la importancia de someter las secuencias al GenBank del NCBI.

4.4 Análisis filogenéticos

La clasificación taxonómica de los peces ha ido cambiando a lo largo del tiempo, esto debido a que la información molecular puede complementar y en muchos casos respaldar lo establecido anteriormente. En cambio, a pesar de todo esto, algunos órdenes como el caso *Carangaria incertae sedis* siguen sin ser resueltos (Betancur *et al.*, 2017).

Inicialmente, cuando se hizo la revisión para los antecedentes, se observaron 12 mitogenomas de especies dentro de la familia Centropomidae *Lates calcarifer* (NC_007439), *Lates japonicus* (NC_034339), *Lates niloticus* (NC_035736) y *Psammoperca waigiensis* (NC_035737), por lo que los mitogenomas de los robalos de este estudio aumentarían este número para la familia. Sin embargo, al realizar el análisis filogenético, se observaron discrepancias, por lo que se profundizó en las relaciones de las especies dentro de la familia. Se observó que en las clasificaciones los géneros *Lates* y *Psammoperca* (WoRMS y Fishbase) están dentro de la familia Latidae y no en Centropomidae.

Lo que hace que los mitogenomas obtenidos en este estudio sean los primeros del género y de la familia. Antes de este registro, no habían mitogenomas completos de la familia Centropomidae y aunque en NCBI aparezcan los genomas mitocondriales completos de las especies como parte de la familia Centropomidae, esta institución no es considerada una autoridad taxonómica. Lo que obliga a realizar búsqueda de relaciones taxonómicas en bases de datos especializadas.

Se han realizado estudios con genes mitocondriales para otras familias de peces como la *Haemulidae*. Sin embargo, en este estudio (Sanciango *et al.*, 2011) como en el de (Lin *et al.*, 2006) solo utilizaron algunos CDS como COI y Citocromo B para realizar el análisis filogenético. (Sanciango *et al.*, 2011) menciona que, a pesar de los análisis moleculares, sigue siendo muy importante verificar las características morfológicas.

A medida que crecen las tecnologías de secuenciación masiva, la confianza en los análisis de inferencia Bayesiana y de probabilidad tomarán cada vez más relevancia (Flanagan & Jones, 2019). De igual manera, se han hecho análisis “filogenómicos” basados en la concatenación de muchos nucleótidos de genes que codifican a proteínas (PCG) como en la investigación de (Colín *et al.*, 2023).

No obstante, esto no significa que sea la única opción a la hora de realizar árboles filogenéticos mediante inferencia Bayesiana. Los resultados de este estudio en robalos han demostrado que el genoma

mitocondrial completo no causa alguna interferencia en el análisis, proporcionando resultados fiables a la hora de comparar taxones. También contribuye a la propuesta del orden Carangaria *incertae sedis*.

4.5 Ontología genética

La Ontología genética se ha utilizado para la predicción de la función genética y el análisis de vías celulares (Harris *et al.*, 2006). Esta herramienta nos fue útil para comprender el resto de los genes obtenidos mediante la secuenciación masiva, facilitando la integración de esta información con la de otros organismos que ya habían sido registrados.

La falta de recursos genéticos abundantes dificulta comprender muchos procesos biológicos y celulares implicados en el desarrollo de los organismos, afortunadamente la NSG está ayudando a resolver esta problemática. Uno de los objetivos de este proyecto fue identificar genes que codifican a proteínas de choque térmico en la secuenciación parcial, estas proteínas son relevantes debido a que tienen funciones que son esenciales en aspectos metabólicos de otras proteínas (Morimoto, 1998).

Gracias a investigaciones realizadas anteriormente y a los datos registrados en GO podemos conocer que la “heat shock protein” 70 (HSP70) ayuda al plegamiento de algunas cadenas polipeptídicas, actúa como mediadora en la reparación y degradación de proteínas alteradas o desnaturalizadas (Kiang & Tsokos, 1998). Mientras que HSP90 activa el soporte de varios componentes del citoesqueleto, los genes que codifican a estas proteínas fueron identificados en este trabajo y podrían ayudar a seleccionar reproductores. Como producto de esta investigación, muy probablemente se tenga la secuencia de estas proteínas en *C. viridis* o *C. nigrescens*

En peces estas proteínas son un punto clave para resolver el estrés celular, la mayoría de estos estudios demuestran una relación entre el aumento de los niveles de expresión en las HSP y la exposición a factores estresantes dentro de un rango ecológicamente relevante, lo que sugiere que la respuesta de estas proteínas desempeña algún papel en la mejora y la supervivencia (Basu *et al.*, 2002).

El poder de los estudios Omicos es importante para lograr una revolución azul y combatir problemas alimentarios mundiales (Pandey & Parhi, 2022). Todo este proceso de análisis, desarrollo y de obtener información ocurre gracias a la combinación con la bioinformática, brindando nuevas oportunidades para mejorar el rendimiento de la acuicultura obteniendo y comprendiendo información de genes que pueden

contribuir a proporcionar características económicamente relevante a los organismos que se utilizan para sistemas de cultivo (Rather *et al.*, 2023).

Es indispensable el uso de estas tecnologías para afrontar algunos desafíos como lo son la baja tasa de producción, pérdida de hábitat y de acervo genético. Tener este tipo de conocimientos nos ayuda a aportar un recurso para lograr mejores prácticas de pesca y acuicultura sostenibles (Bernatchez *et al.*, 2017).

4.6 Microsatélites

La información de genes mitocondriales es de mucha ayuda para evaluar la estructuración genética de varios organismos y su variabilidad, y son herramientas que ayudan a evaluar la diversidad genética, a elaborar mapas genéticos y a estudios de genómica comparativa. Por lo tanto, estos marcadores moleculares aportan un valor significativo en la investigación (Wattanadilokchatkun *et al.*, 2022).

Con la información obtenida mediante la secuenciación masiva, se pudieron identificar microsatélites que puedan ser evaluados posteriormente, ya que no es tan costoso como antes encontrar microsatélites. A pesar de que existan nuevos marcadores como los SNP, los microsatélites seguirán utilizándose debido a que son considerados superiores en cuanto a temas de análisis de parentesco en poblaciones naturales o cultivadas, la evaluación de androgénesis y Ginogénesis (Wenne, 2023).

Además, los posibles microsatélites identificados en este estudio pueden proporcionar ayuda para caracterización genética de linajes en especies de robalo contribuyendo a obtener información para elaborar sistemas de cultivo. Los análisis de parentesco en la acuicultura han proporcionado datos fundamentales sobre el comportamiento, la ecología y la evolución de los peces (Flanagan & Jones, 2019).

Existen estudios que demuestran que los SNP son más eficaces que los microsatélites, sin embargo el desarrollo creciente de programas y protocolos basados en secuencias para la genotipificación ha aumentado considerablemente el número de alelos que puede analizarse en comparación con los SNP (Zhan *et al.*, 2017).

Capítulo 5. Conclusiones

Se lograron obtener los genomas mitocondriales completos en un solo contig de *Centropomus viridis* y *Centropomus nigrescens*.

Se necesita obtener información genética para ayudar a actualizar las clasificaciones taxonómicas.

La secuenciación masiva en organismos no modelo ayuda a obtener información valiosa para mejorar las propuestas realizadas en sistemas de cultivo.

Es de gran importancia expandir los estudios mitogenómicos para especies de interés acuicultural y en especial del orden Carangaria para resolver las posiciones taxonómicas de las especies que en él se encuentran.

Se propone seguir con investigaciones que aporten información de otros genes de interés a GenBank para facilitar los estudios genéticos posteriores y con la información obtenida ayudar a una gestión pesquera más eficaz

Se necesita conocer la diversidad genética para poder ayudar a conservarla.

Con la secuenciación parcial se lograron identificar genes que codifican a proteínas de choque térmico y genes de mantenimiento.

También se logró conocer y obtenerla región 18S ARNr de *Centropomus viridis* y *Centropomus nigrescens*.

Generar información para la acuicultura es relevante, puesto que contribuye a tener una acuicultura sostenible.

Se requiere trabajar en conjunto para proveer de alimento de calidad a la población.

Literatura citada

- Abdelrahman, H., ElHady, M., Alcivar-Warren, A., Allen, S., Al-Tobasei, R., Bao, L., Beck, B., Blackburn, H., Bosworth, B., Buchanan, J., Chappell, J., Daniels, W., Dong, S., Dunham, R., Durland, E., Elaswad, A., Gomez-Chiarri, M., Gosh, K., Guo, X., ... Zhou, T. (2017). Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genomics*, *18*, 191. <https://doi.org/10.1186/s12864-017-3557-1>
- Ahyong, S., Boyko, C. B., Bailly, N., Bernot, J., Bieler, R., Brandão, S. N., Daly, M., De Grave, S., Gofas, S., Hernandez, F., Hughes, L., Neubauer, T. A., Paulay, G., Boydens, B., Decock, W., Dekeyzer, S., Vandepitte, L., Vanhoorne, B., Adlard, R., ... Zullini, A. (2023). World Register of Marine Species (WoRMS). WoRMS Editorial Board. WoRMS - World Register of Marine Species
- Alvarez-Lajonchère, L., & Tsuzuki, M. Y. (2008). A review of methods for *Centropomus* spp. (snooks) aquaculture and recommendations for the establishment of their culture in Latin America. *Aquaculture Research*, *39*(7), 684–700. <https://doi.org/10.1111/j.1365-2109.2008.01921.x>
- Arce-Valdés, L. R., Abadía-Cardoso, A., Arteaga, M. C., Peñaranda-Gonzalez, L. V., Ruiz-Campos, G., & Enríquez-Paredes, L. M. (2023). No effects of fishery collapse on the genetic diversity of the Gulf of California Corvina, *Cynoscion othonopterus* (Perciformes: Sciaenidae). *Fisheries Research*, *261*(December 2022), 106608. <https://doi.org/10.1016/j.fishres.2023.106608>
- Baeza, J. A. (2018). The complete mitochondrial genome of the Caribbean spiny lobster *Panulirus argus*. *Scientific Reports*, *8*(1), 17690. <https://doi.org/10.1038/s41598-018-36132-6>
- Barriga-Sosa, I. de L. A., García De León, F. J., & Del Río-Portilla, M. A. (2016). The complete mitochondrial DNA of the endemic shortfin silverside, *Chirostoma humboldtianum* (Valenciennes, 1835). *Mitochondrial DNA*, *27*(2), 1545–1546. <https://doi.org/10.3109/19401736.2014.953130>
- Basu, N., Todgham, A. E., Ackerman, P. A., Bibeau, M. R., Nakano, K., Schulte, P. M., & Iwama, G. K. (2002). Heat shock protein genes and their functional significance in fish. *Gene*, *295*(2), 173–183. [https://doi.org/10.1016/S0378-1119\(02\)00687-X](https://doi.org/10.1016/S0378-1119(02)00687-X)
- Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., Maes, G. E., Martinsohn, J. T., Miller, K. M., Naish, K. A., Ovenden, J. R., Primmer, C. R., Young Suk, H., Therkildsen, N. O., & Withler, R. E. (2017). Harnessing the Power of Genomics to Secure the Future of Seafood. *Trends in Ecology & Evolution*, *32*(9), 665–680. <https://doi.org/10.1016/j.tree.2017.06.010>
- Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., & Ortí, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, *17*, 162. <https://doi.org/10.1186/s12862-017-0958-3>
- Biotium. (2009). GelRed Nucleic Acid Gel Stain, 10,000X in Water. *Biotium*, 1–2. <https://biotium.com/wp-content/uploads/2013/11/PI-41001.pdf>
- Boison, S. A., Gjerde, B., Hillestad, B., Makvandi-Nejad, S., & Moghadam, H. K. (2019). Genomic and Transcriptomic Analysis of Amoebic Gill Disease Resistance in Atlantic Salmon (*Salmo salar* L.). *Frontiers in Genetics*, *10*(68), 1–15. <https://doi.org/10.3389/fgene.2019.00068>

- Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, 27(8), 1767–1780. <https://doi.org/10.1093/nar/27.8.1767>
- Castillo-Páez, A., del Río-Portilla, M. A., Oñate-González, E., & Rocha-Olivares, A. (2014). The mitochondrial genome of the banded guitarfish, *Zapteryx exasperata* (Jordan and Gilbert, 1880), possesses a non-coding duplication remnant region. *Mitochondrial DNA*, 27(3), 1–3. <https://doi.org/10.3109/19401736.2014.958721>
- Chan, P. P., & Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In *Methods in Molecular Biology*. https://link.springer.com/protocol/10.1007/978-1-4939-9173-0_1
- Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, 255(1–4), 1–29. <https://doi.org/10.1016/j.aquaculture.2005.11.031>
- CIBNOR. (1994). *Desarrollo Científico y Tecnológico del Cultivo de Robalo*. Recuperado el 25 de agosto del 2022 de [https://cibnor.repositorioinstitucional.mx/jspui/bitstream/1001/1858/1/Desarrollo Científico y Tecnológico del Cultivo de Robalo_1994.pdf](https://cibnor.repositorioinstitucional.mx/jspui/bitstream/1001/1858/1/Desarrollo_Científico_y_Tecnológico_del_Cultivo_de_Robalo_1994.pdf)
- Colín, A., Del Río-Portilla, M. A., Lafarga-De la Cruz, F., Ingle-De la Mora, G., & García-De León, F. J. (2023). Assembly, Characterization, and Phylogenetic Relationships of Mitogenomes of Two Species of Mexican Trout (*Oncorhynchus chrysogaster* and *O. mykiss nelsoni*). *Fishes*, 8(4), 178. <https://doi.org/10.3390/fishes8040178>
- CONAPESCA. (2018). *Anuario estadístico de acuacultura y pesca*. Recuperado el 18 de agosto del 2022 de <https://www.gob.mx/conapesca/documentos/anuario-estadistico-de-acuacultura-y-pesca>
- CONAPESCA. (2020). *Anuario estadístico de acuacultura y pesca 2020*. Recuperado el 18 de agosto del 2022 de <https://www.gob.mx/conapesca/documentos/anuario-estadistico-de-acuacultura-y-pesca>
- CONAPESCA. (2021). *Anuario Estadístico de Acuacultura y Pesca*. 44(8), 085201. Recuperado el 18 de agosto del 2022 de <https://www.gob.mx/conapesca/documentos/anuario-estadistico-de-acuacultura-y-pesca>
- Del Río-Portilla, M. A., Vargas-Peralta, C. E., Farfán, C., Barriga-Sosa, I. de L. A., & García-De-León, F. J. (2016). The complete mitochondrial DNA of the bay snook, *Petenia splendida* , a native Mexican cichlid. *Mitochondrial DNA*, 27(2), 1381–1382. <https://doi.org/10.3109/19401736.2014.947590>
- Del Río-Portilla, M. A., Vargas-Peralta, C. E., Lafarga-De La Cruz, F., Arias-Rodriguez, L., Delgado-Vega, R., Galván-Tirado, C., & García-de-León, F. J. (2016). The complete mitochondrial DNA of the tropical gar (*Atractosteus tropicus*) . *Mitochondrial DNA*, 27(1), 557–558. <https://doi.org/10.3109/19401736.2014.905856>
- Dierckxsens, N., Mardulyn, P., & Smits, G. (2016). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, 45(4), gkw955. <https://doi.org/10.1093/nar/gkw955>
- Donaldson, K. A., & Wilson, R. R. (1999). Amphi-Panamic Geminates of Snook (Percoidei: *Centropomidae*) Provide a Calibration of the Divergence Rate in the Mitochondrial DNA Control Region of Fishes. *Molecular Phylogenetics and Evolution*, 13(1), 208–213. <https://doi.org/10.1006/mpev.1999.0625>

- Du, L., Zhang, C., Liu, Q., Zhang, X., & Yue, B. (2018). Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. *Bioinformatics*, 34(4), 681–683. <https://doi.org/10.1093/bioinformatics/btx665>
- Escárcega-Rodríguez, S. (2010). "Aclimatación al cautiverio del robalo prieto, *Centropomus nigrescens* (Gunther, 1864), en estanquería rústica en la costa de Michoacán México ". [Tesis de maestría, Universidad Michoacana de San Nicolás de Hidalgo Instituto de Investigaciones sobre los recursos naturales (INIRENA-UMSNH)]. http://bibliotecavirtual.dgb.umich.mx:8083/xmlui/handle/DGB_UMICH/5781.
- Escárcega-Rodríguez, S. (2018). Primeras pruebas de adaptación al cultivo extensivo del robalo (*Centropomus nigrescens*) en estanques en la costa de Michoacán, México. *CIENCIA Ergo Sum*, 25(3), 1–13. <https://doi.org/10.30878/ces.v25n3a8>
- Faircloth, B. C. (2008). MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*, 8(1), 92–94. <https://doi.org/10.1111/j.1471-8286.2007.01884.x>
- Flanagan, S. P., & Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology*, 28(3), 544–567. <https://doi.org/10.1111/mec.14988>
- Formenti, G., Rhie, A., Balacco, J., Haase, B., Mountcastle, J., Fedrigo, O., Brown, S., Capodiferro, M. R., Al-Ajli, F. O., Ambrosini, R., Houde, P., Koren, S., Oliver, K., Smith, M., Skelton, J., Betteridge, E., Dolucan, J., Corton, C., Bista, I., ... Bukhman, Y. (2021). Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biology*, 22(1), 1–22. <https://doi.org/10.1186/s13059-021-02336-9>
- Galván-Tirado, C., del Río-Portilla, M. A., Delgado-Vega, R., & García-De León, F. J. (2015). Genetic variability between complete mitochondrion genomes of the sablefish, *Anoplopoma fimbria* (Pallas, 1814). *Mitochondrial DNA Part A*, 27(4), 2429–2430. <https://doi.org/10.3109/19401736.2015.1030628>
- Giovanni, B., Apolinar, S.-M., Manuel, M.-B. J., & Leonardo, I.-C. (2022). Technical-economic viability of white snook *Centropomus viridis* culture in floating cages in a coastal lagoon in northwestern Mexico. *Aquaculture Reports*. <https://doi.org/10.1016/j.aqrep.2022.101048>
- Granados Amores Esperanza. (2018). Análisis taxonómico del género *Centropomus* (Perciformes: Centropomidae) del sureste del Golfo de California, México. [Tesis de maestría, Universidad Autónoma De Nayarit Posgrado en ciencias biológico-agropecuarias] <http://dspace.uan.mx:8080/jspui/handle/123456789/2292>.
- Greenwood, P. H. (1976). A review of the family Centropomidae (Pisces, Perciformes). *Bulletin of the British Museum (Natural History) Zoology*, 29(1), 1–81. <https://doi.org/10.5962/p.271710>
- Hahn-von Hessberg, C. M., Vélez-Marín, M., & Quintero, A. G. (2015). Utilización de la biología molecular como medio para optimizar la producción piscícola y repoblamiento de medios naturales. *Boletín Científico Del Centro de Museos*, 19(1), 85–102. <https://doi.org/10.17151/bccm.2015.19.1.6>
- Harris, M. A., Clark, J. I., Ireland, A., Lomax, J., Ashburner, M., Collins, R., Eilbeck, K., Lewis, S., Mungall, C., Richter, J., Rubin, G. M., Shu, S. Q., Blake, J. A., Bult, C. J., Diehl, A. D., Dolan, M. E., Drabkin, H. J., Eppig, J. T., Hill, D. P., ... Westerfield, M. (2006). The Gene Ontology (GO) project in 2006. *Nucleic*

- Acids Research*, 34(90001), D322–D326. <https://doi.org/10.1093/nar/gkj021>
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- INAPESCA(2020). *Avances tecnológicos en el cultivo de Robalo blanco en México Avances en investigación y adaptación del Robalo blanco*. 2–4. Recuperado el 18 de diciembre del 2022 de <https://www.gob.mx/inapesca/es/articulos/avances-tecnologicos-en-el-cultivo-de-robalo-blanco-en-mexico?idiom=es>
- Ju, J., Kim, D. H., Bi, L., Meng, Q., Bai, X., Li, Z., Li, X., Marma, M. S., Shi, S., Wu, J., Edwards, J. R., Romu, A., & Turro, N. J. (2006). Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proceedings of the National Academy of Sciences of the United States of America*, 103(52), 19635–19640. <https://doi.org/10.1073/pnas.0609513103>
- Kiang, J. (1998). Heat Shock Protein 70 kDa Molecular Biology, Biochemistry, and Physiology. *Pharmacology & Therapeutics*, 80(2), 183–201. [https://doi.org/10.1016/S0163-7258\(98\)00028-X](https://doi.org/10.1016/S0163-7258(98)00028-X)
- Kirchner, S., & Ignatova, Z. (2015). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nature Reviews Genetics*, 16(2), 98–112. <https://doi.org/10.1038/nrg3861>
- León-Pech, M. G., Castillo-Páez, A. Y., Bisbal-Pardo, C. I., Rocha-Olivares, A., & del Río-Portilla, M. A. (2016). Complete mitochondrial genome of the beaubrummel Damsel fish, *Stegastes flavilatus* (Pisces: Perciformes, Pomacentridae). *Mitochondrial DNA Part A*, 27(6), 4136–4138. <https://doi.org/10.3109/19401736.2014.1003880>
- Li, Z., Ma, B., Li, X., Lv, Y., Jiang, X., Ren, C., Hu, C., & Luo, P. (2022). The Complete Mitochondrial Genome of *Stichopus naso* (Aspidochirotrida: Stichopodidae: Stichopus) and Its Phylogenetic Position. *Genes*, 13(5), 825. <https://doi.org/10.3390/genes13050825>
- Lin, G., Lo, L. C., Zhu, Z. Y., Feng, F., Chou, R., & Yue, G. H. (2006). The Complete Mitochondrial Genome Sequence and Characterization of Single-Nucleotide Polymorphisms in the Control Region of the Asian Seabass (*Lates calcarifer*). *Marine Biotechnology*, 8(1), 71–79. <https://doi.org/10.1007/s10126-005-5051-z>
- Liu, Z. J., & Cordes, J. F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238(1–4), 1–37. <https://doi.org/10.1016/j.aquaculture.2004.05.027>
- Lopera-Barrero, N. M., Povh, J. A., Ribeiro, R. P., Gomes, P. C., Jacometo, C. B., & Silva Lopes, T. da. (2008). Comparación de protocolos de extracción de ADN con muestras de aleta y larva de peces: extracción modificada con cloruro de sodio. *Ciencia e Investigación Agraria*, 35(1), 77–86. <https://doi.org/10.4067/S0718-16202008000100008>
- Luo, L.-F., Xu, Z.-S., Li, D.-Y., Hu, Z., & Gao, Z.-X. (2022). Comparative transcriptome profiles of four sexually size dimorphic fish. *Scientific Data*, 9(1), 774. <https://doi.org/10.1038/s41597-022-01887-1>
- Magallón-Gayón, E., del Río-Portilla, M. Á., & de los Angeles Barriga-Sosa, I. (2019). The complete mitochondrial genomes of two octopods of the eastern Pacific Ocean: *Octopus mimus* and *Octopus fitchi* (Cephalopoda: Octopodidae) and their phylogenetic position within Octopoda. *Molecular Biology Reports*, 47(2), 943–952. <https://doi.org/10.1007/s11033-019-05186-8>

- Martínez-Brown Juan M, Navarro-Flores Jaime, G.-R. F. J., Ibarra-Castro Leonardo, V.-P. C. E., Ángel, D. R.-P. M., & Rodrigo, M.-M. (2021). Revision of the diagnostic characters of two morphologically similar snook species, *Centropomus viridis* and *C. nigrescens* (Carangiformes: Centropomidae). *Zootaxa*, 4915(3), 326–338. <https://doi.org/10.11646/zootaxa.4915.3.2>
- Miranda, D. F.-F. J., Marceniuk, A. P., Anderson Feijó, Siccha-Ramirez, R., Ribeiro, G. S., Oliveira Caludio, & S., R. R. (2021). Taxonomy of *Centropomus Lacépède, 1802* (Perciformes: Centropomidae), with focus on the Atlantic species of the genus. *Zootaxa*, 4942(3), 301–338. <https://doi.org/10.11646/zootaxa.4942.3.1>
- Morimoto, R. I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes & Development*, 12(24), 3788–3796. <https://doi.org/10.1101/gad.12.24.3788>
- Nunes Pazdiora, B. R. C., Selvaraj, J. J., Marmolejo, D. F., Muñoz, J. E., & Guzmán, Á. I. (2016). Una Metodología no invasiva para la extracción de adn de *coryphaena hippurus*. *Bulletin of Marine and Coastal Research*, 44(2), 395–399. <https://doi.org/10.25268/bimc.invemar.2015.44.2.33>
- Oliveira, E. J., Pádua, J. G., Zucchi, M. I., Vencovsky, R., & Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, 29(2), 294–307. <https://doi.org/10.1590/S1415-47572006000200018>
- Pandey, P. K., & Parhi, J. (2022). Correction to: Advances in Fisheries Biotechnology. In *Advances in Fisheries Biotechnology* (pp. C1–C1). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-3215-0_30
- Peregrino-Uriarte, A., Pacheco-Aguilar, R., Varela-Romero, A., & Yepiz-Plascencia, G. (2007). Differences in the 16SrRNA and cytochrome c oxidase subunit I genes in the mullets *Mugil cephalus* and *Mugil curema*, and snooks *Centropomus viridis* and *Centropomus robalito*. *Ciencias Marinas*, 33(1), 95–104. <https://doi.org/10.7773/cm.v33i1.1104>
- Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53(5), 793–808. <https://doi.org/10.1080/10635150490522304>
- Prabhu, V. R., Singha, H. S., Kumar, R. G., Gopalakrishnan, A., & Nagarajan, M. (2020). Characterization of the complete mitochondrial genome of *Barilius malabaricus* and its phylogenetic implications. *Genomics*, 112(3), 2154–2163. <https://doi.org/10.1016/j.ygeno.2019.12.009>
- Rather, M. A., Agarwal, D., Bhat, T. A., Khan, I. A., Zafar, I., Kumar, S., Amin, A., Sundaray, J. K., & Qadri, T. (2023). Bioinformatics approaches and big data analytics opportunities in improving fisheries and aquaculture. *International Journal of Biological Macromolecules*, 233(July 2022), 123549. <https://doi.org/10.1016/j.ijbiomac.2023.123549>
- Rivas, L. R. (1986). Systematic Review of the Perciform Fishes of the Genus *Centropomus*. *Copeia*, 1986(3), 579. <https://doi.org/10.2307/1444940>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3), 539–542. <https://doi.org/10.1093/sysbio/sys029>

- Rubio, S., Pacheco-Orozco, R. A., Gómez, A. M., Perdomo, S., García-Robles, R., Rubio, S., Pacheco-Orozco, R. A., Milena Gómez, A., Perdomo, S., & García-Robles, R. (2020). Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica*, *61*(2), 49–63. <https://doi.org/10.11144/Javeriana.umed61-2.sngs>
- Salem, M., Paneru, B., Al-Tobasei, R., Abdouni, F., Thorgaard, G. H., Rexroad, C. E., & Yao, J. (2015). Transcriptome Assembly, Gene Annotation and Tissue Gene Expression Atlas of the Rainbow Trout. *PLOS ONE*, *10*(3), e0121778. <https://doi.org/10.1371/journal.pone.0121778>
- Sanciango, M. D., Rocha, L. A., & Carpenter, K. E. (2011). A molecular phylogeny of the Grunts (Perciformes: Haemulidae) inferred using mitochondrial and nuclear genes. *Zootaxa*, *2966*(1), 37. <https://doi.org/10.11646/zootaxa.2966.1.4>
- Santorum, J. M., Darriba, D., Taboada, G. L., & Posada, D. (2014). jmodeltest .org: selection of nucleotide substitution models on the cloud. *Bioinformatics*, *30*(9), 1310–1311. <https://doi.org/10.1093/bioinformatics/btu032>
- Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Sherry, S. T., Yankie, L., & Karsch-Mizrachi, I. (2023). GenBank 2023 update. *Nucleic Acids Research*, *51*(D1), D141–D144. <https://doi.org/10.1093/nar/gkac1012>
- Sharma, A., Siva, C., Ali, S., Sahoo, P. K., Nath, R., Laskar, M. A., & Sarma, D. (2020). The complete mitochondrial genome of the medicinal fish, Cyprinion semiplotum: Insight into its structural features and phylogenetic implications. *International Journal of Biological Macromolecules*, *164*, 939–948. <https://doi.org/10.1016/j.ijbiomac.2020.07.142>
- Shen, Y., & Yue, G. (2019). Current status of research on aquaculture genetics and genomics-information from ISGA 2018. *Aquaculture and Fisheries*, *4*(2), 43–47. <https://doi.org/10.1016/j.aaf.2018.11.001>
- Smith, W. L., & Wheeler, W. C. (2006). Venom Evolution Widespread in Fishes: A Phylogenetic Road Map for the Bioprospecting of Piscine Venoms. *Journal of Heredity*, *97*(3), 206–217. <https://doi.org/10.1093/jhered/esj034>
- Thermo Scientific. (2016). Guía de Usuario del NanoDrop. *Espectrofotómetros Micro-UV/Vis NanoDrop*, 1–342. www.thermoscientific.com/nanodrop
- Tringali, M. D., Bert, T. M., Seyoum, S., Bermingham, E., & Bartolacci, D. (1999). Molecular Phylogenetics and Ecological Diversification of the Transisthmian Fish Genus *Centropomus* (Perciformes: *Centropomidae*). *Molecular Phylogenetics and Evolution*, *13*(1), 193–207. <https://doi.org/10.1006/mpev.1999.0624>
- Vargas-Peralta, C. E., Farfán, C., Lafarga-De La Cruz, F., Barón-Sevilla, B., & Del Río-Portilla, M. A. (2020). Genoma mitocondrial completo del lenguado de California, *Paralichthys californicus*. *Ciencias Marinas*, *46*(4), 297–306. <https://doi.org/10.7773/cm.v46i4.3121>
- Vergara-chen, C. (2014). Los robalos (Pisces, *Centropomidae*) del pacífico de Panamá: Desafíos emergentes en investigación y conservación. *Tecnociencia*, *16*(1), 15–40. Recuperado el 18 de julio del 2022 de <https://revistas.up.ac.pa/index.php/tecnociencia/article/view/1183>
- Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research*, *77*(2), 220–225.

<https://doi.org/10.1016/j.fishres.2005.08.011>

- Ward, C. M., To, T. H., & Pederson, S. M. (2020). NgsReports: A Bioconductor package for managing FastQC reports and other NGS related log files. *Bioinformatics*, 36(8), 2587–2588. <https://doi.org/10.1093/bioinformatics/btz937>
- Wattanadilokchatkun, P., Panthum, T., Jaisamut, K., Ahmad, S. F., Dokkaew, S., Muangmai, N., Duengkae, P., Singchat, W., & Srikulnath, K. (2022). Characterization of Microsatellite Distribution in Siamese Fighting Fish Genome to Promote Conservation and Genetic Diversity. *Fishes*, 7(5), 251. <https://doi.org/10.3390/fishes7050251>
- Wen, B., He, M., Song, J., Zong, B., Lu, Y., & Shu, X. (2000). Study on DeNOx performance of Ce-Mg-Al mixed oxides. *Chinese Journal of Catalysis*, 21(1), 34. https://www.researchgate.net/publication/288541920_Study_on_DeNOx_performance_of_Ce-Mg-Al_mixed_oxides
- Wenne, R. (2023). *Microsatellites as Molecular Markers with Applications in Exploitation and Conservation of Aquatic Animal Populations*. <https://doi.org/10.3390/genes14040808>
- Xu, J., Liu, L., Xu, Y., Chen, C., Rong, T., Ali, F., Zhou, S., Wu, F., Liu, Y., Wang, J., Cao, M., & Lu, Y. (2013). Development and Characterization of Simple Sequence Repeat Markers Providing Genome-Wide Coverage and High Resolution in Maize. *DNA Research*, 20(5), 497–509. <https://doi.org/10.1093/dnares/dst026>
- Yue, G. H., Wang, L., Yang, Z., Sun, F., Tay, Y. X., Wong, J., & Yeo, S. (2023). Genomic resources and their applications in aquaculture of Asian seabass (*Lates calcarifer*). *Reviews in Aquaculture*, 15(2), 853–871. <https://doi.org/10.1111/raq.12764>
- Zhan, L., Paterson, I. G., Fraser, B. A., Watson, B., Bradbury, I. R., Nadukkalam Ravindran, P., Reznick, D., Beiko, R. G., & Bentzen, P. (2017). <sc>scpg</sc> : automated inference of microsatellite genotypes from sequence data. *Molecular Ecology Resources*, 17(2), 247–256. <https://doi.org/10.1111/1755-0998.12561>
- Zhang, J.-H., Wang, F., & Wang, T.-Y. (2011). A simple and effective SuperBuffer for DNA agarose electrophoresis. *Gene*, 487(1), 72–74. <https://doi.org/10.1016/j.gene.2011.05.018>

Anexos

Anexo A. Protocolo de Extracción de ADN del laboratorio de genética acuícola del CICESE:

Preparar la muestra

- 1.- Se pesan en una balanza analítica 20 mg de tejido, para eliminar los excesos de etanol se coloca en un papel secante, una vez seco el tejido se deposita en un mortero previamente esterilizado para macerar con nitrógeno líquido.
- 2.- Una vez macerada la muestra se debe depositar en un tubo Eppendorf estéril de 1.7 ml.
- 3.- Agregar 500 μ l de buffer A con un pH de 7.5, después agitar la muestra y colocar 10 μ l de proteinasa K, incubar los tubos con buffer A y proteinasa K por 2 horas a 65 °C.
- 4.- Después de incubar la muestra, agregar 570 μ l de LiCl y 230 μ l de KAC e incubar nuevamente.
- 5.- Centrifugar la muestra por 15 min a 12000 rpm
- 6.- Después de centrifugar se transfieren 900 μ l del sobrenadante de cada tubo, a nuevos tubos Eppendorf estériles de 1.7 ml, en este paso se debe evitar transferir material particulado.
- 7.- Se colocan 200 μ l de Fenol-Cloroformo en cada tubo, después se debe agitar y centrifugar la muestra a 15 min por 12,000 rpm
- 8.- Concluido el proceso de centrifugación, transferir 800 μ l de sobrenadante evitando mezclar las fases, a tubos Eppendorf nuevos y estériles. Para precipitar el ADN se coloca al sobrenadante 550 μ l de Isopropanol frío y se agitan las muestras enérgicamente, después de agregar el Isopropanol centrifugar por 15 min, 12,000 rpm y 13.4 grav, en este paso se puede observar el pellet que contiene el ADN.
- 9.- Lavar el pellet con 500 μ l de Etanol al 70% y centrifugar por 15 min a 12,000 rpm.

10.- Se decanta el etanol de los tubos de cada muestra, y se deja secar el pellet a temperatura ambiente sobre un papel secante colocando los tubos boca abajo para que el etanol precipite por gravedad.

11.- Se agrega 50 μ l de agua libre de nucleasas a cada tubo de muestra.

12.- Finalmente se agregan 1.5 μ l de RNasa y se incuban los tubos a 37 °C por 40 min, una vez que pasa el tiempo de incubación se guardan las muestras en un congelador a -20°C.

Recomendaciones:

Rotular cada tubo con la siguiente información: método de extracción, tipo de muestra, tejido, organismo y fecha de extracción.

Anexo B Protocolo para elaborar Gel de electroforesis al 1%

- 1.- Se pesan 0.25 g de agarosa en una balanza analítica y se coloca en un matraz Erlenmeyer de 200 ml.
- 2.- Posteriormente, se colocan 25 ml de Super Buffer al 1x(Zhang et al., 2011) al matraz y agitar suavemente.
- 3.- Meter el matraz al microondas durante intervalos de 10 s, agitar suavemente entre cada intervalo hasta que la solución se vea traslúcida.
- 4.- Colocar el contenido del matraz en la cámara electroforética, cuidando que no esté muy caliente y que la solución no forme burbujas y dejar solidificar el gel durante 40 min a temperatura ambiente.
- 5.- Una vez gelificado, ponerlo dentro de la cámara con "Buffer SB" (Super Buffer) al 1x para retirar los peines más fáciles.
- 6.- Para cargar la escalera, se colocan 2 μ l de escalera y 4 μ l de H₂O libre de nucleasas, para cargar las muestras en un papel Parafilm se coloca la cantidad de ADN y buffer de carga 5x a cada muestra (para este experimento se cargó 1 μ l de muestra, 4 μ l de H₂O y 1 μ l de buffer de carga para tener 6 μ l en total en cada pozo). Se programa el equipo de electroforesis a 90 Volts por 60 min y después de este tiempo se puede observar en un fotodocumentador Gel Doc TM XR + Imager (Bio-Rad).

Recomendaciones: Armar la cámara antes de pesar la agarosa

No poner la agarosa caliente, ya que la cámara de electroforesis se puede agrietar

Poner los pocillos en el polo negativo

Usar una punta distinta para cada muestra y cada reactivo

Limpiar el equipo solo con agua destilada después de su uso

Anexo C.

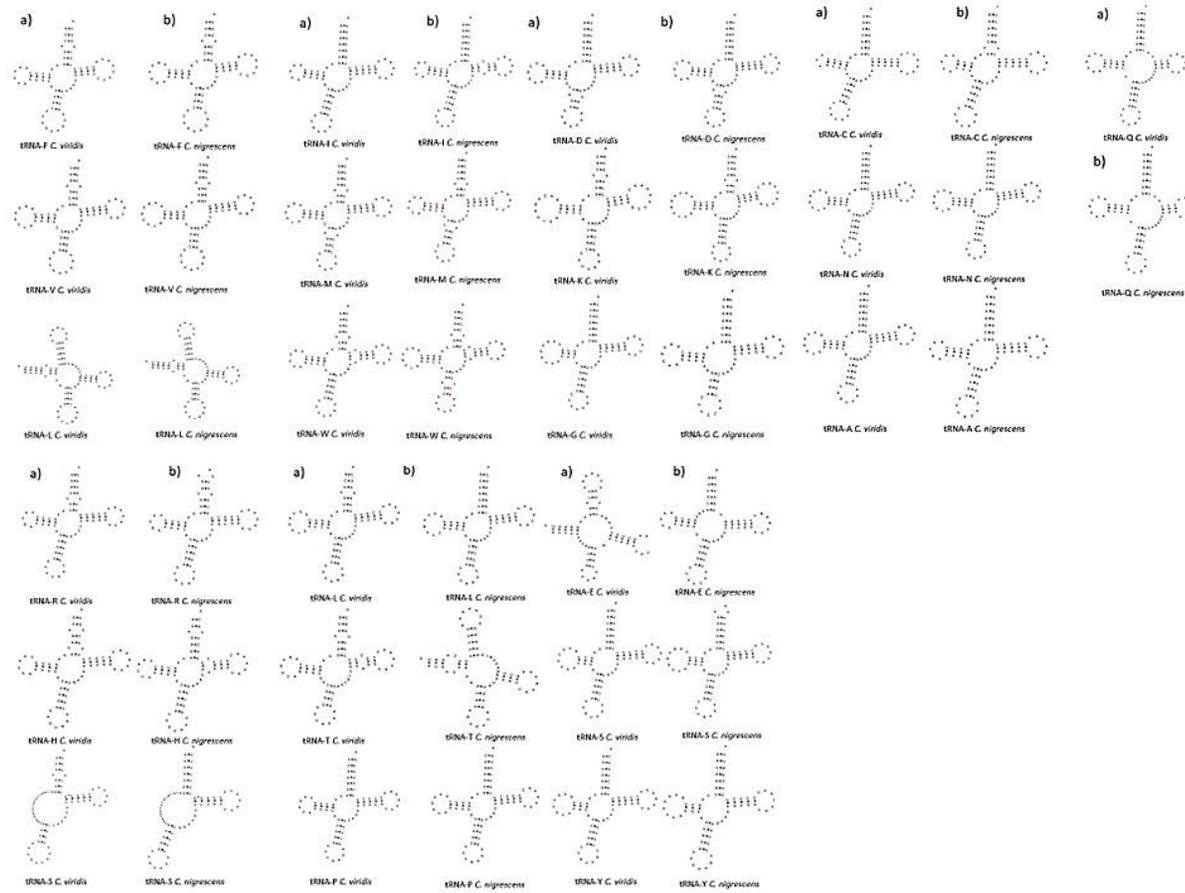


Figura 17. Estructura de los ARNt de las especies *Centropomus viridis* (a) y *Centropomus nigrescens* (b).

Anexo D. Código utilizado para FastQC y Trimmomatic utilizando el servidor

Omica

```
fout = open("Calidad_FastQC.sh", "w")
linea=""
linea=""#!/bin/sh
#
#SBATCH -p cicese
#SBATCH --job-name=FastQC
#SBATCH -e FastQC.err
#SBATCH -o FastQC.log
#SBATCH -t 1-00:00:00
#SBATCH --memory=5G
#
#SBATCH -N 1
#SBATCH -n 5
#
#SBATCH --exclusive
cd $SLURM_SUBMIT_DIR
#
export PATH=$PATH:/LUSTRE/apps/bioinformatica/FastQC_v0.11.5/
cd /LUSTRE/bioinformatica_data/Iga/AlejandraCC/CVI
time /LUSTRE/apps/bioinformatica/FastQC_v0.11.5/fastqc -O
/LUSTRE/bioinformatica_data/Iga/AlejandraCC/CVI/ CVI_R1_001.fastq.gz CVI_R2_001.fastq.gz
exit 0
""
fout.write(linea)
fout.close()
```

Código para trimmomatic usando el servidor Omica

```
fout = open("trim_cvi.sh", "w")
lineas= ""
lineas= ""#!/bin/bash
#
```

```
#SBATCH -p cicese
#SBATCH --job-name=FastQC
#SBATCH -e FastQC.err
#SBATCH -o FastQC.log
#SBATCH -t 1-00:00:00
#SBATCH --memory=5G
#
#SBATCH -N 1
#SBATCH -n 4
#
#SBATCH --exclusive
cd $SLURM_SUBMIT_DIR
#
export PATH=$PATH:/LUSTRE/bioinformatica_data/genomica_funcional/bin/Trimmomatic-0.36/
cd /LUSTRE/bioinformatica_data/lga/AlejandraCC/CVI

time java -jar /LUSTRE/bioinformatica_data/genomica_funcional/bin/Trimmomatic-0.36/trimmomatic-0.36.jar PE -threads 4 -phred33 CVI_R1_001.fastq.gz CVI_R2_001.fastq.gz cvi_1_pair.fastq
cvi_1_unpair.fastq cvi_2_pair.fastq cvi_2_unpair.fastq ILLUMINACLIP:Sequencing_adaptors.fa:2:30:10
HEADCROP:10 SLIDINGWINDOW:20:15 MINLEN:100

exit 0
''''

fout.writelines(lineas)
fout.close()
```

Anexo E. Código para realizar ensamblaje *de Novo* con TRINITY

```
fout = open("Ens_trincni2.sh", "w")

lineas= ""#!/bin/bash

#

#SBATCH -p cicese

#SBATCH --job-name=Trinity

#SBATCH -e FastQC.err

#SBATCH -o FastQC.log

#SBATCH -t 1-00:00:00

#SBATCH --memory=25G

#

#SBATCH -N 1

#SBATCH -n 12

#

#SBATCH --exclusive

cd $SLURM_SUBMIT_DIR

#

export PATH=$PATH:/LUSTRE/apps/bioinformatica/trinityrnaseq-2.4.0/
```

```
export PATH=$PATH:/LUSTRE/apps/bioinformatica/bowtie2-2.2.9/
```

```
export PATH=$PATH:/LUSTRE/apps/bioinformatica/samtools-1.7/bin/
```

```
cd /LUSTRE/bioinformatica_data/lga/AlejandraCC/robalo_cni/
```

```
time /LUSTRE/apps/bioinformatica/trinityrnaseq-2.4.0/Trinity --seqType fq --SS_lib_type RF --left  
cni2_1_pair.fastq.gz --right cni2_2_pair.fastq.gz --CPU 12 --max_memory 25G
```

```
exit 0
```

```
""""
```

```
fout.writelines(lineas)
```

```
fout.close()
```

Anexo F. Código para hacer ensamblaje mediante NOVOPlasty

```
fout =open("config.txt", "w") lines = "" Project: Project name      = prueba_alejandra  Type      =
mito Genome Range      = 16000-17000 K-mer          = 39 Max memory      = Extended log  =
Save assembled reads = Seed Input          = seed_cvi.fasta Reference sequence = Variance detection
= Chloroplast sequence = Dataset 1: Read Length      = 151Insert size      = 412 Platform      =
illumina Single/Paired      = PE Combined reads      = Forward reads      =
CVI_S1_L007_R1_001.fastq.gzReverse reads      = 1CVI_S1_L007_R2_001.fastq.gz Heteroplasmy:-----
-----MAF          =HP exclude list    = PCR-free          =Optional:-----Insert size auto
= yes Insert Range      = 1.9 Insert Range strict = 1.3 Use Quality Scores ="" fout.write(lines) fout.close()
```

Anexo G. Código para realizar árboles filogenéticos con Mr. Bayes

La explicación de este código se encuentra en materiales y métodos

```
fnex = open("carangaria_mtdna.nex", "a")
```

```
lineas = ""
```

```
lset nst=6 rates=invgamma;
```

```
mcmcp ngen=30000 relburnin=yes burnfrac=0.25 samplefreq=100 printfreq=500 diagnfreq=1000  
nchains=4;
```

```
mcmc;
```

```
sump;
```

```
sumt;
```

```
""
```

```
fnex.write(lineas)
```

```
fnex.close()
```

Tabla 9. Tabla de genes identificados para *C. viridis* utilizando Blastx, los números de acceso presentado son de secuencias de otras especies que previamente tenían anotación.

<i>Centropomus viridis</i>				
Grupo	Gen	Contig	Longitud en pares de bases (pb)	Número de acceso NCBI
HSP	Heat shock protein HSP 90-alpha	1994	332 pb	A5A6K9.1
	Hsp90 co-chaperone Cdc37	2051	609 pb	Q9DGQ7.1
SCM	Structural maintenance of chromosomes protein 6	804	629pb	Q802R8.1
	Structural maintenance of chromosomes protein 3	2221	882 pb	P48746.1
18S	18S ribosomal	248	3400 pb	-

Nota: La longitud en pares de bases se verificó en el programa CLC Genomics Workbench, los números de acceso corresponden a las secuencias que ya presentan un registro previo y que se obtuvieron del resultado mediante Blastx.

Tabla 10. Tabla de genes de interés identificados para *C. nigrescens* utilizando Blastx, los números de acceso presentado son de secuencias de otras especies que previamente tenían anotación.

Centropomus nigrescens

Grupo	Gen	Contig	Longitud en pares de bases (pb)	Número de acceso NCBI
HSP	Heat shock protein HSP 90-alpha 1	21829	1589 pb	Q90474.3
	HSP 70	18551	2123 pb	P34932.4 Q07139.2
	HSP 30	7277	2306 pb	P42931.1
SCM	Structural maintenance of chromosomes protein 6	1070	2314 pb	Q802R8.1
	Structural maintenance of chromosomes protein 3	6811	2616pb	P97690.1
18S	18S ribosomal	15880	2857 pb	-

Tabla 11 Comparación de costos entre protocolos de extracción de ADN

Protocolo	Costo por Muestra Aproximado (\$MX)	Tiempo (Horas, Minutos)
Thermofisher - PrepSEQ™ 1-2-3 Nucleic Acid Extraction Kit	\$226.93	1 hora, 45 Minutos
Método de sales	\$3.50	6 horas, 30 minutos