

Caracterización de la historia demográfica del tiburón ballena (*Rhincodon typus*) a partir de
ADN mitocondrial.

Katherin Ximena Lopez Lizarazo

Trabajo de Grado para Optar al Título de Bióloga

Director

Sergio Andrés Marchant Rojas

Doctor en Ciencias Biológicas

Codirectora

Dalia Carolina Barragán Barrera

Doctora en Ciencias Biológicas

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Biología

Bucaramanga

2026

Agradecimientos

Agradezco a mi mamá, a mi abuela, a mis hermanos y a mi papá, Sergio, por todo el apoyo, la ayuda a lo largo de toda mi carrera profesional y por ser el motor que me motiva a avanzar en mi vida académica.

También quiero agradecer a mi novio, por siempre tener una palabra de aliento y por no dejarme desfallecer en todo este proceso.

A mis mascotas, Bonita y Katara por sr mis acompañantes fieles durante todo el transcurso de mi vida universitaria.

A mis amigos y compañeros por ser siempre los confidentes de mis alegrías y tristezas, en los cuales siempre encontré una actitud de escucha y apoyo incondicional.

A mi Director y Co-directora del trabajo de grado, por ser mis guías y aportarme todo su conocimiento y ayuda en el desarrollo de este trabajo.

A los miembros del Laboratorio de Laboratorio de Fisiología, Genómica y transcriptómica por su compañía, ayuda y experiencias compartidas.

A la Fundación Malpelo y Otros Ecosistemas Marinos por las muestras suministradas para el desarrollo de este proyecto.

Finalmente, agradecer al grupo de música y danzas folclóricas UIS, y al director Néstor Enrique Sánchez, por hacer mi paso por la universidad toda una experiencia llena de color, pasión y alegría.

Tabla de Contenido

	Pág.
Introducción.....	10
1. Objetivos.....	13
1.1 Objetivo General	13
1.2 Objetivos Específicos	13
2. Cuerpo del Trabajo.....	14
2.1 Marco Teorico.....	14
2.1.1 Historia de vida del tiburón ballena	14
2.1.2 Principales amenazas del tiburón ballena	15
2.1.3 Inferencia demográfica y teoría coalescente en genética de poblaciones.....	17
2.2 Metodología	19
2.2.1. Área de estudio, Extracción de ADN, PCR y secuenciación.....	19
2.2.1.1 Área de estudio y colecta de muestras.....	19
2.2.1.2 Extracción de ADN, PCR y secuenciación.....	20
2.2.1.3 Procesamiento Bioinformático.....	21
2.2.2 Análisis de datos.....	21
2.2.2.1 Unidades evolutivas.....	22
2.2.2.2 Historia demográfica.	23
2.2.2.2.1 Variación haplotípica y nucleotídica	23
2.2.2.2.2 Teoría coalescente.	24
2.3 Resultados.....	26
2.3.1 Área de estudio, Extracción de ADN, PCR y secuenciación.....	26

2.3.1.1 Área de estudio y colecta de muestras.....	26
2.3.1.2 Extracción de ADN, PCR y secuenciación.....	27
2.3.1.3 Procesamiento Bioinformático.....	27
2.3.2 Análisis de datos.....	28
2.3.2.1 Unidades evolutivas.....	28
2.3.2.2 Historia demográfica.....	33
2.3.2.2.1 Variación haplotípica y nucleotídica.....	33
2.3.2.2.2 Teoría coalescente.....	36
2.4 Discusión.....	39
3. Conclusiones.....	44
4. Recomendaciones.....	46
Referencias Bibliográficas.....	47
Apéndices.....	55

Lista de Tablas

	Pág.
Tabla 1. Medidas de diversidad genética y de neutralidad mitocondrial	39
Tabla 2. Estimaciones de tamaños poblacionales efectivos (N_e) y ancestral (NANC) del tiburón ballena por cuenca oceánica	39

Lista de Figuras

	Pág.
Figura 1. Puntos de colecta de las muestras de tiburón ballena (<i>Rhicondon typus</i>) en el Santuario de Fauna y Flora de la isla Malpelo	20
Figura 2. Mapa con localidades de las secuencias de región control mitocondrial	28
Figura 3. Árbol filogenético basado en las secuencias de Región Control mitocondrial de tiburón ballena (<i>Rhicondon typus</i>), para el conjunto de datos curado con Gblocks, realizado bajo el método de Máxima Verosimilitud (ML).	30
Figura 4. Árbol filogenético basado en las secuencias de Región Control mitocondrial de tiburón ballena (<i>Rhicondon typus</i>), para el conjunto de datos curado con Gblocks, realizado bajo el método de Inferencia Bayesiana (BI).	31
Figura 5. Análisis discriminante de Componentes principales (DAPC) de las secuencias de Región Control mitocondrial de tiburón ballena (<i>Rhicondon typus</i>)	32
Figura 6. Valores obtenidos de diferenciación genética entre poblaciones para el estadístico PhiST.	33
Figura 7. Resultado del Análisis de Distribución de Discordancias (MDA) para el Océano Atlántico, el Océano Índico, el Pacífico Occidental y el Pacífico Occidental	35
Figura 8. Modelo K=4 definido como el modelo óptimo por inferencia coalescente basada en el Espectro de Frecuencias Alélicas multidimensional (mSFS).....	38

Lista de Apéndices

	pág.
Apéndice A. Información sobre las secuencias de tiburón ballena (<i>Rhincodon typus</i>) obtenidas de GenBank.....	55
Apéndice B. Información sobre la extracción de ADN de las muestras de tiburón ballena (<i>Rhincodon typus</i>), correspondientes al Santuario de Fauna y Flora de Malpelo, en la isla Malpelo (Colombia).....	57
Apéndice C. Datos obtenidos durante la colecta de muestras de tiburón ballena (<i>Rhincodon typus</i>) dentro del Santuario de Fauna y Flora de Malpelo para los 4 puntos de muestreo en inmediaciones a la isla Malpelo (Colombia)	56
Apéndice D. Árboles filogenéticos basados en las secuencias de Región Control de ADN mitocondrial de tiburón ballena (<i>Rhincodon typus</i>), para los conjuntos de datos con gaps completos y con gaps reducidos realizados bajo el método de Máxima Verosimilitud.	57
Apéndice E. Árbol filogenético basado en las secuencias de Región Control Mitocondrial de tiburón ballena (<i>Rhincodon typus</i>), para el conjunto de datos con gaps reducidos curado con BMGE, realizado bajo el método de Inferencia Bayesiana.....	59
Apéndice F. Numero de clústeres genéticos inferidos para el tiburón ballena (<i>Rhincodon typus</i>), a partir del algoritmo k-means.....	59
Apéndice G. Valores obtenidos de diferenciación genética entre poblaciones para los estadísticos JostD y Gst.....	60

Resumen

Título: Caracterización de la historia demográfica del tiburón ballena (*Rhincodon typus*) a partir de ADN mitocondrial. *

Autor: Katherin Ximena Lopez Lizarazo. **

Palabras Clave: Unidades evolutivas, Historia demográfica, Frecuencia de alelos, Teoría coalescente

Descripción: El tiburón ballena (*Rhincodon typus*) es una especie altamente migratoria y amenazada, cuya estructura poblacional global y conectividad histórica han sido objeto de debate. Para llenar vacíos de información críticos, se secuenció la Región Control de ADN mitocondrial (ADNmt) de muestras provenientes del Santuario de Fauna y Flora Malpelo (Pacífico colombiano) integrándolos con secuencias globales. Metodológicamente, se combinaron análisis de genética poblacional clásica (distribución de discordancias y pruebas de neutralidad) con inferencia coalescente utilizando el Espectro de Frecuencias Alélicas multidimensional (mSFS). Los resultados revelaron que el Océano Atlántico está fuertemente diferenciado del Indo-Pacífico, exhibiendo firmas genéticas propias de una población estable. El modelo coalescente óptimo rechazó contundentemente la hipótesis de panmixia en el Indo-Pacífico, demostrando que la especie se estructura en cuatro unidades demográficas independientes: Atlántico, Índico, Pacífico Occidental y Pacífico Oriental. Adicionalmente, se detectaron firmas de expansiones poblacionales masivas en las cuencas del Indo-Pacífico. Las reconstrucciones demográficas indican que las expansiones y las subsecuentes fragmentaciones poblacionales están intrínsecamente ligadas a las barreras oceanográficas impuestas por los ciclos glaciales del Pleistoceno, destacando un reciente efecto fundador en el Pacífico Oriental asociado al Último Máximo Glacial. Estos hallazgos descartan la conectividad global reciente, redefinen las unidades evolutivas de *R. typus* y proveen un marco evolutivo esencial para orientar estrategias de manejo y conservación regionalizadas.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Biología. Director: Sergio Andrés Marchant Rojas. Doctor en Ciencias Biológicas. Codirectora: Dalia Barragán Barrera. Doctora en Ciencias Biológicas.

Abstract

Title: Demographic history characterization of Whale shark (*Rhincodon typus*) using mitochondrial DNA *

Author(s): Katherin Ximena Lopez Lizarazo **

Key Words: Evolutionary units, Demographic history, Allele frequency, Coalescent theory

Description: The whale shark (*Rhincodon typus*) is a highly migratory and threatened species, whose global population structure and historical connectivity have been the subject of debate. To fill critical information gaps, the Mitochondrial Control Region (ADNmt) of samples from the Malpelo Fauna and Flora Sanctuary (Colombian Pacific) was sequenced and integrated with global sequences. Methodologically, classical population genetics analyses (mismatch distribution and neutrality tests) were combined with coalescent inference using the multidimensional allele frequency spectrum (mSFS). The results revealed that the Atlantic Ocean is strongly differentiated from the Indo-Pacific, exhibiting genetic signatures typical of a stable population. The optimal coalescent model strongly rejected the hypothesis of panmixia within the Indo-Pacific, demonstrating that the species is structured into four independent demographic units: Atlantic, Indian, Western Pacific, and Eastern Pacific. Additionally, signatures of massive population expansions were detected across the Indo-Pacific basins. Demographic reconstructions indicate that population expansions and subsequent fragmentations are intrinsically linked to oceanographic barriers imposed by Pleistocene glacial cycles, highlighting a recent founder effect in the Eastern Pacific associated with the Last Glacial Maximum. These findings rule out recent global connectivity, redefine the evolutionary units of *R. typus*, and provide an essential evolutionary framework to guide regionalized management and conservation strategies.

* Degree Work

**Faculty of Science. Biology Mayor. Director: Sergio Andrés Marchant Rojas. Doctor in Biological Science.

Codirector: Dalia Barragán Barrera. Doctor in Biological Science.

Introducción

El tiburón ballena (*Rhincodon typus*) es el pez más grande del mundo y una especie emblemática de los ecosistemas marinos tropicales y subtropicales. A pesar de su amplia distribución y alta capacidad de dispersión, su biología lo hace particularmente vulnerable a las perturbaciones antrópicas. Su crecimiento lento, maduración tardía, baja frecuencia reproductiva, comportamiento dócil en superficie y abundancia relativamente baja incrementan su susceptibilidad a la sobreexplotación y a otras amenazas humanas (Stewart & Wilson, 2005). En este contexto, y dada su creciente importancia ecológica, económica y cultural, particularmente en actividades como el ecoturismo, el tiburón ballena ha sido reconocido como una especie prioritaria para la conservación a nivel global (Pierce et al, 2025).

Sin embargo, el desarrollo de estrategias de conservación efectivas para esta especie enfrenta importantes limitaciones, principalmente asociadas a la dificultad de observación, muestreo y seguimiento a lo largo de su distribución global. Estas limitaciones se reflejan en el reducido número de estudios genéticos disponibles, lo cual se refleja en la literatura científica: solo existen nueve estudios genéticos a nivel global, de los cuales apenas tres incluyen poblaciones de América (Pacífico de Panamá y Golfo de California, México) esto ha proporcionado una visión aún incompleta de su estructura poblacional. Hasta hace poco, la comunidad científica indicaba la presencia de al menos dos poblaciones globales que rara vez se mezclan: la del Indo-Pacífico, para la cual se ha reportado una expansión histórica, pero con evidencia reciente de posibles reducciones en su diversidad genética (Vignaud et al., 2014); y la del Atlántico, significativamente diferenciada, con una historia demográfica distinta, contacto secundario entre linajes y un historial de coalescencia más corto (Yagishita et al., 2020).

No obstante, evidencia más reciente sugiere un escenario más complejo. En particular, se ha propuesto la existencia de cuatro unidades poblacionales genéticamente diferenciadas a escala global (Atlántico, Índico, Pacífico Occidental y Pacífico Oriental) lo que indicaría una estructuración más fina de la especie asociada a cuencas oceánicas (Saavedra Valderrama et al., 2025). Este modelo no solo cuestiona la visión tradicional de una división binaria, sino que también resalta la importancia de incorporar enfoques analíticos más robustos que permitan integrar procesos históricos y contemporáneos, tales como la inferencia coalescente y la reconstrucción de la historia demográfica.

En ese sentido, la reconstrucción de la historia demográfica permite comprender cómo eventos históricos, como las fluctuaciones climáticas del Pleistoceno, han influido en la distribución y estructura genética actual de las poblaciones. La integración de métodos tradicionales (e.g., estadísticos de diversidad y neutralidad) con enfoques modernos basados en teoría coalescente y espectros de frecuencias alélicas ofrece una aproximación más robusta para inferir procesos como expansiones poblacionales, cuellos de botella y eventos de divergencia.

En última instancia, la delimitación de unidades evolutivamente significativas (ESU) es fundamental para diseñar estrategias de conservación adecuadas que aseguren la persistencia de la diversidad genética y la viabilidad a largo plazo de especies altamente migratorias, donde la conectividad aparente puede ocultar procesos históricos de diferenciación. Por consiguiente, determinar si existen dos o cuatro poblaciones globales ayudará a comprender mejor la historia demográfica de la especie y su capacidad de resiliencia además la carencia de una delimitación precisa de estas unidades y de una reconstrucción de sus fluctuaciones históricas constituye un vacío crítico que impide un manejo efectivo frente a las presiones antropogénicas actuales.

En el contexto del Pacífico Oriental Tropical, la información genética disponible es particularmente limitada, con pocos estudios que incluyan localidades de América, como el Pacífico de Colombia, Panamá y el Golfo de California (México). En este escenario, el Santuario de Fauna y Flora (SFF) Malpelo representa un sitio de agregación de gran relevancia ecológica, pero hasta ahora carecía de información genética que permitiera entender su papel dentro de la estructura poblacional global de la especie.

Por lo tanto, el presente estudio contó con el objetivo de caracterizar la historia demográfica del tiburón ballena (*Rhincodon typus*) a nivel global para inferir su estructura y dinámica poblacional. Se integraron muestras de 14 localidades a nivel mundial, incluyendo por primera vez datos para Colombia. Mediante el uso del marcador mitocondrial de la Región Control y aplicación de métodos filogenéticos, análisis de diferenciación genética e inferencia coalescente, se busca responder a la pregunta: ¿Cuáles son las trayectorias demográficas y la estructuración genética de tiburón ballena (*Rhincodon typus*) a partir de la integración de datos genéticos globales y regionales?

1. Objetivos

1.1 Objetivo General

Caracterizar la historia demográfica del tiburón ballena (*Rhincodon typus*) a nivel global para inferir su estructura y dinámica poblacional.

1.2 Objetivos Específicos

Establecer las unidades evolutivas independientes presentes en tiburón ballena a partir de información disponible en bases de datos de secuencias genéticas de acceso público y de muestras colectadas en el Pacífico colombiano.

Inferir la historia demográfica de las unidades evolutivas utilizando métodos con base en frecuencia de alelos y teoría coalescente.

2. Cuerpo del Trabajo

2.1 Marco Teórico

2.1.1 *Historia de vida del tiburón ballena*

El tiburón ballena (*Rhincodon typus*) es el más grande de los tiburones y de todas las especies existentes de peces (Vignaud et al., 2014). Es una especie epipelágica oceánica y costera, distribuido en los mares templados-cálidos y tropicales (Yagishita et al., 2020), la cual se alimenta de una diversidad de organismos. Son en su gran mayoría solitarios, pero se ven agregaciones estacionales de hasta 100 individuos relacionadas a eventos de alta disponibilidad de alimento como desoves masivos (Schmidt et al., 2009). A nivel mundial, se han podido identificar 12 lugares de agregación, los cuales están dominadas por machos juveniles (Hoffmayer et al., 2021). Esta especie es altamente migratoria (Schmidt et al., 2009), con reportes de movimientos transoceánicos estacionales, aunque se desconoce si la frecuencia y distancia de estas migraciones son debido a eventos de alimentación o si están vinculadas a otros aspectos de su historia de vida (Hueter et al., 2013). Debido a que estos tiburones pasan gran parte de su ciclo vital en océanos abiertos o profundos, donde se dificulta la observación y muestreo por parte de los investigadores, varios aspectos de su biología son desconocidos (Vignaud et al., 2014). A pesar de su gran tamaño, es considerada una especie inofensiva pues buceadores han entrado en contacto con ellos sin que los tiburones reaccionen de forma agresiva, solamente huyen o se sumergen repentinamente cuando les molesta la proximidad de los buceadores; esta docilidad y fácil acceso en aguas poco profundas, han generado una popularidad en el buceo ecoturístico (Compagno, 2001).

El tiburón ballena posee marcas corporales como manchas claras, rayas horizontales y barras verticales sobre un fondo oscuro, las cuales forman un patrón en tablero de ajedrez, exclusivo para cada individuo (Stewart & Wilson, 2005). Su peso está por encima de los 21,000

kg y un adulto puede medir entre 10 a 12 m, alcanzando generalmente una longitud máxima de 17 a 18 m aunque algunos pueden llegar hasta los 21 m, siendo los especímenes sobre 12 m poco comunes (Pierce et al, 2025). El tiburón ballena puede ser especialmente longevo, viviendo entre 60 a 100 años (Stevens, 2007); posee una tolerancia a temperaturas ambientales entre 5 a 30 °C y puede sumergirse a más de 1,000 m y permanecer allí entre horas y semanas, realizando cambios regulares entre la superficie y la profundidad (Eckert & Stewart, 1996).

El tiburón ballena es una especie ovovivípara, donde los embriones se desarrollan en cajas de huevos marrones y córneos, que eclosionan en el útero y luego nacen (Stevens, 2007). En Taiwán fue capturada una hembra que tenía 300 embriones en el útero en diversas fases de desarrollo (Stewart & Wilson, 2005), siendo la camada más grande registrada para cualquier especie de tiburón. Aunque el periodo de gestación de esta especie es desconocido, se sugiere que puede reproducirse cada dos años o más (Compagno, 2001). Las crías al nacer son de vida libre, miden entre 55 y 60 cm de longitud, y pesan 1 kg; pasados cuatro meses del nacimiento pueden alcanzar los 1.4 m de longitud y 20 kg de peso (Stewart & Wilson, 2005), mostrando un crecimiento rápido en los primeros años de vida, pero este crecimiento se ralentiza, alcanzándose la madurez sexual entre los 10 y 30 años de vida donde pueden medir entre 6 y 8 m (Stevens, 2007).

2.1.2 Principales amenazas del tiburón ballena

Aparte de los humanos, los depredadores del tiburón ballena son poco conocidos; se han registrado unos pocos eventos de depredación de tiburón ballena en etapa juvenil por tintorera o tiburón azul (*Prionace glauca*) y por Marlín azul (*Makaira nigricans*) (Stevens, 2007). En la etapa adulta, y debido a sus gruesas pieles y gran tamaño, tienen pocos depredadores naturales, siendo las orcas (*Orcinus orca*) y el tiburón blanco (*Carcharodon carcharias*) las especies con

posibilidades de depredación, pero con pocos eventos registrados, por lo que se considera que el principal depredador del tiburón ballena es el ser humano (Compagno, 2001).

La mortalidad de tiburones ballenas se produce principalmente por la captura incidental, especialmente por las pesquerías que trabajan en áreas de agregación de tiburones ballena con redes de enmalle y cerco (Stevens, 2007). Por otra parte, la caza furtiva ha registrado muertes con arpón y red en Taiwán, Pakistán, Indonesia, Maldivas, Filipinas, India, donde en los tres últimos países fue prohibida en los 1995, 1998 y 2001, respectivamente (Stewart & Wilson, 2005), siendo impulsada por las nuevas demandas comerciales de los subproductos del tiburón ballena como la carne, aletas y aceites (Guzmán, Beaver, & Díaz-Ferguson, 2021), principalmente en Asia donde es consumido como sopas, carne seca, entre otros platos. Las capturas fueron mayores entre 1980 e inicios de 1990, donde se registra un promedio anual de 158 individuos capturados en redes fijas y 114 individuos en pesquerías con arpón (Stevens, 2007). Sobre la década de 2000 se disminuye la captura pues la especie fue protegida por las legislaciones de diferentes países al ser clasificada por la Unión Internacional para la Conservación de la Naturaleza (UICN) como "Vulnerable a la extinción", basándose en la probabilidad de que entre el 20 % y el 50 % de la especie se perdiera en las próximas tres generaciones, y en el año 2002 fue incluida en el Apéndice II de la Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres (CITES) (Vignaud et al., 2014).

Quizás las mayores amenazas provengan de la pesca (incidental y furtiva), pero existen otras amenazas como la colisión contra embarcaciones, la destrucción de zonas de alimentación del hábitat del arrecife coralino por la sobreexplotación de los recursos, los episodios de blanqueamiento coralino y el cambio climático que está acelerando el calentamiento de las aguas marinas (Stewart & Wilson, 2005).

2.1.3 Inferencia demográfica y teoría coalescente en genética de poblaciones

El estudio de los patrones geográficos de variación genética, como reflejo de la conectividad y la demografía, es una herramienta fundamental para comprender las respuestas de las especies ante cambios ecológicos o factores de estrés ambientales (Mohd Yusoff et al., 2021). En este sentido, las reconstrucciones demográficas históricas permiten relacionar el tamaño efectivo poblacional (N_e) con cambios ambientales pasados, o las diferencias entre especies con otros factores ecológicos, en la especie *Mustelus schmitti* la reconstrucción demográfica histórica permitió relacionar una baja diversidad genética con las oscilaciones del Pleistoceno debidas a La Gran Glaciación Patagónica (Pereyra, García, Miller, Oviedo, & Domingo, 2010). De manera similar, en la especie *Trineustes gratilla* se identificó una disminución en el Tamaño efectivo poblacional (N_e) que puede estar relacionada con una presión antropogénica anterior no documentada, o a cambios ambientales durante el Pleistoceno tardío (Casilagan, Juinio-Meñez, & Crandall, 2013). Estas inferencias no solo explican el pasado, sino que proporcionan rangos de estimación críticos para interpretar la heterocigosidad, la pérdida de variación genética o riesgos de extinción (Stanhope et al., 2023). Adicionalmente, los estudios de diversidad genética son usados para definir poblaciones aisladas reproductivamente, evaluar el intercambio génico entre poblaciones y determinar la historia evolutiva (Galván-Tirado et al 2013).

Para ejecutar estas reconstrucciones, la secuencia de la región control (CR) del ADN mitocondrial ha sido tradicionalmente el marcador de elección en estudios filogenéticos y de demografía histórica. Su utilidad radica en que tiene una alta tasa de mutación, se hereda por vía materna sin recombinación, y el tamaño efectivo de la población (N_e) es cuatro veces menor que el de marcadores nucleares, por lo que los efectos de deriva génica y cuello de botella son más fáciles de detectar (Wang et al, 2013). A partir de estudios filogenéticos se ha reportado una tasa

de sustitución de CR aproximada de 0.025 a 0.26 sitio/Myr incluidos intervalos de confianza (Parsons et al, 1997). Un estudio en humanos que arroja una de las estimaciones más rápidas situó la tasa de sustitución de las regiones hipervariables CR en 0.118 ± 0.031 /sitio/Myr, suponiendo un tiempo de generación de 20 años, corresponde a $\sim 1/600$ generaciones y una edad para el ADNmt MRCA de 133,000 años (Parsons et al, 1997), aunque estudios en tiburones indican que las tasas de sustitución de genes mitocondriales pueden ser entre siete a ocho veces más lentas, en tiburones que en primates o ungulados (Martin et al, 1992). Para el caso específico del tiburón ballena, la tasa de sustitución de la región control mitocondrial sigue siendo desconocida, aunque un estudio realizado por Weber et al (2020) mostró que el tiburón ballena es el segundo vertebrado de evolución más lenta caracterizado hasta ahora, teniendo una tasa evolutiva más lenta que las del tiburón bambú de bandas marrones, el tiburón gato nublado, y otros vertebrados examinados como el celacanto.

Por otra parte, los estudios demográficos y de coalescencia ayudan a la inferencia de la historia evolutiva de una especie permitiendo realizar pruebas sobre las causas de la variación genética y la estimación de parámetros demográficos, donde los métodos basados en teoría coalescente permiten manejar modelos complejos que incorporan fenómenos como la migración, la selección y la recombinación (Rosenberg & Nordborg, 2002). Estos estudios ofrecen perspectivas importantes para la genética de la conservación al identificar eventos de contracción o expansión poblacional (Spence et al., 2018), siendo fundamentales en la generación de estrategias para mantener la diversidad genética en poblaciones en peligro de extinción.

Actualmente, este campo ha experimentado una revolución gracias a los avances en las técnicas genómicas, el acceso a vastas cantidades de información genética en bases de datos públicas y la gran cantidad de herramientas bioinformáticas avanzadas desarrolladas en los

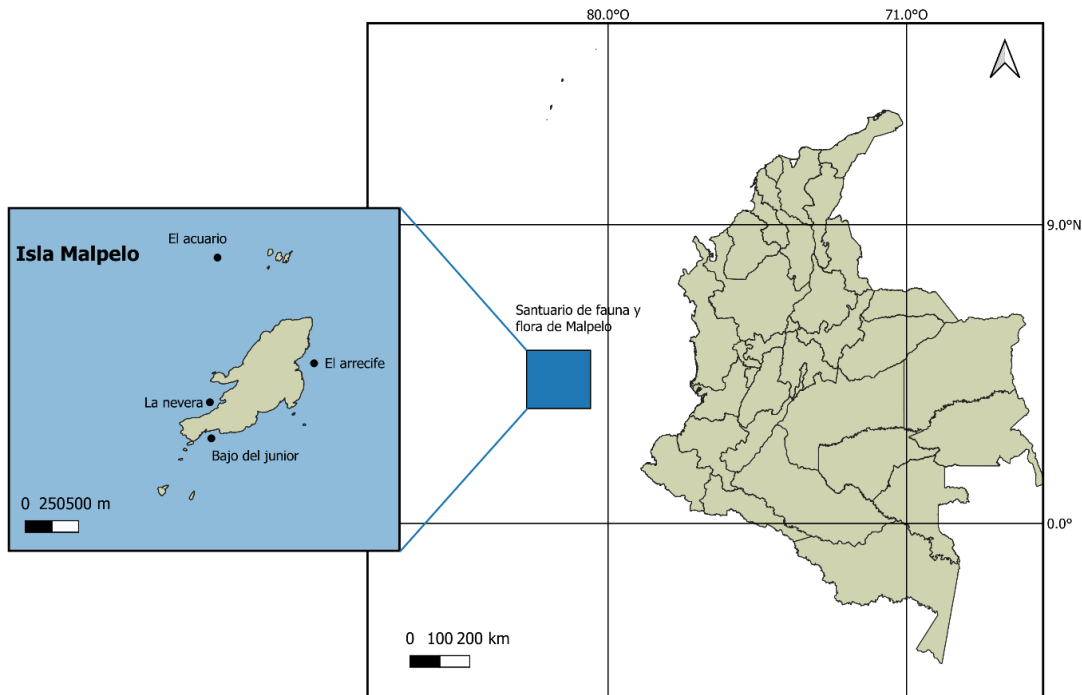
últimos años (Manunza et al, 2023). Esto ha permitido el desarrollo de nuevas técnicas para el estudio de la historia demográfica, como los métodos coalescentes, que incorporan factores como el crecimiento de la población, la recombinación genética, selección natural, mutación, entre otros, al considerarse acontecimientos independientes del proceso genealógico (Mazet & Noûs, 2023). Así, los patrones históricos de flujo génico y demografía son fortalecidos con modelos de simulación que proporcionan estimaciones más realistas del tamaño histórico y la tasa de expansión. Estos modelos superan las limitaciones de la estadística resumida tradicional, la cual puede verse alterada por múltiples fuerzas evolutivas; por ejemplo, la D de Tajima es vulnerable a la recombinación y puede no detectar la selección natural, mientras que el FST puede verse afectado por la diversidad variable entre loci al intentar detectar flujo génico (Kuhner, 2009).

2.2 Metodología.

2.2.1 Área de estudio, Extracción de ADN, PCR y secuenciación

2.2.1.1 Área de estudio y colecta de muestras. Las muestras del pacífico colombiano fueron obtenidas mediante muestreos realizados por la Fundación Malpelo y otros ecosistemas marinos en inmediaciones del Santuario de Fauna y Flora Malpelo (Pacífico Colombiano) durante los años 2015, 2017, 2018 y 2019, en los puntos de buceo El arrecife, El acuario, La nevera y Bajo del Junior (Figura 1). Las muestras obtenidas por Fundación Malpelo y otros ecosistemas marinos, eran tipo biopsias de tejido de individuos de nado libre mediante apnea o buceo, mientras que las características morfológicas se identificaron a través de observación directa por parte de los investigadores de la fundación. Finalmente, las muestras de tejido obtenidas durante las colectas, fueron preservadas en Etanol al 70% y posteriormente refrigeradas para su procesamiento posterior.

Figura 1. Puntos de colecta de las muestras de tiburón ballena (*Rhicondon typus*) en el Santuario de Fauna y Flora de la isla Malpelo.



2.2.1.2 Extracción de ADN, PCR y secuenciación. El ADN de las 16 muestras de tejido correspondientes a la isla Malpelo (Colombia) se extrajo con los kits comerciales DNeasy Blood and Tissue y DNeasy PowerSoil Pro Kit (Qiagen), siguiendo las instrucciones respectivas del fabricante. Cada muestra extraída fue cuantificada mediante NanoPhotometer™ NP80 para conocer su concentración de ADN y así corroborar la efectividad de la extracción.

Posteriormente, se realizó la amplificación de la región control mitocondrial mediante PCR, usando los cebadores WSCR1-F (5'-TTGGCTCCCAAAGCCAAGATTCTTC-3') y WSCR2-R (5'-CTTAATATTTAT TGTCCTGGTTTCAGCC-3') diseñados por Castro et al. (2007). La reacción se realizó en un volumen final de 20µL, compuesto por: 10µL de buffer 10x, 0.8µL de MgCl₂ (50mM), 1µL de dNTPs (4mM), 2µL de Taq polimerasa, 0.5µL de cada cebador y 2µL de

ADN. Se emplearon condiciones de desnaturalización inicial a 95 °C por 1 min, seguida de 35 a 40 ciclos de: 95 °C durante 45 s, 50 °C por 60 s y 72 °C por 90 s; finalizando con una extensión a 72 °C por 10 min. Los productos PCR fueron observados y verificados mediante electroforesis en gel de agarosa al 0.8%. Finalmente, la secuenciación se realizó en un equipo PromethION P2 utilizando el kit Rapid barcoding 96 (SQL-RBK 114.96). Tanto el procesamiento húmedo como el análisis de datos fueron realizados en el Laboratorio de Fisiología, Genómica y Transcriptómica adscrito a la escuela de Biología de la Universidad Industrial de Santander.

2.2.1.3 Procesamiento Bioinformático. Los datos obtenidos por el secuenciador PromethION P2 fueron procesados para la decodificación de bases (basecalling) mediante el software Dorado v.1.1.1(Oxford Nanopore Technologies PLC, 2023). Seguido a esto, las secuencias obtenidas fueron demultiplexadas y asignadas a cada individuo mediante el flujo de trabajo wf-amplicon v.1.1.4 (Oxford Nanopore Technologies PLC, 2024), disponible en el paquete EPI2ME Labs. Con el propósito de contextualizar el estudio a escala global, se obtuvieron 860 secuencias de acceso libre y disponibles en GenBank, para 14 regiones a nivel global, de otros estudios realizados previamente (Apéndice A). Finalmente, estas secuencias fueron alineadas junto con las de Colombia, bajo el algoritmo ClustalW (Thompson et al., 1994) implementado en el software BioEdit v.7.7.1 (Hall, 1999).

2.2.2 Análisis de datos

Al set de datos global se le realizó limpieza y curación mediante los algoritmos de Gblock (Castresana, 2000) y de BMGE (Block Mapping and Gathering with Entropy) (Criscuolo &

Gribaldo, 2010) implementados en línea en el portal NGphylogeny.fr. (Lemoine et al, 2019). Esta curación se ejecutó empleando dos métodos diferentes para obtener 3 grupos de datos, un conjunto de datos con gaps completos (alineamiento en crudo), otro conjunto de datos con gaps reducidos (alineamiento curado con BMGE), y un último conjunto de datos con los gaps reducidos al mínimo posible (alineamiento curado con Gblocks). Esta metodología se llevó a cabo con el fin de investigar el efecto de estas regiones en la señal filogenética y facilitar la comparación con estudios anteriores (Castro et al., 2007; Vignaud et al., 2014; Yagishita et al., 2020) pues la mayoría de ellos emplearon un enfoque dual de datos, con un "alineamiento crudo" con inclusión total de gaps y un "alineamiento modificado" con reducción de gaps. Cabe destacar que, mientras Meekan et al. (2017) redujeron por completo los gaps, estudios más recientes (Hardenstine et al., 2022), han unificado las metodologías anteriores usando los diferentes sets de datos mencionados.

2.2.2.1 Unidades evolutivas. Para caracterizar las unidades evolutivas, se realizaron arboles filogenéticos por Inferencia Bayesiana (BI) y Máxima Verosimilitud (ML). En cada método se realizaron 3 corridas independientes para cada conjunto de datos. Los análisis ML se realizaron utilizando IQTREE v.3.0.1 (T.K.F. Wong et al, 2015) donde el soporte nodal se estimó a partir de 1,000 réplicas rápidas de Bootstrap y 1,000 réplicas de SH-aLRT, con elección automática del mejor modelo de evolución por parte del programa.

Los análisis BI se realizaron usando MRBAYES v.3.2.7 (Ronquist & Huelsenbeck, 2003)), con dos ejecuciones independientes de cuatro cadenas de Markov durante un máximo de 40 millones de generaciones, con muestreo cada 100 generaciones, usando la regla de parada automática con un criterio de convergencia de 0.01 basado en la desviación estándar promedio de las frecuencias

de partición. Se descartaron 2,500 muestras como burn-in y los valores de calentamientos fueron los predeterminados por el programa.

Los árboles obtenidos fueron editados en R v.4.5.3 (R core team, 2020), para una mejor visualización e interpretación de los datos, donde fueron asignados colores a las secuencias según la cuenca geográfica a la que pertenecen, además fueron destacadas las secuencias de Colombia. El soporte de los nodos también fue destacado y coloreado según un intervalo de valores.

De igual forma, se realizó un Análisis discriminante de Componentes principales (DAPC) el cual permite identificar agrupamientos genéticos sin asumir equilibrio de Hardy-Weinberg ni modelos evolutivos específicos, donde se realizó la selección de clústeres genéticos a partir del algoritmo k-means. Además, se estimó la diferenciación genética entre poblaciones de manera pareada empleando los estadísticos PhiST, Gst y JostD, siendo los últimos dos menos influenciados por el número desbalanceado de muestras. Estos análisis se realizaron en R v.4.5.3 (R core team, 2020) únicamente para el conjunto de datos con los gaps reducidos al mínimo posible (alineamiento curado con Gblocks).

2.2.2.2 Historia demográfica

2.2.2.2.1 Variación haplotípica y nucleotídica. Para caracterizar la historia demográfica del tiburón ballena (*Rhincodon typus*), se utilizó las secuencias de la región control del ADN mitocondrial (590 pb) agrupadas en cuatro cuencas oceánicas. Se estimó el número de sitios segregantes (S), la diversidad haplotípica (Hd) y la diversidad nucleotídica (π) siguiendo las aproximaciones teóricas de Nei (1987). Las desviaciones del modelo neutral de mutación-deriva se evaluaron mediante dos estadísticos: la D de Tajima (Tajima, 1989), orientada a detectar excesos de mutaciones de baja frecuencia, y el estadístico R2 de

Ramos-Onsins y Rozas (Ramos-Onsins & Rozas, 2002), el cual presenta mayor poder estadístico para identificar expansiones poblacionales recientes bajo distintos tamaños de muestra. La significancia estadística de R^2 se determinó empíricamente mediante 1,000 simulaciones coalescentes neutras.

La historia demográfica de cada linaje oceánico se reconstruyó inicialmente mediante un Análisis de Distribución de Discordancias (MDA; Rogers & Harpending, 1992), el cual examina la distribución de frecuencias de las diferencias nucleotídicas observadas entre pares de secuencias. La suavidad de las distribuciones empíricas se cuantificó utilizando el Índice de Rugosidad de Harpending (r ; Harpending, 1994), estableciendo un umbral de $r < 0.05$ para rechazar la hipótesis nula de poblaciones estacionarias frente a un escenario de expansión súbita. El tiempo absoluto transcurrido desde la expansión poblacional (t , en generaciones) se calculó a partir del estimador del momento de la expansión (τ) derivado del MDA, utilizando la relación $t = \tau/2u$. La tasa de mutación por secuencia (u) se definió como $u = \mu \times k$, empleando una tasa de mutación nucleotídica (μ) de 7.5×10^{-8} sustituciones/sitio/generación y una longitud de fragmento (k) de 590 pb. Todos los cálculos de diversidad, neutralidad y estadísticos de expansión se ejecutaron en el entorno estadístico R, utilizando los paquetes ape (Paradis et al., 2004) y pegas (Paradis, 2010).

2.2.2.2 Teoría Coalescente. Para reconstruir la historia evolutiva a mayor resolución y determinar las unidades evolutivas independientes de *R. typus*, se infirieron parámetros demográficos históricos utilizando un marco basado en la teoría de la coalescencia implementado en fastsimcoal2 v2.8 (Excoffier et al., 2021). Este enfoque utiliza el Espectro de Frecuencias Alélicas multidimensional (mSFS) derivado de los datos genómicos para estimar tamaños efectivos poblacionales (N_e), tiempos de divergencia (TDIV) y cambios

históricos en el tamaño poblacional (cuellos de botella y expansiones). El mSFS observado se construyó proyectando las frecuencias alélicas a tamaños de muestra maximizados utilizando la herramienta easySFS (Gutenkunst et al., 2009; código disponible en <https://github.com/isaacovercast/easySFS>, asegurando la retención óptima de sitios polimórficos. Se evaluaron formalmente dos hipótesis biogeográficas contrastantes descritas en la literatura:

- Modelo K=2 (Panmixia parcial): Un escenario que asume una divergencia profunda entre el Océano Atlántico y una única metapoblación global que agrupa de manera panmíctica a las cuencas del Océano Índico y el Océano Pacífico (Occidental y Oriental).
- Modelo K=4 (Subdivisión oceánica profunda): Un escenario de aislamiento estructurado que asume cuatro trayectorias evolutivas y demográficas independientes (Atlántico, Índico, Pacífico Occidental y Pacífico Oriental), con divergencias secuenciales hacia el pasado dictadas por barreras geológicas e hidrográficas.

Ambos modelos fueron parametrizados permitiendo que cada linaje experimentara cambios de tamaño efectivo independientes (eventos de cuello de botella seguidos de expansiones demográficas) antes y después de sus respectivos tiempos de divergencia. Se asumió una tasa de mutación de 7.5×10^{-8} sustituciones por sitio por generación.

Para asegurar una convergencia robusta hacia el máximo global de verosimilitud y evitar óptimos locales, se ejecutaron 100 réplicas independientes por cada escenario. Cada réplica consistió en 20 ciclos de optimización utilizando el algoritmo de Maximización Condicional Iterativa (ECM-

Brent). En cada ciclo, se generaron 10,000 simulaciones coalescentes para aproximar el mSFS esperado y evaluar su ajuste con el mSFS observado. La selección del modelo evolutivo más probable se llevó a cabo utilizando el Criterio de Información de Akaike (AIC), el cual penaliza la adición de parámetros complejos en los modelos (Akaike, 1974). El peso de la evidencia empírica a favor del modelo óptimo se evaluó mediante el diferencial de Akaike (ΔAIC), donde valores de $\Delta AIC > 10$ se interpretaron como soporte estadístico decisivo frente a los modelos alternativos (Burnham & Anderson, 2002).

Una vez seleccionado el modelo biogeográfico con el mejor ajuste empírico, se procedió a estimar la incertidumbre estadística de sus parámetros topológicos y demográficos calculando los intervalos de confianza del 95% (95% CI) mediante un enfoque de bootstrapping paramétrico (Excoffier et al., 2013). Utilizando los parámetros de máxima verosimilitud del modelo ganador, se simularon 100 conjuntos de datos genéticos sintéticos (pseudo-SFS). Posteriormente, se reestimaron los parámetros para cada pseudo-SFS realizando 10 búsquedas independientes bajo las mismas condiciones de optimización originales. Finalmente, para la interpretación biológica conjunta de ambos enfoques (MDA y coalescencia), todos los parámetros de tiempo coalescente obtenidos en generaciones fueron transformados a años cronológicos absolutos asumiendo un tiempo generacional estimado de 25 años para *R. typus* (Castro et al, 2007).

2.3 Resultados

2.3.1 Área de estudio, Extracción de ADN, PCR y secuenciación

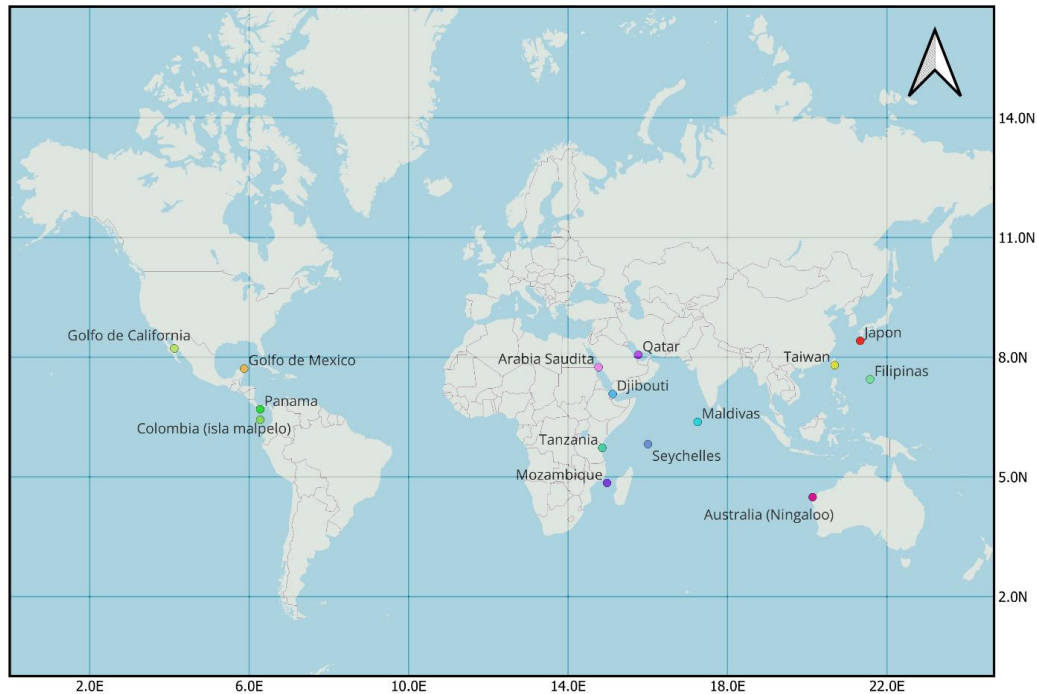
2.3.1.1 Área de estudio y colecta de muestras. Se recolectaron en total, 16 muestras de *Rhincodon typus* dentro del Santuario de Fauna y Flora de Malpelo por parte de la Fundación Malpelo y otros ecosistemas marinos, distribuidas en los cuatro puntos de

muestreo en inmediaciones a la isla Malpelo (Colombia) (Figura 1). Durante la recolecta, se lograron identificar seis juveniles (cinco machos y una hembra) y cuatro adultos (dos machos, una hembra y un sin determinar). Los datos detallados de cada espécimen, incluyendo talla y estadio de desarrollo, se presentan en el Apéndice B.

2.3.1.2 Extracción de ADN, PCR y secuenciación. De las 16 muestras procesadas, se obtuvo una tasa de éxito del 75%, logrando la extracción efectiva de 12 muestras con concentraciones de ADN que oscilaron entre 1.25 ng/uL y 28.2 ng/uL (Apéndice C). Posteriormente, la amplificación mediante PCR y la secuenciación permitieron recuperar fragmentos de la región control mitocondrial entre 661 a 764 pb.

2.3.1.3 Procesamiento Bioinformático. A partir de las 860 secuencias obtenidas de estudios previos, se depuraron 24 registros que presentaban ambigüedades o un tamaño inferior a 590 pb. Al set resultante de 836 secuencias globales se integraron las 12 secuencias correspondientes al Santuario de Fauna y Flora de Malpelo (Colombia), consolidando un alineamiento final de 848 secuencias de diversas localidades mundiales (Figura 2).

Figura 2. Mapa con localidades de las secuencias de región control mitocondrial, incluidas las secuencias de *Rhincodon typus* de estudios previos como las nuevas secuencias de Colombia usadas para este estudio.



2.3.2 Análisis de datos

Se obtuvieron 3 conjuntos de datos que incluyeron las secuencias de región control mitocondrial de los registros procedente de las 15 localidades a nivel mundial (Figura 2). Un primer conjunto de datos de 1,130 pb que contenía la totalidad de los gaps (datos en crudo), un segundo conjunto de datos de 596 pb que contenía los gaps reducidos (alineamiento curado con BMGE) y un último conjunto de datos de 590 pb que no tenía gaps (alineamiento curado con Gblocks).

2.3.2.1 Unidades evolutivas. La reconstrucción filogenética reveló que el tratamiento bioinformático de los datos influyó significativamente en la resolución y el soporte nodal, siendo el conjunto curado con Gblocks (590 pb) el que proporcionó la topología más

robusta y clara. Los resultados de Máxima Verosimilitud (ML) e Inferencia Bayesiana (BI) para los otros conjuntos de datos pueden ser observados en los apéndices D y E, respectivamente, en los cuales se pueden observar topologías menos claras, con grupos que presentaban bajo soporte y definición, además los alineamientos con mayor cantidad de información ambigua, como el set en crudo de 1,130 pb, presentaron dificultades técnicas considerables, al punto de que el análisis mediante Inferencia Bayesiana (BI) no alcanzó la convergencia estocástica tras dos semanas de ejecución. Los filogramas obtenidos mediante Máxima Verosimilitud (ML) e Inferencia Bayesiana (BI) coinciden en identificar una estructuración global definida por la separación del Océano Atlántico como un clado monofilético altamente diferenciado del resto de las regiones mundiales. Esta diferenciación presentó un soporte estadístico decisivo, con valores de probabilidad posterior (BPP=1.0) y soportes de SH-aLRT=98.4 y UFboot=100. (Figuras 3 y 4).

Figura 3. *Árbol filogenético basado en las secuencias de Región Control mitocondrial de tiburón ballena (*Rhincodon typus*), para el conjunto de datos curado con Gblocks, realizado bajo el método de Máxima Verosimilitud (ML).*

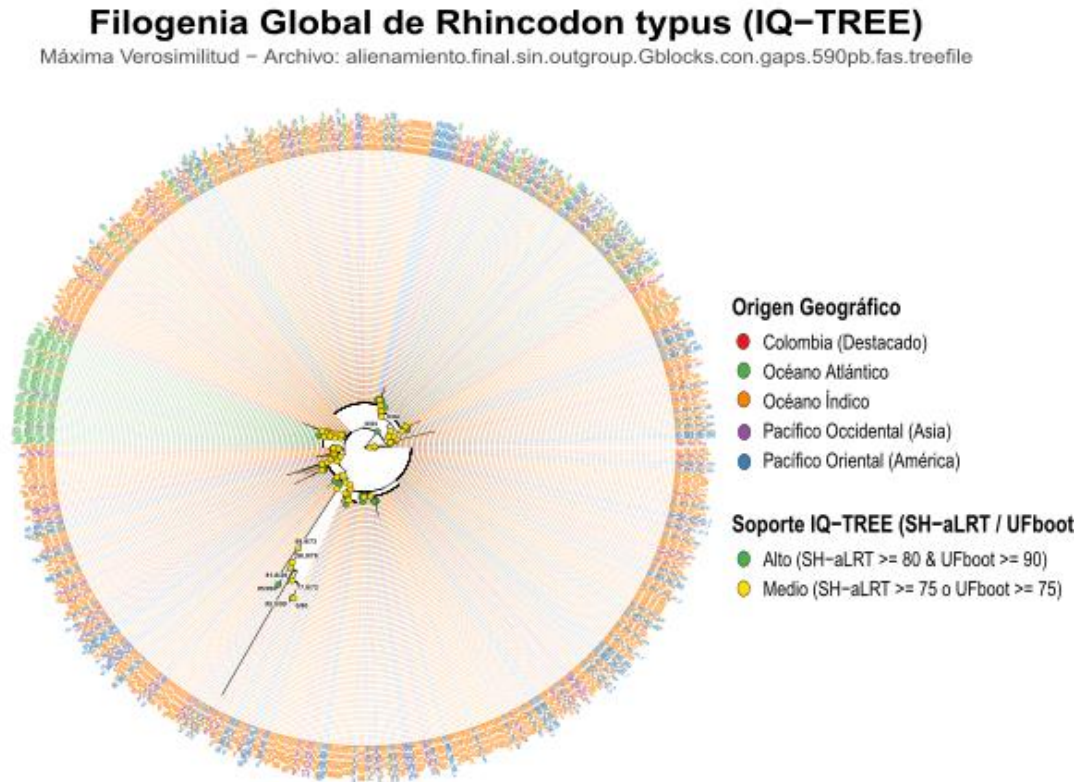
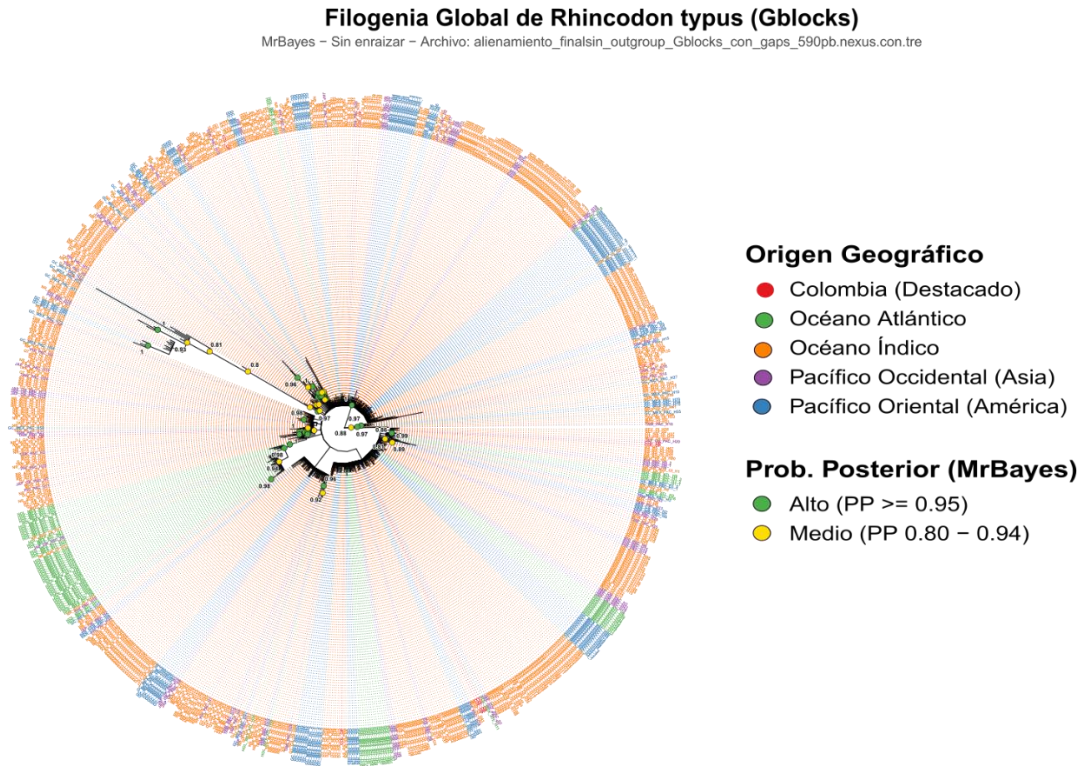
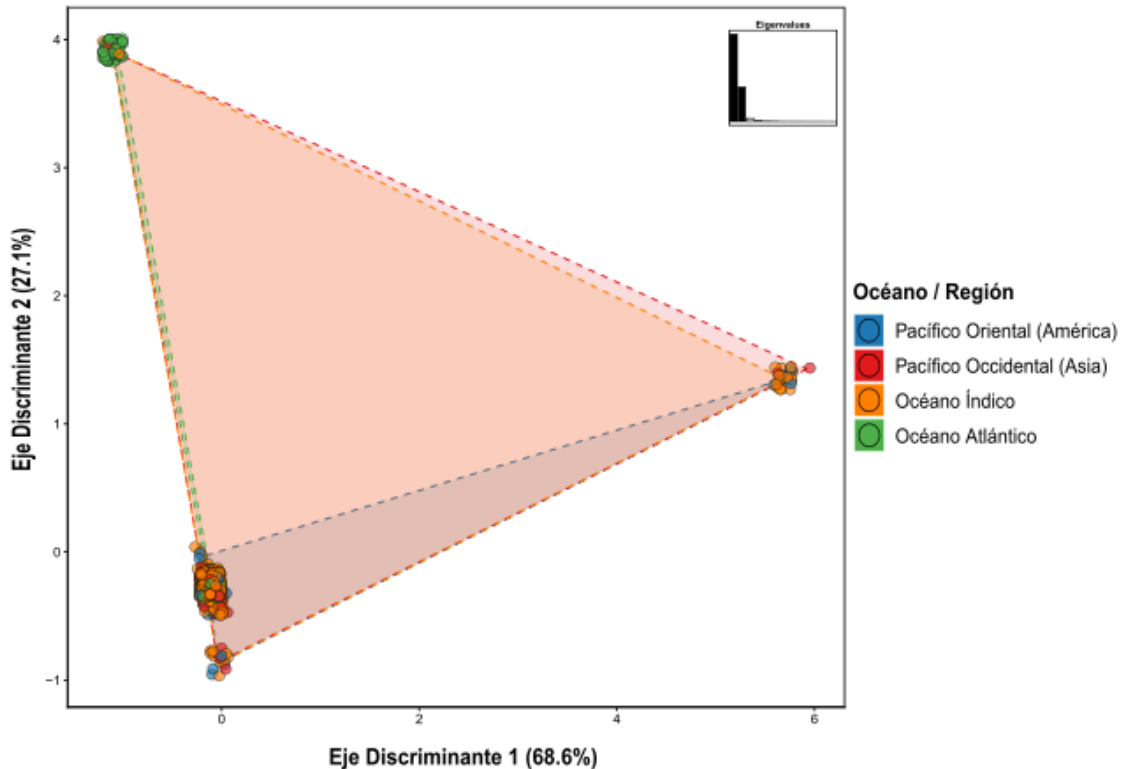


Figura 4. *Árbol filogenético basado en las secuencias Región Control mitocondrial de tiburón ballena (*Rhincodon typus*), para el conjunto de datos curado con Gblocks, realizado bajo el método de Inferencia Bayesiana (BI).*



Por otra parte, la determinación de clústeres genéticos mediante el algoritmo k-means mostro que existen 14 grupos (Apéndice F), sin embargo, el gráfico de dispersión del DAPC muestra que la mayoría de los grupos se solapan, siendo las secuencias del Atlántico las que mejor se agrupan en un eje discriminante más diferenciado que los demás grupos (Figura 5).

Figura 5. Gráfico de dispersión del análisis discriminante de Componentes principales (DAPC) de las secuencias de Región Control mitocondrial de tiburón ballena (*Rhicondon typus*), realizado con el conjunto de datos curado con Gblocks.



Finalmente, los análisis de diferenciación genética entre poblaciones muestran valores muy similares entre los valores los estadísticos Φ_{ST} , G_{ST} y J_{ST} , donde los valores que corresponden a Atlántico frente a las demás poblaciones son altamente significativos para los tres casos (Figura 6 & Apéndice G), mostrando una clara diferenciación de la región del Atlántico frente a las demás regiones, aunque se puede observar un poco de estructura genética adicional dentro de las demás regiones, la cual no es muy clara, pero que podría indicar que hay más de 2 poblaciones diferenciadas a nivel global, las cuales pueden estar agrupadas por cuencas oceánicas (Figura 6).

Figura 6. Valores obtenidos de diferenciación genética entre poblaciones para el estadístico *PhiST*. Los cuadros de color verde indican las agrupaciones del Indo-Pacífico. A. Pacífico Oriental. B. Pacífico Occidental. C. Océano Índico



2.3.2.2 Historia demográfica

2.3.2.2.1 Variación haplotípica y nucleotídica. El análisis del fragmento mitocondrial en las cuatro unidades oceánicas evidenció altos niveles de diversidad haplotípica (Hd) en contraste con valores reducidos de diversidad nucleotídica (π) (Tabla 1). La diversidad haplotípica (Hd) osciló entre 0.8136 en el Océano Atlántico y 0.9604 en el Pacífico Occidental. Por su parte, la diversidad nucleotídica presentó su valor mínimo en el Océano Índico ($\pi=0.00459$, S=140) y el máximo en el Pacífico Occidental ($\pi=0.00615$, S=110).

Las pruebas de neutralidad arrojaron valores negativos para la D de Tajima en todos los grupos evaluados. Estas desviaciones fueron estadísticamente significativas ($p < 0.05$) y consistentes con expansiones demográficas en el Pacífico Occidental ($D = -2.496$), Pacífico Oriental ($D = -2.568$) y Océano Índico ($D = -2.299$). Sin embargo, en el Océano Atlántico el estadístico no alcanzó significancia estadística ($D = -1.742$, $p = 0.082$), sugiriendo una dinámica demográfica distinta o un estado más cercano al equilibrio mutación-deriva. En contraste, el estadístico R2 mostró valores bajos y estadísticamente significativos en las cuatro cuencas oceánicas, variando desde 0.0083 ($p < 0.001$) en el Océano Índico hasta 0.0435 ($p = 0.044$) en el Océano Atlántico (Tabla 1).

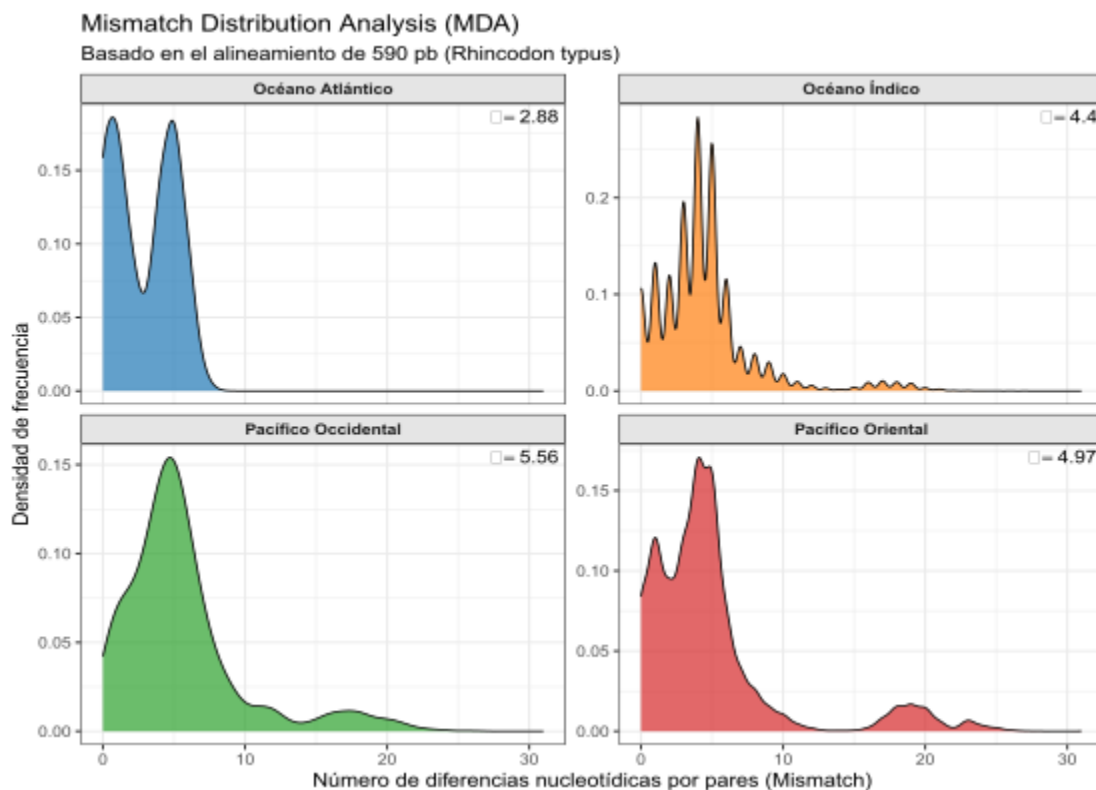
Tabla 1. *Medidas de diversidad genética y de neutralidad mitocondrial*, donde N = tamaño poblacional, S = número de sitios segregantes, Hd = diversidad haplotípica π = diversidad nucleotídica, Tajimas_D = D de Tajima, p_Tajimas = significancia de la D de Tajima, p_R2 = significancia de R2, τ = parámetro mutacional tau, r = índice de rugosidad de Harpending.

Oceano	N	S	Hd	π	Tajimas_D	p_Tajimas	R2	p_R2	τ	r	Expansión_años
Pacífico											
Occidental	87	11	0.9604	0.00615	-2.496	0.013*	0.0329	0.01*	5.562	0.0114	1571078
Océano											
Índico	536	140	0.9372	0.00459	-2.299	0.022*	0.0083	< 0.001***	4.404	0.0203	1244171
Pacífico											
Oriental	145	137	0.9137	0.00546	-2.568	0.01*	0.0209	0.003**	4.966	0.0176	1402878
Océano											
Atlántico	80	32	0.8136	0.00506	-1.742	0.082	0.0435	0.044*	2.882	0.0479	814024

El Análisis de Distribución de Discordancias (MDA) reveló firmas claras de expansión demográfica histórica principalmente en la región del Indo-Pacífico. Las frecuencias de diferencias nucleotídicas pareadas mostraron curvas unimodales clásicas en el Pacífico Occidental, Pacífico Oriental y Océano Índico. En concordancia, el Índice de Rugosidad de Harpending (r) para estas tres cuencas fue notablemente bajo (r entre 0.0114 y 0.0203).

Por el contrario, la distribución de discordancias para el Océano Atlántico exhibió un patrón bimodal (Figura 7), acompañado por el mayor índice de rugosidad de todo el conjunto ($r=0.0479$). La presencia de múltiples picos en la distribución del Atlántico, sumada a la falta de significancia en la D de Tajima, es indicativa de una población estructurada con presencia de linajes mitocondriales divergentes, o de una historia demográfica más estable en comparación con la expansión explosiva evidenciada en las cuencas del Indo-Pacífico.

Figura 7. Resultado del Análisis de Distribución de Discordancias (MDA) para el Océano Atlántico, el Océano Índico, el Pacífico Occidental y el Pacífico Oriental.



Las estimaciones del parámetro mutacional τ para los clados con soporte de expansión (Indo-Pacífico) situaron el inicio de su crecimiento poblacional durante el Pleistoceno temprano (edad

Calabriense), desde hace ~1.24 millones de años en el Índico hasta ~1.57 millones de años en el Pacífico Occidental.

2.3.2.2.2 Teoría Coalescente. La evaluación de las hipótesis biogeográficas mediante inferencia coalescente basada en el Espectro de Frecuencias Alélicas multidimensional (mSFS) proporcionó un soporte estadístico decisivo a favor de la subdivisión oceánica profunda. El modelo $K=4$, que asume cuatro trayectorias demográficas independientes (Océano Atlántico, Océano Índico, Pacífico Occidental y Pacífico Oriental), mostró el mejor ajuste a los datos empíricos. Este modelo superó ampliamente al modelo de panmixia parcial ($K=2$), obteniendo el valor más bajo del Criterio de Información de Akaike ($AIC = 467.70$) y un diferencial ΔAIC de 0.00. Por el contrario, el escenario que agrupaba al Indo-Pacífico como una única metapoblación ($K=2$) recibió una penalización severa y un soporte nulo ($\Delta AIC = 26.32$), permitiendo rechazar formalmente la hipótesis de conectividad global reciente.

Bajo el modelo óptimo ($K=4$), la estimación de los parámetros históricos reveló una divergencia secuencial profunda. Asumiendo un tiempo generacional de 25 años, la separación más antigua correspondió al aislamiento del linaje del Océano Atlántico respecto al clado ancestral Indo-Pacífico, un evento datado hace aproximadamente 381,325 años [95% CI: 4,325 - 9,545,372]. Posteriormente, la cuenca del Océano Índico y el Océano Pacífico divergieron hace 305,700 años [95% CI: 3,050 - 7,413,825]. Finalmente, la separación más reciente ocurrió dentro de la cuenca del Pacífico, donde las poblaciones del Pacífico Occidental y Oriental se dividieron hace apenas 20,850 años [95% CI: 1,775 - 4,942,550], un periodo que coincide cronológicamente con el Último Máximo Glacial (LGM).

Las inferencias del tamaño efectivo poblacional (N_e) mostraron trayectorias demográficas altamente contrastantes entre las cuencas oceánicas en el presente. Las mayores estimaciones de N_e se observaron en el Pacífico Occidental (862,473 individuos [95% CI: 108 - 1,642,918]) y el Océano Atlántico (792,029 individuos [95% CI: 735,480 - 835,480]), seguidas por el Océano Índico (149,325 individuos [95% CI: 109 - 4,016,078]) (Tabla 2). En drástico contraste, el Pacífico Oriental exhibió un N_e contemporáneo marcadamente reducido de solo 946 individuos [95% CI: 108 - 1,642,918].

El modelo genómico también capturó firmas de variaciones drásticas en el tamaño poblacional a lo largo del tiempo, concordantes con las firmas ancestrales detectadas en el marcador mitocondrial. Se infirió un tamaño poblacional ancestral (N_{ANC}) de 1,463 individuos [95% CI: 109 - 4,016,078] (Tabla 2) previo a las divergencias oceánicas. Tras su separación, los linajes principales experimentaron periodos de cuellos de botella seguidos de expansiones demográficas asimétricas. La reducción del tamaño poblacional en el Pacífico Oriental, acoplada a su reciente tiempo de divergencia, sugiere fuertemente un efecto fundador post-glaciación que dio origen a esta unidad demográfica marginal. Las estimaciones completas de todos los parámetros demográficos, incluyendo tiempos de expansión y severidad de los cuellos de botella, se detallan en la Tabla 1 y se ilustran de manera esquemática en la Figura 8.

Figura 8. Modelo $K=4$ definido como el modelo óptimo por inferencia coalescente basada en el Espectro de Frecuencias Alélicas multidimensional (mSFS).

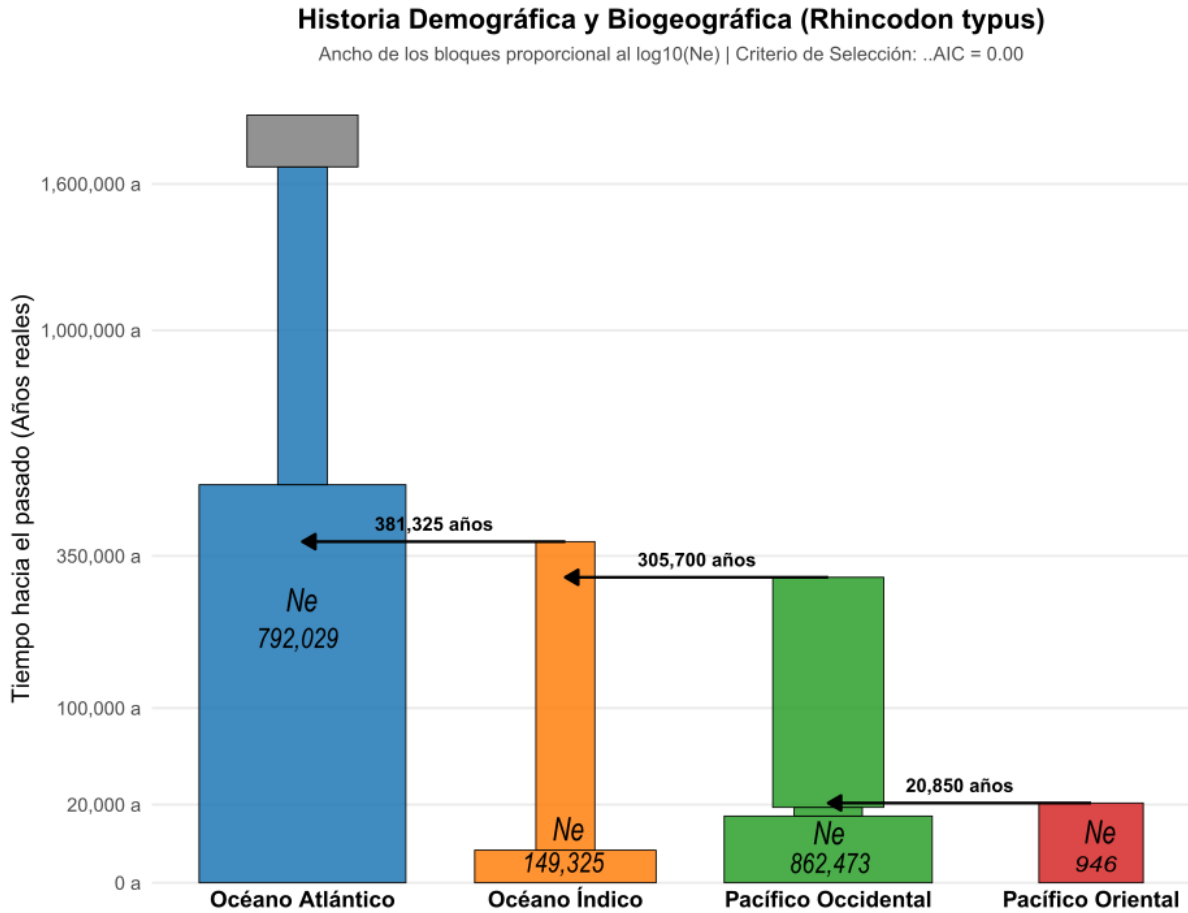


Tabla 2. *Estimaciones de tamaños poblacionales efectivos (N_e) y ancestral (N_{ANC}) del tiburón ballena por cuenca oceánica, donde N_e = Tamaño efectivo poblacional, IC = Intervalo de confianza.*

Región/Parámetro	N_e (individuos)	IC (95%)
Océano Atlántico	792,029	735,480 - 835,480
Océano Indico	149,325	109 - 4,016,078
Pacífico Oriental	946	108 - 1,642,918
Pacífico Occidental	862,473	108 - 1,642,918
NANC (Ancestral)	1,463	109 - 4,016,078

2.4 Discusión

El presente estudio aporta las primeras secuencias de la Región Control de ADNmt de tiburón ballena (*R. typus*) para el sitio de agregación ubicado en el Santuario de Fauna y Flora Malpelo, constituyendo los primeros registros moleculares para Colombia y para la región suroccidental del Océano Pacífico. Por este motivo, los resultados aquí obtenidos representan un avance significativo para llenar los vacíos de información en el Pacífico Occidental, contribuyendo al entendimiento global y regional para la conservación de la especie.

Las reconstrucciones filogenéticas y los análisis de agrupamiento (DAPC) evidenciaron una diferenciación marcada entre la cuenca del Atlántico y las demás regiones oceánicas del mundo, respaldada por altos valores de soporte en el nodo divergente. Si bien la monofilia del clado atlántico no fue absoluta —dado que unas pocas secuencias del Atlántico se agruparon con linajes del Indo-Pacífico y viceversa—, la separación en un eje discriminante distinto confirma su

aislamiento predominante. Este patrón es congruente con investigaciones previas que sugieren una conectividad sumamente restringida entre el Atlántico y el resto del mundo (Castro et al., 2007). Esta restricción parece estar dictada por barreras geográficas y térmicas, vinculadas a la preferencia ecológica de la especie por aguas tropicales y templado-cálidas (Rowat & Brooks, 2012; Wilson et al., 2006). Específicamente, el intercambio esporádico de individuos entre el Índico y el Atlántico está fuertemente limitado por la corriente fría de Benguela en la costa occidental de Sudáfrica (Gibbons & Thibault-Botha, 2002), un sistema de afloramiento donde las aguas frías superficiales impiden el comportamiento de termorregulación de la especie en inmersiones profundas (Castro et al., 2007).

La falta de monofilia completa del Atlántico puede explicarse por eventos históricos de conectividad interoceánica. Por ejemplo, la intrusión ocasional de aguas cálidas del Índico hacia el Atlántico a través de las dinámicas de la corriente de Benguela, la cual inmediatamente después de cada era de hielo recibe una entrada de agua cálida debió a la surgencia de las aguas frías (Peeters et al., 2004), podría facilitar intercambios esporádicos. Además, los giros del núcleo cálido del Océano Índico que se ven ocasionalmente arrastrados por la corriente de Benguela, y que se desplaza hacia el norte para desembocar en el Océano Atlántico central (Castro et al., 2007), también facilitarían estos intercambios. Sin embargo, estos eventos distantes serían insuficientes para contrarrestar los efectos de la deriva genética, manteniendo así la diferenciación observada en los resultados obtenidos.

A una escala más fina, la diferenciación genética evaluada mediante los estadísticos PhiST, Jost's D y GST reveló una estructura geográfica adicional. Aunque estudios previos (Vignaud et al., 2014; Yagishita et al., 2020; Hardenstine et al., 2022) habían descartado variaciones subpoblacionales dentro del Indo-Pacífico atribuyéndolas a sensibilidades estadísticas por tamaños

de muestra desbalanceados, la convergencia de los tres índices —incluyendo a Jost's D, que es altamente robusto ante disparidades muestrales (Barbosa et al., 2018)— sugiere que esta leve estructura es un fenómeno biológico real y no un artefacto metodológico. Esta sutil partición genética es consistente con las observaciones recientes de Saavedra Valderrama et al. (2025), quienes reportaron un mayor número de pasos mutacionales en los individuos del Pacífico Oriental respecto a las cuencas del Índico y Pacífico Occidental. Considerando el lento ritmo de mutación inferido para *R. typus* y la necesidad de efectos acumulativos a gran escala temporal para fijar diferencias genéticas (Junge et al., 2011; Saavedra Valderrama et al., 2025), se propone que la estructura genética débilmente detectada en el Indo-Pacífico refleja las etapas iniciales de una especiación incipiente o diferenciación poblacional impulsada por aislamiento por distancia y barreras oceanográficas difusas.

Esta hipótesis de subdivisión fue robustamente respaldada por la inferencia coalescente basada en el Espectro de Frecuencias Alélicas (mSFS). El ajuste óptimo al modelo de cuatro trayectorias demográficas independientes ($K=4$) apoya la visión de un Pacífico Oriental demográficamente aislado del resto de las cuencas, coincidiendo con Saavedra Valderrama et al. (2025). Este hallazgo ofrece un contraste estadístico formal frente a los escenarios de alta conectividad o agrupaciones panmícticas globales descritas en estudios anteriores (Castro et al., 2007; Vignaud et al., 2014; Yagishita et al., 2020; Hardenstine et al., 2022). En conjunto, estos resultados sugieren que la estructura poblacional del tiburón ballena es más compleja de lo previamente reconocido, con una diferenciación que podría estar organizada a nivel de cuencas oceánicas.

Las inferencias del tamaño efectivo poblacional (N_e) corroboran lo postulado por Vignaud et al. (2014) respecto a que el tamaño contemporáneo de la especie supera ampliamente al

ancestral. Se estimó un tamaño ancestral (NANC) de ~1,463 individuos, el cual fue ampliamente superado por los linajes del Pacífico Occidental, Atlántico e Índico tras procesos asimétricos de expansión. De manera notable, el Pacífico Oriental fue la única unidad demográfica que exhibió un N_e contemporáneo marcadamente reducido (~946 individuos), lo cual, acoplado a su reciente tiempo de divergencia, sugiere fuertemente que esta población periférica es el resultado de un efecto fundador post-glaciación.

Las señales de expansión demográfica histórica fueron inequívocas para las regiones del Pacífico Occidental, Oriental e Índico. Este patrón es inferido por los altos niveles de diversidad haplotípica (H_d) y valores reducidos de diversidad nucleotídica (π) (Saavedra Valderrama et al, 2025), además de valores significativamente negativos en las pruebas de neutralidad (Tajima's D y R_2), y distribuciones de discordancias (MDA) unimodales con bajos índices de rugosidad (r). La expansión de las poblaciones del Indo-Pacífico ya había sido registrada en la literatura (Vignaud et al., 2014; Yagishita et al., 2020; Hardenstine et al., 2022), las cuales podrían seguir creciendo bajo las condiciones climáticas actuales. Por el contrario, el linaje del Océano Atlántico presentó un patrón bimodal, un mayor índice de rugosidad y un D de Tajima no significativo. Este comportamiento diferencial es congruente con los reportes de Yagishita et al. (2020) y Hardenstine et al. (2022), indicando que el Atlántico alberga una población estructurada con linajes mitocondriales divergentes y una historia demográfica sustancialmente más estable.

A pesar de los claros reportes documentando que los tiburones ballena han sido objeto de explotación intensiva durante el último siglo (Vignaud et al., 2014), lo que ha provocado disminuciones poblacionales estimadas en >50% en el Indo-Pacífico y >30% en el Atlántico durante las últimas tres generaciones (Pierce & Norman, 2016), los datos genéticos aquí obtenidos no revelaron señales de cuellos de botella recientes. Esta aparente discrepancia genético-

demográfica es esperable; estudios en otras especies longevas y amenazadas han demostrado que los tiempos generacionales prolongados actúan como un efecto amortiguador que retrasa significativamente el ritmo de erosión de la diversidad genética (Lippé et al., 2006). Adicionalmente, la conectividad residual entre subpoblaciones puede facilitar la recuperación o el mantenimiento de la diversidad tras reducciones repentinas (Jangjoo et al., 2016). En este sentido, se coincide con Vignaud et al. (2014) y Yagishita et al. (2020) en que la ausencia de cuellos de botella detectables en el ADN mitocondrial de *R. typus* responde primordialmente a su largo tiempo generacional (~25 años) y a la conectividad histórica dentro de cada gran cuenca oceánica.

Las estimaciones temporales del parámetro τ situaron el inicio del crecimiento explosivo de las poblaciones del Indo-Pacífico durante el Pleistoceno temprano (hace ~1.24 a ~1.57 millones de años), coincidiendo cronológicamente con las reconfiguraciones oceanográficas de las primeras grandes glaciaciones del Cuaternario. Asimismo, las divergencias poblacionales estimadas mediante el modelo coalescente muestran una notable congruencia temporal con los ciclos glaciales de esta época. La separación basal entre el Atlántico y el clado ancestral Indo-Pacífico (hace ~381,000 años) y la subsecuente división del Índico y el Pacífico (hace ~305,000 años) ocurrieron durante el Pleistoceno medio, un periodo caracterizado por intensas fluctuaciones climáticas globales. Finalmente, la divergencia más reciente que aisló al Pacífico Oriental (hace ~20,850 años) coincide directamente con el Último Máximo Glacial (LGM).

3. Conclusiones

Este estudio se constituye como una de las evaluaciones más exhaustivas hasta la fecha sobre la estructura genética e historia demográfica del tiburón ballena (*Rhincodon typus*) a nivel global. La inclusión de nuevas secuencias provenientes de Colombia (Santuario de Fauna y Flora Malpelo) y Panamá llena un vacío geográfico crítico, ampliando significativamente la resolución filogeográfica en el Pacífico Oriental.

Los árboles filogenéticos construidos por medio de Inferencia Bayesiana y Máxima Verosimilitud mostraron una clara separación de la población del Atlántico frente a las demás localidades del Indo-Pacífico. Este hallazgo es congruente con los resultados de los análisis de diferenciación genética, donde el mayor valor de aislamiento se le asigna al Atlántico. Sin embargo, las métricas de diferenciación también revelaron una estructura genética subyacente dentro del Indo-Pacífico, desafiando la noción tradicional de una panmixia global en esta vasta región oceánica. Esta subdivisión fue corroborada estadísticamente mediante inferencia coalescente. Al evaluar formalmente modelos de 2 y 4 poblaciones, el escenario que asumía cuatro trayectorias evolutivas y demográficas independientes (Atlántico, Índico, Pacífico Occidental y Pacífico Oriental) fue el que demostró un ajuste muy superior a los datos empíricos, redefiniendo las unidades demográficas de la especie.

Tanto los análisis clásicos de genómica poblacional (distribución de discordancias y neutralidad) como los basados en teoría coalescente coinciden en demostrar la expansión poblacional de las regiones del Indo-Pacífico (Océano Índico, Pacífico Occidental y Pacífico Oriental), consolidando los reportes de estudios previos (Vignaud et al., 2014; Yagishita et al., 2020; Hardenstine et al., 2022). Por el contrario, la población del Atlántico exhibió firmas genéticas características de una población estable y con una historia demográfica marcadamente

diferente, un patrón que respalda las observaciones de Yagishita et al. (2020) y Hardenstine et al. (2022).

Finalmente, se demuestra que la arquitectura poblacional contemporánea del tiburón ballena está intrínsecamente ligada a los ciclos climáticos del Cuaternario. Mientras que la gran expansión del Indo-Pacífico se originó en el Pleistoceno temprano, las divergencias oceánicas secuenciales —y particularmente el reciente efecto fundador en el Pacífico Oriental— coinciden cronológicamente con las barreras oceanográficas impuestas por los diferentes eventos de glaciación ocurridos durante el Pleistoceno.

4. Recomendaciones

Se sugiere priorizar el esfuerzo de muestreo en regiones geográficas subrepresentadas, particularmente añadiendo individuos de otras localidades del Pacífico Oriental (ej. costas de Sudamérica), del Pacífico Sur y sectores inexplorados del Atlántico. Esto permitirá mapear con mayor precisión las zonas de transición o contacto secundario para seguir entendiendo las dinámicas actuales e históricas de las poblaciones del tiburón ballena.

Adicionalmente, dado que las inferencias demográficas de este estudio se fundamentan en un marcador de herencia materna, resulta fundamental que futuras investigaciones transiten hacia enfoques de genómica poblacional (como secuenciación de representación reducida [ddRADseq] o genomas completos). La incorporación de miles de marcadores nucleares permitirá capturar el flujo genético mediado por machos, refinar los intervalos de confianza de las divergencias del Pleistoceno, y evaluar con mayor resolución el impacto de las presiones antropogénicas recientes sobre el tamaño efectivo poblacional.

Referencias Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Barbosa, L. M., Barros, B. C., de Souza Rodrigues, M., Silva, L. K., Reis, M. G., & Blanton, R. E. (2018). *The effect of sample size on estimates of genetic differentiation and effective population size for schistosoma mansoni populations*. *International Journal for Parasitology*, 48(14), 1149–1154. <https://doi.org/10.1016/j.ijpara.2018.10.001>.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.
- Casilagan, I. L. N., Junio-Meñez, M. A., & Crandall, E. D. (2013). *Genetic diversity, population structure, and demographic history of exploited sea urchin populations (tripneustes gratilla) in the philippines*. *Journal of Experimental Marine Biology and Ecology*, 449, 284-293. doi: 10.1016/j.jembe.2013.09.012.
- Castro, A. L. F., Stewart, B. S., Wilson, S. G. et al. (2007). *Population genetic structure of Earth's largest fish, the whale shark (Rhincodon typus)*. *Molecular Ecology*, 16(24), 5183–5192. doi:10.1111/j.1365-294x.2007.03597.x.
- Compagno, L. J. V.(2001). *Sharks of the world: An annotated and illustrated catalogue of shark species known to date*. Vol. 2. Bullhead, mackerel and carpet sharks. FAO.
- Castresana, J. (2000). *Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis*, *Molecular Biology and Evolution*, Volume 17, Issue 4, April 2000, Pages 540–552, <https://doi.org/10.1093/oxfordjournals.molbev.a026334>

- Criscuolo, A., Gribaldo, S. (2010). *BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments*. BMC Evol Biol 10, 210. <https://doi.org/10.1186/1471-2148-10-210>
- Eckert, S.A & B. Stewart. (1996). *Migration and movements of the whale shark (Rhincodon typus) in the Sea of Cortez as determined by satellite telemetry*. Hubbs-Sea World Res. Inst., Tech. Rep., 96-269, pp. 1-22, figs 1-10, tabs 1-2, app. 1-3.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). *Robust demographic inference from genomic and SNP data*. PLoS Genetics, 9(10), e1003905.
- Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). *fastsimcoal2: demographic inference under complex evolutionary scenarios*. Bioinformatics, 37(24), 4882-4885.
- Galván-Tirado, C., Díaz-Jaimes, P., García-de León, F. J., Galván-Magaña, F., & Uribe-Alcocer, M. (2013). *Historical demography and genetic differentiation inferred from the mitochondrial DNA of the silky shark (carcharhinus falciformis) in the pacific ocean*. Fisheries Research, 147, 36-46. doi:10.1016/j.fishres.2013.03.020.
- Gibbons MJ, Thibault-Botha D (2002). *The match between ocean circulation and zoogeography of epipelagic siphonophores around southern Africa*. Journal of the Marine Biological Association of the UK, 82, 801–810.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). *Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data*. PLoS Genetics, 5(10), e1000695.
- Guzmán H., Beaver C., Díaz-Ferguson E. (2021). *Novel Insights Into the Genetic Population Connectivity of Transient Whale Sharks (Rhincodon typus) in Pacific Panama Provide*

- Crucial Data for Conservation Efforts*. *Frontiers in Marine Science*. vol 8. ISSN 2296-7745. <https://doi.org/10.3389/fmars.2021.744109>.
- Hall, T.A. (1999) *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. *Nucleic Acids. Symp. Ser.*, 41, pp. 95-98.
- Hardenstine RS, He S, Cochran JEM, Braun CD, Cagua EF, Pierce SJ, Prebble CEM, Rohner CA, Saenz-Angudelo P, Sinclair-Taylor TH, Skomal GB, Thorrold SR, Watts AM, Zakroff CJ, Berumen ML. (2022). *Pieces in a global puzzle: Population genetics at two whale shark aggregations in the western Indian Ocean*. *Ecol Evol*. Jan 25;12(1):e8492. doi: 10.1002/ece3.8492. Erratum in: *Ecol Evol*. 2022 Mar 23;12(3):e8727. PMID: 35127024; PMCID: PMC8796955.
- Harpending, H. (1994). *Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution*. *Human Biology*, 66(4), 591-600.
- Hoffmayer E., McKinney J., Franks J., Hendon J., Driggers W., Falterman B., Galuardi B, Byrne M. (2021). *Seasonal Occurrence, Horizontal Movements, and Habitat Use Patterns of Whale Sharks (Rhincodon typus) in the Gulf of Mexico*. *Frontiers in Marine Science*. Vol 7. ISSN 2296-7745. <https://doi.org/10.3389/fmars.2020.598515>.
- Hueter RE, Tyminski JP, de la Parra R (2013). *Horizontal Movements, Migration Patterns, and Population Structure of Whale Sharks in the Gulf of Mexico and Northwestern Caribbean Sea*. *PLOS ONE* 8(8): e71883. <https://doi.org/10.1371/journal.pone.0071883>.
- Jangjoo, M., S. E Matter, J. Roland, and N. Keyghobadi. (2016). *Connectivity rescue genetic diversity after a demographic bottleneck in a butterfly population network*. *Proc. Natl. Acad. Sci. U. S. A.* 113:10914-10919.

- Junge, C., Vøllestad, L., Barson, N. et al. (2011). *Fuerte flujo genético y falta de estructura poblacional estable ante la rápida adaptación a la temperatura local en un salmónido de desove primaveral, el timalo europeo (Thymallus thymallus)*. Heredity 106, 460–471. <https://doi.org/10.1038/hdy.2010.160>
- Kuhner, M. K. (2009). *Coalescent genealogy samplers: Windows into population history*. Trends in Ecology & Evolution, 24(2), 86-93. doi: 10.1016/j.tree.2008.09.007.
- Lemoine, F. and Correia, D. and Lefort, V. and Doppelt-Azeroual, O. and Mareuil, F. and Cohen-Boulakia, S. and Gascuel, O. (2019). *NGPhylogeny.fr: new generation phylogenetic services for non-specialists*. Nucleic acids research, 47:W260-W265
- Lippé, C., P. Dumont, and L. Bernatchez. (2006). *High genetic diversity and no inbreeding in the endangered copper redhorse, Moxostoma hubbsi (Catostomidae, Pisces): the positive sides of a long generation time*. Mol. Ecol. 15:1769-1780.
- Manunza, A., Ramírez-Díaz, J., Cozzi, P. et al. (2023). *Diversidad genética y demografía histórica de las razas caprinas infrautilizadas en el noroeste de Europa*. Sci Rep 13, 20728. doi:10.1038/s41598-023-48005-8.
- Martin, A. P., Naylor, G. J., & Palumbi, S. R. (1992). *Rates of mitochondrial DNA evolution in sharks are slow compared with mammals*. Nature, 357(6374), 153–155. <https://doi.org/10.1038/357153a0>.
- Mazet, O., Noûs, C. (2023). *Population genetics: coalescence rate and demographic parameters inference*. Peer Community Journal, Volume 3, article no. e53. doi:10.24072/pcjournal.285.
- Meekan M, Austin CM, Tan MH, Wei N-WV, Miller A, Pierce SJ, Rowat D, Stevens G, Davies TK, Ponzo A and Gan HM (2017). *iDNA at Sea: Recovery of Whale Shark (Rhincodon*

- typus*) Mitochondrial DNA Sequences from the Whale Shark Copepod (*Pandarus rhincodonicus*) Confirms Global Population Structure. *Front. Mar. Sci.* 4:420. doi: 10.3389/fmars.2017.00420
- Mohd Yusoff, N.I.S., Mat Jaafar, T.N.A., Vilasri, V. et al. (2021) *Diversidad genética, estructura poblacional y demografía histórica del astrónomo de cola amarilla de dos espinas (Uranoscopus cognatus)*. *Sci Rep* 11, 13357. <https://doi.org/10.1038/s41598-021-92905-6>.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Oxford Nanopore Technologies PLC. (2023). *Dorado*. Available online: <https://github.com/nanoporetech/dorado>
- Oxford Nanopore Technologies PLC. (2024). *EPI2ME workflows: wf-amplicon*. Available online: <https://github.com/epi2me-labs/wf-amplicon>.
- Paradis, E. (2010). *pegas: an R package for population genetics with an integrated-modular approach*. *Bioinformatics*, 26(3), 419-420.
- Paradis, E., Claude, J., & Strimmer, K. (2004). *APE: Analyses of Phylogenetics and Evolution in R language*. *Bioinformatics*, 20(2), 289-290.
- Parsons, T. J., Muniec, D. S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M. R., Berry, D. L., Holland, K. A., Weedn, V. W., Gill, P., & Holland, M. M. (1997). *A high observed substitution rate in the human mitochondrial DNA control region*. *Nature Genetics*, 15(4), 363–368. <https://doi.org/10.1038/ng0497-363>.
- Pereyra, S., García, G., Miller, P., Oviedo, S., & Domingo, A. (2010). *Low genetic diversity and population structure of the narrownose shark (mustelus schmitti)*. *Fisheries Research*, 106(3), 468-473. doi: 10.1016/j.fishres.2010.09.022.

- Pierce, S. J., and B. Norman. (2016). *Rhincodon typus*. The IUCN red list of the threatened species 2016:eT19488A2365291, [http://dx. doi.org/10.2 305/IUCN.UK.2016-1 .RLTS. T19488A2365291.en](http://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T19488A2365291.en).
- Pierce, S.J., Rohner, C.A., Perry, C.T., Jabado, R.W., Norman, B., Reynolds, S., Womersley, F., Robinson, D., Graham, R. & Araujo, G. (2025). *Rhincodon typus*. *The IUCN Red List of Threatened Species* 2025: e.T19488A126673248. <https://dx.doi.org/10.2305/IUCN.UK.2025-2.RLTS.T19488A126673248.en>.
- Ramos-Onsins, S. E., & Rozas, J. (2002). *Statistical properties of new neutrality tests against population growth*. *Molecular Biology and Evolution*, 19(12), 2092-2100.
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rogers, A. R., & Harpending, H. (1992). *Population growth makes waves in the distribution of pairwise genetic differences*. *Molecular Biology and Evolution*, 9(3), 552-569.
- Ronquist F, Huelsenbeck JP. (2003) *MrBayes 3: Bayesian phylogenetic inference under mixed models*. *Bioinformatics*. 19(12):1572–4. doi:10.1093/ bioinformatics/btg180.
- Rosenberg, N., Nordborg, M. (2002) *Árboles genealógicos, teoría coalescente y análisis de polimorfismos genéticos*. *Nat Rev Genet* 3, 380–390. <https://doi.org/10.1038/nrg795>.
- Rowat, D. y Brooks, KS (2012). *Una revisión de la biología, la pesca y la conservación del tiburón ballena Rhincodon typus*. *Journal of Fish Biology*, 80, 1019–1056. 10.1111/j.1095-8649.2012.03252.x
- Saavedra Valderrama, N. A., Marchant Rojas, S.A., Barragán Barrera, D. (2025). *Diversidad y estructura genética del tiburón ballena (Rhincodon typus) estimada a partir de ADN*

mitochondrial. Universidad Industrial de Santander.

<https://noesis.uis.edu.co/handle/20.500.14071/46435>.

Schmidt JV, Schmidt CL, Ozer F, Ernst RE, Feldheim KA, Ashley MV, et al. (2009) *Low Genetic Differentiation across Three Major Ocean Populations of the Whale Shark, Rhincodon typus*. PLoS ONE 4(4): e4988. <https://doi.org/10.1371/journal.pone.0004988>.

Spence, J. P., Steinrücken, M., Terhorst, J., & Song, Y. S. (2018). *Inference of population history using coalescent HMMs: Review and outlook*. Current Opinion in Genetics & Development, 53, 70-76. doi:10.1016/j.gde.2018.07.002.

Stanhope, M. J., Ceres, K. M., Sun, Q., Wang, M., Zehr, J. D., Marra, N. J., Shivji, M. S. (2023). *Genomes of endangered great hammerhead and shortfin mako sharks reveal historic population declines and high levels of inbreeding in great hammerhead*. iScience, 26(1), 105815. doi:10.1016/j.isci.2022.105815.

Stevens, J. D. (2007). *Whale shark (Rhincodon typus) biology and ecology: A review of the primary literature*. Fisheries Research, 84(1), 4-9. doi:10.1016/j.fishres.2006.11.008.

Stewart, B.S., Wilson, S.G. Threatened. (2005). *Fishes of the World: Rhincodon typus (Smith 1828) (Rhincodontidae)*. Environ Biol Fish 74, 184–185. <https://doi.org/10.1007/s10641-005-2229-1>.

Tajima, F. (1989). *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism*. Genetics, 123(3), 585-595.

Thompson, D. Higgins, T. Gibson .(1994). *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic Acids Res., 22, pp. 4673-4680.

- T.K.F. Wong, N. Ly-Trong, H. Ren, H. Baños, A. Roger, E. Susko, C. Bielow, N. De Maio, N. Goldman, M.W. Hahn, G. Huttley, R. Lanfear, B.Q. Minh (2025) *IQTREE 3: Phylogenomic Inference Software using Complex Evolutionary Models*. *ecoevorxiv*. <https://ecoevorxiv.org/repository/view/8916/>
- Vignaud TM, Maynard JA, Leblois R, Meekan MG, Vázquez-Juárez R, Ramírez-Macías D, Pierce SJ, Rowat D, Berumen ML, Beeravolu C, Baksay S, Planes S. (2014) *Genetic structure of populations of whale sharks among ocean basins and evidence for their historic rise and recent decline*. *Molecular Ecology*, 23(10), 2590–2601. <https://doi.org/10.1111/mec.12754>.
- Wang, L., Shi, X., Su, Y., Meng, Z., & Lin, H. (2013). *Genetic divergence and historical demography in the endangered large yellow croaker revealed by mtDNA*. *Biochemical Systematics and Ecology*, 46, 137-144. doi: 10.1016/j.bse.2012.09.021.
- Weber, J. A., Park, S. G., Luria, V., Jeon, S., Kim, H. M., Jeon, Y., Bhak, Y., Jun, J. H., Kim, S. W., Hong, W. H., Lee, S., Cho, Y. S., Karger, A., Cain, J. W., Manica, A., Kim, S., Kim, J. H., Edwards, J. S., Bhak, J., & Church, G. M. (2020). *The whale shark genome reveals how genomic and physiological properties scale with body size*. *Proceedings of the National Academy of Sciences of the United States of America*, 117(34), 20662–20671. <https://doi.org/10.1073/pnas.1922576117>
- Wilson SG, Polovina JJ, Stewart BS, Meekan MG (2006) *Movement of whale sharks (Rhincodon typus) tagged at Ningaloo Reef, Western Australia*. *Marine Biology*, 148, 1157–1166.
- Yagishita, N., Ikeguchi, S., & Matsumoto, R. (2020). *Re-estimation of genetic population structure and demographic history of the whale shark (Rhincodon typus) with additional japanese*

samples, inferred from mitochondrial DNA sequences 1. Pacific Science, 74(1), 31-47,47A-47B. <https://doi.org/10.2984/74.1.3>.

Apéndices

Apéndice A. Información sobre las secuencias de tiburón ballena (*Rhincodon typus*) obtenidas de GenBank. Las columnas indican la cantidad de secuencias obtenidas para cada región de cada uno de los estudios realizados anteriormente. La última columna indica los códigos de acceso del GenBank de las secuencias, y la última fila indica el total general de secuencias obtenidas además del total de secuencias para cada región. N/A= No están en GenBank, secuencias de acceso libre en formato Excel.

	TOTAL	Arabia Saudita (AS)	Tanzania (TAN)	Panamá (PAN)	Australia (AUS)	México Pacífico (MEX_PAC) Golfo Califor	Filipinas (PHI)	Taiwan (TAW)	Mozambique (MOZ)	Maldivas (MAL)	México Atlántico (MEX_ATL) Quin Roo	Japón (JAP)	Djibouti (DJ)	Seychelles (SEY)	Qatar (QT)	Código acceso genbank
Hardenstine et al (2022)	117	60	57													OL782199 - OL782316
Guzmán, Beaver & Díaz-Ferguson, (2021)	15			15												OK094076 - OK094090
Castro et al (2007)	70				12	8	12		18	1	19					EU182401 - EU18244
Ramírez-Macías et al (2007)	36					36										DQ395261 - DQ 395274
Yagishita, Ikeguchi & Matsumoto (2020)	28											28				LC466742 - LC466769
Vignaud et al (2014)	506	26			146	76	21	25	33	9	61		78	31		N/A
Sisgaard et al (2016)	61														61	KX944487 - KX944547
Meehan et al (2017)	27				5				13	2				7		MF872682 - MF872725
TOTAL	860	86	57	15	163	120	33	25	64	12	80	28	78	38	61	

Apéndice B. Datos obtenidos durante la colecta de muestras de tiburón ballena (*Rhincodon typus*) dentro del Santuario de Fauna y Flora de Malpelo para los 4 puntos de muestreo en inmediaciones a la isla Malpelo (Colombia), donde se indica la fecha y la ubicación de la colecta, además de la talla, sexo y estadio de desarrollo del individuo del cual se obtuvo la muestra. Símbolo (-) = indica los datos que no fueron determinados.

Muestra	Fecha	Ubicación	Talla (m)	Sexo	Estadio
1	09/08/2015	Acuario	8	M	Juvenil
2	09/08/2015	Acuario	8	F	Juvenil
3	09/09/2015	Arrecife	5	F	Juvenil
4	09/09/2015	Nevera	11	F	Adulto
5	9/11/2015	Acuario	10	M	Adulto
6	9/12/2015	Acuario	7	F	Juvenil
7	9/14/2015	Acuario	7	F	Juvenil
8	6/12/2017	Nevera	6	F	-
9	6/13/2017	Nevera	6	F	-
10	6/13/2017	Acuario	10	-	-
11	7/26/2018	Junior	6	F	-
12	7/26/2018	Arrecife	6	F	-
13	5/28/2019	-	-	F	Adulto
14	06/02/2019	-	-		Adulto
15	06/01/2019	-	-	-	Adulto
16	8/13/2019	Arrecife	4	F	Juvenil

Apéndice C. Información sobre la extracción de ADN de las muestras de tiburón ballena (*Rhincodon typus*), correspondientes al Santuario de Fauna y Flora de Malpelo, en la isla Malpelo (Colombia). Las columnas indican los datos de la fecha de la extracción, la concentración de ADN obtenida y el kit de extracción usado para cada muestra.

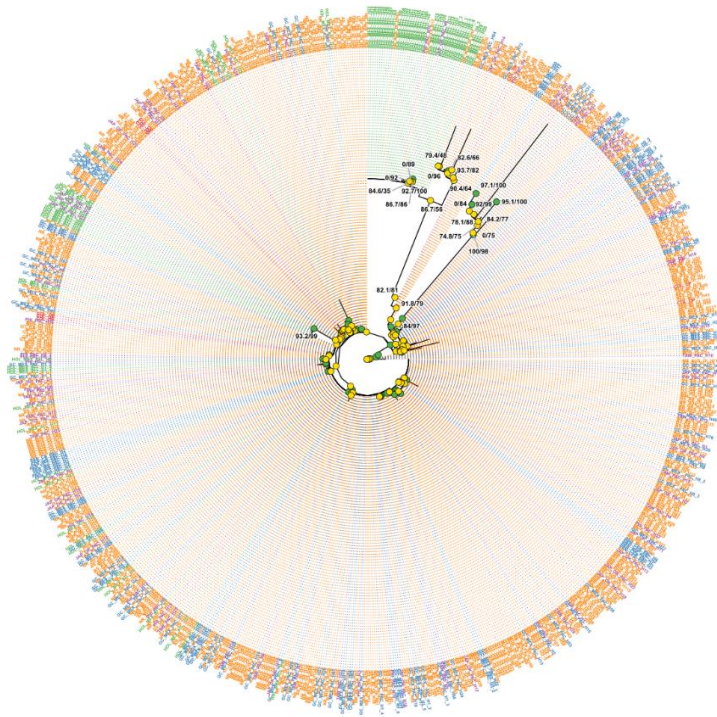
Muestra	Fecha	Concentración	Kit extracción
1	18/06/2025	1.5 ng	DNeasy PowerSoil Pro Kit
2	5/06/2025	1.45 ng	DNeasy Blood and Tissue
3	9/06/2025	28.2 ng	DNeasy Blood and Tissue
4	5/06/2025	4.8 ng	DNeasy Blood and Tissue
5	5/06/2025	4.85 ng	DNeasy Blood and Tissue
6	9/06/2025	9.95 ng	DNeasy Blood and Tissue
7	9/06/2025	3.4 ng	DNeasy Blood and Tissue
8	5/06/2025	3.75 ng	DNeasy Blood and Tissue
9	5/06/2025	3.45 ng	DNeasy Blood and Tissue
10	17/06/2025	4.75 ng	DNeasy Blood and Tissue
11	6/06/2025	4 ng	DNeasy Blood and Tissue
12	17/06/2025	1.25 ng	DNeasy Blood and Tissue

Apéndice D. Árboles filogenéticos basados en las secuencias de Región Control de ADN mitocondrial de tiburón ballena (*Rhincodon typus*), para los conjuntos de datos con gaps completos y con gaps reducidos realizados bajo el método de Máxima Verosimilitud. A). árbol correspondiente al conjunto de datos en crudo, sin curación. B). árbol correspondiente al conjunto de datos con gaps reducidos, curado con BMGE.

Filogenia Global de Rhincodon typus (IQ-TREE)

Máxima Verosimilitud - Archivo: alineamiento.final.sin.outgroup.1130pb.fas.treefile

A



Origen Geográfico

- Colombia (Destacado)
- Océano Atlántico
- Océano Índico
- Pacífico Occidental (Asia)
- Pacífico Oriental (América)

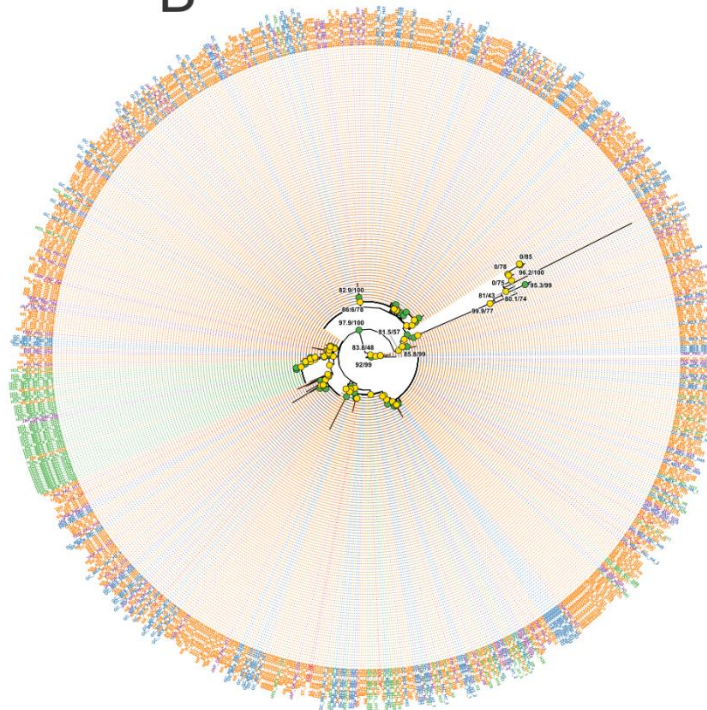
Soporte IQ-TREE (SH-aLRT / UFboot)

- Alto (SH-aLRT \geq 80 & UFboot \geq 90)
- Medio (SH-aLRT \geq 75 o UFboot \geq 75)

Filogenia Global de Rhincodon typus (IQ-TREE)

Máxima Verosimilitud - Archivo: alineamiento.final.sin.outgroup.BMGE.596pb.fas.treefile

B



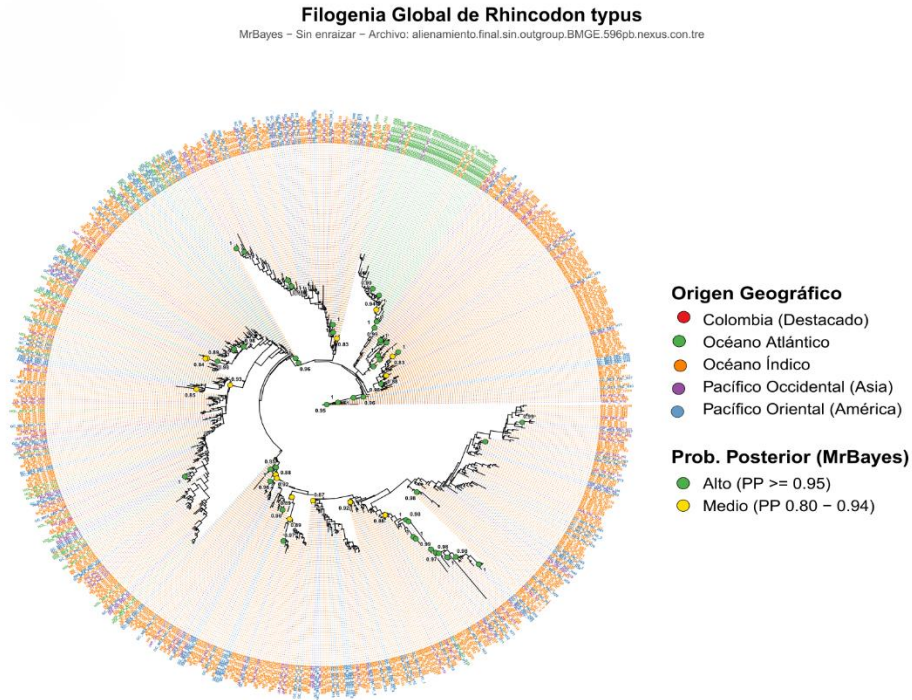
Origen Geográfico

- Colombia (Destacado)
- Océano Atlántico
- Océano Índico
- Pacífico Occidental (Asia)
- Pacífico Oriental (América)

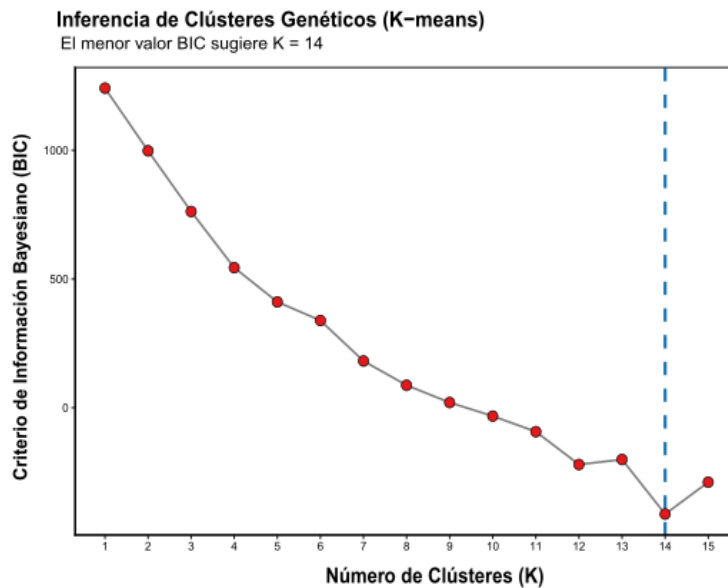
Soporte IQ-TREE (SH-aLRT / UFboot)

- Alto (SH-aLRT \geq 80 & UFboot \geq 90)
- Medio (SH-aLRT \geq 75 o UFboot \geq 75)

Apéndice E. Árbol filogenético basado en las secuencias de Región Control Mitocondrial de tiburón ballena (*Rhincodon typus*), para el conjunto de datos con gaps reducidos curado con BMGE, realizado bajo el método de Inferencia Bayesiana.



Apéndice F. Numero de clústeres genéticos inferidos para el tiburón ballena (*Rhincodon typus*), a partir del algoritmo k-means.



Apéndice G. Valores obtenidos de diferenciación genética entre poblaciones para los estadísticos JostD y G_{st}. A). Tabla de valores obtenidos para JostD. B). Tabla de valores obtenidos para G_{st}.

