INFORME TÉCNICO

Ensamble de genómas de las especies nativas Bocachico *(Prochilodus Magdal enae),* Dorada *(Brycon Moorei),* Capaz *(Pimelodus Grosskopfii)* y Pataló *(Ichthyolephas Longirostris)*

AUNAP
AUTORIDAD NACIONAL DE ACUICULTURA Y PESCA

"Acuicultura y Pesca con Responsabilidad"

ENSAMBLE DE GENÓMAS DE LAS ESPECIES NATIVAS BOCACHICO (*prochilodus magdalenae),* DORADA *(Brycon moorei),* CAPAZ *(Pimelodus grosskopfii)* y PATALÓ *(Ichthyolephas longirostris)*

Equipo de autores y colaboradores

® Universidad del Magdalena	® Autoridad Nacional de Acuicultura y Pesca
Gilberto Junior Orozco Berdugo Grupo de Investigación en Biodiversidad y Ecología Aplicada - GIBEA	Maria Rosa Angarita Peñaranda David Felipe Rivas Sánchez Pedro Julián Contreras Castro Wilberto Angulo Sarina Milena Robles Cristian Armando Rodriguez Gustavo Salazar Ariza

---

Esta publicación, es un producto resultado del convenio de cooperación No. 215 de 2019 cuyo objeto: Realizar un estudio genético-poblacional y genómico sobre especies de peces nativas y tilapia, con propósitos de conformación de lotes para repoblamiento y mejoramiento genético en las estaciones piscícolas de Gigante, Huila y Repelón, Atlántico, suscrito entre la Autoridad Nacional de Acuicultura y Pesca y La Universidad del Magdalena en el año 2019.

Citación sugerida: Orozco-Berdugo. G. (2019). Ensamble de genomas de las especies nativas bocachico *(prochilodus magdalenae*), dorada (*brycon moorei)*, capaz (*pimelodus grosskopfii)* y pataló *(ichthyolephas longirostris*). Convenio 215 de 2019. Autoridad Nacional de Acuicultura y Pesca – AUNAP. 22 p.

ENSAMBLAJE DE GENÓMAS DE LAS ESPECIES NATIVAS BOCACHICO (*Prochilodus magdalenae*), DORADA (*Brycon moorei*), CAPAZ (*Pimelodus grosskopfii*) y PATALÓ (*Ichthyolephas longirostris*)

Tabla de contenido

Lista de Tablas

Lista de Figuras

Glosario

N50: El N50 se define como la longitud mínima de contig / andamio (scaffolds) necesaria para cubrir el 50% del genoma.

N75: N75 es la longitud mínima de contig / andamio (scaffold) para cubrir el 75 % del genoma.

L50: Dado un conjunto de contigs / andamio (scaffolds), cada uno con su propia longitud, el recuento L50 se define como el número más pequeño de contigs / andamio (scaffolds) cuya suma de longitudes constituye la mitad del tamaño del genoma.

L75: Dado un conjunto de contigs / andamio (scaffolds), cada uno con su propia longitud, el recuento de L50 se define como el número más pequeño de contigs / andamio (scaffolds) cuya suma de longitud constituye el 75 % del tamaño del genoma.

1. RESUMEN

El inicio de la era genómica a principios del año 2000 ha abierto la posibilidad de tener genomas secuenciados y ensamblados de muchas especies, incluso sin genomas de referencia, cada vez en lapsos de tiempos más cortos, esto debido al avance en el desarrollo tecnológico de plataformas de secuenciación masiva que en la actualidad han evolucionado a tercera generación, con un poder para procesar millones de fragmentos de manera simultánea y de longitudes cada vez más grandes. Este avance en el desarrollo de la genética en especies con alguna importancia en sistemas de cultivo tiene un gran interés debido al potencial que puede ser descifrado y aplicado en este tipo de especies con el fin de su mejoramiento. Por ello, el propósito de este informe fue presentar los resultados de secuenciación y ensamblaje del genoma de las especies de peces nativos *Prochilodus magdalenae, Pimelodus grosskopfii, Ichtyolephas longirostris y Brycon moorei.* Se obtuvieron alrededor de 400 millones de contigs para cada especie, con un tamaño genómico que fluctúo entre 804.6 Mgb en *P. grosskopfii* a 1274.9 Mgb en *P. magdalenae,* con duplicaciones de regiones menor al 25 % y con una alta calidad enla secuenciación obtenida.

2. INTRODUCCIÓN

En el país, según la Autoridad Nacional de Acuicultura y Pesca (AUNAP, 2014) la acuicultura ha tenido un crecimiento equiparable al del crecimiento mundial de esta actividad, siendo en promedio el 13 por ciento anual durante las últimas dos décadas. También reporta que esta actividad ha venido reemplazando la producción pesquera nacional, llevando al país a establecerse dentro de los 10 países de importancia en acuicultura de latino américa.

Las herramientas de secuenciación masiva (*next-generation sequencing*) han permitido el surgimiento de técnicas de reducción que suministran mucha información representativa del genoma de las especies de interés acuícola. Una manera de conocer gran parte del genoma de los organismos a bajo costo y menor tiempo es reduciéndolo con enzimas de restricción, método utilizado en la estrategia llamada double digest restriction-site associated DNA (ddRAD-seq; Peterson et al., 2012). Una vez que el ADN ha sido fragmentado con las enzimas de restricción (una de corte frecuente y otra de corte raro), durante la preparación de la librería se ligan a los fragmentos de cada individuo un código genético (tag) individual para su identificación lo cual guiará a la PCR para la amplificación de los fragmentos. Esto se hace porque en el proceso de secuenciación se pueden depositar Rad-tags de varios individuos en un solo line que después deberán ser identificados los tags de cada individuo durante los análisis bioinformáticos.

La ventaja de utilizar la técnica de ddRAD-seq es que se puede evaluar el polimorfismo de una población a partir de la obtención de marcadores denominados polimorfismo de un sólo nucleótido (SNP). Esta será la herramienta que se empleará en este proyecto de investigación, la cual ha sido validada con éxito en estudios de varias especies de interés para la acuicultura mundial (leer revisión de Robledo et al., 2018). Por ejemplo, Palaiokostas et al. (2013b) identificaron un marcador molecular en el genoma que está asociado a los genes de determinación del sexo para el pez *Hippoglosus hippoglosus*. Esto fue logrado a partir de 5 genomas de referencia de otros peces y les permitió diferenciar genéticamente los peces machos y de las hembras. En este trabajo se identificaron 59 Rad-tags que se propusieron para la determinación del sexo en ese pez.

Todo lo anterior mencionado, demuestra que es necesario el uso y aplicación de diferentes herramientas como las moleculares para aumentar los conocimientos sobre la biología de estos individuos, y con ello influir en su productividad y conservación, la FAO (FAO, 2011) establece los lineamientos para la implementación de técnicas moleculares para planes de mejoramiento genético y bancos de genes en pro de mejorar su producción y rentabilidad, así como en la conservación y protección de recursos naturales.

El repoblamiento es una de las principales estrategias de conservación de peces es un que garantiza la composición y abundancia de especies de importancia ecología en los ecosistemas acuáticos abiertos y mantener la seguridad alimentaria y la calidad de vida de los pescadores (Merino et al., 2013; AUNAP, 2019). La calidad del conjunto de alevinos que se produce en un cultivo sea con fines de repoblación o para consumo está determinada por las condiciones genéticas y el estado fisiológico de los reproductores, además de las condiciones ambientales del lugar (FAO, 2019) debido a que los sistemas de reproducción masivos de especies parten de un pequeño número de reproductores que se mantienen bajo condiciones aparentemente óptimas, sin embargo, el encierro, la alimentación, la competencia interespecíficas y la edad puede afectar la salud de los animales generando disminución en su potencial reproductivo y mala calidad de los alevinos (ICA, 2019a).

La AUNAP cuenta con dos estaciones piscícolas ubicadas en la cuenca del rio Magdalena: la Estación Piscícola del Alto Magdalena (EPAM) en el municipio de Gigante (departamento del Huila) y la Estación Piscícola del Bajo Magdalena (EPBM) en el municipio de Repelón (departamento del Atlántico) (AUNAP et al., 2014; AUNAP, 2019).

En ambas estaciones se desarrollan actividades relacionadas con la producción masiva de alevinos de especies nativas como el bocachico (*Prochilodus magdalenae*), y especies introducidas como la tilapia plateada (*Oreochromis niloticus*) y la tilapia roja (*Oreochromis sp.*); adicionalmente se encuentran en fase experimental otras especies nativas como la dorada (*Brycon moorei*), el pátalo (*Ichthyeolephas longirotris*) y el capaz (*Pimelodus grosskopfii*). En estos sitios también se realizan investigaciones que contribuyen al incremento de la actividad acuícola en el país impulsando al sostenimiento de especies de importancia comercial y programas de fomento piscícola y repoblamiento de cuerpos de agua mediante la siembra de especies nativas (ICA, 2019a; ICA, 2019b).

Estos centros de investigación son fundamentales para el mantenimiento de especies de importancia ecológica, comercial y nativas con altos índices de productividad como lo son el bocachico (Povh et al., 2008) y especies introducidas como la tilapia roja, gracias a los programas de generación y validación de nuevas tecnologías que obedecen a las necesidades regionales y a las exigencias del mercado (ICA, 2019a).

Por todo lo descrito, el presente tuvo como objetivo establecer ensamblar el genoma de cuatro especies nativas (*Prochilodus magdalenae, Brycon moorei, Pimelodus grosskopfi, Ichtyolephas longirostris*) en las Estaciones de Repelón Atlántico en el Bajo Magdalena y Estación Piscícola del Alto Magdalena EPAM en Gigante Huila, y con ello la realización de análisis moleculares para conocer las características del genoma de las especies anteriormente mencionadas, con la finalidad de generar información sobre su biología, conservación y proponer planes de mejoramiento que permitan mejorar su producción y rentabilidad.

3. OBJETIVOS

3.1 Objetivo General

Establecer un stock de reproductores de especies nativas y de tilapia con criterios genéticos poblacionales en las Estaciones de Repelón Atlántico en el Bajo Magdalena y Estación Piscícola del Alto Magdalena EPAM en Gigante Huila.

3.2 Objetivos específicos

- Hacer un análisis exploratorio de las secuencias de la especie bocachico en un secuenciador Illumina HiSeq con base en lecturas en doble sentido, inserciones de 350 pares de bases y profundidad de cobertura 50x

- Con base en las lecturas hechas con el secuenciador Illumina HiSeq, seleccionar e implementar la mejor estrategia de ensamblaje del genoma e implementarla

4. METODOLOGÍA

4.1 Fase de campo

La fase de campo consistió en la colecta de muestras de reproductores de Bocachico (*Prochilodus magdalenae*) y Dorada (*Brycon moorei*) en la Estación de Repelón Atlántico en el Bajo Magdalena, y de Capaz (*Pimelodus grosskopfii*), y Pataló (*Ichtyolephas longirostris*) en la Estación Piscícola del Alto Magdalena EPAM en Gigante Huila. Debido al valor que representan estas especies en los sistemas de cultivo de cada estación, todo el proceso y la logistica desarrollada para la toma de muestras garantizó el menor estrés posible de los animales y por ende su supervivencia.

Para ello, todos los peces fueron anesteciados con Eugenol y con la ayuda de equipo de disección se retiró una pequeña porción de aleta caudal la cual se introdujo en un tubo eppendorf (1.5 ml) conteniendo etanol absoluto para asegurar su preservación. Así mismo, cada animal fue marcado para su identificación mediante el uso de transponders tipo Glass, el cual está dotado de un microchip provisto con un código alfanumérico que permitirá identificar al ejemplar mediante el uso de un lector electrónico, sin necesidad de contacto físico con el reproductor. El número de ejemplares que fueron usados para el estudio, la especie y la estación piscícola aparecen consignados en la tabla xx.

4.2 Fase de laboratorio

Las muestras obtenidas fueron envíadas al laboratorio Australian Genome Research Facility Ltd. (Melbourne, Australia) para su procesamiento y obtención de datos genómicos, el cual consistió la preparación de las librerías mediante la técnica de doble digestión de sitios de restricción asociados al ADN (ddRAD: Double digest restriction-site associated DNA) usando el protocolo propuesto por Peterson et al. (2012) e incluyendo los siguientes pasos:

4.3 Digestión con enzimas de restricción

Digestión de ADN usando un set de 6 enzimas (PstI, MspI, MseI, NlaIII, HpyCh4IV, EcoRI) y realizando una preparación de ocho bibliotecas de doble digestión (PstI/MspI, Pst/MseI, Pst/NlaIII, Pst/HpyCh4IV, EcoRI/MspI, EcoRI/MseI, EcoRI/NlaIII y EcoRI/HpyCh4IV) a partir de un conjunto de 3 muestras representativas en cada especie. El propósito para la preparación de estas librerías es evitar las regiones repetidas que aparecen como una banda en el gel y como un pico en el electroferograma, idealmente debería haber un perfil uniforme dentro de la ventana de selección de tamaño (280 pb - 342 pb o 280 pb - 375 pb). La presencia de una región de repetición dentro de la ventana de selección de tamaño puede reducir la cantidad de secuencia utilizable que se puede interrogar para la similitud de SNP. Cuando hubo múltiples combinaciones que producen un perfil aceptable, se seleccionó la combinación que muestra el nivel más alto de amplificación. Esta selección del tamaño de contigs ligados digeridos combinados se seleccionó utilizando el método Blue Pippin.

4.4 Preparación de librerías y secuenciación de genomas

Las bibliotecas usadas para secuenciación genómica de cada especie se generaron en Australian Genome Research Facility Ltd (AGRF) y las librerías preparadas para ese propósito se desarrollaron mediante el protocolo TruSeq Nano PCR-Free. La secuenciación se llevó a cabo en una plataforma NovaSeq 6000 en una configuración de ambos sentidos usando contigs de 150 pb. Cabe resaltar que secuenciación adicional para las especies *P. magadalenae* y *B. moorei* fue necesaria debido a que solo se generaron contigs de 270 M y 240 M, respectivamente. En total, se obtuvo aproximadamente 400 millones de contigs por muestra (Tabla 1) y como las especies tienen un tamaño de genoma esperado de ~900 Mb, la cobertura del genoma según el tamaño esperado con una pureza de muestra fue óptima en 133x.

Tabla 1. Rendimiento en cantidad de contigs obtenido con la plataforma de secuenciación para las especies nativas.

Especie	Rendimiento
Ichtyolephas longirostris	398841814
Pimelodus grosskopfii	410206452
Prochilodus magadalenae	441728969
Brycon moorei	374271541

4.5 Pre-procesamiento de los contigs

Las secuencias replicadas se fusionaron antes del análisis. Las contigs se recortaron con Trim-Galore v0.5.0 (Krueger 2015) para eliminar adaptadores / cebadores y secuencias de baja calidad (longitud <30 pb o calidad <10). Los errores de secuenciación y las llamadas de base ambiguas se eliminaron mediante SGA (Simpson 2014). La función 'preproceso' se aplicó con la configuración predeterminada. Las contigs preprocesadas se indexaron y se ingresaron en el módulo 'correcto' para realizar la corrección de errores. Esto implementa un algoritmo de corrección basado en k-mer donde el umbral de corrección de k-mer se detecta automáticamente. Los k-mers largos se utilizan para descartar errores de singleton, después de lo cual se pueden eliminar errores adicionales utilizando cálculos inexactos a partir de un consenso de múltiples alineaciones. Se conservaron las contigs sin corregir. Como se utilizó un método sin PCR, las secuencias duplicadas también se conservaron y no se calcularon. Las cualidades de contig y las características del genoma resultantes se investigaron utilizando el módulo "preqc" de SGA.

4.6 Ensamblaje del genoma

Las contigs de secuenciación limpias se ensamblaron usando SPAdes v3.14.1 (Bankevich et al. 2012) con rangos de 5 kmer "-k 33,55,71,101,121" y las configuraciones "--only-assembly" y "--cov-cutoff auto". El pulido se llevó a cabo en Pilon v1.23 utilizando secuencias de más de 1 Kb y el indicador "--chunksize 10000000".

4.7 Filtrado y estadísticas de ensamblajes

Se evaluó la calidad del montaje. Primero, se pasaron andamios a través del alineador BLAST para identificar contaminantes en el ensamblaje. 69-79% de los contigs coincidieron con un registro de peces en la base de datos nt. Para evaluar la calidad del ensamblaje se utilizó el programa QUAST v5.0.2 (Gurevich et al. 2013).

4.8 Anotación del genoma

El consenso final del ensamblaje consenso se evalúo para lograr integridad y coherencia mediante el uso de la eficiencia del mapeo y la evaluación comparativa de ortólogos universales de copia única. Se utilizó BUSCO v3.0.2 (Simão et al.2015) para identificar ortólogos de copia única universales y estimar la integridad del ensamblaje utilizando la base de datos Actinopterygii v10.

4.9 Control de calidad de la información obtenida

Las contigs de secuencia de las 4 especies se analizaron de acuerdo con las medidas de control de calidad de Australian Genome Research Facility Ltd (AGRF), para ello se utilizó

el programa FASTQC (Andrews 2010) para evaluar la calidad de las contigs sin procesar utilizando un tamaño de k-mer de 7. Además, se seleccionaron al azar mil contigs sin procesar y se alinearon con la base de datos de nucleótidos no redundante en NCBI con Blast + v2.6.0 (Camacho et al.2009).

5. RESULTADOS

Patrones de corte con enzimas de restricción y elección de las dos mejores

El experimento combinado de las ocho enzimas para seleccionar las que mejor patrón de bandeo y mayor polimorfismo generaban durante la contig se muestra a continuación. En general, las especies nativas (*Prochilodus magdalenae*, *Brycon moorei, Pimelodus grosskopfii, Ichtyolephas longirostris*) presentaron un comportamiento similar en el patrón de digestión con las enzimas donde se observa que el resultado obtenido fue bueno en todas las combinaciones, exceptuando el experimento C3 (Pst/NlaIII) donde se evidenció poca calidad en el patrón de digestión de las enzimas; mientras que B3 (Pst/MseI) y F3 (EcoRI/MseI) mostraron los mejores resultados. Por otro lado, las especies exóticas (*O. niloticus* y *O. sp.*) presentaron los mejores resultados en la digestión en el experimento F3 (EcoRI/MseI) y H3 (EcoRI/HpyCh4IV), con baja calidad en los experimentos B3 (Pst/MseI), C3 (Pst/NlaIII) y G3 (EcoRI/NlaIII) en *O. niloticus;* y C3 (Pst/NlaIII) y G3 (EcoRI/NlaIII) en *O. sp*.

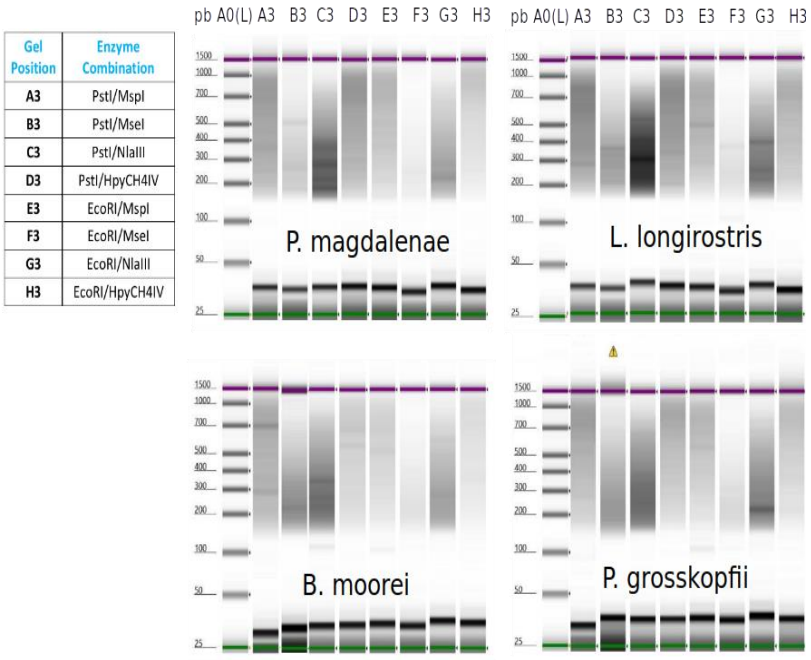

Figura 1. Pruebas con enzimas de restricción para la digestión de genoma de especies nativas.

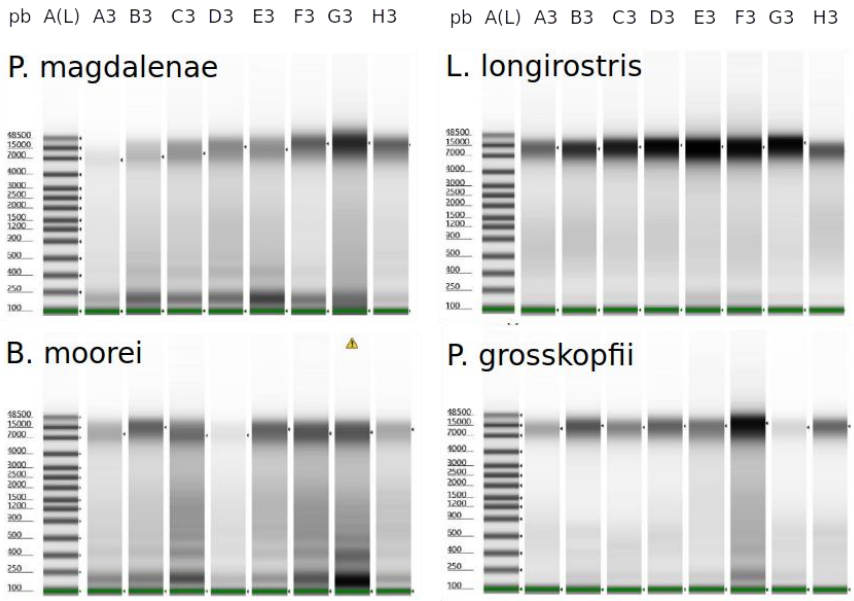

Figura 2. Patrones de digestión para elegir las dos enzimas más con los mejores resultados y su posterior utilización en la técnica ddRAD sequencing.

Resultados en la calidad de ensamblaje del genoma de especies nativos de peces nativos

En la Tabla 2 se muestra un resumen de las métricas de montaje, incluido el número de andamios (scatfolds) y la longitud total: N50, N75, L50, L75. Los nombres de andamios en los archivos FASTA de salida de SPAdes tienen el siguiente formato:> NODE_97_length_6237_cov_11.9819, en el que 97 es el número de la transcripción, 6237 es la longitud de su secuencia en nucleótidos y 11.9819 es la cobertura de k-mer. Se debe tener en consideración que la cobertura de k-mer es siempre menor que la cobertura de contig (por base). En la tabla se puede ver detallado el número de contigs obtenidos para el ensamblaje del genoma de cada especie según el tamaño y longitud del contig, según su representatividad de acuerdo con el 50 o 75 % del genoma (N50 y N75), entre otros parámetros (Tabla 2). En general y como es de esperarse, se observa un aumento en el número de contigs a medida que se disminuyen los tamaños de los mismos.

Tabla 2. Descriptores estadísticos de la calidad del ensamblaje del genoma de especies de peces nativos

Descriptor de ensamblaje	*I. longirostris*	*P. grosskopfii*	*P. magadalenae*	*B. moorei*
# contigs (>= 0 bp)	401273	245878	399534	196068
# contigs (>= 1000 bp)	401035	245774	399112	195946
# contigs (>= 5000 bp)	22423	52844	45677	32304
# contigs (>= 10000 bp)	4969	12393	12914	14866
# contigs (>= 25000 bp)	834	550	2252	4862
# contigs (>= 50000 bp)	134	80	472	1980
Total length (>= 0 bp)	929155239	900374477	1190338978	874207229
Total length (>= 1000 bp)	928923870	900272728	1189927548	874089155
Total length (>= 5000 bp)	205947574	457697842	469019831	562187172
Total length (>= 10000 bp)	92029877	180789338	249503915	441367872
Total length (>= 25000 bp)	32035113	20402489	94429411	290056445
Total length (>= 50000 bp)	8940703	5374624	34768142	190998091
# contigs	401273	245878	399534	196068
Largest contig	138955	168798	323135	702793
Total length	929155239	900374477	1190338978	874207229
GC (%)	40.48	39.02	42.14	41.43
N50	2573	5091	3760	10263
N75	1591	2778	1990	2951
L50	98150	51356	74976	14445
L75	214386	111249	185371	56980
# N's per 100 kbp	196.19	43.1	390.96	363.16

Los contigs limpios se asignaron al primer ensamblaje con una longitud mínimo de 100 y una longitud de alineación mínima de 50. La tasa de reasignación de los conjuntos primarios fue superior al 94% para todas las muestras excepto *I. longirostris* (Tabla 3). Para las contigs correctamente emparejados, la tasa de reasignación fue del 67-78%.

Tabla 3. Tasas de alineamiento en la herramienta en línea BLAST y remapeo en ensamblajes filtrados en el genoma de peces nativos.

Sample	BLAST (% fish)	Remap (%)	Remap-paired (%)
Ichtyolephas longirostris	77.96	89.13	67.57
Pimelodus grosskopfii	79.6	94.02	78.8
Prochilodus magadalenae	69.54	94.39	75.47
Brycon moorei	79.28	94.27	78.93

Anotación del genoma

La evaluación comparativa de ortólogos universales de copia única (análisis BUSCO) mostró que algunos de los genes ortólogos conservados que están presentes en los peces se pueden encontrar en el ensamblaje consenso (Tabla 4).

Tabla 4. Resultados de BUSCO para conjuntos de genomas generados con Spades. (C: completo [D: duplicado], F: fragmentado, M: faltante, n: número de genes) en las especies de peces nativos.

Sample	Complete BUSCOS's	BUSCOS results
Ichtyolephas longirostris	627	C:17.3%[S:17.0%,D:0.3%],F:16.8%,M:65.9%,n:3640
Pimelodus grosskopfii	1412	C:38.8%[S:27.7%,D:11.1%],F:19.6%,M:41.6%,n:3640
Prochilodus magadalenae	1283	C:35.2%[S:34.1%,D:1.1%],F:19.8%,M:45.0%,n:3640
Brycon moorei	1718	C:47.2%[S:46.7%,D:0.5%],F:15.9%,M:36.9%,n:3640

Descripción de algunos parámetros del genoma de especies de peces nativos

El tamaño de los genomas de las especies de peces nativos estuvo alrededor de 1000 Mpb, con un tamaño mínimo para *P. grosskopfii* de 804.6 Mpb y un máximo de 1263.4 y 1274.9 Mpb para *I. longirostris* y *P. magdalenae,* respectivamente (Figura 3).

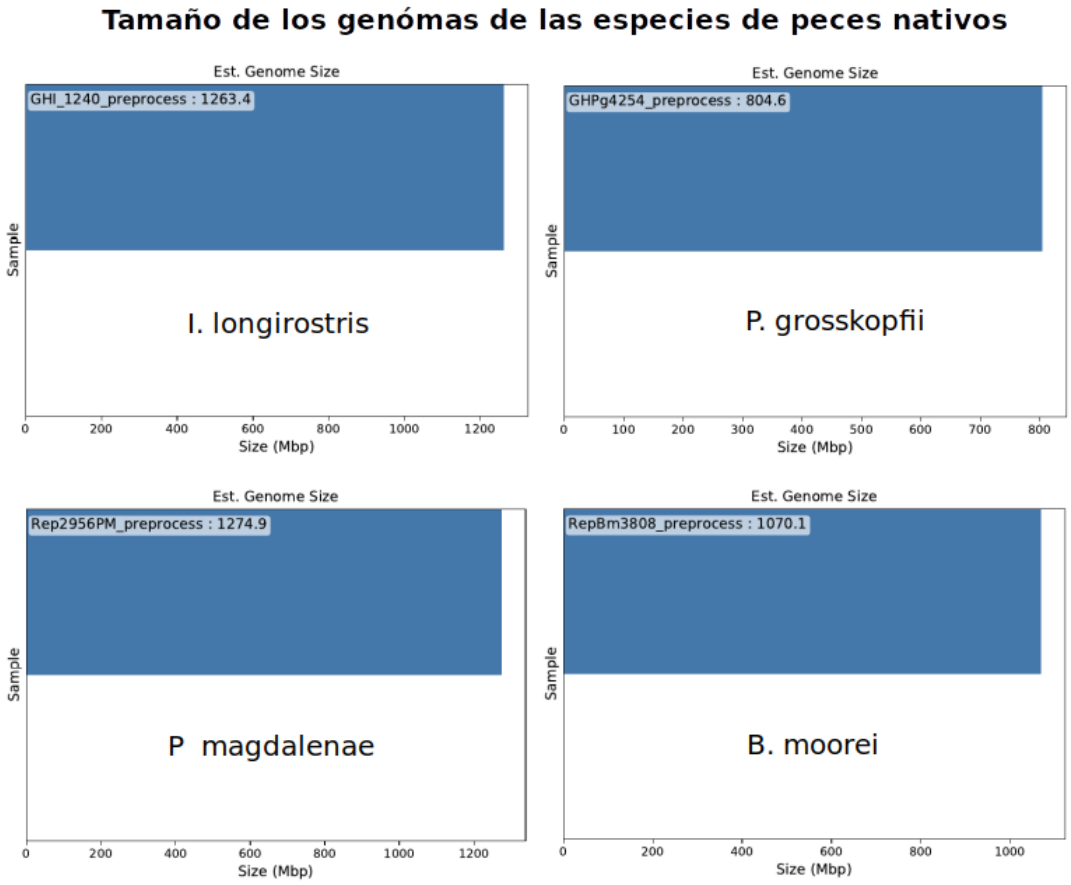

Figura 3. Tamaño del genoma en megabases obtenido para cada una de las especies de peces nativos.

La proporción de regiones genómicas duplicadas estuvo por debajo del 25 %, con valores menores para *P. magdalenae* y mayores para *P. grosskopfii* (Figura 4); mientras que el tamaño de los contigs obtenidos y usados para el ensamblaje de los genomas de las cuatro especies de peces nativos se agruparon alrededor de un promedio de 400 pb (Figura 5) en mayor proporción.

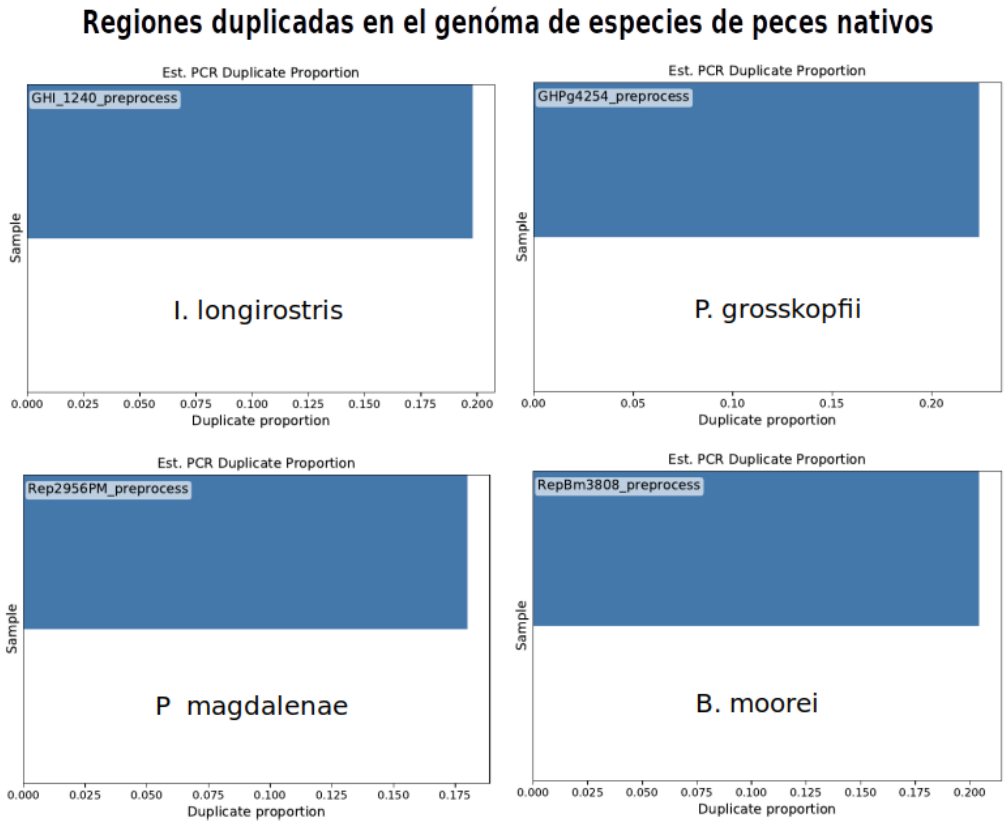

Figura 4. Proporción de regiones duplicadas en el genoma de las especies de peces nativos.

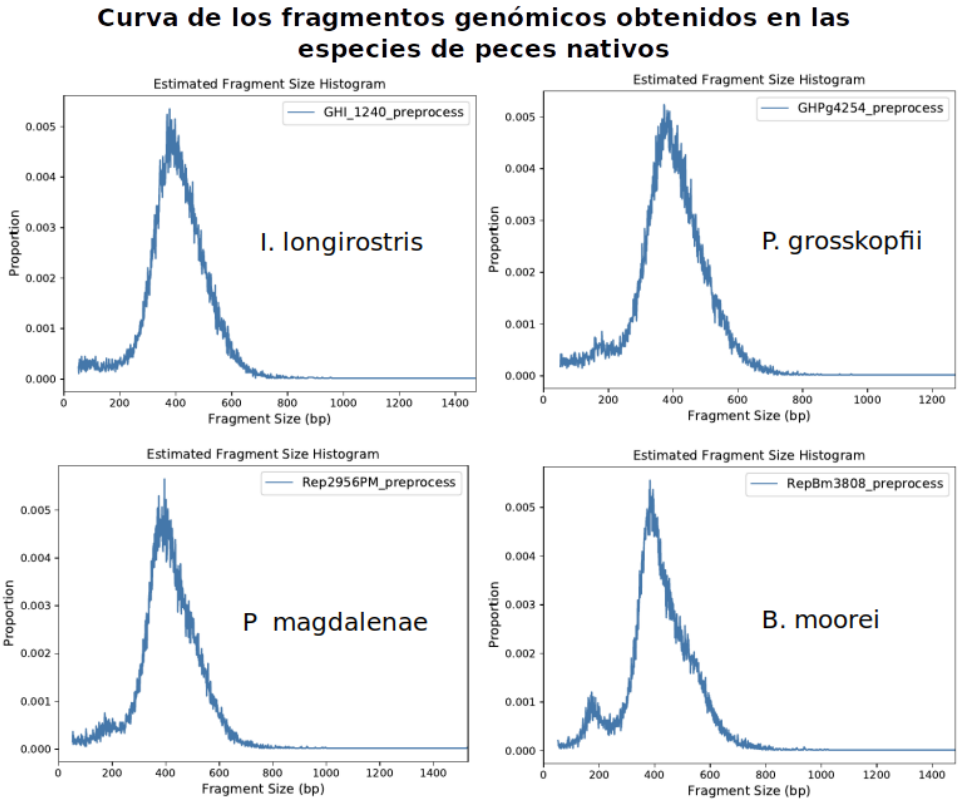

Figura 5. Curva de acumulación promedio en del tamaño de los contigs obtenidos en pares de bases(bp) en cada una de las especies de peces nativos.

Descripción del control de calidad de los contigs obtenidos y del ensamblaje de los genomas de las especies de peces nativos

La calidad de la secuencia por base para las muestras mostró excelentes resultados con calidad mayor a 89,52% de bases superiores con un el índice de calidad Q30. No hubo evidencia de contaminación viral o bacteriana significativa en los datos; no obstante, pequeños residuos de contaminación del adaptador estuvieron presente en *P. grosskopfii*. En general, los resultados del control de calidad mostraron que los resultados de la secuenciación fueron muy buenos, teniendo en cuenta que las curvas se distribuyen en las áreas verdes de las gráficas que representan una alta calidad en la información obtenida de la secuenciación (Figura 6).

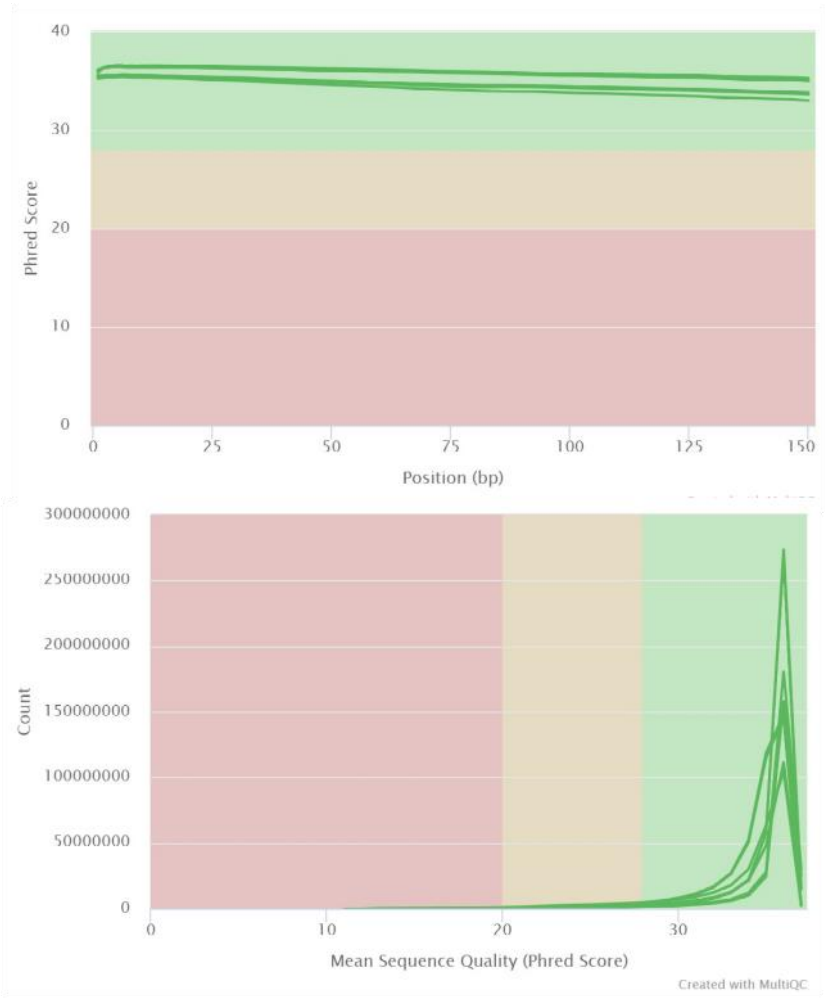

Figura 6. Control de calidad de los contigs obtenidos por posición y por secuencia en los genomas de las especies de peces nativos.

El contenido de bases Guanina-Citocina (GC) de la especie osciló entre el 40% y el 43%, estando *P. magdalenae* y *B. moorei* más cerca del 40% y el hecho de que menos del 25% de las secuencias fueran duplicadas (Figura 7) sugiere que hay una alta diversidad en las bibliotecas obtenidas, lo que sugiere buenos resultados en la preparación y obtención de las mismas.

Actividad 1.5 y 1.6_Hacer un análisis exploratorio de las secuencias de la especie bocachico en un secuenciador Illumina HiSeq con base en lecturas en doble sentido, inserciones de 350 pares de bases y profundidad de cobertura 50x. Con base en las lecturas hechas con el secuenciador Illumina HiSeq, seleccionar e implementar la mejor estrategia de ensamblaje del genoma e implementarla

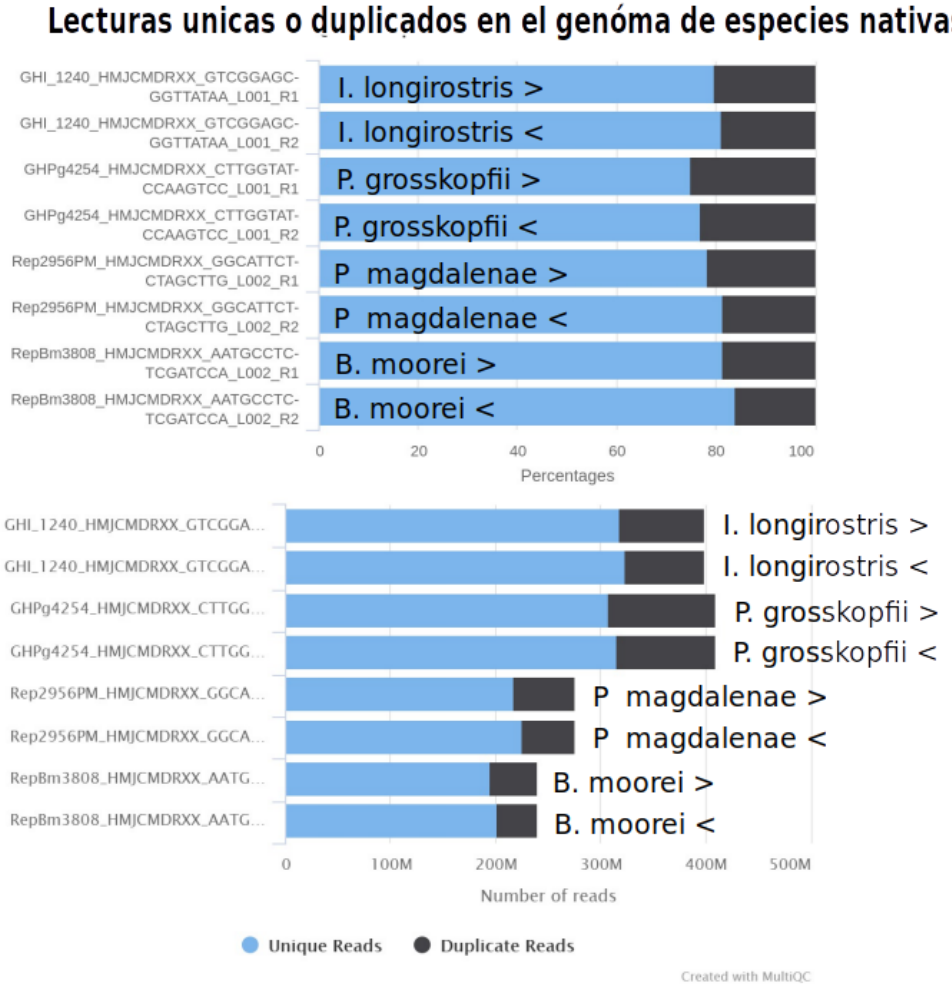

Figura 7. Cantidad de contigs unicos y duplicados encontradas en los genomas de las especies de peces nativos.

6. DISCUSIÓN

El bocachico (*P. magdalenae*) el capaz (*P. grosskopfii*), el pataló (*I. longirostris*) y la dorada (*B. moorei*) son es especies de peces nativos de Colombia, los cuales tienen un alto valor pesquero y cuya explotación ha sido bastante intensiva en los últimos años, reduciendo las abundancias de sus poblaciones y generando efectos que no han sido evaluados.

Una estrategia que ha mostrado tener buenos resultados en otras especies de peces en estas condiciones, es la creación de paquetes tecnológicos para su cultivo, lo cual propicia una alternativa para explotar poblaciones en sistemas de cautiverio y minimizar la presión sobre las poblaciones silvestres; en este sentido, el conocer la información genómica de una especie y mapear genes de importancia adaptativa pueden brindar una gran ventaja a la hora de iniciar procesos de mejoramiento genético en sistemas de cultivo. Por ello, en este informe se presentan los resultados de secuenciación y ensamblaje del genoma de estas especies de peces nativos.

Debido a que no hay información sobre los genomas de estas especies de peces nativos de Colombia y, por lo tanto, no aparecen registros en la base de datos de GeneBank; sólo fue posible explorar la relación de porcentajes de los genomas ensamblados de estas especies de peces nativos mediante la herramienta BLAST. Por ejemplo, *B. moorei* mostró cierta similitud de sus secuencias con *Astyanax mexicanus* (5.76% de las contigs) seguido por la piraña de vientre rojo *Pygocentrus nattereri* (2.51% de las contigs), mientras que *I. longirostris* también tuvo cierta similitud con estas especies, pero con más contigs asignados a la piraña de vientre rojo *Pygocentrus nattereri* (6,77%) que al *Astyanax mexicanus* (2,16%). También se identificó similitudes en la secuencia del chano *Chanos chanos* con 1,52% de las contigs de *I. longirostris*. Así mismo, resultados de BLAST de *P. magadalenae* fueron principalmente relacionadas con la piraña de vientre rojo *Pygocentrus nattereri* (4.02%) y *Astyanax mexicanus* (1.81%) y con pequeñas cantidades de curimbata *Prochilodus lineatus* y chano *Chanos chanos*. Por el contrario, las secuencias de *P. grosskopfii* tuvieron una mayor similitud con la especie de bagre rayado *Pseudoplatystoma fasciatum* y bagre amarillo *Pimelodus maculatus* con un 6,79% de las contigs, las cuales son más cercanas filogeneticamente a esta especie nativa. Un aspecto importante a considerar en esta especie fue la gran cantidad de regiones duplicadas identificadas. Diferentes autores han sugerido que la morfología y la diversidad de especies en peces teleósteos, podría estar relacionada con duplicaciones de genes independientes o para un genoma completo.

7. REFERENCIAS

Andrews, Simon. 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data.” Manual. https://doi.org/citeulike-article-id:11583827.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” Journal of Computational Biology. https://doi.org/10.1089/cmb.2012.0021.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. “BLAST+: Architecture and Applications.” BMC Bioinformatics. https://doi.org/10.1186/1471-2105-10-421.

FAO (2011). Desarrollo de la acuicultura. Enfoque ecosistémico a la acuicultura: Orientaciones Técnicas para la Pesca Responsable, No. 5, Supl. 4. Roma, ISSN 1020-5314, 60 pp.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. “QUAST: Quality Assessment Tool for Genome Assemblies.” Bioinformatics.

Krueger, F. 2015. “Trim Galore!: A Wrapper Tool around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files.” Babraham Institute. 2015. https://doi.org/10.1002/maco.200603986.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” Bioinformatics. https://doi.org/10.1093/bioinformatics/btv351.