

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

**SECUENCIACIÓN DEL GENOMA COMPLETO DEL MONO
ARAÑA DE CABEZA MARRÓN ECUATORIANO *ATELES
FUSCICEPS FUSCICEPS*, UNA ESPECIE EN PELIGRO
CRÍTICO DE EXTINCIÓN**

Lizbeth Alejandra Larreátegui Bolagay

Ingeniería en Biotecnología

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniera en Biotecnología

Quito, 22 de diciembre de 2022

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

**Secuenciación del genoma completo del mono araña de cabeza marrón
ecuatoriano *Ateles fusciceps fusciceps*, una especie en peligro crítico de
extinción**

Lizbeth Alejandra Larreátegui Bolagay

Nombre del profesor, Título académico

Gabriela Pozo Andrade, M.B.S.

Quito, 22 de diciembre de 2022

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Lizbeth Alejandra Larreátegui Bolagay

Código: 00209197

Cédula de identidad: 1721440269

Lugar y fecha: Quito, 22 de diciembre de 2022

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

El mono araña de cabeza marrón ecuatoriano (*Ateles fusciceps fusciceps*) es una especie en peligro crítico de extinción que figura en la Lista Roja de la UICN como uno de los 25 primates más amenazados del mundo. A pesar de esta situación los esfuerzos de conservación actuales se basan en regulaciones de extracción maderera y en la protección de reservas ecológicas, sin considerar la necesidad urgente de información genómica para desarrollar planes de acción efectivos para su conservación. Es así que, mediante el aislamiento de ADN genómico, la secuenciación de lecturas largas generadas por Oxford Nanopore Technologies y su ensamblaje *de novo*, el presente estudio generó por primera vez un genoma de referencia para esta subespecie. Este es un primer borrador catalogado como adecuado posterior a la evaluación de su calidad y continuidad, no obstante, este es aún un primer acercamiento para generar un genoma de referencia más preciso y completo.

Palabras clave: *Ateles fusciceps fusciceps*, enfoque genómico, genoma de referencia, ensamblaje *de novo*, continuidad, calidad, estrategias de conservación, lecturas largas, Oxford Nanopore Technologies.

ABSTRACT

The Ecuadorian brown-headed spider monkey (*Ateles fusciceps fusciceps*) is a critically endangered species and is classified one of the world's 25 most endangered primates by the IUCN Red List. Despite this situation, current conservation efforts focus on logging regulations and ecological reserves protection. Without sufficient consideration to the urgent need for genomic information to develop effective action plans for its conservation. Through the isolation of genomic DNA, the sequencing of long reads by Oxford Nanopore Technologies and their *de novo* assembly, the present study generated for the first time a reference genome for this subspecies. This first draft is cataloged as appropriate after the evaluation of its quality and continuity, however, this is still a first approach to generate a more precise and complete reference genome.

Keywords: *Ateles fusciceps fusciceps*, genomic approach, reference genome, *de novo* assembly, continuity, quality, conservation strategies, long reads, Oxford Nanopore Technologies.

TABLA DE CONTENIDO

INTRODUCCIÓN	12
Características generales.....	12
Hábitat y estado de conservación	12
Estudios genómicos	13
Métodos de secuenciación	13
Métodos de ensamblaje.....	14
Importancia de conocer el genoma completo de <i>Ateles fusciceps fusciceps</i>	14
Objetivos.....	15
MÉTODOS.....	16
Obtención de las muestras	16
Extracción de ADN	16
Preparación de bibliotecas genómicas	16
Secuenciación.....	17
Análisis Bioinformáticos	17
RESULTADOS	19
Extracción, secuenciación y evaluación de las lecturas.....	19
Ensamblado y Pulido del Genoma	19
Evaluación de la integridad del ensamblado	20
Comparación de Ensamblados	21
DISCUSIÓN.....	22
Extracción de ADN genómico	22

Secuenciación de lecturas largas	22
Ensamblaje inicial del genoma.....	23
Eficiencia del pulido	23
Contigüidad del ensamblaje en notación BUSCO	24
Evaluación del genoma frente a una referencia	25
CONCLUSIONES	26
TABLAS	27
FIGURAS	29
REFERENCIAS BIBLIOGRÁFICAS.....	30
ANEXOS	37

ÍNDICE DE TABLAS

Tabla 1. Resumen de Nanoplot para el total de lecturas obtenidas mediante secuenciación con ONT posterior al filtrado con Nanofilt en base a un límite de calidad de Q7	27
Tabla 2. Resumen de QUAST antes y después del pulido del Genoma	27
Tabla 3. Evaluación de la integridad del ensamblaje del genoma de <i>Ateles fusciceps fusciceps</i> en notación BUSCO	28
Tabla 4. Comparación de estadísticos para los genomas ensamblados de <i>Ateles Geoffroyi</i> y <i>Ateles fusciceps fusciceps</i>	28

ÍNDICE DE FIGURAS

Figura 1. Diagrama de puntos: Longitud de lectura vs Calidad media de lectura para las secuencias obtenidas por la tecnología ONT.....	29
---	----

ÍNDICE DE ANEXOS

Anexo 1. Información general de la concentración de las muestras extraídas, sus bibliotecas y cantidad cargada para el secuenciamiento.....	37
Anexo 2. Resumen de la información generada y análisis en nanoplot por corrida.....	37

INTRODUCCIÓN

Características generales

El mono araña de cabeza marrón ecuatoriano (*Ateles fusciceps fusciceps*) es un primate diurno perteneciente a la familia Atelidae. Su dieta se basa principalmente en frutos maduros, optando por hojas, semillas, brotes, cortezas, hongos, barro e insectos, cuando estos escasean. Habitan en grupos de hasta 35 individuos, evitan establecimientos humanos, y se distribuyen en subgrupos de 1 a 10 individuos para buscar alimento. Además, viven aproximadamente hasta los 24 años y las hembras alcanzan la madurez sexual a los 4 y 5 años. Generalmente, tienen una sola cría posterior a un periodo de gestación de 230 días y el cuidado lo realiza únicamente la hembra durante los primeros 20 meses (Tirira et al., 2017; Romero, 2018).

Hábitat y estado de conservación

Las zonas bajas al oeste del Ecuador conforman la región Tumbes-Chocó-Magdalena, un punto clave de biodiversidad y endemismo que enfrenta una extensa destrucción y fragmentación, lo que afecta principalmente a especies que dependen de grandes áreas de bosque continuo, como *A. fusciceps fusciceps* (Cervera & Griffith, 2016; Tirira, 2017). Esta subespecie está catalogada en peligro crítico de extinción y figura en la Lista Roja de la UICN como uno de los 25 primates más amenazados del mundo (Schwitzer et al., 2015). Romero (2018), determinó que solo 280 ejemplares sobreviven al noroeste de Ecuador, repartidos en poblaciones con limitados individuos. La única población sana de esta subespecie se encuentra en la provincia de Esmeraldas (Morelos-Juárez et al., 2015).

La situación de este primate es alarmante dado que el Chocó ecuatoriano depende en gran medida de la tala comercial de los bosques, además, la cacería ilegal y el tráfico de especies agravan su amenaza de extinción (Tirira, 2017). En el contexto de conservación de *A. fusciceps fusciceps*, los esfuerzos actuales se basan en las regulaciones de extracción maderera

y en la protección de reservas ecológicas (Morelos-Juárez et al., 2015). Aunque esta orientación es importante, diversos estudios han reconocido la necesidad urgente de información adicional que permita dilucidar la diversidad genética para desarrollar planes de acción efectivos para su conservación (Frankham, 2003; Morelos-Juárez et al., 2015; Romero, 2018).

Estudios genómicos

El genoma de referencia esclarece la representación de la diversidad genética al facilitar la localización y visualización de la heterocigosidad de marcadores moleculares, lo que constituye una estrategia para conservar la biodiversidad (Totikov et al., 2021). Este enfoque genómico, a diferencia de estudios genéticos, proporciona estimaciones de mayor precisión lo que es relevante dado que las recomendaciones y conclusiones pueden variar cuando los datos son más exactos (Frankham, 2010; Supple & Shapiro, 2018).

En la actualidad, los enfoques genómicos son utilizados en investigaciones de conservación de especies. Por ejemplo, Sutton y colaboradores (2018) secuenciaron y ensamblaron un genoma de referencia de alta calidad para el cuervo hawaiano ‘Alalā (*Corvus hawaiiensis*), una especie extinta en estado silvestre. Esto permitió realizar emparejamientos estratégicos para disminuir la tasa de endogamia y aumentar la tasa de crecimiento poblacional. Este escenario permitió preservar la diversidad genética restante y maximizó la probabilidad de que los individuos en cautiverio puedan adaptarse a diversos cambios ambientales.

Métodos de secuenciación

Para obtener un genoma se requiere de un proceso de secuenciación en donde se generen lecturas que contengan el orden de las 4 bases que conforman el material genético de uno o varios individuos (Feng et al., 2015). Existen diversas tecnologías para obtener esta información, como Oxford Nanopore Technologies (ONT), una tecnología de cuarta generación que emplea nanoporos para secuenciar material genético en tiempo real, generando

lecturas ultra largas que pueden ensamblarse (Feng et al., 2015). La longitud de lectura de esta tecnología proporciona una visión más completa de la variación genética en la caracterización de regiones repetidas y de variaciones estructurales (Nanopore Technology, 2022).

La precisión de secuenciación de ONT es baja a comparación de otras tecnologías, pero su capacidad de secuenciación directa, el alto rendimiento, y la longitud de lectura proporcionada es significativamente superior (Wang et al., 2021). En adición, los dispositivos ONT son portables, económicamente accesibles y no requieren de habilidades o recursos informáticos de alto nivel para el procesamiento de datos. Estas características contribuyen a que esta tecnología sea compatible con una amplia gama de aplicaciones como en este caso la recopilación independiente de datos (Feng et al., 2015; Wang et al., 2021).

Métodos de ensamblaje

Las secuencias se ensamblan, es decir, se reconstruyen para obtener el orden adecuado de las bases (Dominguez Del Angel et al., 2018). Este proceso puede realizarse en base a una referencia que es un genoma previamente ensamblado y curado de la misma especie o una cercanamente emparentada (Liao et al., 2019). Otra opción, y la que se usó en este caso, es ensamblar *de novo* sin la intervención de una referencia; empleando algoritmos de ensamblaje como el gráfico de Bruijn o el consenso de diseño superpuesto (Dominguez Del Angel et al., 2018). Los genomas de referencia además se usan para el ordenamiento de contigs, la anotación de genes y su análisis funcional (Ballouz et al., 2019). Para el presente estudio, el ensamblaje de *Ateles geoffroyi* se seleccionó como referencia de comparación, considerando la cercanía evolutiva y la disponibilidad de un genoma de libre acceso, (Birren, 2019).

Importancia de conocer el genoma completo de *Ateles fusciceps fusciceps*

Los cimientos de los sistemas biológicos son el material genético y sus variaciones, que permiten la evolución de las poblaciones y su adaptación a cambios ambientales (Geffen et al.,

2007). El tamaño limitado de las poblaciones de *A. fusciceps fusciceps* los hace susceptibles a la depresión endogámica y la pérdida de diversidad genética. Por ende, los esfuerzos de conservación se beneficiarían de un genoma de referencia (Frankham, 2003; Romero, 2018), que otorgue una visión integral de la diversidad genómica, como las variantes estructurales, variaciones de un nucleótido y repeticiones. Estos rasgos permiten discernir de forma más precisa una asociación fenotipo-genotipo (Nanopore Technology, 2022).

Un genoma de referencia puede usarse para estudiar marcadores asociados a mejorar la aptitud de supervivencia, análisis de genética poblacional, selección de emparejamiento y otros enfoques (Wang et al., 2021). Además de este aprendizaje directo, el genoma ensamblado conforma una referencia que puede proporcionar datos de alta calidad para futuros estudios comparativos. Por esta razón, su reconstrucción debe realizarse de forma precisa, ya que la posición adecuada de las bases y la continuidad del ensamblaje pueden afectar los resultados de los análisis posteriores. El valor de este estudio aumenta al considerar que la falta de un enfoque genómico en las estrategias de conservación de *A. fusciceps fusciceps*, se debe a la carencia total de esta información (Romero, 2018). Por ende, este proyecto representa un primer acercamiento para generar información genómica y vincularla a estas estrategias.

Objetivos

Este estudio forma parte del proyecto piloto ORG.one, una iniciativa en cooperación con ONT, cuyo fin es apoyar la secuenciación rápida de especies en peligro crítico de extinción (Org.one, 2021). Bajo esta iniciativa, se busca otorgar herramientas genómicas para mejorar las estrategias de conservación de esta subespecie. Específicamente el objetivo de este estudio es obtener el primer genoma de referencia para *A. fusciceps fusciceps* mediante la secuenciación y ensamblaje *de novo* de lecturas largas generadas por ONT.

MÉTODOS

Obtención de las muestras

Se extrajeron 5 ml de sangre de Mishki, un ejemplar macho de *A. fusciceps fusciceps* que fue atendido en TUERI-USFQ por un ataque que sufrió de parte de otro mono. El individuo tiene 6 años de edad y habita en un grupo de monos juveniles, en Jambelí, Naranjal, provincia del Guayas. La extracción de sangre se realizó en TUERI durante un control veterinario regular. La muestra se preservó en K2-EDTA a una temperatura de 4°C y fue transportada al Laboratorio de Biotecnología Vegetal de la USFQ, donde se almacenó a -20°C.

Extracción de ADN

En el aislamiento de ADN, se empleó el Qiagen Dneasy Blood and Tissue Kit para un total de 16 extracciones individuales. En la última sección del protocolo, el método de elución final se modificó, optando por emplear 30 µl de agua ultrapura para obtener una elución total de 60 µl después de 2 pasos de elución. Posterior a la extracción, se cuantificó el ADN mediante Qubit™ Fluorometric Quantitation y con NanoDrop 2000 se analizó la calidad. En cada proceso se requirió 1 µl de ADN por reacción (2 µl totales).

Preparación de bibliotecas genómicas

El protocolo para la preparación de las bibliotecas genómicas siguió el flujo de trabajo del Ligation Sequencing Kit SQK-LSK112 (Oxford Nanopore Technologies). Se inició con una concentración promedio de 2000 ng y en total se generaron 14 bibliotecas. Este proceso se divide en tres partes: la primera sección involucró la reparación del ADN y preparación de los extremos, en la segunda sección se ligan los adaptadores y se purifica el ADN y la última sección, involucra la carga de las bibliotecas en las celdas de flujo.

Después de las dos primeras secciones se cuantificó la concentración de ADN mediante Qubit™ Fluorometric Quantitation, tomando 1 µl de cada biblioteca. Esto se realizó para identificar pérdidas relevantes de ADN durante las diferentes etapas del protocolo y para conocer la cantidad que se carga en la celda de flujo. Las bibliotecas preparadas se almacenaron a 4°C hasta su secuenciación.

Secuenciación

El proceso de secuenciación se realizó en el equipo MinION Mk1C y se contó con un total de 6 celdas de flujo. A partir de las 14 bibliotecas genómicas se llevaron a cabo 21 corridas. Por lo general cada celda de flujo se empleó en 3 corridas, a excepción de 2 en las que se realizaron 4 corridas. Las muestras que poseían una concentración de ADN alta (superior a 800 ng), se sometieron a 2 corridas de secuenciación. Así mismo, dependiendo de la concentración final de cada biblioteca, se optó por cargar 6, 7 o 12 µl en la celda de flujo, hasta arribar a una concentración aproximada de 500 ng de ADN (Anexo 1). En cuanto al base calling en tiempo real, este se ejecutó con el software Guppy (v5.1.13) de ONT.

Análisis Bioinformáticos

Los análisis de las secuencias se llevaron a cabo mediante un pipeline que consta de 8 pasos. Inicialmente, se realizó un filtrado de lecturas mediante NanoFilt v1.1.0 (De Coster et al., 2018) que filtra secuencias en base a su calidad o longitud. Este proceso se realizó para recuperar las lecturas con un índice de calidad superior a Q7. A continuación, se retiraron los adaptadores de los extremos de las lecturas colocados en la preparación de librerías a través de Porechop v0.2.4 (Wick et al., 2017). Esta es una herramienta específica para lecturas de Oxford Nanopore que identifica la secuencia de los adaptadores y las elimina de las lecturas.

Para analizar el output de la secuenciación se empleó NanoPlot v1.30.1 (De Coster et al., 2018) que genera un archivo resumen con estadísticos y una serie de gráficos que

simplifican diferentes aspectos de las lecturas de secuenciación. Con Nanoplot se obtuvo un análisis holístico de todas las corridas y se analizó cada biblioteca individualmente (**Anexo 2**). La información generada en cada corrida se unificó en un solo archivo.

El archivo con todas las lecturas se usó para realizar el ensamblaje con SMARTdenovo (Liu et al., 2021). Este es un ensamblador *de novo*, rápido y específico para lecturas largas de PacBio y Oxford Nanopore, que no incluye una etapa de corrección de errores. El ensamblaje se produce al alinear un set completo de lecturas contra un set completo de lecturas sin procesar. El output resultante fue un conjunto de contigs, que se mapeó al genoma de referencia (*A. geoffroyi*) para obtener una secuencia consenso mediante Minimap2 v2.24 (Li, 2018). Este programa de alineación de secuencias ordena lecturas de ADN o ARNm en base a una referencia. La referencia seleccionada fue el ensamblaje a nivel de contigs perteneciente al mono araña de manos negras (*Ateles geoffroyi* GCA_024741035.1).

El ensamblaje mapeado pasó por un proceso de pulido mediante Medaka v1.7.2 (ONT Ltd., 2022), una herramienta que crea secuencias consenso y variantes de llamada empleando datos de secuenciación de Oxford Nanopore. El output se genera en base a redes neuronales, entre un ensamblaje inicial contra la acumulación de lecturas de secuenciación individual. En adición, se usó QUAST v5.2.0 (Gurevich et al., 2013), para evaluar ensamblajes individuales y comparar y contrastar diferentes ensamblajes. La evaluación con QUAST se llevó a cabo en 2 ocasiones para el output de Minimap y el de Medaka. Para la comparación de ensamblajes, se empleó en ambos casos la referencia de *Ateles geoffroyi* (GCA_024741035.1).

Finalmente, se usó BUSCO v5.3.2 (Manni et al., 2021) para evaluar la integridad del ensamblaje en base a la detección de genes frente a una base de datos de referencia. El conjunto de datos del linaje seleccionado fue: primates_odb10 (Número de genomas: 25, número de BUSCO: 13780) y el predictor genético utilizado fue metaeuk (Manni et al., 2021).

RESULTADOS

Extracción, secuenciación y evaluación de las lecturas

Se obtuvieron 16 reacciones de extracción que arrojaron buenos índices de calidad y concentración, de las cuales se seleccionaron 14 en base a su concentración de ADN (ng/μl). La concentración media de estas reacciones fue de alrededor 49.9 ng/μl. Posterior a la preparación de bibliotecas genómicas se contó con una concentración final media de 74.8 ng/μl (Anexo 1). Para secuenciar se requiere alrededor de 500 ng de ADN por corrida, por ende, para ajustar esta concentración se cargaron 6, 7 y 12 μl de cada biblioteca en el dispositivo MinION Mk1C (Anexo 1). Se desarrollaron 21 corridas que generaron 77.26 Gb de información total, de la cual el 69% (53.50 Gb) se clasificó como “*passed*” aprobada (Anexo 2).

Posterior al paso inicial de filtrado (>Q7) se obtuvieron 8,960,201 lecturas que corresponden a 55.95 Gb de información total. Las lecturas constaron con un tamaño N50 de 12,971 pb de extensión y una longitud y calidad de lectura media de 6,244.50 pb y 10.9, respectivamente (Tabla 1). Así mismo, la distribución individual de lecturas se observó de forma gráfica en el diagrama de puntos (Figura 1). Es evidente que existe una mayor concentración de lecturas entre las longitudes de 0 a 40 kb y en el intervalo de 7 a 15 para la calidad, aunque se generaron lecturas con una longitud de hasta 160 kb.

Ensamblado y Pulido del Genoma

El análisis individual con QUAST mostró que SMARTdenovo y Minimap2 generaron un ensamblaje inicial de 2.57 Gb de longitud total en 6,856 contigs. El N50 de las lecturas fue de 795,891 pb y la longitud del contig más largo de 5.14 Mb. Posteriormente este genoma se pulió con Medaka dando como resultado un ensamblaje de 6,856 contigs con una longitud total de 2.59 Gb, en donde, el contig más largo ascendió a 5,164,154 pb y el N50 a 799,988 pb de longitud (Tabla 2). A pesar de que el proceso de pulido no disminuyó la cantidad de contigs

totales, aumentó el tamaño total del ensamblaje en un 0.54% (0.01 Gb adicionales). En cuanto al valor N50 y el tamaño del contig más largo aumentaron un 0.52% (4,097 pb) y 0.43% (22,182 pb), respectivamente (Tabla 2).

En adición, QUAST alineó ambos ensamblajes frente al genoma a nivel de contig de *Ateles geoffroyi* (GCA_024741035.1), para su comparación. El ensamblaje mapeado presentó una fracción del genoma alineado de 92.65% con un alineamiento más largo de 1.14 Mb y 25,194 *misassemblies*. Por otro lado, el genoma pulido presentó una fracción cercana a 93.5%, con una extensión de 1.15 Mb para el alineamiento más largo y 26,309 *misassemblies*. En ambos casos la longitud total alineada, 2.50 Gb para el ensamblaje mapeado y 2.54 Gb para el ensamblaje mapeado y pulido se asemejó a la longitud total del ensamblaje de 2.57 Gb y 2.59 Gb respectivamente (Tabla 2).

Los resultados evidenciaron que, para los 4 estadísticos analizados, el ensamblaje pulido presentó valores más altos. Específicamente en la fracción del genoma, se obtuvo un aumento del 0.84% después del pulido. Así mismo, el tamaño del alineamiento más largo aumentó en un 0.41% (4,661 pb), la longitud alineada total se incrementó un 1.59% (13.9 Gb) y el número de *missassemblies*, el estadístico con mayor diferencia ascendió un 4.42% (4.097 *missassemblies* adicionales) (Tabla 2).

Evaluación de la integridad del ensamblado

Se realizó una comparación del genoma ensamblado y pulido con el conjunto de 25 genomas del linaje primates que dispone de 13,780 marcadores BUSCO. Se identificaron cerca de 10,602 BUSCOS completos que equivalen a 77% de genes esperados (Tabla 3). En adición, los parámetros de BUSCO colocaron este primer ensamblaje del genoma como de buena calidad, debido a que en su mayoría los genes de copia única 75.4% poseen una puntuación de duplicación baja de únicamente el 1.6%.

Por otro lado, la cantidad de BUSCOs fragmentados constituyeron un 5.4% del total, categoría que considera a los genes que se encuentran de forma parcial en la secuencia. Finalmente, los BUSCOs perdidos corresponden al 17.6% y abarcan a los genes ortólogos ausentes o que son demasiado divergentes para ser localizados e identificados como completos o parciales en la base de datos (Tabla 3). Estos dos últimos parámetros son un indicador de la continuidad del ensamblaje.

Comparación de Ensamblados

En la Tabla 4 se compararon 2 ensamblajes de genomas a nivel de contig de especies cercanamente emparentadas, el generado en el estudio para *Ateles fusciceps fusciceps* y la referencia *Ateles geoffroyi* como una base de evaluación. En el caso de *A. geoffroyi*, el método de ensamblaje y la tecnología de secuenciación, fueron Wtdbg2 v.2 y PacBio RSII, respectivamente. Mientras que para el ensamblaje de *A. fusciceps fusciceps* se empleó SMARTdenovo y MinION Mk1C.

Fue evidente la similitud de ciertos parámetros como el tamaño del genoma, que corresponde a 2.68 Gb para el ensamblaje de *A. geoffroyi* y 2.59 Gb para *A. fusciceps fusciceps*. Del mismo modo, el porcentaje de GC posee valores cercanos con una diferencia mínima de 0.3% entre ambos ensamblajes. No obstante, parámetros como la cobertura y el tamaño del contig N50 son significativamente más bajos en el genoma generado en este estudio, con una diferencia de 36X y 28.4 Mb, respectivamente. En cuanto al número de contigs, el ensamblaje de *A. fusciceps fusciceps* posee 6,856 contigs, lo que corresponde a una cantidad 2.52 veces mayor al ensamblaje de *A. geoffroyi* que cuenta con 2,723 contigs totales (Tabla 4).

DISCUSIÓN

Extracción de ADN genómico

El proceso de extracción arrojó buenos resultados con una concentración media de 49.9 ng/μl (Anexo 1). Dado que las mediciones de Qubit no se ven influenciados por contaminación de ARN y cuantifica únicamente dsDNA, los valores obtenidos son confiables (Dang et al., 2016). De la misma manera, se evidenció que la preparación de bibliotecas genómicas se desarrolló correctamente debido a que la concentración promedio fue de 74.8 ng/μl (Anexo 1) y para la secuenciación de ADN genómico con ONT, empleando el Kit SQK-LSK12 se requiere una concentración mínima de 25 ng/μl (MSU, 2022).

Secuenciación de lecturas largas

Los dispositivos MinION clasifican a las secuencias como aprobadas y fallidas en base a la calidad (Magi et al., 2017). En este caso las lecturas aprobadas poseían una puntuación Q media > 8 (valor por defecto del dispositivo). Para recuperar lecturas se optó por un paso de filtrado en base a una puntuación Q media > 7 , una calidad suficiente para ensamblar un genoma según la literatura revisada (Siregar et al., 2021; Esnault et al., 2022).

La secuenciación de bibliotecas con ONT genera lecturas con tamaños de 10 a 30 kb (Amarasinghe et al., 2020) y una calidad media de 10 (Magi et al., 2017). Los resultados para estos parámetros fueron adecuados considerando la técnica de secuenciación, la química de los kits empleados que mejoran constantemente (MSU, 2022), y el almacenamiento inicial de la sangre por un periodo de 3 meses previo a obtener los permisos genéticos para su uso.

En cuanto al estadístico N50, es una medida de contigüidad que involucra la longitud mediana y media de un conjunto de secuencias. Su valor indica la longitud de la lectura más corta que pertenece al grupo de lecturas más largas, un grupo que representa el 50% de nucleótidos totales en el conjunto (De Coster et al., 2018). En el presente estudio, posterior al

filtrado se obtuvo un N50 cercano a 13 kb, lo que indica que la longitud mínima de los fragmentos más largos que representan el 50% de los datos es de 13 kb o más (Tabla 1). Esta longitud alcanzada es conveniente para el ensamblaje *de novo* (Liu et al., 2021).

Ensamblaje inicial del genoma

El ensamblador SMARTdenovo y el software Minimap2, generaron un primer borrador del genoma de 2.57 Gb de extensión en 6,856 contigs con resultados buenos para la fracción del genoma alineado y la longitud total del ensamblaje (Tabla 2). No obstante, el número de contigs fue elevado considerando que debe ser cercano al número de cromosomas. Esto pudo deberse a la carencia de secuencias para conectar contigs (Koren & Phillippy, 2015) o al ensamblador. Ya que SMARTdenovo no incluye un módulo de corrección de errores, lo que puede generar menor contigüidad y un bajo nivel de integridad (Wang et al., 2021).

En cuanto al mapeo posterior al ensamblaje, este permitió establecer el orden y dar una orientación relativa a los contigs (Sadasivan et al., 2022). En este caso, Minimap2 permitió determinar un origen aproximado de la secuencia de consulta en el destino que fue el genoma de referencia (*A. geoffroyi*). Este proceso tiene una mayor precisión cuando se trabaja con lecturas largas y produce una alineación de significancia biológica (Sadasivan et al., 2022).

Eficiencia del pulido

Las lecturas largas potencian procesos, como el ensamblaje de contigs, el andamiaje y la llamada de variantes estructurales, pero son muy ruidosas. Por esto, la corrección de errores es un paso importante posterior al ensamblaje (Morisse et al., 2021). Debido a la carencia de lecturas cortas, se empleó Medaka, un pulidor de autocorrección, que solo emplea lecturas largas (Morisse et al., 2021). Lee y colaboradores (2021) compararon 8 herramientas de pulido de última generación para ONT y mostraron que Medaka produce la calidad más similar a usar lecturas cortas en el ensamblaje, de un genoma microbiano. A pesar de que el número de

contigs no disminuyó, el resultado del pulido bajo los parámetros de evaluación de QUAST, fue un ensamblaje más grande y continuo, N50 más extenso y cercano a 1 Mb (De Coster et al., 2018), con un porcentaje de identidad consenso mayor frente a la referencia (Tabla 2).

Contigüidad del ensamblaje en notación BUSCO

La evaluación de la calidad del ensamblaje en medidas de contigüidad y recuento total de genes, se obtuvo mediante BUSCO. La resolución de este análisis depende del número de genes utilizados, en este caso el linaje: primates_odb10 que consta de 25 genomas y 13,780 BUSCOs es catalogada como de alta resolución, lo que otorga un nivel de confianza alto para estas evaluaciones (Simão et al., 2015; Waterhouse et al., 2018).

Los genes se clasifican como completos cuando sus longitudes se encuentran en un rango de 2 desviaciones estándar de la longitud media del grupo BUSCO y constituyen al menos el 90% de genes (Simão et al., 2015; Waterhouse et al., 2018). En este caso se identificaron 77% de genes completos (Tabla 3), un valor apropiado considerando la cobertura de 20.87X alcanzada (Sims et al., 2014). Por otro lado, los genes completos duplicados, son raros ya que los BUSCOs son genes que evolucionan bajo el control de una sola copia. Por ende, su recuperación mínima indica un ensamblaje preciso (Simão et al., 2015; Manni et al., 2021) que fue lo obtenido, ya que solo el 1.6% de genes completos fue duplicado (Tabla 3).

Los genes recuperados de forma parcial o fragmentados fueron 5.4%, y aquellos no recuperados “faltantes” correspondieron a un 17.6% (Tabla 3). Si bien el porcentaje de genes faltantes no es bajo, existe la posibilidad de que el conjunto de genes anotados carezca de coincidencias que si estén en el genoma (Simão et al., 2015). Debido a que las limitaciones técnicas en la predicción de genes pueden inflar las proporciones de BUSCOs fragmentados y faltantes, cuando se trabaja con genomas grandes como el de *A. fusciceps fusciceps* (Manni et al., 2021). Además, las secuencias de ONT tienen tasas de error de 10 al 30% y se componen

principalmente de indels (Morisse et al., 2021). Por la naturaleza de las lecturas ciertos genes pueden poseer inconsistencias de distintas extensiones, lo que pudo disminuir el porcentaje de BUSCOs completos e incrementar los faltantes y fragmentados (Simão et al., 2015).

Por último, en primates puede haber grandes diferencias en el ADN nuclear de taxones relacionados (Puntieri et al., 2018). En *Ateles* las diferencias en el tamaño del genoma reflejan su variabilidad cromosómica y se relacionan con regiones hetero y eucromáticas (Fantini et al., 2016). Por ende, ciertos genes “faltantes” pueden haber sido potencialmente bastante divergentes (Carter & Hussain, 2018). Aun así, estos resultados pueden optimizarse mejorando el ensamblaje con otros ensambladores, pulidores y combinaciones (Lee et al., 2021).

Evaluación del genoma frente a una referencia

En primates la variación en el tamaño del genoma es pequeña (Fantini et al., 2016) y el contenido promedio de GC se encuentra en 40.83% (Qi et al., 2016). Por ende, la similitud en estos valores entre el genoma generado para *A. fusciceps fusciceps* y la referencia *A. geoffroyi*, es un indicador de un ensamblaje apropiado. Las diferencias de cobertura y N50 en parte se deben a los métodos y recursos de secuenciación (Tabla 4). Ya que a pesar de la similitud en la longitud de lectura de ONT y PacBio, las tasas de error e indels de ONT son más altas (Laver et al., 2015; Lu et al., 2016). Además, la profundidad de cobertura, se ve afectada por la precisión de algoritmos de ensamblaje y la profundidad de secuenciación (Sims et al., 2014).

Finalmente, aun cuando el N50 es significativamente inferior en los resultados obtenidos, Jauhal & Newcomb (2021) afirman que se puede obtener una puntuación BUSCO alta a partir de ensamblajes con un N50 bajo. Esto refuerza que, a pesar de su ubicuidad, el N50 no es el mejor indicador para medir la precisión del genoma y para este caso en particular, los porcentajes de BUSCOs completos son una mejor evaluación que indica el ensamblaje adecuado del genoma generado en este estudio (Jauhal & Newcomb, 2021).

CONCLUSIONES

El aislamiento del ADN genómico completo de Mishki, un ejemplar de la subespecie *Ateles fusciceps fusciceps* y su posterior secuenciación mediante Oxford Nanopore Technologies, permitió generar 55.95 Gb de información contenidas en 8,960,201 lecturas. Con esto se produjo el primer ensamblaje del genoma *de novo* de lectura larga para uno de los primates más amenazados de la costa ecuatoriana.

Mediante el ensamblador SMARTdenovo, se produjo un primer borrador ensamblado de 2.57 Gb que aumentó a 2.59 Gb posterior al pulido con Medaka. El incremento en los parámetros de evaluación con QUAST demostró la eficiencia del pulidor, al generar un ensamblaje más grande y completo en términos de contigüidad. Así mismo, mediante la notación BUSCO el ensamblaje de este primer borrador se considera adecuado al identificar 77% de genes completos en el genoma. Este parámetro de evaluación permitió medir la precisión del ensamblaje de forma más acertada que otras métricas.

Al evaluar el genoma, se observaron similitudes significativas entre el ensamblaje propuesto y la referencia (*A. Geoffroyi*), mientras que la cobertura estimada de 20.87X refleja la profundidad de secuenciación alcanzada. Para estudios posteriores se recomienda evaluar diferentes métodos de ensamblaje, pulido y posibles combinaciones para definir un pipeline más eficiente que genere un ensamblaje más preciso y completo.

Por último, a pesar de las estrategias ecológicas altamente adaptables de los primates, es evidente que *A. fusciceps fusciceps* es sensible a amenazas antropogénicas, sobre todo la deforestación de su hábitat. Por esto los datos genómicos derivados de este proyecto son un componente esencial para la recuperación actual y futura de esta subespecie. La integración de datos genómicos como parte del esfuerzo de gestión en la conservación permitirá maximizar el crecimiento y la salud genética de la subespecie en el futuro.

TABLAS

Tabla 1. Resumen de Nanoplot para el total de lecturas obtenidas mediante secuenciación con ONT posterior al filtrado con Nanofilt en base a un límite de calidad de Q7

Input	Longitud media	Calidad media	Número de lecturas	N50	Bases Totales
<i>Lecturas filtradas (Q7)</i>	6,244.50	10.90	8,960,201	12.97	55.95 Gb

Tabla 2. Resumen de QUAST antes y después del pulido del Genoma

	Ensamblaje alineado*	Ensamblaje Pulido**
Estadísticos sin referencia		
<i>Número de contigs</i>	6,856	6,856
<i>Contig más largo</i>	5,141,972	5,164,154
<i>Longitud total</i>	2,572,944,418	2,586,824,631
<i>N50</i>	795,891	799,988
Estadísticos con referencia		
<i>Fracción del Genoma (%)</i>	92.646	93.490
<i>Alineamiento más largo</i>	1,142,326	1,146,987
<i>Longitud alineada total</i>	2,497,337,383	2,537,097,083
<i>Número de misassemblies</i>	25,194	26,309

* Output de Minimap2: Genoma ensamblado posterior a su mapeo en base a un genoma de referencia (Ateles geoffroyi GCA_024741035.1)

** Output de Medaka: Genoma mapeado y pulido mediante el software Medaka

Tabla 3. Evaluación de la integridad del ensamblaje del genoma de *Ateles fusciceps fusciceps* en notación BUSCO

Parámetro	Resultados	Porcentaje (%)
<i>BUSCOs completos:</i>	10,602	77
- BUSCOs copia única	10,384	75.4
- BUSCOs duplicados	218	1.6
<i>BUSCOs fragmentados</i>	742	5.4
<i>BUSCOs perdidos</i>	2,436	17.6
<i>BUSCOs totales</i>	13,780	100

Tabla 4. Comparación de estadísticos para los genomas ensamblados de *Ateles geoffroyi* y *Ateles fusciceps fusciceps*

Estadísticos	<i>Ateles geoffroyi</i> *	<i>Ateles fusciceps fusciceps</i> **
<i>Tamaño del Genoma</i>	2.68 Gb	2.59 Gb
<i>Cobertura</i>	56.87X	20.87X
<i>Número de Contigs</i>	2,723	6,856
<i>Contig N50</i>	29.2 Mb	0.8 Mb
<i>Porcentaje GC</i>	40.5	40.8
<i>Nivel de ensamblaje</i>	Contig	Contig

* *Ateles geoffroyi* (GCA_024741035.1)

** Output de Medaka: Genoma mapeado y pulido mediante el software Medaka

FIGURAS

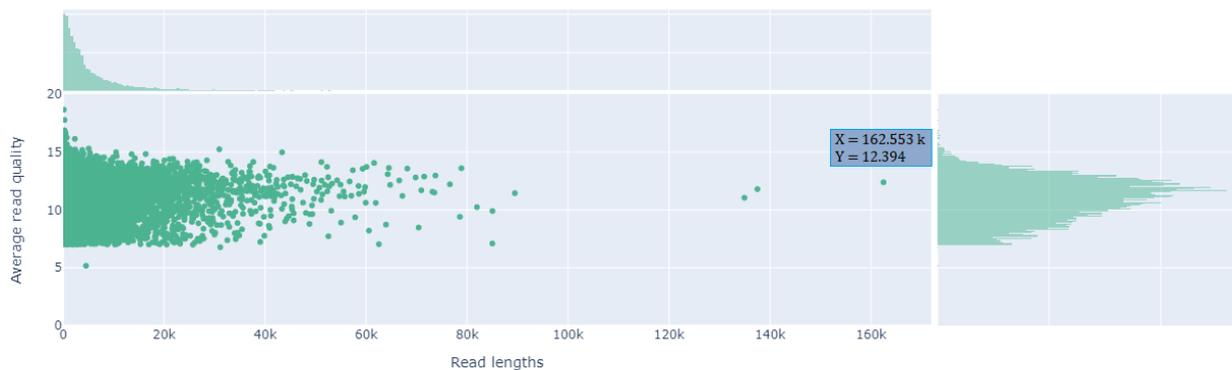


Figura 1. Diagrama de puntos: Longitud de lectura vs Calidad media de lectura para las secuencias obtenidas por la tecnología ONT. Se tiene la distribución de las lecturas totales, en el eje de las abscisas se tiene la longitud de lectura en kb y en el eje de las ordenadas la calidad de lectura media, con límites de calidad del 0 – 20. Adicionalmente, el gráfico superior muestra el histograma evaluando únicamente la longitud de las lecturas, mientras que el gráfico inferior derecho evalúa únicamente la calidad. Se señaló la lectura obtenida más extensa de 1662,553 kb con una calidad de 12.394

REFERENCIAS BIBLIOGRÁFICAS

- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21(1). <https://doi.org/10.1186/s13059-020-1935-5>
- Birren, B. W. (2016, September 28). *The 200 mammals project: Sequencing genomes by a novel cost-effective method, yielding a high resolution annotation of the human genome*. Grantome. Retrieved July 19, 2022, from <https://grantome.com/grant/NIH/R01-HG008742-01A1>
- Carter, J.-M., & Hussain, S. (2018). Robust long-read native DNA sequencing using the ONT CsgG Nanopore System. *Wellcome Open Research*, 2, 23. <https://doi.org/10.12688/wellcomeopenres.11246.3>
- Cervera, L., & Griffith, D. M. (2016). New Population and Range Extension of the critically endangered Ecuadorian Brown-headed spider monkey (*ateles fusciceps fusciceps*) in western Ecuador. *Tropical Conservation Science*, 9(1), 167–177. <https://doi.org/10.1177/194008291600900109>
- Dang, J., Mendez, P., Lee, S., Kim, J. W., Yoon, J.-H., Kim, T. W., Sailey, C. J., Jablons, D. M., & Kim, I.-J. (2016). Development of a robust DNA quality and quantity assessment qPCR assay for targeted next-generation sequencing library preparation. *International Journal of Oncology*, 49(4), 1755–1765. <https://doi.org/10.3892/ijo.2016.3654>
- De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15), 2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>

- Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., Bouri, L., Bocs, S., Klopp, C., Gibrat, J.-F., Vlasova, A., Leskosek, B. L., Soler, L., Binzer-Panchal, M., & Lantz, H. (2018). Ten steps to get started in genome assembly and annotation. *F1000Research*, 7, 148. <https://doi.org/10.12688/f1000research.13598.1>
- Esnault, G., Earley, B., Cormican, P., Waters, S. M., Lemon, K., Cosby, S. L., Lagan, P., Barry, T., Reddington, K., & McCabe, M. S. (2022). Assessment of Rapid Minion Nanopore DNA virus meta-genomics using calves experimentally infected with bovine herpes virus-1. *Viruses*, 14(9), 1859. <https://doi.org/10.3390/v14091859>
- Fantini, L., Jeffery, N. W., Pierossi, P., Gregory, T. R., & Nieves, M. (2016). Qualitative and quantitative analysis of the genomes and chromosomes of spider monkeys (primates: Atelidae). *Biological Journal of the Linnean Society*, 118(4), 752–762. <https://doi.org/10.1111/bij.12775>
- Feng, Y., Zhang, Y., Ying, C., Wang, D., & Du, C. (2015). Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics & Bioinformatics*, 13(1), 4–16. <https://doi.org/10.1016/j.gpb.2015.01.009>
- Frankham, R. (2003). Genetics and conservation biology. *Comptes Rendus Biologies*, 326(SUPPL. 1), 22–29. [https://doi.org/10.1016/S1631-0691\(03\)00023-4](https://doi.org/10.1016/S1631-0691(03)00023-4)
- Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, 143(9), 1919–1927. <https://doi.org/10.1016/j.biocon.2010.05.011>
- Geffen, E., Luikart, G., & Waples, R. (2007) Impacts of modern molecular genetic techniques on conservation biology. *Publications, Agencies and Staff of the U.S.*

Department of Commerce, 461.

<https://digitalcommons.unl.edu/usdeptcommercepub/461>

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). Quast: Quality Assessment Tool for Genome Assemblies. *Bioinformatics*, *29*(8), 1072–1075.

<https://doi.org/10.1093/bioinformatics/btt086>

Jauhal, A. A., & Newcomb, R. D. (2021). Assessing genome assembly quality prior to downstream analysis: N50 Versus Busco. *Molecular Ecology Resources*, *21*(5), 1416–1421. <https://doi.org/10.1111/1755-0998.13364>

Koren, S., & Phillippy, A. M. (2015). One chromosome, one contig: Complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology*, *23*, 110–120. <https://doi.org/10.1016/j.mib.2014.11.014>

Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies Minion. *Biomolecular Detection and Quantification*, *3*, 1–8.

<https://doi.org/10.1016/j.bdq.2015.02.001>

Lee, J. Y., Kong, M., Oh, J., Lim, J. S., Chung, S. H., Kim, J.-M., Kim, J.-S., Kim, K.-H., Yoo, J.-C., & Kwak, W. (2021). Comparative evaluation of nanopore polishing tools for microbial genome assembly and polishing strategies for downstream analysis.

Scientific Reports, *11*(1). <https://doi.org/10.1038/s41598-021-00178-w>

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>

- Liao, X., Li, M., Zou, Y., Wu, F.-X., Yi-Pan, & Wang, J. (2019). Current challenges and solutions of de Novo Assembly. *Quantitative Biology*, 7(2), 90–109.
<https://doi.org/10.1007/s40484-019-0166-9>
- Liu, H., Wu, S., Li, A., & Ruan, J. (2021). SMARTdenovo: A de novo assembler using long noisy reads. *Gigabyte*, 2021, 1–9. <https://doi.org/10.46471/gigabyte.15>
- Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore Minion Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, 14(5), 265–279.
<https://doi.org/10.1016/j.gpb.2016.05.004>
- Magi, A., Giusti, B., & Tattini, L. (2017). Characterization of minion nanopore data for resequencing analyses. *Briefings in Bioinformatics*, 18(6), 940–953.
<https://doi.org/10.1093/bib/bbw077>
- Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Current Protocols*, 1, e323. doi: 10.1002/cpz1.323
- Morisse, P., Marchet, C., Limasset, A., Lecroq, T., & Lefebvre, A. (2021). Scalable long read self-correction and assembly polishing with multiple sequence alignment. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-020-80757-5>
- MSU. (2022). *Oxford Nanopore sample requirements*. Research Technology Support Facility. Retrieved November 26, 2022, from https://rtsf.natsci.msu.edu/genomics/sample-requirements/oxford-nanopore-sample-requirements/#Sample_Mass_and_Concentration_Requirements
- Nanopore Technology. (2022). *New insights into large genomes* (1st ed., Vol. 4, Ser. Whole-genome Sequencing). White Paper.

- Org.one. (2021, October 20). *Org.one*. Oxford Nanopore Technologies. Retrieved July 19, 2022, from <https://org.one/oo>
- Oxford Nanopore Technologies Ltd. (2022, September 21). *Medaka*. PyPI. Retrieved October 7, 2022, from <https://pypi.org/project/medaka/>
- Puntieri, F., Andrioli, N. B., & Nieves, M. (2018). Association between genomic instability and evolutionary chromosomal rearrangements in Neotropical primates. *Genome Biology and Evolution*, *10*(7), 1647–1656. <https://doi.org/10.1093/gbe/evy119>
- Qi, W.-H., Yan, C.-chao, Li, W.-J., Jiang, X.-M., Li, G.-Z., Zhang, X.-Y., Hu, T.-Z., Li, J., & Yue, B.-S. (2016). Distinct patterns of simple sequence repeats and GC distribution in intragenic and intergenic regions of primate genomes. *Aging*, *8*(11), 2635–2654. <https://doi.org/10.18632/aging.101025>
- Romero, V. (2018). *Ateles fusciceps*. In J. Brito, M. A. Camacho, V. Romero, & A. F. Vallejo (Eds.), *Mamíferos del Ecuador (Version 20)*. Museo de Zoología, Pontificia Universidad Católica del Ecuador. https://bioweb.bio/faunaweb/mammaliaweb/FichaEspecie/Ateles_fusciceps
- Sadasivan, H., Maric, M., Dawson, E., Iyer, V., Israeli, J., & Narayanasamy, S. (2022). Accelerating minimap2 for accurate long read alignment on gpus. *BioRxiv*. <https://doi.org/10.1101/2022.03.09.483575>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). Busco: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>

- Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121–132. <https://doi.org/10.1038/nrg3642>
- Siregar, I. Z., Dwiyantri, F. G., Pratama, R., Matra, D. D., & Majiidu, M. (2021). Generating long-read sequences using Oxford Nanopore technology from *Diospyros celebica* genomic DNA. *BMC Research Notes*, *14*(1). <https://doi.org/10.1186/s13104-021-05484-0>
- Smith, A. D., Xuan, Z., & Zhang, M. Q. (2008). Using quality scores and longer reads improves accuracy of Solexa read mapping. *BMC Bioinformatics*, *9*(1). <https://doi.org/10.1186/1471-2105-9-128>
- Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology* 2018 19:1, *19*(1), 1–12. <https://doi.org/10.1186/S13059-018-1520-3>
- Sutton, J. T., Helmkampf, M., Steiner, C. C., Bellinger, M. R., Korlach, J., Hall, R., Baybayan, P., Muehling, J., Gu, J., Kingan, S., Masuda, B. M., & Ryder, O. A. (2018). A high-quality, long-read de Novo Genome Assembly to aid conservation of Hawaii's last remaining Crow species. *Genes*, *9*(8), 393. <https://doi.org/10.3390/genes9080393>
- Tirira, D. G., Méndez-Carvajal, P. G., & Morales-Jiménez, A.-L. (2017). Brown-Headed Spider Monkey. In C. Schwitzer, R. A. Mittermeier, A. B. Rylands, F. Chiozza, E. A. Williamson, E. J. Macfie, J. Wallis, & A. Cotton (Eds.), *Primates in Peril: The World's 25 Most Endangered Primates 2016-2018* (pp. 85–88). *IUCN SSC Primate Specialist Group*. <https://portals.iucn.org/library/sites/library/files/documents/2017-059.pdf>
- Totikov, A., Tomarovsky, A., Prokopov, D., Yakupova, A., Bulyonkova, T., Derezanin, L., Rasskazov, D., Wolfsberger, W. W., Koepfli, K.-P., Oleksyk, T. K., & Kliver, S.

- (2021). Chromosome-level genome assemblies expand capabilities of Genomics for Conservation Biology. *Genes*, *12*(9), 1336. <https://doi.org/10.3390/genes12091336>
- Wang, J., Chen, K., Ren, Q., Zhang, Y., Liu, J., Wang, G., Liu, A., Li, Y., Liu, G., Luo, J., Miao, W., Xiong, J., Yin, H., & Guan, G. (2021). Systematic comparison of the performances of de Novo Genome Assemblers for Oxford Nanopore technology reads from piroplasm. *Frontiers in Cellular and Infection Microbiology*, *11*(696669). <https://doi.org/10.3389/fcimb.2021.696669>
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, Bioinformatics and Applications. *Nature Biotechnology*, *39*(11), 1348–1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2017). Busco applications from quality assessments to gene prediction and Phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548. <https://doi.org/10.1093/molbev/msx319>
- Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Completing bacterial genome assemblies with Multiplex Minion Sequencing. *Microbial Genomics*, *3*(10). <https://doi.org/10.1099/mgen.0.000132>

ANEXOS

Anexo 1. Información general de la concentración de las muestras extraídas, sus bibliotecas y cantidad cargada para el secuenciamiento.

Muestra	Conc. Inicio	Conc. Libreria	µl cargados
<i>MaM01</i>	56.8 ng/µl	94.0 ng/µl	7
<i>MaM02</i>	79.7 ng/µl	150 ng/µl	6
<i>MaM03</i>	54.8 ng/µl	31.0 ng/µl	7
<i>MaM04</i>	48.0 ng/µl	65.6 ng/µl	7
<i>MaM05/06</i>	38.12 ng/µl	65.6 ng/µl	6
<i>MaM06</i>	48.2 ng/µl	49.0 ng/µl	6
<i>MaM07</i>	52.6 ng/µl	100 ng/µl	6
<i>MaM08</i>	52.2 ng/µl	95.4 ng/µl	6
<i>MaM09</i>	48.0 ng/µl	98.2 ng/µl	7
<i>MaM10</i>	57.8 ng/µl	86.6 ng/µl	7
<i>MaM12</i>	55.0 ng/µl	70.6 ng/µl	7
<i>MaM15</i>	29.6 ng/µl	51.2 ng/µl	7
<i>MaM16</i>	33.6 ng/µl	50 ng/µl	7
<i>MaMvarios</i>	44.0 ng/µl	39.4 ng/µl	12
Promedio	49.9 ng/µl	74.8 ng/µl	

Anexo 2. Resumen de la información generada y análisis en nanoplot por corrida

Muestra	Corrida	Total Bases	Total Passed	Longitud Media	Calidad Media	N50
<i>MaM01</i>	1	7.30 Gb	5.35 Gb	5,214.3	11.3	11,216.0

	2	4.87 Gb	3.34 Gb	5,767.8	11.0	12,225.0
<i>MaM02</i>	1	7.58 Gb	5.82 Gb	7,203.4	11.3	14,774.0
	2	4.76 Gb	3.35 Gb	5,934.1	10.9	11,528.0
<i>MaM03</i>	1	3.08 Gb	2.00 Gb	5,880.0	10.7	12,185.0
	2	2.34 Gb	1.58 Gb	6,001.7	10.7	12,636.0
<i>MaM04</i>	1	3.23 Gb	2.02 Gb	4,848.0	11.1	9,785.0
	2	1.24 Gb	0.75 Gb	4,617.9	10.9	9,964.0
<i>MaM05/06</i>	1	0.62 Gb	0.46 Gb	7,623.7	11.7	18,515.0
<i>MaM06</i>	1	1.94 Gb	1.18 Gb	7,566.6	11.2	17,298.0
<i>MaM07</i>	1	2.16 Gb	1.60 Gb	7,135.7	11.4	12,314.0
<i>MaM08</i>	1	0.436 Gb	0.315 Gb	6,669.2	11.4	12,757.0
<i>MaM09</i>	1	6.33 Gb	4.24 Gb	5,071.8	10.8	10,731.0
	2	1.22 Gb	0.800 Gb	7,315.0	11.2	14,146.0
<i>MaM10</i>	1	7.43 Gb	5.46 Gb	10,203.7	11.4	21,517.0
	2	2.23 Gb	1.56 Gb	8,478.9	11.1	18,249.0
<i>MaM12</i>	1	7.82 Gb	5.37 Gb	6,065.7	11.3	11,242.0
	2	6.12 Gb	4.13 Gb	6,121.6	11.4	11,329.0
<i>MaM15</i>	1	2.20 Gb	1.38 Gb	5,832.6	10.7	12,004.0
<i>MaM16</i>	1	1.50 Gb	0.900 Gb	5,770.0	10.6	12,963.0
<i>MaMvarios</i>	1	2.85 Gb	1.89 Gb	5,108.0	10.9	9,927.0
Total	21	77.26 Gb	53.50 Gb			