

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

**PRIMER GENOMA DE REFERENCIA DEL MONO
CAPUCHINO ECUATORIANO *CEBUS AEQUATORIALIS***

Francis Gabriela Rodríguez Jiménez

Ingeniería en Biotecnología

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

Quito, 18 de diciembre de 2024

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

**HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA**

**Primer genoma de referencia del mono capuchino ecuatoriano
*Cebus aequatorialis***

Francis Gabriela Rodríguez Jiménez

Nombre del profesor, Título académico

Gabriela Pozo Andrade, M.B.S.

Quito, 18 de diciembre de 2024

© 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: Francis Gabriela Rodríguez Jiménez

Código: 00321749

Cédula de identidad: 1753859378

Lugar y fecha: Quito, 18 de diciembre de 2024

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UNPUBLISHED DOCUMENT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 capuchino ecuatoriano (*Cebus aequatorialis*) es un primate pequeño y ágil, que se caracteriza por su comportamiento diurno, arborícola y gregario. Habita principalmente en los bosques secos y húmedos de Ecuador y Perú. Actualmente, se encuentra en peligro crítico de extinción debido a actividades antropogénicas como la cacería ilegal y la pérdida de sus hábitats por fragmentación. Con el fin de establecer estrategias de conservación más eficientes en esta especie se ha destacado la necesidad de realizar estudios genómicos, que brinden información acerca de su taxonomía, evolución o su dinámica poblacional. Para esto es clave tener un genoma de referencia. De esta manera, el objetivo de este trabajo fue secuenciar, ensamblar y anotar el primer genoma de referencia de *Cebus aequatorialis*. Para esto, se realizó una extracción de ADN con el DNeasy Blood and Tissue Kit de Qiagen, y se prepararon 8 librerías con el Ligation Sequencing Kit v.14 de ONT. La secuenciación se realizó con Oxford Nanopore MinION, y el ensamblaje se realizó con dos herramientas diferentes, SMARTdenovo y Flye, siendo esta última la mejor opción en términos de calidad. Finalmente se mapeo el ensamblado en base a una referencia y se pulió con las lecturas crudas. El ensamblado final tiene un tamaño total de 2.7 Gb, 2,997 contigs y un contig más largo de 55.3 Mb. Para la anotación, se identificaron 21,710 genes y el gen más largo demostró tener 1.12 Mb, todos estos valores se asemejan al de otras especies de primates. En este proyecto se logró obtener un buen genoma de referencia, con contigs largos y una buena continuidad. Se espera que los resultados de este trabajo sean una herramienta para realizar futuros estudios genómicos en la especie, que permitan focalizar los esfuerzos de conservación en las necesidades de esta.

Palabras Clave: secuenciación, genoma de referencia, mono capuchino ecuatoriano, *Cebus aequatorialis*, conservación, ensamblaje, anotación estructural, Oxford Nanopore Technologies.

ABSTRACT

The Ecuadorian capuchin monkey (*Cebus aequatorialis*) is a small, agile primate characterized by its diurnal, arboreal, and gregarious behavior. It primarily inhabits dry and humid forests in Ecuador and Peru. Currently, it is classified as critically endangered due to anthropogenic activities such as illegal hunting and habitat loss caused by fragmentation. To establish more effective conservation strategies for this species, the necessity of conducting genomic studies has been emphasized, providing critical insights into its taxonomy, evolution, and population dynamics. A reference genome is fundamental for such efforts. Therefore, the objective of this study was to sequence, assemble, and annotate the first reference genome of *Cebus aequatorialis*. DNA extraction was performed using the DNeasy Blood and Tissue Kit by Qiagen, and eight libraries were prepared with the ONT Ligation Sequencing Kit v.14. Sequencing was conducted with the Oxford Nanopore MinION platform, and assembly was carried out using two different tools, SMARTdenovo and Flye, the latter offering superior quality. The assembled genome was mapped to a reference and polished using raw reads. The final assembly has a total size of 2.7 Gb, 2,997 contigs, and a longest contig of 55.3 Mb. Annotation identified 21,710 genes, with the longest gene measuring 1.12 Mb, values comparable to those of other primate species. This project successfully produced a high-quality reference genome, featuring long contigs and good continuity. The results of this study are expected to serve as a valuable tool for future genomic research on the species, enabling targeted conservation efforts that address its specific needs

Key words: sequencing, reference genome, Ecuadorian white-fronted capuchin, *Cebus aequatorialis*, conservation, assembly, structural annotation, Oxford Nanopore Technologies.

TABLA DE CONTENIDO

INTRODUCCIÓN	12
Características generales	12
Estado de conservación	12
Genómica en la conservación de una especie	13
Métodos de secuenciación y ensamblaje.....	14
Importancia de establecer un genoma de referencia para <i>Cebus aequatorialis</i>	15
Objetivos	15
MÉTODOS	16
Obtención de muestras	16
Extracción de ADN	16
Preparación de librerías genómicas	16
Secuenciación	17
Procesamiento de datos	17
RESULTADOS	19
Evaluación de la calidad del ADN y de las lecturas crudas.....	19
Evaluación de la calidad de los ensamblados	19
Comparación entre la integridad de los ensamblados	20
Evaluación de la anotación estructural	20
Comparación final entre ensamblados	21
DISCUSIÓN	22

Extracción de ADN	22
Evaluación de la secuenciación	22
Comparación de la calidad de los ensamblados	23
Comparación de la integridad y contigüidad de los ensamblajes	23
Evaluación de la anotación estructural	24
Mejor ensamblado para <i>Cebus aequatorialis</i>	25
CONCLUSIONES	26
TABLAS	27
FIGURAS.....	30
REFERENCIAS BIBLIOGRÁFICAS	32
ANEXOS.....	39

ÍNDICE DE TABLAS

Tabla 1. Evaluación de calidad de todas las lecturas crudas, obtenidas de las 16 corridas de secuenciación, usando Nanoplot v1.30.1	27
Tabla 2. Comparación de QUAST v5.2.0 entre los ensamblados probados en este proyecto.....	27
Tabla 3. Evaluación de la integridad del genoma usando BUSCO v5.4.4 para los ensamblados realizados en este estudio	28
Tabla 4. Anotación estructural de genes para las ensamblados y comparación de herramientas de anotación.	28

ÍNDICE DE FIGURAS

Figura 1. Diagrama de dispersión de puntos obtenido a partir de Nanoplot v1.30.1 para todas las lecturas crudas	30
Figura 2. Gráfico de cobertura del mejor ensamblado de <i>Cebus aequatorialis</i>	30
Figura 3. BlobToolKit Snailplot de Flye 2000	31

ÍNDICE DE ANEXOS

Anexo 1. Datos de Qubit de la extracción de ADN para antes y después de la creación de bibliotecas de secuenciación	39
--	----

INTRODUCCIÓN

Características generales

El mono capuchino ecuatoriano (*Cebus aequatorialis*) es un primate mediano, caracterizado por la coloración rojiza marrón de su pelaje y su aspecto delgado. Cuenta con una cola prensil de color marrón, casi siempre enrollada; y con extremidades alargadas de un color más claro (Allen, 1914). La especie tiene comportamiento diurno, arborícola y suele vivir en grupos, que están conformados por entre cinco a veinte individuos. Su dieta consiste principalmente de frutas, como las pertenecientes a los géneros *Ficus*, *Inga* y *Sheelea*; pero también pueden comer artrópodos pequeños (Jack & Campos, 2012). Se ha estimado que el tiempo generacional de la especie es de 16 años, pero no existe mucha información acerca de su comportamiento reproductivo (Moscoso et al., 2021). Hasta el 2012, *Cebus aequatorialis* era considerada una subespecie de *Cebus albifrons*, pero se decidió realizar la separación taxonómica por la diferencia en sus territorios geográficos (Boubli et al., 2012).

La especie se encuentra distribuida mayoritariamente en la costa ecuatoriana, pero se han observado algunas poblaciones en el Noroccidente de Perú, que llegan hasta el Parque Nacional Cerros de Amotape. Es un primate que puede adaptarse a varios tipos de hábitats, pero los más comunes son los bosques tropicales o los secos tumbesinos, donde se lo observa en todos los estratos vegetales (Cervera et al., 2018).

Estado de conservación

Cebus aequatorialis ha sido clasificada como una especie En Peligro Crítico de Extinción de acuerdo con la IUCN debido a que su hábitat es uno de los más amenazados por actividades antropogénicas, agrícolas o ganaderas (Dodson & Gentry, 1991). La disminución de su hábitat, en conjunto con la cacería ilegal, dada en parte como un método de retaliación de aquellos que consideran a la especie como una plaga para los cultivos, han puesto en riesgo sus poblaciones en el país. De acuerdo con los últimos monitoreos de la especie, se cree que

existen alrededor de 6,250 individuos, pero se estima que la población seguirá disminuyendo, en especial si no se toman medidas preventivas focalizadas en su hábitat y en reducir la cacería ilegal (Moscoso et al., 2021). Actualmente, la especie se encuentra protegida por las leyes ecuatorianas, por lo que se prohíbe la tenencia, comercialización y cacería en el territorio nacional (Campos & Jack, 2013). Los esfuerzos de conservación están enfocados en mejorar las regulaciones de protección de sus hábitats naturales, ya que la mayoría se encuentra en reservas, pero la especie no se encuentra dentro de un plan de manejo. Aun así, cualquier esfuerzo de conservación podría ser fortalecido con estudios genómicos que aporten conocimiento acerca de la salud y diversidad genética de la especie, permitiendo establecer estrategias más focalizadas a sus necesidades (Cervera et al., 2018).

Genómica en la conservación de una especie

El incremento de extinciones de especies en las últimas décadas ha destacado la importancia de generar estrategias de conservación más eficientes, siendo uno de los principales enfoques, el de los estudios genómicos, ya que el objetivo de los planes de recuperación de especies amenazadas suele enfocarse en mantener en el tiempo la viabilidad de una población (Linck & Cadena, 2024). Tal y como se ha mencionado anteriormente, el genoma de referencia de una especie es la base para empezar los estudios genéticos de la misma, y en sí, se ha demostrado que puede ayudar a identificar una posible endogamia, a comprender su dinámica y crecimiento poblacional, y a identificar el efecto de determinadas fuerzas; a largo plazo, esto permite un mejor manejo de la especie y ayuda a establecer estrategias de conservación más eficientes (Brandies et al., 2019).

En diversos estudios se ha demostrado que un genoma de referencia puede acoplarse a otras herramientas que permitan comprender las dinámicas de una especie, por ejemplo, la genómica poblacional o el desarrollo de marcadores moleculares y microsatélites, que permiten investigar endogamia, la estructura poblacional de la especie y el efecto y presencia de

fenómenos como cuellos de botella o migración (Brandies et al., 2019). Un genoma de referencia puede servir también como base en un GWAS, para identificar variedades genéticas de una misma especie; en análisis y anotación de SNPs; o en estudios metagenómicos para entender factores como la dieta completa de la especie (Formenti et al., 2022). De igual manera, este nuevo genoma constituye una referencia para futuros estudios comparativos.

Métodos de secuenciación y ensamblaje

La secuenciación es un proceso que permite conocer la secuencia exacta de las bases nitrogenadas timina, adenina, guanina y citosina, en las moléculas de ADN de cualquier especie. En la actualidad, las técnicas más avanzadas de secuenciación son las de NGS (next-generation sequencing), que se caracterizan por realizar análisis paralelos y masivos, es decir, permiten obtener las secuencias de millones de fragmentos de ADN en una sola corrida (Hu et al., 2021).

Oxford Nanopore Technologies (ONT) es una tecnología de secuenciación NGS de lectura larga, que puede detectar cada una de las bases nitrogenadas de una secuencia; cuando estas pasan por un nanoporo. Esto es posible ya que cada base genera un cambio diferente en la corriente eléctrica, lo cual se detecta en un secuenciador. La tecnología presentada por ONT se ha destacado frente a otros tipos de tecnología, a pesar de que la precisión de secuenciación es más baja; esto se debe a sus bajos costos, portabilidad y por la posibilidad de obtener lecturas mucho más largas (Lu et al., 2016).

Todas las técnicas de secuenciación deben integrar un análisis bioinformático, que permita procesar las grandes cantidades de información que se obtienen de los datos crudos. Uno de los pasos más importantes en este procesamiento es el ensamblaje de secuencias, que significa reordenar todos los datos crudos para reconstruir fragmentos significativos de la secuencia de DNA de interés (Liao et al., 2019). Un buen ensamblaje permite que los pasos

posteriores, tal como el mapeado y el pulido, presenten una buena calidad, la cual se puede revisar con diversas herramientas de evaluación de genoma.

Importancia de establecer un genoma de referencia para *Cebus aequatorialis*

En base al estado de conservación de *Cebus aequatorialis* y debido a la reducción considerable en su densidad poblacional durante los últimos años, se decidió incluir esta especie al proyecto ORG.one, una iniciativa de Oxford Nanopore Technologies, en la cual se busca generar genomas de referencia con métodos de secuenciación rápidos, en especies que se encuentran amenazadas o en peligro crítico de extinción (Org.one, 2021). El objetivo de ORG.one es proteger la biodiversidad global al generar un mayor entendimiento acerca de la genética de las especies, lo que a largo plazo permitirá realizar estudios en herramientas de genómica, como las ya antes mencionadas

Cebus aequatorialis es una especie amenazada, cuya baja densidad poblacional la pone en riesgo de sufrir eventos de endogamia o pérdida de diversidad genética (Adams & Kasakoff, 1976), por este motivo y ya que no existen muchos estudios genéticos en esta especie, este proyecto constituye una base para establecer esfuerzos de conservación más eficientes y focalizados a este primate.

Objetivos

El objetivo de este trabajo es secuenciar, ensamblar y anotar el primer genoma de referencia para el mono capuchino ecuatoriano (*Cebus aequatorialis*), mediante la secuenciación de lecturas largas generadas con Oxford Nanopore Technologies.

MÉTODOS

Obtención de muestras

Las muestras se obtuvieron de un individuo adulto de mono capuchino ecuatoriano que se encuentra en el zoológico la Isla del Tapir (0°07'01.98"S 79°27.18"O), ubicado en la provincia de Santo Domingo, en el noroeste de Ecuador. En un control veterinario regular del individuo, dentro del mismo zoológico, se extrajo 1 ml de sangre, la cual fue preservada en K2-EDTA a 4°C, para luego ser transportada al Laboratorio de Biotecnología Vegetal dentro de la USFQ. En el laboratorio, la muestra fue almacenada a -80°C.

Extracción de ADN

Para la extracción de ADN de la muestra obtenida, se utilizó el Dneasy Blood and Tissue Kit de QIAGEN, con una modificación al protocolo original, las eluciones se realizaron con agua ultrapura, además se llevaron a cabo dos eluciones en el mismo tubo de 30 µl para un total de 50 µl. Con este kit se hicieron ocho extracciones de ADN. Finalmente, se realizó la cuantificación y la evaluación de calidad del ADN extraído utilizando Qubit™ Fluorometric Quantitation y Nanodrop 2000. Para este último paso, se utilizó 1 µl de ADN en cada una de las reacciones.

Preparación de librerías genómicas

Para la construcción de librerías se siguió el protocolo del Ligation Sequencing Kit V14 SQK-LSK114 de ONT. El protocolo empezó con una concentración aproximada de 2000 ng por reacción y resultó en la creación de 8 librerías genómicas. El proceso se divide en dos secciones, la reparación de ADN y preparación de extremos para unión de adaptadores, y la ligación de adaptadores. En la reparación de ADN y preparación de extremos se incubó el ADN con las enzimas durante 15 minutos a 20°C y 15 minutos a 65°C. El siguiente paso es la ligación de adaptadores, en este paso también se realizó otro cambio en el protocolo, la reacción de ligación se incubó durante 20 minutos en lugar de 10. Adicionalmente, en la purificación con

beads de cada etapa se realizó un solo lavado, en lugar de dos, como indicaba el protocolo. Para cada una de las librerías se realizó una cuantificación de ADN con Qubit™ Fluorometric Quantitation, después de ambas etapas (Anexo 1). Las librerías se almacenaron a 4°C previas a la secuenciación.

Secuenciación

La secuenciación se llevó a cabo en el secuenciador MinION Mk1C, que cuenta con el software MinKNOW integrado dentro de su sistema. Se utilizaron 6 celdas de flujo y en cada una se realizó hasta 3 corridas de secuenciación. De esta manera, en total se realizaron 16 corridas de secuenciación, usando las ocho librerías obtenidas previamente. Cada corrida de secuenciación fue de aproximadamente 24 horas. Todas las librerías se secuenciaron al menos una vez, pero aquellas con mayor cantidad de ADN (mayor a 800 ng) se secuenciaron hasta 3 veces. En cada una de las corridas se agregó 160ng de cada librería, realizando modificaciones en base a la concentración de ADN inicial. El base-calling en tiempo real se realizó mediante el software Guppy v5.1.13 de ONT.

Procesamiento de datos

Para el análisis bioinformático de las secuencias, el primer paso fue retirar los adaptadores que se ligan en el extremo de las lecturas durante la preparación de librerías, esto se realiza mediante un programa específico diseñado por ONT, que se llama Porechop v0.2.4 (Wick et al., 2017). Luego, se realiza un filtrado de lecturas usando NanoFilt v1.1.0 (De Coster et al., 2018), con el cual se realizaron dos filtraciones de tamaño, quitando lecturas menores a 1000 y menores a 2000 pb. También se realizó un análisis de las lecturas crudas con NanoPlot v1.30.1 (De Coster et al., 2018). Se analizó cada corrida de secuenciación de manera independiente, y al final se generó un análisis para las lecturas unificadas.

El siguiente paso del pipeline fue el ensamblado, para el cual se utilizaron dos herramientas, SMARTdenovo (Liu et al., 2021) y Flye v2.7.1 (Kolmogorov et al. 2019).

Ambos son ensambladores *de novo* que permiten alinear las lecturas crudas para formar contigs. Los ensamblados obtenidos se mapean a una referencia en el siguiente paso, donde se emplea Minimap2 v2.24 (Li, 2018) para ordenar los contigs. La referencia empleada para este mapeo fue el genoma de *Cebus albifrons*, que cuenta con un tamaño de 2.9 Gb, y una cobertura de genoma de 46.1X; fue secuenciado usando Oxford Nanopore PromethION, y la herramienta de ensamblaje fue NextDenovo v. 1 (GCA_023783575.1; BGI-Shenzhen, 2022). Posteriormente, las secuencias mapeadas se pulieron usando Medaka v1.7.2 (ONT, 2018) y Racon v1.5.0 (Vaser et al., 2017). Los ensamblados de Flye v2.7.1 y SMARTdenovo se evaluaron empleando QUAST v5.2.0 (Mikheenko et al. 2018), con el genoma de *C. albifrons* como referencia para la comparación; y BUSCO v5.4.4 (Manni et al. 2021), donde se usó la base de datos primates_odb10, que cuenta con 13,780 regiones conservadas, para evaluar la calidad del genoma. A partir de la evaluación de calidad se escogió solamente el ensamblado de mejor calidad para Flye v2.7.1, el que tenía una filtración quitando las lecturas menores a 2000 pb, y para SMARTdenovo, con una filtración de lecturas menores a 1000 pb.

Finalmente, se realizó una comparación entre dos pipelines de anotación estructural de genes. El primero requería de RepeatModeler v2.0.4 (Flynn et al. 2020) y de Maker v2.31.9 (Campbell et al. 2014). En las primeras rondas de anotación se hace una predicción *ab initio*, usando evidencias de proteínas y EST, estas se obtuvieron de las bases de datos del NCBI, de especies del mismo género taxonómico de *Cebus aequatorialis*. El segundo pipeline de anotado requería del software bioinformático OmicsBox, el cual tiene acoplado varias herramientas de predicción de genes, una de estas es Augustus v3.5.0 (Stanke et al., 2005), que también emplea las mismas evidencias de proteínas y EST. El análisis de ambos pipelines se realizó con InterProScan v5.61 (Jones et al. 2014), para identificar dominios Pfam y filtrar los genes con valores AED (annotation edit distance), mayores a 0.5, y con el script de agat_sp_statistics.pl (Dainat 2020), para obtener los estadísticos del anotado.

RESULTADOS

Evaluación de la calidad del ADN y de las lecturas crudas

En este trabajo se realizaron ocho extracciones de ADN a partir de un individuo de mono capuchino ecuatoriano, la concentración promedio de estas extracciones se identificó alrededor de 27.38 ng/μl y los índices de calidad y pureza, mediante espectrofotometría, estaban alrededor de 1.8 - 2.0 para la relación 260/280, y entre 2.0 - 2.2 para la relación 260/230; es decir se mantuvieron dentro del rango aceptado (Anexo 1). Para la secuenciación se realizaron 16 corridas, en 6 celdas de flujo, en donde se generó aproximadamente 86 Gb de información. De toda la información generada, aproximadamente un 99.6% tiene una calidad de Q7 o superior.

En base a la evaluación de calidad de las lecturas crudas con Nanoplot v1.30.1, se identificó que de la secuenciación se obtuvieron 18,442,482 lecturas crudas, con un tamaño medio de 4,665.5 pb, una calidad media de 10.1 y un N50 de 9,635 pb. Toda esta información estaba contenida en aproximadamente 86.000 millones de pb y generó una fracción de genoma de 31.8X (Tabla 1). Esta información se visualiza en el diagrama de dispersión, en donde se evidenció que las lecturas tienen valores de calidad entre 7 y 15, pero, aun así, se pueden observar algunas lecturas con calidades mayores a 15 y cercanas a 18 (Figura 1).

Evaluación de la calidad de los ensamblados

Mediante la evaluación de calidad de los ensamblados en QUAST se identificaron los siguientes valores para el ensamblado de SMARTdenovo (SdN), 1,795 contigs, con un contig más largo de aproximadamente 16 Mb, un N50 de 3,507,773 pb, y un L50 de 244 contigs (Tabla 2). Por otro lado, el ensamblado de Flye v2.7.1, presenta 2,997 contigs, un contig más largo de 5,344,372 pb, un N50 de 14,933,979 pb y un L50 de 55 contigs (Tabla 2). En adición, al alinear los ensamblados con la referencia *Cebus albifrons*, se puede observar que el

porcentaje de fracción de genoma cubierto es de 90.011% para SMARTdenovo y de 90.132% para Flye v2.7.1.

Comparación entre la integridad de los ensamblados

Al evaluar la integridad del genoma con los 13,780 genes ortólogos y regiones conservadas, se obtuvo, para el ensamblado de SMARTdenovo, un 87.8% de BUSCOs completos, un 4.6% de fragmentados y 7.6% de faltantes. De entre todos los BUSCOs completos, hay 85.6% de copia única y 2.2% duplicados (Tabla 3). Para el ensamblado de Flye v2.7.1, se observan 97.9% de BUSCOs completos, lo que constituye alrededor de 13,505 de los 13,780 grupos BUSCO de la base de datos; 1.2% de fragmentados y 0.8% de faltantes. De entre todos los BUSCOs completos, un 96.3% son de copia única, y 1.6% son duplicados (Tabla 3).

Evaluación de la anotación estructural

Tal y como se mencionó previamente, la anotación estructural se realizó con dos pipelines de anotación, el de Maker v2.31.9 y el de Augustus v3.5.0. Para el ensamblado de SMARTdenovo y usando Maker v2.31.9 como anotador se obtuvieron 28,952 genes, con un tamaño medio de 16,675 pb, y con un gen más largo de 1,087,428 pb. Por otro lado, para el mismo ensamblado, usando el protocolo de Augustus v3.5.0, se obtuvieron 22,479 genes, con un tamaño medio de 41,627 pb, y el del gen más largo está en alrededor de 835,785 pb (Tabla 4). En el caso del ensamblado de Flye v2.7.1, la anotación se realizó solamente usando Augustus v3.5.0, ya sabiendo que la calidad de anotado es mejor en base a los resultados obtenidos con SMARTdenovo. Para este ensamblado se identificaron 21,710 genes, con un tamaño promedio de 44,103 pb y un gen más largo de 1,119,867 pb (Tabla 4).

Comparación final entre ensamblados

Al evaluar las dos herramientas de ensamblaje, y como se discutirá en la siguiente sección, se puede observar que la calidad más alta se obtuvo con Flye v2.7.1. De esta manera, se realizó un gráfico de cobertura para este ensamblado con el fin de identificar cuántas lecturas se alinean en el contig más largo, siendo esto otro método de evaluación de completeness y de calidad del ensamblado y del mapeado. En general, se puede observar que las lecturas crudas se reparten de forma homogénea en todas las posiciones del contig, con excepción de una barra más larga alrededor de la posición de la base 0.70 (Figura 2).

Para tener una mejor representación visual de como se ve el mejor ensamblado, se realizó un BloobToolKit Snailplot, en el cual se pueden observar varios parámetros de calidad de forma gráfica. Se observa que el tamaño total del genoma ensamblado es de 2.7 Gb, y el contig más largo, señalado en rojo, cubre 55.3 Mb de este. Se puede identificar también el valor de N50, que es 14.9 M, y el contenido de GC (40.8%) y AT (59.2%) en la parte inferior. En la parte superior derecha se observan los 97.9% de BUSCOs completos previamente mencionados, al igual que los porcentajes respectivos de BUSCOs fragmentados y faltantes (Figura 3).

DISCUSIÓN

Extracción de ADN

Se evidenció que el procedimiento de extracción de ADN dio buenos resultados de calidad y cantidad. De acuerdo con la literatura, para la secuenciación con Nanopore, usando la misma cantidad de μl indicadas en el Ligation Sequencing Kit (ONT, 2022), se sugiere empezar con una concentración de 20 ng/ μl para una masa total de 1000 ng, pero se ha encontrado que el proceso puede empezar con hasta 6 ng/ μl para un total de 300 ng (Karst et al., 2021). En este caso se obtuvo una concentración promedio de ADN de 33.8 ng/ μl , con una masa total de 850-1000 ng, un valor suficiente para empezar la secuenciación. De igual manera, el kit indica iniciar el proceso de secuenciación con 10-20 fentomoles (ONT, 2022), en este proyecto se hizo una estimación de la cantidad de ADN para empezar con 14-20 fmol. Se calculó que la cantidad de ADN necesaria para empezar era de 160 ng, de acuerdo con el tamaño promedio de lecturas que se obtuvieron para otros primates (Pozo et al., 2024).

Evaluación de la secuenciación

En la evaluación de calidad de las lecturas crudas, el primer dato que destaca es la cantidad total de bases, que se encuentra en 86 Gb, un valor alto en esta sección indica más información para empezar con el proceso de ensamblaje. Se obtuvieron 18,442,482 lecturas, con una calidad promedio de 10.1. Estos valores coinciden con lo que se indica en la literatura, ya que para Nanopore, dependiendo de la celda de flujo utilizada, se espera lecturas con una calidad promedio de Q10 (Wang et al., 2021). Toda esta información generó una cobertura de genoma de 31.8X. Se ha reportado que este parámetro depende de la aplicación que se le vaya a dar a la secuenciación, para realizar WGS o establecer un genoma de referencia, lo adecuado es tener una cobertura de 30X (Barret et al., 2006). Con respecto al número de lecturas, hay estudios que indican que de la secuenciación con Nanopore, usualmente se obtienen entre 10,000 a 30,000 lecturas (De Coster et al., 2018), el valor obtenido se encuentra dentro de lo esperado. De igual manera, el tamaño medio de las lecturas y su calidad son comparables a lo

que se observa en literatura para primates. En el artículo de Pozo et al., (2024), se realizó la secuenciación de *Ateles fusciceps fusciceps*, y se identificaron alrededor de 8 millones de lecturas, con un tamaño medio de 6,244 pb. Esto es comparable a lo obtenido en este proyecto, ya que el tamaño medio de lectura fue de 4,665.5 pb, la diferencia no es significativa, y ya que en este trabajo la cantidad de lecturas es mayor, se sabe que la información es suficiente para obtener un ensamblado de buena calidad.

Comparación de la calidad de los ensamblados

Se puede observar que SdN muestra menos contigs, pero los valores de contig más largo, N50, y L50 son menores a los presentados con Flye v2.7.1. Esto nos indica que con Flye v2.7.1 se obtiene un ensamblado de mejor calidad, más continuo y menos fragmentado. De acuerdo con la literatura, se debe buscar un N50 alto y un L50 bajo, ya que esto demuestra menos gaps de información y por lo tanto más continuidad (Liem et al., 2021). El ensamblado de Flye v2.7.1 muestra un N50 mayor y un L50 menor, en comparación al de SdN, lo que indica que el ensamblado es de mejor calidad y más continuo. Esta información es comparable a la encontrada en literatura, en el genoma de referencia de *Cebus albifrons* (BGI-Shenzhen, 2022) se reportan 2,079 contigs, un N50 de 18.5 Mb y un L50 de 44. Los datos de calidad de este proyecto son comparables a lo que hay reportado en literatura. En la referencia se reporta un tamaño de genoma de 2.9 Gb, mientras que el ensamblado de Flye v2.7.1 presenta un tamaño de 2.7 Gb, por lo que el valor está dentro de lo esperado para el genoma de un primate (Rogers & Gibbs, 2014). El porcentaje de GC para el ensamblado de Flye v2.7.1 fue de 40.8%, teniendo en cuenta que este valor es consistente en primates (Qi et al., 2016), se observan los siguientes valores en *C. imitator* (40.5%) y en *C. albifrons* (41%).

Comparación de la integridad y contigüidad de los ensamblajes

El ensamblado de SMARTdenovo presenta un porcentaje significativamente menor de BUSCOs completos en comparación al ensamblado de Flye v2.7.1, lo que indica que este

último es más integro, más continuo y de mejor calidad, ya que la mayoría de las regiones ortólogas se encuentran presentes. Se reporta que un ensamblado de excelente calidad debería tener más de 90% de BUSCOs completos, y que un valor inferior al 50% representa una calidad baja (Tvedte et al., 2021). Con el ensamblado de Flye v2.7.1 se logra superar al 90% indicado en la literatura. De igual manera, se observa que casi todos los BUSCOs son de una copia y solo aproximadamente 1.5% son duplicados, el cual podría deberse a una región repetitiva o a algún fallo en la secuenciación que haya duplicado la región (Simao et al., 2015). El porcentaje de BUSCOs faltantes y fragmentados para el ensamblado de Flye v2.7.1 es significativamente bajo, lo que indica calidad alta. Estos datos son comparables a lo reportado para diversos genomas de primates. En la referencia de *Sapajus apella*, se reportan 97.8% de BUSCOs completos (Culibrik et al., 2019), mientras que para *Cebus imitator*, se reportan 97.5% de BUSCOs completos (GCF_001604975.1; Orkin et al., 2021). Los datos del ensamblado de Flye v2.7.1, son mayores a lo que hay reportado en literatura.

Evaluación de la anotación estructural

En la literatura se ha reportado que el genoma de los primates tiene alrededor de 20,000-25,000 genes (Pennisi, 2007), por lo que se puede observar que el pipeline de Maker v2.31.9, con 28,952 genes, tiene un número demasiado elevado al esperado. Esto posiblemente se debe a que el pipeline utiliza soft masking, lo cual no es lo mejor para primates, ya que estos genomas tienen cantidades altas de regiones repetitivas, que podrían haberse predicho como genes (Saenko et al., 2021). Por otro lado, ambas anotaciones realizadas con Augustus v3.5.0 presentan valores más cercanos a lo que hay reportado en literatura para el número de genes, de exones e intrones, en genomas de referencia de primates, siendo Flye v2.7.1, la opción con más parámetros similares, lo que indica una mejor calidad de anotación. La anotación de Flye v2.7.1 con Augustus v3.5.0, presenta valores comparables a lo que se ha reportado para especies taxonómicamente cercanas a *Cebus aequatorialis*, por ejemplo, se obtuvieron 21,710

genes, con un tamaño promedio de 44,103 pb, mientras que para *Cebus imitator* (Orkin et al., 2021) se han reportado 21,544 genes, con un tamaño promedio de 40,827 pb; y para *Sapajus apella* (GCF_009761245.1; Culibrk et al. 2019), se ha reportado 20,350 genes, con un tamaño medio de 38,604 pb. Con respecto al número de exones y sus tamaños promedios, *C. imitator* (254,797 exones) y *S. apella* (274,829 exones), se alejan un poco al valor obtenido en el proyecto (200,769 exones), posiblemente debido a que en ambos se usa una tecnología de secuenciación diferente a ONT, aun así, la diferencia no es grande y el valor se encuentra dentro de lo esperado.

Mejor ensamblado para *Cebus aequatorialis*

En base a la evaluación de calidad de ensamblado y anotación se puede identificar que Flye v2.7.1 genera un genoma menos fragmentado y de mejor calidad, por lo que fue escogido como el genoma de referencia. En la literatura se ha observado que SMARTdenovo es útil para ensamblar genomas pequeños, mientras que Flye suele generar mejores resultados en genomas más grandes (Wang et al., 2021). Esto se evidencia en el artículo de Pozo et al., (2024), donde se recalca que Flye es una buena herramienta para genomas de primates ya que genera resultados altos de calidad en genomas grandes y con varias regiones repetitivas.

De esta manera, los resultados de este proyecto no solo demuestran la calidad de los métodos utilizados para la secuenciación y ensamblaje del ADN, sino que también resaltan el potencial del uso de un genoma de referencia como una buena herramienta para estudios genómicos en *Cebus aequatorialis* y de otros primates. El genoma generado, que presenta estándares altos de calidad y continuidad, es una base sólida para explorar la diversidad genética, identificar características funcionales o taxonómicas y desarrollar estrategias de conservación específicas para la especie. Este trabajo contribuye de forma significativa al campo de la genómica de primates y se espera que los resultados puedan ser aplicados en estudios de otras especies con desafíos genómicos y de conservación similares.

CONCLUSIONES

En este trabajo se presenta el primer genoma de referencia para la especie, al igual que sus datos se ensamblaje y anotación estructural. La secuenciación de lectura larga de Oxford Nanopore, a través de 16 corridas, permitió obtener 86 Gb de información, en 18.44 Mb lecturas. De igual manera, se realizó una comparación de diferentes ensamblados, con el fin de definir cuáles son las herramientas que generan una mejor calidad en términos de contigüidad, completeness e integridad. Se identificó que Flye v2.7.1, en comparación con SMARTdenovo, produce mejores parámetros de calidad en el ensamblado. El ensamblado de Flye v2.7.1 presentó 2.997 contigs, siendo el más largo de 55.3 Mb, lo que indica un genoma más completo, menos fragmentado y de mejor calidad. Mediante BUSCO v5.4.4, se realizó un análisis de completeness del ensamblado y se identificó que Flye v2.7.1 mostró más regiones conservadas, teniendo 97.8% de BUSCOs completos, un dato comparable y ligeramente mayor a lo presentado en literatura para especies similares, esto indica que el genoma de referencia presentado es de buena calidad, poco fragmentado y continuo.

Para la anotación estructural se identificó que el pipeline de Augustus v3.5.0 genera una mejor calidad de anotación en relación con Maker v2.31.9. La anotación final se realizó con el ensamblado de Flye v2.7.1 y con la herramienta Augustus v3.5.0, se reportó 21,710 genes en el ensamblado, con un tamaño promedio de 44,103 pb. El número de genes, exones e intrones es similar a lo que hay reportado en genomas de referencia de especies de primates, comprobando una buena calidad de ensamblado y anotación para este genoma. *Cebus aequatorialis* es una especie en peligro crítico de extinción, que requiere de estrategias de conservación más focalizadas, se espera que estos resultados puedan ser usados como base para realizar más estudios genómicos en *Cebus aequatorialis*, y así promover la conservación de la especie.

TABLAS

Tabla 1. Evaluación de calidad de todas las lecturas crudas, obtenidas de las 16 corridas de secuenciación, usando Nanoplot v1.30.1

Input	Longitud Media	Calidad Media	Número De Lecturas	Cobertura De Genoma	Bases Totales
Ca01,02,03,04,05,06,07,08	4,665.5	10.1	18,442,482	31.8X	86,044,031,537

Tabla 2. Comparación de QUAST v5.2.0 entre los ensamblados probados en este proyecto

Parámetros	SMARTdenovo*	Flye**
# Contigs	1,795	2,997
Tamaño Del Genoma	2.685 GB	2.704 GB
Contig Más Largo	16,862,666	55,344,372
N50	3,507,773	14,933,679
N90	961,766	2,825,112
L50	244	55
L90	797	210
Fracción Del Genoma % ***	90.011	90.132

*Genoma ensamblado con SMARTdenovo y que contiene solo lecturas mayores a 1000 pares de bases, de acuerdo con la filtración con Nanofilt

**Ensamblaje realizado con Flye, y con lecturas filtradas por tamaño, eliminando las menores a 2000 pb.

***Al alinear los ensamblados a la referencia, *Cebus albifrons*

Tabla 3. Evaluación de la integridad del genoma usando BUSCO v5.4.4 para los ensamblados realizados en este estudio

Parámetro	SMARTdenovo	Flye
BUSCOS completos*	87.8	97.9
Copia Única	65.6	96.3
Duplicados	2.2	1.6
BUSCOS fragmentados*	4.6	1.2
BUSCOS faltantes*	7.6	0.9
Total de Grupos BUSCO	13,780 (100%)	13,780 (100%)

*Información de los parámetros presentada como porcentajes

Tabla 4. Anotación estructural de genes para las ensamblados y comparación de herramientas de anotación.

Parámetros	SdN Maker	SdN Augustus	Flye Augustus
# Genes	28,952	22,479	21,710
# Exones	160,748	204,274	200,769
# Intrones	131,677	172,014	168,461
Tamaño Promedio De Genes	16,675	41,627	44,103
Tamaño Promedio De Exones	187	268	237

Tamaño Promedio De Intrones	3,424	4,972	5,207
Gen Más Largo	1,087,428	835,785	1,119,867

FIGURAS

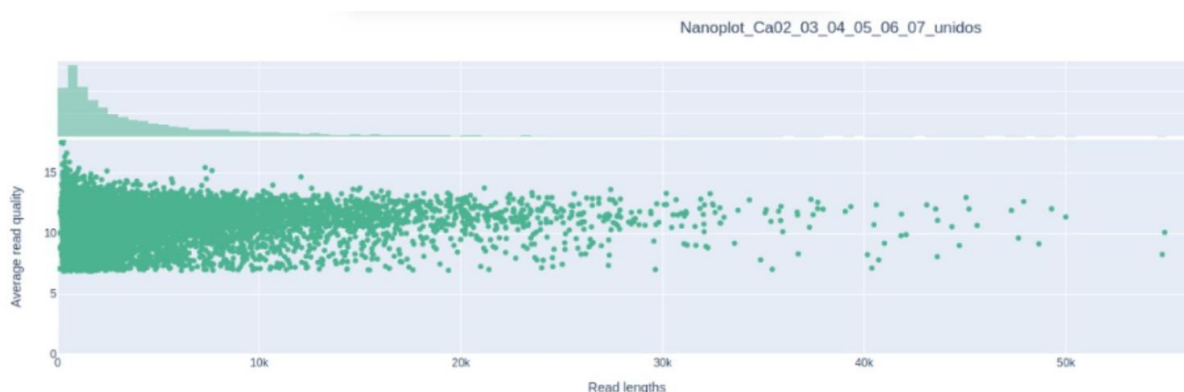


Figura 1. Diagrama de dispersión de puntos obtenido a partir de Nanoplot v1.30.1 para todas las lecturas crudas obtenidas de todas las corridas de secuenciación. El gráfico evalúa la calidad media de lectura (Q) en relación con el tamaño de la lectura (kb). En el gráfico superior se observa un histograma donde se evalúa solamente la longitud de las lecturas.

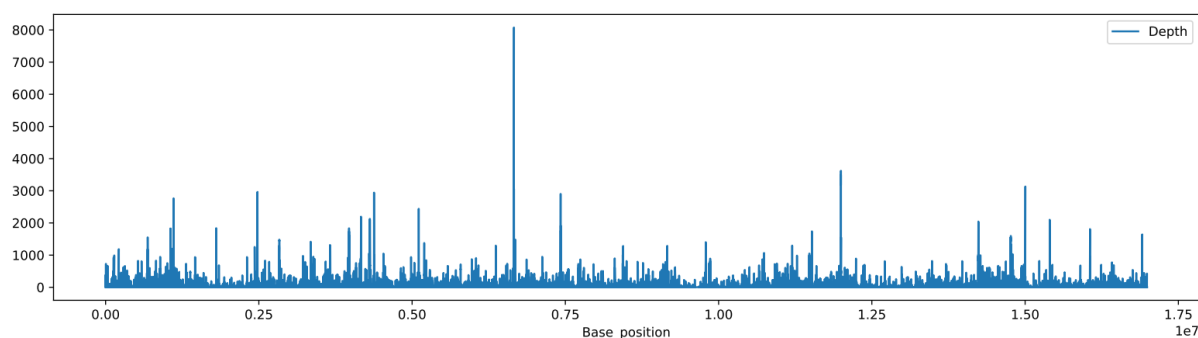


Figura 2. Gráfico de cobertura del mejor ensamblado de *Cebus aequatorialis*, que es el de Flye 2000. Se mapea un 5% de las lecturas crudas, seleccionadas aleatoriamente, sobre el contig más largo. El gráfico indica las diferentes bases del contig más largo y el número de lecturas crudas que se alinean en esa posición en específico, por lo que en general, se busca que la cobertura de las lecturas sea homogénea en todas las posiciones.

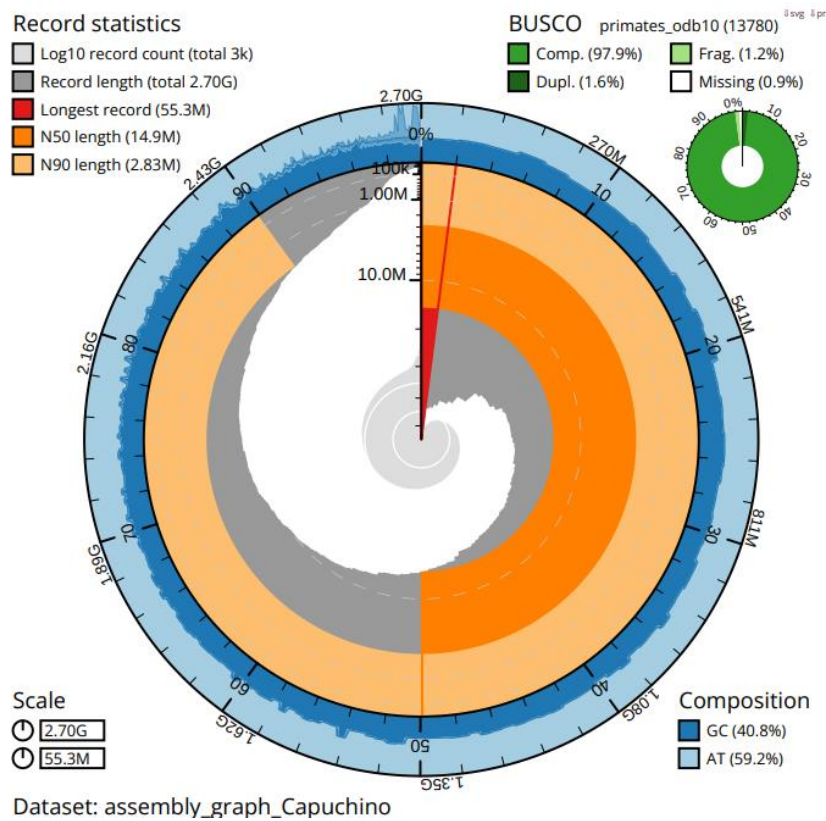


Figura 3. BlobToolKit Snailplot de Flye 2000, el mejor ensamblado de *Cebus aequatorialis*.

El gráfico muestra varias métricas de la evaluación del ensamblado, principalmente de QUAST v5.2.0 y BUSCO v5.4.4, en donde se destaca un tamaño total del genoma de 2.7 Gb; y el contig más largo, que mide 55.3 Mb. En la sección derecha superior del gráfico se puede observar un gráfico de resumen del número de BUSCOS completos, duplicados y fragmentados del ensamblado.

REFERENCIAS BIBLIOGRÁFICAS

- Adams, J. W., & Kasakoff, A. B. (1976). Factors underlying endogamous group size. In *Regional analysis* (pp. 149-173). Academic Press.
<https://www.sciencedirect.com/science/article/pii/B9780126521023500141>
- Allen, J. A., Miller, L. E., Carriker, M. A., & Richardson, W. B. (1914). New South American monkeys. *Bulletin of the AMNH*; v. 33, article 43.
<https://digitallibrary.amnh.org/bitstreams/9454f728-516d-4d32-95d8-8b75d027d9f9/download>
- Barrett, J. C., & Cardon, L. R. (2006). Evaluating coverage of genome-wide association studies. *Nature genetics*, 38(6), 659-662. https://www.researchgate.net/profile/Lon-Cardon/publication/7066849_Evaluating_coverage_of_genome-wide_association_studies/links/553e24a50cf294deef6fbab9/Evaluating-coverage-of-genome-wide-association-studies.pdf
- Boubli, J. P., Rylands, A. B., Farias, I. P., Alfaro, M. E., & Alfaro, J. L. (2012). C ebus Phylogenetic Relationships: A Preliminary Reassessment of the Diversity of the Untufted Capuchin Monkeys. *American Journal of Primatology*, 74(4), 381-393.
https://www.academia.edu/download/50770660/Cebus_Phylogenetic_Relationships_A_Preli20161207-23405-rmm5kn.pdf
- BGI-Shenzhen (2022). The genome of the white-fronted capuchin (*Cebus albifrons*). NCBI. [accessed 2024 October 17].
https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_023783575.1/

- Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes*, *10*(11), 846.
<https://www.mdpi.com/2073-4425/10/11/846/pdf>
- Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinformatics*, *48*(1).
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4286374/>
- Campos, F. A., & Jack, K. M. (2013). A potential distribution model and conservation plan for the critically endangered Ecuadorian capuchin, *Cebus albifrons* *aequatorialis*. *International Journal of Primatology*, *34*, 899-916.
- Cervera, L., Solórzano, M. F., Alfonso-Cortes, F., de la Torre, S., Fuentes, N., & Tirira, D. G. (2018). Capuchino ecuatoriano *Cebus aequatorialis* JA Allen, 1914. *PRIMATES DEL ECUADOR*, *55*. https://www.researchgate.net/profile/Galo-ZapataRios/publication/330533190_Estado_de_Conservacion_de_los_Primates_del_Ecuador/links/5c469ad492851c22a3870367/
- Culibrk L, Leelakumari S, Tse K, Cheng D, Chuah E, Kirk H, Pandoh P, Troussard A, Zhao Y, Mungall A, et al. 2019. The genome of the tufted capuchin (*Sapajus apella*). *NCBI*. [accessed 2024 October 17].
<https://www.ncbi.nlm.nih.gov/nuccore/WRPQ000000000.1/>
- Dainat J. 2020. Another Gff analysis toolkit to handle annotations in any GTF/GFF format (v0.8.0). *Zenodo*.
- 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>

- Dodson, C. H., & Gentry, A. H. (1991). Biological extinction in western Ecuador. *Annals of the Missouri Botanical Garden*, 273-295. <https://www.jstor.org/stable/2399563>
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 117(17), 9451–9457.
- Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., ... & Zammit, G. (2022). The era of reference genomes in conservation genomics. *Trends in ecology & evolution*, 37(3), 197-202.
- 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>
- Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, 82(11), 801-811.
- Jack, K. M., & Campos, F. A. (2012). Distribution, abundance, and spatial ecology of the critically endangered Ecuadorian capuchin (*Cebus albifrons aequatorialis*). *Tropical Conservation Science*, 5(2), 173-191. <https://journals.sagepub.com/doi/pdf/10.1177/194008291200500207>
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome biology*, 17, 1-11.

- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 30(9)
- Karst, S. M., Ziels, R. M., Kirkegaard, R. H., Sørensen, E. A., McDonald, D., Zhu, Q., ... & Albertsen, M. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nature methods*, 18(2), 165-169. <https://www.nature.com/articles/s41592-020-01041-y>
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 37(5):540–546
- 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., & Wang, J. (2019). Current challenges and solutions of de novo assembly. *Quantitative Biology*, 7(2), 90-109.
- Liem, M., Regensburg-Tuïnk, T., Henkel, C., Jansen, H., & Spaijk, H. (2021). Microbial diversity characterization of seawater in a pilot study using Oxford Nanopore Technologies long-read sequencing. *BMC Research Notes*, 14(1), 1-7. <https://bmcresearchnotes.biomedcentral.com/articles/10.1186/s13104-021-05457-3>
- 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 and Bioinformatics*, 14(5), 265-279. https://academic.oup.com/gpb/articlepdf/14/5/265/5088651/gpb_14_5_265.pdf

- Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Current Protocols*, 1, e323.
- Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*. 34(13)
- Moscoso, P., De la Torre, S., Cornejo, F., Mittermeier, R., Lynch, J., Heymann, E. (2021). *Cebus aequatorialis* (amended version of 2020 assessment). The IUCN Red List of Threatened Species 2021: e.T4081A191702052.
- Orkin, J. D., Montague, M. J., Tejada-Martinez, D., De Manuel, M., Del Campo, J., Cheves Hernandez, S., ... & Melin, A. D. (2021). The genomics of ecological flexibility, large brains, and long lives in capuchin monkeys revealed with fecalFACS. *Proceedings of the National Academy of Sciences*, 118(7), e2010632118. <https://www.pnas.org/doi/pdf/10.1073/pnas.2010632118>
- Org.one. (2021, October 20). Org.one. Oxford Nanopore Technologies. Retrieved September 19, 2024, from <https://org.one/oo>
- Oxford Nanopore Technologies Ltd. (2022, September 21). Medaka. *PyPI*. Retrieved October 13, 2024, from <https://pypi.org/project/medaka/>
- Pennisi, E. (2007). Boom time for monkey research. *Science*, 316(5822), 216-218. <https://www.science.org/doi/full/10.1126/science.316.5822.216>
- Ponti, G., Maccaferri, M., Manfredini, M., Kaleci, S., Mandrioli, M., Pellacani, G., ... & Tomasi, A. (2018). The value of fluorimetry (Qubit) and spectrophotometry (NanoDrop) in the quantification of cell-free DNA (cfDNA) in malignant melanoma and prostate cancer patients. *Clinica Chimica Acta*, 479, 14-19.

<https://www.academia.edu/download/90464161/j.cca.2018.01.00720220830-1-wicpz2.pdf>

- Pozo, G., Albuja-Quintana, M., Larreátegui, L., Gutiérrez, B., Fuentes, N., Alfonso-Cortés, F., & Torres, M. D. L. (2024). First whole-genome sequence and assembly of the Ecuadorian brown-headed spider monkey (*Ateles fusciceps fusciceps*), a critically endangered species, using Oxford Nanopore Technologies. *G3: Genes, Genomes, Genetics*, 14(3). <https://academic.oup.com/g3journal/article/14/3/jkae014/7581940>
- Rogers, J., & Gibbs, R. A. (2014). Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature Reviews Genetics*, 15(5), 347-359. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4113315/>
- Saenko SV, Groenenberg DSJ, Davison A, Schilthuizen M. 2021. The draft genome sequence of the grove snail *Cepaea nemoralis*. *G3 (Bethesda)*. 11
- Sangiovanni, M., Granata, I., Thind, A. S., & Guarracino, M. R. (2019). From trash to treasure: detecting unexpected contamination in unmapped NGS data. *BMC bioinformatics*, 20, 1-12.
- 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 (Oxford, England)*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Tvedte, E. S., Gasser, M., Sparklin, B. C., Michalski, J., Hjelman, C. E., Johnston, J. S., ... & Dunning Hotopp, J. C. (2021). Comparison of long-read sequencing technologies in interrogating bacteria and fly genomes. *G3*, 11(6), <https://academic.oup.com/g3journal/article/11/6/jkab083/6188627>

- Versoza, C. J., & Pfeifer, S. P. (2024). A hybrid genome assembly of the endangered aye-aye (*Daubentonia madagascariensis*). *G3: Genes, Genomes, Genetics*, 14(10),
- Wang, Y., Zhao, Y., Bolas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature biotechnology*, 39(11), 1348-1365. <https://www.nature.com/articles/s41587-021-01108-x>
- 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. Datos de Qubit de la extracción de ADN para antes y después de la creación de bibliotecas de secuenciación.

Muestra	Concentración Inicial (ng/μl)	Concentración Librería (ng/μl)
Ca01	28.2	35.7
Ca02	34.6	41
Ca03	47.2	37.2
Ca04	39.4	21
Ca05	15.36	27.8
Ca06	22.2	41.2
Ca07	17.72	31.8
Ca08	14.32	25.4
<i>Promedio</i>	27.38	33.8