

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

Genoma de referencia del mono araña de vientre amarillo *Ateles Belzebuth*, una especie en peligro de extinción

Juan Sebastian Moscoso Galarza

Ingeniería en Biotecnología

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero 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

**Genoma de referencia del mono araña de vientre amarillo *Ateles belzebuth*,
una especie en peligro de extinción**

Juan Sebastian Moscoso Galarza

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: Juan Sebastian Moscoso Galarza

Código: 00320743

Cédula de identidad: 0106657232

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UNPUBLISHED DOCUMENT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 araña de vientre amarillo (*Ateles belzebuth*) es un primate que se encuentra en peligro de extinción debido a la pérdida y fragmentación de su hábitat por actividades antropogénicas, como la tala, la minería, la caza para consumo y la expansión agrícola. Estas amenazas comprometen su supervivencia, lo que hace indispensable fortalecer los esfuerzos de conservación. En este contexto, los estudios genómicos se han convertido en una herramienta clave, ya que permiten comprender la diversidad genética y la conectividad poblacional, información esencial para diseñar estrategias de conservación efectivas. Sin embargo, estos estudios requieren un genoma de referencia de alta calidad. El genoma de referencia actual de *A. belzebuth* se encuentra altamente fragmentado en 110,868 contigs, por lo que este estudio tuvo como objetivo obtener un genoma de referencia de mejor calidad y menos fragmentado de *A. belzebuth*. Para ello, se realizó la extracción de ADN de una muestra sanguínea de *A. belzebuth* mediante el kit Dneasy Blood & Tissue de Qiagen, seguido por la preparación de liberías mediante Ligation Sequencing Kit v14 de ONT y la generación del ensamblado de referencia utilizando secuenciación de lecturas largas ONT y técnicas de ensamblaje *de novo*, como SMARTdenovo y Flye. La secuenciación generó 72,23 Gb de información comprendidas en alrededor de 7.3 millones de lecturas con una calidad promedio de 15.9. El ensamblado de Flye demostró ser superior, logrando un N50 de 12,780,818 bp, 2,088 contigs, y una mayor integridad genómica con un 98% de BUSCOs completos, superando significativamente al ensamblado actual disponible. Esto se realizó mediante ORG.one, una iniciativa de Oxford Nanopore Technologies (ONT) que promueve la secuenciación eficiente y ensamblado de alta calidad de especies amenazadas o en peligro de extinción.

Palabras clave: *Ateles belzebuth*, genómica de conservación, ORG.one, secuenciación de lecturas largas, Oxford Nanopore Technologies, ensamblaje *de novo*.

ABSTRACT

The white-bellied spider monkey (*Ateles belzebuth*) is an endangered primate due to habitat loss and fragmentation from anthropogenic activities such as logging, mining, hunting for consumption, and agricultural expansion. These threats seriously compromise its survival, making it essential to strengthen conservation efforts. In this context, genomic studies have become a key tool, as they allow understanding of genetic diversity and population connectivity, essential information for designing effective conservation strategies. However, these studies require a high-quality reference genome. The current reference genome of *A. belzebuth* is highly fragmented into 110,868 contigs, therefore this study aimed to obtain a better quality and less fragmented reference genome of *A. belzebuth*. To achieve this, DNA was extracted from a blood sample of *A. belzebuth* using the DNeasy Blood & Tissue Kit from Qiagen, followed by library preparation with the Ligation Sequencing Kit v14 from ONT, and the generation of the reference assembly using long-read sequencing with ONT and *de novo* assembly techniques, such as SMARTdenovo and Flye. The sequencing generated 72.23 Gb of data comprising around 7.3 million reads with an average quality of 15.9. The Flye assembly proved to be superior, achieving an N50 of 12,780,818 bp, 2,088 contigs, and greater genomic completeness with 98% of complete BUSCOs, significantly surpassing the currently available assembly. This was done through ORG.one, an Oxford Nanopore Technologies (ONT) initiative that promotes efficient sequencing and high-quality assembly of threatened or endangered species.

Key words: *Ateles belzebuth*, conservation genomics, ORG.one, long-read sequencing, Oxford Nanopore Technologies, *de novo* assembly.

TABLA DE CONTENIDO

Introducción	11
Métodos.....	15
Resultados	18
Discusión.....	20
Conclusiones	24
Tablas	25
Figuras.....	27
Referencias bibliográficas.....	28
Anexos	32

ÍNDICE DE TABLAS

Tabla 1. Resumen de NanoPlot para las lecturas obtenidas mediante la secuenciación ONT después del filtrado con NanoFilt, utilizando un umbral de calidad de Q9 y longitud de 1000 bases.....	25
Tabla 2. Comparación quast de ensamblado con SMARTdenovo, Flye y un ensamblado de referencia de NCBI	25
Tabla 3. Comparación busco de ensamblado con SMARTdenovo, Flye y ensamblado de NCBI.....	26
Tabla 4. Comparación de anotado estructural de ensamblados con SMARTdenovo y Flye..	26

ÍNDICE DE FIGURAS

Figura 1. Snail plot.....	27
----------------------------------	----

ÍNDICE DE ANEXOS

Anexo 1. Concentración de ADN en la extracción y en la librería	32
--	----

INTRODUCCIÓN

Características generales y distribución

Ateles belzebuth, comúnmente conocido como el mono araña de vientre amarillo, es uno de los primates más grandes y longevos que existen actualmente en el neotrópico (Stevenson, 2011). Se distingue físicamente por su cola prensil, una mancha triangular en la frente y un pelaje de color negro con tonalidades amarillentas o blanquecinas en la región ventral (Defler, 2010). *A. belzebuth* tiene una dieta principalmente frugívora, aunque en ciertas regiones, como el Yasuní, complementan su alimentación con insectos para suplir deficiencias proteicas (Pozo, 2004). Asimismo, vive en grupos numerosos de entre 15 y 56 individuos, en donde predominan las hembras adultas frente a los machos adultos (Shimooka et al., 2008).

Este mono araña se distribuye en los bosques de tierras bajas de la Amazonía occidental de Colombia, Ecuador y Perú, así como en la Amazonía norte en Venezuela y el noroeste de Brasil. Su distribución es irregular, con poblaciones dispersas en el noroeste de Colombia, el este de Ecuador, el norte de Perú y ciertas zonas de Venezuela y Brasil. En Ecuador, la especie habita principalmente al sur del Río Napo y se extiende hacia el piedemonte andino, aunque parece estar ausente hacia el sur de las áreas del Río Putumayo (Link et al., 2021).

Estado de conservación

Según la Lista Roja de la UICN, *A. belzebuth* está clasificado como una especie "En Peligro" de extinción, ya que se proyecta una reducción poblacional de al menos un 50% entre los años 2005 y 2050. Esta disminución se atribuye principalmente a la pérdida significativa de su entorno natural por actividades antropogénicas y a la presión continua de la caza excesiva. Específicamente, esta especie enfrenta una intensa amenaza de cacería para el consumo, el comercio ilegal de mascotas y la destrucción de hábitat provocada por la expansión agrícola, la tala y la minería (Rodríguez Bolaños, 2012). Estos factores podrían comprometer a las

poblaciones de Ecuador, Perú, Colombia y Brasil a nivel generacional, afectando aún más su estado de conservación en el futuro (Link et al., 2021).

Importancia del genoma de referencia

Tener un genoma de referencia de buena calidad es crucial para entender y abordar su conservación, ya que puede proporcionar información importante sobre su variabilidad genética, estructura poblacional e incluso adaptaciones evolutivas (Llanos et al., 2021). De hecho, a partir de la genómica de conservación se pueden identificar genes que son esenciales para la supervivencia de la especie y su capacidad de adaptación a cambios ambientales; además, contar con un genoma bien caracterizado ayuda a detectar problemas de consanguinidad y planificar cruzamientos con el fin de prevenir fenómenos como la endogamia y la pérdida de diversidad genética (ORG.one, 2021). Un genoma de referencia también facilita la identificación de áreas vulnerables en el genoma, lo que apoya decisiones informadas en la conservación de especies amenazadas. Este recurso es especialmente útil en regiones con alta biodiversidad y pocos recursos de investigación, ayudando a cerrar el déficit de información genética y permitiendo enfoques de conservación basados en datos precisos para la protección y recuperación de estas especies en riesgo (Linck & Cadena, 2024).

Genoma de referencia actual

El Instituto de Biología Evolutiva de la Universidad Pompeu Fabra, en colaboración con instituciones internacionales, construyeron un catálogo de diversidad genómica que incluye información de la secuenciación y ensamblado de 233 especies de primates, entre ellos, *A. belzebuth*. Este trabajo representó el primer ensamblado para la especie y fue realizado mediante la secuenciación con Illumina NovaSeq 6000 y el método de ensamblado Megahit (Kuderna et al., 2023). A pesar de que constituye una base sólida para el estudio de *A. belzebuth*, presenta los siguientes parámetros: 111,455 contigs, un valor N50 de 51,9 kb y L50

de 14,646 bp (National Center for Biotechnology Information, 2023). Estos valores indican un ensamblado altamente fragmentado con contigs relativamente pequeños, lo cual representa una menor continuidad y calidad del genoma (Hernández et al., 2020). Por lo tanto, para mejorar el ensamblado actual del mono araña, es de interés la implementación de diferentes tecnologías de secuenciación y herramientas de ensamblado.

ORG.one

ORG.one es una iniciativa de Oxford Nanopore Technologies que promueve la secuenciación rápida y eficiente de genomas de especies amenazadas o en peligro de extinción (Oxford Nanopore Technologies, 2021). La finalidad es generar genomas de referencia de alta calidad que servirán como herramientas para el desarrollo de estudios genómicos a futuro para mejorar los esfuerzos de conservación y preservación de las especie amenazadas. Además, estos datos genómicos son compartidos en una base de datos de libre acceso, en donde científicos a nivel internacional podrán desarrollar nuevos estudios genómicos o propuestas y sugerencias para guiar la conservación de las especies (ORG.one, 2021).

Secuenciación con Oxford Nanopore Technologies

Oxford Nanopore Technologies (ONT) es un método de secuenciación que emplea una tecnología basada en nanoporos. En este proceso, los ácidos nucleicos, como el ADN y ARN, atraviesan poros de tamaño nanométrico embebidos en una membrana, donde se detectan cambios específicos en la corriente eléctrica en función del tipo de nucleótido (Lin et al., 2021).

La tecnología ONT genera lecturas más largas y continuas en comparación con los métodos clásicos de lecturas cortas, permitiendo la construcción de ensamblados genómicos de mayor cobertura, calidad y completitud. De la misma manera, sus equipos de secuenciación son compactos, portátiles y tienen una gran capacidad de procesamiento continuo (Lin et al.,

2021). Por lo tanto, para mejorar la calidad del ensamblado actual de *A. belzebuth*, se puede emplear este tipo de tecnología de secuenciación.

Métodos de ensamblado

El ensamblado de genomas es un proceso mediante el cual se reconstruye una secuencia genómica completa a partir de fragmentos más cortos de ADN. Existen dos enfoques principales para este proceso: el ensamblado por comparación, que emplea un genoma de referencia que ha sido previamente estudiado, y el ensamblado *de novo*, que no emplea un genoma de referencia, sino se basa únicamente en los datos de secuenciación obtenidos (Aguilar & Falquet, 2015). En cuanto a las ventajas, el ensamblado por comparación requiere una menor capacidad de procesamiento, mientras que el ensamblado *de novo* se destaca por evitar sesgos de la referencia. Por este motivo, el ensamblado *de novo* es de interés en la generación de genomas de referencia (Reinaudi, 2020).

Objetivos

El objetivo de este estudio es obtener un genoma de referencia de mejor calidad y menos fragmentado de *Ateles belzebuth*. Para ello, en participación de la iniciativa ORG.one, se empleará Oxford Nanopore Technologies como tecnología de secuenciación y herramientas de ensamblado *de novo* como Flye y SMARTdenovo. En las siguientes secciones se detallan los métodos de secuenciación y ensamblado utilizados, así como los análisis posteriores para evaluar la calidad del genoma y su comparación con el genoma de referencia actual de *Ateles belzebuth*.

MÉTODOS

Obtención de muestra

Se obtuvo una muestra sanguínea de Corni, un ejemplar macho de *Ateles belzebuth* en el Centro de Rescate de Animales Silvestres amaZONico (Tena, Napo). La muestra fue recolectada durante un control veterinario bajo sedación cuando el individuo tenía 5 años de edad. La muestra fue dispensada en tubos K2 EDTA y se almacenó a -80 °C hasta su extracción.

Extracción de ADN

Se realizaron dos extracciones individuales de ADN a partir de la muestra sanguínea de Corni con el protocolo Dneasy Blood and Tissue Kit de Quiagen. Posteriormente, se realizó la cuantificación de las concentraciones de ADN mediante el fluorómetro Qubit, empleando 1 µL de ADN de cada extracción.

Preparación de librerías genómicas

Se utilizó el protocolo Ligation Sequencing Kit V14 (SQK-LSK114) con ciertas modificaciones. Primero, se cambió el tiempo de incubación de la preparación de extremos y reparación de ADN a 15 minutos a 20°C y 65°C. Segundo, en la ligación de adaptadores y limpieza del ADN, se cambió el tiempo de incubación a temperatura ambiente (20 minutos). Adicionalmente, después de cada paso se realizó un único lavado de beads.

Se cuantificó el ADN en el fluorómetro Qubit para verificar que las concentraciones se encuentren en un rango adecuado al final de cada paso. Por último, se preparó la celda de flujo y se cargaron las librerías genómicas para la secuenciación en PromethION.

Secuenciación

Se realizó la secuenciación de las dos librerías utilizando el equipo PromethION 2 Solo. Ambas librerías genómicas fueron procesadas en una única corrida, utilizando una sola celda

de flujo (FLO-PRO114M) durante 18 y 22 horas, respectivamente. Por último, se realizó el base calling con el software Guppy v 5.1.13 de Oxford Nanopore Technologies.

Análisis bioinformático

El pipeline bioinformático comenzó con el uso del software NanoFilt v2.8.0 (De Coster et al., 2018), el cual permitió filtrar las lecturas según parámetros de calidad y longitud. Para ello, se estableció un umbral con una calidad superior a Q9 y una longitud superior a 1,000 bases. Se utilizó el software Porechop v 0.2.4 (Wick et al., 2017), que identificó y eliminó los adaptadores que fueron ligados en los extremos de las lecturas durante la preparación de librerías (Andrade González & Pazmiño Carrera, 2024). Una vez eliminados los adaptadores, se utilizó el software NanoPlot v1.42.0 (De Coster & Rademakers, 2023). Esto permitió evaluar la calidad de las lecturas a través de resúmenes estadísticos y representaciones gráficas de los datos de secuenciación (Villalta et al., 2024). (Tabla 1).

Para el ensamblado se emplearon dos softwares especializados: Flye v 2.7.1 (Kolmogorov et al., 2019) y SMARTdenovo v1.0.0 (Liu et al., 2021). Estos programas, han sido diseñados para manejar lecturas largas obtenidas por tecnologías de secuenciación como Oxford Nanopore Technologies (Liu et al., 2021). Además, son ensambladores *de novo*, lo que significa que no requieren de una secuencia de referencia o conocimientos previos de la organización del genoma (Aguilar & Falquet, 2015). Luego, se utilizó el software Minimap2 v2.24 (Li, 2018) para ordenar los contigs en base a un genoma de referencia, mejorando de esta forma la organización y estructura del ensamblado (Leiva Sandoval et al., 2022). Para ello, se empleó la referencia de *Ateles fusciceps fusciceps* (GCA_036589455.1).

El ensamblado generado con SMARTdenovo fue pulido con Medaka v.1.7.2 (Oxford Nanopore Technologies, 2021). Este software corrige errores en los ensamblados a partir de un algoritmo que emplea una red neuronal recurrente de tipo Long Short-Term Memory (Luan et

al., 2024). El ensamblado generado con Flye empleó Racon v.1.5.0 (Vaser et al., 2017), una herramienta que alinea las lecturas largas al ensamblado, filtra alineamientos de baja calidad y deriva un consenso mediante grafos de alineamiento de orden no parcial (Luan et al., 2024).

Se realizó un segundo pulido a los dos ensamblados con POLCA MaSuRCA v.3.4.1 (Zimin & Salzberg, 2020). Este software permite corregir errores en los ensamblados utilizando lecturas cortas obtenidas con Illumina (Zimin & Salzberg, 2020). En este caso, se utilizaron las lecturas cortas de Illumina de *Ateles belzebuth* disponibles en NCBI (GCA_963573605.1).

Posteriormente, se utilizó el software QUAST v5.2.0 (Mikheenko et al., 2018). Este software permitió evaluar la calidad de los ensamblados a partir de un amplio rango de métricas y gráficos (Gurevich et al., 2013). En este caso, se empleó el software al ensamblado generado con SMARTdenovo, generado con Flye, y el ensamblado de referencia actual disponible en NCBI (GCA_963573605.1) Además, se utilizó el genoma de *Ateles fusciceps fusciceps* (GCA_036589455.1) para obtener estadísticos en base a una referencia.

Para medir la integridad genómica, se utilizó el software BUSCO v.5.7.1 (Simão et al., 2015), que evalúa la presencia de genes ortólogos conservados evolutivamente. Para ello, se empleó el set de datos del linaje de primates llamado primates_odb10 (Manni et al., 2021).

Por último, se realizó el anotado estructural mediante OmicsBox v.3.3.2 (Hoff & Stanke, 2019). Este software permite obtener datos genómicos a través de una combinación de métodos como BLAST, InterProScan y análisis de vías metabólicas (Coll, 2020). En este caso, la anotación estructural se realizó para los ensamblados genómicos de SMARTdenovo y Flye.

RESULTADOS

Extracción de ADN

Las concentraciones de ADN obtenidas tras la extracción de las muestras sanguíneas fueron de 62.0 ng/μl y 81.2 ng/μl en un volumen total de 60 μl (Anexo 1). Tras finalizar el protocolo de preparación de librerías, se obtuvo valores de 447.84 y 499.20 ng totales de ADN (Anexo 1).

Análisis de las lecturas obtenidas de la secuenciación con ONT

Los resultados de NanoPlot obtenidos de las lecturas filtradas con un umbral de calidad de Q9 y un tamaño mínimo de 1000 bases (Tabla 1) arrojaron un total de 7,355,881 lecturas, que corresponden a 72,23 Gb de data y una cobertura de 27.7X. Estas lecturas presentaron una longitud media de 9,819.80 bases, una calidad promedio de 15.90 y un valor N50 de 15,745 bases.

Evaluación calidad del ensamblado

El ensamblado con SMARTdenovo cubrió el 98.045% del genoma de referencia, obtuvo 2,180 contigs con una longitud total de 2,668,804,955 bp, un valor N50 de 3,476,143 bases, 40.87% de contenido GC y su contig más largo fue 15,233,524 bp (Tabla 2). En contraste, Flye cubrió el 98.407% del genoma de referencia, generó 2,088 contigs, con una longitud total de 2,648,976,60 bp, un N50 de 12,780,818, 40.82% de contenido GC y su contig más largo fue de 60,944,116 bp (Tabla 2). El ensamblado de NCBI (GCA_963573605.1) cubrió el 96,578% del genoma, generó 110,868 contigs con una longitud total de 2,589,204,212 bases, un N50 de 51,919 bases, un 40.79% de contenido GC y su contig más largo fue de 460,063 bases (Tabla 2).

Evaluación integridad del ensamblado

Para evaluar la integridad de los tres ensamblados se utilizó el set de datos de primates_odb10 que contiene 13,780 genes obtenidos a partir de 25 especies de primates (Manni et al., 2021). En la Tabla 3 se pueden observar los resultados de BUSCO de los tres ensamblados genómicos, en donde Flye presentó un 98.0% de genes completos con un 96.1% de genes de copia única y 1.9% de genes duplicados. SMARTdenovo presentó un 97.1% de genes completos con un 95.1% de genes de copia única y 2.0% de genes duplicados. El ensamblado de NCBI (GCA_963573605.1) obtuvo 69.5% de genes completos, 68.3% de genes de copia única y 1.2% de genes duplicados. Por otra parte, en cuanto a los genes fragmentados, Flye presentó un 1.3%, SMARTdenovo un 1.6% y el ensamblado de NCBI un 13.9%. Finalmente, en cuanto a los genes faltantes, Flye presentó un 0.7%, SMARTdenovo un 1.3% y el ensamblado de NCBI un 16.6%.

Anotación estructural:

En cuanto a los resultados de anotación estructural obtenidos a partir de OmicsBox (Tabla 4), se identificaron 24,396 genes en el ensamblado de SMARTdenovo y 24,252 en Flye. Respecto a los intrones, se encontraron 157,137 en SMARTdenovo y 157,960 en Flye. Con respecto a los exones, el resultado obtenido con Flye (182,212) fue ligeramente mayor a los resultados obtenidos con SMARTdenovo (181,203). En cuanto al tamaño promedio de genes, en SMARTdenovo se obtuvo un valor de 32,276 bases y Flye de 32,314. Por último, el tamaño promedio de los exones también mostró una ligera variación, con 205 para SMARTdenovo y 190 para Flye.

DISCUSIÓN

Secuenciación

La secuenciación mediante Oxford Nanopore Technologies generó alrededor de 7.3 millones de lecturas y 72.23 Gb, lo que corresponde a una cobertura de 27.7X del genoma de *A. belzebuth*. Esta cobertura es adecuada para obtener un ensamblado de referencia de alta calidad según lo demostrado en estudios previos de secuenciación ONT en primates (Pozo et al., 2024). Además, se ha observado que valores cercanos a 30X pueden cubrir hasta un 95% de la fracción del genoma (Sims et al., 2014). Por otra parte, los datos obtenidos se destacan por su excelente calidad, reflejada en un Q score de 15.9, que indica un basecalling de lecturas altamente confiable y una baja probabilidad de error por base (Edgar & Flyvbjerg, 2015). Asimismo, este valor supera ampliamente a la calidad observada en otros estudios de secuenciación en primates, en donde los valores Q score oscilan entre 10 y 11 (Lee et al., 2024; Pozo et al., 2024). Por último, las lecturas presentaron una longitud media de 9,819 bases y un N50 de 15,745 bp, indicando la presencia de fragmentos largos que permiten ensamblar genomas grandes de manera más continua y precisa (Chaisson et al., 2009).

Ensamblado

El objetivo de investigación se cumplió exitosamente. Se obtuvieron dos ensamblados de alta calidad, que resultaron ser significativamente superiores al genoma de referencia actual de *A. belzebuth*. En este caso, los ensamblados de Flye y SMARTdenovo muestran una mejor calidad y continuidad por varias razones. En primer lugar, presentaron un menor número de contigs, lo que indica que el ensamblado presenta un menor grado de fragmentación, que corresponde a un parámetro indicativo de mayor continuidad genómica (Andrade González & Pazmiño Carrera, 2024; Aguilar-Bultet & Falquet, 2015). En segundo lugar, los valores de N50 obtenidos son bastante altos, lo que refleja la presencia de contigs

de mayor tamaño y una menor fragmentación del ensamblado (Mäkinen et al., 2012). Por último, los valores del contig más largo demuestran que estos ensamblados lograron abarcar secciones más extensas y continuas del genoma (Baker, 2012; Sorber et al., 2008).

Al comparar los resultados de los ensamblados, Flye sobresalió en todos los aspectos, pues reportó el mayor valor N50 (12,780,818 bp), el contig más largo (60.94 Mb) y el menor número de contigs (2,088 contigs), superando al ensamblado de SMARTdenovo, que presentó un N50 de 3,476,143 bp, un contig máximo de 15.23 Mb y 2,180 contigs. El ensamblado de NCBI mostró una calidad significativamente inferior, con un N50 de 51,919 bp, un contig máximo de 460 kb y 110,868 contigs. Por otra parte, en cuanto a los resultados basados en la referencia de *A. fusciceps fusciceps* (GCA_036589455.1), Flye obtuvo una mayor fracción de este genoma (98.407%), superando tanto al ensamblado de SMARTdenovo (98.045%) como al de NCBI (96.578%). Además, los valores obtenidos para su longitud total (2,648,975,690 bp) y contenido GC (40.82%) fueron valores cercanos a los reportados en la referencia, con una longitud total de 2,639,265,159 bp y un 40.85% de contenido GC (Pozo et al., 2024).

Integridad de ensamblados con respecto a BUSCO

BUSCO corresponde a otro parámetro importante para la evaluación de la calidad de ensamblados en términos de integridad genómica (Simão et al., 2015). Bajo este enfoque, los ensamblados de SMARTdenovo y Flye evidencian una calidad superior en comparación con el genoma de referencia actual de *A. belzebuth* (Tabla 3). En primer lugar, estos ensamblados superaron significativamente el umbral del 95% de BUSCOs completos, lo cual corresponde a un indicador de alta calidad en genomas de referencia de organismos modelo (Seppey et al., 2019). En segundo lugar, estos ensamblados presentaron una mayor cantidad de BUSCOs completos de copia única y una menor cantidad de BUSCOs completos duplicados, lo que sugiere que ambos ensamblados han capturado de manera precisa y sin redundancia los genes

esenciales que normalmente se encuentran en una sola copia en el 90% de las especies de la base de datos de primates_odb10 (Simão et al., 2015). Por último, presentaron una menor cantidad de BUSCOs fragmentados y faltantes, lo que significa que la calidad de los datos de secuenciación y las estrategias de ensamblado fueron adecuadas (BioBam, 2024)

Es importante destacar que, al contrastar los resultados obtenidos (Tabla 3), Flye presenta la mayor integridad genómica. Este ensamblado reportó el mayor número de BUSCOs completos (98.0%), seguido por el ensamblado de SMARTdenovo (97.1%) y el de NCBI (69.5%). Además, en cuanto a los BUSCOs de copia única y duplicados, Flye presentó un 96.1% y 1.9%, respectivamente, superando al ensamblado de SMARTdenovo (95.1% y 2%) y NCBI (68.3% y 1.2%). Por último, Flye presentó la menor cantidad de BUSCOs fragmentados y faltantes con 1.3% y 0.7% respectivamente, mientras que el ensamblado de SMARTdenovo obtuvo 1.6% y 1.3% y el de NCBI presentó porcentajes más altos de 13.9% y 16.6%.

Anotación estructural

Se evaluaron los elementos genómicos presentes en los ensamblados de SMARTdenovo y Flye mediante la anotación estructural con OmicsBox, en donde se observaron diferencias leves. Por una parte, el ensamblado de Flye reportó 24,252 genes, mientras que SMARTdenovo 24,396 genes. Estos resultados se encuentran dentro del rango observado en primates del Nuevo Mundo, que varía desde aproximadamente 20,000 genes, como en *S. apella* (GCA_009761245.1; Culibrk et al., 2021) y *S. boliviensis boliviensis* (GCA_016699345.2; Chiou et al., 2011), hasta los 31,417 genes registrados en *A. fusciceps fusciceps* (Pozo et al., 2024).

Por otra parte, el ensamblado con Flye reportó 157,960 intrones y 182,212 exones y, el ensamblado con SMARTdenovo 157,137 y 181,203. Estos valores son similares a los reportados por Pozo et al. (2024) para *A. fusciceps*, con 149,050 intrones y 183,970 exones, lo

que sugiere una consistencia en la anotación entre especies de una misma familia. Por otra parte, Flye reportó un tamaño promedio de genes de 32,314 bp y SMARTdenovo 32,276, aproximándose a lo reportado en otras especies de primates del Nuevo Mundo, en donde se ha observado un tamaño promedio que varía aproximadamente entre los 17,000 y 40,000 bp (Culibrk et al., 2019; Pozo et al., 2024). Por último, Flye reportó un tamaño promedio de exones de 190 bp, más cercano al valor de 176 bp descrito para *A. fusciceps fusciceps* (Pozo et al., 2024) en comparación con SMARTdenovo, el cual mostró un valor más distante de 205 bp.

Finalmente, tras la evaluación de los ensamblados a partir de los distintos parámetros, se destaca que el ensamblado de Flye constituye un nuevo genoma de referencia de alta calidad, significativamente superior al ensamblado actual de *A. belzebuth*. De hecho, las características sobresalientes de este nuevo ensamblado se resumen en el Snail plot (Figura 1), donde su mayor continuidad se representa visualmente por el menor número de contigs, el valor N50 más alto y el contig más largo. Adicionalmente, presenta una gráfica de BUSCO, que confirma su mayor integridad genómica a través del mayor porcentaje de genes completos.

El nuevo genoma de referencia de Flye es muy importante, ya que puede ser utilizado como una herramienta para el desarrollo de estudios genómicos a futuro, como, por ejemplo, estudios de diversidad genética (Hidalgo, 2003). En este contexto, un genoma de referencia de alta calidad, como el obtenido con Flye, puede ser utilizado para alinear las secuencias de nuevos ejemplares de *A. belzebuth* para la identificación de SNPs (Vazquez, 2019). Estos marcadores moleculares son muy importantes, ya que permiten estudiar la diversidad entre poblaciones, identificar de cuellos de botella, endogamia, entre otros (Llanos et al., 2021). Una vez realizados los estudios genómicos, es posible diseñar estrategias de conservación específicas para salvaguardar la especie, que van desde programas de manejo genético, hasta la restauración de hábitats (Linck & Cadena, 2024; ORG.one, 2021).

CONCLUSIONES

Los 72,23 Gb de datos obtenidos mediante la secuenciación ONT, que representan una cobertura de 27.7X del genoma de *Ateles belzebuth*, constituyen un volumen significativo de información y una cobertura adecuada para generar un ensamblado de referencia de alta calidad, como se ha demostrado en estudios previos en primates. Los parámetros de calidad de las lecturas, como la longitud media de 9,819.80 bases y el puntaje de calidad promedio de 15.90, sugieren una buena calidad con fragmentos largos y basecalling confiable.

Los análisis de ensamblaje demostraron que el método de Flye superó al ensamblado con SMARTdenovo y al de referencia disponible en NCBI. Flye obtuvo el menor número de contigs con 2,088, un valor N50 de 12,780,818 bp y el contig más largo de 60,944,116 bp. Además, el ensamblado de Flye cubrió el 98.407% del genoma de referencia de *Ateles fusciceps fusciceps*, y presentó una alta integridad con 98.0% de genes completos y bajos porcentajes de genes fragmentados y faltantes según los resultados de BUSCO.

En cuanto a la anotación estructural, los ensamblados generados con Flye y SMARTdenovo mostraron resultados similares en parámetros como, el número de genes, intrones, exones y tamaños promedios. Estos valores se encuentran dentro de los rangos esperados para especies de primates del Nuevo Mundo, lo que sugiere una consistencia en la anotación entre especies de una misma familia.

La obtención de un genoma de referencia de alta calidad e integridad para *Ateles belzebuth* representa un avance significativo para la conservación de esta especie en peligro de extinción. Este recurso constituye el primer paso para desarrollar estrategias efectivas de conservación, siendo necesario continuar con estudios de diversidad genética, estructura poblacional y adaptaciones evolutivas para implementar programas de manejo y protección más precisos.

TABLAS

Tabla 1. Resumen de NanoPlot para las lecturas obtenidas mediante la secuenciación ONT después del filtrado con NanoFilt, utilizando un umbral de calidad de Q9 y longitud de 1000 bases.

Input	Longitud media (bp)	Calidad media	Número de lecturas	N50 (bp)	Bases Totales (Gb)	Cobertura
<i>Lecturas filtradas (Q9 y 1000 bases)</i>	9,819.80	15.90	7,355,881.00	15,745.00	72,23	27.7X

Tabla 2. Comparación QUAST de ensamblado con SMARTdenovo, Flye y un ensamblado de referencia de NCBI

	SMARTdenovo	Flye	Ensamblado NCBI
<i>Fracción del genoma (%)</i>	98.045	98,407	96,578
<i>Longitud Total (bp)</i>	2,668,804,955	2,648,975,690	2,589,204,212
<i>Número de contigs</i>	2,180	2,088	110,868
<i>Contig más largo (bp)</i>	15,233,524	60,944,116	460,063
<i>N50 (bp)</i>	3,476,143	12,780,818	51,919
<i>GC (%)</i>	40.87	40.82	40,79

Tabla 3. Comparación BUSCO de ensamblado con SMARTdenovo, Flye y ensamblado de NCBI

Parámetro	SMARTdenovo		Flye		NCBI	
<i>BUSCOs completos</i>	13,381	97.1%	13,506	98.0%	9,580	69.5%
- <i>BUSCOs copia única</i>	13,106	95.1%	13,247	96.1%	9,414	68.3%
- <i>BUSCOs duplicados</i>	275	2.0%	259	1.9%	166	1.2%
<i>BUSCOs fragmentados</i>	225	1.6%	174	1.3%	1,922	13.9%
<i>BUSCOs faltantes</i>	174	1.3%	100	0.7%	2,278	16.6%
<i>BUSCOs totales</i>	13,780	100%	13,780	100%	13780	100%

Tabla 4. Comparación de anotado estructural de ensamblados con SMARTdenovo y Flye

Parámetro	SMARTdenovo	Flye
<i>Número de genes</i>	24,396	24,252
<i>Número de intrones</i>	157,137	157,960
<i>Número de exones</i>	181,203	182,212
<i>Tamaño promedio de genes</i>	32,276	32,314
<i>Tamaño promedio de exones</i>	205	190

FIGURAS

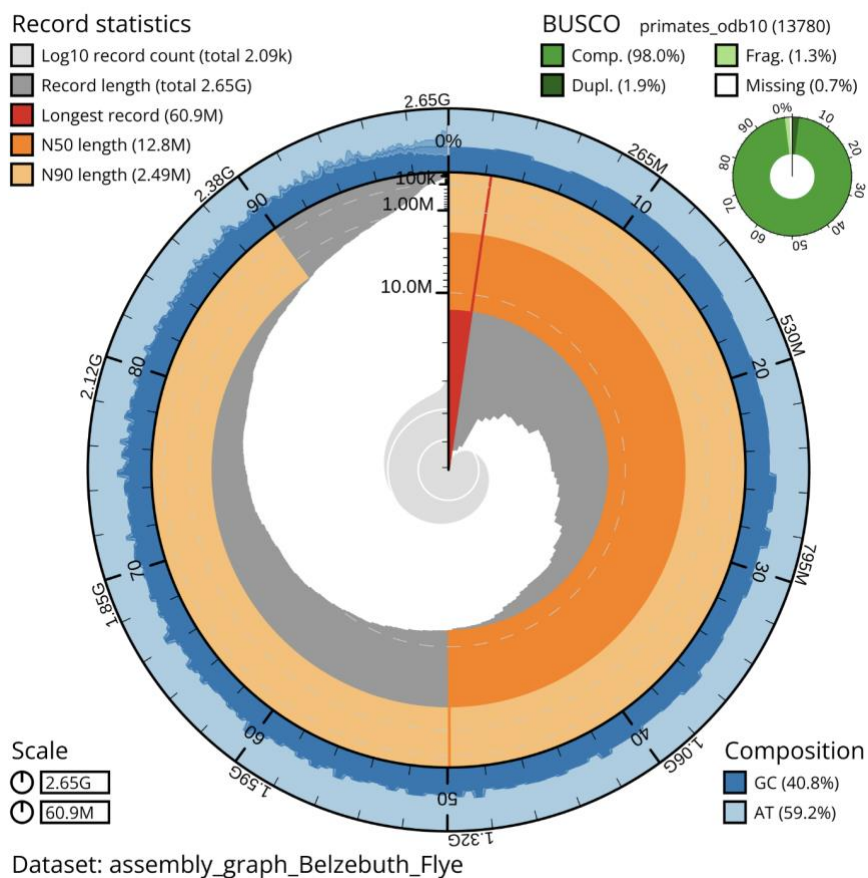


Figura 1. Snail Plot.

El gráfico Snailplot de BlobToolKit muestra las métricas del ensamblado final de Flye del genoma de *Ateles belzebuth*. En este gráfico, se observa un tamaño total del genoma de 2.65 Gb, con un contig más largo de 60.9 Mb. Además, se presentan los valores de N50 y N90, que alcanzan longitudes de 12.8 Mb y 2.49 Mb, respectivamente. La composición del genoma en términos de bases nitrogenadas corresponde a un 40.8% de GC y un 59.2% de AT. En la esquina superior derecha, se incluyen los resultados de la evaluación de integridad genómica mediante el análisis de genes BUSCO, utilizando la base de datos primates_odb10. Los resultados indican un 98% de genes completos, un 1.3% fragmentados y un 0.7% ausentes. Esto sugiere una alta calidad del ensamblado, lo cual es fundamental para los análisis posteriores de conservación y variabilidad genética.

REFERENCIAS BIBLIOGRÁFICAS

- Aguilar-Bultet, L., & Falquet, L. (2015). Secuenciación y ensamblaje de novo de genomas bacterianos: una alternativa para el estudio de nuevos patógenos. *Revista de salud animal*, 37(2), 125-132. http://scielo.sld.cu/scielo.php?pid=S0253-570X2015000200008&script=sci_arttext&tlng=pt
- Andrade González, K. S., & Pazmiño Carrera, C. M. (2024). *Ensamblaje de genoma de Sphingobium yanoikuyae a partir de datos de secuenciación de Illumina y PACBIO archivados en la base de datos de NCBI* (Bachelor's thesis). <https://dspace.ups.edu.ec/bitstream/123456789/28673/1/TTQ1739.pdf>
- Baker, M. (2012). De novo genome assembly: what every biologist should know. *Nature methods*, 9(4), 333-337. <https://www.nature.com/articles/nmeth.1935/1000>
- BioBam. (2024). *Genome Completeness Assessment with BUSCO*. BioBam. <https://www.biobam.com>
- Chaisson, M. J., Brinza, D., & Pevzner, P. A. (2009). *De novo fragment assembly with short mate-paired reads: Does the read length matter?*. *Genome research*, 19(2), 336-346. <https://genome.cshlp.org/content/19/2/336.short>
- Chiou, K. L., Pozzi, L., Alfaro, J. W. L., & Di Fiore, A. (2011). Pleistocene diversification of living squirrel monkeys (*Saimiri* spp.) inferred from complete mitochondrial genome sequences. *Molecular Phylogenetics and Evolution*, 59(3), 736-745. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_016699345.2/
- Coll, M. B. (2020). *De novo genome assembly and annotation of fungus in OmicsBox*. BioBam. <https://www.biobam.com/de-novo-genome-assembly-and-annotation-of-sheath-rot-fungus-sarocladium-oryzae/>
- Culibrk, L., Leelakumari, S., Tse, K., Cheng, D., Chuah, E., Kirk, H., Pandoh, P., Troussard, A., Zhao, Y., Mungall, A., Moore, R., Marra, M. A., Sinclair-Smith, T. and Jones, S. J. M. (2019). *The genome of the Tufted Capuchin (Sapajus apella)*. NCBI. <https://academic.oup.com/g3journal/article/14/3/jkae014/7581940>
- 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://academic.oup.com/bioinformatics/article-abstract/34/15/2666/4934939>
- De Coster, W., & Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*, 39(5), btad311. <https://academic.oup.com/bioinformatics/article-abstract/39/5/btad311/7160911>
- Defler, T. R. (2010). *Historia natural de los primates colombianos*. Universidad Nacional de Colombia. https://books.google.com/books?hl=es&lr=&id=2VWtDwAAQBAJ&oi=fnd&pg=PP1&dq=Primates+of+Colombia+defler&ots=BfpmjZaxi4&sig=zlcE_ne7WXoZkgXCq rPNsbbCwTs
- Edgar, R. C., & Flyvbjerg, H. (2015). *Error filtering, pair assembly and error correction for next-generation sequencing reads*. *Bioinformatics*, 31(21), 3476-3482. <https://academic.oup.com/bioinformatics/article-abstract/31/21/3476/194979>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075. <https://academic.oup.com/bioinformatics/article-abstract/29/8/1072/228832>
- Hernández, M., Quijada, N. M., Rodríguez-Lázaro, D., & Eiros, J. M. (2020). Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico

- clínico. *Revista argentina de microbiología*, 52(2), 150-161.
<https://www.sciencedirect.com/science/article/pii/S0325754119300811>
- Hidalgo, R. (2003). Variabilidad genética y caracterización de especies vegetales. *Boletín técnico*, (8), 2-26. <https://books.google.com/books?hl=es&lr=&id=B55X-G3WiugC&oi=fnd&pg=PA2&dq=genoma+de+referencia+estudios+de+diversidad+genetica&ots=i61tsTHEZG&sig=YHfdSuAFPrDA9qgMWFGDyFWbJB0>
- Hoff, K. J., & Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. *Current protocols in bioinformatics*, 65(1), e57.
<https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpbi.57>
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology*, 37(5), 540-546.
<https://www.nature.com/articles/s41587-019-0072-8>
- Kuderna, L. F., Gao, H., Janiak, M. C., Kuhlwilm, M., Orkin, J. D., Bataillon, T., ... & Marques Bonet, T. (2023). A global catalog of whole-genome diversity from 233 primate species. *Science*, 380(6648), 906-913.
<https://www.biorxiv.org/content/10.1101/2023.05.02.538995.full.pdf>
- Lee, F. C., Sitam, F. T., & Tan, L. P. (2024). Little Influence of DNA Quality on the Direct Sequencing Output of Non-human Primates' Faecal Samples. *Journal of Virological Methods*, 115074.
<https://www.sciencedirect.com/science/article/pii/S016609342400199X>
- Leiva Sandoval, A. M., Gil Ordoñez, A., Bartolini, I., Olortegui, J. A., Velasquez, R., & Cuéllar, W. J. (2022). Evaluación del sistema flongle (Oxford Nanopore) de bajo costo para secuenciación de patógenos.
<https://cgspace.cgiar.org/bitstream/handle/10568/128183/report.pdf?sequence=1>
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094-3100. <https://academic.oup.com/bioinformatics/article-abstract/34/18/3094/4994778>
- Lin, B., Hui, J., & Mao, H. (2021). Nanopore technology and its applications in gene sequencing. *Biosensors*, 11(7), 214. <https://www.mdpi.com/2079-6374/11/7/214>
- Linck, E. B., & Cadena, C. D. (2024). A latitudinal gradient of reference genomes. *Molecular Ecology*, e17551. <https://onlinelibrary.wiley.com/doi/10.1111/mec.17551>
- Link, A., Palacios, E., Stevenson, P.R., Boubli, J.P., Mittermeier, R.A., Shantee, S., Urbani, B., de la Torre, S., Cornejo, F.M., Moscoso, P., Mourthé, Í., Muniz, C.C. & Rylands, A.B. (2021). Ateles belzebuth (amended version of 2019 assessment). The IUCN Red List of Threatened Species 2021: e.T2276A191684587.
<https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T2276A191684587.en>
- Liu, H., Wu, S., Li, A., & Ruan, J. (2021). SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte*, 2021.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9632051/>
- Llanos, Y. V., Toro, E. J. R., Paz, W. B., Sarmiento, R. M., Gallego, R. O., & Cerón-Muñoz, M. (2021). *Estructura genética de poblaciones de ganado Blanco Orejinegro en Colombia*.
- Luan, T., Commichaux, S., Hoffmann, M., Jayeola, V., Jang, J. H., Pop, M., ... & Luo, Y. (2024). Benchmarking short and long read polishing tools for nanopore assemblies: achieving near-perfect genomes for outbreak isolates. *BMC genomics*, 25(1), 679.
<https://link.springer.com/article/10.1186/s12864-024-10582-x>
- Mäkinen, V., Salmela, L., & Ylinen, J. (2012). Normalized N50 assembly metric using gap-restricted co-linear chaining. *BMC bioinformatics*, 13, 1-5.
<https://link.springer.com/article/10.1186/1471-2105-13-255>

- Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: assessing genomic data quality and beyond. *Current Protocols*, 1(12), e323. <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpz1.323>
- National Center for Biotechnology Information. (2023). Ateles belzebuth genome assembly, GCA_963573605.1. NCBI Datasets. https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_963573605.1/
- ORG.one. (2021). ORG.one. Oxford Nanopore Technologies. Retrieved October 7, 2024, from <https://nanoporetech.com/es/oo>
- Oxford Nanopore Technologies. (19 de abril de 2021). *Introducing ORG.one: A project to support rapid DNA sequencing of critically endangered species*. Oxford Nanopore Technologies Recuperado el 28 de septiembre de 2024 de <https://nanoporetech.com/news/news-introducing-orgone-project-support-rapid-dna-sequencing-critically-endangered-species>
- Oxford Nanopore Technologies Ltd. (11 de octubre de 2024). Medaka. PyPI. Recuperado el 29 de octubre de 2024 de <https://pypi.org/project/medaka/>
- 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), jkae014. <https://academic.oup.com/g3journal/article-abstract/14/3/jkae014/7581940>
- Pozo, W. E. (2004). Agrupación y dieta de *Ateles belzebuth belzebuth* en el Parque Nacional Yasuní, Ecuador. *Anuar Investigación Científica*, 2, 77-102. https://www.researchgate.net/profile/Wilmer-E-Pozo-R/publication/263082933_AGRUPACION_Y_DIETA_DE_ATELES_BELZEBUTH_BELZEBUTH_EN_EL_PARQUE_NACIONAL_YASUNI_ECUADOR/links/0deec539b419bbb3a8000000/AGRUPACION-Y-DIETA-DE-ATELES-BELZEBUTH-BELZEBUTH-EN-EL-PARQUE-NACIONAL-YASUNI-ECUADOR.pdf
- Reinaudi, C. P. (2020). *Eficiencia en el ensamblado De Novo de secuencias de RNA-Seq. Estudio de la performance de algoritmos* (Doctoral dissertation, Universidad Nacional de Luján).
- Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. *Gene prediction: methods and protocols*, 227-245. https://link.springer.com/protocol/10.1007/978-1-4939-9173-0_14
- Shimooka, Y., Campbell, C. J., Di Fiore, A., Felton, A. M., Izawa, K., Link, A., ... & Wallace, R. B. (2008). 12 Demography and group composition of *Ateles*. *Spider monkeys: Behavior, ecology and evolution of the genus Ateles*, 55, 329. <https://books.google.com/books?hl=es&lr=&id=Ys7LCgAAQBAJ&oi=fnd&pg=PA329&dq=demography+and+group+composition+in+Ateles&ots=5FdXgC4TDdb&sig=Mae6n2DII7KI0B-fHA1FNiSMgMU>
- 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://academic.oup.com/bioinformatics/article-abstract/31/19/3210/211866>
- 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://www.nature.com/articles/nrg3642/1000>
- Rodríguez Bolaños, A. (2012). Modelos predictivos de distribución para los micos atelinos (*Lagothrix* y *Ateles*) en Colombia.

- <https://repositorio.uniandes.edu.co/entities/publication/60c124c7-2c0f-4232-a2dd-765e95f4e822>
- Sorber, K., Chiu, C., Webster, D., Dimon, M., Ruby, J. G., Hekele, A., & DeRisi, J. L. (2008). The long march: a sample preparation technique that enhances contig length and coverage by high-throughput short-read sequencing. *PLoS One*, 3(10), e3495. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0003495>
- Stevenson, P. R. (2011). The abundance of large ateline monkeys is positively associated with the diversity of plants regenerating in neotropical forests. *Biotropica*, 43(4), 512-519. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1744-7429.2010.00708.x><https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1744-7429.2010.00708.x>
- Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome research*, 27(5), 737-746. <https://genome.cshlp.org/content/27/5/737.short>
- Vazquez, D. V. (2019). *Alineado y comparación de secuencias genómicas obtenidas de grupos discrepantes para la detención de regiones cromosómicas que controlan caracteres de frutos de tomate* (Master's thesis). <https://rephip.unr.edu.ar/bitstreams/d6c28fa8-c705-4564-9d15-457fc0b81a3e/download>
- Villalta, D. R., Villegas, D. B., Hidalgo, B. A., Solórzano, L. M., & Tubito, C. C. (2024). Caracterización de las tecnologías de secuenciación genética de segunda y tercera generación. *Tecnología en Marcha*, 37(2), 83-94. <https://dialnet.unirioja.es/servlet/articulo?codigo=9435488>
- 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), e000132. <https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000132>
- Zimin, A. V., & Salzberg, S. L. (2020). The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS computational biology*, 16(6), e1007981. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007981>

ANEXOS

Anexo 1. Concentración de ADN en la extracción y en la librería

Muestra	Concentración extracción ADN	Concentración librería
<i>BeC07</i>	62.0 ng/ μ l en 60 μ l	447,84 ng
<i>BeC08</i>	81.2 ng/ μ l en 60 μ L	499,20 ng