

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

Primer genoma de referencia de *Scalesia gordilloi*, una especie endémica de la isla San Cristóbal, Galápagos

Emilio Vélez Darquea

Biología

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

Quito, 9 de mayo de 2025

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 de *Scalesia gordilloi*, una especie endémica de
la isla San Cristóbal, Galápagos****Emilio Vélez Darquea****Nombre del profesor, Título académico****Gabriela Pozo Andrade, M.B.S**

Quito, 9 de mayo de 2025

© 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: Emilio Vélez Darquea

Código: 00326442

Cédula de identidad: 0105974059

Lugar y fecha: Quito, 9 de mayo de 2025

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UNPUBLISHED DOCUMENT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

Scalesia gordilloi es una especie endémica de la isla San Cristóbal perteneciente a una única población localizada en la costa suroeste. Debido a diferentes amenazas, como la introducción de especies herbívoras, fragmentación de hábitat, y el turismo, actualmente se encuentra catalogada por la UICN como una especie en peligro crítico de extinción. Actualmente, no se tiene datos genéticos de la especie, por lo que el objetivo de este estudio fue secuenciar y ensamblar el primer genoma de referencia de *S. gordilloi* para tener una herramienta que permita realizar estudios genéticos y genómicos. Para ello, se extrajo ADN del tejido foliar de un único individuo mediante el protocolo CTAB, y se prepararon las librerías genómicas con el kit Ligation Sequencing (SQK-LSK114) de Oxford Nanopore Technologies. Después de la secuenciación, se utilizaron las herramientas bioinformáticas SMARTdenovo y Flye para el ensamblaje, y Quast y BUSCO para el análisis de calidad y completitud. Se obtuvo una cobertura de 32X en 14,790,278 de lecturas con un tamaño medio de 6,985 pb y una calidad media de 12.9. Los resultados de Flye fueron superiores a los de SMARTdenovo, consiguiendo una menor fragmentación, un N50 de 341,921 pb, y una completitud alta, ya que se identificó el 97.12% de los genes ortólogos. Los resultados obtenidos representan un avance significativo para futuros estudios relacionados con la ecología, evolución, y conservación de la *S. gordilloi*. Sin embargo, el ensamblaje todavía presenta un alto índice de fragmentación, por lo que es necesario continuar trabajando en su mejora para conseguir una mejor calidad a futuro.

Palabras clave: Oxford Nanopore Technologies, secuenciación de tercera generación, conservación, ORG.one, peligro crítico de extinción, *Scalesia gordilloi*.

ABSTRACT

Scalesia gordilloi is an endemic species of San Cristóbal Island, belonging to a single population located on the southwest coast. Due to various threats, such as the introduction of herbivorous species, habitat fragmentation, and tourism, it is currently listed by the IUCN as a critically endangered species. Currently, there is no genetic data for the species, so the objective of this study was to sequence and assemble the first reference genome of *S. gordilloi* to provide a tool for genetic and genomic studies. To this end, DNA was extracted from leaf tissue of a single individual using the CTAB protocol, and genomic libraries were prepared with the Ligation Sequencing kit (SQK-LSK114) from Oxford Nanopore Technologies. After sequencing, the bioinformatics tools SMARTdenovo and Flye were used for assembly, and Quast and BUSCO for quality and completeness analysis. A 32X coverage was obtained in 14,790,278 reads with an average read size of 6,985 bp and an average quality score of 12.9. Flye's results were superior to those of SMARTdenovo, achieving lower fragmentation, an N50 of 341,921 bp, and high completeness, as 97.12% of orthologous genes were identified. The results obtained represent a significant advance for future studies related to the ecology, evolution, and conservation of *S. gordilloi*. However, the assembly still exhibits a high level of fragmentation, so further work is needed to improve it to achieve better quality in the future.

Keywords: Oxford Nanopore Technologies, third-generation sequencing, conservation, ORG.one, critically endangered, *Scalesia gordilloi*.

TABLA DE CONTENIDO

Introducción	10
Islas Galápagos y el género <i>Scalesia</i>	10
Desafíos en la conservación de especies amenazadas	11
Iniciativa ORG.one y ONT	12
Metodología	14
Toma de muestras	14
Extracción de ADN y purificación	14
Preparación de librerías.....	14
Secuenciación	15
Análisis bioinformático.....	15
Resultados	18
Extracción	18
Secuenciación	18
Calidad del ensamblado	19
Compleitud del genoma.....	19
Anotación estructural	20
Discusión.....	22
Secuenciación	22
Calidad del ensamblaje	22
Compleitud del genoma.....	24
Anotación estructural	26
Conclusiones	27

Referencias bibliográficas.....	28
Anexo A: Concentración de ADN de la extracción y de la librería final	35

ÍNDICE DE TABLAS

Tabla 1. Resultados de la secuenciación de las muestras de un individuo de <i>Scalesia gordilloi</i> , provenientes de la isla San Cristóbal (Galápagos), obtenidos mediante ONT. Se presentan los parámetros del número de lecturas generadas, el tamaño medio de las lecturas, calidad y la cobertura promedio con relación al tamaño esperado.....	18
Tabla 2. Resultados de la calidad del ensamblado de SMARTdenovo y Flye. Se presentan los parámetros relacionados con la fracción cubierta del genoma con relación a <i>S. atracyloides</i> , cantidad total de contigs, longitud del contig más largo, longitud total del ensamblaje, y las métricas de calidad N50 y L50.	19
Tabla 3. Comparación de la completitud del genoma determinada con BUSCO para SMARTdenovo y Flye. Se destaca la cantidad de genes completos encontrados, el número de copias únicas, cantidad de genes duplicados, la cantidad de genes fragmentados, y genes no encontrados.....	20
Tabla 4. Anotación estructural del ensamblaje de Flye. Se presentan los parámetros reportados, que incluyen la cantidad de genes, exones e intrones identificados, así como el tamaño promedio de genes y exones.....	20

INTRODUCCIÓN

Islas Galápagos y el género *Scalesia*

Las Islas Galápagos son un archipiélago perteneciente al territorio ecuatoriano. Están localizadas en el Océano Pacífico a una distancia de aproximadamente 1000 km del continente (Lanteri, 1997). Su aislamiento ha favorecido procesos biológicos adaptativos y de especiación, ya que los organismos que logran llegar deben acomodarse y adaptarse a los nuevos ambientes desafiantes (Kelley et al., 2019). Un ejemplo de ello es el género *Scalesia* presente en las Galápagos. Este género pertenece a la familia de plantas Asteraceae y está conformado por un total de 15 especies, de las cuales 12 son arbustivas y tres son arbóreas (Itow, 1995). Al igual que los pinzones de Darwin, las especies del género *Scalesia* son el resultado de una diversificación y especiación rápida que tuvo lugar a partir de un ancestro común (Fernández-Mazuecos et al., 2020). Dicho proceso se conoce como radiación adaptativa, y ha sido un foco de interés para estudios evolutivos en tiempo real (Fernández-Mazuecos et al., 2020). Las poblaciones del género *Scalesia* se encuentran separadas unas de otras, e incluso algunas están limitadas a islas específicas, como el caso de *Scalesia gordilloi* (Itow, 1995)..

La especie *S. gordilloi* es endémica de la isla de San Cristóbal, y se encuentra restringida a una única población localizada en la costa suroeste (Tye & Nykanen, 2015). La especie es evolutivamente cercana a *S. divisa* y *S. incisa*, y se encuentra en la zona seca baja de la isla cerca de la playa La Lobería, donde crece junto a matorrales pequeños y rocas volcánicas sueltas (Hamann & Wium-Andersen, 1986). Sin embargo, factores como la pérdida de hábitat e impactos generados por la introducción de especies herbívoras y el turismo amenazan a su población (Tye & Nykanen, 2015). Las cabras ferales representan una de las principales amenazas debido al pastoreo intensivo (Itow, 1995). El turismo, que

contaba con un manejo sustentable hasta los años 80, ha aumentado significativamente y, debido a su estrecha relación con la introducción de especies no nativas, actualmente constituye una de las mayores amenazas para la biodiversidad en las islas (Muñoz Barriga, 2015). Como resultado, la *S. gordilloi* fue catalogada por la IUCN como una especie en peligro crítico de extinción (Tye & Nykanen, 2015).

Desafíos en la conservación de especies amenazadas

La pérdida de especies es uno de los problemas más grandes que enfrenta el mundo actualmente, y diferentes actividades antropogénicas, como la producción de gases de efecto invernadero, contaminación, y deforestación son algunos de los principales factores que han influido en el proceso (Kaiho, 2022). Actualmente el 28% de todas las especies del planeta se encuentran amenazadas, y es probable que, dentro de 40 a 60 años, entre el 20 y 50% lleguen a extinguirse si no se reduce el impacto humano (Kaiho, 2022).

Las regiones del Sur Global poseen la mayor biodiversidad del planeta, sin embargo también presentan la mayor cantidad de especies amenazadas (Tydecks et al., 2018). Esto dificulta la conservación de especies en peligro, ya que mientras el Norte Global lidera la investigación científica, las latitudes más bajas sufren una desigualdad en los esfuerzos investigativos, lo que causa vacíos de conocimiento sobre la distribución de especies y sesgos en la comprensión de los patrones de biodiversidad y diversificación (Linck & Cadena, 2024).

Esta desigualdad investigativa no solo afecta a la comprensión de la biodiversidad, sino también limita la generación de recursos clave para la conservación, como el desarrollo de investigaciones relacionadas a la genética y genómica. Como consecuencia de distintos factores históricos que llevaron a que los esfuerzos investigativos se concentren en el Norte Global, las latitudes más bajas enfrentan un déficit de genomas de referencia en comparación

con su alta diversidad biológica (Linck & Cadena, 2024). Esto se debe principalmente a limitantes como el desarrollo económico en la región y la concentración del capital educativo en países desarrollados (Linck & Cadena, 2024). Por lo tanto, para reducir esta brecha, y generar un impacto significativo en la conservación del planeta, es necesario incentivar la investigación en los países del Sur Global mediante la colaboración y financiamiento por parte de instituciones que dispongan de la capacidad económica (Linck & Cadena, 2024).

En este sentido, la obtención de genomas de referencia se ha convertido en una herramienta crucial para la conservación. La importancia de un genoma de referencia radica en que en el ADN se encuentra almacenada toda la información necesaria para el desarrollo y funcionamiento de un organismo vivo (Oxford Nanopore Technologies, s. f.). La secuenciación significa desglosar esta información y contribuir a la conservación de una especie, ya que permite descifrar incertidumbres como su historia evolutiva, diversidad genética o alelos deletéreos (Oxford Nanopore Technologies, s. f.). Un genoma de referencia sirve como una herramienta crucial para posteriores estudios comparativos de secuencias específicas que serían muy difíciles o imposibles de realizar si no se dispone de datos genómicos previos (Worley et al., 2017). De esta manera se puede comprender el nivel de variación genética en una población, así como también facilitar comparaciones filogenéticas entre especies cercanas (Worley et al., 2017).

Iniciativa ORG.one y ONT

Ante estos desafíos, han surgido diferentes proyectos como ORG.one, cuyo objetivo es secuenciar genomas completos de especies amenazadas o en peligro crítico mediante tecnologías de Oxford Nanopore (Oxford Nanopore Technologies, s. f.). Una vez obtenidos, los genomas serán publicados en bases de datos de libre acceso para que cualquier persona

pueda desarrollar estudios centrados en su conservación (Oxford Nanopore Technologies, s. f.).

Oxford Nanopore Technologies (ONT) es una tecnología de secuenciación de lecturas largas de tercera generación que funciona a partir de nanoporos (Weirather et al., 2017). Cuando una molécula de ADN atraviesa los nanoporos, cada base nitrogenada produce un cambio único en la corriente eléctrica que se interpretará en una señal específica (Weirather et al., 2017). Una vez realizada la secuenciación, el ensamblaje del genoma se puede ejecutar con un genoma de referencia o *de novo* mediante herramientas bioinformáticas. El ensamblaje *de novo* consiste en construir el genoma de un organismo desde cero a partir de las lecturas generadas durante la secuenciación (Liao et al., 2019). Las lecturas se superponen a medida que las secuencias coincidan, formando fragmentos más grandes denominados contigs, que luego se agruparán para formar scaffolds (Liao et al., 2019).

Por lo tanto, como parte de la iniciativa de ORG.one, el objetivo principal de este proyecto consiste en secuenciar y ensamblar el primer genoma de referencia de *Scalesia gordilloi*. Para llevar a cabo este proceso, se utilizarán tecnologías de secuenciación de Oxford Nanopore, y se ensamblará el genoma mediante las herramientas bioinformáticas Flye y SMARTdenovo.

Secuenciar el genoma de *S. gordilloi* no solo es una contribución a la comprensión genética de esta especie, sino también al género *Scalesia*. La disposición del genoma de referencia de *S. gordilloi* significa un gran avance para su conservación, además de facilitar el entendimiento de la historia evolutiva y biodiversidad de este grupo de plantas.

METODOLOGÍA

Toma de muestras

Se colectaron muestras de tejido foliar de siete individuos de *S. gordilloi* pertenecientes a la única población existente de la especie cerca de la playa La Lobería en San Cristóbal. Se trabajó con las muestras de un único individuo. Las muestras fueron refrigeradas a -20°C al momento la recolección. Posteriormente, se trasladaron al laboratorio de Biotecnología Vegetal de la Universidad San Francisco de Quito, donde se conservaron a -80 °C hasta el inicio de la extracción.

Extracción de ADN y purificación

El protocolo de extracción se realizó en base a la metodología CTAB propuesta por Doyle & Doyle (1987) con modificaciones. Según el artículo, la metodología es efectiva para extraer fragmentos largos de ADN de tejidos frescos, y ha funcionado correctamente con diferentes grupos de angiospermas. Después de la extracción se empleó un fluorómetro de Qubit 4 (Thermo Fisher Scientific, s. f.) para determinar la concentración de ADN, donde se utilizó 1 uL de la muestra y 199 uL de Working Solution. Finalmente se realizó un paso de purificación con Zymo Genomic Cleanup Kit (Zymo Research, s. f.) previo a la preparación de librerías para eliminar fragmentos pequeños y cualquier tipo de contaminación que pueda interferir en la secuenciación.

Preparación de librerías

La librería se preparó siguiendo el protocolo Ligation Sequencing Kit V14 (SQK-LSK114) de Oxford Nanopore Technologies con modificaciones. Durante este proceso se repararon los fragmentos de ADN y se prepararon sus extremos para la adición de los adaptadores con las enzimas NEBNext FFPE DNA Repair Mix y Ultra II End Repair Mix. Al

finalizar, se incubaron las muestras en un termociclador a 20°C por 5 min, y después a 65°C por 5 min. Finalmente, se realizó una purificación del ADN utilizando AMPure XP Beads.

Después de preparar el ADN, se utilizó el reactivo NEBNext Quick T4 DNA Ligase para unir los adaptadores de secuenciación a los fragmentos. Se incubó la muestra por 20 min a temperatura ambiente, y se realizó una limpieza con AMPure XP Beads y el buffer LFB. Finalmente, se cuantificó la concentración de ADN de la librería con el fluorómetro Qubit 4 (Thermo Fisher Scientific, s. f.) para determinar que la concentración sea adecuada y proseguir con la secuenciación.

Secuenciación

Se realizaron dos secuenciaciones de la librería mediante el instrumento de secuenciación PromethION 2 Solo en dos celdas de flujo R.10.4.1. En cada celda se dispensaron 300 ng de la librería junto con el buffer y perlas de secuenciación, en un volumen total de 200 uL. Durante la secuenciación, se programó el software MinKNOW para filtrar las lecturas con una calidad superior o igual a Q7. La primera secuenciación se realizó durante 48 horas, mientras que la segunda corrió durante 28 horas y 40 min. Durante las secuenciaciones se llevó a cabo el basecalling de alta precisión (HAC, por sus siglas en inglés) mediante el software MinKNOW. Este tipo de basecalling es más lento que otros modelos, pero permite obtener lecturas más fidedignas y con menos errores (Pater et al., 2021).

Análisis bioinformático

El primer paso del análisis bioinformático consistió en eliminar los adaptadores de las lecturas con la herramienta Porechop, ya que no pertenecen a la secuencia de ADN original (Quan et al., 2023). Posteriormente, se filtraron las lecturas menores a 2000 pb mediante el

software Nanofilt, y se determinó la calidad de las lecturas mediante el programa NanoPlot (Mock et al., 2023).

El ensamblaje del genoma se realizó con dos programas, Flye y SMARTdenovo. Ambos programas bioinformáticos son utilizados para el ensamblaje *de novo* de fragmentos largos de ADN, pero funcionan con algoritmos diferentes (Kolmogorov et al., 2019; Liu et al., 2021). Estas herramientas se encargan de agrupar los fragmentos generados de la secuenciación para formar lecturas más largas y continuas que contribuirán a la construcción del genoma (Liao et al., 2019). Al finalizar el ensamblaje, se utilizó Minimap2, una herramienta bioinformática que permite mapear lecturas con bases de datos referenciales (Li, 2018), para alinear los contigs con el genoma de referencia de *S. atracyloides* (GCA_947069175.1), una especie cercana a *S. gordilloi*. El genoma de *S. atracyloides* tiene una alta calidad. Fue secuenciado con la tecnología PacBio Sequel, y tiene un tamaño aproximado de 3.22 Gb representados en 1,329 scaffolds (Cerca et al., 2022). Además, representa una alta completitud e integridad, con un 93.3% de genes esperados identificados con BUSCO, y 43,093 genes anotados (Cerca et al., 2022). Posteriormente, se utilizó la herramienta de pulido Racon para corregir los posibles errores del ensamblado. Este programa se basa en utilizar las secuencias mapeadas para construir gráficos de alineamiento de orden parcial, que definirán la secuencia consenso (Vaser et al., 2017). Después se utilizó Quast, un programa que emplea métricas de calidad propias y de diferentes softwares (Gurevich et al., 2013), para determinar la calidad de los ensamblados realizados con SMARTdenovo y Flye.

Debido a que existen estudios que sugieren que todas las especies de *Scalesia* son tetraploides (Cerca et al., 2022), se realizó un paso previo al análisis de completitud para generar un ensamblaje haploide mediante el software Purge Haplotigs. Usualmente cuando se realizan ensamblajes *de novo* con alta heterocigosidad en determinadas regiones, los

algoritmos suelen ensamblarlas como contigs independientes en vez de fusionarlos en uno solo, lo que da como resultado un genoma más grande de lo esperado que puede producir complicaciones en futuros estudios (Roach et al., 2018). Purge Haplontigs funciona adecuadamente con lecturas de tercera generación y genomas grandes (Roach et al., 2018). Esta herramienta se encarga de identificar y eliminar contigs duplicados que pueden generarse debido a la heterocigosidad del genoma, lo que mejora la calidad del ensamblaje (Roach et al., 2018).

Posteriormente se utilizó BUSCO para analizar la completitud del genoma. Este software permite identificar la presencia de genes ortólogos conservados que deberían estar presentes en el ensamblaje, basándose en grupos taxonómicos de la base de datos OrthoDB (Seppey et al., 2019). En este caso se utilizó la base de datos de eudicotiledóneas (Eudicots_Odb10), que cuenta con un total de 2,326 regiones ortólogas.

Finalmente, se utilizó OmicsBox para realizar el anotado estructural del genoma. La anotación es un proceso que permite localizar e identificar genes presentes en la secuencia genómica (Guigó, 2023). OmicsBox trabaja con diferentes recursos bioinformáticos como CloudBlast, InterProScan, o Blast2GO que se encargan de identificar la función de un determinado gen (Coll, 2020). Los genes son regiones de ADN que generan productos funcionales, ya sea en forma de proteínas provenientes de regiones codificantes, o de diferentes tipos de ARN no codificante (Ejigu & Jung, 2020). Además, la anotación también puede identificar pseudogenes, que, aunque anteriormente no se consideraban funcionales, ahora se sabe que pueden participar en la regulación de la expresión génica (Ejigu & Jung, 2020). En este caso, las evidencias utilizadas para la anotación genómica fueron proteínas de *Helianthus annuus*, una especie perteneciente a la familia Asteraceae, igual que *S. gordilloi* (Hamann & Wium-Andersen, 1986; Nicolson & Human, 2013).

RESULTADOS

Extracción

La concentración de ADN obtenida durante la extracción para la muestra fue de 302 ng/uL en un volumen total de 50 uL. Después de la purificación mediante el kit Zymo Genomic Cleanup Kit (Zymo Research, s. f.), la concentración se redujo a 41.2 ng/uL en un volumen total de 49.8 uL. Finalmente, al preparar las librerías se obtuvo una concentración de 680 ng totales en aproximadamente 25 uL.

Secuenciación

La cantidad de lecturas generadas de ambas secuenciaciones se pudo observar mediante un análisis con Nanoplot. En total se produjeron 14,790,278 lecturas con un N50 de 9,707 pb, y un tamaño medio de 6,985 pb. La calidad media de las lecturas generadas fue de 12.9, y se obtuvieron 103,311,749,249 bases totales (**Tabla 1**). Esto equivale a una cobertura aproximada de 32X, tomando como referencia el tamaño del genoma de *S. atracyloides* (GCA_947069175.1) que mide 3.2 Gb (Cerca et al., 2022).

Tabla 1. Resultados de la secuenciación de las muestras de un individuo de *Scalesia gordilloi*, provenientes de la isla San Cristóbal (Galápagos), obtenidos mediante ONT. Se presentan los parámetros del número de lecturas generadas, el tamaño medio de las lecturas, calidad y la cobertura promedio con relación al tamaño esperado.

Parámetros	Nanoplot
<i>Lecturas generadas</i>	14,790,278
<i>Tamaño medio de lecturas (pb)</i>	6,985
<i>N50 (pb)</i>	9,707
<i>Calidad media</i>	12.9
<i>Cantidad de bases (Gb)</i>	103.31
<i>Cobertura</i>	32X

Calidad del ensamblado

El análisis de calidad con Quast del ensamblado realizado con SMARTdenovo evidenció que se cubrió el 68.89% del tamaño esperado en comparación con el genoma de *S. atracyloides* (GCA_947069175.1), con una longitud total de 2,947,442,928 pb. La cantidad total de contigs generados fue de 48,093, donde el más largo llegó a medir 538,187 pb. El valor de N50 resultó en un tamaño de 77,564 pb, mientras que el L50 estuvo representado en 11,819 contigs (**Tabla 2**).

Por otra parte, los resultados obtenidos con Flye indicaron una mayor cobertura del genoma, alcanzando el 84.04% de lo esperado, con una longitud total de 3,993,992,042 pb distribuida en 41,617 contigs. En este caso, el fragmento más largo fue de 8,697,309 pb, el N50 de 341,921 pb, y el L50 estuvo conformado por 2586 contigs (**Tabla 2**).

Tabla 2. Resultados de la calidad del ensamblado de SMARTdenovo y Flye. Se presentan los parámetros relacionados con la fracción cubierta del genoma con relación a *S. atracyloides*, cantidad total de contigs, longitud del contig más largo, longitud total del ensamblaje, y las métricas de calidad N50 y L50.

Parámetros	SMARTdenovo_2000	Flye_2000
Fracción del genoma	68.89%	84.04%
Contigs	48,093	41,617
Contig más largo (pb)	538,187	8,697,309
Longitud total (Gb)	2.95	3.99
N50 (pb)	77,564	341,921
L50	11,819	2,586

Compleitud del genoma

Los resultados obtenidos de BUSCO evidencian que se lograron identificar la mayoría de los genes ortólogos, tanto para el ensamblaje de Flye como de SdN. La base de datos

Eudicots_Odb10 alberga 2,326 genes, de los cuales se identificaron 2,259 (97.12%) en Flye, y 2,239 (96.27%) en SdN. De los genes identificados, 802 (34.48%) eran de copia única en Flye, y 883 (37.96%) en SdN. Por otra parte, los genes duplicados fueron 1,457 (62.64%) en Flye, y 1,356 (58.30%) en SdN. De los genes que se buscaron, solamente 11 (0.47%) resultaron fragmentados para Flye, y 56 (2.41%) no se encontraron, mientras que en SdN, 17 (0.73%) estaban fragmentados, y 70 no se encontraron (3.01%) (**Tabla 3**).

Tabla 3. Comparación de la completitud del genoma determinada con BUSCO para SMARTdenovo y Flye. Se destaca la cantidad de genes completos encontrados, el número de copias únicas, cantidad de genes duplicados, la cantidad de genes fragmentados, y genes no encontrados.

Genes ortólogos	SMARTdenovo_200	Flye_2000
Completos	2,239	2,259
Únicos	883	802
Duplicados	1,356	1,457
Fragmentados	17	11
No encontrados	70	56

Anotación estructural

La anotación estructural con OmicsBox se realizó únicamente para el ensamblaje de Flye, ya que los resultados del ensamblaje eran superiores a los de SMARTdenovo (Tabla 4). En total se identificaron 619,766 genes, 1,575,101 exones, y 958,012 intrones. El tamaño medio de los genes fue de 1,337 pb, mientras que el de los exones fue de 365 pb.

Tabla 4. Anotación estructural del ensamblaje de Flye. Se presentan los parámetros reportados, que incluyen la cantidad de genes, exones e intrones identificados, así como el tamaño promedio de genes y exones.

Parámetros	Anotación Flye_2000
Genes	619,766

<i>Tamaño promedio de genes</i>	1,337 pb
<i>Exones</i>	1,575,101
<i>Tamaño promedio de exones</i>	365 pb
<i>Intrones</i>	958,012

DISCUSIÓN

Secuenciación

En base a los resultados obtenidos, se puede concluir que las lecturas obtenidas durante la secuenciación son adecuadas para realizar el ensamblaje del genoma. En total se obtuvieron 103.31 Gb de información, equivalente a una cobertura del 32X con respecto al tamaño estimado del genoma, tomando como referencia el genoma de *S. atracyloides* (GCA_947069175.1) que tiene un tamaño de 3.2 Gb. La cantidad de información es suficiente, ya que en general, una cobertura de 30X se considera aceptable para ensamblar a partir de lecturas secuenciadas con tecnologías Next Generation Sequencing (NGS) (Cliften, 2015).

La calidad media de las lecturas fue de 12.9, equivalente a una precisión aproximada del 94.9%. Este valor se considera adecuado, ya que supera ampliamente a lo reportado en estudios similares de secuenciación con ONT, como el de la secuenciación de *Vaccinium floribundum*, donde se obtuvo una calidad promedio de 8.9 (Albuja-Quintana et al., 2024). Por otra parte, en total se generaron casi 15 millones de lecturas con un tamaño medio de 6,985 pb y un N50 de 9,707 pb, representando un tamaño considerablemente alto. Esto se debe a que mientras más largas sean las lecturas de secuenciación los resultados serán mejores, ya que permiten un ensamblaje más preciso, sobre todo cuando el genoma tiene muchas regiones repetitivas (Reinert et al., 2015).

Calidad del ensamblaje

El ensamblaje del genoma se realizó utilizando tanto SMARTdenovo como Flye. Sin embargo, los resultados de Flye son significativamente superiores. Con Flye se logró cubrir el 84.04% del tamaño esperado con respecto al genoma de *S. atracyloides* (GCA_947069175.1), mientras que SdN cubrió el 68.89%, demostrando una mayor cobertura

por parte de Flye. Por otra parte, Flye presenta 41,617 contigs, en comparación a SMARTdenovo que tiene 48,093. Una mayor cantidad de contigs, significa un genoma más fragmentado (Alneberg et al., 2013), por lo que los resultados de Flye son mejores en comparación con SMARTdenovo. En cuanto al contig más largo, Flye otorgó un valor de 8,697,309 pb, superando significativamente al de SMARTdenovo que mide 538,187 pb. Mientras más largo sea un contig, los resultados serán mejores, ya que significa una mayor continuidad del ensamblado (Chakraborty et al., 2016).

En cuanto a la longitud total, se puede apreciar que los resultados de Flye no solo superan a los de SMARTdenovo, sino también al tamaño del genoma de referencia de *S. atractyloides* (GCA_947069175.1). Esto posiblemente se debe a que el ensamblaje todavía presenta un alto índice de fragmentación (41,617 contigs), donde los contigs no ensamblados se repiten en el ensamblaje, generando una sobreestimación del tamaño real del genoma. La fragmentación del genoma podría deberse a solapamientos erróneos de contigs, por lo general producidos por instrumentos de secuenciación de una sola molécula, que suelen tener altas tasas de error (Koren & Phillippy, 2015). Además, diferentes estudios han evidenciado que las plantas presentan varias dificultades durante el proceso de secuenciación y ensamblaje genómico, principalmente por la complejidad de sus genomas, que por lo general presentan tamaños extensos y varias regiones repetitivas (Kubis, 1998; Liao et al., 2019).

Sin embargo, a pesar de que el genoma se encuentre fragmentado en aproximadamente 40,000 contigs, esto no constituye una limitación. En la base de datos NCBI existen reportes de genomas de referencia con índices de fragmentación mucho más altos, inclusive superando los 100 mil contigs. Un ejemplo de ello es la especie *Solanum phureja* (GCA_009849755.1), cuyo genoma está fragmentado en 229,084 contigs.

Aparte del tamaño del ensamblaje, es importante considerar el N50 y L50 para comprender su continuidad. El N50 es una medida que indica el tamaño del contig más pequeño de los contigs más grandes que abarcan el 50% del genoma, mientras que el L50 es una medida que indica en cuantos contigs se encuentra representado el 50% del genoma (Jayakumar & Sakakibara, 2019; Thrash et al., 2020). Un N50 grande y un L50 pequeño significará una mayor continuidad del ensamblado, ya que un genoma continuo se basa en obtener la menor cantidad posible de contigs, pero con longitudes amplias (Jayakumar & Sakakibara, 2019; Thrash et al., 2020). En este caso, los resultados de Flye son superiores a los de SMARTdenovo, ya que la mitad del genoma se encuentra representado en contigs superiores a 341,921 pb, y solamente se requieren de 2,586 contigs para cubrir el 50% del ensamblado. La razón por la que los resultados de Flye son superiores a los de SMARTdenovo, podría deberse a que funciona con un enfoque distinto de ensamblaje, específicamente al trabajar con regiones repetitivas. Flye inicia agrupando las lecturas para formar secuencias preliminares sujetas a errores, las cuales después son utilizadas para crear un “gráfico de repeticiones” (Wick & Holt, 2021). En este gráfico se identificarán las regiones repetitivas para determinar su orden y cantidad correcta y formar los contigs finales (Wick & Holt, 2021). Probablemente este genoma alberga una cantidad alta de regiones repetitivas, lo cual puede explicar porque Flye funcionó mejor.

Compleitud del genoma

Al analizar la completitud del genoma mediante BUSCO, se puede concluir que los resultados obtenidos son favorables tanto para el ensamblaje realizado con SMARTdenovo, como para el ensamblaje realizado con Flye. En total se buscaron 2,326 genes ortólogos disponibles en la base de datos Eudicots_Odb10 correspondientes al grupo taxonómico de eudicotiledóneas.

Para ambos ensamblados se logró identificar cerca del 100% de los genes ortólogos (BUSCOs), y solo un pequeño porcentaje se encontraban fragmentados o ausentes, lo que evidencia una completitud alta. Sin embargo, de los genes encontrados más de la mitad se encuentran duplicados. Estos resultados son similares a los reportados para el genoma de *S. atracyloides*, con la única diferencia que se utilizó la base de datos Virdiplantae_Odb10 (Cerca et al., 2022). De igual manera, la gran mayoría de los genes BUSCO fueron identificados como completos (93.3%), mientras que solamente el 2.8% fueron fragmentados, y el 3.9% no se encontraron (Cerca et al., 2022). Además, en el mismo estudio, más del 50% de los BUSCOs se encontraban duplicados (Cerca et al., 2022)..

Una posible explicación a la alta duplicación de genes puede deberse a la alta fragmentación del ensamblaje, y la posible tetraploidía de la especie, considerando que el ancestro en común de las especies de *Scalesia* era tetraploide (Cerca et al., 2022). Esto puede generar dificultades durante el ensamblaje, principalmente en regiones con alta heterocigosidad, donde es probable que las diferentes copias se ensamblen como regiones distintas en lugar de una sola (Guan et al., 2020), lo que complica los análisis posteriores como la identificación de genes con BUSCO. Además, debido a la alta fragmentación, es probable que un mismo gen esté repetido varias veces en más de un contig (Han et al., 2013), aumentando el número de genes duplicados identificados por BUSCO.

Finalmente, aunque se utilizó Purge Haplotig previo al análisis con BUSCO para evitar errores relacionados a la poliploidía, otros estudios han evidenciado que en comparación con herramientas similares, Purge Haplotigs tiende a conservar más genes y regiones duplicadas (Guan et al., 2020), lo cual pudo haber influido en los resultados obtenidos.

Anotación estructural

Los resultados obtenidos de la anotación estructural no fueron favorables, ya que en comparación con estudios similares en otras plantas de la familia Asteraceae, la cantidad de genes, intrones, y exones obtenidos es inusualmente alta. Por ejemplo, se ha reportado que la *S. atracyloides* tiene aproximadamente 43,000 genes, *Lactuca sativa* alrededor de 38,000 genes codificantes, y *Helianthus annus* cerca de 52,000 (Badouin et al., 2017; Cerca et al., 2022; Reyes-Chin-Wo et al., 2017). Estos valores son considerablemente menores a los reportados en este estudio, donde se identificaron 619,766 genes para *S. gordilloi*. De igual manera, la cantidad de intrones y exones reportados tanto para *Helianthus annus* como para *Lactuca sativa*, no superan los 400 mil (NCBI, s. f.-b, s. f.-a), mientras que, en este estudio, se identificaron 958,012 intrones, y 1,575,101 exones.

Como se puede apreciar, los resultados obtenidos difieren en gran medida a los reportados en estudios similares. Una de las posibles explicaciones a estas inconsistencias puede estar relacionada con el alto grado de fragmentación del genoma. La fragmentación puede provocar que un solo gen se encuentre representado en varios contigs que no se han ensamblado correctamente, lo que genera resultados imprecisos (Han et al., 2013), similar a lo que ocurre con BUSCO. En este estudio se logró secuenciar y ensamblar un primer borrador del genoma de referencia de *S. gordilloi*. Además, si bien las lecturas generadas de la secuenciación eran considerablemente grandes, no fueron lo suficientemente largas como para cubrir de manera eficiente un genoma tan extenso y con posibles regiones repetitivas. Es por esta razón que es necesario continuar trabajando en el ensamblado para conseguir una menor fragmentación y mejor calidad.

CONCLUSIONES

Se logró secuenciar y ensamblar el primer borrador del genoma de referencia de *S. gordilloi* con una completitud del 97.12% respecto a los genes ortólogos encontrados, en aproximadamente 40,000 contigs. La cantidad de información obtenida durante la secuenciación supera el 30X de cobertura, y la calidad de las lecturas es mayor a lo reportado en estudios de secuenciación con ONT, evidenciando una buena calidad y cantidad de datos.

Durante el ensamblaje, se evidenció que los resultados de Flye fueron superiores a los de SMARTdenovo. Flye cubrió cerca de 85% del genoma de referencia de *S. atractyloides* (GCA_947069175.1), tuvo una menor cantidad de contigs, el contig más largo fue de 8,697,309 pb, y los parámetros N50 y L50 superaban ampliamente a los de SMARTdenovo. Por otra parte, los resultados de BUSCO fueron similares para ambos ensamblados, ya que la mayoría de los genes fueron encontrados, y que a pesar de que más de la mitad se encontraban duplicados, muy pocos se encontraban fragmentados o perdidos.

Si bien el genoma se encuentra fragmentado en 41,617 contigs, los resultados obtenidos son favorables al haber obtenido el primer genoma de referencia de esta especie emblemática de las Islas Galápagos que está en peligro de extinción. En base a lo anterior, se puede concluir que el objetivo del proyecto fue alcanzado, pero es importante continuar con la corrección y mejora al ensamblaje para que, en el futuro, este genoma pueda ser utilizado como un instrumento para el desarrollo de estudios relacionados a la ecología, evolución, y conservación, no solo de esta especie sino de todo el género *Scalesia*.

REFERENCIAS BIBLIOGRÁFICAS

- Albuja-Quintana, M., Pozo, G., Gordillo-Romero, M., Armijos, C. E., & Torres, M. D. L. (2024). Genome report: First reference genome of *Vaccinium floribundum* Kunth, an emblematic Andean species. *G3: Genes, Genomes, Genetics*, 14(8), jkae136. <https://doi.org/10.1093/g3journal/jkae136>
- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Loman, N. J., Andersson, A. F., & Quince, C. (2013). *CONCOCT: Clustering cONTigs on COverage and ComposiTion* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.1312.4038>
- Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., Lelandais-Brière, C., Owens, G. L., Carrère, S., Mayjonade, B., Legrand, L., Gill, N., Kane, N. C., Bowers, J. E., Hubner, S., Bellec, A., Bérard, A., Bergès, H., Blanchet, N., ... Langlade, N. B. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, 546(7656), 148-152. <https://doi.org/10.1038/nature22380>
- Cerca, J., Petersen, B., Lazaro-Guevara, J. M., Rivera-Colón, A., Birkeland, S., Vizueta, J., Li, S., Li, Q., Loureiro, J., Kosawang, C., Díaz, P. J., Rivas-Torres, G., Fernández-Mazuecos, M., Vargas, P., McCauley, R. A., Petersen, G., Santos-Bay, L., Wales, N., Catchen, J. M., ... Martin, M. D. (2022). The genomic basis of the plant island syndrome in Darwin's giant daisies. *Nature Communications*, 13(1), 3729. <https://doi.org/10.1038/s41467-022-31280-w>
- Chakraborty, M., Baldwin-Brown, J. G., Long, A. D., & Emerson, J. J. (2016). Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Research*, 44(19), e147. <https://doi.org/10.1093/nar/gkw654>

- Cliften, P. (2015). Base Calling, Read Mapping, and Coverage Analysis. En *Clinical Genomics* (pp. 91-107). Elsevier. <https://doi.org/10.1016/B978-0-12-404748-8.00007-1>
- Coll, M. B. (2020, marzo 13). *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/>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin*, 19(1), 11-15.
- Ejigu, G. F., & Jung, J. (2020). Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology*, 9(9), 295. <https://doi.org/10.3390/biology9090295>
- Fernández-Mazuecos, M., Vargas, P., McCauley, R. A., Monjas, D., Otero, A., Chaves, J. A., Andino, J. E. G., & Rivas-Torres, G. (2020). The Radiation of Darwin's Giant Daisies in the Galápagos Islands. *Current Biology*, 30(24), 4989-4998.e7. <https://doi.org/10.1016/j.cub.2020.09.019>
- Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 36(9), 2896-2898. <https://doi.org/10.1093/bioinformatics/btaa025>
- Guigó, R. (2023). Genome annotation: From human genetics to biodiversity genomics. *Cell Genomics*, 3(8), 100375. <https://doi.org/10.1016/j.xgen.2023.100375>
- 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>

- Hamann, O., & Wium-Andersen, S. (1986). *Scalesia gordilloi* sp. Nov. (Asteraceae) from the Galápagos Islands, Ecuador. *Nordic Journal of Botany*, 6(1), 35-38.
<https://doi.org/10.1111/j.1756-1051.1986.tb00858.x>
- Han, M. V., Thomas, G. W. C., Lugo-Martinez, J., & Hahn, M. W. (2013). Estimating Gene Gain and Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. *Molecular Biology and Evolution*, 30(8), 1987-1997.
<https://doi.org/10.1093/molbev/mst100>
- Itow, S. (1995). Phytogeography and Ecology of *Scalesia* (Compositae) Endemic to the Galapagos Islands. *Pacific Science*, 49(1), 17-30.
- Jayakumar, V., & Sakakibara, Y. (2019). Comprehensive evaluation of non-hybrid genome assembly tools for third-generation PacBio long-read sequence data. *Briefings in Bioinformatics*, 20(3), 866-876. <https://doi.org/10.1093/bib/bbx147>
- Kaiho, K. (2022). Extinction magnitude of animals in the near future. *Scientific Reports*, 12(1), 19593. <https://doi.org/10.1038/s41598-022-23369-5>
- Kelley, D., Page, K., Quiroga, D., & Salazar, R. (2019). The Origins and Ecology of the Galapagos Islands. En D. Kelley, K. Page, D. Quiroga, & R. Salazar (Eds.), *In the Footsteps of Darwin: Geoheritage, Geotourism and Conservation in the Galapagos Islands* (pp. 67-93). Springer International Publishing. https://doi.org/10.1007/978-3-030-05915-6_3
- 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://doi.org/10.1038/s41587-019-0072-8>
- Koren, S., & Phillippy, A. M. (2015). One chromosome, one contig: Complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology*, 23, 110-120. <https://doi.org/10.1016/j.mib.2014.11.014>

- Kubis, S. (1998). Repetitive DNA Elements as a Major Component of Plant Genomes. *Annals of Botany*, 82, 45-55. <https://doi.org/10.1006/anbo.1998.0779>
- Lanteri, A. A. (1997). Islas Galápagos: Un paraíso amenazado. *Museo*, 10, 75-83.
- 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., Yi-Pan, & Wang, J. (2019). Current challenges and solutions of *de novo* assembly. *Quantitative Biology*, 7(2), 90-109. <https://doi.org/10.1007/s40484-019-0166-9>
- Linck, E., & Cadena, C. (2024). A Latitudinal Gradient of Reference Genomes. *Molecular ecology*, e17551. <https://doi.org/10.1111/mec.17551>
- Liu, H., Wu, S., Li, A., & Ruan, J. (2021). SMARTdenovo: A de novo assembler using long noisy reads. *GigaByte*, 2021, gigabyte15. <https://doi.org/10.46471/gigabyte.15>
- Mock, A., Braun, M., Scholl, C., Fröhling, S., & Erkut, C. (2023). Transcriptome profiling for precision cancer medicine using shallow nanopore cDNA sequencing. *Scientific Reports*, 13(1), 2378. <https://doi.org/10.1038/s41598-023-29550-8>
- Muñoz Barriga, A. (2015). La contradicción del turismo en la conservación y el desarrollo en Galápagos—Ecuador. *Estudios y perspectivas en turismo*, 24(2), 399-413.
- NCBI. (s. f.-a). *Helianthus annuus Annotation Release 101* [Dataset]. NCBI. https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/Helianthus_annuus/101/
- NCBI. (s. f.-b). *Lactuca sativa Annotation Release GCF_002870075.4-RS_2022_12* [Dataset]. NCBI. https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/Lactuca_sativa/GCF_002870075.4-RS_2022_12/

- Nicolson, S. W., & Human, H. (2013). Chemical composition of the 'low quality' pollen of sunflower (*Helianthus annuus*, Asteraceae). *Apidologie*, 44(2), 144-152.
- <https://doi.org/10.1007/s13592-012-0166-5>
- Oxford Nanopore Technologies. (s. f.). *ORG.one. Our Goal Is to Support Rapid Sequencing of Any Endangered Species, Anywhere, by Anyone*. Recuperado 4 de febrero de 2025, de <https://nanoporetech.com/oo>
- Pater, A. A., Bosmeny, M. S., White, A. A., Sylvain, R. J., Eddington, S. B., Parasrampuria, M., Ovington, K. N., Metz, P. E., Yinusa, A. O., Barkau, C. L., Chilamkurthy, R., Benzinger, S. W., Hebert, Madison. M., & Gagnon, K. T. (2021). High throughput nanopore sequencing of SARS-CoV-2 viral genomes from patient samples. *Journal of Biological Methods*, 8(COVID 19 Spec Iss), e155.
- <https://doi.org/10.14440/jbm.2021.360>
- Quan, Z.-J., Li, S.-A., Yang, Z.-X., Zhao, J.-J., Li, G.-H., Zhang, F., Wen, W., Cheng, T., & Zhang, X.-B. (2023). GREPore-Seq: A Robust Workflow to Detect Changes After Gene Editing Through Long-Range PCR and Nanopore Sequencing. *Genomics, Proteomics & Bioinformatics*, 21(6), 1221-1236.
- <https://doi.org/10.1016/j.gpb.2022.06.002>
- Reinert, K., Langmead, B., Weese, D., & Evers, D. J. (2015). Alignment of Next-Generation Sequencing Reads. *Annual Review of Genomics and Human Genetics*, 16(Volume 16, 2015), 133-151. <https://doi.org/10.1146/annurev-genom-090413-025358>
- Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., Xia, L., Froenicke, L., Lavelle, D. O., Truco, M.-J., Xia, R., Zhu, S., Xu, C., Xu, H., Xu, X., Cox, K., Korf, I., Meyers, B. C., & Michelmore, R. W. (2017). Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nature Communications*, 8, 14953. <https://doi.org/10.1038/ncomms14953>

- Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplots: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19(1), 460. <https://doi.org/10.1186/s12859-018-2485-7>
- Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. En M. Kollmar (Ed.), *Gene Prediction: Methods and Protocols* (pp. 227-245). Springer. https://doi.org/10.1007/978-1-4939-9173-0_14
- Thermo Fisher Scientific. (s. f.). *Qubit 4 Fluorometer*. Qubit 4 Fluorometer. Recuperado 18 de marzo de 2025, de <https://www.thermofisher.com/ht/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/models/qubit-4.html>
- Thrash, A., Hoffmann, F., & Perkins, A. (2020). Toward a more holistic method of genome assembly assessment. *BMC Bioinformatics*, 21(4), 249. <https://doi.org/10.1186/s12859-020-3382-4>
- Tydecks, L., Jeschke, J. M., Wolf, M., Singer, G., & Tockner, K. (2018). Spatial and topical imbalances in biodiversity research. *PLOS ONE*, 13(7), e0199327. <https://doi.org/10.1371/journal.pone.0199327>
- Tye, A., & Nykanen, M. (2015). IUCN Red List of Threatened Species: Scalesia gordilloi. *IUCN Red List of Threatened Species*. <https://www.iucnredlist.org/en>
- 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://doi.org/10.1101/gr.214270.116>
- Weirather, J. L., De Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., & Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*, 6, 100. <https://doi.org/10.12688/f1000research.10571.2>

Wick, R. R., & Holt, K. E. (2021). Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research*, 8, 2138.

<https://doi.org/10.12688/f1000research.21782.4>

Worley, K. C., Richards, S., & Rogers, J. (2017). The value of new genome references. *Experimental Cell Research*, 358(2), 433-438.

<https://doi.org/10.1016/j.yexcr.2016.12.014>

Zymo Research. (s. f.). *DNA Sequencing Clean-up Kits*. DNA Sequencing Clean-up Kits. Recuperado 18 de marzo de 2025, de <https://zymoresearch.eu/collections/dna-sequencing-clean-up-kits>

ANEXO A: CONCENTRACIÓN DE ADN DE LA EXTRACCIÓN Y DE LA LIBRERÍA FINAL

Muestra	Extracción	Librería final
<i>SgJ02</i>	302 ng/uL en 50 uL	680 ng totales en 25 uL