

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

**Ensamblaje *de novo* y anotación del genoma mitocondrial del mortiño
(*Vaccinium floribundum* Kunth.)**

Paula Camila González Corrales

Ingeniería en Biotecnología

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

Quito, 16 de mayo de 2022

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

**Ensamblaje *de novo* y anotación del genoma mitocondrial del mortiño
(*Vaccinium floribundum* Kunth.)**

Paula Camila González Corrales

Nombre del profesor, Título académico

María de Lourdes Torres, PhD

Quito, 16 de mayo de 2022

© DERECHOS DE AUTOR

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

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

Nombres y apellidos: Paula Camila González Corrales

Código: 00200140

Cédula de identidad: 1718395542

Lugar y fecha: Quito, 16 de mayo de 2022

ACLARACIÓN PARA PUBLICACIÓN

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

UNPUBLISHED DOCUMENT

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

RESUMEN

Vaccinium floribundum o mejor conocido como mortiño, es un arbusto nativo de la región andina, conocido por su importancia económica, cultural y ecológica dentro del Ecuador. Sin embargo, hasta el día de hoy no existe estudios genómicos de esta especie. En esta investigación se buscó ensamblar y anotar el genoma de la mitocondria de *V. floribundum* a partir de lecturas largas. El genoma de la mitocondria es una fuente de información sobre evolución y adaptación ambiental. Para la búsqueda del mejor protocolo de ensamblaje se utilizó programas bioinformáticos donde se evaluó las diferentes combinaciones posibles con el fin de obtener un tamaño aproximado de 459 kb nucleótidos perteneciente al género *Vaccinium*. Se determinó que SMARTdenovo y Hinge en combinación con Nanopolish produjeron las secuencias óptimas para la anotación y búsqueda de genes, gracias a su mayor cercanía al genoma de referencia de la mitocondria de *V. macrocarpon* (KF314261.1). A partir del ensamblaje de lecturas largas, se encontró que es posible el ensamblado de genomas continuos y de buena calidad. Sin embargo, es necesario incorporar más información, mayor número de secuencias largas, implementación de secuencias cortas, nuevas referencias, con el fin de obtener resultados de mejor calidad. Este es el primer estudio donde se reporta el genoma de la mitocondria del mortiño, especie silvestre nativa de los andes ecuatorianos. Información clave para posteriores estudios genómicos sobre evolución y adaptabilidad de especies de plantas al páramo.

Palabras clave: mortiño, *Vaccinium floribundum*, genoma de la mitocondria, secuenciación, lecturas largas, ensamblaje, anotación,

ABSTRACT

Vaccinium floribundum or better known as mortiño, is a native shrub of the Andean region, known for its economic, cultural, and ecological importance within Ecuador. However, to date, there are no genomic studies of this species based on the blueberry. In this research, the assembly and genome annotation of the mitochondria of *V. floribundum* was tried using long reads. The mitochondrial genome is a source of information on evolution and environmental adaptation. To search for the best assembly protocol, bioinformatic programs were used where the different possible combinations were evaluated in order to obtain an approximate size of 459 kb nucleotides belonging to the closest size to the *Vaccinium* genus. It was determined that SMARTdenovo and Hinge in combination with Nanopolish produced the optimal sequences for gene annotation and searching, thanks to their closer proximity to the reference genome of the reference *V. macrocarpon* mitochondria (KF314261.1). From the assembly of long reads, it was found that the structuring of continuous and good quality genomes is possible. However, it is necessary to incorporate more information, a greater number of long sequences, implementation of short sequences, and new references, in order to obtain better quality results. Finally, to carry out the first study where the genome of the mitochondria of the mortiño, a wild species native to the Ecuadorian Andes, is reported. Key information for subsequent genomic studies of evolution and adaptability of the plant to the páramo zone. It represents promoting, an understanding of the functioning and adaptation of the species to the Andean moors. As well as granting a genetic resource for further studies within this variety.

Keywords: Andean blueberry, *Vaccinium floribundum*, mitochondrial genome, sequencing, long reads, assembly, annotation

Tabla de contenido

1	INTRODUCCIÓN.....	11
1.1	Generalidades de <i>Vaccinium Floribundum</i> Kunth.....	11
1.1.1.	Características, distribución, propiedades destacadas	11
1.1.2.	Importancia económica, cultural, y ecológica.....	11
1.1.3.	Estudios moleculares preliminares de <i>V. floribundum</i> en el Ecuador	12
1.2	Secuenciación con Oxford Nanopore Technologies	12
1.3	Mitocondria	13
1.4	Secuenciación del genoma mitocondrial de <i>V. floribundum</i>	14
2	METODOLOGÍA.....	15
2.1	Recolección de muestras de <i>V. floribundum</i>	15
2.2	Extracción de ADN y purificación.....	15
2.3	Secuenciación en MinION del genoma de <i>V. floribundum</i>	15
2.4	Análisis bioinformático.....	16
2.4.1	Pre-procesamiento de lecturas de la secuenciación de <i>V. floribundum</i> ...	16
2.4.2	Extracción de lecturas de la mitocondria de <i>V. floribundum</i>	16
2.4.3	Ensamblaje del genoma de la mitocondria de <i>V. floribundum</i>	16
2.4.4	Pulido del genoma de la mitocondria de <i>V. floribundum</i>	16
2.4.5	Evaluación de los genomas de la mitocondria de <i>V. floribundum</i>	17
2.4.6	Anotación del genoma de la mitocondria de <i>V. floribundum</i>	17
3	RESULTADOS	18
3.1	Secuenciación en MinION del genoma de <i>V. floribundum</i>	18
3.2	Evaluación de los ensamblados del genoma de la mitocondria de <i>V. floribundum</i>	18
3.2.1	Ensamblaje del genoma de la mitocondria de <i>V. floribundum</i>	18
3.2.2	Pulido del genoma de la mitocondria de <i>V. floribundum</i>	19
3.2.3	Llamado de variantes de los ensamblajes de la mitocondria de <i>V. floribundum</i>	21
3.2.4	Anotación del genoma final de la mitocondria de <i>V. floribundum</i>	21
4	DISCUSIÓN.....	22
4.1	Secuenciación en MinION de <i>V. floribundum</i>	22

4.2	Evaluación y comparación de los ensamblajes de la mitocondria de <i>V. floribundum</i>	22
4.2.1	Ensamblaje del genoma de la mitocondria de <i>V. floribundum</i>	22
4.2.2	Pulido del genoma de la mitocondria de <i>V. floribundum</i>	23
4.2.3	Llamado de variantes	24
4.3	Anotación	24
5.	CONCLUSIONES	26
6.	TABLAS	27
7.	FIGURAS	31
8.	REFERENCIAS	33
9.	ANEXOS	38

ÍNDICE DE TABLAS

Tabla 1. Estadísticas sobresalientes del desempeño de la secuenciación de <i>Vaccinium floribundum</i>	27
Tabla 2. Parámetros considerados para la evaluación del genoma de la mitocondria de <i>Vaccinium floribundum</i> en los distintos ensambladores.....	28
Tabla 3. Parámetros considerados para la evaluación del genoma de la mitocondria de <i>Vaccinium floribundum</i> , en el ensamblado y pulido.	29
Tabla 4. Parámetros evaluados con respecto a la referencia (mitocondria <i>V. macrocarpon</i> , KF386162.1) en el llamado de variantes para anotación de genomas de la mitocondria del mortiño.....	30

ÍNDICE DE FIGURAS

Figura 1. Protocolo de análisis bioinformático para el ensamblaje del genoma de la mitocondria de <i>Vaccinium floribundum</i>	31
Figura 2. Estructura de los genes encontrados en el genoma de la mitocondria de <i>Vaccinium floribundum</i> a. Nanopolish+SMARTdenovo b. Nanopolish+HIInge.	32

ÍNDICE DE ANEXOS

ANEXO 1. Características de los ensamblajes más destacados de <i>V. floribundum</i> con respecto a la referencia.	38
---	----

1 INTRODUCCIÓN

1.1 Generalidades de *Vaccinium Floribundum* Kunth

1.1.1. Características, distribución, propiedades destacadas

Vaccinium floribundum Kunth, conocido normalmente como mortiño o uva de monte pertenece a la familia Ericaceae. Es un arbusto silvestre nativo de los páramos andinos. Se distribuye en el noroccidente de Suramérica dentro de Venezuela, Colombia, Ecuador y Perú. Su distribución recorre de los 2500 y 4300 msnm en zonas con climas fríos, templados de 7-18 °C con alta radiación solar. Su crecimiento llega de los 2.0 – 2.5 m de altura, con hojas pequeñas, flores de menos de 1 cm las cuales pueden crecer en racimos o solitarias, con flor de color blanca o rosada, su fruto es una baya esférica lisa de 5 a 8 mm de color morado oscuro, azul y (Coba et al., 2012). *V. floribundum* se caracteriza por sus propiedades, en hojas y frutos que poseen un potencial antimicrobiano. Sus bayas se caracterizan por ser antioxidantes por la alta presencia de polifenoles y antocianinas beneficiosos para prevenir oxidación molecular, enfermedades neurológicas y cardíacas (Chiluisa, 2015, Coba et al., 2012).

1.1.2. Importancia económica, cultural, y ecológica

Dentro de la sociedad y economía ecuatoriana juega un papel importante, debido a que sus frutos son utilizados para la preparación de la colada morada (bebida tradicional), vinos, postes, además de ser una fuente económica de comunidades andinas indígenas (Coba et al., 2012). Su recolección es silvestre debido a su dificultad de propagación por la baja tasa de germinación de sus semillas y la falta de investigación de protocolos de propagación con insumos agropecuarios (Coba et al., 2012, Chamorro, 2019)

V. floribundum, tiene un rol ecológico importante gracias a que brinda protección y alimento a la fauna de los páramos, después de incendio y deforestaciones se caracteriza por su recuperación rápida de tallos y hojas (Chiluisa, 2015). El mortiño se encuentra amenazado por deforestaciones, incendios, el cambio de la zona de páramo por zonas dedicadas al

pastoreo, malas prácticas de recolección de frutos dando como resultado una eliminación en la estacionalidad de la planta (Chamorro, 2020).

1.1.3. Estudios moleculares preliminares de *V. floribundum* en el Ecuador

Se ha buscado entender de mejor manera el comportamiento de esta especie mediante la determinación de la diversidad genética y estructura poblacional de *V. floridundum* en los páramos del Ecuador (Cobo et al., 2016, Vega-Polo et al., 2020,). En la investigación de Vega-Polo et al, 2020, se determinó la estructura poblacional del mortiño dentro del Ecuador, donde esta especie presento un alto grado de diversidad genética, y una estructura poblacional en base a la interacción de dinámicas de aislamiento por distancia y altitud geográfica, distintivo de la zona de páramo.(Vega-Polo et al., 2020).

1.2 Secuenciación con Oxford Nanopore Technologies

El primer secuenciador por nanoporo, fue el MinION que es una tecnología de Oxford Nanopore Technologies (ONT), la cual se basa en la lectura de moléculas de ADN mediante su paso por un poro de proteína a nano escala que se encuentra incrustado dentro de una membrana resistente a la electricidad. MinION sirve como un biosensor capaz de soportar una corriente de voltaje constante (Wang et al., 2021, Jain et al, 2016). La lectura de las moléculas de ADN se basa en los cambios de corriente iónica, lo cual representa una secuencia de nucleótidos que se detecta y codifica a algoritmos bioinformáticos (Wang et al., 2021). Los datos recopilados son procesados por MinKNOW, un software operativo de ONT el cual establece parámetros y da seguimiento a la muestra. Es capaz de realizar un llamado de bases el cual a partir información cruda traduce cargas eléctricas a una base nucleotídica, base para el ensamblado y anotado posteriormente (Wang et al., 2021).

MinION es el primer secuenciador de lecturas largas, 10000-100000 pares de bases, permite la producción de datos en una menor cantidad de tiempo, con presupuestos más bajos sin la necesidad de una capacidad computacional alta gama (Jain et al, 2016, Wang et al., 2021). Estas características lo diferencian de otros secuenciadores siendo un dispositivo útil para investigación debido a su portabilidad sin límites. El uso de MinION desencadenó una ola de secuenciamiento de diferentes genomas de especies con interés potencial (Jain et al., 2016). Teniendo como base la generación de una secuencia consenso la cual posee toda la información genética de un individuo u organelo en la cual se puede identificar regiones de interés. El proceso para la estructuración, ensamblaje, de cualquier genoma se da mediante los siguientes pasos: a. La generación de *contigs* mediante el solapamiento de lecturas. b. La unión de *contigs* mediante solapamientos que da un andamio de estructura *scaffolds*. c. Relleno de zonas sin coincidencias *gaps* (Wang et al., 2021). El anotado permite que, a partir de una estructura ordenada o secuencia consenso, se identifique regiones de interés dentro del material genético que tiene una función dentro del individuo (Jung et al., 2020). Con la información obtenida se realiza una búsqueda en bases de datos como el *NCBI*, *Gene Ontology Resource*, o *Kyoto*, donde se obtiene accesiones de las regiones de interés que se puede relacionar a procesos metabólicos (Lathe et al., 2008).

1.3 Mitocondria

La mitocondria es un organelo que actúa como central energética de la célula, produce la energía necesaria para activar rutas metabólicas (Aanen et al., 2014). Esta energía se obtiene mediante la oxidación de metabolitos sintetizando trifosfato de adenosina (ATP) (Aanen et al., 2014). La mitocondria contiene material genético propio el cual, posee información de la evolución de cada especie y su adaptabilidad ambiental (Aanen et al.,

2014). Dentro del género *Vaccinium*, la mitocondria tiene un tamaño alrededor de los 400 kb, con un total de 33 genes identificados codificantes de proteínas, ARNr, ARNt útiles en la oxidación de molecular y adaptación evolutiva (Fajardo et al., 2014).

1.4 Secuenciación del genoma mitocondrial de *V. floribundum*

A partir de la secuenciación, ensamblaje y anotación de genomas mitocondriales se ha logrado conocer información de la evolución, adaptabilidad ambiental, análisis filogenético para determinar la similitud funcional entre varios niveles taxonómicos (Palmer et al., 2000, Xu et al., 2021). Se conoce un número reducido de genomas mitocondriales. Dentro del género *Vaccinium*, tan solo existe una referencia, *Vaccinium macrocarpon* (KF386162.1) perteneciente a la mitocondria (Xu et al., 2021).

Este proyecto tiene como objetivo buscar y evaluar el mejor protocolo para realizar el ensamblaje y, anotación del genoma de la mitocondria de *Vaccinium floribundum*.

2 METODOLOGÍA

2.1 Recolección de muestras de *V. floribundum*

En la localidad de Lloa, Pichincha, Ecuador (S0° 11.55122' W78° 35.216') bajo el contrato marco MAEDNB-CM-2016-0046-M-0002, se recolectaron hojas jóvenes de un arbusto de mortiño (muestra P079). Las hojas fueron almacenadas en sílica gel y posteriormente transportadas al Laboratorio de Biotecnología Vegetal de la USFQ. Su almacenamiento inicial se lo realizó a -20 °C y luego de aproximadamente un día se conservó el material vegetal a -80 °C hasta la extracción de ADN.

2.2 Extracción de ADN y purificación

La extracción de ADN se realizó de hojas jóvenes mediante el protocolo *High-Molecular-Weight DNA Using Organic Solvents* de Green y Sambrook (2021). Después, se realizó una eliminación de fragmentos (< 10 Kb) mediante el kit *Short Read Eliminator* (Jones et al., 2020). Para posteriormente cuantificar el ADN obtenido mediante NanoDrop y Qubit con el fin de ver su calidad para la secuenciación.

2.3 Secuenciación en MinION del genoma de *V. floribundum*

Con el ADN extraído se realizó la preparación de librerías mediante el kit de ligación SQK-LSK109 de ONT (Reiling et al., 2020). Al protocolo propuesto se le realizaron algunas modificaciones como: el uso de 2 µg en vez de 1 µg de ADN, dentro del paso de ligación de adaptadores se incubó el buffer de ligación, adaptadores y la enzima ligasa por 20 minutos en vez de 10. Para la carga de la librería se utilizaron dos celdas de flujo en el equipo MinION Mk1B. Se utilizó MinKNOW v.7 y Guppy v1.1 alpha13-0glec7786 para el llamado de bases y LongQC para el control de calidad de las lecturas (Wick et al., 2019).

2.4 Análisis bioinformático

Se siguió el protocolo que se observa en la Figura 1.

2.4.1 Pre-procesamiento de lecturas de la secuenciación de *V. floribundum*

En el pre-procesamiento de lecturas se usó Porechop v0.2.4 para la remoción de adaptadores. Posteriormente se utilizó Nanofilt v2.8.0 para eliminar lecturas con una calidad menor a $Phred=9$ y se filtró las de tamaño menor a 1kpb (Wang et al., 2018)

2.4.2 Extracción de lecturas de la mitocondria de *V. floribundum*

Al realizar la secuenciación se abarca todo el genoma de *V. floribundum*, nuclear, del cloroplasto y de la mitocondria. Se filtró el material genético de la mitocondria mediante el software Brasr v5.3.5 (Wang et al., 2018), en base a la referencia de *V. macrocarpon* (código de acceso: KF386162.1). Para el análisis de las lecturas se realizó un reporte de calidad mediante LongQC (Fukasawa et al., 2020).

2.4.3 Ensamblaje del genoma de la mitocondria de *V. floribundum*

Se emplearon cuatro ensambladores de lecturas largas para el ensamblaje *de novo* de las secuencias mitocondriales filtradas: Hinge v.0.41 (Kamath et al, 2017), Canu v.1.4. (Koren et al, 2017), Flye v. 2.7-0 (Kolmogorov et al., 2019), y SMARTdenovo v.1.0.0 (Liu et al., 2021). En el ensamblador Flye, se aumentó el parámetro `--asm-coverage=50` para minimizar el uso de memoria (Kolmogorov et al., 2019). Se especificó para todos los softwares un tamaño esperado del género *Vaccinium* de 459kb con base a la referencia (Fajardo et al., 2014) y se complementó el resto de los parámetros con opciones default.

2.4.4 Pulido del genoma de la mitocondria de *V. floribundum*

El pulido de un genoma se busca minimizar la cantidad de errores que no se pudieron identificar en pasos anteriores para obtener una secuencia consenso de mejor calidad

(Morisse et al., 2021). Se realizó un reordenamiento de los *contigs* generados por cada ensamblador en base a la referencia de *V. macrocarpon* (KF386162.1) mediante el script de Wang et al. (2018). En base al reordenamiento de los *contigs*, se obtuvo una sola secuencia consenso por ensamblador, las cuales se pulieron con Racon v1.4.20, Nanopolish v0.13.2, y Racon v1.4.20 + Nanopolish v0.13.2 (Wang et al., 2018, Senol Cali et al., 2019).

2.4.5 Evaluación de los genomas de la mitocondria de *V. floribundum*

La evaluación de las combinaciones entre ensambladores y pulidores se realizó mediante un llamado de variantes/SNPs, donde se analizó las mejores combinaciones en base a secuencias consensos, determinando la cantidad de nucleótidos diferentes entre la secuencia generada con la referencia. Se realizó un llamado de variantes mediante Nanopolish v0.13.2 con el comando de *nanopolish variants* (Shikov., 2021). Estas secuencias fueron evaluadas posteriormente por métricas en base a tamaño del genoma, alineamiento más largo, menor cantidad de indels, y ratio de duplicación en QUILT (Gurevich et al, 2013). En Geneious Prime 2022.1.1 mediante la herramienta “Pairwise Alignment”, se alineó las secuencias de interés con la referencia con el fin de analizar el % de identidad, # de semejanzas, # de diferencias para determinar así la mejor secuencia consenso para su anotación posterior (Biomatters Ltd, 2022).

2.4.6 Anotación del genoma de la mitocondria de *V. floribundum*

Se anotaron las combinaciones de ensamblado y pulido en base a la secuencia consenso más larga, alineamiento más largo, menor cantidad de indels, ratio de duplicación, menor % de similitud, mayor número de diferencias y menor número de similitudes con respecto a la referencia (mitocondria, *V. macrocarpon*, KF386162.1) en base a los programas QUILT y Geneious Prime 2022.1.1 (Gurevich et al., 2013, Biomatters Ltd., 2022, Wang et al., 2018).

3 RESULTADOS

3.1 Secuenciación en MinION del genoma de *V. floribundum*

En la Tabla 1 se resume las estadísticas más importantes en el ensayo de secuenciación. Donde se generó 20GB de información con 2 964 735 lecturas, de las cuales 20729 lecturas son pertenecientes a la mitocondria. Generando así una cobertura promedio de lectura de 195.171 X. En el tamaño de las lecturas se obtuvo que la mayor fue de 115 091 pb, siendo un valor que concuerda con lo esperado por la secuenciación en ONT. También se evaluó el parámetro N50 o mediana de la cantidad de nucleótidos encontrados que fue de 19 864.00 pb lo cual indica que las lecturas generadas en la secuenciación se parecen al menos un 50% a la referencia (mitocondria, *V. macrocarpon*, KF386162.1). (Parra, 2021).

3.2 Evaluación de los ensamblados del genoma de la mitocondria de *V. floribundum*

3.2.1 Ensamblaje del genoma de la mitocondria de *V. floribundum*

En la Tabla 2 se observa las métricas obtenidas en base a los cuatro ensambladores. Estos valores fueron calculados por QUAST en base al genoma de referencia. Comparando todos los genomas obtenidos con cada ensamblador se determinó que el genoma más largo fue el obtenido utilizando SMARTdenovo (318 595 pb), el mismo que presentó el alineamiento más largo de (44 989 pb); el segundo más largo fue Canu (317 057 pb) y después Hinge (316 250 pb) con un alineamiento de 44 931 pb y 44 672 pb, respectivamente. Seguido por Hinge (316 250 pb) y Flye (103 067 pb). Se analizó también la cantidad de NGA50, parámetro da información sobre la cantidad de puntos de interrupción dentro de los ensamblajes con respecto a la referencia lo cual permite determinar los puntos que representan verdaderas diferencias estructurales entre la secuencia ensamblada y pulida con la referencia (Gurevich et al., 2013). El parámetro NGA50 de menor valor fue Flye con la

ausencia del mismo , seguido por Canu (9082), Hinge (9903) y SMARTdenovo (10 823). Con respecto al ratio de duplicación se encontró que Flye presenta un valor de 1.001 y los otros ensambladores entre 0.993-0.998. Wang et al, 2018 reportan que los valores mayores a 1.00 representan eventos de duplicaciones. Por lo tanto, el ensamblaje con Flye presentó regiones duplicadas que no fueron identificadas en la referencia. Finalmente, se obtuvo la presencia de indels, inserciones o deleciones, por cada 100 000 pb. La secuencia ensamblada por Hinge presenta una mayor cantidad de zonas con nucleótidos eliminados o añadidos (829.49), seguido por Canu (437.92), SMARTde novo (306.81) y Flye (257.94).

3.2.2 Pulido del genoma de la mitocondria de *V. floribundum*

En la Tabla 3 se presentan las métricas calculadas del genoma de la mitocondria del mortino en base a las diferentes combinaciones de ensambladores y pulidores.

Racon fue uno de los programas usados para pulir las versiones ensambladas. Se comparó el resultado con todas las secuencias sin pulir y se obtuvo una disminución en el tamaño de las secuencias consenso alrededor de 16-19kb. Por ejemplo, el tamaño de la mitocondria obtenido por SMARTdenovo paso de 318 595 pb a 301 975 pb. Por lo tanto, el acortamiento del tamaño del genoma provoca un menor número de nucleótidos al momento de alinear con respecto a la referencia y se posee una menor cantidad de información. Tanto antes como después del pulido los ensambladores Hinge, Canu y Flye poseen un tamaño menor a SMARTdenovo. También se disminuye el valor de NGA50 en todos los ensambladores, existiendo un mayor número de puntos de interrupción, obteniendo fragmentos de un menor tamaño que no abarcan el tamaño de genoma esperado. Del mismo modo, se observó que el ratio de duplicación pasó de 0.993-1.001 a 0.997 en promedio, acercándose a 1.00 valor reportado como óptimo. Por otro lado, el número de indels por 100

000 pb disminuyó en todos los ensambladores de 829.49-257.94 pb a 323.08-293.33 pb. Lo que indica una mayor cercanía al genoma de referencia.

Como segunda opción de pulidor se usó Nanopolish, con el fin de mejorar la calidad de la secuencia consenso. Se obtuvo que Canu (585 pb) , SMARTdenovo (813 pb) , Flye (5 pb) disminuyeron su tamaño de longitud comparándolo con la secuencia obtenida en el ensamblado. En el genoma ensamblado por Hinge se observó que la secuencia aumentó 725 pb en el tamaño del genoma comparándolo con el no pulido. De la misma manera, al evaluar el parámetro NGA50, se obtuvo que en Canu (9082 a 9067) y SMARTdenovo (10 823 a 10 794), disminuyó su valor de puntos de interrupción dentro de la secuencia pulida. En Hinge existió un aumento (9903 a 9950) y Flye no presentó un cambio en este parámetro. También, disminuyó el tamaño del alineamiento más largo aproximadamente 100 kb en todos los ensambladores. Al analizar el parámetro ratio de duplicación se determinó que este aumentó de 0.993-1.001 a 0.996 como promedio en todos los ensambladores pulidos. El número de indels disminuyó con el ensamblador Hinge (829.49 a 598.85) y aumentó su número en Canu (437.92 a 522.99), SMARTdenovo (306.81 a 447.95) y Flye (257.94 a 260.98).

Se uso también Racon + Nanopolish, donde disminuyó el tamaño de todos los genomas ensamblados, siendo SMARTdenovo (301 837) el más largo, seguido por Hinge (300 069 pb), Canu (298 957 pb) y Flye (84 439). Donde se ve una reducción en el tamaño de alineamiento más largo ya que disminuyó aproximadamente de 2-4 kb por ensamblador. Se puede determinar que, tanto el parámetro NGA50 como el ratio de duplicación se mantienen con los mismos valores que en el pulido individual de Racon debido al uso de la misma secuencia como archivo de entrada para el pulido combinado. Las inserciones y deleciones por 100 000 pb disminuyeron en este pulido, siendo SMARTdenovo (283.18 pb) el de menor valor seguido por Canu (307.39 pb), Flye (315.2 pb) y Hinge (320.37 pb).

3.2.3 Llamado de variantes de los ensamblajes de la mitocondria de *V. floribundum*

En la Tabla 4 se observa el porcentaje de identidad, el número de pares de bases semejantes y diferentes de las distintas combinaciones con mejores parámetros en el ensamblado. Los valores obtenidos provienen del llamado de variantes realizado por Nanopolish v0.13.2 y visualizado en Geneious Prime 2022.1 con respecto a la referencia (mitocondria, *V. macrocarpon*, KF386162.1). Donde se destaca que la combinación con mejores parámetros es la obtenida con Nanopolish + SMARTdenovo (312 896 pb), seguido de Nanopolish+Hinge (312 751 pb). Teniendo un porcentaje de identidad del 66.803% y 66.588% respectivamente. Con respecto al número de semejanzas Nanopolish + SMARTdenovo posee 307.726 pb y Nanopolish+Hinge 307.043 pb con un número de diferencias de 152.920 pb y 154.063 pb respectivamente.

3.2.4 Anotación del genoma final de la mitocondria de *V. floribundum*

En la Figura 2 se muestra los resultados obtenidos de la anotación de los genomas de Nanopolish + SMARTdenovo y Nanopolish+Hinge, con el programa Geneious Prime 2022.1. Para la anotación de las combinaciones se utilizó un parámetro de similitud del 70% en Geneious Prime 2022.1, dando como resultado en ambos genomas 35 secuencias codificantes de proteínas, 37 genes, 3 genes codificantes de ARNr y 17 genes codificantes a ARNt. Analizando estos datos y comprándolos con la referencia (mitocondria, *V. macrocarpon*, KF386162.1), se encontró un igual número de genes y ARNr, un mayor número de secuencias codificantes de proteínas y un menor número de genes codificantes de ARNt. *V. macrocarpon* posee 33 secuencias codificantes de proteínas, 35 genes, 3 ARNr, 18 ARNt (ANEXO 1.) (Fajardo et al., 2014).

4 DISCUSIÓN

4.1 Secuenciación en MinION de *V. floribundum*

A partir de la secuenciación con MinION se obtuvo alrededor del 0.69% de lecturas pertenecientes a la mitocondria de *V. floribundum*. En base a este porcentaje se pudo alcanzar una cobertura de 195. 171 X con respecto al genoma de referencia (mitocondria, *V. macrocarpon*, KF386162.1) que posee un tamaño de 459 Kb (Fajardo et al., 2014). A pesar de tener una buena cobertura, ésta no es suficiente para lograr reconstruir todo el genoma debido a su extenso tamaño. Comparando la cobertura necesaria (1000 X) requerida para ensamblar genomas mitocondriales completos, la data recolectada cubre lo necesario para obtener un primer borrador del genoma mitocondrial de *V. floribundum*. (Lee et al., 2018). Por lo tanto, se puede reportar que el uso de MiniION como tecnología de secuenciación es una opción viable para la obtención de data. Esto basado en que ONT reporta que su dispositivo puede generar lecturas desde 500 pb hasta 2.3 Mb por lo que se espera que parámetros como el N50 y el tamaño de la lectura más larga sobresalgan (Jain et al., 2016).

4.2 Evaluación y comparación de los ensamblajes de la mitocondria de *V. floribundum*

4.2.1 Ensamblaje del genoma de la mitocondria de *V. floribundum*

Para el ensamblaje del genoma de la mitocondria del mortiño se partió desde un ensamblaje *de novo*. El cual se basa en la estructuración de *contigs* mediante la alineación de nucleótidos y solapamiento (Baker, 2012). Ensambladores como Hinge, Canu y SMARTdenovo usan un el algoritmo de *Over Lap Consensus (OLC)* contrario a Flye que se basa en la Gráfica de Bruijin. El algoritmo de OLC se basa en el solapamiento de lecturas para la construcción de gráficos en base al solapamiento con el fin de unir todos estos

fragmentos en una sola secuencia consenso (Dida, F., & Yi, G., 2021). Por otro lado, la Gráfica de Bruijin, genera *contigs* en base a sobrelapamientos aleatorios para después construir gráficos donde se combinan los *contigs* en los diferentes ensamblajes posibles con el fin de generar una secuencia consenso final (Kolmogorov et al., 2019).

A pesar que el ensamblador Canu es el procesador predilecto por el secuenciador para el aprovechamiento máximo de lecturas en base a la comparación de Wang et al. 2018, entre Canu y Hinge, Canu es el ensamblador con mejor capacidad para estructuración de lecturas. Sin embargo, Liu et al. 2021, sustenta que SMARTdenovo representa en comparación a Canu y Flye, un mejor análisis de ensamblaje dentro de los algoritmos en base al ensamblaje del genoma del tomate silvestre. Dentro del análisis de comparación de ensamblajes realizado en este proyecto se destacó que SMARTdenovo es el algoritmo que menos información pierde al momento de ensamblar, produciendo una secuencia consenso de mayor longitud, menor número de indels, ratio de duplicación lejano a 1 con relación a los otros ensambladores.

4.2.2 Pulido del genoma de la mitocondria de *V. floribundum*

En base al ensamblado del genoma, sin incluir el pulido, se conocía que las secuencias consenso obtenidas tenían una identidad máxima de sobrelapamiento del 68% con respecto a la referencia (mitocondria, *V. macrocarpon*, KF386162.1). Tomando en cuenta que la tasa de error de ONT va alrededor de 10-15%, es necesario realizar el paso de pulido para la corrección de errores con el fin de tener una secuencia consenso de mejor calidad. Dentro de este paso, el programa Nanopolish es el pulidor que dio como resultado ensamblajes más precisos abarcando más información. Comparando este resultado con el estudio de Wang et al, 2018, se puede confirmar que este pulidor abarca más información debido a que su tasa

de error es de 0.0022 por base de lectura, lo cual valida una mejor calidad de secuencia consenso.

Después de una evaluación en base a las diferentes combinaciones de ensamblado y pulido, se llegó a la conclusión que con las combinaciones SMARTdenovo y Hinge más el pulidor Nanopolish se obtiene un mayor número tamaño de secuencias por lo que al compararlo con la referencia (mitocondria, *V. macrocarpon*, KF386162.1) abarca un mayor tamaño del genoma, que las otras combinaciones probadas.

4.2.3 Llamado de variantes

En base a los genomas destacados después del pulido se realizó el llamado de variantes que tiene como objeto la búsqueda de polimorfismo, inserciones o deleciones dentro el genoma en base a la referencia, en Nanopolish (Simpson Lab, 2019). Para su posterior visualización en Geneious Prime 2022.1. Esto se determina a partir de un alineamiento de las secuencias sobresalientes con respecto a la referencia (mitocondria, *V. macrocarpon*, KF386162.1) donde estas abarcaron tan solo un 68% de 459 kb de *V. macrocarpon*. Tomando en cuenta que se utilizó *Next Generation Sequencing* (NGS) para la obtención de la información, se reconoce que las desviaciones, indels y SNPs, con respecto a la referencia (mitocondria, *V. macrocarpon*, KF386162.1), pueden estar dadas o un porcentaje de estas están atadas a la tasa de error aleatorio, generada por el secuenciador al no tener una tasa de error uniforme en todos los sitios, ya que hay zonas donde el secuenciador es más propenso a errores (Muzzey, et al..2015).

4.3 Anotación

A partir de la estructura y mapeo del genoma de la mitocondria del mortino *V. floribundum* se puede destacar que el uso de secuencias largas provenientes de ONT, es una

manera factible, viable de ensamblar y anotar un genoma, a pesar de que sigue siendo un desafío en cubrir zonas menores a los 10 000 pb (Jain et al, 2016, Wang et al., 2021). Wang et al, 2018, destaca que probablemente el ensamblaje de lectura largas con la secuenciación PacBio sea más preciso por nucleótido. Sin embargo, es importante resaltar que la cantidad de información obtenida está relacionada directamente con el enfoque de ensamblaje y pulido que se pone a prueba. Comparando la longitud de tamaño obtenido con *V. macrocarpon* (459 678 pb) se logró abarcar más del 50% del genoma de la mitocondria tan solo con lecturas largas y una identificación de entre 37-67 genes codificantes de este genoma (Mower et al., 2012). Lo cual permite destacar que se logró obtener el ensamblaje *de novo* y anotación del genoma mitocondrial de *V. floribundum*.

5. CONCLUSIONES

El genoma de la mitocondria de *V. floribundum* fue secuenciado y ensamblado de manera exitosa utilizando lecturas provenientes de ONT. Demostrando que existe la posibilidad de la obtención de genomas continuos de una calidad media mediante el uso de lecturas largas. La tecnología representó también un reto al tener que minimizar lo más posible la tasa de error que se tiene con ONT, mejorando la calidad de las secuencias mediante el paso de pulido. Cuando se comparó diferentes combinaciones de ensamblajes y pulidos se destacó que el ensamblado con SMARTdenovo tanto antes como después del pulido representó la menor pérdida de información, seguido del ensamblador Hinge en base a la comparación de todos los ensambladores utilizados. El mejor pulidor fue Nanopolish ya que con el mismo se obtuvo las secuencias con una mayor longitud de secuencias consenso y mayor precisión de cercanía con la referencia (mitocondria, *V. macrocarpon*, KF386162.1). Sin embargo, para un mejor ensamblaje es necesario la generación de más lecturas largas o realizar una secuenciación de lecturas cortas para completar los espacios y el tamaño de secuencia esperada, para obtener así una secuencia más confiable. A partir de esta investigación se reporta el primer ensamblaje y anotación del genoma de la mitocondria del mortiño, especie silvestre nativa de los Andes ecuatorianos. Información clave para posteriores estudios genómicos de evolución y adaptabilidad de especies de plantas que habitan los páramos ecuatorianos.

6. TABLAS

Tabla 1. Estadísticas del desempeño de la secuenciación de *Vaccinium floribundum*

Estadísticas	Valor
Información obtenida (GB)	20
Número de lecturas totales (# lecturas)	2 964 735
Número de lecturas de la mitocondria (# lecturas)	20 729
Cobertura promedio por lectura (X)	195. 171
Lectura más larga (pb)	115 091
N50 (pb)	19 864.00

Tabla 2. Parámetros considerados para la evaluación del genoma de la mitocondria de *Vaccinium floribundum* en los distintos ensambladores.

Ensamblador	Tamaño	Alineamiento más largo	NGA50	Ratio de duplicación	Indels por 100 000 pb
Hinge	316 250	44 672	9903	0.993	829.49
Canu	317 053	44 931	9082	0.997	437.92
SMARTde novo	318 595	44 989	10 823	0.998	306.81
Flye	103 067	18 975	.	1.001	257.94

Tabla 3. Parámetros considerados para la evaluación del genoma de la mitocondria de *Vaccinium floribundum*, en el ensamblado y pulido.

Programas para pulido de genomas	Ensamblador	Tamaño (pb)	Alineamiento más largo (pb)	NGA50	Radio de duplicación	Indels por 100 000 (pb)
Racon	Hinge*	300 062	36 323	7190	0.999	323.08
	Canu	298 975	36 791	6609	0.998	305.69
	SMARTdenovo*	301 975	43 539	7192	0.999	293.33
	Flye	84 445	17 398	-	0.995	311.49
Nanopolish	Hinge*	316 875	44 814	9950	0.995	598.85
	Canu	316 468	44 804	9067	0.995	522.99
	SMARTdenovo*	317 782	44 894	10 794	0.996	447.95
	Flye	103 062	18 975	.	1.000	260.98
Racon + Nanopolish	Hinge*	300 069	36 323	7190	0.999	320.37
	Canu	298 957	36 791	6609	0.998	307.39
	SMARTdenovo*	301 837	43 542	7192	0.999	283.18
	Flye	84 439	17 398	-	0.995	315.2

*Genomas con mejores parámetros para siguientes pasos.

Tabla 4. Parámetros evaluados con respecto a la referencia (mitocondria *V. macrocarpon*, KF386162.1) en el llamado de variantes para anotación de genomas de la mitocondria del mortino.

Pulidores	Ensambladores	Tamaño (pb)	% identidad	Número de semejanzas (pb)	Número de diferencias (pb)
Racon	Hinge	300 062	64.140	295.872	165.421
	SMARTdenovo	301 824	64.557	297.776	163.486
Nanopolish	Hinge*	312 751	66.588	307.043	154.063
	SMARTdenovo*	312 896	66.803	307.726	152.920
Racon+Nanopolish	Hinge	300 069	64.146	295.896	165.387
	SMARTdenovo	301 837	64.567	297.819	163.435

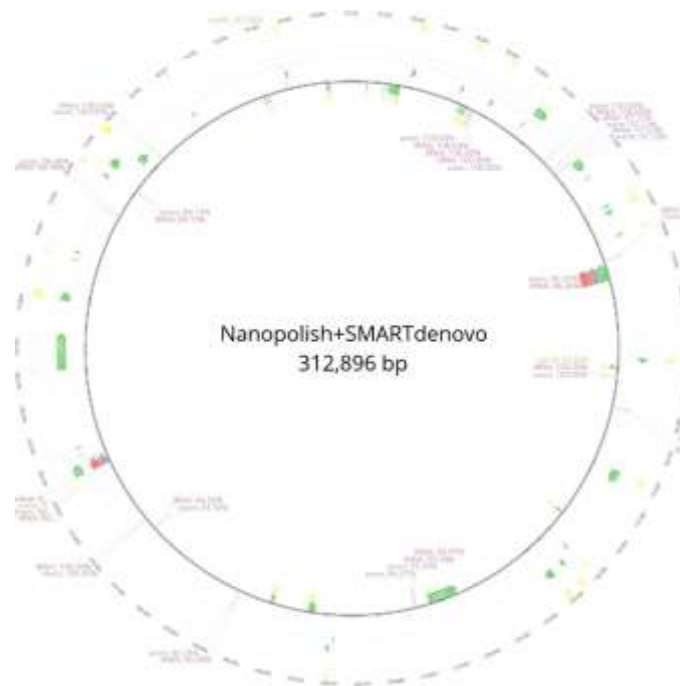
*Genomas con mejores parámetros para anotación

7. FIGURAS



Figura 1. Protocolo de análisis bioinformático para el ensamblaje del genoma de la mitocondria de *Vacinnium floribundum*

a.



b.

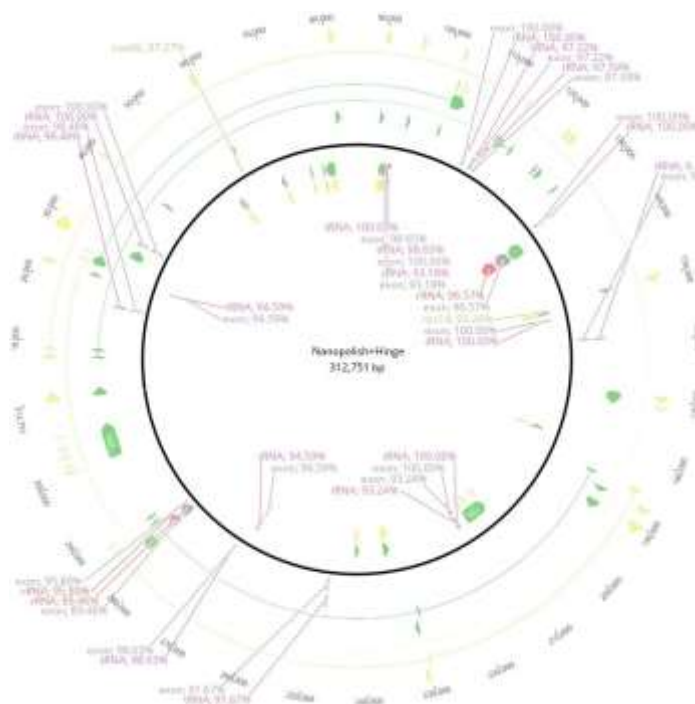


Figura 2. Estructura y mapa de los genes encontrados en el genoma de la mitocondria de *Vaccinium floribundum*. Ensamblados en a. SMARTdenovo b. Hinge y pulidos con Nanopolish.

8. REFERENCIAS

- Aanen, D. K., Spelbrink, J. N., & Beekman, M. (2014). What cost mitochondria? The maintenance of functional mitochondrial DNA within and across generations. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 369, Issue 1646, p. 20130438). *The Royal Society*. <https://doi.org/10.1098/rstb.2013.0438>
- Baker, M. (2012). De novo genome assembly: what every biologist should know. In *Nature Methods* (Vol. 9, Issue 4, pp. 333–337). *Springer Science and Business Media LLC*. <https://doi.org/10.1038/nmeth.1935>
- Biomatters Ltd. (2022). Geneious Prime 2022.1 User Manual. (PDF). <https://assets.geneious.com/documentation/geneious/GeneiousPrimeManual.pdf>
- Chamorro, G. (2020). Determinación de métodos de propagación sexual y asexual del mortiño (*Vaccinium floribundum*) con fines de conservación de la especie. Previo a la obtención de título de ingeniería en ciencias ambientales y ecodesarrollo. <https://dspace.pucesi.edu.ec/bitstream/11010/423/1/1.%20TESIS%20%28GRACE%20CAROLINA%20CHAMORRO%20GARCIA%29.pdf>
- Chiluisa, P. (2015). *Vaccinium floribundum* Kunth, reserva de antocianinas en los páramos. <http://repositorio.puce.edu.ec/handle/22000/9930>
- Coba, P., Coronel, D., Verdugo, K., Paredes, M., Yugsi, E., Huachi, L. (2012). ESTUDIO ETNOBOTÁNICO DEL MORTIÑO (*Vaccinium floribundum*) COMO ALIMENTO ANCESTRAL Y POTENCIAL ALIMENTO FUNCIONAL. *La Granja*, 16, 2. ISSN: 1390-3799
- Cobo, M. M., Gutiérrez, B., Torres, A. F., & Torres, M. de L. (2016). Preliminary analysis of the genetic diversity and population structure of mortiño (*Vaccinium floribundum*

- Kunth). *Biochemical Systematics and Ecology*, 64, 14–21.
doi:10.1016/j.bse.2015.11.008
- Dida, F., & Yi, G. (2021). Empirical evaluation of methods for de novo genome assembly. In *PeerJ Computer Science* (Vol. 7, p. e636). *PeerJ*. <https://doi.org/10.7717/peerj-cs.636>
- Fajardo, D., Schlautman, B., Steffan, S., Polashock, J., Vorsa, N., & Zalapa, J. (2014). The American cranberry mitochondrial genome reveals the presence of selenocysteine (tRNA-Sec and SECIS) insertion machinery in land plants. In *Gene* (Vol. 536, Issue 2, pp. 336–343). Elsevier BV. <https://doi.org/10.1016/j.gene.2013.11.104>
- Fukasawa, Y., Ermini, L., Wang, H., Carty, K., & Cheung, M. S. (2020). LongQC: A Quality Control Tool for Third Generation Sequencing Long Read Data. *G3* (Bethesda, Md.), 10(4), 1193–1196. <https://doi.org/10.1534/g3.119.400864>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. In *Bioinformatics* (Vol. 29, Issue 8, pp. 1072–1075). *Oxford University Press* (OUP). <https://doi.org/10.1093/bioinformatics/btt086>
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. In *Genome Biology* (Vol. 17, Issue 1). *Springer Science and Business Media LLC*. <https://doi.org/10.1186/s13059-016-1103-0>
- Jung, H., Ventura, T., Chung, J. S., Kim, W.-J., Nam, B.-H., Kong, H. J., Kim, Y.-O., Jeon, M.-S., & Eyun, S. (2020). Twelve quick steps for genome assembly and annotation in the classroom. In F. Ouellette (Ed.), *PLOS Computational Biology* (Vol. 16, Issue 11, p. e1008325). *Public Library of Science* (PLOS). <https://doi.org/10.1371/journal.pcbi.1008325>

- Kamath, G. M., Shomorony, I., Xia, F., Courtade, T. A., & Tse, D. N. (2017). HINGE: long-read assembly achieves optimal repeat resolution. In *Genome Research* (Vol. 27, Issue 5, pp. 747–756). Cold Spring Harbor Laboratory. <https://doi.org/10.1101/gr.216465.116>
- 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. doi:10.1038/s41587-019-0072-8
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. In *Genome Research* (Vol. 27, Issue 5, pp. 722–736). Cold Spring Harbor Laboratory. <https://doi.org/10.1101/gr.215087.116>
- Lathe, W., Williams, J., Mangan, M. & Karolchik, D. (2008) Genomic Data Resources: Challenges and Promises. *Nature Education* 1(3):2
- Lee, H.-O., Choi, J.-W., Baek, J.-H., Oh, J.-H., Lee, S.-C., & Kim, C.-K. (2018). Assembly of the Mitochondrial Genome in the Campanulaceae Family Using Illumina Low-Coverage Sequencing. In *Genes* (Vol. 9, Issue 8, p. 383). MDPI AG. <https://doi.org/10.3390/genes9080383>
- Liu, H., Wu, S., Li, A., & Ruan, J. (2021). SMARTdenovo: a de novo assembler using long noisy reads. In *Gigabyte* (Vol. 2021, pp. 1–9). GigaScience Press. <https://doi.org/10.46471/gigabyte.15>
- Mower, J. P., Jain, K., & Hepburn, N. J. (2012). The Role of Horizontal Transfer in Shaping the Plant Mitochondrial Genome. *Mitochondrial Genome Evolution*, 41–69. doi:10.1016/b978-0-12-394279-1.00003-x

- Muzzey, D., Evans, E. A., & Lieber, C. (2015). Understanding the Basics of NGS: From Mechanism to Variant Calling. In *Current Genetic Medicine Reports* (Vol. 3, Issue 4, pp. 158–165). *Springer Science and Business Media LLC*. <https://doi.org/10.1007/s40142-015-0076-8>
- Morisse, P., Marchet, C., Limasset, A., Lecroq, T., & Lefebvre, A. (2021). Scalable long read self-correction and assembly polishing with multiple sequence alignment. In *Scientific Reports* (Vol. 11, Issue 1). *Springer Science and Business Media* <https://doi.org/10.1038/s41598-020-80757-5>
- Palmer, J. D., Adams, K. L., Cho, Y., Parkinson, C. L., Qiu, Y.-L., & Song, K. (2000). Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. In *Proceedings of the National Academy of Sciences* (Vol. 97, Issue 13, pp. 6960–6966). *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.97.13.6960>
- Parra, M. (2021). Ensamblaje *de novo*, anotación, y filogenia del cloroplasto del mortiño (*Vaccinium floribundum* Kunth.). Tesis previa a la obtención del título de Ingeniera en Biotecnología,
- Reiling, S., Chen, S., Ragoussis, I. (2020). McGill Nanopore Ligation LibPrep Protocol SQK-LSK109. *McGill University*. [dx.doi.org/10.17504/protocols.io.bpegmjbw](https://doi.org/10.17504/protocols.io.bpegmjbw)
- Senol Cali, D., Kim, J. S., Ghose, S., Alkan, C., & Mutlu, O. (2019). Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. *Briefings in bioinformatics*, 20(4), 1542–1559. <https://doi.org/10.1093/bib/bby017>
- Shikov, A., Tsay, V., Fedyakov, M., Eismont, Y., Rudnik, A., Urasov, S., Sherbak, S., & Glotov, O. (2021). The application of Nanopore sequencing for variant calling on the

- human mitochondrial DNA. *Biological Communications*, 66(2), 109–123.
<https://doi.org/10.21638/spbu03.2021.202>
- Simpson Lab. (2019). Nanopolish Documentation.
<https://readthedocs.org/projects/nanopolish/downloads/pdf/latest/>
- Vega-Polo, P., Cobo, M. M., Argudo, A., Gutierrez, B., Rowntree, J., & Torres, M. de L. (2020). Characterizing the genetic diversity of the Andean blueberry (*Vaccinium floribundum* Kunth.) across the Ecuadorian Highlands. *Plos One*, 15(12), e0243420. doi:10.1371/journal.pone.0243420
- Wang, W., Schalamun, M., Morales-Suarez, A., Kainer, D., Schwessinger, B., & Lanfear, R. (2018). Assembly of chloroplast genomes with long- and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case. *BMC genomics*, 19(1), 977. <https://doi.org/10.1186/s12864-018-5348-8>
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. In *Nature Biotechnology* (Vol. 39, Issue 11, pp. 1348–1365). *Springer Science and Business Media LLC*.
<https://doi.org/10.1038/s41587-021-01108-x>
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. In *Genome Biology* (Vol. 20, Issue 1). *Springer Science and Business Media LLC*. <https://doi.org/10.1186/s13059-019-1727-y>
- Xu, J., Luo, H., Nie, S., Zhang, R. G., & Mao, J. F. (2021). The complete mitochondrial and plastid genomes of *Rhododendron simsii*, an important parent of widely cultivated azaleas. *Mitochondrial DNA. Part B, Resources*, 6(3), 1197–1199.
<https://doi.org/10.1080/23802359.2021.1903352>

9. ANEXOS

ANEXO 1. Características de los ensamblajes más destacados de *V. floribundum* con respecto a la referencia.

Parámetro	Mitocondria de <i>V. floridundum</i> Nanopolish + SMARTdenovo	Mitocondria de <i>V. floridundum</i> Nanopolish + Hinge	Mitocondria de <i>V. macrocarpon</i>
Tamaño (pb)	312 896	312 751	459 678
Número de genes	37	37	33
Secuencias codificantes (CDS)	35	35	35
Genes ARNt	17	17	18
Genes ARNr	3	3	3