

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

**Ensamblaje *de novo*, anotación, y filogenia del cloroplasto del
mortiño (*Vaccinium floribundum* Kunth.)**

Manuela Salomé Parra Vaca

Ingeniería en Biotecnología

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

Quito, 17 de diciembre del 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias biológicas y Ambientales

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

**Ensamblaje *de novo*, anotación, y filogenia del cloroplasto del
mortiño (*Vaccinium floribundum* Kunth.)**

Manuela Salomé Parra Vaca

Nombre del profesor, Título académico

María de Lourdes Torres, PhD

Quito, 17 de diciembre de 2021

© 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: Manuela Salomé Parra Vaca

Código: 00203123

Cédula de identidad: 1724225238

Lugar y fecha: Quito, 17 de diciembre de 2021

ACLARACIÓN PARA PUBLICACIÓN

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

UNPUBLISHED DOCUMENT

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

RESUMEN

El mortiño (*Vaccinium floribundum* Kunth.) es un arbusto silvestre, endémico de la región andina con importancia cultural, económica, y ecológica para el Ecuador. A pesar de esto, hasta el momento no se ha reportado estudios genómicos de la especie. El presente trabajo reporta el genoma del cloroplasto de *V. floribundum*, ensamblado a partir de lecturas largas, y anotado. El genoma del cloroplasto es de gran interés ya que es una fuente de marcadores moleculares útiles para estudios de identificación y diversidad genética, además gracias a su estructura y genes altamente conservados es posible su uso en análisis taxonómicos y evolutivos aportando así a la resolución filogenética del género *Vaccinium*. Para el ensamblaje del genoma del cloroplasto del mortiño se utilizó diferentes programas bioinformáticos, y se evaluó el desempeño de cada flujo de trabajo. Se encontró que un ensamblaje en los programas Hinge o SMARTdenovo más una fase de pulido en Nanopolish resulta en genomas del cloroplasto del mortiño precisos y cercanos al cloroplasto de *Vaccinium macrocarpon*, usado como genoma de referencia. Con el genoma anotado se logró realizar un análisis filogenético usando tres de los genes encontrados (*matK*, *rbcL*, *ndhF*). Esto demuestra la posibilidad de obtener genomas continuos y de calidad con el presente flujo de trabajo, sin embargo, se recomienda incorporar al análisis bioinformático lecturas cortas, más genomas de referencia, y un mayor número de genes en el análisis filogenético con el fin de obtener resultados aún más confiables. Finalmente, este estudio reporta el primer genoma del cloroplasto de una especie silvestre andina del Ecuador, y hace disponible un recurso genético útil para una variedad de estudios posteriores.

Palabras clave: mortiño, *Vaccinium floribundum*, genoma del cloroplasto, secuenciación, ensamblaje de genomas, anotación de genes.

ABSTRACT

Vaccinium floribundum Kunth., commonly known as mortiño, is a wild shrub endemic to the Andean region, with cultural, economic, and ecologic value for Ecuador. Despite of that, there are no reports about genomic studies of *V. floribundum*. This study reports the chloroplast genome of *V. floribundum* assembled with long reads, and annotated. The chloroplast genome is of great interest, as it is a source of molecular markers for identification and genetic diversity studies. Moreover, thanks to its highly conserved structure and genes, chloroplast genomes are useful for inferring taxonomic and evolutive relationships, thus contributing to the phylogenetic resolution of the *Vaccinium* genus. Different bioinformatic programs were used for the assembly of the mortiño chloroplast genome, and the performance of each pipeline was evaluated. It was found that an assembly in Hinge or SMARTdenovo, plus polishing in Nanopolish produces a *V. floribundum* chloroplast genome that is accurate and close to the chloroplast genome of *Vaccinium macrocarpon*, used as a reference sequence. With the annotated genome, a phylogenetic analysis was performed using three of the genes founded (*matK*, *rbcL*, *ndhF*). This demonstrates that continuous and quality genomes can be achieved with the present pipeline, however, it is recommended to incorporate short reads, more reference genomes, and more genes into the phylogenetic analysis in order to obtain even more reliable results. Finally, this study reports the first chloroplast genome of a wild Ecuador Andean species, and makes available a useful genetic resource for a variety of subsequent studies.

Key words: mortiño, *Vaccinium floribundum*, chloroplast genome, sequencing, genome assembly, gene annotation.

TABLA DE CONTENIDOS

1	INTRODUCCIÓN	12
1.1	Generalidades de <i>Vaccinium Floribundum</i> Kunth	12
1.1.1	Generalidades, distribución, propiedades.	12
1.1.2	Importancia económica, cultural, y ecológica.	12
1.1.3	Estudios moleculares preliminares de <i>V. floribundum</i> en el Ecuador.	13
1.2	Secuenciación, anotación, y ensamblaje de genomas.....	13
1.3	Cloroplasto.....	14
1.4	Secuenciación del cloroplasto de <i>V. floribundum</i>	15
2	METODOLOGÍA	16
2.1	Recolección de muestras de <i>V. floribundum</i>	16
2.2	Secuenciación en MinION del genoma de <i>V. floribundum</i>	16
2.3	Análisis bioinformático	16
2.3.1	Pre-procesamiento de lecturas de la secuenciación de <i>V. floribundum</i>	16
2.3.2	Extracción de las lecturas del cloroplasto de <i>V. floribundum</i>	17
2.3.3	Ensamblaje del genoma del cloroplasto de <i>V. floribundum</i>	17
2.3.4	Pulido del genoma del cloroplasto de <i>V. floribundum</i>	17
2.3.5	Evaluación de los genomas del cloroplasto de <i>V. floribundum</i>	17
2.3.6	Llamado de Variantes en los genomas del cloroplasto de <i>V. floribundum</i>	18
2.3.7	Anotación de los genomas del cloroplasto de <i>V. floribundum</i>	18
2.4	Filogenia de <i>V. floribundum</i>	18
3	RESULTADOS	19
3.1	Secuenciación en MinION del genoma de <i>V. floribundum</i>	19
3.2	Evaluación y comparación de los genomas del cloroplasto de <i>V. floribundum</i>	19
3.2.1	Ensamblaje del genoma del cloroplasto de <i>V. floribundum</i>	19
3.2.2	Ensamblajes del cloroplasto del <i>V. floribundum</i> después del pulido.....	20
3.2.3	Llamado de Variantes de los ensamblajes con métricas sobresalientes.	21
3.3	Estructura y genes del genoma del cloroplasto del mortiño	21
3.4	Filogenia de <i>V. floribundum</i>	22
4	DISCUSIÓN	23
4.1	Secuenciación en MinION de <i>V. floribundum</i>	23
4.2	Evaluación y comparación de los ensamblajes del cloroplasto de <i>V. floribundum</i>	23
4.2.1	Ensamblaje del genoma del cloroplasto de <i>V. floribundum</i>	23
4.2.2	Ensamblaje del genoma del cloroplasto de <i>V. floribundum</i> después del pulido....	24
4.3	Características del genoma del cloroplasto de <i>V. floribundum</i>	25
4.4	Filogenia de <i>V. floribundum</i>	26
5	CONCLUSIONES	27
6	TABLAS	28
7	FIGURAS	32

8 REFERENCIAS.....37

ÍNDICE DE TABLAS

Tabla 1. Estadísticas importantes del desempeño de la secuenciación de <i>Vaccinium floribundum</i>	28
Tabla 2. Parámetros considerados para la evaluación del genoma del cloroplasto de <i>Vaccinium floribundum</i> ensamblado en cuatro distintos programas.	29
Tabla 3. Parámetros considerados para la evaluación del genoma del cloroplasto de <i>Vaccinium floribundum</i> ensamblado y pulido en dos programas.	30
Tabla 4. Características de los ensamblajes óptimos de <i>V. floribundum</i> (H+N y S+N) y el genoma de referencia (<i>V. macrocarpon</i> JQ757046.1)	31

ÍNDICE DE FIGURAS

Figura 1. Flujo de trabajo en el análisis bioinformático para el ensamblaje del cloroplasto de <i>Vaccinium floribundum</i>	32
Figura 2. Alineamiento entre el genoma de referencia (cloroplasto de <i>V. macrocarpon</i> JQ757046.1) con los ensamblajes del genoma del cloroplasto de <i>Vaccinium floribundum</i>	33
Figura 3. Estructura y mapa de los genes encontrados en el genoma del cloroplasto de <i>Vaccinium floribundum</i>	34
Figura 4. Análisis filogenético de <i>V. floribundum</i> utilizando tres genes representativos del cloroplasto.....	35

ÍNDICE DE ANEXOS

ANEXO 1: Especies y genes usados en el análisis filogenético de *Vaccinium floribundum*..... 44

1 INTRODUCCIÓN

1.1 Generalidades de *Vaccinium Floribundum* Kunth

1.1.1 Generalidades, distribución, propiedades.

Vaccinium floribundum Kunth. (*V. floribundum*), más conocido como mortiño o uva de monte, es un arbusto silvestre, endémico de la región andina que pertenece a la familia Ericaceae. Se distribuye en el noroccidente de Suramérica, principalmente en Venezuela, Colombia, Ecuador, y Perú (Coba et al., 2012). En el Ecuador, el mortiño crece en los páramos, ecosistemas caracterizados por climas templados y fríos (7-18°C), vegetación no arbórea, baja presión atmosférica, y alta radiación solar (Giné & Sánchez, 2015). El mortiño se desarrolla entre los 2-3 metros de altura, posee una inflorescencia formada por 6-10 flores rosadas, y un fruto en forma de baya esférica, lisa, y carnosa de un color violeta-azul (Coba et al., 2012; Chiluisa, 2015). Sus hojas y frutos poseen compuestos bioactivos como polifenoles, antocianinas, antioxidantes, y flavonoides que brindan varios beneficios a la salud cuando se consumen (Coba et al., 2012; Llivisaca et al., 2018).

1.1.2 Importancia económica, cultural, y ecológica.

El mortiño juega un rol cultural y económico importante en el Ecuador. Sus frutos son usados en la preparación de bebidas tradicionales, mermeladas, vinos, postres, y demás productos que son una fuente relevante de ingresos para comunidades indígenas (Coba et al., 2012; Roldán, 2013). A pesar de esto, no es posible cultivar mortiño debido a la dificultad de su propagación y escasa información sobre sus requerimientos agronómicos, es así que toda su demanda se satisface a través de la recolección silvestre del mismo durante los meses de octubre a diciembre (Cerón, 2019). *Vaccinium floribundum* también posee un rol ecológico clave en el ecosistema alpino de altura al que pertenece (Ramsay & Oxley, 1996). Se ha reportado que durante los últimos años el páramo, ecosistema de vital importancia para el país por ser fuente de recursos hídricos, ha estado bajo presión antropogénica lo que ha causado su fragmentación, y pérdida de hábitats (Chuncho y

Chuncho, 2019). No obstante, se ha visto que tras eventos de quema y deforestación el mortiño es una de las primeras especies en recuperarse (Ramsay & Oxley, 1996; De la Torre et al., 2008).

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

Teniendo en cuenta la importancia etnobotánica y ecológica del mortiño, se han realizado esfuerzos por describir la población silvestre de esta especie. Es así que desde el 2016 se han desarrollado estudios para determinar la diversidad genética y estructura poblacional de *V. floribundum* en los páramos del Ecuador (Cobo et al., 2016; Vega-Polo et al., 2020). En estas investigaciones se encontró que la población del mortiño en Ecuador tiene un grado alto de diversidad genética, y su estructura poblacional responde a la interacción entre las dinámicas de aislamiento por distancia, y aislamiento por altitud, creadas por la geografía característica de la región andina (Vega-Polo et al., 2020).

1.2 Secuenciación, anotación, y ensamblaje de genomas

Durante las últimas décadas ha existido un desarrollo significativo de técnicas de secuenciación y herramientas bioinformáticas lo que ha provocado mayor accesibilidad al secuenciamiento, ensamblaje, y anotación de genomas, convirtiéndose en una herramienta clave para la investigación dentro de las ciencias biológicas (Campos De Quiroz, 2002; Jung et al., 2020). La secuenciación por nanoporo es la técnica más actual y pertenece a las tecnologías de secuenciamiento de tercera generación (TGS) (Heather & Chain, 2015; Kchouk et al., 2017). Este sistema consiste en el paso de cadenas simples de ADN a través de un canal proteico de escala nanométrica ubicado en una membrana delgada que separa dos cámaras (cis y trans) (Agah et al., 2016). Dentro del nanoporo existe una corriente de iones, y cuando el ADN lo atraviesa crea una interferencia con la misma. Un sensor lo detecta y genera medidas de corriente y tiempo de translocación. Con el uso de un programa de llamado de bases (*base calling*) se transforman todas

estas señales en la secuencia de nucleótidos de la molécula de ADN (Jurkowski, 2020; Lin et al., 2021). En el 2014, Oxford Nanopore Technologies (ONT) desarrolló la primera y más pequeña plataforma de secuenciación por nanoporo, MinION (90 gramos, 10 cm de largo). Este es un equipo portable y asequible, capaz de producir datos en tiempos cortos. Se conecta a laptops con pocos requisitos de hardware, y no requiere equipos de imagen para detectar nucleótidos (Bleidorn, 2015; Jain et al., 2016). Todas estas propiedades han permitido su uso en diversas aplicaciones tanto para enseñanza como investigación, e incluso gracias a su portabilidad se ha usado en ambientes tan extremos como estaciones espaciales y regiones polares (Lu et al., 2016).

Gracias a esta tecnología se ha logrado ensamblar y anotar los genomas de varias especies de bacterias, animales, y plantas de interés (Lu et al., 2016; Jain et al., 2018; Dumschott et al., 2020). El ensamblaje se refiere al proceso mediante el cual se determina el orden correcto de todas las lecturas generadas en la secuenciación. La estrategia tras el ensamblaje consta de tres pasos: 1) formación de *contigs* a través del solapamiento de lecturas, 2) generación de *scaffolds* al unir *contigs* de forma ordenada, y 3) relleno de *gaps* (Sohn & Nam, 2016). La anotación, por otro lado, se refiere a identificar y describir las regiones de interés biológico dentro del genoma, tanto estructural como funcionalmente. Esto se basa en una búsqueda de homología y mapeo de ontología genética utilizando bases de datos, así como el *Gene Ontology Resource* (GO) que posee información sobre la función de genes, o la enciclopedia de genes y genomas de Kyoto (KEGG) que relaciona cada gen con la vía metabólica en la que participa, entre otros repositorios electrónicos (Domingues del Angel et al., 2018; Jung et al., 2020).

1.3 Cloroplasto

El cloroplasto es el centro metabólico de las plantas y juega un papel importante en la fisiología, desarrollo, y evolución de las mismas (Freudenthal et al., 2020; Daniell et al., 2021). Su genoma posee un tamaño de 120-220 kilobases (kb) y está organizado en cuatro regiones: región

de copia única grande, región de copia única pequeña, y un par de repeticiones invertidas; más conocidas como LSC, SSC, IRa e IRb, respectivamente, por sus siglas en inglés (Scheunert et al., 2020; Lin et al., 2020). Posee entre 110-113 genes, los que en su mayoría codifican proteínas y enzimas involucradas en la fotosíntesis (Lin et al., 2020).

1.4 Secuenciación del cloroplasto de *V. floribundum*

En los últimos años se han logrado secuenciar y ensamblar genomas de cloroplastos de varias especies de interés agronómico y biológico como los de *Oryza sativa*, *Solanum tuberosum*, *Arabidopsis thaliana*, entre otros. (Wang & Gao, 2016; Zhao et al., 2019; Park et al., 2020). Sin embargo, hay pocos reportes de secuenciación de cloroplastos de especies nativas del Ecuador (Su et al., 2018; Tao et al., 2020), el mortiño no se incluye dentro de estas. Es así que la secuenciación, ensamblaje, y anotación del cloroplasto de *Vaccinium floribundum* resulta interesante, ya que es una fuente valiosa de marcadores moleculares útiles para diversidad genética e identificación molecular (Daniell et al., 2016). Así también, la secuencia del genoma del cloroplasto es altamente conservada en términos de estructura, tamaño, y genes (Lin et al., 2020), por lo que el genoma del cloroplasto del mortiño permitiría realizar una reconstrucción confiable de la historia evolutiva de la especie, aportando así en el entendimiento del género polifilético al que pertenece (*Vaccinium*) (Kron et al., 2002; Wei et al., 2020).

Este trabajo propone evaluar diferentes estrategias de ensamblaje combinando distintos programas, para de esta forma definir el análisis bioinformático más adecuado que permita obtener el genoma del cloroplasto del mortiño, y así proporcionar el primer recurso genómico disponible de *Vaccinium floribundum*.

2 METODOLOGÍA

2.1 Recolección de muestras de *V. floribundum*

Se colectaron hojas de un arbusto de mortiño (individuo denominado P079) ubicado en la localidad de Lloa, Pichincha-Ecuador (S0° 11.55122' W78° 35.216') bajo el contrato marco MAE-DNB-CM-2016-0046-M-0002. La muestra se almacenó en sílica gel durante el transporte al Laboratorio de Biotecnología Vegetal de la USFQ. Una vez en el laboratorio, la muestra se conservó a -20 °C durante esa noche, y al siguiente día se transfirió al ultracongelador de -80 °C para su almacenamiento hasta la extracción de ADN.

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

El ADN se extrajo usando el protocolo de ADN de alto peso molecular (HMW) propuesto por Green y Sambrook (2017). Luego, se eliminaron los fragmentos pequeños de ADN (< 10Kb) con el kit *Short Read Eliminator* (Jones et al., 2020). Posteriormente, se prepararon las librerías de secuenciación con el kit de ligación SQK-LSK109 de ONT (Reiling et al., 2020). Se siguió el protocolo especificado por el fabricante con algunas modificaciones: se inició con 2 μ g de ADN en lugar de 1 μ g, además en el paso de ligación de adaptadores se incubó el ADN con buffer de ligación, mezcla de adaptadores, y ligasa por 20 minutos en lugar de 10. La librería se cargó en dos celdas de flujo del equipo MinION Mk1B para su secuenciación. El software MinKNOW v.7. y Guppy v1.1.alpha13-0-g1ec7786 se usaron para el llamado de bases. Finalmente, se generó un reporte del estado de las lecturas en LongQC (Fukasawa et al., 2020).

2.3 Análisis bioinformático

El flujo de trabajo se encuentra ilustrado en la Figura 1.

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

Se utilizó Porechop v0.2.4 para remover los adaptadores de las lecturas. Después con Nanofilt v2.8.0 se filtró y eliminó todas las lecturas con calidad menor a 9 y tamaño menor a 1 kb (Wang et al., 2018).

2.3.2 Extracción de las lecturas del cloroplasto de *V. floribundum*.

Debido a que el ADN extraído pertenecía al genoma total del se realizó una extracción de lecturas del cloroplasto utilizando la herramienta Blasr v5.3.5 (Wang et al., 2018). Este programa permitió el alineamiento de las lecturas con un genoma de referencia, en este caso fue el genoma del cloroplasto de *Vaccinium macrocarpon* (código de acceso: JQ757046.1) (Chaisson & Tesler, 2012). Esta especie se escogió como referencia por ser una de las pocas especies del género *Vaccinium* con cloroplasto ensamblado, anotado, y verificado por el NCBI; además por presentar la distribución geográfica más cercana a la de *V. floribundum* (Fajardo et al., 2013).

2.3.3 Ensamblaje del genoma del cloroplasto de *V. floribundum*.

Se usó cuatro programas de ensamblaje *de novo* de lecturas largas: 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). Se utilizaron los parámetros pre-establecidos, excepto en el ensamblador Flye, en el cual se aumentó el parámetro `--asm-coverage=50` para minimizar el uso de memoria (Kolmogorov et al., 2019). Para todos los programas se especificó un tamaño de genoma esperado de 210 Kb en base a la cobertura obtenida y los tamaños reportados de cloroplastos del género *Vaccinium* (Fajardo et al., 2012; Kim et al., 2018; Guo et al., 2019).

2.3.4 Pulido del genoma del cloroplasto de *V. floribundum*.

Se reordenó los *contigs* de cada ensamblaje en base a la estructura y orientación del genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1) utilizando el script de Wang et al. (2018). El reordenamiento generó secuencias consenso formadas por un solo *contig*, las cuales fueron pulidas en Racon v1.4.20, y Nanopolish v0.13.2 (Wang et al., 2018).

2.3.5 Evaluación de los genomas del cloroplasto de *V. floribundum*.

Para evaluar y comparar los ensamblajes generados se alineó cada genoma obtenido con el genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1) usando Mummer v3.23. Adicionalmente, se eligió el ensamblaje óptimo en base a las siguientes métricas calculadas por el programa QUAST: NGA50/ tamaño de alineamiento más largo, tamaño del genoma, radio de duplicación, cantidad de indels por cada 100 000 pb, y número de caracteres genómicos (Gurevich et al., 2013; Wang et al., 2018).

2.3.6 Llamado de Variantes en los genomas del cloroplasto de *V. floribundum*.

Se realizó el llamado de variantes/SNPs de los ensamblajes con las mejores métricas en Geneious Prime v.2021.0.3 con la herramienta "Find Variations/SNPs" (Biomatters Ltd, 2021).

2.3.7 Anotación de los genomas del cloroplasto de *V. floribundum*.

De igual manera, solo se anotaron los ensamblajes con parámetros sobresalientes en el programa en línea Chloe (Zhong, 2020). Para identificar las secuencias repetidas invertidas (IR) del genoma se utilizó el plug-in "Find repeats" de Geneious Prime v.2021.0.3 (Gibss, 2019).

2.4 Filogenia de *V. floribundum*

En base a la literatura se escogieron tres genes representativos del genoma del cloroplasto (*matK*, *rbcL*, *ndhF*) de los cuales se ha reportado su uso frecuente en la resolución de relaciones filogenéticas (Hilu et al., 2008; Li, 2008; Peterson et al., 2010; Llivisaca et al., 2018; Amar, 2020). Con el uso de BLAST se identificaron secuencias homólogas de estos genes en 12 especies del género *Vaccinium* (Anexo 1), las cuales fueron las más cercanas a *V. floribundum* (Kron et al., 2002). Especies de la familia Actinidiaceae, del orden Ericales, se eligieron como grupo externo (Kim et al., 2018). Se concatenaron los tres genes de cada especie, y se alinearon en Aliview v.1.27 (Larsson, 2014). Luego, en la versión web de IQtree se construyeron árboles filogenéticos con el método de máxima verosimilitud (ML) utilizando los parámetros pre-establecidos (1000 corridas de bootstrap) (Trifinopoulos et al., 2016).

3 RESULTADOS

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

En la Tabla 1 se resumen las estadísticas importantes del ensayo de secuenciación. Este ensayo generó 20 GB de información con 2 964 735 lecturas de las cuales el 2.37% (70 195 lecturas) pertenecen al cloroplasto. Esto generó una cobertura promedio por lectura de 1962.388 X. Es decir, cada base generada se secuenció 1962.388 veces. En cuanto al tamaño de las lecturas, la de mayor tamaño fue de 121 871 pb, que es lo esperado al usar la tecnología ONT (Amarasinghe et al., 2020). Finalmente, el N50 o mediana obtenida fue de 12 710 pb, lo que indica que lecturas de ese tamaño o mayores contienen al menos al 50% de los nucleótidos del conjunto total de secuencias generadas (Kahlke, 2020).

3.2 Evaluación y comparación de los genomas del cloroplasto de *V. floribundum*

3.2.1 Ensamblaje del genoma del cloroplasto de *V. floribundum*.

En la Tabla 2 se muestra las métricas calculadas de los cuatro ensamblajes del genoma del cloroplasto del mortño. Entre estos, el genoma ensamblado en Flye fue el más grande (164 439 pb) y presentó el alineamiento más largo (119 447 pb); mientras que el genoma de Hinge presentó el tamaño de genoma y de alineamiento más corto (163 686 pb y 118 689 pb, respectivamente). Sobre el radio de duplicación, el genoma ensamblado por Canu presentó un valor igual a 1.149 mientras que el resto de genomas ensamblados presentaron radios de entre 0.995 a 0.998. Se ha reportado que valores mayores a 1.00 indican eventos de duplicación en el genoma (Wang et al., 2018). Es decir, el ensamblaje de Canu presenta regiones duplicadas las cuales no se reportan en el genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1). En cuanto a la cantidad de indels (inserciones y deleciones) por cada 100 000 pb, el genoma de Hinge presenta la mayor cantidad (1085.36), seguido por Canu (732.27), SMARTdenovo (599.14), y Flye (437.09). Finalmente, respecto al número de genes presentes en el genoma de referencia (cloroplasto de *V.*

macrocarpon JQ757046.1) que se encontraron en cada genoma del cloroplasto del mortño, Canu presentó 297 genes. Este número es bajo al comparar con los 332 genes que se encontraron en el resto de genomas del cloroplasto de *V. floribundum*.

3.2.2 Ensamblajes del cloroplasto del *V. floribundum* después del pulido.

En la Tabla 3 se resumen la métricas calculadas para los genomas del cloroplasto del mortño luego de la fase de pulido.

El uso del programa Racon para pulir los genomas del cloroplasto del mortño causó una disminución en los tamaños de las secuencias en alrededor de 7-8 kb. Por ejemplo, el tamaño del genoma cloroplasto obtenido en Hinge pasó de 163 686 pb a 156 686 pb. Aquel acortamiento del tamaño de los genomas, a su vez, provocó que el tamaño del alineamiento con el genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1) y el número de caracteres genómicos encontrados disminuyan. Los parámetros como el radio de duplicación y número de indels, fueron los únicos que mejoraron en todos los casos. Por un lado, el radio de duplicación pasó de 0.995-1.149 a 1.00 (siendo 1.00 el valor óptimo reportado). Por otro lado, el número de indels disminuyó de valores entre 599.14-1 085.36, a menos de 570, lo que indica una mayor cercanía con el genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1).

El empleo del programa Nanopolish aumentó el tamaño de todos los genomas en 1-2 kb, siendo el genoma de Hinge el más grande (165 116 pb) y el de Canu el más corto (164 705 pb). De la misma forma, NGA50 aumentó en todos los casos, en donde el genoma de Hinge presenta el alineamiento más largo (120 215 pb), y el de Flye el más corto (119 911 pb). Sobre el radio de duplicación, a excepción del genoma de Canu, todos poseen valores muy cercanos a 1.00. Sobre el número de indels, este disminuyó en todos los casos, donde SMARTdenovo posee los valores más pequeños (óptimo), y Canu los más grandes (no óptimo). Finalmente, en todos los genomas se encontró la misma cantidad de caracteres genómicos que antes del pulido.

Al comparar los resultados obtenidos, se identificó a dos genomas como los ensamblajes con parámetros sobresalientes, estos son: genoma ensamblado en Hinge y pulido en Nanopolish (H+N), y genoma ensamblado en SMARTdenovo y pulido en Nanopolish (S+N) (Tabla 3).

3.2.3 Llamado de Variantes de los ensamblajes con métricas sobresalientes.

Entre los genomas Hinge + Nanopolish (H+N) y SMARTdenovo + Nanopolish (S+N) existen 3 594 variantes (indels y SNPs) lo que se traduce a un 98.9% de identidad (similitud). Por otro lado, al compararlos con el genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1) se encontró que el genoma H+N posee 2 966 variantes, mientras que el genoma S+N 2 633 variantes. Es decir, los porcentajes de no identidad con *V. macrocarpon* son 12.4% y 9.2%, respectivamente, confirmando que se trata de dos taxones diferentes, pero cercanos.

3.3 Estructura y genes del genoma del cloroplasto del mortiño

En la Figura 2 se observa el alineamiento de los ensamblajes con el genoma de referencia, todos los genomas presentan perfiles de coincidencia bastante similares, es decir, tienen un alineamiento continuo (representado de color violeta) hasta alrededor de los 120 Kb, y un evento de inversión (representado de color celeste) en los últimos 60 kb. Este patrón hace referencia a la estructura cuadripartida característica del cloroplasto de plantas: región LSC y región SSC, separadas por dos regiones repetidas invertidas (IRa e IRb) (Lin et al., 2020). Además, los tamaños de alineamiento coinciden con los tamaños esperados de las regiones del cloroplasto del mortiño y reportados para *V. macrocarpon* por Fajardo et al. (2013): >100 Kb para la región LSC, y ~60 Kb para IRa + IRb.

En la Tabla 4 se muestra que la estructura del genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1) está organizada de la siguiente forma: LSC de 104 544 pb, IR de 34 232 pb, y SSC de 3 029 pb (Fajardo et al., 2013). Por otro lado, el genoma del cloroplasto del mortiño H+N y S+N posee: una región LSC de 105 262 pb, y 104 807 pb; IR de 27 665 pb y 28

458 pb; y región SSC de 4 524 pb y 3 270 pb, respectivamente. Es decir, en comparación con *V. macrocarpon* el LSC resulta ser ligeramente más largo (287-718 pb más), la región IR más pequeña (6 567-5 774 pb menos) y la región SSC bastante similar.

En la figura 3 se observan los resultados obtenidos en la anotación de los genomas H+N y S+N. Ambos poseen 74 secuencias codificantes de proteínas, 5 genes codificantes de ARNr, y 36 genes codificantes de ARNt. El número de genes identificados es apenas mayor al encontrado en el genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1), pues el mismo posee 73 genes codificantes de proteínas, 28 ARNt, y 4 ARNr (Fajardo et al., 2013) (Tabla 4).

3.4 Filogenia de *V. floribundum*

Se obtuvieron dos árboles filogenéticos a partir de los dos ensamblajes del cloroplasto del mortiño (H+S y S+N). Los genes concatenados de estos dos genomas indican las mismas relaciones evolutivas entre las especies analizadas pues los dos árboles comparten topología (Figura 3). Los árboles muestran que *Vaccinium corymbosum*, y *Vaccinium uliginosum* forman un clado, en el que su ancestro común es cercano a *Vaccinium erythrocarpum*. El ancestro de este clado, a su vez, se presenta como la especie hermana del mortiño. Esta relación está soportada con 93% y 96% de evidencia de bootstrap para los genomas H+N y S+N, respectivamente. *V. macrocarpon* también es cercano al ancestro del clado del mortiño, sin embargo, el soporte de esta relación es menor al 70% lo que indica una rama no tan estable (Yao et al. 2020). El resto de especies muestran un patrón de relación bastante similar a lo reportado en el estudio de Kron et al. (2002), y Tsutsumi (2011). Por otro lado, también se muestra que el grupo externo permite reconocer la raíz del árbol tal cual sucede con el estudio de Kim et al. (2018).

4 DISCUSIÓN

4.1 Secuenciación en MinION de *V. floribundum*

El cloroplasto posee cientos de copias de su genoma, por lo que es frecuente que proyectos de secuenciación de genomas de plantas tengan lecturas de cloroplastos como subproducto (Freudhental et al., 2020). Es el caso del presente trabajo, en donde de más de dos millones de lecturas obtenidas el 2.37% corresponden al cloroplasto. Teniendo en cuenta el tamaño esperado del genoma, 210 Kb, este pequeño porcentaje de lecturas se traduce a una cobertura de 1962.38 X la cual según lo reportado es suficiente para ensamblar un genoma completo. De hecho, según Ekblom y Wolf (2014) una cobertura de 45-50 X suele ser la requerida para ensamblajes de genomas. En cuanto al tamaño de las lecturas obtenidas, tanto el N50 (mediana) como el tamaño de la lectura más larga son los esperados pues se ha reportado que ONT genera lecturas desde 500 pb hasta 2.3 Mb (Amarashinghe et al., 2020).

4.2 Evaluación y comparación de los ensamblajes del cloroplasto de *V. floribundum*

4.2.1 Ensamblaje del genoma del cloroplasto de *V. floribundum*.

El enfoque utilizado para el ensamblaje del genoma del cloroplasto del mortiño fue un ensamblaje *de novo*. Con este acercamiento se generan *contigs* únicamente en base al solapamiento de lecturas (Lischer & Shimizu, 2017). Es así que, Canu, Hinge, y SMARTdenovo se basan en algoritmos de ensamblaje OLC (*Overlap Layout Consensus*), mientras que Flye emplea un algoritmo generalizado de Bruijn Graph. El algoritmo OLC posee tres etapas: (1) solapamiento de las lecturas, (2) construcción de gráficos en base al solapamiento, (3) inferencia de una secuencia consenso (Commins et al., 2009; Li et al., 2017). El algoritmo de Flye, por otro lado, (1) genera *contigs* en base a solapamientos inexactos, (2) construye gráficos donde combina los *contigs* en diferentes ensamblajes posibles, (3) genera un *contig* final preciso (Kolmogorov et al., 2019).

Hay que tener en cuenta que el desempeño de un algoritmo de ensamblaje depende de varios factores como cobertura, composición de la librería, complejidad del genoma, entre otros (Jung et al., 2020). Por ejemplo, Murigneux et al. (2020) reportan a Canu y Flye como los mejores ensambladores para el genoma nuclear de *Macadamia janseni* mientras que Wang et al. (2018) proponen a Hinge como el mejor ensamblador para el cloroplasto de *Eucalyptus pauciflora*. En este caso, la evaluación de los genomas del cloroplasto del mortiño, indica que el algoritmo de Flye es el adecuado para el presente set de datos, pues produce el genoma más largo, sin eventos de duplicación, y con el menor número de indels comparado con el resto de ensamblajes.

4.2.2 Ensamblaje del genoma del cloroplasto de *V. floribundum* después del pulido.

Se sabe que la precisión de los primeros borradores de genomas ensamblados a partir de lecturas ONT es menor al 98%. Esto debido a las altas tasas de error (10-15%) que caracterizan a la tecnología. Por esta razón es necesario corregir errores y mejorar la precisión de los genomas mediante un proceso conocido como pulido (Shin et al., 2019). En este caso se utilizaron dos programas: Racon v1.4.20, y Nanopolish v0.13.2. Racon usa un algoritmo que le permite generar una nueva secuencia consenso en base al mapeo de lecturas crudas al ensamblaje borrador (Vaser et al., 2017). Por otro lado, el algoritmo de Nanopolish propone ediciones al azar del genoma borrador, les da un puntaje en base a las señales crudas ONT, y la secuencia con mayor puntaje es el mejor ensamblaje de genoma posible (Wick et al., 2019; Simpson, 2021).

Tras evaluar cada genoma, se encontró que Racon en todos los casos generó secuencias de menor tamaño y con parámetros inferiores comparados a los genomas antes del pulido. La disminución de tamaño se explica porque su algoritmo corta regiones con baja cobertura (Vaser et al., 2017), y en el caso de genomas con regiones repetidas como el cloroplasto, se ha visto que la cobertura no es constante (Gao et al., 2019). Al contrario, tras el pulido en Nanopolish se evidencia una mejora en la precisión y calidad de todos los genomas borradores, siendo los producidos con los programas Hinge y SMARTdenovo aquellos que destacan al presentar parámetros superiores

en relación al resto de genomas (incluidos los previos al pulido). El desempeño superior de Nanopolish es esperado ya que utiliza los datos crudos de la secuenciación lo cual disminuye los errores introducidos en el llamado de bases (Senol et al., 2017). Sin embargo, su algoritmo es costoso computacionalmente y requiere tiempos de corrida mucho más largos que Racon. Por esta razón, se recomienda combinar a Racon y Nanopolish para reducir el tiempo de corrida de Nanopolish pero seguir obteniendo la precisión y calidad característica de su algoritmo (Senol et al., 2017; Vaser et al., 2017).

4.3 Características del genoma del cloroplasto de *V. floribundum*

En cuanto al tamaño, el genoma más largo obtenido para *V. floribundum* es de 165 116 pb, mientras que el genoma del cloroplasto del *V. macrocarpon* (JQ757046.1) posee 176 045 pb, el cloroplasto de *Vaccinium bracteatum* (LC521967.1) 174 404 pb, y el de *Vaccinium vitis-idaea* (LC521969.1) 173 967 pb (Fajardo et al., 2013; NCBI, 2021). Es decir, el ensamblaje del cloroplasto del mortiño posee alrededor de 10 kb menos que especies cercanas del mismo género. A pesar de que la diferencia de tamaños de genomas en plantas (cercanas o no) se explica por varias razones (elementos transponibles, poliploidización, entre otros) (Ren et al., 2018), se ha visto que los algoritmos de ensamblaje también afectan en el tamaño del genoma obtenido (Kooig & Pellicer, 2020). En este caso, el cloroplasto de *V. bracteatum* y, *V. vitis-idaea*, se ensamblaron en NOVOPlasty v 2.7.1. a partir de lecturas cortas de la tecnología Illumina (Kim et al., 2020). Por lo tanto, usar un enfoque híbrido (lecturas largas y cortas) podría aumentar el tamaño del genoma del cloroplasto del mortiño a lo reportado en cloroplastos de especies cercanas (Amarasingue et al., 2020). Por otro lado, sobre la estructura del genoma, el cloroplasto del mortiño posee las cuatro regiones conservadas características de este organelo: LSC, SSC, las cuales están separadas por dos regiones repetidas invertidas (IRa e IRb) (Lin et al., 2020).

4.4 Filogenia de *V. floribundum*

Vaccinium es un género con clasificación compleja, pues en base a estudios cladísticos moleculares se ha sugerido que se trata de un género polifilético (Kron et al., 1999). Sin embargo, en base a caracteres morfológicos, se lo ha dividido a en dos subgéneros *Oxycoccus* y *Vaccinium* (Kron et al., 2002; Johnson., 2012). *Oxycoccus* encierra a los arándanos rojos (*cranberries*), mientras que *Vaccinium* se clasifica a su vez en secciones como *Cyanococcus*, *Pyxothamnus*, entre otras. Las especies consideradas para este análisis filogenético pertenecen al subgénero *Vaccinium* a excepción de *V. macrocarpon* y *V. erythrocarpum* que son de *Oxycoccus* (Song & Hancock, 2011). *V. floribundum*, por su parte, se ha clasificado dentro de la sección *Pyxothamnus* junto a *V. ovatum* (Kron et al., 2002).

Los árboles obtenidos de los dos genomas H+N y S+N (Figura 4) comparten misma topología pero diferentes soportes de bootstrap (Sinharay, 2010). Las relaciones mostradas coinciden en su mayoría con lo reportado en la taxonomía tradicional y estudios filogenéticos del género *Vaccinium* (Powell & Kron, 2003; y Tsutsumi, 2011). Sin embargo, los árboles indican que *V. floribundum* es más cercano a especies del subgénero *Oxycoccus* cuando la literatura lo clasifica como parte de *Vaccinium*. Esto se explica porque aunque se realizó un ensamblaje *de novo* se usó a *V. macrocarpon* como genoma de referencia tanto para ordenar y orientar los *contigs*. Al usar una secuencia de referencia existe la posibilidad de introducir sesgos en el ensamblaje, y evitar la reconstrucción de regiones muy divergentes con el genoma de referencia. Por esta razón, se recomienda a futuro utilizar múltiples secuencias de referencia de especies cercanas para minimizar el sesgo (Lischer & Simizu, 2017). Así también, para tener una mejor resolución sobre la filogenia del mortiño, se puede aumentar el número de genes analizados como lo realizado por Fajardo et al. (2013) para la filogenia de *V. macrocarpon*. Se podría usar también la secuencia del cloroplasto completa como en los estudios filogenéticos de los géneros *Dracocephalum*, *Alchemilla*, y *Achyranthes* (Yao et al., 2020; Rono et al., 2020; Xu et al., 2020).

5 CONCLUSIONES

El genoma del cloroplasto de *V. floribundum* fue secuenciado y ensamblado satisfactoriamente utilizando lecturas ONT. Esto demuestra la posibilidad de obtener genomas continuos y de calidad usando únicamente lecturas largas a pesar de que existan regiones de difícil resolución como las repeticiones invertidas del genoma del cloroplasto. La tecnología ONT posee altas tasas de error, sin embargo, se logró mejorar la precisión de las secuencias incorporando la fase de pulido. Se compararon diferentes enfoques de ensamblaje y se demostró que la combinación de un ensamblaje en SMARTdenovo o en Hinge (los cuales utilizan el algoritmo OLC) más un pulido en Nanopolish (S+N o H+N) resulta en el genoma de cloroplasto del mortiño más preciso y cercano al genoma de referencia utilizado. No obstante, para mejorar el ensamblaje se recomienda incorporar al análisis bioinformático lecturas cortas y más secuencias de referencia con el fin obtener un genoma más confiable. Este trabajo reporta el primer genoma del cloroplasto de una especie silvestre andina del Ecuador, y hace disponible un recurso genético muy útil para una variedad de estudios posteriores. Justamente, en el presente, se demostró que el flujo de trabajo elegido logró producir secuencias completas de genes (*rbcl*, *matK*, *ndhF*) con las cuales se pudo realizar un análisis filogenético del mortiño. Se recomienda aumentar la cantidad de genes analizados para establecer más claramente la filogenia de *V. floribundum*.

6 TABLAS

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

Estadística	Valor
Información Obtenida (GB)	20
Número de lecturas totales (# lecturas)	2 964 735
Número de lecturas del cloroplasto (# lecturas)	70 195
Cobertura promedio por lectura (X)	1962.388
Lectura más larga (pb)	121 871
N50 (pb)	12 710

Tabla 2. Parámetros considerados para la evaluación del genoma del cloroplasto de *Vaccinium floribundum* ensamblado en cuatro distintos programas

Ensamblador	Tamaño (pb)	NGA50/ Alineamiento más largo (pb)	Radio de duplicación	Indels por 100 000 pb	Número de caracteres genómicos
Hinge	163 686	118 689	0.995	1085.36	332
Canu	163 851	119 252	1.149	732.27	287
SMARTdenovo	164 352	119 425	0.998	599.14	332
Flye	164 439	119 447	0.998	437.09	332

Tabla 3. Parámetros considerados para la evaluación del genoma del cloroplasto de *Vaccinium floribundum* ensamblado y pulido en dos distintos programas

Programa de pulido de genomas	Ensamblador	Tamaño (pb)	NGA50/ Alineamiento más largo (pb)	Ratio de duplicación	Indels por 100 000 pb	Número de caracteres genómicos
Racon	Hinge	156 686	114 136	1.008	565.83	291
	Canu	157 878	113 649	1.002	538.24	295
	SMARTdenovo	154 807	114 352	1.003	539.42	297
	Flye	157 740	117 483	1.002	535.21	311
Nanopolish	Hinge *	165 116	120 215	1.002	491.62	332
	Canu	164 705	120 056	1.155	486.07	287
	SMARTdenovo*	164 993	120 041	1.002	438.89	332
	Flye	164 880	119 911	1.001	390.96	332

*Genomas óptimos

Tabla 4. Características de los ensamblajes óptimos de *V. floribundum* (H+N y S+N) y el genoma de referencia (*V. macrocarpon* JQ757046.1)

Parámetro	Cloroplasto de <i>V. floribundum</i> Hinge + Nanopolish (H+N)	Cloroplasto de <i>V. floribundum</i> SMARTdenovo + Nanopolish (S+N)	Cloroplasto de <i>V. macrocarpon</i> (JQ757046.1)
Tamaño (pb)	165 116	164 993	176 037
Tamaño SSC (pb)	4 524	3 270	3 029
Tamaño LSC (pb)	105 262	104 807	104 544
Tamaño IR (pb)	27 665	28 458	34 232
Número genes	115	115	105
Secuencias codificantes (CDS)	74	74	73
Genes ARNt	36	36	28
Genes ARNr	5	5	4

7 FIGURAS

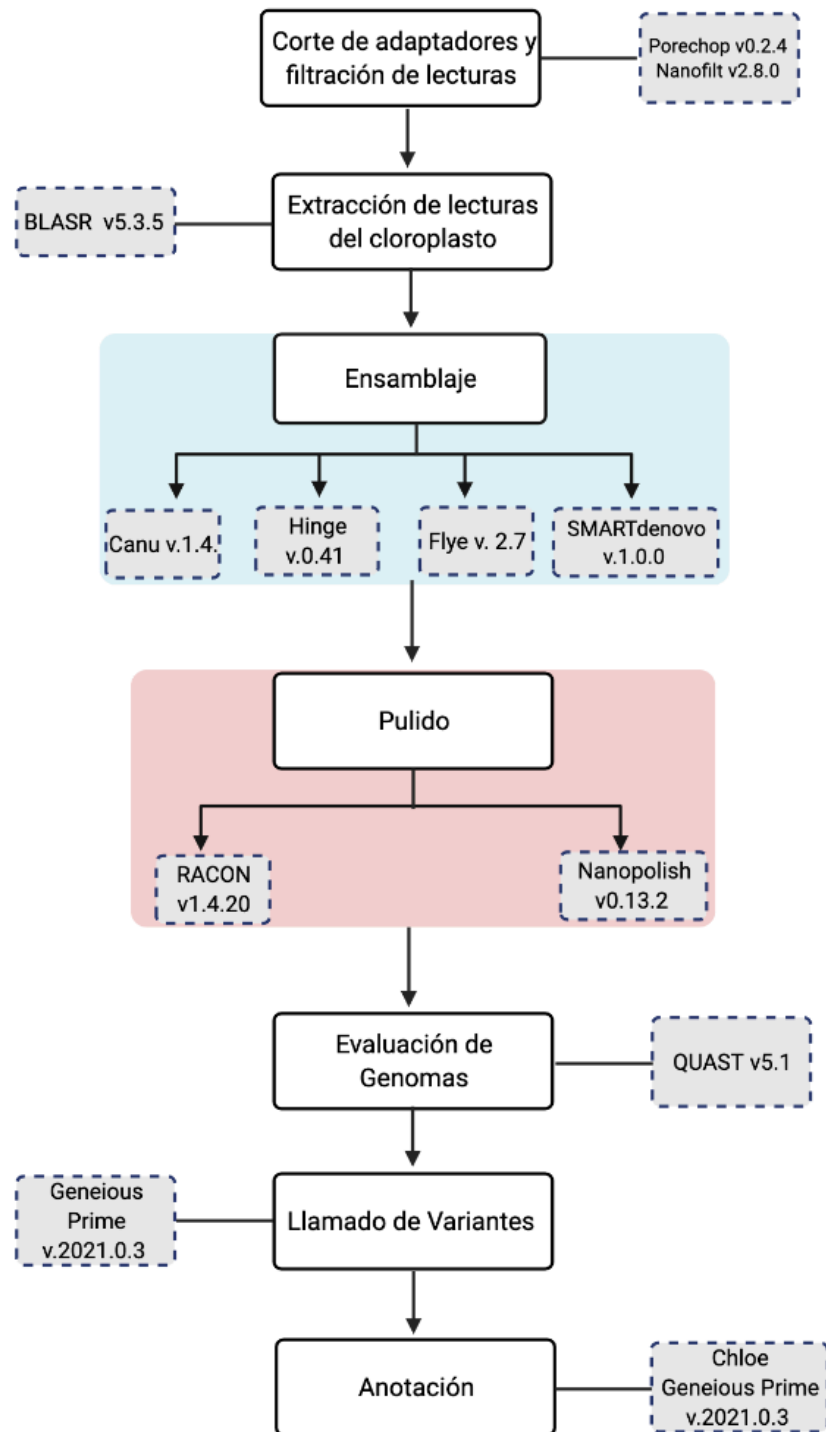


Figura 1. Flujo de trabajo en el análisis bioinformático para el ensamblaje del cloroplasto de *Vaccinium floribundum*. Esquema realizado en BioRender.

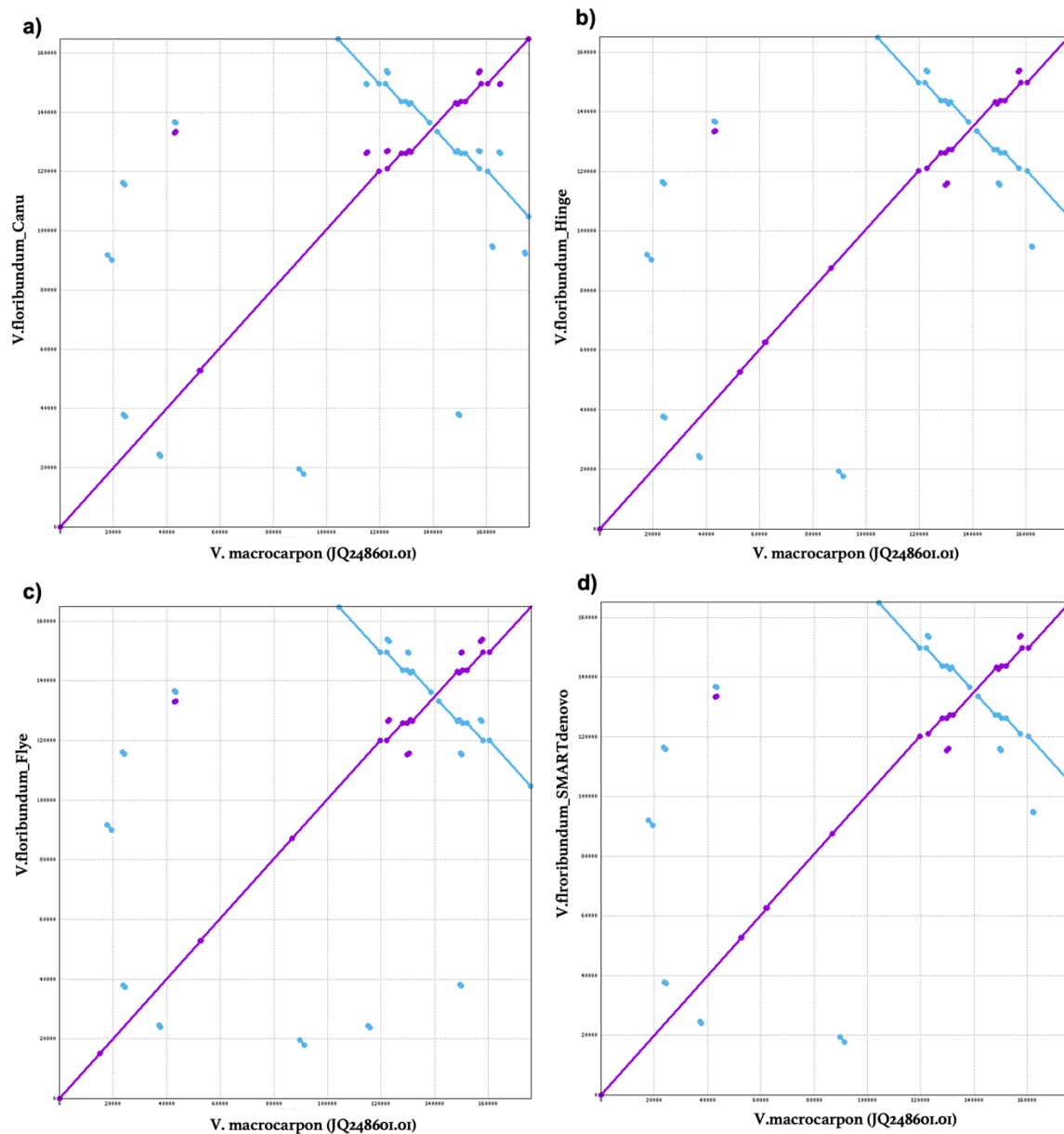


Figura 2. Alineamiento entre el genoma de referencia (cloroplasto de *V. macrocarpon* JQ757046.1) con el genoma del cloroplasto de *Vaccinium floribundum*. a) Presenta el alineamiento con el genoma obtenido en Canu, mientras que b) en Hinge, c) en Flye, d) en SMARTdenovo; y pulidos en Nanopolish. Los alineamientos en sentido *forward* se observan de color morado, mientras que los alineamientos en sentido *reverse* de color celeste.

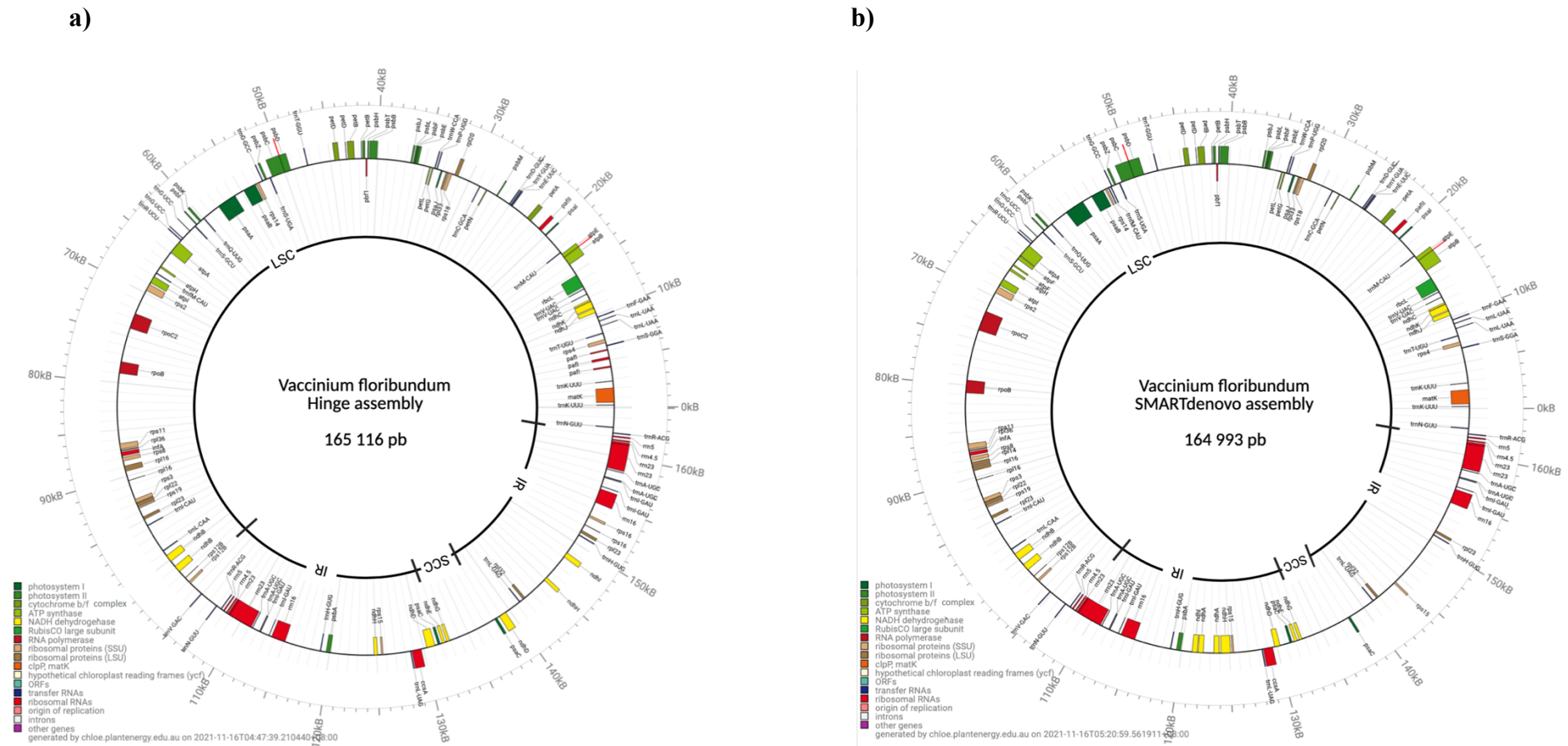
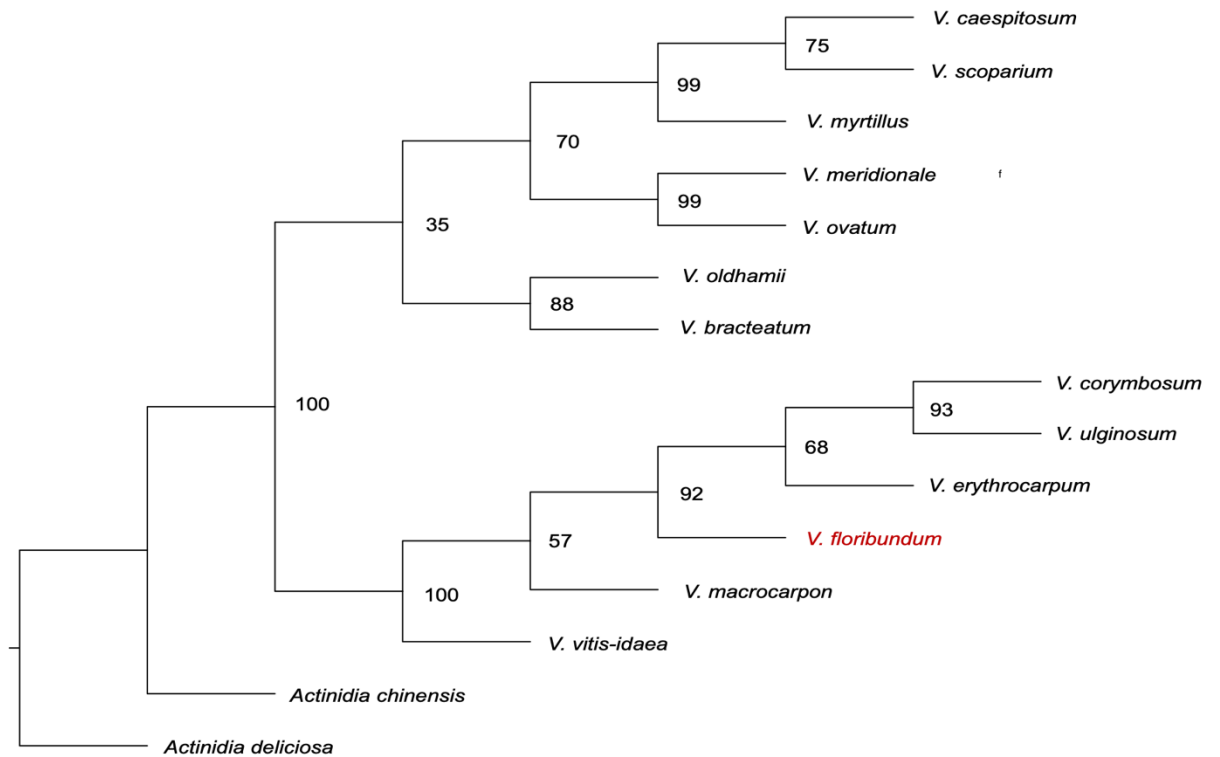


Figura 3. Estructura y mapa de los genes encontrados en el genoma del cloroplasto de *Vaccinium floribundum*. a) Presenta el genoma ensamblado en Hinge, mientras b) en SMARTdenovo; y pulidos en Nanopolish.

a)



b)

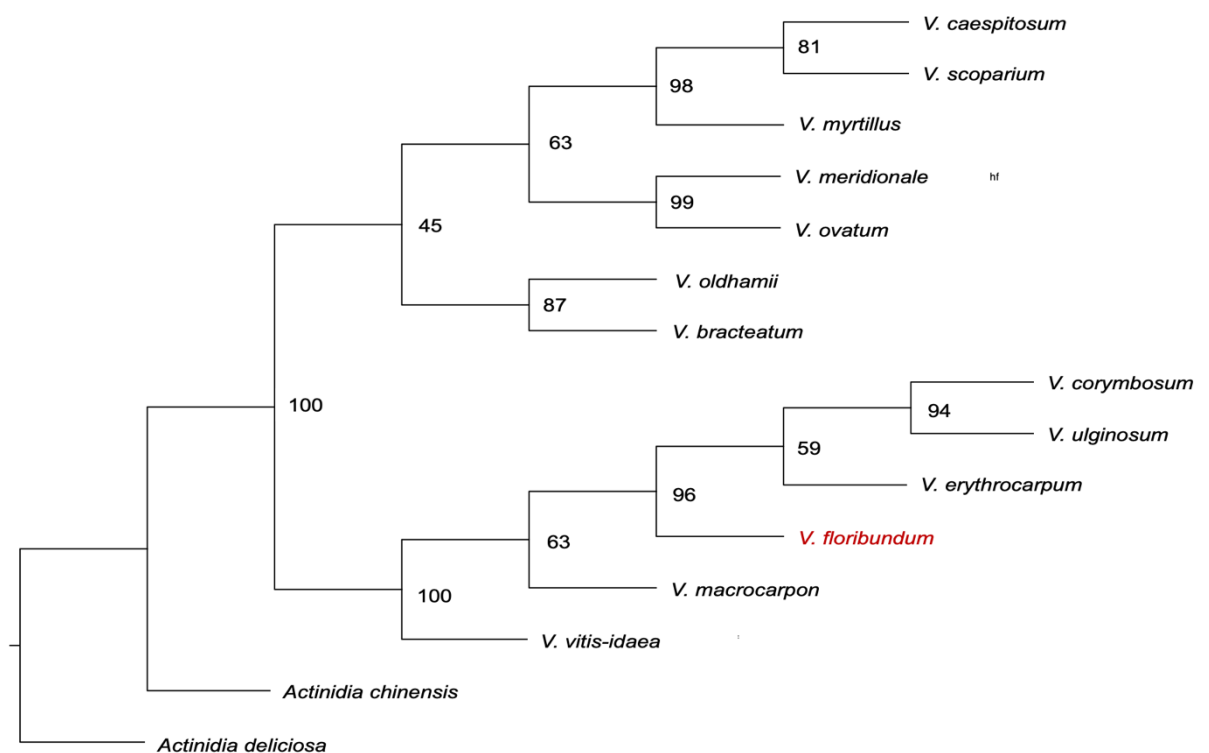


Figura 4. Análisis filogenético de *V. floribundum* utilizando tres genes representativos del cloroplasto. a) Presenta el árbol obtenido con el ensamblaje de Hinge+Nanopolish mientras b)

presenta el árbol con el ensamblaje de SMARTdenovo+Nanopolish. En cada nodo se presentan los valores de confiabilidad de la relación presentada en base a 1000 replicaciones de bootstrap.

8 REFERENCIAS

- Agah, S., Zheng, M., Pasquali, M., & Kolomeisky, A. B. (2016). DNA sequencing by nanopores: advances and challenges. *Journal of Physics D: Applied Physics*, *49*(41), 413001. doi:10.1088/0022-3727/49/41/413001
- Amar, M. H. (2020). ycf1-ndhF genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in *Prunus persica*. *Journal, Genetic Engineering & Biotechnology*, *18*(1), 42. doi:10.1186/s43141-020-00057-3
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*(1), 30. doi:10.1186/s13059-020-1935-5
- Biomatters Ltd. (2021). Geneious Prime 2021.2 User Manual. (PDF). <https://assets.geneious.com/documentation/geneious/GeneiousPrimeManual.pdf>
- Bleidorn, C. (2016). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity*, *14*(1), 1–8. doi:10.1080/14772000.2015.1099575
- Campos-de Quiroz, H. (2002). Plant genomics: an overview. *Biological Research*, *35*(3–4), 385–399. doi:10.4067/s0716-97602002000300013
- Cerón, B. (2019). Caracterización agronómica y determinación del tiempo térmico en las etapas iniciales de mortiño (*Vaccinium floribundum* Kunth) para prospección de producción forzada en la hacienda El Prado IASA I-ESPE. (Tesis Carrera Ingeniería en Ciencias Agropecuarias). <http://repositorio.espe.edu.ec/handle/21000/21020>
- Chaisson, M. J., & Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*, *13*, 238. doi:10.1186/1471-2105-13-238
- Chiluisa, P. (2015). *Vaccinium floribundum* Kunth, reserva de antocianinas en los páramos. Tesis previa a la obtención del título de Magíster en Biología de la Conservación. <http://repositorio.puce.edu.ec/handle/22000/9930>
- Chuncho, C., Chuncho, G. (2019). Páramos del Ecuador, importancia y afectaciones: Una revisión. *Bosques Latitud Cero* 9 (2). Retrieved September 20, 2021, from <https://revistas.unl.edu.ec/index.php/bosques/issue/view/61>
- 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

- Commins, J., Toft, C., & Fares, M. A. (2009). Computational biology methods and their application to the comparative genomics of endocellular symbiotic bacteria of insects. *Biological Procedures Online*, *11*, 52–78. doi:10.1007/s12575-009-9004-1
- Daniell, H., Jin, S., Zhu, X.-G., Gitzendanner, M. A., Soltis, D. E., & Soltis, P. S. (2021). Green giant—a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. *Plant Biotechnology Journal*, *19*(3), 430–447. doi:10.1111/pbi.13556
- Daniell, H., Lin, C.-S., Yu, M., & Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*, *17*(1), 134. doi:10.1186/s13059-016-1004-2
- De la Torre, L., Navarrete, H., Muriel, P., Macía, M. J., & Balslev, H. (2008). *Enciclopedia de las Plantas Útiles del Ecuador (con extracto de datos)*. Herbario QCA de la Escuela de Ciencias Biológicas de la Pontificia Universidad Católica del Ecuador & Herbario AAU del Departamento de Ciencias Biológicas de la Universidad de Aarhus.
- Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Lantz, H. (2018). Ten steps to get started in Genome Assembly and Annotation. [version 1; peer review: 2 approved]. *F1000Research*, *7*. doi:10.12688/f1000research.13598.1
- Dumschott, K., Schmidt, M. H.-W., Chawla, H. S., Snowdon, R., & Usadel, B. (2020). Oxford Nanopore sequencing: new opportunities for plant genomics? *Journal of Experimental Botany*, *71*(18), 5313–5322. doi:10.1093/jxb/eraa263
- Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), 1026–1042. doi:10.1111/eva.12178
- Fajardo, D., Senalik, D., Ames, M., Zhu, H., Steffan, S. A., Harbut, R., ... Zalapa, J. E. (2013). Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genetics & Genomes*, *9*(2), 489–498. doi:10.1007/s11295-012-0573-9
- Freudenthal, J. A., Pfaff, S., Terhoeven, N., Korte, A., Ankenbrand, M. J., & Förster, F. (2020). A systematic comparison of chloroplast genome assembly tools. *Genome Biology*, *21*(1), 254. doi:10.1186/s13059-020-02153-6
- 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. doi:10.1534/g3.119.400864
- Gao, S., Tran, Q., & Phan, V. (2019). Understand Effective Coverage by Mapped Reads using Genome Repeat Complexity. EPiC Series in Computing 60.
- Gibbs, M. (2019). De novo assembly and reconstruction of complete circular chloroplast genomes using Geneious Prime. *Application Note (PDF)*.

- Giné, D., Sánchez, R. (2015). El páramo andino: características territoriales y estado ambiental. Aportes interdisciplinarios para su conocimiento. *Estudios Geográficos* 76(278). DOI:10.3989/estgeogr.201513
- Green, M. R., & Sambrook, J. (2017). Isolation of High-Molecular-Weight DNA Using Organic Solvents. *Cold Spring Harbor Protocols*, 2017(4), pdb.prot093450. doi:10.1101/pdb.prot093450
- Guo, W., Luo, L., Huang, Y., Li, G., Wang, X., Cheng, T., Li, W. (2019). The complete chloroplast genome of *Vaccinium fragile* (Vacciniaceae), a shrub endemic to China. *Mitochondrial DNA. Part B, Resources*, 4(2), 2310–2311. doi:10.1080/23802359.2019.1627948
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075. doi:10.1093/bioinformatics/btt086
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1–8. doi:10.1016/j.ygeno.2015.11.003
- Hilu, K. W., Black, C., Diouf, D., & Burleigh, J. G. (2008). Phylogenetic signal in matK vs. trnK: a case study in early diverging eudicots (angiosperms). *Molecular Phylogenetics and Evolution*, 48(3), 1120–1130. doi:10.1016/j.ympev.2008.05.021
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36(4), 338–345. doi:10.1038/nbt.4060
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1), 239. doi:10.1186/s13059-016-1103-0
- Jones, A., Torkel, C., Stanley, D., Nasim, J., Borevitz, J., & Schwessinger, B. (2021). High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing. *Plos One*, 16(7), e0253830. doi:10.1371/journal.pone.0253830
- Jung, H., Jeon, M.-S., Hodgett, M., Waterhouse, P., & Eyun, S.-I. (2020). Comparative Evaluation of Genome Assemblers from Long-Read Sequencing for Plants and Crops. *Journal of Agricultural and Food Chemistry*, 68(29), 7670–7677. doi:10.1021/acs.jafc.0c01647
- Jurkowski, T. P. (2020). Technologies and applications for the assessment of 5-hydroxymethylcytosine. In *Epigenetics Methods* (pp. 261–278). Elsevier. doi:10.1016/B978-0-12-819414-0.00013-6
- Kahlke, T. (2020). Long-Read, long reach Bioinformatics Tutorials. https://timkahlke.github.io/LongRead_tutorials/APP_MET.html

- Kamath, G. M., Shomorony, I., Xia, F., Courtade, T. A., & Tse, D. N. (2017). HINGE: long-read assembly achieves optimal repeat resolution. *Genome Research*, 27(5), 747–756. doi:10.1101/gr.216465.116
- Kchouk, M., Gibrat, J. F., & Elloumi, M. (2017). Generations of sequencing technologies: from first to next generation. *Biology and Medicine (Aligarh)*, 09(03). doi:10.4172/0974-8369.1000395
- Kim, S.-C., Baek, S.-H., Lee, J.-W., & Hyun, H. J. (2019). Complete chloroplast genome of *Vaccinium oldhamii* and phylogenetic analysis. *Mitochondrial DNA Part B*, 4(1), 902–903. doi:10.1080/23802359.2019.1579067
- Kooij, P. W., & Pellicer, J. (2020). Genome size versus genome assemblies: are the genomes truly expanded in polyploid fungal symbionts? *Genome Biology and Evolution*, 12(12), 2384–2390.
- 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. *Genome Research*, 27(5), 722–736. doi:10.1101/gr.215087.116
- Kron, K. A., Fuller, R., Crayn, D. M., Gadek, P. A. and Quinn, C. J. (1999). Phylogenetic relationships of epacrids and vaccinioids (Ericaceae s.l.) based on *matK* sequence data. *Plants Systematics and Evolution* 218: 55–65.
- Kron, K. A., Powell, E. A., & Luteyn, J. L. (2002). Phylogenetic relationships within the blueberry tribe (Vaccinieae, Ericaceae) based on sequence data from MATK and nuclear ribosomal ITS regions, with comments on the placement of Satyria. *American Journal of Botany*, 89(2), 327–336. doi:10.3732/ajb.89.2.327
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22), 3276–3278. doi:10.1093/bioinformatics/btu531
- Li, C., Lin, F., An, D., Wang, W., & Huang, R. (2017). Genome sequencing and assembly by long reads in plants. *Genes*, 9(1). doi:10.3390/genes9010006
- Lin, Z., Zhou, P., Ma, X., Deng, Y., Liao, Z., Li, R., & Ming, R. (2020). Comparative analysis of chloroplast genomes in *Vasconcellea pubescens* A.DC. and *Carica papaya* L. *Scientific Reports*, 10(1), 15799. doi:10.1038/s41598-020-72769-y
- Lin, B., Hui, J., Mao, H. (2021). Nanopore Technology and Its Applications in Gene Sequencing. *Biosensors*, 11, 214. <https://doi.org/10.3390/bios11070214>
- Lischer, H. E. L., & Shimizu, K. K. (2017). Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*, 18(1), 474. doi:10.1186/s12859-017-1911-6

- Liu, H., Wu, S., Li, A., & Ruan, J. (2021). SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte*, 2021, 1–9. doi:10.46471/gigabyte.15
- Lluisaca, S., Manzano, P., Ruales, J., Flores, J., Mendoza, J., Peralta, E., & Cevallos-Cevallos, J. M. (2018). Chemical, antimicrobial, and molecular characterization of mortiño (*Vaccinium floribundum* Kunth) fruits and leaves. *Food Science & Nutrition*, 6(4), 934–942. doi:10.1002/fsn3.638
- Lu, H., Giordano, F., & Ning, Z. (2016). Oxford nanopore minion sequencing and genome assembly. *Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute*, 14(5), 265–279. doi:10.1016/j.gpb.2016.05.004
- Murigneux, V., Rai, S. K., Furtado, A., Bruxner, T. J. C., Tian, W., Harliwong, I., ... Henry, R. J. (2020). Comparison of long-read methods for sequencing and assembly of a plant genome. *GigaScience*, 9(12). doi:10.1093/gigascience/giaa146
- National Center for Biotechnology Information (NCBI).(1988). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information.).[Internet].Citado el 18 de noviembre del 2021 de <https://www.ncbi.nlm.nih.gov/>
- Park, J., Xi, H., & Kim, Y. (2020). The Complete Chloroplast Genome of *Arabidopsis thaliana* Isolated in Korea (Brassicaceae): An Investigation of Intraspecific Variations of the Chloroplast Genome of Korean *A. thaliana*. *International Journal of Genomics*, 2020, 3236461. doi:10.1155/2020/3236461
- Peterson, P. M., Romaschenko, K., & Johnson, G. (2010). A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Molecular Phylogenetics and Evolution*, 55(2), 580–598. doi:10.1016/j.ympev.2010.01.018
- Powell, E. A., & Kron, K. A. (2003). Molecular systematics of the northern andean blueberries (vaccinieae, vaccinioideae, ericaceae). *International Journal of Plant Sciences*, 164(6), 987–995. doi:10.1086/378653
- Ramsay, P. M., & Oxley, E. R. B. (1996). Fire temperatures and postfire plant community dynamics in Ecuadorian grass páramo. *Vegetatio*.
- Reiling, S., Chen, S., Ragoussis, I. (2020). McGill Nanopore Ligation LibPrep Protocol SQK- LSK109. McGill University. dx.doi.org/10.17504/protocols.io.bpegmjwb
- Ren, L., Huang, W., Cannon, E. K. S., Bertioli, D. J., & Cannon, S. B. (2018). A mechanism for genome size reduction following genomic rearrangements. *Frontiers in genetics*, 9, 454.
- Roldán, F. (2013). Caracterización molecular, funcional y estudio del comportamiento post cosecha del mortiño (*Vaccinium floribundum* Kunth) de la comunidad de Quinticusig del cantón Sigchos de la provincia de Cotopaxi. (Tesis ingeniería Agroindustrial). <http://bibdigital.epn.edu.ec/handle/15000/5738>

- Rono, P. C., Dong, X., Yang, J.-X., Mutie, F. M., Oulo, M. A., Malombe, I., ... Wang, Q.-F. (2020). Initial complete chloroplast genomes of alchemilla (rosaceae): comparative analysis and phylogenetic relationships. *Frontiers in Genetics, 11*, 560368. doi:10.3389/fgene.2020.560368
- Scheunert, A., Dorfner, M., Lingl, T., & Oberprieler, C. (2020). Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing. *Plos One, 15*(3), e0226234. doi:10.1371/journal.pone.0226234
- Simpson, J. (2021). Signal Analysis using Nanopolish. Oxford Nanopore Technologies. (video). <https://nanoporetech.com/resource-centre/jared-simpson-signal-analysis-using-nanopolish>
- 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. doi:10.1093/bib/bby017
- Shin, S. C., Kim, H., Lee, J. H., Kim, H.-W., Park, J., Choi, B.-S., ... Kim, S. (2019). Nanopore sequencing reads improve assembly and gene annotation of the *Parochlus steinenii* genome. *Scientific Reports, 9*(1), 5095. doi:10.1038/s41598-019-41549-8
- Sohn, J.-I., & Nam, J.-W. (2018). The present and future of de novo whole-genome assembly. *Briefings in Bioinformatics, 19*(1), 23–40. doi:10.1093/bib/bbw096
- Song, G.-Q., & Hancock, J. F. (2011). *Vaccinium*. In C. Kole (Ed.), *Wild crop relatives: genomic and breeding resources* (pp. 197–221). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-16057-8_10
- Su, Y., Kang, M., Zhang, Z., Wu, H., & Xu, X. (2019). The complete chloroplast genome of *Banara guianensis* from Ecuador. *Mitochondrial DNA Part B, 4*(1), 359–360. doi:10.1080/23802359.2018.1542988
- Tao, L., Shi, Z.-G., & Long, Q.-Y. (2021). Complete chloroplast genome sequence and phylogenetic analysis of a wild species of abiu fruit, *Pouteria caimito* (Ruiz & Pavon.) Radlk. *Mitochondrial DNA. Part B, Resources, 6*(1), 138–139. doi:10.1080/23802359.2020.1852897
- Tsutsumi, C. (2011). The Phylogenetic Positions of Four Endangered *Vaccinium* Species in Japan. *Bull. Natl. Mus. Nat. Sci.* 37 (2). https://www.kahaku.go.jp/research/publication/botany/download/37_2/BNMNS_B370204.pdf
- Trifinopoulos, L., Nguyen, A., Minh, B. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis.. *Nucleic Acids Res.*, 44:W232-W235. <https://doi.org/10.1093/nar/gkw256>
- 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. doi:10.1101/gr.214270.116

- 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, S., & Gao, L.-Z. (2016). Complete Chloroplast Genome Sequence and Annotation of the Tropical japonica Group of Asian Cultivated Rice (*Oryza sativa* L.). *Genome Announcements*, *4*(1). doi:10.1128/genomeA.01703-15
- 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. doi:10.1186/s12864-018-5348-8
- Wei, F., Tang, D., Wei, K., Qin, F., Li, L., Lin, Y., ... Miao, J. (2020). The complete chloroplast genome sequence of the medicinal plant *Sophora tonkinensis*. *Scientific Reports*, *10*(1), 12473. doi:10.1038/s41598-020-69549-z
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, *20*(1), 129. doi:10.1186/s13059-019-1727-y
- Xu, J., Shen, X., Liao, B., Xu, J., & Hou, D. (2020). Comparing and phylogenetic analysis chloroplast genome of three *Achyranthes* species. *Scientific Reports*, *10*(1), 10818. doi:10.1038/s41598-020-67679-y
- Yao, J., Zhao, F., Xu, Y., Zhao, K., Quan, H., Su, Y., ... Lan, X. (2020). Complete chloroplast genome sequencing and phylogenetic analysis of two dracocephalum plants. *BioMed Research International*, *2020*, 4374801. doi:10.1155/2020/4374801
- Zhao, C., Chen, S., Sun, K., Li, D., Meng, J., Tang, Y., & Song, S. (2019). Sequencing and characterization the complete chloroplast genome of the potato, *Solanum tuberosum* L. *Mitochondrial DNA Part B*, *4*(1), 953–954. doi:10.1080/23802359.2019.1579061
- Zhong, X. (2020). Assembly, annotation and analysis of chloroplast genomes. Tesis Doctoral para Ciencias Moleculares. <https://doi.org/10.26182/5f333d9ac2bee>

ANEXO

ANEXO 1: Especies y genes usados en el análisis filogenético de *Vaccinium floribundum*Tabla 5. Especies y códigos de accesoión de los tres genes utilizados para el análisis filogenético del cloroplasto de *Vaccinium floribundum*

Especie	Código de accesoión del gen		
	<i>matK</i>	<i>rbcL</i>	<i>ndhF</i>
<i>Vaccinium ovatum</i>	MT593012.1	KX678497.1	MT593007.1
<i>Vaccinium caespitosum</i>	MT593010.1	MG224041.1	MT593005.1
<i>Vaccinium corymbosum</i>	MZ328079.1	MZ328079.1	MZ328079.1
<i>Vaccinium uliginosum</i>	LC521968.1	LC521968.1	LC521968.1
<i>Vaccinium scoparium</i>	AF419716.1	MG222739.1	AF419766.1
<i>Vaccinium oldhamii</i>	AB623174.1	NC_042713.1	NC_042713.1
<i>Vaccinium macrocarpon</i>	JQ757046.1	L12625.2	JQ757046.1
<i>Vaccinium erythrocarpum</i>	MW006668.1	MW006668.1	MW006668.1
<i>Vaccinium myrtillus</i>	AF382810.1	KM361028.1	MN150132.1
<i>Vaccinium vitis-idaea</i>	LC521969.1	KF602162.1	LC521969.1
<i>Vaccinium meridionale</i>	U89759.1	AF124576.1	AF419756.1
<i>Vaccinium bracteatum</i>	AB623177.1	LC521967.1	LC521967.1
Grupo externo			
<i>Actinidia deliciosa</i>	JQ218264.1	NC_026691.1	NC_026691.1
<i>Actinidia chinensis</i>	AY491644.1	NC_026690.1	NC_026690.1