

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

Caracterización de polimorfismos en los genes LaCAO (*Lupinus angustifolius* copper amine oxidase) y RAP2-7 (factor de respuesta al etileno/APETALA2), y su influencia en la concentración de alcaloides en las semillas del chocho (*Lupinus mutabilis*)

Daniel Alejandro Dávila Báez

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

Caracterización de polimorfismos en los genes LaCAO (*Lupinus angustifolius* copper amine oxidase) y RAP2-7 (factor de respuesta al etileno/APETALA2), y su influencia en la concentración de alcaloides en las semillas del chocho (*Lupinus mutabilis*)

Daniel Alejandro Dávila Báez

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: Daniel Alejandro Dávila Báez

Código: 00204439

Cédula de identidad: 1723114193

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UNPUBLISHED DOCUMENT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

Lupinus mutabilis, conocido como chocho o tarwi, es una leguminosa de los valles interandinos. En la actualidad ha despertado gran interés a nivel mundial como alimento funcional por su alto contenido de proteínas y ácidos grasos. Sin embargo, sus semillas poseen un contenido considerable de alcaloides quinolizidínicos (QAs), metabolitos secundarios que pueden inducir toxicidad al ser ingeridos. Esto, ha limitado el interés en la producción a gran escala del chocho por los elevados costos asociados a la remoción de alcaloides. Hay esfuerzos de mejoramiento genético para desarrollar variedades de chocho con baja cantidad de alcaloides. En el presente estudio, se analizó la secuencia de ADN de los genes LaCAO y RAP2-7 (involucrados en la ruta de biosíntesis de alcaloides en lupinos) en 14 accesiones de chocho provenientes del Banco de germoplasma del INIAP con el objetivo de identificar polimorfismos asociados a una alta y baja producción de alcaloides. Se realizó un diseño de primers y amplificación de los genes LaCAO y RAP2-7. Se llevó a cabo un secuenciamiento por nanoporo obteniendo secuencias consenso que sirvieron para observar la presencia de polimorfismos específicos para accesiones altas o bajas en QAs. Se encontró 2 polimorfismos específicos para accesiones con bajo contenido de QAs en el gen RAP2-7. Mientras que en el caso del gen LaCAO, no se encontró polimorfismos que sean específicos para accesiones con bajo contenido de QAs. Estos resultados son interesantes para el desarrollo de futuros proyectos enfocados en la funcionalidad de marcadores moleculares dirigidos al mejoramiento genético del chocho.

Palabras clave: *Lupinus mutabilis*, LaCAO, RAP2-7, alcaloides quinolizidínicos, secuenciamiento por nanoporo

ABSTRACT

Lupinus mutabilis, known as chocho or tarwi, is a legume from the inter-Andean valleys. Currently, it has raised great interest worldwide as a functional food due to its high protein and fatty acid content. However, its seeds have a considerable content of quinolizidine alkaloids (QAs), secondary metabolites that can induce toxicity when ingested. This has limited interest in the large-scale production of chocho due to the high costs associated with the removal of alkaloids. For this reason, there are breeding efforts to develop lupine varieties with low amount of alkaloids. In the present study, the DNA sequence of the LaCAO and RAP2-7 genes (involved in the alkaloid biosynthesis pathway in lupines) was analyzed in 14 lupine accessions from the INIAP Germplasm Bank with the aim of identifying polymorphisms associated with high and low alkaloid production. A primer design and amplification of the LaCAO and RAP2-7 genes were carried out. Nanopore sequencing was performed for obtaining consensus sequences that were used to observe the presence of specific polymorphisms for accessions high or low in QAs. Two specific polymorphisms were found for accessions with low content of QAs in RAP2-7. While, in the case of the LaCAO gene, no polymorphisms were found that are specific for accessions with low content of QAs. These results are interesting for the development of future projects focused on the functionality of molecular markers aimed at the genetic improvement of lupine.

Key words: *Lupinus mutabilis*, LaCAO, RAP2-7, quinolizidine alkaloids, *Oxford Nanopore Sequencing*

TABLA DE CONTENIDO

1. INTRODUCCIÓN	12
2. METODOLOGIA.....	16
_2.1 Material Vegetal	16
_2.2 Extracción y Cuantificación de Alcaloides Quinolizidínicos (QAs).....	16
_2.3 Diseño de Primers para la Amplificación de los Genes LaCAO y RAP2-7.....	16
_2.4 Amplificación de los Genes LaCAO y RAP2-7	17
_2.5 Preparación de Librerías y Secuenciación de Amplicones LaCAO y RAP2-7 mediante Nanopore Sequencing.....	17
_2.6 Secuenciación de Amplicones RAP2-7 mediante Sanger Sequencing	18
_2.7 Procesamiento de las secuencias obtenidas	18
_2.8 Mapeo de Lecturas e Identificación de Polimorfismos	18
_2.9 Comparación de Secuencias Obtenidas por Nanopore y Sanger.....	18
3. RESULTADOS	19
_3.1 Cuantificación de Alcaloides QAs	19
_3.2 Diseño de Primers y Amplificación de Genes LaCAO y RAP2-7	19
_3.3 Secuenciación y Mapeo de Lecturas LaCAO y RAP2-7 tras Mapeo.....	19
_3.4 Identificación de Polimorfismos.....	20
_3.5 Comparación de las Secuencias Consenso de RAP2-7 entre la Secuenciación por Oxford Nanopore y Sanger.....	21
4. DISCUSIÓN.....	22
_4.1 Cuantificación de alcaloides	22

_4.2	Diseño de Primers y Amplificación.....	22
_4.3	Secuenciación	23
_4.4	Mapeo y generación de secuencias consenso	23
_4.5	Detección de Polimorfismos.....	24
_4.6	Comparación entre secuencias ONT y Sanger sequencing	24
5.	CONCLUSIONES	26
6.	TABLAS.....	27
7.	FIGURAS	30
8.	REFERENCIAS BIBLIOGRÁFICAS	37
9.	ANEXOS	43

ÍNDICE DE TABLAS

Tabla 1. Porcentaje de alcaloides promedio de accesiones bajas en alcaloides (PROINPA P5 y PROINPA P8) y altas en alcaloides (FR9 P3 y FR11 P8) obtenidas mediante una titulación ácido base con NaOH y fenolftaleína por triplicado.	27
Tabla 2. Primers propuestos para la amplificación de los genes LaCAO_ fragmento 1, LaCAO _ fragmento 2, LaCAO _ fragmento 3 y RAP2-7 obtenidos en Pimer3.	27
Tabla 3. Resultados generales de la secuenciación del pool de 7 genes en las 14 accesiones obtenidas por el software MinION™.	27
Tabla 4. Cobertura tras el mapeo de los genes RAP2-7 y LaCAO en las 14 accesiones analizadas.....	28

ÍNDICE DE FIGURAS

Figura 1. Amplificación de los genes RAP2-7, LaCAO _ fragmento 1, LaCAO _ fragmento 2 y LaCAO _ fragmento 3 en las accesiones 176, 188, 190, 192, 193, 707, 707, 712, 723 y 742.	30
Figura 2. Polimorfismos en RAP2-7 de las 14 accesiones.....	32
Figura 3. Polimorfismos en LaCAO de las 14 accesiones.	33
Figura 4. Discrepancias entre secuencias consenso obtenidas por la secuenciación de Oxford Nanopore y Sanger	34
Figura 5. Árboles filogenéticos de las secuencias consenso obtenidas por Oxford Nanopore y Sanger.	35

ÍNDICE DE ANEXOS

Anexo 1: Cuantificación de alcaloides quinolizidínicos en accesiones 176, 188, 190, 192, 193, 706, 707, 712, 723 y 742.	43
Anexo 2: Tamaño, calidad, número de reads y porcentaje de recuperación de la secuenciación del pool de 7 genes en las 14 accesiones analizadas	44
Anexo 3: Gráfico de barras de cobertura de los genes RAP2-7 y LaCAO en las 14 accesiones analizadas.....	46

1. INTRODUCCIÓN

Lupinus mutabilis, comúnmente conocido como chocho o tarwi, es una leguminosa de origen andino que crece en un rango altitudinal de 1.500 a 3800 msnm, principalmente en los valles interandinos de Ecuador, Perú y Bolivia (Chirinos-Arias, 2015). Su interés en la industria alimenticia se debe al alto valor nutricional de sus semillas. Entre sus componentes se encuentran ácidos grasos esenciales (ácido oleico y ácido linoleico) y aminos bio-activos (leucina, lisina, e isoleucina) con varias propiedades de defensa inmunológica, anti-cancerígenas y anti-hipertensivas (Chirinos-Arias, 2015). Además, se conoce que la cantidad de proteínas varía del 41 al 51% de su peso seco, siendo superior al contenido proteico de otros cultivos de interés alimenticio como la soya (Jacobsen & Mujica, 2006). Sin embargo, hay una gran prevalencia de otros componentes anti-nutricionales en las semillas del chocho. Se ha identificado la presencia de alcaloides, ácidos fítics, e inhibidores de tripsina que ejercen efectos fisiológicos indeseables y toxicidad anti-colinérgica (Villacrés et al., 2020).

El componente anti-nutricional más destacable en las semillas del chocho son los alcaloides quinolizidínicos (QAs). Estos son metabolitos secundarios derivados de lisina sintetizados en las partes aéreas de varias especies del género *Lupinus*, y acumuladas en las semillas maduras (Kroc et al., 2019). Su presencia aporta un sabor amargo en las semillas, y funciona como mecanismo de defensa contra herbívoros y patógenos (Frick et al., 2017). Se ha reportado que en general, el porcentaje de alcaloides presentes en las semillas de *L. mutabilis* se encuentra en un rango entre 2.56% y 4.14%; siendo la lupanina, esparteína, y 1,3-hidroxilupanina los más prevalentes (Ortega-David et al., 2010).

Para un consumo seguro, se busca llegar a una concentración menor al 0.02% de alcaloides (Frick et al., 2017), el mismo que se logra mediante diferentes tipos de

procesamientos a nivel industrial; que incluyen remojo o deslupinación, mejoramiento agrícola y/o fermentación (Villacrés et al., 2020). La deslupinación es la metodología más utilizada; y consiste en una hidratación por 12 horas y cocción a presión de las semillas (Jacobsen & Mujica, 2006). Debido a este tedioso procesamiento, la producción industrial del chocho se ha visto limitada (Villacrés et al., 2009). Por ende, es importante el desarrollo de alternativas que potencien la producción del chocho, como lo es el mejoramiento genético y desarrollo de variedades con bajo o nulo contenido de alcaloides.

Se han realizado varios estudios a nivel de expresión genética en otras especies del género *Lupinus* (como *L. angustifolius* y *L. albus*) con el fin de determinar los principales genes asociados a la producción de alcaloides (Kroc et al., 2019). Se han identificado genes cuya expresión difiere de forma significativa entre accesiones amargas (alto contenido de alcaloides) y dulces (bajo contenido de alcaloides) de *L. angustifolius*. Los mismos se han categorizado en genes relacionados al metabolismo secundario, factores de transcripción del metabolismo secundario, y genes relacionados a la biosíntesis de L-lisina (Kroc et al., 2019). Para fines de esta investigación, se analizará los genes LaCAO y RAP2-7.

LaCAO (*L. angustifolius*-copper amine oxidase) es un gen relacionado al metabolismo secundario en especies del género *Lupinus*, cuya proteína, ubicada en los peroxisomas, está involucrada en la catálisis de poli-aminas referentes a procesos de muerte celular, reestructuración de la pared celular, y desarrollo (Tavladoraki et al., 2016). En la síntesis de QAs, se involucra en la formación de la molécula 1-piperidina a partir de cadaverina, durante primeros pasos de la biosíntesis de la esparteína (el QAs más simple y del que derivan otros QAs). En *L. angustifolius*, este gen se expresa en tejidos como tallo,

hojas y pedicelos, y los alcaloides resultantes del proceso de biosíntesis, son transportados a la semilla mediante el floema (Yang et al., 2017).

RAP2-7 (factor de respuesta al etileno/APETALA2) es un factor de transcripción relacionado al metabolismo secundario de especies de *Lupinus*. Se segrega con el locus *iucundus*, un QTL involucrado en la biosíntesis de alcaloides (Czepiel et al., 2021). Su expresión en órganos aéreos varía entre accesiones amargas (alta expresión) y dulces (baja expresión); excepto en el tallo y vainas, donde existe una alta expresión en accesiones dulces de *L. angustifolius* (Czepiel et al., 2021).

La identificación de polimorfismos es una forma útil en la determinación de alelos específicos que aportan en algún fenotipo de interés (Morgil, Can Gercek & Tulum, 2020), como es el caso del rasgo cuantitativo de concentración de alcaloides. El desarrollo de marcadores moleculares permite la detección de dichos alelos, y para ello utiliza técnicas de hibridación a zonas polimórficas para una identificación molecular (Morgil, Can Gercek & Tulum, 2020). Para lograr identificar los polimorfismos a nivel de secuencia de ADN, una herramienta muy efectiva es la secuenciación. En este sentido, el secuenciamiento por nanoporo, desarrollado por *Oxford Nanopore Technologies* es un método de secuenciación de tercera generación que emplea una membrana con nanoporos, la cual es atravesada por la molécula de ADN y la secuencia es decodificada en tiempo real. (Slatko, Gardner & Ausubel, 2018). Este método ofrece varias ventajas frente a otros métodos de secuenciación como Sanger o Illumina al ser más económico y enfocado a lecturas largas con un tamaño superior a 1.000 bp (Petersen et al., 2019). De igual forma, esta tecnología comprende de sistemas *multiplexing*, que consiste en el uso de barcodes para el etiquetado de diferentes ciertas muestras de DNA, con el objetivo de ser analizadas juntas dentro de una misma corrida (Freed et al., 2020).

El objetivo de esta investigación fue determinar la presencia de polimorfismos específicos en los genes LaCAO y RAP2-7 en accesiones de *Lupinus mutabilis* con alto y bajo contenido de alcaloides. Para ello, se cuantificó alcaloides en accesiones provenientes del Banco de Germoplasma del INIAP, con el fin de definir las accesiones a analizar. Se realizó un diseño de primers y amplificación de estos genes en las accesiones seleccionadas. Posteriormente, estos genes fueron analizados mediante secuenciación por nanoporo y se realizó un procesamiento de lecturas para obtener las secuencias consenso correspondientes a cada accesión. Finalmente, se comparó dichas secuencias para la identificación de polimorfismos específicos para accesiones altas y bajas en concentración de alcaloides. Los resultados pueden ser útiles para la identificación de marcadores moleculares que puedan participar en programas de mejoramiento genético del chocho para promover el desarrollo de variedades de chocho con bajo contenido de alcaloides.

2. METODOLOGIA

2.1 Material Vegetal

Para los ensayos de cuantificación de alcaloides, se analizó 4 accesiones provistas por el Banco de germoplasma del INIAP: 2 catalogadas como amargas: FR9-P3 (Flor rosada 9, planta 3), y FR11-P8 (Flor rosada 11, planta 8); y 2 catalogadas como dulces: PR-P5 (PROINPA, planta 5) y PR-P8 (PROINPA, planta 8).

Para los ensayos de amplificación y secuenciación de los genes RAP2-7 y LaCAO, se agregó 10 accesiones seleccionadas por un estudio previo del Laboratorio de Biotecnología Vegetal: 5 con alto contenido de alcaloides (4.65% y 4.91%), denominadas accesiones amargas: 176, 188, 190, 192, 193; y 5 con bajo contenido de alcaloides (2.84%-3.61%) denominadas accesiones dulces: 706, 707, 712, 723, 742.

2.2 Extracción y Cuantificación de Alcaloides Quinolizidínicos (QAs)

Para la extracción de alcaloides presentes en las semillas de las muestras FR9-P3, FR11-P8, PR-P5, y PR-P8, se realizó una titulación ácido-base utilizando fenolftaleína como indicador y NaOH 0,1N como neutralizador. Se midió el volumen de NaOH requerido para causar el viraje de la solución de incoloro a rosado. Se utilizó la siguiente ecuación:

$$\% \text{ Alcaloides} = \frac{\text{Normalidad NaOH (0.1N)} * \text{Volumen NaOH} * \text{Peso equivalente de lupanina (24,8 g)}}{\text{Peso de muestra (0,2 g)}}$$

2.3 Diseño de Primers para la Amplificación de los Genes LaCAO y RAP2-7

Se hizo una búsqueda por homologías de los genes reportados en *L. angustifolius* (RAP2-7: MK834264.1; LaCAO: MF152953.1) con secuencias del transcriptoma de *L. mutabilis* (SRR3423049, SRR3423050, SRR3423051). Las secuencias homólogas encontradas fueron usadas para el diseño de primers mediante *Primer 3 v.4.0*. Se usó el software *UNAFold* para seleccionar los primers con: %GC: 40-60; hairpin: 0; $\Delta G > -6$.

LaCAO, al tener un tamaño aproximado de 6.8 Kb, se lo dividió en 3 fragmentos solapantes: fragmento 1 (2336 bp), fragmento 2 (2794 bp) y fragmento 3 (1669 pb). Se generaron primers para amplificar cada uno de los 3 fragmentos de manera independiente. Para el gen RAP2-7, de tamaño aproximado de 3 Kb, se diseñó un único par de primers, con el objetivo de amplificar la totalidad del gen en un solo fragmento.

2.4 Amplificación de los Genes LaCAO y RAP2-7

Las condiciones de amplificación de los 3 fragmentos de LaCAO fueron: Buffer 1X, MgCl₂ 3.0mM, dNTPs 0.3mM, primers 0.5uM y 1U de Taq Platinum (Invitrogen). Las condiciones de ciclado fueron: Denaturación inicial a 94 °C/5 min., 35 ciclos de denaturación a 94 °C/30 seg., annealing a 55 °C/30 seg. y extensión a 72 °C/3 min., y extensión final a 72 °C/5 min. Las condiciones de amplificación de RAP2-7 fueron: Buffer 1X, MgCl₂ 2.0mM, dNTPs 0.3mM, primers 0.3uM y 1 U de Taq Platinum (Invitrogen). Las condiciones de ciclado fueron las mismas que los fragmentos del gen LaCAO. Los amplicones fueron almacenados a -20 °C hasta el momento de la secuenciación.

2.5 Secuenciación de Amplicones LaCAO y RAP2-7 mediante Nanopore Sequencing

Para la preparación de librerías de secuenciamiento, se utilizó el protocolo *Ligation sequencing amplicons - native barcoding*, disponible en la página web de ONT (<https://nanoporetech.com/community>) que utiliza los kits SQK-LSK109 EXP-NBD104, EXP-NBD114. Se midió la cantidad de ADN presente en la librería mediante fluorometría utilizando el Qubit 3.0. El pool de ADN fue cargado en una *flow cell* R.9.4.1 instalada dentro del instrumento *MinION™ Mk1C*, y la adquisición de datos se realizó mediante el software *MinKNOW™ v.21.10.8* utilizando el algoritmo de *high-accuracy basecalling*.

2.6 Secuenciación de Amplicones RAP2-7 mediante Sanger Sequencing

Para comparar las secuencias obtenidas del secuenciamiento por nanoporo, con secuencias de tecnologías tradicionales como Sanger, se envió amplicones del gen RAP2-7 para su secuenciamiento en la empresa Macrogen, Corea, utilizando el servicio *Standard Sequencing*. Se seleccionó las accesiones 190 y 707 como representantes de muestras altas en alcaloides y bajas en alcaloides, respectivamente.

2.7 Procesamiento de las secuencias obtenidas

El demultiplexing fue realizado por el equipo *MinION™ Mk1C* y el software *Guppy* v6.0.1. Para la remoción de adaptadores se utilizó el software *Porechop* v.0.2.4 (Wick et al., 2018). Se realizó un filtrado por calidad y tamaño con el software *BBmap* (Joint Genome Institute, 2020). Las lecturas con calidad $q < 10$, y tamaños < 200 pb fueron removidas.

2.8 Mapeo de Lecturas e Identificación de Polimorfismos

Para aislar las secuencias de LaCAO y RAP2-7 del pool total de secuencias, se utilizó la estrategia de separación por mapeo. Se empleó la herramienta *Map to reference* del software *Geneious prime*, usando como referencia la secuencia de los genes en *L. angustifolius*. Para la identificación de polimorfismos asociados a un fenotipo amargo o dulce, se alineó y comparó cada secuencia de cada accesión, con el software *Mega X*.

2.9 Comparación de Secuencias Obtenidas por Nanopore y Sanger

Para determinar la calidad de las secuencias se hizo una comparación entre la secuencia del gen RAP2-7 del secuenciamiento por Nanopore y la obtenida por Sanger. Para determinar discrepancias se realizó un alineamiento con el software *Mega X*. Se identificó discrepancias de nucleótidos entre ambos métodos y se hizo un árbol filogenético en *MEGAX* con la opción *Phylogeny* y el algoritmo *Maximum Likelihood* (Hall, 2013).

3. RESULTADOS

3.1 Cuantificación de Alcaloides QAs

Se logró cuantificar los QAs de las semillas de las accesiones FR9-P3, FR11-P8, PROINPA-P5, y PROINPA-P8 por triplicado. Los valores promedio obtenidos fueron: 4,55%, 4,57%, 3,47% y 3,85%, respectivamente. Los resultados, con su respectiva desviación estándar Stv. Dev., se presentan en la Tabla 1.

3.2 Diseño de Primers y Amplificación de Genes LaCAO y RAP2-7

Los primers diseñados para la amplificación de los 3 fragmentos del gen LaCAO y el gen RAP2-7 se presentan en la Tabla 2. Todos los primers cumplen con las condiciones descritas en la sección 2.4. El ΔG es generalmente positivo, a excepción del primer forward de RAP2-7 (-1.15) y el primer forward de LaCAO_ fragmento 1 (-0.72).

Respecto a los resultados de amplificación, todos los primers diseñados generaron productos en el tamaño esperado: LaCAO_fragmento 1 (~2 Kb), LaCAO_fragmento 2 (~2 Kb), LaCAO_fragmento 3 (~1.5 Kb) y RAP2-7 (~3 Kb) (Figura 1). Para los fragmentos 2 y 3 del gen LaCAO y el gen RAP2-7 se obtuvo bandas claras e intensas. En el caso del fragmento 1 del gen LaCAO se observó bandas tenues en el tamaño esperado; sin embargo, también se observó bandas inespecíficas de ~500 bp en todas las accesiones, y una banda inespecífica de 1.5-2.0 Kb en la accesión 190.

3.3 Secuenciación y Mapeo de Lecturas LaCAO y RAP2-7 tras Mapeo

La cantidad total de ADN determinada en la librería mediante fluorometría previo a la secuenciación fue de 101.20 ng (65,51 fmol). Este valor se encuentra dentro del rango de 50 a 100 fmol que recomienda el protocolo *Native Barcoding Amplicons*. Los resultados generales de la secuenciación y basecalling se observan en la Tabla 3. Se generó un total de

4.23 Gb de información, para un total 6.21 millones de lecturas con una calidad promedio de 12.9 Phred.

Respecto a la cantidad de información recuperada luego de aplicar los filtros de calidad y tamaño, el Anexo 2 muestra el porcentaje de recuperación de información, la calidad y tamaño promedio de las lecturas obtenidas para cada barcode antes y después del procesamiento de los datos. El porcentaje de recuperación promedio de las lecturas es del 79%.

Los resultados obtenidos luego del mapeo en *Geneious Prime* para la recuperación de las lecturas de los genes LaCAO y RAP2-7 se muestran en la Tabla 4. En RAP2-7, la cobertura más baja corresponde a la accesión PROINPA P8 (bajo en alcaloides) con 23 (0,023X) y la accesión con más cobertura es la 712 con 14,775 (15X); mientras que en el caso de LaCAO, la accesión con más baja cobertura es la 742 con 15,405 (15X), y la accesión con más cobertura es la FR9 P3 con 55,918 (56X). En el Anexo 3 se observa un gráfico de barras sobre la cobertura de ambos genes. En general, LaCAO alcanza una mayor cobertura.

3.4 Identificación de Polimorfismos

En la Figura 2 se presentan los polimorfismos vistos en el gen RAP2-7. Se observó varios SNPs y deleciones no específicas para accesiones con bajo y alto contenido de alcaloides. Se encontró que las accesiones 192 y 193 tuvieron un patrón de deleción de citosina, adenina, guanina, adenina, y guanina en los nucleótidos 1.623, 1.624, 1.629, 1.638 y 1.639, respectivamente. Finalmente, se identificó una deleción de citosina en el nucleótido 8 de 3 accesiones bajas en alcaloides (706, 707, y 742), y una deleción de citosina en el nucleótido 1528 en 5 accesiones bajas en alcaloides (707, 712, 723, 742, y PROINPA P5).

Por otro lado, en la Figura 3 se observan los polimorfismos asociados al gen LaCAO. Gran parte del gen no presenta polimorfismos entre todas las accesiones analizadas. Se encontró muchos polimorfismos asociados a la accesión 188 en los nucleótidos 4898, 4900, 4905, 4907 y 4913 (guanina a adenina, guanina a citosina, adenina a citosina, guanina a adenina, y guanina a adenina, respectivamente), además de una inserción en la posición 5341-5349. Finalmente, se identificó la presencia de polimorfismos específicos en el gen LaCAO comunes en las accesiones PROINPA P5, PROINPA P8, FR9 P3, FR11 P8, y 712.

3.5 Comparación de las Secuencias Consenso de RAP2-7 entre la Secuenciación por Oxford Nanopore y Sanger

La secuenciación de RAP2-7 por *Sanger* en la accesión 190 generó un archivo de 1.326 pb producto del primer forward 1 y otra secuencia de 1382 del primer reverse. Mientras que en la accesión 707 se generó un archivo de 1374 nt del primer forward y otro de 1401 nt del primer reverse. Los resultados de la alineación en *MEGAX* se observan en la Figura 4. Se encontró discrepancias entre las secuencias analizadas. Las mismas fueron comunes en regiones homopoliméricas. De igual forma, el árbol filogenético que compara las secuencias consenso generadas por *Sanger* y *Oxford Nanopore* se encuentra en la Figura 5. En el caso de las accesiones bajas en alcaloides, se observa que las secuencias de las accesiones 723 y 742 se encuentran más relacionadas mediante una longitud de rama de 0,06 (6 sustituciones por cada 100 nucleótidos). Además, la secuencia 707 obtenida de *Sanger* no es igual a la secuencia 707 obtenida de *Oxford Nanopore*; incluso, la misma tiene 2 sustituciones por cada 100 nucleótidos (0,02 de longitud de rama) al compararla con el grupo de secuencias 706, 707, 712, 723 y 742 de *Oxford Nanopore*. Por otra parte, la secuencia 190 de *Sanger* está más relacionada a la accesión 188 que a la accesión 190 generada por *Oxford Nanopore*.

4. DISCUSIÓN

4.1 Cuantificación de alcaloides

El porcentaje de alcaloides en las accesiones amargas FR9 P3 y FR11 P8, fue 4,55% y 4,57%, respectivamente. Mientras que, las accesiones dulces PROINPA P5/P8 tuvieron una concentración de 3,47% y 3,85%, respectivamente. Esto no coincide con los resultados planteados por la literatura. Es sabido que las semillas manifiestan una percepción dulce o carente de amargor cuando alcanzan una concentración inferior al 0,1% de alcaloides (Gutiérrez et al., 2016), difiriendo con lo obtenido en las accesiones PROINPA P5/P8. Una posible explicación sería la presencia de posibles compuestos ácidos de las semillas como equivalentes del ácido gálico (Ruiz et al., 2019). Estos compuestos podrían disminuir el pH, generando una neutralización no dirigida a alcaloides durante la titulación ácido-base. Esta falta de precisión sugiere que los resultados no sean válidos. Con el fin de solucionar la falta de precisión en la neutralización, se puede proponer el uso de herramientas como la cromatografía de gas líquido acoplado a espectrometría de masas para futuros proyectos (Fadhil et al., 2007).

4.2 Diseño de Primers y Amplificación

Los resultados de la PCR mostraron que todos los primers amplificaron bandas en el tamaño esperado para los 3 fragmentos del gen LaCAO y el gen RAP2-7. Sin embargo, el gen LaCAO_ fragmento 1 presentó bandas tenues. Esto se puede deber a una baja complementariedad de los primers con la secuencia del ADN genómico que podría reducir la eficiencia del anclaje de la Taq polimerasa. De igual forma, es posible que las secuencias de los primers presenten cierto grado de auto-complementariedad de acuerdo a lo observado en el análisis con el software *UNAFold*. El análisis indicó que la secuencia del primer forward

del fragmento 1 presenta una energía espontánea ΔG de -0,71 (UNAFold, 2013). Este valor de ΔG negativo podría formar acoplamiento por puentes de hidrógeno dentro de la secuencia del primer, generando estructuras secundarias que reducen la eficiencia del anclaje a la secuencia target y, por tanto, una menor tasa de amplificación (Rychlik, 1993).

4.3 Secuenciación

En la Tabla 3 se observa que se generó alrededor de 4,23 Gb y un valor Phred de 12,9. Anteriores estudios basados en la secuenciación de amplicones por *Oxford Nanopore* mostraron que un valor Phred de 12,2 puede alcanzar hasta un 94% de precisión y una disminución en la aparición de errores en la secuenciación (Tabata et al., 2022). Con ello, se infiere que las lecturas generadas fueron menos susceptibles a errores de la secuenciación.

4.4 Mapeo y generación de secuencias consenso

En el caso de RAP2-7, se obtuvo una cobertura promedio de 8.760 X. LaCAO_ fragmento 1 tuvo una cobertura promedio de 13X, LaCAO_ fragmento 2 una cobertura de 8.520X, y LaCAO_ fragmento 3 una cobertura de 21.537 X. Se ha reportado que valores de cobertura en el rango 40X a 139X inducen una tasa de error máxima de 0,008% en las secuencias obtenidas de mediante *Oxford Nanopore* (Whitford et al., 2021). Mientras que una cobertura de 15 X manifiesta una tasa de error mayor al 0,042% y una inducción de quimeras del 0,009% (Karst et al., 2021). Por lo tanto, los altos valores de cobertura de RAP2-7, LaCAO_ fragmento 2/3 indican que las secuencias consenso presentarían tasas de error muy bajas. Sin embargo, el caso de LaCAO_ fragmento 1 sugiere la posibilidad de generar secuencias consenso con mayor porcentaje de error y formación de quimeras. Debido a los bajos niveles de cobertura, los resultados de este fragmento fueron descartados. Las lecturas obtenidas en *Geneious Prime* indican una alta presencia de adeninas y timinas en el

fragmento 1 de LaCAO. Las regiones homopoliméricas son secuencias de DNA de bases idénticas (Ivády et al., 2018), y presentan problemas en la identificación de polimorfismos debido a la dificultad de *Oxford Nanopore* en determinar si una o muchas bases repetidas pasan por el nanoporo (Yang et al., 2012), generando así posibles regiones erróneas que se suelen manifestar con baja cobertura en el mapeo (Menegon et al., 2017).

4.5 Detección de Polimorfismos

No se logró encontrar polimorfismos asociados a accesiones con bajo o alto contenido de alcaloides en el gen LaCAO. La elevada cobertura de los fragmentos 2 y 3 del gen LaCAO permitió identificar ciertas mutaciones propias para cada secuencia individualmente.

Curiosamente, al analizar las secuencias de las muestras del grupo PROINPA y Flor Rosada 9 y 11, se pudo detectar una inserción de 13 nucleótidos en la posición 3441. La presencia de polimorfismos específicos es un indicativo de un origen común de las accesiones, que diferencian estas 4 accesiones de las demás (Lucht, 2015).

Con respecto al gen RAP2-7 se logró detectar polimorfismos específicos para accesiones bajas en alcaloides (Figura 2). La detección de estos polimorfismos en accesiones dulces podría sugerir que estos alelos estarían relacionados a un fenotipo dulce como se ha planteado en anteriores estudios en *L. angustifolius* (Kroc et al., 2019). Se ha reportado en *L. angustifolius* que los alelos de fenotipo amargo y dulce son sustituciones no sinónimas que cambian una serina (alelo amargo) a arginato (alelo dulce). Esto propone futuros enfoques que relacionen los polimorfismos encontrados con cambios en la secuencia de aminoácidos.

4.6 Comparación entre secuencias ONT y Sanger sequencing

Se observó que las secuencias obtenidas para las accesiones 707 y 190 mediante secuenciamiento por nanoporo y por secuenciamiento Sanger difieren mucho entre sí. En la

Figura 4 se presentó discrepancias entre las secuencias consenso de ambas secuenciaciones en regiones homopoliméricas. Esto es un indicativo de la limitación de *Oxford Nanopore* para determinar polimorfismos en estas regiones, ya que es posible que dichas discrepancias no sean producto de mutaciones sino por errores en la secuenciación (Slatko et al., 2018). Posibles soluciones son el uso de primers que excluyan áreas homopoliméricas durante la amplificación (Applied Biosystems, 2016), uso de ensamblados *de novo* a partir de referencias obtenidas de secuenciamiento en *Sanger* (Menegon et al., 2017), y herramientas de corrección como Denoiser (Ivady et al., 2018), donde se ha logrado aumentar la cobertura y precision de bases en homopolimeros (Ivady et al., 2018; Quince et al., 2011).

Finalmente, la identificacion de polimorfismos en los genes RAP2-7 y LaCAO se resumen en la Figura 6. A pesar de que se identifico polimorfismos especificos que diferencian accesiones amargas y dulces en RAP2-7, no se puede concluir facilmente sobre su reproducibilidad como marcador molecular. Las limitaciones en la secuenciacion de homopolimeros mediante *Oxford Nanopore* indica la necesidad de otras alternativas en la preparacion de libreras, secuenciacion y procesamiento (Jain et al., 2015). En enfoques futuros, se sugiere analizar la funcionalidad del marcador mediante pruebas de deteccion como hibridacion de sonda con tag o metodos basados en PCR con primers dirigidos a sitios polimorficos (Lin et al., 1996). Adicionalmente, es necesario tomar en cuenta la concentracion de alcaloides como rasgo cuantitativo (Sharma et al., 2012). Esto plantea la necesidad de estudiar QTLs asociados a este rasgo, como la region genomica *iucundus* al que pertenece RAP2-7 en *L. angustifolius* (Czepiel et al., 2021), con el fin de estudiar mas genes y desarrollar marcadores con potencial asociacion a este rasgo (Baack et al., 2008).

5. CONCLUSIONES

En este estudio, se logró determinar polimorfismos en los genes LaCAO y RAP2-7 entre accesiones con baja y alta concentración de alcaloides quinolizidínicos en *L. mutabilis*, mediante secuenciación con *Oxford Nanopore*. Se obtuvo 4,23 Gb de información y una cobertura promedio de 31155,21X en LaCAO y 8760,36X para RAP2-7. No se encontró polimorfismos específicos para accesiones amargas o dulces en el gen LaCAO. Sin embargo, se identificó una región polimórfica en el nucleótido 3441 a 3453 del gen LaCAO que es exclusivo para las accesiones PROINPA P5 y PROINPA P8 (bajo contenido de alcaloides); y FR9 P3, FR11 P8, 193 (alto contenido de alcaloides). Por otra parte, sí se observó polimorfismos específicos en RAP2-7 en el nucleótido 8 (una deleción de citosina), y en el nucleótido 1528 (una deleción de citosina), que son comunes en las accesiones de bajo contenido de alcaloides 706, 707 y 742, y en 707, 712, 723, 742, y PROINPA P5, respectivamente. En general, para futuros proyectos asociados a la búsqueda de polimorfismos, se recomienda utilizar estrategias de secuenciamiento apoyado con *Sanger* para regiones homopoliméricas acoplada a *NGS*. Además, se sugiere la realización de pruebas moleculares basadas en PCR como CAPS o dCAPS para determinar la funcionalidad de RAP2-7 como marcador molecular. Finalmente, se recomienda enfocar otros estudios en el locus *iucundus* en *L. mutabilis*, con la finalidad de determinar otros genes asociados al rasgo cuantitativo de síntesis de alcaloides. Los resultados obtenidos en esta investigación son relevantes para establecer potenciales marcadores moleculares dirigidos a accesiones con bajo contenido de alcaloides, con el fin de potenciar programas de fito-mejoramiento para la selección de estas variedades de chocho.

6. TABLAS

Tabla 1. Porcentaje de alcaloides promedio de accesiones bajas en alcaloides (PROINPA P5 y PROINPA P8) y altas en alcaloides (FR9 P3 y FR11 P8) obtenidas mediante una titulación ácido base con NaOH y fenolftaleína por triplicado. Std. Dev.: Desviación estándar.

Muestra	Accesiones Bajas en Alcaloides		Accesiones Altas en Alcaloides		
	% Alcaloides	Std. Dev.	Muestra	% Alcaloides	Std. Dev.
PROINPA P5	3,47	0,047	FR9 P3	4,55	0,068
PROINPA P8	3,85	0,245	FR11 P8	4,57	0,109

Tabla 2. Primers propuestos para la amplificación de los genes LaCAO_ fragmento 1, LaCAO_ fragmento 2, LaCAO_ fragmento 3 y RAP2-7 obtenidos en Primer3. Tm: Temperatura de melting. %GC: porcentaje de guaninas y citosinas. Hairpin: Estabilidad de estructuras secundarias. ΔG : Energía espontánea de formación de estructuras secundarias.

Gen	Primer	Tm	%GC	Hairpin	ΔG	Secuencia	Extensión	Tamaño de producto esperado (nt)
LaCAO_ fragmento 1	Forward	59,70	40,00	0	-0,72	GCAATCTTCTTAGCCATCATTTCT	Exón 1- Exón 2	2336
	Reverse	60,35	50,00	0		GAGTTTGTGCAGCATCCTTAC		
LaCAO_ fragmento 2	Forward	59,51	40,00	0	1,48	TGGACCCTTTATCTCCTTCTGAAAT	Exón 2- Exón 8	2794
	Reverse	59,24	41,67	0		AGCTTCACCAGGTTTAGAATCAAC		
LaCAO_ fragmento 3	Forward	59,18	40,00	0	0,93	TAAGCTAACTGGAATTCTGAGCATG	Exón 8- Exón 12	1669
	Reverse	57,69	40,00	0		GACTCCAATTCACATTTACTAGGTG		
RAP2-7	Forward	57	36,00	0	-1,15	ATCTCAACCTATGGTAAAGATTTGG	Exón 1- Exón 8	2926
	Reverse	63	45,00	0		TCTCCCTGATGCAATGTTCA		

Tabla 3. Resultados generales de la Secuenciación del pool de 7 genes en las 14 accesiones obtenidas por el software MinION™.

Rendimiento de la Secuenciación

<i>Tiempo de corrida</i>	24 horas
<i>Tiempo de basecalling</i>	75 horas
<i>Información generada</i>	4,23 Gb
<i>Número de lecturas</i>	6,21 M
<i>Calidad promedio</i>	12,9 Phred

Tabla 4. Cobertura tras el mapeo de los genes RAP2-7 y LaCAO en las 14 accesiones analizadas.

	Accesión	Cobertura
RAP2-7	706 (Bajo en alcaloides)	14.607
	707 (Bajo en alcaloides)	8.588
	712 (Bajo en alcaloides)	14.775
	723 (Bajo en alcaloides)	14.042
	742 (Bajo en alcaloides)	8.050
	PROINPA P5 (Bajo en alcaloides)	1.481
	PROINPA P8 (Bajo en alcaloides)	23
	176 (Alto en alcaloides)	8.074
	188 (Alto en alcaloides)	10.193
	190 (Alto en alcaloides)	11.026
	192 (Alto en alcaloides)	12.522
	193 (Alto en alcaloides)	13.014
	FR11 P3 (Alto en alcaloides)	3.103
	FR11 P8 (Alto en alcaloides)	3.147
LaCAO	706 (Bajo en alcaloides)	21.537
	707 (Bajo en alcaloides)	19.451
	712 (Bajo en alcaloides)	27.231
	723 (Bajo en alcaloides)	28.482

742 (Bajo en alcaloides)	15.405
PROINPA P5 (Bajo en alcaloides)	47.247
PROINPA P8 (Bajo en alcaloides)	39.821
176 (Alto en alcaloides)	20.423
188 (Alto en alcaloides)	29.092
190 (Alto en alcaloides)	25.988
192 (Alto en alcaloides)	28.984
193 (Alto en alcaloides)	27.754
FR9 P3 (Alto en alcaloides)	55.918
FR11 P8 (Alto en alcaloides)	48.840

7. FIGURAS

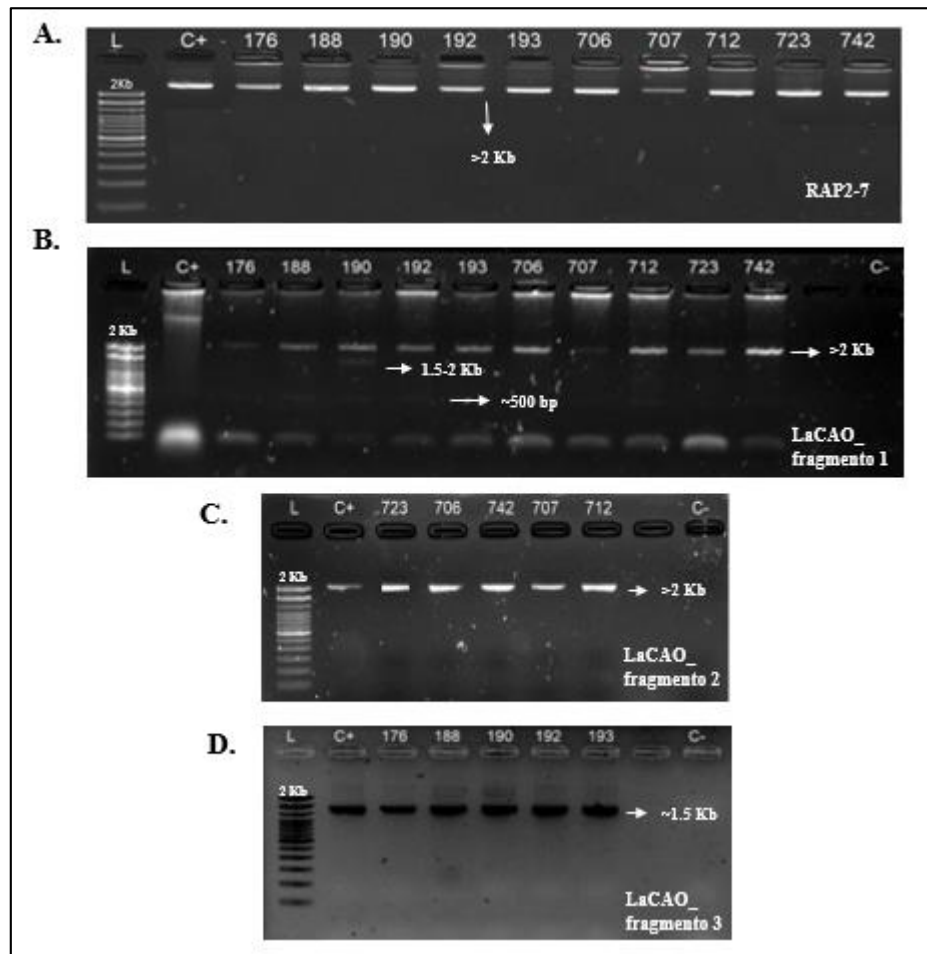


Figura 1. Amplificación de los genes RAP2-7, LaCAO_ fragmento 1, LaCAO_ fragmento 2 y LaCAO_ fragmento 3 en las accesiones 176, 188, 190, 192, 193, 707, 707, 712, 723 y 742. A. Amplificación de RAP2-7 en accesiones 176-742. B. Amplificación de LaCAO_ fragmento 1 en accesiones 176-742. C. Amplificación de LaCAO_ fragmento 2 en accesiones dulces (706, 707, 712, 723 y 742). D. Amplificación de LaCAO_ fragmento 3 en accesiones amargas (176, 188, 190, 192 y 193). L: Ladder 2Kb. C+: Control positivo. C-: Control negativo.

1. 706 (Bajo en Alcaloides)	A A C T C A A - C T A T G G T A A A G A
2. 707 (Bajo en Alcaloides)	A A C T C A A - C T A T G G T A A A G A
3. 712 (Bajo en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
4. 723 (Bajo en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
5. 742 (Bajo en Alcaloides)	A A C T C A A - C T A T G G T A A A G A
6. PROINPA P5 (Bajo en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
7. PROINPA P8 (Bajo en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
8. 176 (Alto en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
9. 188 (Alto en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
10. 190 (Alto en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
11. 192 (Alto en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
12. 193 (Alto en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
13. FR9 P3 (Alto en Alcaloides)	A A C T C A A C C T A T G G T A A A G A
14. FR11 P8 (Alto en Alcaloides)	A A C T C A A C C T A T G G T A A A G A

Polimorfismo en el nucleótido 8. Deleción de una citosina en las accesiones 706, 707 y 742.

1. 706 (Bajo en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
2. 707 (Bajo en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
3. 712 (Bajo en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
4. 723 (Bajo en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
5. 742 (Bajo en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
6. PROINPA P5 (Bajo en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
7. PROINPA P8 (Bajo en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
8. 176 (Alto en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
9. 188 (Alto en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
10. 190 (Alto en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
11. 192 (Alto en Alcaloides)	T A T G T G G T A T C T A T C A G T G T T
12. 193 (Alto en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
13. FR9 P3 (Alto en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T
14. FR11 P8 (Alto en Alcaloides)	T A T G T G G T A T A T A T C A G T G T T

Polimorfismo de 1 sólo nucleótido en la posición 473. Cambio de adenina a citosina en la accesión 192.

1. 706 (Bajo en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
2. 707 (Bajo en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
3. 712 (Bajo en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
4. 723 (Bajo en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
5. 742 (Bajo en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
6. PROINPA P5 (Bajo en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
7. PROINPA P8 (Bajo en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
8. 176 (Alto en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
9. 188 (Alto en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
10. 190 (Alto en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
11. 192 (Alto en Alcaloides)	T T G A T A A T T T G G A C T A A T T G A
12. 193 (Alto en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
13. FR9 P3 (Alto en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A
14. FR11 P8 (Alto en Alcaloides)	T T G A T A A A T T G G A C T A A T T G A

Polimorfismo de 1 solo nucleótido en la posición 1011. Cambio de adenina a timina en la accesión 192.

1. 706 (Bajo en Alcaloides)	C C T T T T T T G G T A C T A C C A A T
2. 707 (Bajo en Alcaloides)	- C T T T T T T T G G T A C T A C C A A T
3. 712 (Bajo en Alcaloides)	- C T T T T T T T G G T A C T A C C A A T
4. 723 (Bajo en Alcaloides)	- C T T T T T T T G G T A C T A C C A A T
5. 742 (Bajo en Alcaloides)	- C T T T T T T T G G T A C T A C C A A T
6. PROINPA P5 (Bajo en Alcaloides)	- C T T T T T T T G G T A C T A C C A A T
7. PROINPA P8 (Bajo en Alcaloides)	C C T T T T T T G G T A C T A C C A A T
8. 176 (Alto en Alcaloides)	C C T T T T T T G G T A C T A C C A A T
9. 188 (Alto en Alcaloides)	C C T T T T T T G G T A C T A C - A A -
10. 190 (Alto en Alcaloides)	C C T T T T T T G G T A C T A C C A A T
11. 192 (Alto en Alcaloides)	C C T T T T T T G G T A C T A C C A A T
12. 193 (Alto en Alcaloides)	C C T T T T T T G G T A C T A C C A A T
13. FR9 P3 (Alto en Alcaloides)	C C T T T T T T G G T A C T A C C A A T
14. FR11 P8 (Alto en Alcaloides)	C C T T T T T T G G T A C T A C C A A T

Polimorfismo de 1 solo nucleótido común en accesiones de bajo contenido de alcaloides (707, 712, 723, 742, PROINPA P5). Deleción de citosina en posición 1528.

1. 706 (Bajo en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
2. 707 (Bajo en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
3. 712 (Bajo en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
4. 723 (Bajo en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
5. 742 (Bajo en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
6. PROINPA P5 (Bajo en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
7. PROINPA P8 (Bajo en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
8. 176 (Alto en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
9. 188 (Alto en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
10. 190 (Alto en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
11. 192 (Alto en Alcaloides)	T C C - - A C A T - G A A C A T G A - - A
12. 193 (Alto en Alcaloides)	T C C - - A C A T - G A A C A T G A - - A
13. FR9 P3 (Alto en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A
14. FR11 P8 (Alto en Alcaloides)	T C C C A A C A T G G A A C A T G A A G A

Polimorfismo en accesiones 192 y 193. Deleción de citosina, adenina, guanina, adenina y guanina en las posiciones 1623, 1624, 1629, 1638 y 1639, respectivamente.

Figura 2. Polimorfismos en RAP2-7 de las 14 accesiones. En la izquierda se encuentra la alineación de las secuencias consenso de todas las accesiones, y en la derecha el tipo de polimorfismo y la posición.

1. 706 (Bajo en Alcaloides)	G G T T A A - - - - - A G A - - A T C G A G A A
2. 707 (Bajo en Alcaloides)	- - - G A A T C C A T G G A - - A A A A A A A A
3. 712 (Bajo en Alcaloides)	G T G A A A - - - - - T G T A - - A T T T A A A A
4. 723 (Bajo en Alcaloides)	T A T T A A T A T T T A G A - - A T A A T A A G
5. 742 (Bajo en Alcaloides)	T A T T T T - - - - - T T A A - - A A A A A A A A
6. PROINPA P5 (Bajo en Alcaloides)	T A T T A T T A T T T G T T - - T T A A A A A A
7. PROINPA P8 (Bajo en Alcaloides)	T T T T A - - - - - T T C A - - A A T A A A A A
8. 176 (Alto en Alcaloides)	T T A A A A T G T A T T T A - - A C T A A A T A
9. 190 (Alto en Alcaloides)	G G T T A A - - - - - A A G A - - A T C G A G A A
10. 192 (Alto en Alcaloides)	- -
11. 193 (Alto en Alcaloides)	C A T T A A T - - - T A C A - - A G C A A A G A
12. FR9 P3 (Alto en Alcaloides)	A G T G G A G T C G T G G A G G A T A G A T G A
13. FR11 P8 (Alto en Alcaloides)	A G T A G A T G T G C A T A - - A C T A A A A A

Región descartada no comparable. 0 a 2727 nucleótidos (Fragmento 1)

1. 706 (Bajo en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
2. 707 (Bajo en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
3. 712 (Bajo en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
4. 723 (Bajo en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
5. 742 (Bajo en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
6. PROINPA P5 (Bajo en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
7. PROINPA P8 (Bajo en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
8. 176 (Alto en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
9. 188 (Alto en Alcaloides)	C T A A A C G T G G C A A T G A T A A A A G
10. 190 (Alto en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
11. 192 (Alto en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
12. 193 (Alto en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
13. FR9 P3 (Alto en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G
14. FR11 P8 (Alto en Alcaloides)	C T A G A G G T G G A A G T G A T A G A A G

Polimorfismo de 1 sólo nucleótido en la accesión 188. Cambio de guanina a adenina, guanina a citosina, adenina a citosina, guanina a adenina y guanina a adenina en las posiciones 4898, 4900, 4905, 4907 y 4913.

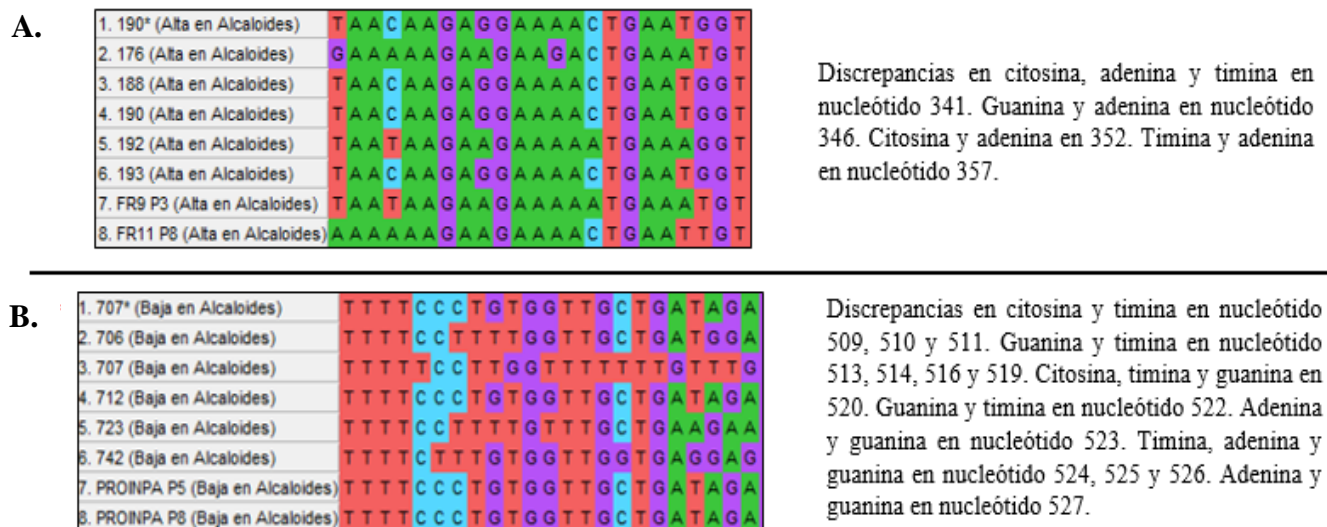


Figura 4. Discrepancias entre secuencias consenso obtenidas por la secuenciación de Oxford Nanopore y Sanger A. Accesiones altas en alcaloides. B. Accesiones bajas en alcaloides. En la izquierda se encuentra la alineación de las secuencias consenso, y en la derecha el tipo de polimorfismo y la posición. *: Secuencia consenso obtenida de *Sanger*.

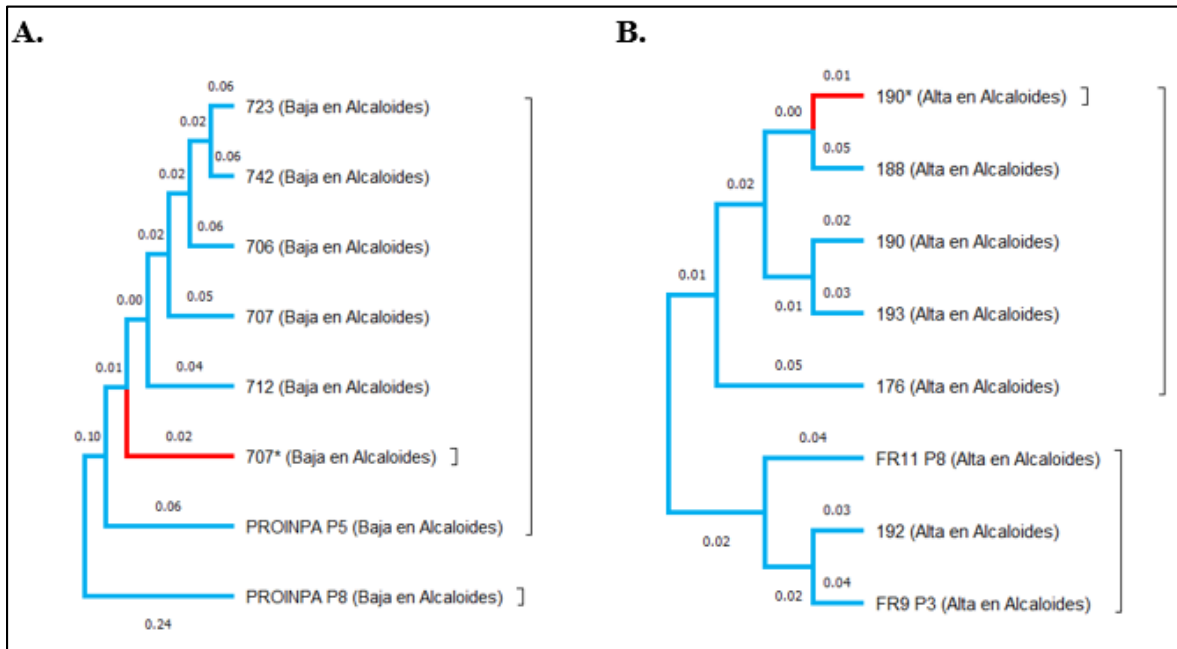


Figura 5. Árboles filogenéticos de las secuencias consenso obtenidas por Oxford Nanopore y Sanger. A. Filogenia de accesiones dulces (706, 707, 712, 723 y 742). B. Filogenia de accesiones amargas (176, 188, 190, 192 y 193). Los números sobre las ramas son el *branch lenght* entre accesiones. Color celeste: Secuencias de *Oxford Nanopore*. Color rojo: Secuencias de *Sanger*.

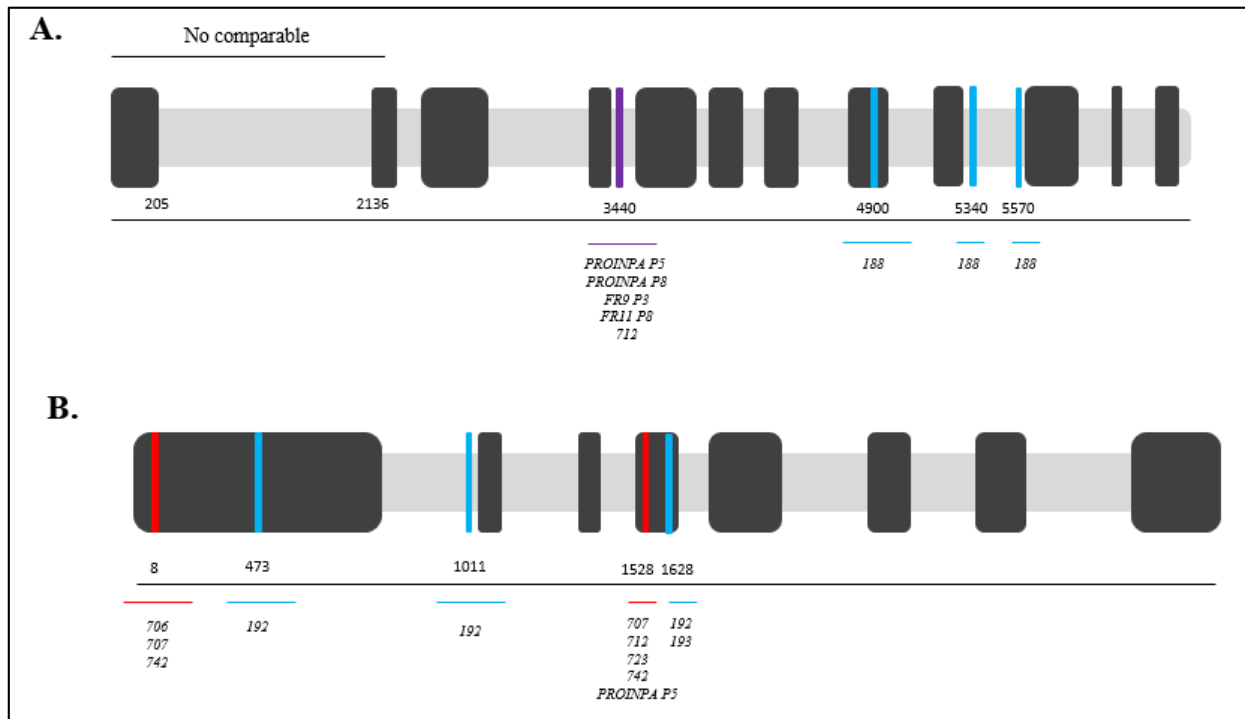


Figura 6. Ubicación de polimorfismos asociados a los genes LaCAO y RAP2-7. A. Gen LaCAO. B. Gen RAP2-7. Color negro: Exones. Color plomo: Intrones. Color celeste: Polimorfismos asociados a una o dos accesiones. Color morado: Polimorfismo asociado a muestras seleccionadas cualitativamente (*PROINPA P5*, *PROINPA P8*, *FR9 P3* y *FR11 P8*). Color rojo: Polimorfismos asumidos como específicos para accesiones de bajo contenido de alcaloides.

8. REFERENCIAS BIBLIOGRÁFICAS

- Applied Biosystems. (2016). *Troubleshooting Sanger sequencing data*. Recuperado de https://tools.thermofisher.com/content/sfs/manuals/MAN0014435_Trbleshoot_Sanger_seq_data_UB.pdf?cid=techcomm92
- Baack, E. J.; Sapir, Y.; Chapman, M.; Burke, J. & Rieseberg, L. H. (2008). Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. *Molecular ecology*, 17(2), 666–677.
- Blawid, R.; Silva, J. & Nagata, T. (2017). Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. *Annals of Applied Biology*, 170(3), 301–314.
- Borah, P. (2011). Primer designing for PCR. *Science Vision*, 11(3), 134-136.
- Chen, Y.; Ye, W.; Zhang, Y. & Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Research*, 43(16), 7762–7768.
- Chirinos-Arias, M. (2015). Tarwi (*Lupinus mutabilis* Sweet) una planta con potencial nutritive y medicinal. *Revista Bio Ciencias*, 3(3), 163-172.
- Czepiel, K.; Krajewski, P.; Wilczura, P.; Bielecka, P.; Świącicki, W. & Kroc, M. (2021). Expression Profiles of Alkaloid-Related Genes across the Organs of Narrow-Leafed Lupin (*Lupinus angustifolius* L.) and in Response to Anthracnose Infection. *International journal of molecular sciences*, 22(5), 2676.
- Fadhil, S.; Hamid, M.; Rouhollah, G. & Mohammad, R. (2007). Spectrophotometric Determination of Total Alkaloids in *Peganum harmala* L. Using Bromocresol Green. *Research Journal of Phytochemistry*, 1, 79-82

- Freed, N.; Vlková, M.; Faisal, M. & Silander, O. K. (2020). Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods and Protocols*, 5(1).
- Frick, K. (2018). Genomic research to improve the quality of narrow-leafed lupin (*Lupinus angustifolius* L.) grain
- Frick, K. M.; Kamphuis, L. G.; Siddique, K. H. M.; Singh, K. B. & Foley, R. C. (2017). Quinolizidine Alkaloid Biosynthesis in Lupins and Prospects for Grain Quality Improvement. *Frontiers in Plant Science*, 8.
- Golparian, D.; Dona, V.; Sánchez-Busó, L.; Foerster, S.; Harris, S.; Endimiani, A.; Low, N. & Unemo, M. (2018). Oxford Nanopore MinION genome sequencer: performance characteristics, optimised analysis workflow, phylogenetic analysis and prediction of antimicrobial resistance in *Neisseria gonorrhoeae*. *The Preprint Server for Biology*, 349316.
- Gutierrez, A.; Infantes, M.; Pascual, G. & Zamora, J. (2016). Evaluación de los factores en el desamargado de tarwi (*Lupinus mutabilis* Sweet). *Agroindustrial Science*, 6(1), 145-149.
- Haines, A.; Tobe, S.; Kobus, H. & Linacre, A. (2015). Properties of nucleic acid staining dyes used in gel electrophoresis. *ELECTROPHORESIS*, 36(6), 941–944.
- Hall, B. (2013). Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution*, 30(5), 1229–1235.
- Invitrogen. (2015). *Invitrogen™ ADN polimerasa Platinum™*. Recuperdo de https://assets.fishersci.com/TFS-Assets/LSG/manuals/Platinum_Taq_DNA_Polymerase_UG.pdf?_ga=2.37384164.554383315.1650349086-918965874.1647983306

- Ishii, K. & Fukui, M. (2001). Optimization of Annealing Temperature to Reduce Bias Caused by a Primer Mismatch in Multitemplate PCR. *Applied and Environmental Microbiology*, 67(8), 3753–3755.
- Iványi, G.; Madar, L.; Dzsudzsák, E.; Koczok, K.; Kappelmayer, J.; Krulisova, V.; Macek, M.; Jr, Horváth, A. & Balogh, I. (2018). Analytical parameters and validation of homopolymer detection in a pyrosequencing-based next generation sequencing system. *BMC genomics*, 19(1), 158.
- Jacobse, S. & Mujica, A. (2006). El tarwi (*Lupinus mutabilis* Sweet.) y sus parientes silvestres. *Botánica Económica de los Andes Centrales*, 458-482.
- Jain, M.; Fiddes, I.; Miga, K.; Olsen, H.; Paten, B. & Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. *Nature Methods*, 12(4), 351–356.
- Joint Genome Institute. (2020). Recuperado de <https://genome.jgi.doe.gov/portal/>
- Karst, S.; Ziels, R.; Kirkegaard, R.; Sørensen, E.; McDonald, D.; Zhu, Q.; Knight, R. & Albertsen, M. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nature Methods*, 18(2), 165–169.
- Kroc, M.; Czepiel, K.; Wilczura, P.; Mokrzycka, M. & Świącicki, W. (2019). Development and Validation of a Gene-Targeted dCAPS Marker for Marker-Assisted Selection of Low-Alkaloid Content in Seeds of Narrow-Leafed Lupin (*Lupinus angustifolius* L.). *Genes*, 10(6), 428.
- Koressaar, T. & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10), 1289–1291.
- Lee, B., Moon, T., Yoon, S. & Weissman, T. (2017). DUDE-Seq: Fast, flexible, and robust denoising for targeted amplicon sequencing. *PLOS ONE*, 12(7).

- Lin, J.; Kuo, J.; Ma, J.; Saunders, J.; Beard, H.; MacDonald, M.; Kenworthy, W.; Ude, G. & Matthews, B. (1996). Identification of molecular markers in soybean comparing RFLP, RAPD and AFLP DNA mapping techniques. *Plant Molecular Biology Reporter*, *14*(2), 156–169.
- Lucht, J. (2015). Public Acceptance of Plant Biotechnology and GM Crops. *Viruses*, *7*(8), 4254–4281.
- Menegon, M.; Cantaloni, C.; Rodriguez-Prieto, A.; Centomo, C.; Abdelfattah, A.; Rossato, M.; Bernardi, M.; Xumerle, L.; Loader, S. & Delledonne, M. (2017). On site DNA barcoding by nanopore sequencing. *PLOS ONE*, *12*(10), 1-18.
- Morgil, H.; Can Gercek, Y. & Tulum, I. (2020). Single Nucleotide Polymorphisms (SNPs) in Plant Genetics and Breeding. *The Recent Topics in Genetic Polymorphisms*.
- Nazarenko, I. (2002). Multiplex quantitative PCR using self-quenched primers labeled with a single fluorophore. *Nucleic Acids Research*, *30*(9), 37–37.
- NewEngland BioLabs Inc. (2022). *Protocol for LongAmp™ Taq 2X Master Mix*.
Recuperado de <https://international.neb.com/protocols/2012/09/05/protocol-for-longamp-taq-2x-master-mix-m0287>
- Ortega-David, E.; Rodríguez, A.; David, A. & Zamora-Burbano, Á. (2010). Caracterización de semillas de lupino (*Lupinus mutabilis*) sembrado en los Andes de Colombia. *Acta Agronómica*, *59*(1), 111-118.
- Petersen, L.; Martin, I.; Moschetti, W.; Kershaw, C. & Tsongalis, G. J. (2019). Third generation sequencing in the clinical laboratory: Exploring the advantages and challenges of nanopore sequencing. *Journal of Clinical Microbiology*, *1*(58).
- Quince, C.; Lanzen, A.; Davenport, R. & Turnbaugh, P. (2011). Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, *12*(38), 1-18.

- Rychlik, W. (1993). Selection of Primers for Polymerase Chain Reaction. *PCR, Protocols*, (15) 31–40.
- Ruiz-López, M.; Barrientos-Ramírez, L.; García-López, P.; Valdés-Miramontes, E.; Zamora-Natera, J.; Rodríguez-Macias, R.; Salcedo-Pérez, E.; Bañuelos-Pineda, J. & Vargas-Radillo, J. (2019). Nutritional and Bioactive Compounds in Mexican Lupin Beans Species: A Mini-Review. *Nutrients*, 11(8), 1785.
- Sharma, V.; Chaudhary, S.; Srivastava, S.; Pandey, R. & Kumar, S. (2012). Characterization of variation and quantitative trait loci related to terpenoid indole alkaloid yield in a recombinant inbred line mapping population of *Catharanthus roseus*. *Journal of genetics*, 91(1), 49–69.
- Slatko, B.; Gardner, A. & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current protocols in molecular biology*, 122(1), e59. <https://doi.org/10.1002/cpmb.59>
- Tabata, Y.; Matsuo, Y.; Fujii, Y. & Hirota, K. (2022). Rapid detection of single nucleotide polymorphisms using the MinION nanopore sequencer: a feasibility study for perioperative precision medicine. *JA Clin Rep*, 8(17), .
- Tavladoraki, P.; Cona, A. & Angelini, R. (2016). Copper-Containing Amine Oxidases and FAD-Dependent Polyamine Oxidases Are Key Players in Plant Tissue Differentiation and Organ Development. *Frontiers in plant science*, 7, 824.
- Thornton, B. & Basu, C. (2015). Rapid and Simple Method of qPCR Primer Design. *PCR Primer Design*, 173–179.
- UNAFold. (2003). UNAFold Forming Fold. *Nucleic Acids Res*, 31(13), 3406-3415.

- Villacrés, E.; Quelal, M. B.; Fernández, E.; García, G.; Cueva, G. & Rosell, C. M. (2020). Impact of debittering and fermentation processes on the antinutritional and antioxidant compounds in *Lupinus mutabilis* sweet. *LWT*, 109745(131).
- Villacrés, E.; Peralta, E.; Cuadrado, L.; Revelo, J.; Addo, S. & Aldaz Berrones, R. F. (2009). Propiedades y aplicaciones de los alcaloides del chocho (*Lupinus mutabilis* Sweet).
- Wick, R.; Judd, L. & Holt, K. (2018). Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLOS Computational Biology*, 14(11)
- Whitford, W.; Hawkins, V.; Moodley, K.; Grant, M.; Lehnert, K.; Snell, R.; & Jacobsen, J. (2021). Optimised multiplex amplicon sequencing for mutation identification using the MinION nanopore sequencer. *bioRxiv*
- Yang, T.; Nagy, I.; Mancinotti, D.; Otterbach, S. L.; Andersen, T. B.; Motawia, M. S.; Asp, T. & Geu-Flores, F. (2017). Transcript profiling of a bitter variety of narrow-leafed lupin to discover alkaloid biosynthetic genes. *Journal of experimental botany*, 68(20), 5527–5537.
- Yang, X.; Chockalingam, S. P. & Aluru, S. (2012). A survey of error-correction methods for next-generation sequencing. *Briefings in Bioinformatics*, 14(1), 56–66.
- Zaretskaya, I.; Cutcutache, I.; Rozen, S. & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13(1), 13-134
- Zamora-Natera, F.; García-López, P.; Ruiz-López, M.; Rodríguez-Macías, R. & Salcedo-Pérez, E. (2009). Composición y concentración de alcaloides en *Lupinus Exaltatus* zucc. durante su crecimiento y desarrollo. *Interciencia*, 34(9), 672-676.

9. ANEXOS

ANEXO 1: CUANTIFICACIÓN DE ALCALOIDES QUINOLIZIDÍNICOS EN

ACCESIONES 176, 188, 190, 192, 193, 706, 707, 712, 723 Y 742

<i>Accesión</i>	<i>Accesiones Bajas en Alcaloides</i>			<i>Accesiones Altas en Alcaloides</i>			
	<i>% Alcaloides</i>	<i>Std. Dev.</i>	<i>País Origen</i>	<i>Accesión</i>	<i>% Alcaloides</i>	<i>Std. Dev.</i>	<i>País Origen</i>
706	3,34	0,344	Perú	176	4,65	0,068	Ecuador
707	2,84	0,111	Perú	188	4,75	0,109	Otros
712	3,47	0,344	Perú	190	4,68	0,325	Ecuador
723	3,61	0,099	Perú	192	4,91	0,19	ExURSS
742	3,15	0,047	Perú	193	4,79	0,036	ExURSS

Descripción: Se observa el porcentaje de alcaloides QAs de las accesiones seleccionadas previamente en el Laboratorio de Biotecnología Vegetal de la USFQ. Se incluye el nombre de la accesión (código propuesto por el INIAP), y su país de origen. Std. Dev.: Desviación estándar.

**ANEXO 2: TAMAÑO, CALIDAD, NÚMERO DE READS Y PORCENTAJE DE
RECUPERACIÓN DE LA SECUENCIACIÓN DEL POOL DE 7 GENES EN LAS 14
ACCESIONES ANALIZADAS**

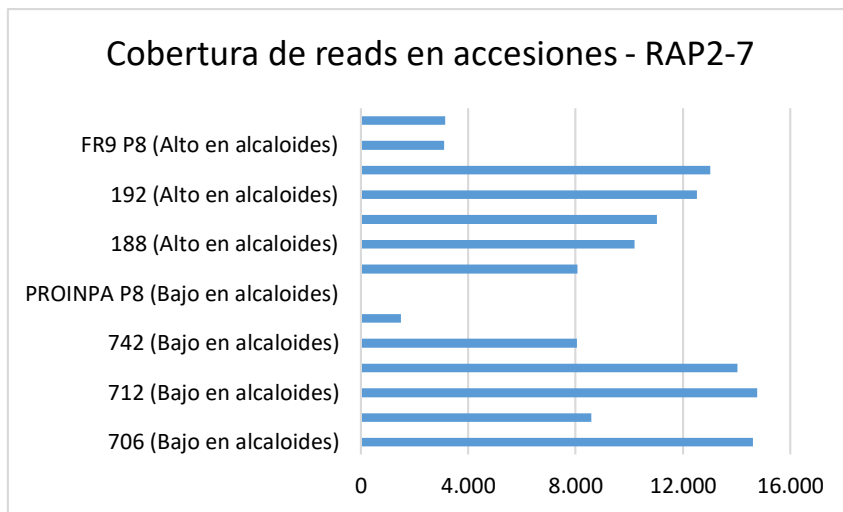
<i>Muestra</i>		Tamaño Medio (pb)	Calidad Media (Phred)	Número de reads	% Recuperación
706 (Bajo en alcaloides)	Sin Filtro	719,1	13,1	266 632	79,73%
	Filtro de calidad y tamaño	650,8	14,6	212 577	
707 (Bajo en alcaloides)	Sin Filtro	553,1	13,1	256 876	71,54%
	Filtro de calidad y tamaño	505,3	14,6	183 769	
712 (Bajo en alcaloides)	Sin Filtro	585,9	13	410 654	86,70%
	Filtro de calidad y tamaño	490,6	14,7	356 044	
723 (Bajo en alcaloides)	Sin Filtro	652,5	12,9	331 754	77,01%
	Filtro de calidad y tamaño	608	14,5	255 479	
742 (Bajo en alcaloides)	Sin Filtro	707,3	12,9	245 227	77,92%
	Filtro de calidad y tamaño	664,4	14,4	191 087	
PROINPA P5 (Bajo en alcaloides)	Sin Filtro	905,8	13,1	218 985	80,78%
	Filtro de calidad y tamaño	853,3	14,4	176 900	
PROINPA P8 (Bajo en alcaloides)	Sin Filtro	1169	13	157 814	81,57%
	Filtro de calidad y tamaño	1130	14,4	128 735	
176 (Alto en alcaloides)	Sin Filtro	466,4	12,4	378 907	85,51%
	Filtro de calidad y tamaño	377,7	14,6	324 005	
188 (Alto en alcaloides)	Sin Filtro	522,1	12,7	377 703	82,23%
	Filtro de calidad y tamaño	444,3	14,4	310 583	
190 (Alto en alcaloides)	Sin Filtro	554,5	13	342 907	72,11%
	Filtro de calidad y tamaño	519,9	14,6	247 261	
192 (Alto en alcaloides)	Sin Filtro	525,1	12,8	404 326	70,44%
	Filtro de calidad y tamaño	489,4	14,6	284 824	
193 (Alto en alcaloides)	Sin Filtro	591,5	12,9	323 587	74,04%
	Filtro de calidad y tamaño	557,6	14,6	239 578	
FR9 P3 (Alto en alcaloides)	Sin Filtro	1070,9	13	196 886	81,94%
	Filtro de calidad y tamaño	1030,5	14,4	161 323	

FR11 P8 (Alto en alcaloides)	Sin Filtro	1018,2	13,1	181 544	80,34%
	Filtro de calidad y tamaño	983	14,4	145 857	

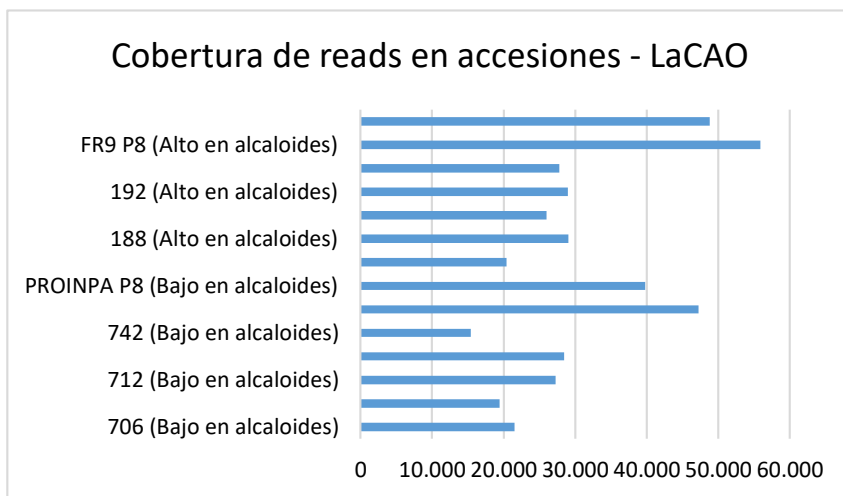
Descripción: Resultados de la secuenciación en las accesiones 176, 188, 190, 192, 193, PROINPA P5, PROINPA P8, 706, 707, 712, 723, 742, FR9 P3 y FR11 P8.

**ANEXO 3: GRÁFICO DE BARRAS DE COBERTURA DE LOS GENES RAP2-7 Y
LACAO EN LAS 14 ACCESIONES ANALIZADAS**

A.



B.



Descripción: Gráfico de barras para la cobertura de los genes en las accesiones 176, 188, 190, 192, 193, PROINPA P5, PROINPA P8, 706, 707, 712, 723, 742, FR9 P3 y FR11 P8. A. RAP2-7. B. LaCAO.

