



Centro de Investigación Científica de Yucatán, A. C.
Posgrado en Ciencias Biológicas

**“Identificación y clonación de genes
relacionados a la biosíntesis del ácido betulínico
en *Pentalinon andrieuxii*”**

Tesis que presenta

Alexa Sharai Aguilar Acevedo

En opción al título de

MAESTRO EN CIENCIAS

(Ciencias Biológicas: Bioquímica y Biología Molecular de
Plantas)

Mérida, Yucatán, México

2022

CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.
POSGRADO EN CIENCIAS BIOLÓGICAS



RECONOCIMIENTO

Por medio de la presente, hago constar que el trabajo de tesis de **Alexa Sharai Aguilar Acevedo** titulado “**Identificación y clonación de genes relacionados a la biosíntesis del ácido betulínico en *Pentalinon andrieuxii***”, fue realizado en la **Unidad de Bioquímica y Biología Molecular de Plantas**, en el laboratorio No. 26 del **Centro de Investigación Científica de Yucatán, A.C.** bajo la dirección del **Dr. Gregorio Godoy Hernández** y la **Dra. Elsa Beatriz Góngora Castillo**, dentro de la opción de Bioquímica y Biología Molecular, perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Atentamente



Dra. Cecilia Hernández Zepeda
Directora de Docencia

Mérida, Yucatán, México, a 2 de marzo de 2022

DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en la sección de Materiales y Métodos, los Resultados y Discusión de este documento proviene de las actividades de investigación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que a razón de lo anterior y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica de Yucatán, A.C., y en el mismo tenor, reconozco que si derivaren de este trabajo productos intelectuales o desarrollos tecnológicos, en lo especial, estos se regirán en todo caso por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.



Alexa Sharai Aguilar Acevedo

Este trabajo se llevó a cabo en la Unidad de Bioquímica y Biología Molecular de Plantas del Centro de Investigación Científica de Yucatán, A.C., y forma parte del proyecto CONACyT No. 257915 titulado: ANÁLISIS DEL TRANSCRIPTOMA RELACIONADO CON LA BIOSÍNTESIS DE ISOPRENOIDES EN *Pentalinon andrieuxii* (APOCYNACEAE). bajo la dirección del Dr. Gregorio Godoy Hernández y la Dra. Elsa Góngora Castillo .

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada No. 747644.

Al Centro de Investigación Científica de Yucatán, a la Unidad de Bioquímica y Biología de Plantas por la oportunidad de realizar el posgrado.

Sin duda alguna, llegar hasta aquí sin la aceptación y responsabilidad de mi director el Dr. Gregorio Godoy Hernández no hubiera sido posible, gracias por permitirme ser parte de su grupo de investigación, por su confianza y su apoyo durante el desarrollo de la tesis.

Quiero agradecer de forma especial a mi coasesora la Dra. Elsa Beatriz Góngora Castillo por el tiempo dedicado, su paciencia, las sugerencias y los importantes aportes en la bioinformática que contribuyeron al desarrollo de esta tesis.

De igual manera agradezco el apoyo, dedicación y sugerencias para la mejora de este trabajo a mi asesor externo el Dr. Ángel Arturo Guevara García.

A mi comité revisor, Dr. Felipe Augusto Vázquez Flota y el Dr. Luis Manuel Peña Rodríguez, que de gran forma contribuyeron en la culminación de la tesis.

A la M.C. Elidé Avilés Berzunza por el apoyo técnico brindado, por los consejos y asistencia en el laboratorio durante la fase experimental.

A mis amigos y compañeros del CICY, con los cuales compartí experiencias, sueños y esfuerzos, gracias por su apoyo y todos los gratos recuerdos.

A mi familia, algunas están aquí conmigo y otras en mis recuerdos y en mi corazón, sin importar en donde estén quiero darles las gracias por formar parte de mí, por todo lo que me han brindado.

DEDICATORIAS

A mis padres, Rocío Erika Acevedo Meunier y Luis Aguilar Martínez, por su apoyo incondicional que siempre me han brindado en todas las metas que me he trazado en la vida. Por enseñarme que la distancia no es un impedimento para demostrarme su cariño y amor, siempre sentí su mano protectora y su voz de aliento y respaldo. Sus consejos siempre oportunos y en el momento indicado. Los amo.

A mis hermanas Erika y Brenda, han sido esenciales a lo largo de este tiempo y me han impulsado en cada momento. Gracias por la confianza que siempre nos hemos tenido.

A mi sobrino Iker, por llenar de alegría mi vida. Su gran cariño y amor me hacen sentir muy afortunada de tenerlo.

¡Gracias por todo!

ÍNDICE

INTRODUCCIÓN	1
CAPÍTULO I	3
ANTECEDENTES	3
1.1. <i>Pentalinon andrieuxii</i> (Müll. Arg.) BF Hansen & Wunderlin.....	3
1.1.2. Clasificación taxonómica.....	4
1.1.2. Características taxonómicas.....	4
1.1.3. Importancia.....	5
1.2. Ácido betulínico.....	7
1.2.1. Biosíntesis del ácido betulínico.....	9
1.3. Triterpenos sintasas.....	12
1.4. Papel de las monooxigenasas CYP450 en la biosíntesis de triterpenos.....	13
1.5. Análisis de transcriptomas en plantas.....	14
1.6. Identificación de genes de interés en transcriptoma a través de análisis bioinformáticos.....	15
1.6.1. Homología de secuencias.....	15
1.6.2. Alineamiento de secuencias.....	16
1.7. Filogenética molecular.....	18
1.8. Secuenciación y ensamblado <i>de novo</i> del transcriptoma de <i>P. andrieuxii</i>	18
1.8.1. Ensamblado <i>de novo</i> de <i>P. andrieuxii</i>	19
JUSTIFICACIÓN	20
HIPÓTESIS	20
OBJETIVOS	21
Objetivo general.....	21
Objetivos específicos.....	21
ESTRATEGIA EXPERIMENTAL	22
CAPÍTULO II	25
MATERIALES Y MÉTODOS	25
2.1. Análisis bioinformático.....	25
2.1.1. Identificación de los genes de biosíntesis del ácido betulínico en los transcriptomas.....	25
2.1.2. Análisis filogenético.....	26
2.1.3. Análisis de expresión de los genes de biosíntesis del ácido betulínico.....	26

2.2. Análisis molecular	28
2.2.1. Extracción de ARN total de plantas de <i>P. andrieuxii</i>	28
2.2.2. PCR punto final de los genes implicados en la biosíntesis del ácido betulínico.	29
2.2.3. Aislamiento y clonación del ADNc del gen SQS en <i>Escherichia coli</i>	29
CAPÍTULO III.....	33
RESULTADOS.....	33
3.1. Análisis bioinformático.	33
3.1.1. Identificación de los genes involucrados en la biosíntesis del ácido betulínico en los transcriptomas de <i>P. andrieuxii</i>	33
3.1.2. Análisis filogenético.....	43
3.1.4. Análisis de expresión de los genes involucrados en la biosíntesis del ácido betulínico.	48
3.2. Análisis molecular.	51
3.2.1. Extracción de ARN total de plantas de <i>P. andrieuxii</i>	51
3.2.2. PCR punto final de genes de biosíntesis del ácido betulínico.....	52
3.2.3. Aislamiento y clonación del ADNc del gen SQS.....	54
CAPÍTULO IV.....	61
DISCUSIÓN GENERAL.....	61
CAPITULO V.....	65
CONCLUSIONES Y PERSPECTIVAS.....	65
CONCLUSIONES.....	65
PERSPECTIVAS.....	66
BIBLIOGRAFÍA.....	67
ANEXOS.....	77

LISTADO DE FIGURAS

Figura 1.1. <i>Pentalinon andrieuxii</i> (Müll. Arg.) B.F. Hansen & Wunderlin.....	3
Figura 1.2. Representación esquemática de las características taxonómicas de <i>Pentalinon andrieuxii</i>	5
Figura 1.3. Estructura química del ácido betulínico.	8
Figura 1.4. Esquema general de la biosíntesis de isoprenoides en células vegetales. .	9
Figura 1.5. Esquema de la biosíntesis del ácido betulínico	11
Figura 1.6. Vías para la biosíntesis de ácido oleanólico, ácido ursólico y ácido betulínico catalizados por las monooxigenasas de la familia CYP716A	14
Figura 1.7. Diagrama de la estrategia experimental.	23
Figura 3.1. Alineamiento múltiple entre las SQS identificadas en el transcriptoma del <i>P. andrieuxii</i> y la SQS obtenidas del NCBI.....	40
Figura 3.2. Alineamiento múltiple entre las SQE identificadas en el transcriptoma de <i>P. andrieuxii</i> y la SQE obtenidas del NCBI	41
Figura 3.3. Alineamiento múltiple entre las OSC's identificadas en el transcriptoma de <i>P. andrieuxii</i> y OSC obtenida del NCBI	41
Figura 3.4. Alineamiento múltiple entre las CYP identificadas en el transcriptoma de <i>P. andrieuxii</i> y la CYP450 obtenida del NCBI.	42
Figura 3.5. Árbol filogenético por el método de máxima verosimilitud con el programa PhyLM (Guindon <i>et al.</i> 2010).	44
Figura 3.6. Análisis filogenético por método de máxima verosimilitud con el programa PhyLM (Guindon <i>et al.</i> 2010).	45
Figura 3.7. Mapa de calor de los patrones de expresión de los genes involucrados en la biosíntesis del ácido betulínico para cada condición.	50
Figura 3.8. Integridad de los ARN extraídos de los diferentes tejidos de <i>P. andrieuxii</i>	52
Figura 3.9. Amplificación de los ORF de los genes de la biosíntesis de triterpenos. .	53
Figura 3.10. Predicción del corte con la enzima <i>EcoRI</i> sobre el plásmido pGEM-SQS.....	54
Figura 3.11. Digestión y PCR del plásmido pGEM-SQS	55

Figura 3.12. Alineamiento entre el producto de traducción del transcrito de SQS del transcriptoma de hoja adulta (SQSHA), contra la secuencia obtenida tras la secuenciación (SQS_macrogen)	56
Figura 3.13. Alineamiento entre el producto de traducción del transcrito de LUP del transcriptoma de hoja joven (LUPHJ), contra la secuencia obtenida tras la secuenciación (LUP_macrogen)	57
Figura 3.14. Alineamiento entre la secuencia SQS de <i>Vitis vinifera</i> (VvSQS) y <i>Betulla platyphylla</i> (BpSQS) contra la SQS recuperada (SQS_macrogen).	59
Figura 3.15. Alineamiento entre las OSC de <i>Vitis vinifera</i> (Vv-bAS y VvLUS) contra la LUP recuperada (LUP_macrogen).....	59

LISTADO DE TABLAS

Tabla 1.1 Clasificación taxonómica de <i>Pentalinon</i>	4
Tabla 1.2 Datos de las muestras para ensamblado de transcriptoma	19
Tabla 1.3. Detalles del ensamblaje del transcriptoma <i>de novo</i>	19
Tabla 3.4. Genes que codifican enzimas involucradas en la vía biosintética del ácido betulínico.	33
Tabla 3.5. Alineamiento de secuencias de los transcriptomas de <i>P. andrieuxii</i> contra las secuencias previamente reportadas de la base de datos de Genbank.	36
Tabla 3.6. Agrupación por clústers de los genes candidatos por CD-HIT.	37
Tabla 3.7. Tamaño de los marcos de lecturas de cada secuencia y la longitud de aminoácidos.	38
Tabla 3.8. Secuencias de cebadores diseñados a partir de la región codificante de los genes candidatos para RT-PCR.	47
Tabla 3.9. Matriz del conteo de lecturas de cada gen.....	49
Tabla 3.10. Cuantificación de ARN total de tejidos de <i>P. andrieuxii</i> y relaciones de absorbancia para determinar la pureza del ARN.....	51
Tabla 3.11. Cuantificación de cDNA de tejidos de <i>P. andrieuxii</i> y relaciones de absorbancia para determinar la pureza.....	52
Tabla 3.12. Secuencias que producen alineamientos significativos contra los resultados de secuenciación mediante BLASTp.	58

ABREVIATURAS

aAs	α -amirina sintasa
ADNc	ADN complementario
ARNm	ARN mensajero
bAs	β -amirina sintasa
BLAST	Basic Local Alignment Search Tool
CDS	Coding Sequence
CYP450	Citocromo P450
DEPC	Dietilpirocarbonato
DMAPP	Dimetilalil difosfato
FDS	Farnesil difosfato sintasa
FPP	Farnesil pirofosfato
GC	Guanina-Citocina
GPP	Geranil pirofosfato
GPPS	Geranil profosfato sintasa
HA	Hoja adulta
HJ	Hoja joven
IDI	Isopentenil profosfato isomerasa
IPP	Isopentil difosfato
IPTG	Isopropil- β -D-1-tiogalactopiranosido
IUPAC	International Union of Pure and Applied Chemistry
LB	Medio Luria-Bertani
LUP	Lupel
LUS	Lupeol sintasa
MEP	Vía del metileritritol fosfato
MFAS	Amirina sintasa de función múltiple
MVP	Vía del mevalonato
NADPH	Nicotiamida-Adenina Dinucleotido fosfato
NCBI	Centro Nacional para la Información Biotecnológica
ng	Nanogramo
ORF	Marco de lectura abierta (open reading frame)
OSC	Oxido escualeno ciclase
pb	Pares de bases
pH	Potencial de hidrógeno

RA	Raíz adulta
RJ	Raíz joven
RNA-seq	Secuenciación del Transcriptoma
rpm	Revoluciones por minuto
RT-PCR	Reacción en cadena de la polimerasa con reverso transcripción
SQE	Escualeno epoxidasa
SQS	Escualeno sintasa
TAE	Tris-Acetato-EDTA
TPM	Transcritos por millón de kilobases
UV	Ultravioleta
x g	Fuerza centrífuga relativa, expresada en gravedad
X-Gal	β -D-galactopiranosido

RESUMEN

Pentalinon andrieuxii es una planta nativa de la península de Yucatán que se ha utilizado comúnmente en la medicina tradicional maya debido a sus propiedades biológicas y representa una fuente natural de metabolitos con potencial aplicación en la industria farmacéutica. Recientemente, se ha detectado que las plantas de *P. andrieuxii* genéticamente transformadas mostraron un aumento en su contenido de ácido betulínico, un triterpeno pentacíclico natural tipo lupano con una variedad de actividades biológicas que incluyen la inhibición del virus de inmunodeficiencia humana (VIH), así como propiedades antibacterianas, antipalúdicas y antitumorales. Sin embargo, muchos de los genes de la biosíntesis del ácido betulínico no han sido identificados debido a la complejidad de las vías biosintéticas de este grupo de metabolitos. El objetivo del presente trabajo consistió en identificar por homología y análisis filogenético los genes involucrados en la biosíntesis del ácido betulínico a partir de los transcriptomas de los tejidos de hoja y raíz de plantas jóvenes y adultas de *P. andrieuxii*, de donde, se identificaron cinco transcritos de genes implicados en la biosíntesis de ácidos triterpénicos, incluyendo los transcritos de SQS (escualeno sintasa), SQE (escualeno epoxidasa), dos OSC (oxidoescualenos ciclasas) incluyendo bAs (β -amirina sintasa) y LUS (lupeol sintasa), así como monooxigenas de la subfamilia CYP716A que funcionan como oxidasas del C-28 de triterpenos. Asimismo, se evaluaron los niveles de expresión *in silico* y se amplificaron los ORF de los genes identificados con oligonucleótidos específicamente diseñados. Finalmente, se comprobó por secuenciación la identidad de los genes LUS y SQS clonados.

ABSTRACT

Pentalinon andrieuxii is a native plant of the Yucatan peninsula that has been commonly used in traditional Mayan medicine due to its biological properties and represents a natural source of metabolites with potential application in the pharmaceutical industry. Recently, it has been detected that genetically transformed *P. andrieuxii* plants show an increase in their content of betulinic acid, a natural pentacyclic triterpene type lupane with a variety of biological activities including the inhibition of the human immunodeficiency virus (HIV), properties antibacterial, antimalarial and antitumor. However, many of the genes of the biosynthesis of betulinic acid have not been identified due to the complexity of the biosynthetic pathways of this group of metabolites. The objective of the present work consisted in identifying by homology and phylogenetic analysis the genes involved in the biosynthesis of betulinic acid from the transcriptomes of the leaf and root tissues of young and adult plants of *P. andrieuxii*, from which, five transcripts of genes involved in the biosynthesis of triterpenic acids, including the transcripts of SQS (squalene synthase), SQE (squalene epoxidase), two OSC (oxidosqualenes cyclases) including bAs (β -amyrin synthase) and LUS (lupeol synthase), as well as monooxygenases of the CYP716A subfamily that function as oxidases of the C-28 of triterpenes. Likewise, in silico expression levels were evaluated and the ORFs of the identified genes were amplified with specifically designed oligonucleotides. Finally, the identity of the cloned LUS and SQS genes was verified by sequencing.

INTRODUCCIÓN

Los terpenoides se encuentran en todas las especies de plantas, se sugieren que existen al menos 80,000 terpenoides diferentes en la naturaleza muchos de los cuales son de origen vegetal (Buckingham *et al.*, 2015). Muchos de ellos son esenciales para el crecimiento, el desarrollo y el metabolismo general de las plantas, pero también son de interés económico para su desarrollo como medicamentos, nutracéuticos, sabores, fragancias, pigmentos, agroquímicos y desinfectantes. En las últimas tres décadas, la biología molecular, la química y la transcriptómica de la biosíntesis de terpenoides han atraído gran interés (Tang *et al.*, 2019).

Por el valor potencial en la ingeniería metabólica de las vías de biosíntesis, los genes que codifican las enzimas que participan en la biosíntesis de terpenoides se han identificado y caracterizado en varias plantas medicinales como *Salvia miltiorrhiza* (Ma *et al.*, 2012), *Dolioscarpus schottianus* (Oliveira *et al.*, 2002) *Salix* spp. y *Populus* spp. (Palo 1987).

Pentalinon andrieuxii es una planta trepadora nativa de la península de Yucatán reconocida por su producción de metabolitos secundarios de interés farmacéutico, de la cual se ha logrado aislar ácido betulínico, un triterpeno pentacíclico tipo lupano que exhibe una gran variedad de propiedades biológicas que incluyen actividades antibacterianas, de antiinmunodeficiencia humana (VIH) y/o antitumorales (Domínguez *et al.*, 2010). Sin embargo, las características moleculares de los genes y las enzimas involucradas en la biosíntesis de ácido betulínico en esta especie no se ha reportado.

La estrategia de RNA-seq o secuenciación de transcriptomas se ha utilizado ampliamente para investigar e identificar los genes críticos involucrados en la biosíntesis de metabolitos secundarios en plantas como *Camptotheca acuminata*, *Catharanthus roseus*, *Rauvolfia serpentina* (Góngora *et al.*, 2012) y *Artemisia annua* (Soetaert *et al.*, 2013). En este trabajo se pretende identificar los genes relacionados con la biosíntesis de ácido betulínico en *P. andrieuxii*, a partir de un análisis de los transcriptomas de plantas jóvenes y adultas que serán validados a través de análisis de expresión en órganos (hoja y raíz) en distintos estadios (joven y adulta).

CAPÍTULO I

ANTECEDENTES

1.1. *Pentalinon andrieuxii* (Müll. Arg.) BF Hansen & Wunderlin

Pentalinon andrieuxii es una planta trepadora perenne, que puede crecer hasta la parte superior de la copa de los árboles (Figura 1.1).

Es nativa de la península de Yucatán y se distribuye en los estados de Campeche, Chiapas, Hidalgo, Oaxaca, Puebla, Querétaro, Quintana Roo, San Luis Potosí, Tabasco y Veracruz, también se puede encontrar desde el sur de Florida hasta Nicaragua (Juárez *et al.*, 2007), habita en bosques secos, en elevaciones de 0-500 m.s.m. Lo especímenes con flores han sido coleccionados entre abril y agosto (Morales, 2006).



Figura 1.1. *Pentalinon andrieuxii* (Müll. Arg.) B.F. Hansen & Wunderlin.

1.1.2. Clasificación taxonómica

Pentalinon andrieuxii antiguamente era llamada *Urechites andrieuxii* Müll. Arg., Linnaea, más tarde, en 1986, los autores Hansen & Wunderlin (WFO, 2021), demostraron que *Pentalinon* es un nombre más antiguo que *Urechites*, por lo que se le renombró en dicho género (Morales, 2005). Es conocida comúnmente como bejuco guaco, bejuco de la víbora o contrayerba, su clasificación taxonómica se muestra en la Tabla 1.1.

Tabla 1.1 Clasificación taxonómica de *Pentalinon*

Reino	Plantae
Phylum	Magnoliophyta
Clase	Magnoliopsida
Orden	Gentianales
Familia	Apocynaceae
Subfamilia	Apocynoideae
Género	<i>Pentalinon</i>
Especie	<i>Pentalinon andrieuxii</i> (Müll. Arg.) B.F. Hansen & Wunderlin, 1986
Nombres comunes	Bejuco guaco, bejuco de la víbora, contrayerba (español), ya'ax juux, xiiw kaan, tip'te' aak' (maya)

1.1.2. Características taxonómicas

Pentalinon andrieuxii es una planta trepadora leñosa o semileñosa que llega a crecer hasta 6 m de alto; tiene tallos piloso-hispidos, sus hojas son opuestas, su peciolo de 8 a 20 mm de largo, variando su forma de ovada a obovada, de 5 a 10 (12) cm de largo y de 2 a 6 (7) cm de ancho; su inflorescencia es en forma de cima escorpioide arqueada, a menudo con algunas ramificaciones dicotómicas, hasta de 25 cm de largo, llegando a tener hasta 15 flores, pedúnculo hasta de 6 cm de largo; los nectarios son concrecentes, excediendo el largo del ovario; folículos de 20 a 28 cm de largo, sin constricciones notables, de colores cafés; con semillas numerosas, subfusiformes, de aproximadamente 7 mm de largo y 1 mm de ancho, cafés, provistas de un estrecho rostro hasta de 10 mm de largo, en cuyo extremo se ubica un penacho de pelos hasta de 2.5 cm de largo, de color café claro (Figura 1.2) (Rzedowski y Calderón, 1998).

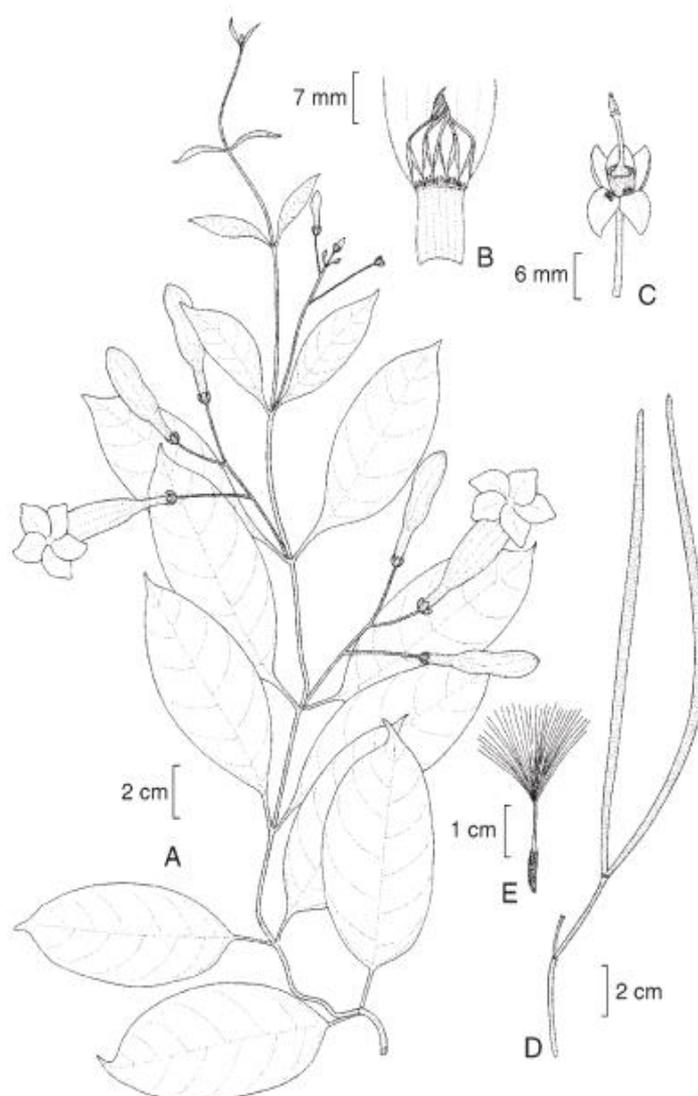


Figura 1.2. Representación esquemática de las características taxonómicas de *Pentalinon andrieuxii* A. Hojas e inflorescencias; B. aspecto del androceo adherido a la corola; C. cáliz; D. vainas; E. semilla. Tomado de Rzedowski y Calderón, 1998

1.1.3. Importancia.

Pentalinon andrieuxii se usa comúnmente en la medicina tradicional maya para el tratamiento de lesiones cutáneas derivadas de la leishmaniasis (úlceras de Chiclero) (Lezama *et al.*, 2014). Los curanderos mayas recomiendan lavar la lesión con una infusión de raíz de *P. andrieuxii* y luego aplicar la raíz en polvo sobre el área afectada, alternativamente, las hojas secas y molidas se pueden aplicar directamente sobre la lesión (Chan *et al.*, 2003). Estudios *in vitro* han demostrado que los extractos solubles en hexano y metanol de las raíces de *P. andrieuxii* mostraron actividad inhibitoria del crecimiento contra varios promastigotes de *Leishmania*,

incluidos L. donovani, L. braziliensis y L. mexicana (Lezama *et al.*, 2007). El compuesto pentalinosterol (PEN), de un extracto soluble en hexano de raíces de *P. andrieuxii* ha mostrado una potente actividad antileishmanial contra promastigotes y amastigotes de *L. mexicana in vitro* (Pan *et al.*, 2012). Estos estudios *in vitro* e *in vivo* han demostrado que los extractos de raíz de *P. andrieuxii* no solo poseen actividad antileishmanial directa, sino que también muestran potentes propiedades inmunomoduladoras, que aumentan notablemente la actividad antiparasitaria de los macrófagos y las células dendríticas (Oghumu *et al.*, 2017).

Por otra parte, se usa para tratar las mordeduras de serpientes y el látex se recomienda para aliviar dolores de cabeza y trastornos nerviosos (Yam-Puc *et al.*, 2012). Además, estudios biológicos previos sobre extractos de esta planta han dado lugar a informes de actividades antiaterogénicas, antiinflamatorias, antileishmaniacas y antidepresivas (Pan *et al.*, 2012).

Un estudio fitoquímico del extracto de la raíz de *P. andrieuxii* reportó el aislamiento y la identificación de dos nuevos trinorsesquiterpenos con una estructura inusual, estos metabolitos se han designado con el nombre de urechitol A y urechitol B, con un esqueleto derivado del guaianólido (Yam-Puc *et al.*, 2009) y se ha reportado el aislamiento de ácido betulínico (Domínguez *et al.*, 2010). Por lo tanto, esta planta representa una fuente natural de metabolitos con aplicación potencial en la industria farmacéutica.

Por otra parte, se ha logrado la inducción de raíces transformadas de *P. andrieuxii* con *Agrobacterium rhizogenes*, cuantificando la producción de ácido betulínico, así como la evaluación del uso de elicitores químicos sobre su producción a escala de matraz, sin embargo, no se observó un aumento en la concentración de ácido betulínico tras haber sido sometido a tratamientos con los elicitores ácido jasmónico y sulfato de vanadilo (Sandoval, 2014), en cambio, con una concentración de 200 μM de jasmonato de metilo se observó un aumento de 28% (0.804 mg/gps de ácido betulínico) en la concentración de ácido betulínico en raíces transformadas de *P. andrieuxii* (De los santos, 2019). Otro trabajo consistió en evaluar la relación entre la edad de plántulas con la metilación global del ADN y la acumulación del ácido betulínico durante el crecimiento de la planta en dos condiciones diferentes (en sombra y media sombra), encontrándose menores niveles de metilación del ADN, menor crecimiento y menor producción de ácido betulínico en plantas con pocos niveles de luz (media sombra) (Almeyda, 2017).

Recientemente se realizó un trabajo para evaluar la relación entre el estado de desarrollo de la planta y la producción de ácido betulínico en *P. andrieuxii*, encontrándose que el contenido de este metabolito se mantuvo constante en el tejido de la hoja durante el desarrollo de la

planta, lo que indica que es un metabolito secundario constitutivo en el órgano, con un posible papel en la defensa de la planta (Hiebert *et al.*, 2016). Recientemente se ha reportado que plantas de *P. andrieuxii* regeneradas a partir de raíces transformadas, mostraron un aumento hasta de 10 veces en su contenido de ácido betulínico (Hiebert *et al.*, 2020).

1.2. Ácido betulínico.

El ácido betulínico (Nombre IUPAC: ácido (3 β)-hidroxi-lup-20(29)-en-28-oico) es un triterpenoide pentacíclico natural de tipo lupano (Figura 1.3) presente en varias especies de plantas, por ejemplo, *Triphyophyllum peltatum*, *Ancistrocladus heyneanus*, *Ziziphi fructus*, *Diospyros leucomelas*, *Tetracera boliviana*, *Syzygium formosanum* y *Doliocarpus schottianus* (Jin *et al.*, 2019; Qian *et al.*, 2011; Oliveira *et al.*, 2002). Es perteneciente a la familia de compuestos llamados isoprenoides (también llamados terpenos o terpenoides), un grupo de metabolitos secundarios con una amplia actividad biológicas. En particular, el ácido betulínico incluye propiedades antirretroviral, antibacterial, antipalúdicas, antiinflamatoria, anticancerígena, antioxidantes y antimalarial (Qian *et al.*, 2011; Yogeewari y Sriram, 2005; Steele *et al.*, 1999). Ha demostrado citotoxicidad selectiva contra varios tipos de tumores específicos, inhibiendo el crecimiento de células cancerosas sin afectar las células normales y la inhibición del virus de inmunodeficiencia humana (VIH) (Einzhammer y Xu, 2014).

Actualmente, el ácido betulínico se obtiene a escala industrial a partir de la extracción de la corteza de abedul o bien mediante modificación química de betulina. Sus principales fuentes se encuentran en plantas de la familia Betulaceae como *Betula alba*, *Betula pendula*, *Corylus avellana* y *Carpinus betulus*, entre otras. El contenido de betulina en la corteza de abedul varía entre 10 y 30%. El color blanco de la corteza de abedul se atribuye a la betulina, que a menudo llena casi todo el interior de las células peridérmicas. La betulina se puede aislar de la corteza de abedul por sublimación o por extracción secuencial con solventes orgánicos (cloroformo, diclorometano, acetona, etanol y otros), repitiendo el proceso un par de veces más con el disolvente orgánico puro. Este compuesto es el componente principal de la corteza de abedul, pero generalmente se acompaña de pequeñas cantidades de lupeol, ácido

betulínico, aldehído betulínico y alobetulina (Patočka, 2003). Por otra parte, se ha logrado extraer ácido betulínico en extracto crudo de hojas de *P. andrieuxii*, demostrando su actividad antiprotozoaria contra *Leishmania amazonensi*, *Leishmania braziliensis*, *Trypanosoma cruzi* Talahuen y *Plasmodium falciparum* (Domínguez *et al.*, 2010).

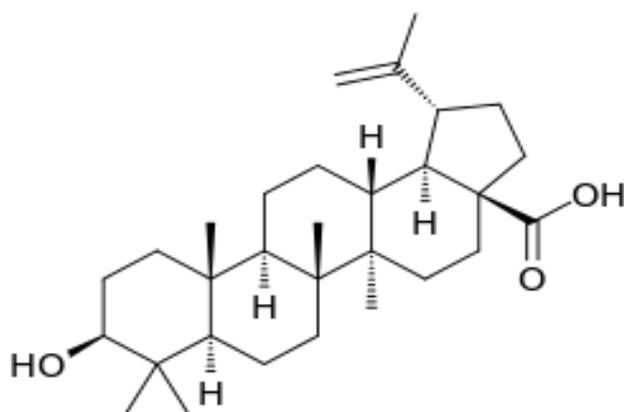


Figura 1.3. Estructura química del ácido betulínico.

1.2.1. Biosíntesis del ácido betulínico.

Los terpenoides o isoprenoides representan uno de los grupos más diversos de productos naturales producidos por las plantas. En ellas, los terpenoides van desde metabolitos primarios esenciales como los esteroides, carotenoides, quinonas y hormonas, que pueden ser empleados para una variedad de funciones básicas en el crecimiento y el desarrollo, pero la mayoría de los terpenoides se usan para interacciones químicas más especializadas y protección en el entorno abiótico y biótico (Tholl, 2015). Los terpenoides son clasificados de acuerdo al número de carbonos que contienen en: Monoterpenos (C10); Sesquiterpenos (C15); Diterpenos (C20); Triterpenos (C30); Tetraterpenos (C40); Politerpenos (más de 45 átomos de carbono) (Kirby y Keasling, 2009). Todos los isoprenoides se derivan del precursor común isopentenil difosfato (IPP) y su isómero estructural dimetilalil difosfato (DMAPP), los cuales son producidos a través de 2 vías diferentes en las plantas: la ruta del mevalonato (MVA) y la ruta del 2- C- metil- d- eritritol 4-fosfato (MEP) (Figura 1.4) (Vranová *et al.*, 2012).

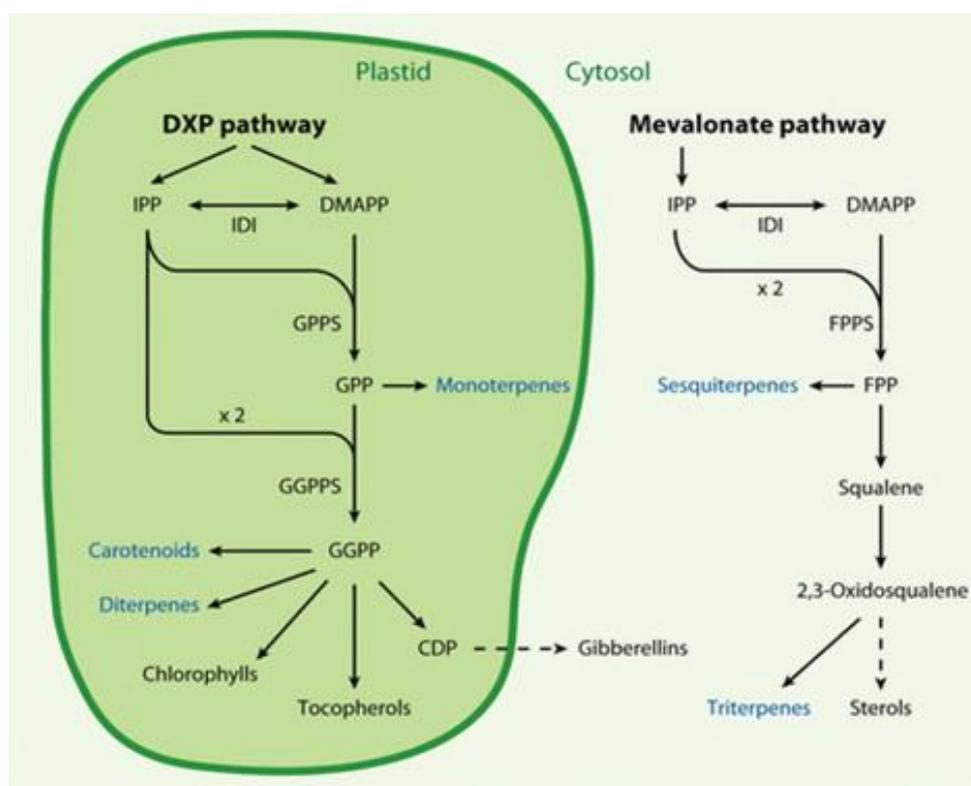


Figura 1.4. Esquema general de la biosíntesis de isoprenoides en células vegetales. DXP, ruta del 2-C-metil-D-eritritol 4-fosfato; IPP, isopentenil difosfato; DMAPP, dimetil difosfato; IDI, isopentenil pirofosfato isomerasa; GPPS, geranil pirofosfato sintasa; GPP, geranil pirofosfato; GGPPS, geranilgeranil pirofosfato sintasa; GGPP, geranilgeranil pirofosfato; CDP, copalil difosfato; FPPS, farnesil difosfato sintasa; FPP, farnesil pirofosfato, (Kirby y Keasling, 2009).

El ácido betulínico se sintetiza a partir de la vía del mevalonato, que ocurre en el citosol y los peroxisomas de las células vegetales, para producir, a partir de la condensación de acetil-CoA, los bloques de construcción de cinco carbonos (IPP y DMAPP), utilizados para fabricar isoprenoides (Vranová *et al.*, 2013). La condensación consecutiva de las unidades IPP y DMAPP conduce a la formación de pirofosfatos prenilados, que son los precursores inmediatos de las diferentes clases de terpenoides. En la formación de los triterpenoides como el ácido betulínico, la molécula de IPP se une con el isómero DMADP para formar geranil pirofosfato (GPP), esta reacción de condensación es catalizada por la geranil difosfato sintasa (GPPS). La molécula de GPP se une con otra molécula de IPP para formar farnesil pirofosfato (FPP), reacción catalizada por la farnesil pirofosfato sintasa. A continuación, la escualeno sintasa (SQS) y el epóxido de escualeno (SQE), catalizan dos reacciones que requieren NADPH para formar escualeno y 2,3-oxidoescualeno, respectivamente. Este último es el precursor común de los triterpenos, que a su vez es ciclado normalmente por oxidoescualeno ciclasas específicas (OSC) hacia estructuras tetra o pentacíclicas, esta ciclación a productos triterpénicos, es una de las reacciones enzimáticas más complejas en la biosíntesis de triterpenos. El 2,3-oxidoescualeno se convierte en lupeol por la lupeol sintasa (LUS), durante este proceso se introducen enlaces internos en la estructura, resultando en la formación de moléculas policíclicas que contienen diversos números de anillos de 5- y 6- miembros. El lupeol se origina a partir de la conformación silla-silla-bote de 2,3-oxidoescualeno a través del catión dammarenilo (Peña *et al.*, 2014).

Finalmente, el ácido betulínico se deriva del lupeol por su oxidación sucesiva en la posición C-28, catalizada por las enzimas de la familia del citocromo P450 y subfamilia CYP716A. Dichas enzimas requieren oxígeno molecular y NADPH para oxidar el sustrato. Se trata de reacciones de monooxigenación en las que sólo uno de los átomos de oxígeno es incorporado en la molécula del sustrato, mientras que el otro es reducido hasta agua (Figura 1.5) (Zhao *et al.*, 2015; Moses *et al.*, 2013; Dubey *et al.*, 2013).

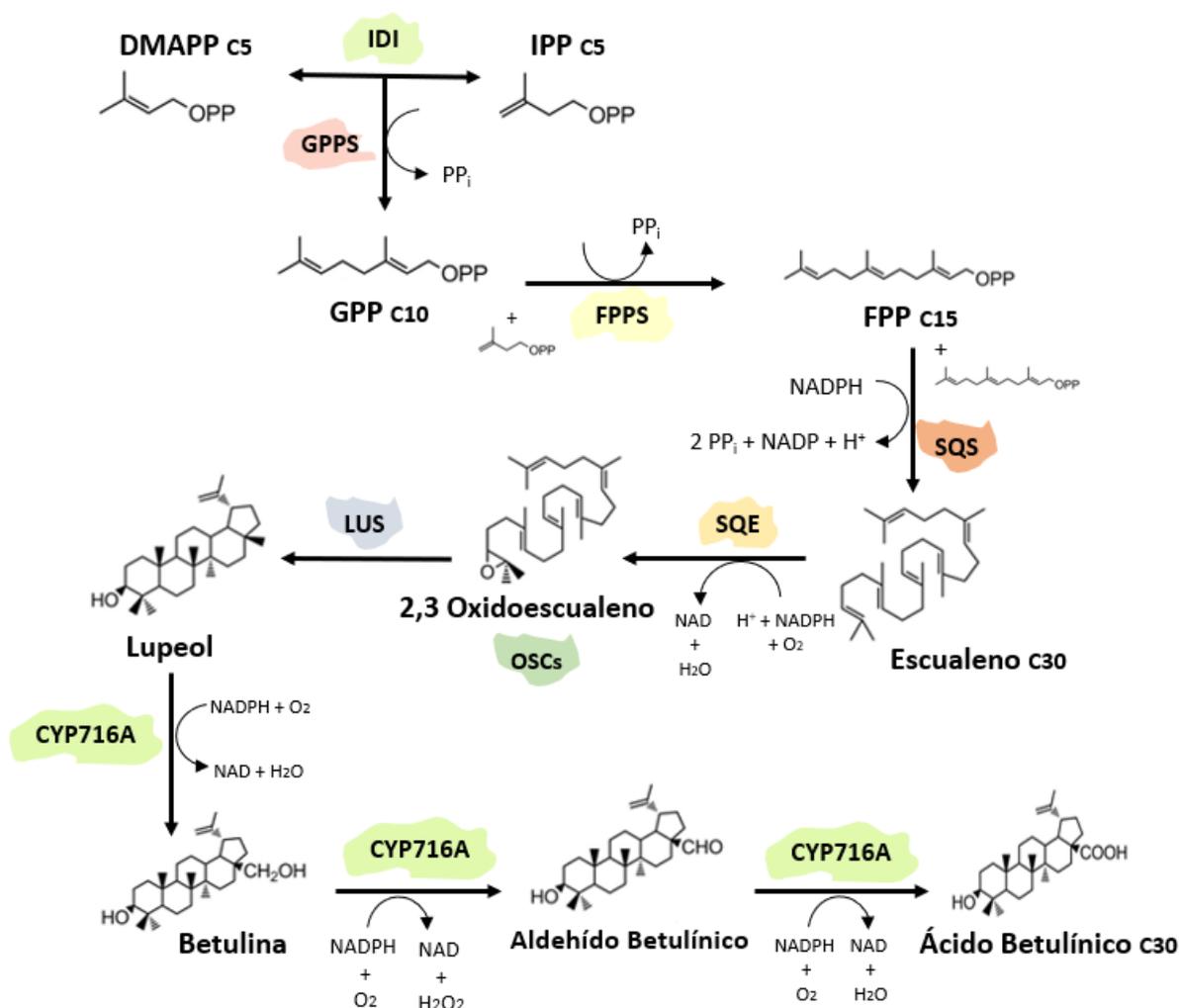


Figura 1.5. Esquema de la biosíntesis del ácido betulínico. DMAPP, dimetil difosfato; IPP, isopentil difosfato; GPP, geranyl pirofosfato; FPP, farnesil pirofosfato; IDI, isopentil pirofosfato Isomerasa; GPPS, geranyl pirofosfato sintasa; FPPS, farnesil difosfato sintasa; SQS, escualeno sintasa; SQE, escualeno epoxidasa; OSCs, oxidoescualeno ciclasas; LUS, lupeol sintasa; CYP716A, subfamilia de citocromo monooxigenasa. (Vranová *et al.*, 2013; Fukushima *et al.*, 2011).

En el ámbito molecular, todos los genes y casi todas las actividades enzimáticas de la biosíntesis de precursores de isoprenoides de plantas se han identificado utilizando mutantes y ensayos de actividad *in vitro*, al menos en la planta modelo *A. thaliana*.

En la mayoría de las especies de plantas, la IDI está codificada por un solo gen. En *A. thaliana* se ha encontrado dos genes IDI, cada uno de los cuales produce dos variantes de proteínas a partir de sitios de inicio de transcripción alternativos en sus promotores (Vranová *et al.*, 2013).

1.3. Triterpenos sintasas.

Entre los metabolitos secundarios de las plantas, los terpenoides son el grupo estructuralmente más diverso, se derivan de unidades de isopreno de cinco carbonos y, por lo tanto, a veces se denominan isoprenoides. Sin embargo, los productos naturales resultantes a menudo se ciclan y algunas veces también se reorganizan, ocasionalmente de una manera tan compleja que confunde el reconocimiento de su origen isoprenoide (Ai-Xia *et al.*, 2007). Gran parte de esta diversidad resulta de las múltiples formas en que las unidades constituyentes de isopreno se pueden unir y luego ciclar y/o reorganizar. Los terpenoides se estratifican más simplemente por el número de unidades de isopreno constituyentes (Gao *et al.*, 2012).

Los triterpenos son una de las clases más grandes de productos naturales de plantas, con más de 20,000 diferentes. La gran mayoría de la diversidad de triterpenos ocurre en el reino vegetal, aunque otros organismos también producen triterpenos (Hill y Connolly, 2015). Los triterpenoides están compuestos por seis unidades C5 (isopentenil difosfato). Sus esqueletos de carbono se sintetizan a partir del precursor común 2,3 - oxidosqualeno por las oxidosqualeno ciclasas (OSC), y estos esqueletos se oxidan aún más por las monooxigenasas (P450s) del citocromo P450 y se glicosilan por glicosiltransferasas dependientes de UDP (UGT), para generar diversos triterpenoides (Yasumoto *et al.*, 2016).

Los triterpenos simples (esteroles) son componentes estructurales importantes de las membranas y también tienen un papel en la señalización (como hormonas esteroideas). En contraste, los triterpenos complejos no se consideran esenciales para el crecimiento y desarrollo normales, y aunque existen en plantas en forma simple no modificada, a menudo se acumulan como conjugados con carbohidratos y otras macromoléculas, especialmente como glucósidos de triterpeno. Los saponinas triterpénicas tienen importantes funciones ecológicas y agronómicas, contribuyendo a la resistencia a plagas y patógenos y a la calidad de los productos de las plantas de cultivo (Shashi *et al.*, 1992). También tienen una amplia gama de aplicaciones comerciales en los sectores de alimentos, cosméticos, farmacéuticos y biotecnología industrial. A pesar de la ya notable diversidad, que existe entre los esqueletos de carbono de los triterpenos, siguen surgiendo nuevas variantes (Moses *et al.*, 2013).

La ciclación de triterpeno puede conducir a una amplia gama de diferentes estructuras, todas derivadas del sustrato isoprenoide lineal simple y ubicuo 2,3-oxidosqualeno. Estos andamios de triterpeno proporcionan la base para una modificación adicional mediante enzimas

modificadoras de triterpeno (por ejemplo, citocromo P450 y transferasas), lo que conduce a una enorme diversidad estructural (Sawai y Saito, 2011).

1.4. Papel de las monooxigenasas CYP450 en la biosíntesis de triterpenos.

Aunque los triterpenos simples como el lupeol son comunes en las plantas, las estructuras de triterpenos a menudo se modifican aún más, a través de reacciones de oxidación, glicosilación y/o acetilación, para producir una diversidad estructural. Las monooxigenasas del citocromo P450 (P450) desempeñan papeles clave en la diversificación estructural de los triterpenoides de las plantas (Tamura *et al.*, 2017). La oxigenación de la estructura mediada por dichas enzimas imparte propiedades funcionales y diversifican el esqueleto básico, por ejemplo, la introducción de grupos oxhidrilo, cetona, aldehído, carboxilo o epoxi. Los grupos funcionales introducidos por tales modificaciones, pueden allanar el camino para una mayor adaptación por enzimas como las transferasas de carbohidratos y las aciltransferasas. Por lo tanto, los P450 juegan un papel clave en la funcionalización de triterpenos. (Nelson y Werck-Reichhart, 2011).

Estas enzimas representan un grupo grande y diverso que se han asignado a 11 clados en función de la similitud de secuencia, dentro de estos clados se dividen en subfamilias. En los últimos años ha habido una actividad considerable sobre los tipos de enzimas P450 involucradas en las modificaciones de triterpeno, de lo que se ha informado de un total de 20 P450 modificadores de triterpeno, 19 de ellos de dicotiledóneas (Thimmappa *et al.*, 2014).

Aunque varios P450 que modifican triterpenoides de diferentes familias representan cinco clanes: CYP51, 71, 72, 85 y 86, estas enzimas permanecen relativamente inexploradas en el reino vegetal. Se ha sugerido que el reclutamiento de P450 para la biosíntesis de triterpenos ocurrió varias veces durante la evolución en diferentes linajes de plantas. La familia CYP716 pertenece al clan CYP85 de P450, esta familia es ubicua y se han encontrado en la mayoría de las plantas terrestres (Sandee *et al.*, 2019). La modificación más común que realizan es la oxidación en tres pasos en la posición C-28 de la β -amirina, la α -amirina y el lupeol, que conduce al ácido oleanólico (tipo oleano), al ácido ursólico (tipo ursano) y al ácido betulínico (tipo lupano), respectivamente (Figura 1.6), que se pueden encontrar en más de 1,600 especies de plantas de más de 140 familias. La subfamilia CYP716A es altamente conservada y funciona principalmente como una oxidasa en el C-28 de triterpenos comunes en las plantas (Miettinen *et al.*, 2017; Tamura *et al.*, 2017).

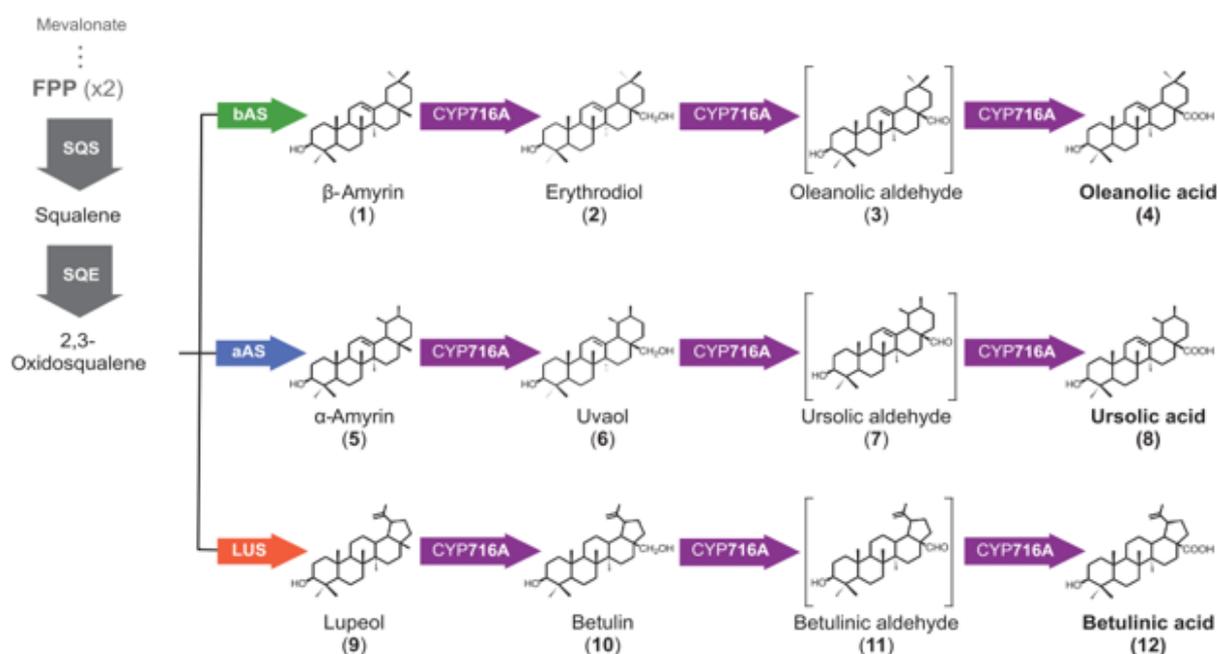


Figura 1.6. Vías para la biosíntesis de ácido oleanólico, ácido ursólico y ácido betulínico catalizados por las monooxigenasas de la familia CYP716A. FPP, farnesil pirofosfato; SQS, escualeno sintasa; SQE, escualeno epoxidasa; bAS, β-amirina sintasa; aAS, α-amirina sintasa; LUS, lupeol sintasa, (Fukushima *et al.*, 2011).

1.5. Análisis de transcriptomas en plantas.

El transcriptoma es el conjunto completo de genes que se están expresando en un momento dado en una célula, tejido u organismo. Los transcriptomas son muy variables, ya que muestran qué genes se están expresando en un momento específico del desarrollo, así su análisis permite estudiar la expresión diferencial en diferentes estados de desarrollo.

Los objetivos claves de la transcriptómica son: catalogar todas las especies de transcripción, incluidos los ARNm, los ARN no codificantes y los ARN pequeños; determinar la estructura transcripcional de los genes, en términos de sus sitios de inicio, extremos 5' y 3', patrones de empalme y otras modificaciones postranscripcionales; y cuantificar los niveles cambiantes de expresión de cada transcrito durante el desarrollo y bajo diferentes condiciones (Wang *et al.*, 2009).

Recientemente, los avances en la tecnología de secuenciación masiva o de alto rendimiento del transcriptoma junto con la bioinformática, han abierto una nueva vía para tales investigaciones. Muchas de ellas emplean etiquetas de secuencias expresadas (EST), secuenciación de transcriptomas (RNA-seq) y técnicas cuantitativas como la PCR en tiempo

real, también conocido como qPCR. La secuenciación de ARN (RNA-Seq) está reemplazando rápidamente a los microarreglos para establecer los perfiles de expresión génica, con una mayor precisión y sensibilidad; esta técnica involucra la conversión del conjunto del transcritos aislados de ADN complementario (ADNc), para después secuenciarlo directamente con un enfoque basado en secuenciación masiva y profunda (Fang *et al.*, 2012; Trapnell *et al.*, 2012).

1.6. Identificación de genes de interés en transcriptoma a través de análisis bioinformáticos.

El desarrollo del RNA-Seq brindó la oportunidad de realizar estudios transcriptómicos detallados, incluso en especies sin un genoma de referencia. El ensamblaje *de novo* en transcriptoma se ha aplicado ampliamente para investigar e identificar los genes críticos involucrados en la biosíntesis de metabolitos secundarios en plantas, e.g. *Camptotheca acuminata*, *Catharanthus roseus*, *Rauvolfia serpentina* (Góngora *et al.*, 2012), *Salvia miltiorrhiza* (Xu *et al.*, 2015), *Cinnamomum camphora* (Chen *et al.*, 2018) y *Panax japonicus* (Zhang *et al.*, 2015). El principal reto para lograrlo es contar con métodos que permitan reconocer y recuperar tales secuencias dentro del transcriptoma. Estos métodos hacen uso a análisis computacionales y estadísticos y de extensas bases de datos para identificar secuencias homólogas que ofrezcan una interpretación funcional hipotética. La disponibilidad de estos datos permite a los usuarios acceder libremente a información sobre genes biosintéticos conocidos que se encuentran en plantas de gran importancia y, lo que es más importante, proporcionar la información necesaria para identificar genes potenciales para los pasos biosintéticos desconocidos restantes para productos de interés (Góngora *et al.*, 2012).

La estrategia más extendida y que mejor se adapta al análisis de datos masivos para validar y anotar los genes de interés es buscar secuencias homólogas en una base de datos. La estrategia para encontrar secuencias homólogas es alinear secuencias de identidad desconocida contra las secuencias depositadas en bases de datos previamente caracterizadas, para evaluar el parecido entre ellas.

1.6.1. Homología de secuencias.

La homología es un concepto fundamental para inferir relaciones biológicas y evolución de genes y proteínas. Es una herramienta fundamental para la biología comparada, especialmente la biología evolutiva (Morrison *et al.*, 2015).

Se infiere homología cuando dos secuencias o estructuras comparten más similitud de lo que cabría esperar por casualidad; cuando existe un exceso de similitud entre dos secuencias, la

explicación más simple es que ambas surgieron de un ancestro común. Sin embargo, las secuencias homólogas no siempre comparten una similitud de secuencia significativa; hay miles de alineaciones de proteínas homólogas que no son significativas, pero son claramente homólogas con base a una similitud estructural estadísticamente significativa o una fuerte similitud de secuencia con una secuencia intermedia. Por lo tanto, cuando una búsqueda de similitud encuentra una coincidencia estadísticamente significativa, se puede inferir con seguridad que las dos secuencias son homólogas; por el contrario si no se encuentra ninguna coincidencia estadísticamente significativa con los contenidos de bases de datos, no se puede asegurar de que no existan homólogos (Pearson, 2013).

Las comparaciones de secuencias están basadas en la identidad, la similitud y la homología entre las secuencias. La identidad se refiere a la suma de residuos idénticos en posiciones equivalentes en dos secuencias alineadas. La similitud de secuencia tiene en cuenta las coincidencias aproximadas y es significativa solo cuando dichas sustituciones se puntúan de acuerdo con alguna medida de "diferencia" o "igualdad", con sustituciones conservadas o muy probablemente asignado puntuaciones más favorables que las no conservadas o improbables. Dichas puntuaciones suelen estar tabuladas e incluir penalizaciones para las inserciones y deleciones (también llamados gaps, porque insertan huecos en el alineamiento). Las tablas de puntuaciones de sustitución de un residuo por otro se denominan matrices de sustitución o de puntuación. Por su parte, el término "homología de secuencia" es el más importante de los tres. Frecuentemente se confunde homología con similitud, pero la homología es cualitativa y solo admite dos estados: o ser homólogo o no serlo, mientras que la similitud es cuantificable. Sin embargo, el grado de similitud entre dos secuencias puede servir para inferir la existencia o no de homología entre ambas (Pertsemlidis y Fondon, 2001).

1.6.2. Alineamiento de secuencias.

La forma de encontrar secuencias homólogas es utilizando bases de datos de secuencias conocidas a través del alineamiento de secuencias y evaluando el parecido entre ellas. Este proceso no es tan sencillo como poner una secuencia encima de otra, dado que las secuencias tienen diferentes longitudes, inserciones, deleciones y sustituciones.

De las diversas herramientas informáticas desarrolladas para realizar esta tarea, la más utilizada es BLAST (Basic Local Alignment Search Tool, por sus siglas en inglés) (Altschul *et al.*, 1990). El algoritmo realiza comparaciones entre pares de secuencias, buscando regiones de similitud local. Para cada comparación calcula el porcentaje de identidad y similitud, y el

el valor "esperado" que indica la probabilidad de que el alineamiento se haya dado al azar (McGinnis y Madden, 2004).

BLAST utiliza el algoritmo Smith-Waterman que se basa en el uso de operaciones de programación dinámica para garantizar que el alineamiento local encontrado es el óptimo con respecto a un determinado sistema de puntuación utilizando matrices tipo BLOSUM (Blocks of Amino Acid Substitution Matrix, por sus siglas en inglés) o PAM (Point Accepted Mutation, por sus siglas en inglés). El uso de este tipo de matrices permite a BLAST dar una puntuación a los alineamientos que realiza. Una matriz de este tipo contiene la puntuación (score) que se le da a un alineamiento particular (Altschul *et al.*, 1990). El algoritmo BLAST tiene el siguiente funcionamiento:

1. Algoritmo heurístico: primero filtra las regiones de baja complejidad y las elimina de la secuencia de consulta (Query). A continuación, divide la secuencia a alinear en subsecuencias (k-meros) de longitud más corta (3 aminoácidos u 11 nucleótidos por defecto) y así se genera una lista de todas las secuencias cortas, o palabras, que componen la secuencia query. Luego, se usa una matriz de puntuación (BLOSUM62, por defecto, para aminoácidos) para determinar todas las palabras coincidentes de alta puntuación para cada palabra en la secuencia de consulta. La lista de coincidencias se reduce tomando solo aquellas que puntuarán por encima de un umbral determinado, llamado "neighborhood score threshold" ("umbral de puntuación del vecindario") (Pertsemlidis y Fondon, 2001).

2. Programación dinámica: BLAST busca en la base de datos de secuencias de destino (Subject) coincidencias exactas con la lista de palabras generadas; cuando encuentra varias subsecuencias en una misma entrada de la base de datos, BLAST intenta extender la alineación de las palabras coincidentes en ambas direcciones siempre que la puntuación continúe aumentando. La extensión termina cuando el valor máximo del "score" disminuye por encima de un umbral determinado. La alineación resultante se denomina par de puntuación alta o HSP.

3. Significación estadística: finalmente calcula el "bit-score" (puntuación de bits) y "E-value" (valor esperado) del alineamiento local extendido que indican la probabilidad de que dicho alineamiento sea fruto del azar, en relación con el tamaño de la base de datos (Kerfeld y Scott, 2011).

La homología se infiere del exceso de similitud, y el exceso de similitud se reconoce a partir de estimaciones estadísticas, las cuales se usan para intentar evitar la inexactitud de las medidas de similitud e identidad. Por su parte, la significación estadística ("E-value" y "bit-

score”) mide la probabilidad de que un alineamiento no sea debido al azar (Kerfeld y Scott, 2011). Si el alineamiento es improbable que sea fruto del azar, entonces ambas secuencias son probablemente homólogas. La puntuación de bits es una puntuación bruta normalizada que representa el tamaño requerido de una base de datos de secuencias, en la que la coincidencia actual podría encontrarse por casualidad (Pearson, 2013).

1.7. Filogenética molecular.

Otra posible ruta para predecir homología de secuencias, incluye un enfoque filogenético para predecir la función molecular. Es una herramienta muy potente para la interpretación de los sistemas biológicos en un contexto evolutivo incorporando a la par información de secuencias homólogas de otras especies ya caracterizadas (Engelhardt *et al.*, 2009). La comparación de secuencias de genes en diferentes organismos, permite agruparlos en familias que corresponden a grupos de genes con cierta similitud, que descienden de un gen ancestral común (genes homólogos); estas relaciones se pueden representar en mayor detalle en forma de un árbol filogenético que captura la estructura jerárquica de las distintas relaciones evolutivas. Un árbol filogenético es simplemente una representación gráfica de posibles relaciones evolutivas inferidas de un conjunto de caracteres homólogos y, por lo tanto, el enfoque principal de la filogenia es en realidad la homología misma (Morrison *et al.*, 2015). Originalmente, el propósito de la mayoría de los árboles filogenéticos moleculares era estimar las relaciones entre las especies representadas por esas secuencias, pero hoy los propósitos se han expandido para incluir la comprensión de las relaciones entre las secuencias en sí mismas sin tener en cuenta la especie ancestral, infiriendo las funciones de genes que no se han estudiado experimentalmente (Hall, 2013).

1.8. Secuenciación y ensamblado *de novo* del transcriptoma de *P. andrieuxii*.

Para la obtención de los transcriptomas, se aisló ARN total de hoja y raíz de plantas adulta y joven de *P. andrieuxii*, para construir 4 bibliotecas de ADNc, que fueron secuenciadas utilizando la plataforma Illumina NextSeq500. La secuenciación se realizó desde ambos extremos de las moléculas de ADN (paired-end). Las muestras se ensamblaron *de novo* con el programa TRINITY (Grabherr *et al.*, 2011) a través de la contratación del servicio comercial de secuenciación masiva en LANGEBIO de Cinvestav Unidad Irapuato.

Al no contar un un genoma de referencia, a partir de los datos de secuenciación, se realizó el pre-procesamiento y filtrado de las lecturas con el programa Trimmomatic (Bolger *et al.*, 2014), con un score de calidad de phred=28, que indica la probabilidad de encontrar un error

de lectura en una base. En la Tabla 1.2 se muestran las lecturas obtenidas para cada una de las bibliotecas construidas.

Tabla 1.2 Número de secuencias obtenidas para cada tejido de *P. andrieuxii*.

Biblioteca	Lecturas totales (R1/R2)
Raíz Joven	206 millones
Hoja Joven	279 millones
Raíz Adulta	240 millones
Hoja Adulta	266 millones

1.8.1. Ensamblado *de novo* de *P. andrieuxii*.

Las muestras presentaron una calidad alta. Además, la remoción de las primeras cinco bases y las últimas 10 bases de todas las lecturas, mejoró notablemente la calidad de las secuencias que posteriormente se usaron para el ensamblaje transcriptómico de las cuatro muestras secuenciadas.

Las lecturas de cada una de las muestras se ensamblaron con el programa TRINITY, teniéndose en consideración la naturaleza del organismo (planta) y los fenómenos de “splicing” alternativo para un óptimo ensamblaje cuyos resultados se muestran en la Tabla 1.3.

Tabla 1.3. Detalles del ensamblaje del transcriptoma *de novo*.

Descripción	Raíz joven	Hoja joven	Raíz adulta	Hoja adulta
% GC	39.59	40.72	39.68	39.94
# Genes	97.696	96,024	62,118	60,522
# Transcritos	210,731	207,685	156,997	147,715

JUSTIFICACIÓN

Aun cuando el conocimiento fitoquímico sobre *P. andrieuxii* ha crecido considerablemente en los últimos años, con informes que describen la producción de metabolitos secundarios como el ácido betulínico, un triterpeno con actividad citotóxica y antiparasitaria, prácticamente no existe información sobre las enzimas encargadas de la biosíntesis de este triterpeno en la especie. Realmente, muchos de los genes que codifican las enzimas involucradas en la biosíntesis no se han identificado, principalmente debido a la complejidad de las vías. Considerando que se ha detectado que plantas genéticamente transformadas de *P. andrieuxii* con *Agrobacterium rhizogenes* que expresan los genes *rol*, muestran un aumento en el contenido de ácido betulínico (Hiebert *et al.*, 2020), surge la necesidad de obtener información acerca de la ruta de biosíntesis de este metabolito en esta especie, lo cual puede lograrse a partir de la identificación, caracterización estructural y la evaluación de los niveles de expresión de los genes potencialmente involucrados con base en los transcriptomas ensamblados de *P. andrieuxii* referidos.

HIPÓTESIS

Si la acumulación del ácido betulínico en plantas de *P. andrieuxii* transformadas, se presenta únicamente en hojas de plantas jóvenes y adultas, los transcritos de los genes implicados en la biosíntesis de dicho metabolito, estarán expresados en los transcriptomas de hojas jóvenes y adultas.

OBJETIVOS

Objetivo general

Identificar los transcritos de los genes potencialmente involucrados en la biosíntesis de ácido betulínico, en los transcriptomas de plantas adultas y jóvenes de *Pentalinon andrieuxii* en los tejidos de hoja y de raíz.

Objetivos específicos

1. Identificar los transcritos de los genes potencialmente involucrados en la biosíntesis de ácido betulínico en los cuatro transcriptomas ensamblados *de novo* de *P. andrieuxii* utilizando estrategias de alineamiento y análisis de secuencias.
2. Caracterizar los genes seleccionados a través de la identificación de motivos, análisis filogenéticos y estimaciones de los niveles de expresión.
3. Corroborar la identidad de los genes seleccionados mediante PCR punto final y clonación.

ESTRATEGIA EXPERIMENTAL

Para poder identificar los genes involucrados en la biosíntesis de ácido betulínico en los transcriptomas ensamblados *de novo* de *P. andrieuxii*, este trabajo partió de una búsqueda por homología utilizando estrategias de alineamiento de secuencias previamente caracterizadas y depositadas en bases de datos contra las secuencias de los transcriptomas. Posteriormente, se realizó un análisis filogenético con el fin de clasificar los transcritos de acuerdo a relaciones evolutivas. Más adelante, se determinaron los patrones de expresión de los genes identificados mediante mapeo y conteo de secuencias a los transcriptomas. A partir de las secuencias seleccionadas, se diseñaron cebadores específicos para corroborar la identidad de los genes por PCR punto final a partir del ARN extraído de cuatro condiciones (plantas tipo silvestre jóvenes y adultas en los tejidos de raíz y hoja). Finalmente, se clonaron los genes con mayor expresión y se secuenciaron para validar la identidad de los genes (Figura 1.7).

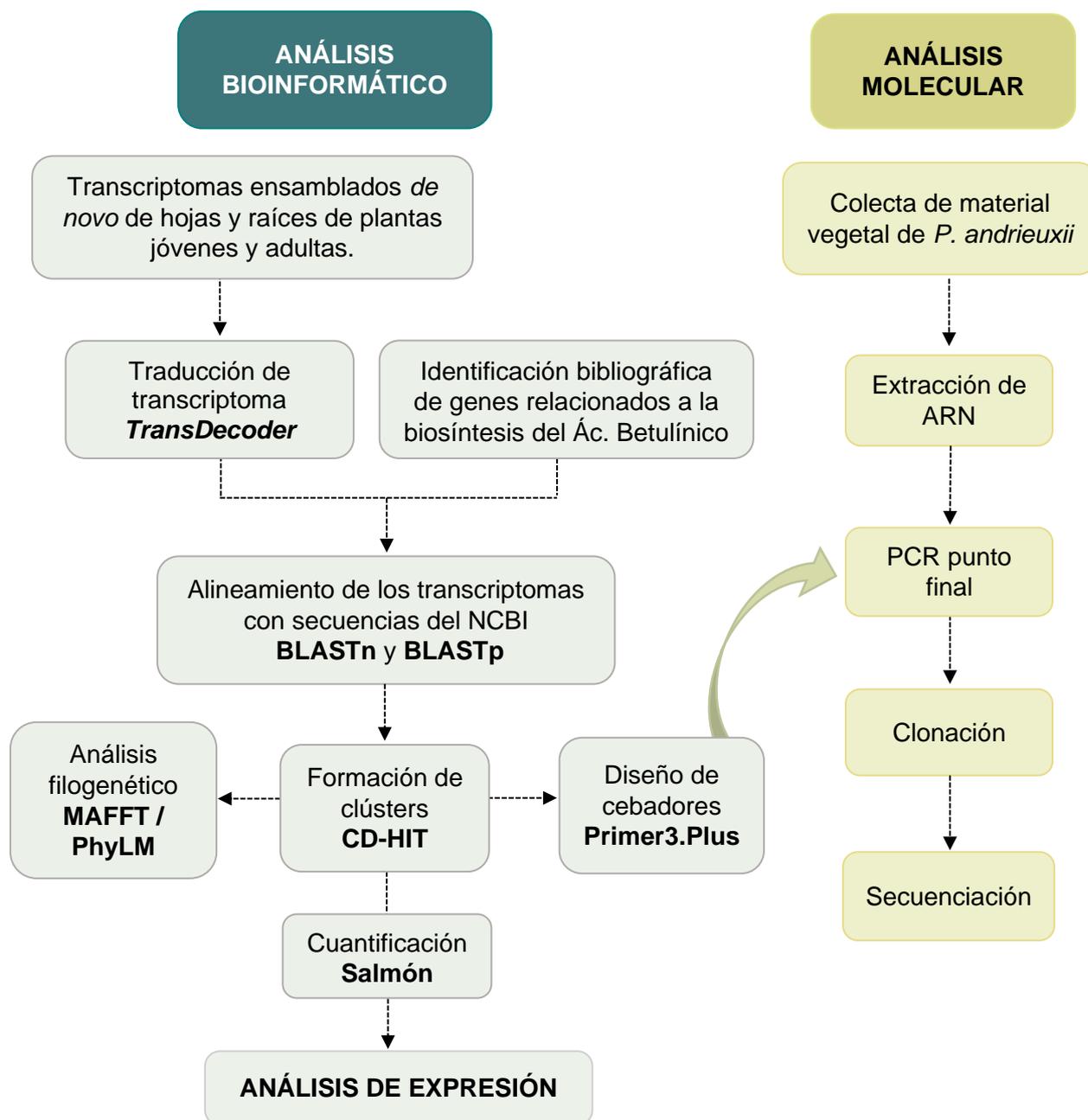


Figura 1.7. Diagrama de la estrategia experimental.

CAPÍTULO II

MATERIALES Y MÉTODOS

2.1. Análisis bioinformático.

2.1.1. Identificación de los genes de biosíntesis del ácido betulínico en los transcriptomas.

Como primer paso para la identificación de los genes en el transcriptoma, se realizó una búsqueda bibliográfica de genes ortólogos de plantas implicadas en la ruta biosintética del ácido betulínico. Posteriormente, las secuencias completas fueron obtenidas a partir de la base de datos Genbank, se descargaron en formato fasta (Tabla 3.1).

Los análisis se realizaron utilizando OS Linux en la distribución UBUNTU. Los 4 transcriptomas de *P. andrieuxii*, secuenciados y ensamblados previamente se tradujeron a proteínas usando el programa TransDecoder (Haas *et al.*, 2013) usando los siguientes comandos:

```
$ TransDecoder.LongOrfs -t transcripts.fasta
```

```
$ TransDecoder.Predict -t transcripts.fasta
```

Las secuencias de aminoácidos se convirtieron en una base de datos Blast usando el comando “makeblastdb”:

```
$ makeblastdb -in transcripts.fasta.transdecoder.pep  
-dbtype prot -out database_name -parse_seqids
```

Se utilizaron los programas BLASTp y BLASTn (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) para alinear las secuencias identificadas en Genbank a los transcriptomas traducidos utilizando la siguiente línea de comandos:

```
$ blastp -query seq.fasta -db database_name -out  
output_file.txt -num_alignments 1 -evalue 1e-10 -outfmt 6  
-num_threads
```

Posteriormente, las isoformas se filtraron con base en los resultados de BLAST usando el programa CD-HIT (Li *et al.*, 2001) (<http://weizhong-lab.ucsd.edu/cd-hit/>) para formar clústers cumpliendo un umbral de similitud definido:

```
$ cd-hit -I db_name -o output_name -c 0.95 -n 5 -d 35
```

Posteriormente, se llevó a cabo un alineamiento múltiple de las secuencias seleccionadas previamente traducidas, contra las secuencias de la Tabla 3.1 (una secuencia por enzima) utilizando la herramienta en línea Clustal Omega (Sievers *et al.*, 2011) (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), con el fin de saber qué tan conservados se encuentran cada uno de los genes entre especies. Una vez identificados los genes, se procedió a diseñar los cebadores para clonación empleando el programa Primer3.Plus (Untergasser *et al.*, 2007) (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>).

2.1.2. Análisis filogenético.

El árbol filogenético se realizó con base en las secuencias de aminoácidos de las enzimas caracterizadas funcionalmente como oxidoescualenos ciclasas recuperadas del GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) y obtenidas de Hayashi *et al.*, 2004, y el árbol de las secuencias de las monooxigenasas CYP450 recuperadas del GenBank y obtenidas de Huang *et al.*, 2019. Dichas secuencias se alinearon con los genes seleccionados de los transcriptomas usando Clustal Omega. Se generaron los árboles filogenéticos utilizando el método de máxima verosimilitud usando 1,000 réplicas de Bootstrap para evaluar la fuerza de los nodos en los árboles usando la herramienta en línea PhyLM 3.0 (Guindon *et al.* 2010) (<http://www.atgc-montpellier.fr/phyml/>).

2.1.3. Análisis de expresión de los genes de biosíntesis del ácido betulínico.

Los 4 transcriptomas fueron analizados con la herramienta FastQC v0.72 (Andrews, 2010) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>) para determinar la calidad de las secuencias, a través de la plataforma web Galaxy (Afgan *et al.*, 2016) (<https://usegalaxy.org/>).

El programa Trimmomatic v0.38.0 (Bolger *et al.*, 2014), se utilizó para el procesamiento para el recorte y filtrado de lecturas. Dentro de esta herramienta se aplicó el recortador Slidingwindow que escanea desde un extremo 5' de la lectura y corta cuando la calidad promedio de un grupo de bases cae por debajo de un umbral específico, en este valor se aplicó un promedio del score de calidad por debajo de 20. Se usó el parámetro MINLEN para

eliminar las lecturas con una longitud menor a 50 nucleótidos. Finalmente, se repitió un análisis con FastQC para verificar el estado de las lecturas procesadas.

El programa Salmon Version 1.3.0 (Patro *et al.*, 2017) se utilizó para la cuantificación de los datos filtrados RNA-seq.

Finalmente, se generó un mapa de calor para visualizar la expresión de los genes biosintéticos del ácido betulínico en los diferentes transcriptomas de *P. andrieuxii*. Se utilizó la herramienta Heatmap2 del paquete R gplots (Version 3.0.1) en Galaxy. El mapa de calor se generó utilizando valores de expresión normalizados, utilizando los valores de expresión en TPM, para esto se realizó una transformación de datos con el método de $\log_2(x+1)$.

2.2. Análisis molecular.

2.2.1. Extracción de ARN total de plantas de *P. andrieuxii*.

A partir de las plantas adultas y jóvenes (6 meses) de *P. Andrieuxii* se colectaron hojas y raíces frescas sin daño aparente, de las cuales 1 g aproximadamente de tejido fueron molidos en morteros con nitrógeno líquido. Fracciones de 0.3g del tejido macerado se transfirió a tubos eppendorf. A cada tubo se le adicionó 700 μ L de amortiguador con CTAB, previamente adicionado con β -mercaptoetanol al 2% y se incubaron a 65 °C por 10 minutos. Entonces se les adicionó 500 μ L de cloroformo: alcohol isoamílico (24:1) y se mezcló perfectamente, después se dejó en reposo por 3 minutos a temperatura ambiente. Las muestras se centrifugaron a 12,000 x g por 10 minutos a 4°C y el sobrenadante se transfirió a un tubo nuevo, este proceso se repitió una vez más adicionando cloroformo:alcohol isoamílico. Después de la segunda recuperación del sobrenadante, se le adicionó un volumen de LiCl 8M y se mezcló perfectamente, esta mezcla se dejó precipitando toda la noche a 4 °C. Al día siguiente, los tubos se centrifugaron a 15,000 x g por 30 minutos a 4 °C. y se desechó el sobrenadante. A la pastilla se le adicionó 500 μ L de amortiguador SSTE (SDS 1%, NaCl 1 M, Tris-HCl 10 mM y EDTA 0.5 M) previamente clarificado, incubándose a 65 °C por 10 minutos. Posteriormente, se les adicionó 500 μ L de cloroformo: alcohol isoamílico (24:1), se homogenizó y se dejó en reposo por 3 minutos a temperatura ambiente. Se centrifugaron a 12,000 x g por 10 minutos a 4 °C, el sobrenadante se recuperó y se le adicionó un volumen de isopropanol. Se dejó precipitar la muestra a 20 °C por 2 horas, se volvió a centrifugar a 15,000 x g por 30 minutos a 4 °C. y se desechó el sobrenadante. La pastilla se lavó con 1 mL de etanol frío al 70% y se centrifugo a 14,000 x g durante 5 minutos. El proceso de centrifugación se repitió, se desechó el etanol por decantación y la pastilla se dejó secar durante 15 minutos a temperatura ambiente, invirtiendo los tubos en papel absorbente. Finalmente, la muestra de ARN se resuspendió en 20 μ L de agua desionizada DEPC y se almacenó a -80 °C.

La calidad del ARN obtenido fue evaluada observando su integridad por electroforesis en TAE (Tris-Acetato-EDTA) 1X DEPC (Dietilpirocarbonato) en un gel de agarosa al 1% a 100 V. El gel se tiñó con 1 μ L de SYBR® Safe DNA gel stain y se observó en un fotodocumentador UV. La cuantificación se realizó mediante el método espectrofotométrico tomando medidas de absorbancia a A260, A280 y A230 en Nanodrop 2000 (Thermo Scientific Inc.).

Con el fin de asegurar que el ARN obtenido esté libre de ADN genómico, se realizó una digestión con DNasa (DNase I, RNase-free, Thermo Scientific, EN0521) y posteriormente se purificó con PureLink® RNA Mini Kit para su posterior análisis.

2.2.2. PCR punto final de los genes implicados en la biosíntesis del ácido betulínico.

Para la síntesis de ADNc se realizaron reacciones de 20 μ l con el kit RevertAid H Minus First Strand cDNA Synthesis (Thermo Scientific, K1632), para ello se agregó en un tubo estéril libre de nucleasas en hielo, 500 ng de cada ARN, 1 μ L de oligo (dT)18 primer y agua libre de nucleasas, hasta un volumen final de 12 μ L, la reacción se incubó a 65°C por 5 minutos, se agregó 4 μ L de 5X reaction Buffer, 1 μ L de inhibidor de RNasa RiboLock (20U/ μ L), 2 μ L de dNTP Mix 10 mM y 1 μ L de la transcriptasa inversa RevertAid H Minus M-MuLV (200 U/ μ L), la reacción se mezcló y centrifugó ligeramente, se incubó a 42°C por 60 minutos, la reacción se terminó calentandola a 70 ° C durante 5 minutos para inactivar la enzima y las muestras se almacenaron a -20 °C.

2.2.3. Aislamiento y clonación del ADNc del gen SQS en *Escherichia coli*.

Con el fin de clonar el gen SQS se realizó una reacción de PCR con el cDNA previamente sintetizado, empleando oligonucleotidos específicos diseñados a partir de la region codificante de los genes relacionados con la biosíntesis del ácido betulínico.

La reacción de amplificación consistió de 2 μ L de Buffer 10X MgCl, 0.4 μ L de dNTP's 10 mM, 4 mM de cebadores, 500 ng de cDNA, 5 U de DreamTaq DNA polymerase y agua libre de nucleasas hasta alcanzar un volumen final de 20 μ L. El programa de amplificación consistió en una activación de la polimerasa a 95°C por 3 minutos, seguida de 31 ciclos de desnaturalización a 95°C por 30 segundos, una hibridación con un gradiente de temperatura (de 50 a 55°C) para determinar la temperatura óptima de alineamiento para cada par de cebadores, de 30 segundos, una extensión a 72°C por 1 minuto, y una extensión final a la misma temperatura por 5 minutos. Los productos amplificados se resolvieron mediante electroforesis en gel de agarosa al 1% teñido con 0.6 μ L de SYBR® Safe y se visualizaron en un transiluminador UV. Los fragmentos amplificados se extrajeron del gel de agarosa cortando las bandas con un bisturí estéril, transfiriendo los pedazos de agarosa a un tubo de microcentrífuga de 1.5 mL, para someterlos a un proceso de purificación con el kit Zymoclean Gel DNA Recovery (Cat. No. D4001), siguiendo las instrucciones del fabricante. Los

productos recuperados se cuantificaron y se verificó su pureza en un NanoDrop 2000 (Thermo Scientific).

Los productos amplificados fueron integrados al vector de clonación pGEM-T Easy (pGEM®-T Easy Vector Systems, Promega, Cat. No. A1360) utilizando una relación molar inserto:vector 5:1. La reacción se llevó a cabo en un tubo Eppendorf de 0.7 mL, añadiendo 5 µL de Buffer de ligación rápida 2X, 1 µL del vector de clonación pGEM-T Easy Vector (50 ng), 1 µL de producto de PCR purificado y 1 µL de la enzima T4 ADN ligase (3 U/ µl), ajustado a un volumen final de 10 µL con agua libre de nucleasas. La mezcla se homogenizó por pipeteo, se incubó a temperatura ambiente por una hora y a 4 °C toda la noche.

Para la transformación de bacterias por electroporación se prepararon células DH5α químicamente competentes con RbCl₂. Para ello, previamente se prepararon los buffers TfbI (RbCl₂ 100 mM, MnCl₂ • 4H₂O 50 mM, acetato de potasio 30 mM, CaCl₂ • 2H₂O 10 mM y glicerol 15% vol., pH 5.8) y TfbII (MOPS 10 mM, RbCl₂ 10 mM, CaCl₂ • 2H₂O 75 mM y glicerol 15% vol., pH 6.8). Se sembraron por estriado las células DH5α de un stock de glicerol congelado en cajas con medio LB y se incubaron a 37°C, durante la noche. Al día siguiente, una colonia se inoculó en 3 mL de medio LB líquido, y se dejó crecer toda la noche a 37°C con agitación vigorosa (200 rpm). El crecimiento de la bacteria se monitoreó hasta alcanzar una DO₆₀₀ a 0.6, se conservó en hielo por 5-10 minutos y se transfirió a dos tubos de centrifuga cónico de 50 mL, se centrifugaron a 4°C por 10 minutos a 4000 rpm y el sobrenadante se desechó de manera cuidadosa, se adicionó 20 mL de buffer TfbI por cada 50 mL de medio y se conservó en hielo por 15 minutos para favorecer su resuspensión, se volvió a centrifugar por 10 minutos a 3000 rpm a 4°C, el sobrenadante se desechó, se adicionó 2.5 mL de buffer TfbII por cada 50 mL de cultivo manteniendo siempre en hielo y resuspendiéndolo cuidadosamente. Se distribuyeron en alícuotas de 50 µL en tubos Eppendorf, se congelaron con nitrógeno líquido y se almacenaron a -80°C.

La transformación se realizó con 2 µL de la reacción de ligación usando 50 µL de células competentes DH5α, la mezcla se mantuvo en hielo por 30 minutos, se sometió a un choque térmico a 42°C por 50 segundos e inmediatamente se trasladó a hielo por 2 minutos. En una campana de flujo laminar se añadieron 950 µL de medio LB líquido sin antibiótico y se incubaron a 37°C por 1.5 horas con una agitación constante a 200 rpm. 100 µL de este cultivo bacteriano se esparcieron en cajas Petri con medio LB sólido suplementado con ampicilina (100 µg/ µL), IPTG (0.5 mM) y X-Gal (80 µg/mL) y las placas se dejaron incubando a 37°C toda la noche.

De las placas de transformación se seleccionaron colonias de color blanco que con ayuda de palillos de madera estériles se inocularon en 5 mL de medio LB líquido con ampicilina (100mg/mL) que se incubó a 37°C con agitación a 200 rpm toda la noche.

De estos cultivos se recuperaron los plásmidos utilizando el kit ZymoPURE Plasmid Miniprep (Cat No. D4209) siguiendo las instrucciones del fabricante.

Al ADN plasmídico se realizó una digestión con las enzimas de restricción EcoRI (FastDigest®) en un volumen final de 20 µL, conteniendo 2 µL de buffer 10X FastDigest® Green, 1 µg de ADN plasmídico y 1 µL de enzima, ajustado el volumen final con agua libre de nucleasas. Las reacciones de digestión se incubaron a 37°C en un bloque de calor durante 5 minutos. Los productos de digestión se resolvieron por electroforesis en gel de agarosa al 1%, teñido con 0.6 µL de SYBR® Safe y se visualizaron en un fotodocumentador ChemiDoc XRS+ de Bio Rad.

El ADN de plásmidos verificados por digestión fue enviado a secuenciar a MacroGen (<https://dna.macrogen.com/#>) en Corea de Sur.

CAPÍTULO III

RESULTADOS

3.1. Análisis bioinformático.

3.1.1. Identificación de los genes involucrados en la biosíntesis del ácido betulínico en los transcriptomas de *P. andrieuxii*.

Se identificaron los genes potencialmente involucrados en la biosíntesis del ácido betulínico por revisión bibliográfica. En la Tabla 3.1 se enlistan los genes seleccionados involucrados en la biosíntesis del ácido betulínico a partir del precursor escualeno y las respectivas especies donde se identificaron. Para la especie escualeno sintasa (SQS) se identificó un gen (*GuSQS1*) en la especie *Glycyrrhiza uralensis*, para la escualeno epoxidasa (SQE) se identificó un gen en la especie *Medicago truncatula* (*SQE2*), en el caso de la lupeol sintasa (LUS) se identificó un gen (*GgLUS1*) en la especie *Glycyrrhiza glabra* que produce ácido betulínico. Por último, para la citocromo monooxigenasa específicamente de la subfamilia CYP716A, se identificó en 8 especies de plantas: *Betula platyphylla*, *Vitis vinífera*, *Glycyrrhiza uralensis*, *Lotus japonicus*, *Lagerstoemia speciosa*, *Lagerstoemia speciosa*, *Avicennia marina*, *Rosamarinus officinalis* y *Medicago truncatula*.

Tabla 3.1. Genes que codifican enzimas involucradas en la vía biosintética del ácido betulínico.

Enzima		Gen	Especie	Referencia
Abreviatura	Nombre			
SQS	Escualeno sintasa	<i>GuSQS1</i> (CAJ77652)	<i>Glycyrrhiza uralensis</i>	Lu <i>et al.</i> , 2008
SQE	Escualeno epoxidasa	<i>SQE2</i> (KEH25842)	<i>Medicago truncatula</i>	Young <i>et al.</i> , 2011
LUS	Lupeol sintasa	<i>GgLUS1</i> (BAD08587)	<i>Glycyrrhiza glabra</i>	Hayashi <i>et al.</i> , 2004
CYP450 Subfamilia CYP716A	Citocromo P450 monooxigenasa	<i>CYP716A180</i>	<i>Betula platyphylla</i>	Zhou <i>et al.</i> , 2016
		<i>CYP716A15</i> (NM_001281186) <i>CYP716A17</i> (NM_001281147)	<i>Vitis vinifera</i>	Fukushima <i>et al.</i> , 2011
		<i>CYP716A179</i> (LC157867)	<i>Glycyrrhiza uralensis</i>	Tamura <i>et al.</i> , 2017
		<i>CYP716A51</i> (AB706297)	<i>Lotus japonicus</i>	Suzuki <i>et al.</i> , 2019
		<i>CYP716A265</i> (LC325360) <i>CYP716A266</i>	<i>Lagerstoemia speciosa</i>	Sandeep <i>et al.</i> , 2019
		<i>CYP716A259</i>	<i>Avicennia marina</i>	Nakamura <i>et al.</i> , 2018
		<i>CYP716A155</i>	<i>Rosamarinus officinalis</i>	Huang <i>et al.</i> , 2019
		<i>CYP716A12</i> (KM978958)	<i>Medicago truncatula</i>	Fukushima <i>et al.</i> , 2011

Las secuencias de aminoácidos completa de las enzimas escualeno sintasa, escualeno epoxidasa, lupeol sintasa y de la familia CYP716A caracterizadas funcionalmente como triterpenoides oxidasas y participantes en la biosíntesis del ácido betulínico se obtuvieron de la base GenBank del NCBI (<http://www.ncbi.nlm.nih.gov/>), descargándolas en formato Fasta para someterlas a comparación con las secuencias del transcriptoma de *P. andrieuxii* con los programas BlastP y BLASTn. En el caso de los genes de las enzimas CYP716A se eligió el mejor alineamiento de los genes previamente identificados de la base de datos. En la Tabla 3.2 se muestran los genes candidatos de los transcriptomas de *P. andrieuxii* resultantes del filtrado de isoformas con base en el resultado de BLASTp y BLASTn. Se identificaron secuencias en los 4 transcriptomas con un porcentaje de similitud e identidad en un rango de 66% a 94%, un porcentaje de cobertura que va de 64% a 99% y el valor esperado de 0.0 para todos los alineamientos. Estos resultados indican que es muy probable que se hayan detectado las secuencias de la biosíntesis del ácido betulínico en los 4 transcriptomas.

Tabla 3.2. Alineamiento de secuencias de los transcriptomas de *P. andrieuxii* contra las secuencias previamente reportadas de la base de datos de Genbank.

Enzima	Gen	Accession	Hoja adulta						Hoja joven					
			Subject	Cover	% Iden.	% Simil.	E-value	Score	Subject	Cover	% Iden.	% Simil.	E-value	Score
ESCUALENO SINTASA	<i>GuSQS1</i>	CAJ77652.1	Seq-DN27839	78%	84%	93%	0.0	613	Seq-DN40610	64%	86%	94%	0.0	510
ESCUALENO EPOXIDASA	<i>SQE2</i>	KEH25842.1	Seq-DN29221	99%	69%	81%	0.0	744	Seq-DN39632	99%	69%	81%	0.0	744
LUPEOL SINTASA	<i>GgLUS1</i>	BAD08587.1	Seq-DN23707	99%	77%	77%	0.0	1020	Seq-DN39113	99%	74%	85%	0.0	1195
CITOCROMO MONOOXIGENASA	<i>CYP716A15</i>	NP_001268115	Seq-DN23331	99%	78%	90%	0.0	792	Seq-DN38732	99%	78%	90%	0.0	791

Enzima	Gen	Accession	Raíz adulta						Raíz joven					
			Subject	Cover	% Iden.	% Simil.	E-value	Score	Subject	Cover	% Iden.	% Simil.	E-value	Score
ESCUALENO SINTASA	<i>GuSQS1</i>	CAJ77652.1	Seq-DN28188	72%	84%	93%	0.0	564	Seq-DN47446	78%	89%	93%	0.0	613
ESCUALENO EPOXIDASA	<i>SQE2</i>	KEH25842.1	Seq-DN29312	86%	80%	87%	0.0	693	Seq-DN46236	96%	66%	79%	0.0	668
LUPEOL SINTASA	<i>GgLUS1</i>	BAD08587.1	Seq-DN25765	99%	70%	77%	0.0	1020	Seq-DN48047	99%	70%	77%	0.0	1020
CITOCROMO MONOOXIGENASA	<i>CYP716A15</i>	NP_001268115	Seq-DN27887	86%	76%	89%	0.0	669	Seq-DN48275	99%	78%	90%	0.0	792

En los cuatro transcriptomas se encontraron todos los genes relacionados a la biosíntesis del ácido betulínico. Sin embargo, se encontraron isoformas en los alineamientos de la familia CYP716A, debido a que el transcriptoma se ensambló por separado, se generaron diferentes nombres para cada gen, por lo que fue necesario analizar cada secuencia manualmente, es decir, eligiendo las secuencias de longitud más larga, de tal manera que se filtraran las isoformas y después someter esos genes al programa CD-HIT para agrupar aquellas secuencias que cumplieran el 95% de similitud.

Las secuencias SQS se agruparon en tres clústers, una de ellas con dos secuencias idénticas derivadas de los transcriptomas de raíz, mientras que las secuencias de los transcriptomas de hoja no se agruparon entre sí. Las secuencias de SQE se agruparon en tres clústers, con dos secuencias idénticas derivadas de los transcriptomas de hojas, mientras que las secuencias de los transcriptomas de raíz no se agruparon entre sí. Las secuencias de LUP formaron dos clústers, una de ellas con tres secuencias idénticas derivadas de los transcriptomas de hoja adulta y de raíces, mientras que la secuencia del transcriptoma de hoja joven no se agrupó con las demás. Por último, las secuencias de la familia CYP716A generó dos clústers, una de ellas con dos secuencias provenientes de los transcriptomas de hoja adulta y raíz joven, y el otro clúster se agruparon secuencias de los transcriptomas de hoja joven y raíz adulta (Tabla 3.3).

Tabla 3.3. Agrupación por clústers de los genes candidatos por CD-HIT.

Secuencia	No. secuencias	No. clústers	No. sec/clus	Clus. por condición	Nombre asignado
SQS	4	3	1-1-2	(HA) (HJ) (RA, RJ)	PaSQS
SQE	4	3	2-1-1	(HA,HJ) (RA) (RJ)	PaSQE
LUP	4	2	3-1	(HA, RA, RJ) (HJ)	PaLUP
CYP716A	4	2	2-2	(HA, RJ) (HJ, RA)	PaCYP

Se realizó un análisis de alineamiento múltiple para determinar si existían diferencias entre las secuencias. Los resultados arrojaron que el ensamblado de transcritos parciales no permitieron el agrupamiento entre ellas. En la Tabla 3.4 se muestran el tamaño de los marcos de lecturas de cada secuencia y la longitud de aminoácidos. Los marcos de lecturas van en un rango de 248 a 437 en aminoácidos para las secuencias de SQS siendo la secuencia derivada del transcriptoma de hoja joven la única secuencia parcial. Las secuencias de SQE

van de un rango de 479 a 531 en aminoácidos, las secuencias de LUP en un rango de 756 a 761 aminoácidos, y las secuencias de la subfamilia CYP716A de 372 a 478 aminoácidos.

Tabla 3.4. Longitud en pares de bases (pb) y en aminoácidos (aá) de los marcos de lecturas de los transcritos de los cuatro genes seleccionados en los cuatro transcriptomas de *P. andrieuxii*.

Gen-tejido	ORF (pb)	Longitud (aá)	Gen-tejido	ORF (pb)	Longitud (aá)
SQS-HA	1,314	437	LUP-HA	2,286	761
SQS-HJ	745	248	LUP-HJ	2,271	756
SQS-RA	1,245	414	LUP-RA	2,286	761
SQS-RJ	1,242	414	LUP-RJ	2,286	761
SQE-HA	1,531	509	CYP-HA	1,437	478
SQE-HJ	1,423	474	CYP-HJ	1,311	436
SQE-RA	1,569	531	CYP-RA	1,117	372
SQE-RJ	1595	522	CYP-RJ	1239	412

A pesar de que se encontraron algunos transcritos parciales, al realizar el alineamiento múltiple entre las secuencias recuperadas del transcriptoma y las secuencias obtenidas del NCBI, se observó que todas las secuencias contienen los motivos de aminoácidos conservados.

En el alineamiento de las secuencias de aminoácidos de las escualeno sintasas obtenidas y la de *Glycyrrhiza uralensis* caracterizada previamente, se observan 5 segmentos altamente conservados incluso en las enzimas vegetales (I- V; Figura 3.1), que pueden estar implicados en el sitio catalítico. El dominio I es el sitio de unión química que consta de aproximadamente 18 aminoácidos distribuidos por todo el polipéptido. El dominio II es el sitio de unión al sustrato - Mg²⁺, que consta de 10 aminoácidos, en el cual está incrustado el dominio V. El dominio III es el sitio activo, que está compuesto por aproximadamente nueve residuos de aminoácidos. El dominio IV es sitio catalítico de aproximadamente 14 aminoácidos distribuidos por todo el polipéptido. Este sitio está compuesto por la gran cavidad central formada por hélices alfa antiparalelas con dos regiones ricas en aspartato (DXXXD) en paredes opuestas (Hata *et al.*, 1997; Rong *et al.*, 2016).

En el caso de las secuencias SQE, los residuos de aminoácidos que se muestran en recuadro rojo representan el sitio de unión al sustrato (escualeno) (Noguera *et al.*, 2018), alineados

con la secuencia de *Medicago truncatula*, lo que indica que es probable que los genes SQE de *P. andrieuxii* codifiquen proteínas con función de escualeno epoxidasa (Figura 3.2).

De la misma manera, para las secuencias OSC's los motivos de aminoácidos incluyeron un DCTAE altamente conservado, que participa en la unión del 2,3-oxidoscualeno y la protonación del grupo epóxido. Se ha indicado que el extremo N-terminal es la región más importante para la especificidad entre lupeol sintasa y β -amirina sintasa; los residuos de triptófano y leucina en el motivo Met-Trp-Cys-Tyr-Cys-Arg (MWCYCR) de β -amirina sintasa se han encontrado en la planta *Panax ginseng* y el motivo Met-Leu-Cys-Tyr-Cys-Arg (MLCYCR) de lupeol sintasa en *Olea europaea*. Por lo tanto, este motivo se ha identificado para la diferenciación de productos (Souza, et al., 2016).

En el alineamiento se pudo detectar que en las secuencias PaLUP-HA, RA y RJ presentan el motivo MWCYCR a diferencia de la OSC de *Glycyrrhiza glabra* (GgLUP) previamente caracterizada, que presenta el motivo MLCYCR. Por lo tanto, PaLUP-HA, RA y RJ codifican la proteína de β -amirina sintasa, debido a la presencia del aminoácido triptofano (W) en lugar de presentar el residuo de leucina (L), por otro lado, la secuencia PaLUP-HJ, si coincidió con el motivo de GgLUP, de modo que codifica la proteína lupeol sintasa (Figura 3.3)

Por último, las secuencias de CYP716A de *P. andrieuxii* con los motivos conservados típicos de los P450 auténticos, incluido el motivo FxxGxRxCxG para la unión al hemo, la tríada ERR (con motivos de ácido glutámico y arginina de EXXR y arginina de PERF), para mantener la estructura terciaria de la proteína y la posición del grupo hemo, así como el motivo A/GGXD/ET para la unión y activación de oxígeno, se alinearon con las CYP de *Vitis vinifera* (Fig 3.4).

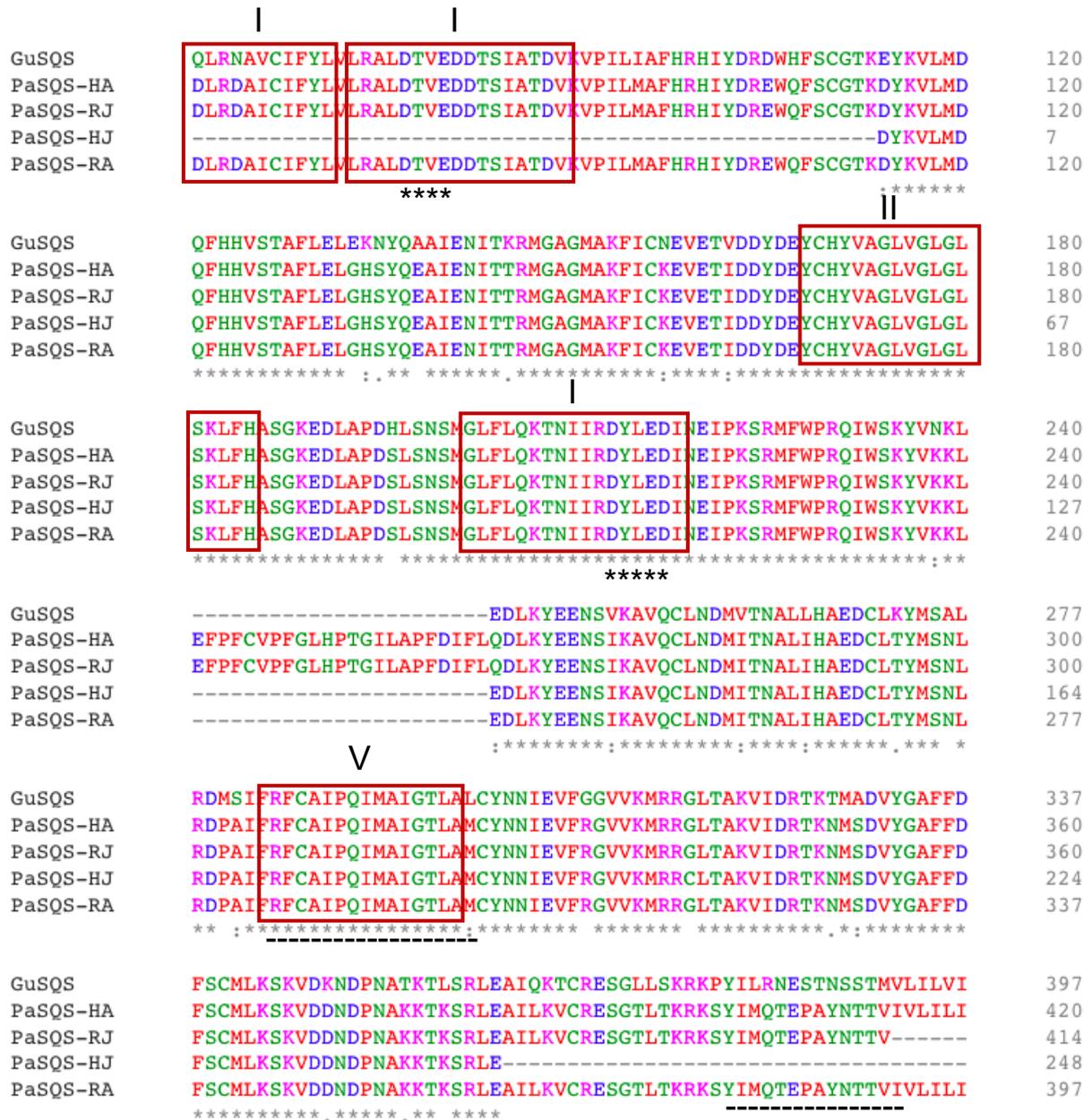


Figura 3.1. Alineamiento múltiple entre las SQS identificadas en el transcriptoma del *P. andrieuxii* y la SQS obtenidas del NCBI. PaSOS-HA secuencia de hoja adulta, PaSOS-HJ secuencia de hoja joven, PaSOS-RA secuencia de raíz adulta, PaSOS-RJ secuencia de raíz joven, por otro lado, GuSOS incluye la secuencia de SQS de *Glycyrrhiza uralensis* previamente caracterizada. I-V: Dominios altamente conservados de escualenos sintasas; *****: Dos regiones ricas en aspartato (DXXXD); "----": dos dominios transmembrana. **Morado:** residuos polares básicos, con carga positiva. **Azul:** residuos polares ácidos con carga negativa. **Verde:** residuos polares neutros. **Rojo:** residuos apolares.

Sitio de unión al sustrato

MtSQE	RKSTSSINTLAGLTYTICIASPDPAKEMREACFNYSLGGVFSDEPIALLAGLNSSSS	458
PaSQE-RJ	RKPVASTINTLAGALYKVFSAINDQARNELRQACFDYLSLGGIFAYGPVLLSGLNPQPM	468
PaSQE-HA	RKPVASTINTLAGALYQVFCASPDPAQEMRQACFDYLSLGGIFASGPVALLSGLNPKPL	439
PaSQE-HJ	RKPVASTINTLAGALYKVFSAINDQARKEMRQACFDYLSLGGVCSSTGPIALLSGLNPRPL	458
PaSQE-RA	RKPVASTINTLAGALYKVFSSSPDQARKEMRQACFDYLSLGGVFSGGPVLLSGLNPRPL	453

** .:***** ** : : * **:*:*:***:*****: : *::**:**

Figura 3.2. Alineamiento múltiple entre las SQE identificadas en el transcriptoma de *P. andrieuxii* y la SQE obtenidas del NCBI. PaSQE-HA secuencia de hoja adulta, PaSQE-HJ secuencia de hoja joven, PaSQE-RA secuencia de raíz adulta, PaSQE-RJ secuencia de raíz joven, por otro lado, MtSQE incluye la secuencia de SQE de *Medicago truncatula* previamente caracterizada. **Morado:** residuos polares básicos, con carga positiva. **Azul:** residuos polares ácidos con carga negativa. **Verde:** residuos polares neutros. **Rojo:** residuos apolares.

MWCYCR

PaLUP-HA	MPPEFWILPSFVPMHPARMWCYCRMVYMPMSYLYGKRFVGPITPLVRQLREELYLEPYDQ	297
PaLUP-RA	MPPEFWILPSFVPMHPARMWCYCRMVYMPMSYLYGKRFVGPITPLVRQLREELYLEPYDQ	297
PaLUP-RJ	MPPEFWILPSFVPMHPARMWCYCRMVYMPMSYLYGKRFVGPITPLVRQLREELYLEPYDQ	297
GgLUP	LPPELWLLPKFAPFHPGMLCYCRMVYMPMSYLYGKRFVGPITPLVRLREELYNEPYNQ	295
PaLUP-HJ	MPPEFWLLPKLSPIHPGMLCYCRMVYMPMSYLYGKRFVGPITELVQSLRQEIYVQPYNK	296

:***:*:*:*.: *:*:* *****:*****:***** *::.*:*:* :***:

DCTAE

PaLUP-HA	HGWQVSDCTAFLGKCCLLFSTMPPEIVGEKLEPERLCDAVNVLLSLQSKNGGLAAWEPAG	537
PaLUP-RA	HGWQVSDCTAFLGKCCLLFSTMPPEIVGEKLEPERLCDAVNVLLSLQSKNGGLAAWEPAG	537
PaLUP-RJ	HGWQVSDCTAFLGKCCLLFSTMPPEIVGEKLEPERLCDAVNVLLSLQSKNGGLAAWEPAG	537
GgLUP	HGWQVSDCTAFLKAAALLSEMPSELVGGKMETERYDAVNVILSLQSSNGGFPAWEPQK	535
PaLUP-HJ	HGWQVSDCTAFLGKCALLFARMPPPELFGEKTETERFYDAVNVILSLQSPNGGFPAWEPQR	536

*****_**:: ** *:* * * * * : *****:***** ** : *****

Figura 3.3. Alineamiento múltiple entre las OSC's identificadas en el transcriptoma de *P. andrieuxii* y OSC obtenida del NCBI. PaLUP-HA secuencia de hoja adulta, PaLUP-HJ secuencia de hoja joven, PaLUP-RA secuencia de raíz adulta, PaLUP-RJ secuencia de raíz joven, por otro lado, GgLUP incluye la secuencia de LUP de *Glycyrrhiza glabra* previamente caracterizada. **Morado:** residuos polares básicos, con carga positiva. **Azul:** residuos polares ácidos con carga negativa. **Verde:** residuos polares neutros. **Rojo:** residuos apolares.

3.1.2. Análisis filogenético.

Con el fin de establecer la posible relación evolutiva de las enzimas SQS, SQE, LUP y CYP de *P. andrieuxii*, se realizaron árboles filogenéticos basados en enzimas triterpénicas y monooxigenasas CYP modificadoras de triterpenos previamente caracterizadas contra las secuencias putativas de *P. andrieuxii* identificadas en los transcriptomas. Se generaron alineamientos múltiples de secuencias usando el programa CLUSTAL W (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

Las enzimas de biosíntesis de triterpenos se pudieron relacionar en los clados correspondientes, excepto tres secuencias: LUPRJ, LUPRA y LUPHA que se agruparon junto con las OSC que forman andamios de triterpenos pentacíclicos de β -amirina. Asimismo, LUPHJ se localizó en el clado de las OSC que producen lupeol (Figura 3.5).

El árbol filogenético de las monooxigenasas reveló que las posibles monooxigenasas se agruparon en los clados relacionados con las CYP pertenecientes al clan CYP85, justo donde pertenece la familia de citocromos CYP716A, sin embargo, inesperadamente no se relacionaron con las monooxigenasas que pueden oxidar el lupeol en ácido betulínico (Figura 3.6).

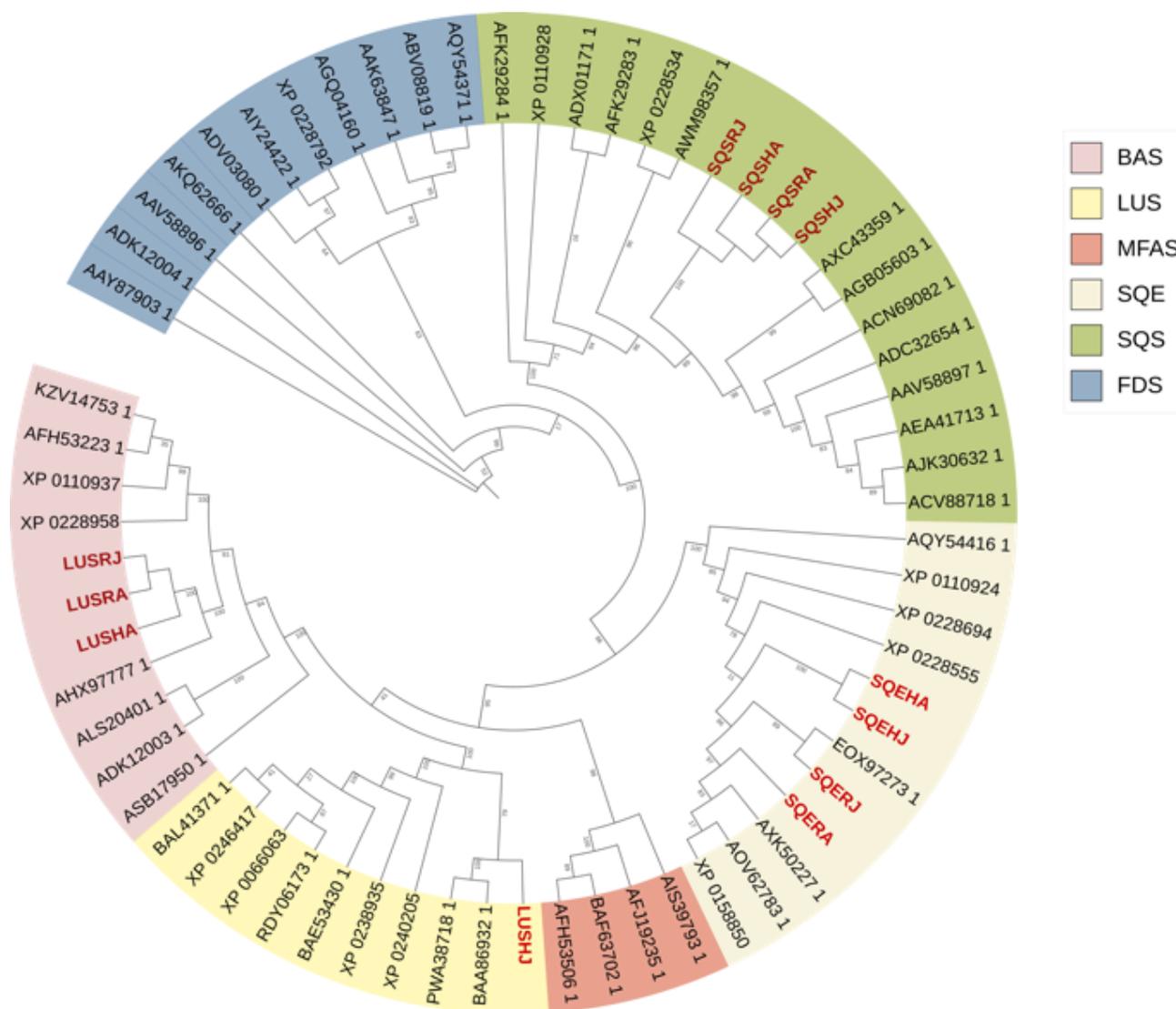


Figura 3.5. Árbol filogenético por el método de máxima verosimilitud con el programa PhyLM (Guindon *et al.* 2010). Relaciones entre las secuencias SQS, SQE Y LUP del transcriptoma y los genes recuperados del NCBI obtenidas de Aminfar *et al.*, 2019. Las secuencias del transcriptoma se representan en color rojo (LUSRJ: lupeol sintasa de raíz joven, LUSRA: raíz adulta, LUSHA: hoja adulta, LUSHJ: hoja joven, SQERA: escualeno epoxidasa de raíz adulta, SQERJ: raíz joven, SQEHJ: hoja joven, SQEHA: hoja adulta, SQSHJ: escualeno sintasa de hoja joven, SQSRA: raíz adulta, SQSHA: hoja adulta, SQSRJ: raíz joven). BAS: β -amirina sintasa, LUS: lupeol sintasa, MFAS: amirina sintasa de función múltiple, SQE: escualeno epoxidasa, SQS: escualeno sintasa, FDS: farnesil sintasa.

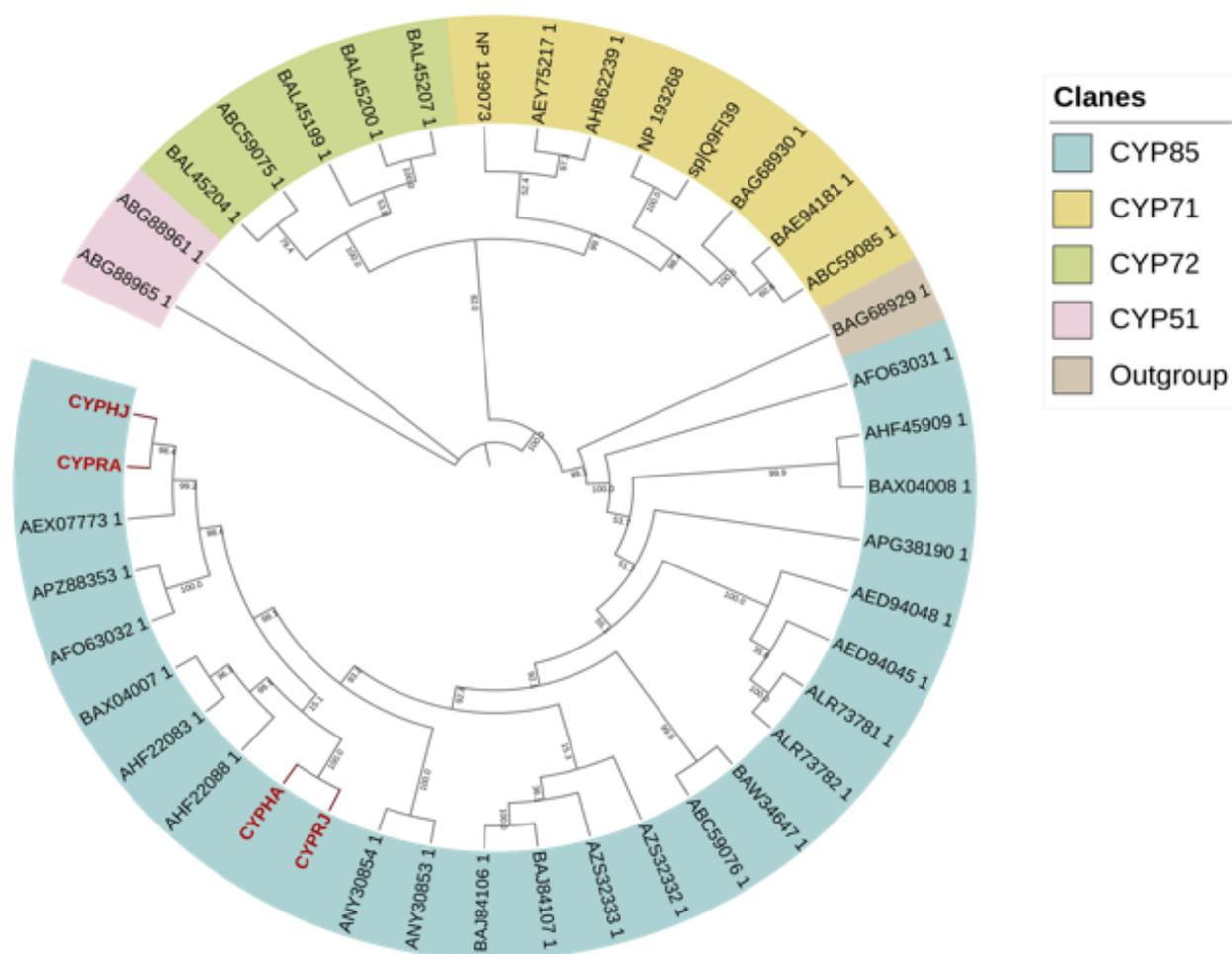


Figura 3.6. Análisis filogenético por método de máxima verosimilitud con el programa PhyLM (Guindon *et al.* 2010). Relaciones entre las secuencias CYP del transcriptoma y los genes previamente caracterizados recuperados del NCBI obtenidas de Huang *et al.*, 2019. Las secuencias marcadas con flecha roja incluyen los CYP que pueden oxidar el lupeol en AB, los CYP identificados en *P. andrieuxii* en este estudio se representan en los recuadros de color rojo. Los CYP en recuadro azul, amarillo, verde y rosado pertenecen a los clados CYP85, CYP71, CYP72 y CYP51, respectivamente. CYPHJ: Hoja joven, CYPRA: Raíz adulta, CYPHA: Hoja adulta, CYPRJ: Raíz joven. Se utilizó BAG68929.1 como grupo externo.

A partir de las secuencias consenso de los clústers y las agrupaciones formadas en los árboles filogenéticos se procedió a diseñar cebadores para la amplificación por PCR punto final. Para el gen SQS se diseñaron tres pares de cebadores, dos de ellos nombrados como S1 y S2 para las secuencias derivadas de los transcriptomas de hoja adulta y hoja joven, respectivamente, y un solo cebador para las secuencias correspondientes a los transcriptomas de raíz adulta y raíz joven (S3). Para el gen SQE se diseñó un par de cebadores para las secuencias de los transcriptomas de hoja adulta y hoja joven (E1) y dos pares de cebadores para las secuencias de raíz adulta (E2) y raíz joven (E3). Para el gen LUP se diseñó un par de cebadores para las secuencias de los transcriptomas de hoja adulta, raíz adulta y raíz joven (L1) y un cebador para la secuencia correspondiente al transcriptoma de hoja joven (L2). Por último, para el gen CYP se diseñaron dos pares de cebadores, el primer par (C1) derivado de las secuencias de los transcriptomas de hoja adulta y raíz joven y el segundo par de cebadores de los transcriptomas de hoja joven y raíz adulta (C2) (Tabla 3.5).

Para el diseño de los cebadores, se extrajo la secuencia codificante (CDS) del marco de lectura más larga con la ayuda del programa ORFfinder del NCBI (<https://www.ncbi.nlm.nih.gov/orffinder/>), las secuencias se sometieron al programa Primer3.Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) optando por un tamaño de 20 bases para cada cebador. Se seleccionaron aquellos cebadores que amplificarán a una temperatura de alineación entre 55°C a 60°C y con un porcentaje de GC entre 40% a 60%.

Tabla 3.5. Secuencias de cebadores para PCR punto final diseñados a partir de la región codificante de los genes seleccionados.

Cebadores ORF			
Gen-Tejido	Nombre del cebador	“Forward” (5'→3')	“Reverse” (5'→3')
SQS-HA	S1	ATGGGTTTGTCTACAGAA	TCAAATATCATTTTGCCGTT
SQS-HJ	S2	ATGGACCAATTCCACCATGT	AAACTAAGAGCAGGCTTGAA
SQS-RA SQS-RJ	S3	ATGGGGAGCTTCGGCGCCAT	TGTTCCAATGGCCATGATCT
SQE-HA SQE-HJ	E1	ATGGTGGTAATAGATCAGTA	TTATGCCTCAGCAGGTGGAG
SQE-RA	E2	ATGGATCACTACATTGTCGG	TCATTTAGCAGGAGGTCCTC
SQE-RJ	E3	ATGGCTTTGCGCCATGGATTT	CTAAAGTATGCTGTCGATGG
LUP-HA LUP-RA LUP-RJ	L1	ATGTGGAGGTTGAAGGTTGC	TTAGACATTTTATAGATGGCA
LUP-HJ	L2	ATGTGGAAGCTGAAGATTGC	TTATGCCAGAAGAACATGTC
CYP-HA CYP-RJ	C1	ATGGAGTTCTTCTATGTCTC	TTATGCTTTATGAGGAAAGA
CYP-HJ CYP-RA	C2	ATGGTCGGGGAGAGTCTTCA	TCATGCTTTGTGGGGATAAA

3.1.4. Análisis de expresión de los genes involucrados en la biosíntesis del ácido betulínico.

Los datos filtrados de RNA-seq de cada transcriptoma secuenciado fueron alineados a su respectivo transcriptoma ensamblado para estimar los niveles de expresión con la herramienta Salmon (Versión 1.5.1) (<https://combine-lab.github.io/salmon/>), como archivo de salida del programa se generó una matriz de datos con el conteo de las lecturas para cada gen, de los cuales se seleccionaron aquellos para el análisis posterior. Dicha matriz contiene dos columnas que representaban el nombre del transcrito proporcionada en el archivo de referencia y la estimación de la abundancia relativa de la transcripción en unidades de transcripciones por millón (TPM). Para el gen SQS, los valores de expresión van de un rango de 25.47 a 408 TPM en los cuatro transcriptomas, teniendo el valor más alto en el transcriptoma de hoja joven y más bajo en el transcriptoma de raíz adulta. El gen SQE presentó un valor de expresión en un rango de 1.87 a 18.95 TPM, teniendo el valor más alto nuevamente en el transcriptoma de hoja joven. Para el gen LUP va de un rango de expresión de 1.27 a 11.11 TPM, y por último, los transcritos de los genes CYP dan un rango de expresión de 0.29 a 38.40 TPM (Tabla 3.6).

En la Figura 3.7 se presenta un mapa de calor general a partir de la matriz de los conteos normalizados (valores de la abundancia relativa), para cada uno de los transcritos evaluados en de cada condición, para esto se realizó una transformación de datos con el método de $\log_2(x + 1)$. En él, se muestra una relación entre los valores de expresión génica y el tejido/inductor específico. Las agrupación jerárquica revela el agrupamiento a nivel de columnas en dos grupos (hoja adulta con hoja joven y raíz adulta con raíz joven). El tejido con mayor expresión de los genes fue en la hoja joven seguido de la raíz adulta. Se observa claramente que SQS fue el gen que presentó mayor expresión de los 4 genes analizados, predominando en el transcriptoma de hoja joven y presentando cantidades mínimas en raíz adulta. Por su parte, SQE mostró poca expresión en raíz joven y de la misma forma que SQS, predominó en hoja joven. Los transcritos de las OSC (LUP) mostraron expresiones superiores en ambos tejidos de plantas jóvenes, que en los tejidos de plantas adultas. Finalmente, los genes con menor expresión fueron las CYP, con cantidades insignificantes en hoja adulta y con mayor expresión en raíces, tanto adulta como joven. En general, los genes para la biosíntesis de ácido betulínico muestran un patrón de expresión espacio-temporal (tejido particular / etapa de desarrollo) (Figura 3.7).

Tabla 3.6. Matriz del conteo de lecturas de cada gen en los transcriptomas de *P. andrieuxii*.
 TPM: Transcrito por millón, unidad para medir la expresión de cada transcrito.

Tejido	Gen	Nombre	TPM
HA	SQS	Seq-DN27839	32.43
	SQE	Seq-DN29221	12.14
	LUP	Seq-DN23707	1.27
	CYP	Seq-DN23331	0.29
HJ	SQS	Seq-DN40610	408.02
	SQE	Seq-DN39632	18.95
	LUP	Seq-DN39113	11.11
	CYP	Seq-DN38732	5.62
RA	SQS	Seq-DN28188	25.47
	SQE	Seq-DN29312	6.28
	LUP	Seq-DN25765	2.97
	CYP	Seq-DN27887	38.40
RJ	SQS	Seq-DN47446	48.72
	SQE	Seq-DN46236	1.87
	LUP	Seq-DN48047	7.59
	CYP	Seq-DN48275	22.45

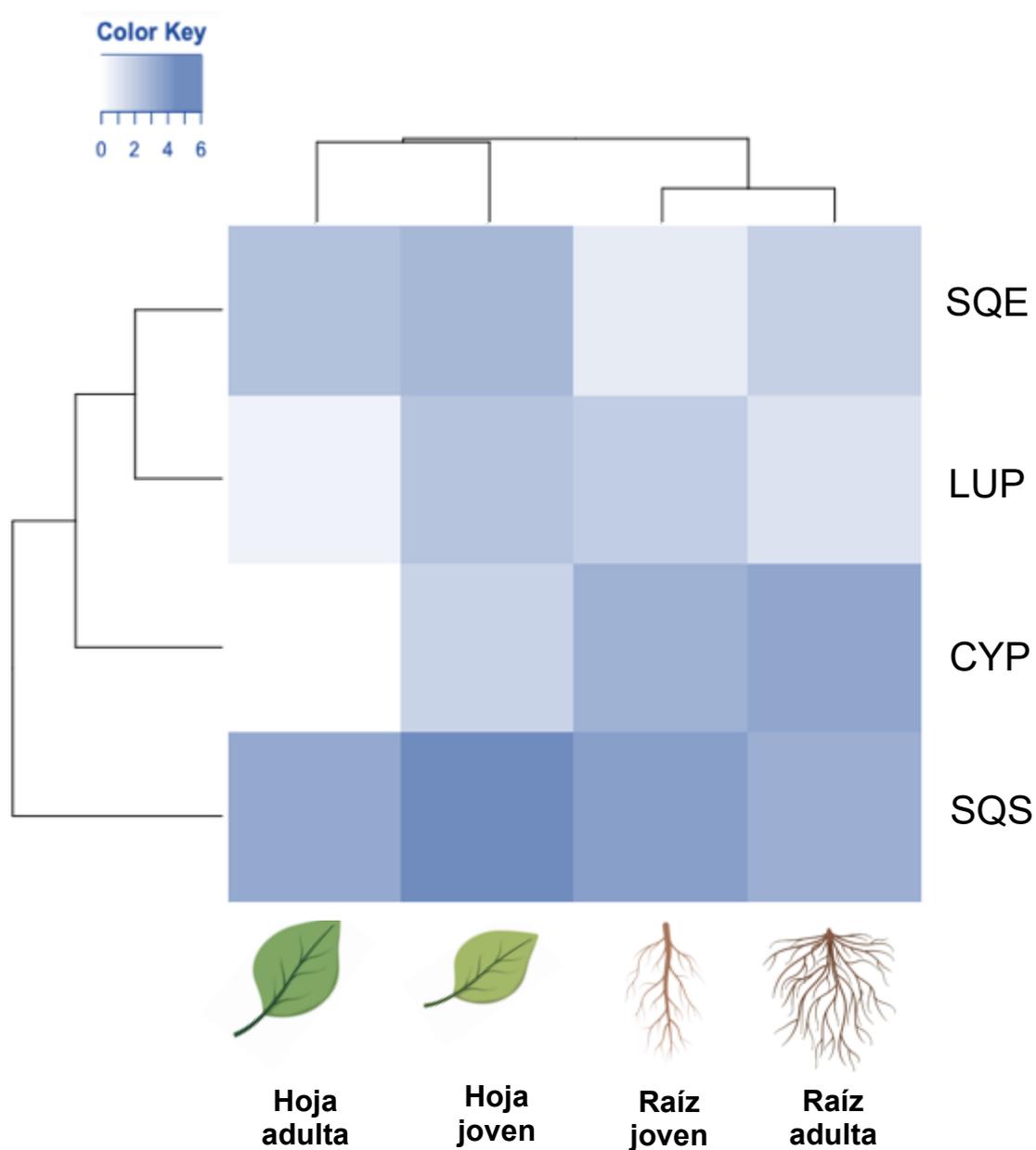


Figura 3.7. Mapa de calor de los patrones de expresión de los genes involucrados en la biosíntesis del ácido betulínico para cada condición. En el eje x representa las diferentes condiciones en las que se elaboraron los transcriptomas y el eje y los genes seleccionados.

3.2. Análisis molecular.

3.2.1. Extracción de ARN total de plantas de *P. andrieuxii*.

La extracción de ARN se realizó a partir de hojas y raíces de plantas adultas y jóvenes de *P. andrieuxii*, cultivadas en el vivero del CICY. Basicamente se siguió el protocolo establecido en el laboratorio empleando la técnica de CTAB, con mínimos ajustes con respecto a los tiempos de incubación. La obtención de ARN de alta calidad es la base de muchos análisis posteriores de biología molecular de plantas, como la RT-PCR, que es más confiable y eficiente cuando se emplea ARN con alta pureza e integridad (Liu *et al.*, 2018). Se procesaron un total de 4 muestras, verificando su integridad en geles de agarosa y midiendo la concentración por absorbancia para determinar la pureza del ARN y del cDNA. Todas presentaron una concentración aceptable, considerando que la cantidad óptima de ARN para realizar la retrotranscripción debe ser mayor a 80 ng. Además, ya que las proporciones de absorbancia A260/A280 se encontraron dentro de los parámetros requeridos para un análisis de este tipo (1.9-2.2), se procedió a trabajar con el ARN obtenido (Tabla 3.7). En cuanto a la integridad de RNA, todas las muestras mostraron dos bandas definidas características de las subunidades 28S y 18S del ARN ribosomal. (Figura 3.8).

Tabla 3.7. Cuantificación y relaciones de absorbancia del ARN total de tejidos de *P. andrieuxii*

Tejido	ng/ μ l	Ratios de calidad OD	
		A ₂₆₀ /A ₂₈₀	A ₂₆₀ /A ₂₃₀
HA	2090.3	1.99	2.11
HJ	351.8	2.03	2.20
RA	335.6	2.00	2.30
RJ	326.4	1.98	2.29

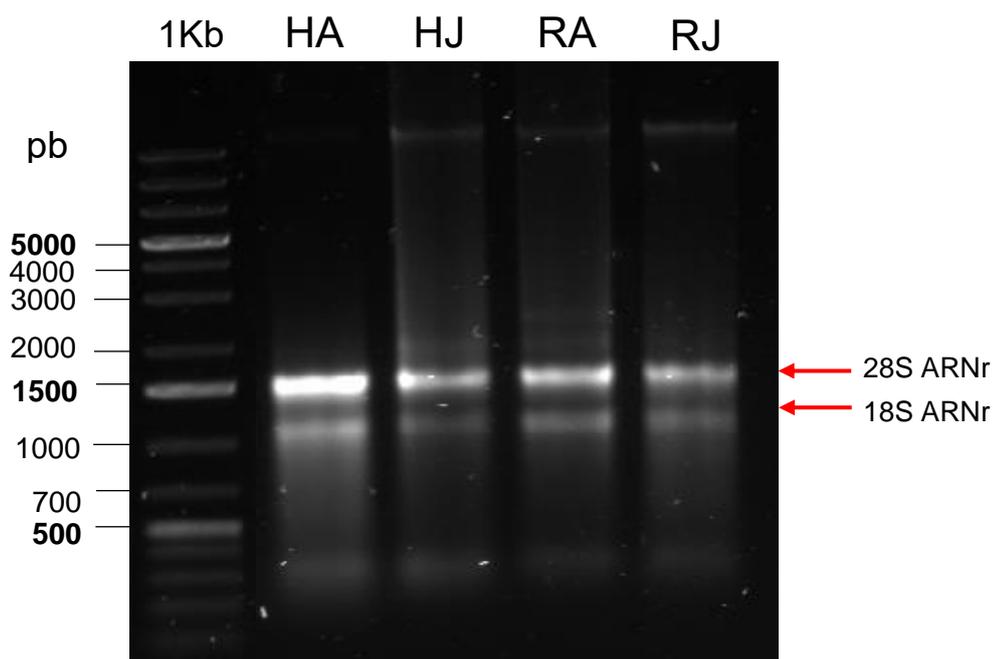


Figura 3.8. Integridad de los ARN extraídos de los diferentes tejidos de *P. andrieuxii*. 1Kb: Marcador de peso molecular. HA: hoja adulta, HJ: hoja joven, RA: raíz adulta, RJ: raíz joven.

3.2.2. PCR punto final de genes de biosíntesis del ácido betulínico.

Después de tratar las muestras con DNasas, se procedió a realizar la síntesis de cDNA. Asimismo, se obtuvo una concentración aceptable y los valores de la relación 260/280 oscilaron entre el rango 1.6-2.0 (Tabla 3.8), lo que permitió que las muestras se pudieron utilizar para el análisis de expresión.

Tabla 3.8. Cuantificación de cDNA de tejidos de *P. andrieuxii* y relaciones de absorbancia para determinar la pureza.

Tejido	ng/ μ l	Radios de calidad OD	
		A ₂₆₀ /A ₂₈₀	A ₂₆₀ /A ₂₃₀
HA	2,512.7	1.68	1.95
HJ	2,650.5	1.80	1.89
RA	2,512.5	1.62	1.87
RJ	2,336.7	1.83	1.94

Para evaluar el rendimiento de los cebadores diseñados para la amplificación de fragmentos ORF de los genes de biosíntesis del ácido betulínico en los cuatro tejidos de *P. andrieuxii*, se realizó una PCR punto final y los productos de amplificación se resolvieron por electroforesis en geles de agarosa (Figura 3.9) Según los resultados, un solo amplicón de ADN coincidió con el tamaño esperado para los genes *SQS*, *SQE*, *LUP* y *CYP* en los cuatro tejidos, lo que indicó que los cebadores trabajaron de manera eficiente. Sin embargo, con los cebadores de *LUP* en tejido de raíz joven no lograron amplificar el gen. Así mismo, en la amplificación con los cebadores de *SQS*, se obtuvo una segunda banda inespecífica.

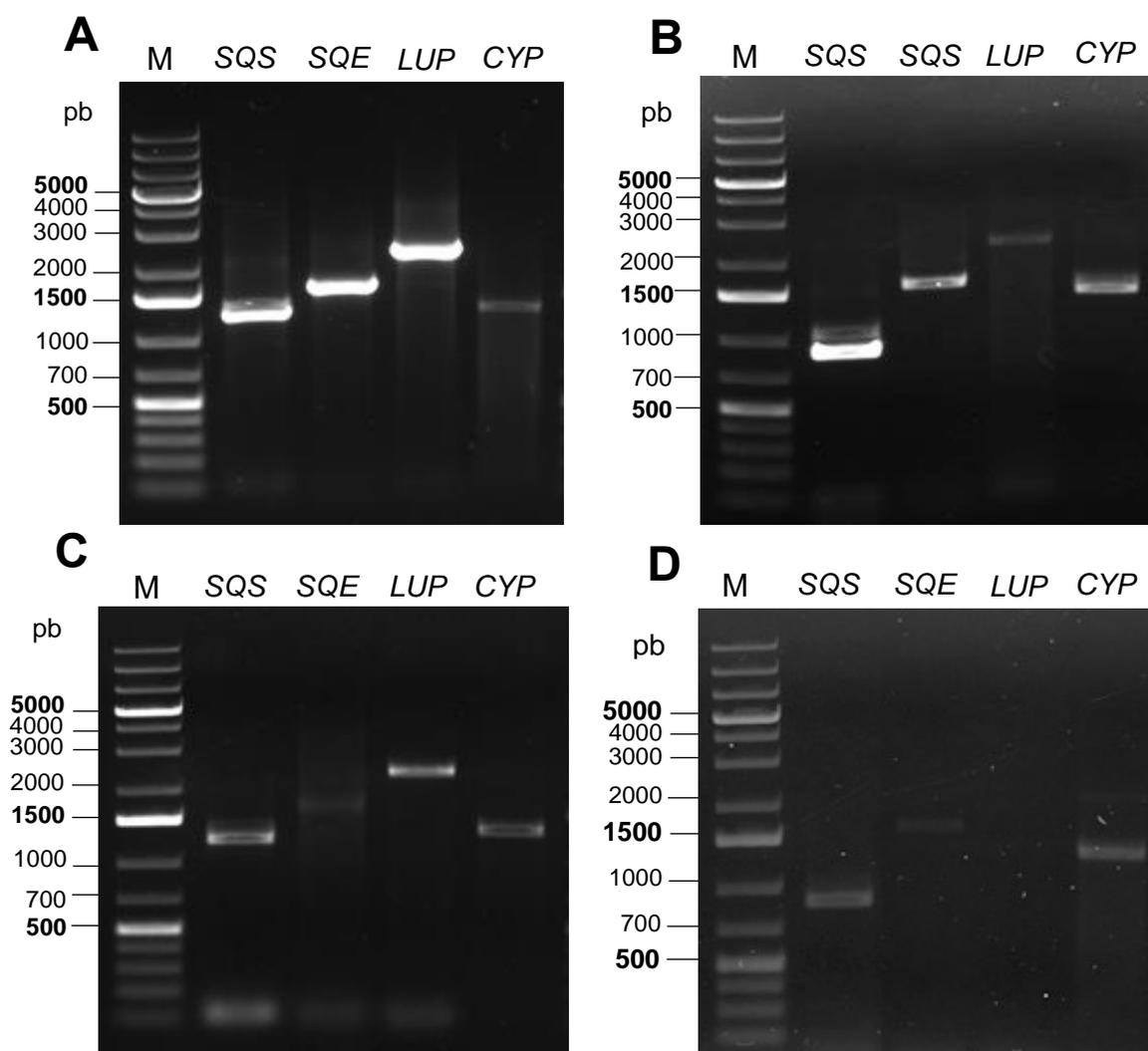


Figura 3.9. Amplificación de los ORF de los genes de la biosíntesis de triterpenos. A) Hoja de planta adulta, **B)** hoja de planta joven, **C)** Raíz de planta adulta, **D)** Raíz de planta joven. **M:** Marcador 1 kb plus. **SQS:** escualeno sintasa; **SQE:** escualeno epoxidasa; **LUP:** lupeol; **CYP:** citocromos P450 subfamilia CYP716A. Los geles utilizados fueron de agarosa al 1% con 0.6 μ l de SYBR Safe.

3.2.3. Aislamiento y clonación del ADNc del gen SQS.

Con el objetivo de validar la identidad de los genes seleccionados, se clonó el marco de lectura del gen SQS. Para ello, primeramente se amplificó el ORF utilizando los cebadores diseñados a partir de la región codificante. Los fragmentos revelados en los geles de agarosa que coincidieron con el tamaño esperado (1314 pb, Tabla 3.4), se clonaron en el vector pGEM-T easy y se analizaron con enzimas de restricción.

Los resultados de la digestión confirmaron la presencia de dos sitios de corte para la enzima *EcoRI* flanqueando el fragmento clonado en el vector recombinante pGEM-SQS (Figura 3.10). La resolución de los fragmentos de la digestión por electroforesis revelaron la presencia de dos bandas esperadas, así como la amplificación del gen SQS en el vector recombinante (Figura 3.11).

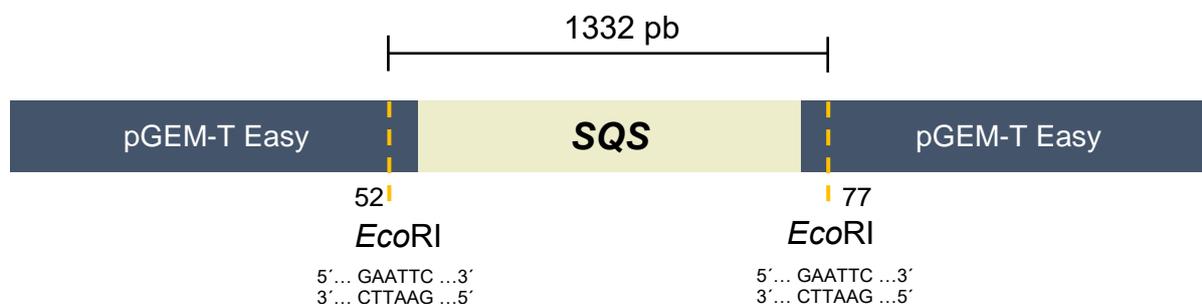


Figura 3.10. Predicción del cortes con las enzima *EcoRI* sobre el plásmido pGEM-SQS.

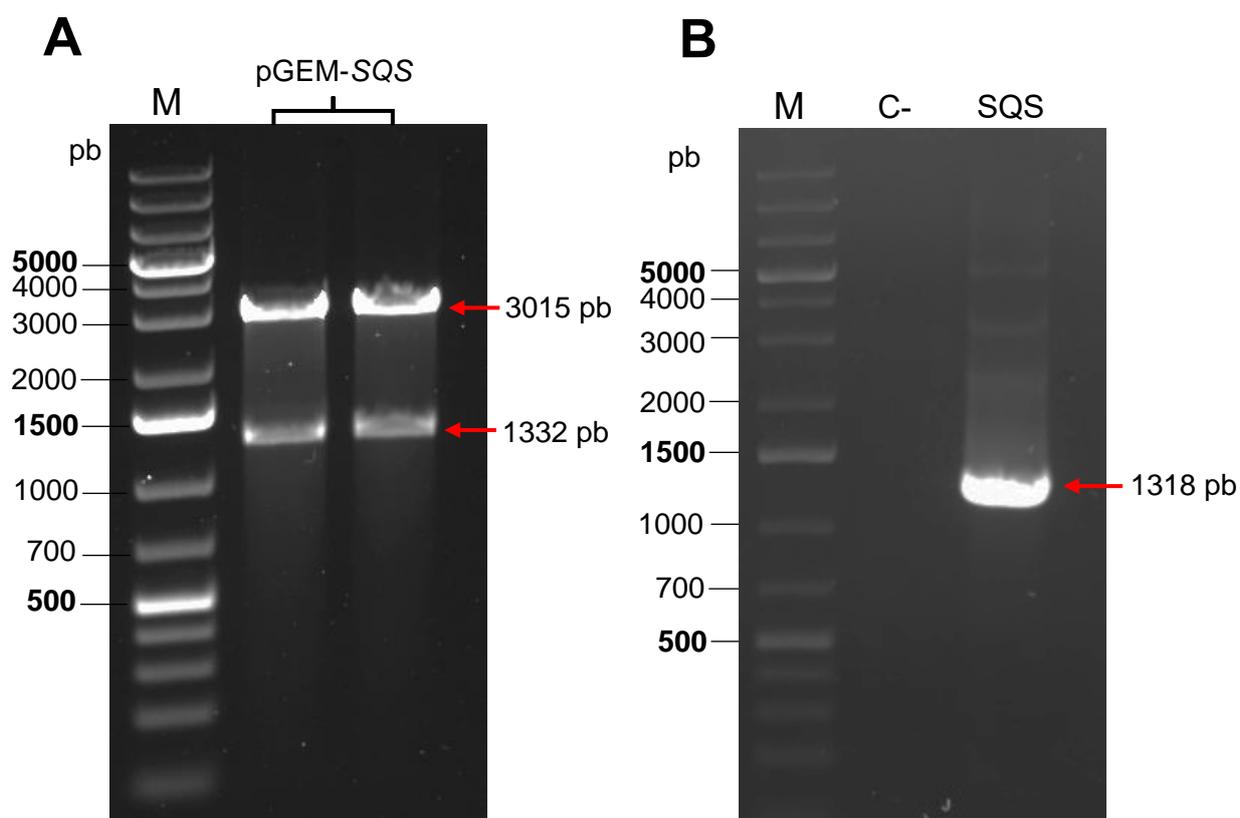


Figura 3.11. Digestión y PCR del plásmido pGEM-SQS. A) Cortes con la enzima *EcoRI*, la banda de 1332 pb corresponde al gen *SQS* y la banda de 3015 pb al vector. **B)** Amplicón de 1318 pb del gen *SQS*. C-: control negativo. M: Marcador 1 kb plus.

Posteriormente, se mandó a secuenciar el plásmido pGEM-SQS para la verificación indiscutible de la identidad del fragmento clonado. También se mandó a secuenciar el producto de PCR para el gen de lupeol de hoja joven.

En la Figura 3.12 se presenta el alineamiento de aminoácidos entre la secuencia del transcrito de *SQS*-HA y la secuencia enviada por MacroGen (*SQS*_macrogen), en el que mostraron un 99.24% de identidad, ya que una parte intermedia no se pudo secuenciar. Por su parte, el resultado de secuenciación de *LUP* mostró falta de aminoácidos en ambos extremos y en posiciones intermedias, mostrándo un 95.08% de identidad entre las secuencias alineadas (Figura 3.13).

```

SQSHA      MGSFGAILRHPDDAYPLMKLKMAARKAEKQIPPQPHWGFCYTMLHKVSRSFALVIQQLDT 60
SQS_macrogen -----MAARKAEKQIPPQPHWGFCYTMLHKVSRSFALVIQQLDT 39
                *****

SQSHA      DLRDAICIFYLVLRALDTVEDDTSIATDVKVPILMAFHRHIYDREWQFSCGTKDYKVLMD 120
SQS_macrogen DLRDAICIFYLVLRALDTVEDDTSIATDVKVPILMAFHRHIYDREWQFSCGTKDYKVLMD 99
                *****

SQSHA      QFHHVSTAFLELGHYSYQEAIIENITTRMGAGMAKFICKEVETIDDYDEYCHYVAGLVGLGL 180
SQS_macrogen QFHHVSTAFLELGHYSYQEAIIENITTRMGAGMAKFICKEVETIDDYDEYCHYVAGLVGLGL 159
                *****

SQSHA      SKLFHASGKEDLAPDSLNSMGLFLQKTNIIRDYLEDINEIPKSRMFWRQIWSKYVKKL 240
SQS_macrogen SKLFHASGKEDLAPDSLNSMGLFLQKTNIIRDYLEDINEIPKSRMFWRQIWSKYVKKL 219
                *****

SQSHA      EFPFCVPFGLHPTGILAPFDIFLQDLKYEENS IKAVQCLNDMITNALIHAEDCLTYMSNL 300
SQS_macrogen -----EDLKYEENS IKAVQCLNDMITNALIHAEDCLTYMSNL 256
                : *****

SQSHA      RDPAIFRFCAIPQIMAI GTLAMCYNNIEVFRGVVKMRRGLTAKVIDRTKNMSDVYGAFPD 360
SQS_macrogen RDPAIFRFCAIPQIMAI GTLAMCYNNIEVFRGVVKMRRGLTAKVIDRTKNMSDVYGAFPD 316
                *****

SQSHA      FSCMLKSKVDDNDPNAKTKSRLEAILKVCRESGTLTKRKS YIMQTEPAYNTTVIVLILI 420
SQS_macrogen FSCMLKSKVDDNDPNAKTKSRLEAILKVCRESGTLTKRKS YIMQTEPAYNTTVIVLILI 376
                *****

SQSHA      ILAILLAYLSSKRQNDI 437
SQS_macrogen IRAILLAYLSSKRQNEI 393
                * *****:

```

Figura 3.12. Alineamiento entre el producto de traducción del transcrito de SQS del transcriptoma de hoja adulta (SQSHA), contra la secuencia obtenida tras la secuenciación (SQS_macrogen). **Morado:** residuos polares básicos, con carga positiva. **Azul:** residuos polares ácidos, con carga negativa. **Verde:** residuos polares neutros. **Rojo:** residuos apolares.

LUPHJ	MWKLKIAEGDGPWLTTTNNYIGRQFWFDPNAGTPYERAEVERLRQEFMNRFKIRQSSD	60
LUP_macrogen	-----PKVLGYTTNNYIGRQFWFDPNAGTPYERAEVERLRQEFMNRFKIRQSSD	51

LUPHJ	LLMRMLQKENPAGSIPPPVRVHEHEDITEEAVTTTLRRAISFYSSIQAHDGHWPAESAG	120
LUP_macrogen	LLMRMLQKENPAGSIPPPVRVHEHEDITEEAVTTTLRRAISFYSSIQAHDGHWPAESAG	111

LUPHJ	PLFFLPPLVIALYIIGAMNEILTPAHQTEIKRYIYNHQEDGGWGLHIEGHSTVFGSCLN	180
LUP_macrogen	PLFFLPPLVIALYIIGAMNEILTPAHQTEIKRYIYNHQEDGGWGLHIEGHSTVFGSCLN	171

LUPHJ	YTALRLLGEGPDGGEDGAVAKGRKWILDHGGAIGIPSWGKFWLTVLGLYEDGLNMPPE	240
LUP_macrogen	YTALRLLGEGPDGGEDGAVAKGRKWILDHGGAIGIPSWGKFWLTVLGLYEDGLNMPPE	231

LUPHJ	FWLLPKLSPIHPGKMLCYCRLVMPMSYLYGKRFVGPITELVQSLRQEIYVQPYNKVDWN	300
LUP_macrogen	FWLLPKLSPIHPGKMLCYCRLVMPMSYLYGKRFVGPITELCFWLSRH-----	279
	***** * ;.	
LUPHJ	KARNTCAKEDLYYPHPLAQDMLWGFNLHFVEPILNRWPFKGLREKALKKAMEHVHYEDEM	360
LUP_macrogen	-----GWQFF-----	284
	* *	
LUPHJ	SRYMCIGCVEKVLCLIACWVEDPNSDVYKRHIARLPDYYWVAEDGLKMQSFGCQTWDAAF	420
LUP_macrogen	-----ANS-----	287
	* ;.	
LUPHJ	AIQAILSSNLAEYVPTLRKAHEFIKASQVRDNPAGDFGAMYRHSKGAWTFMQDHGWQ	480
LUP_macrogen	-----GYLIKFRDNPAGDFGAMYRHSKGAWTFMQDHGWQ	323
	:.*****	
LUPHJ	VSDCTAEGLKCALLFARMPPPELFGEKTEETERFYDAVNVILSLQSPNGGPPAWEPQRAFRW	540
LUP_macrogen	VSDCTAEGLKCALLFARMPPPELFGEKTEETERFYDAVNVILSLQSPNGGPPAWEPQRAFRW	383

LUPHJ	LESNPTEFFEDVLIKEYVECTSSAIQGLALFKKWHHPNYRRKEIESSIVRAAKYIEDIQ	600
LUP_macrogen	LESNPTEFFEDVLIKEYVECTSSAIQGLALFKKWHHPNYRRKEIESSIVRAAKYIEDIQ	443

LUPHJ	EPDGWSGSWGICYTYGTWFAVEGLATCGKTYQNC PALRKACKFLLSKQLPDGGWGENYL	660
LUP_macrogen	EPDGWSGSWGICYTYGTWFAVEGLATCGKTYQNC PALRKACKFLLSKQLPDGGWGENYL	503

LUPHJ	SNANEVYTNIEGNKSNLVQTSWALLSLIDAGQAEIDPTPIHRGIRLLINSQERDGFPPQQ	720
LUP_macrogen	SNANEVYTNIEGNKSNLVQTSWALLSLIDAGQAEIDPTPIHRGIRLLINSQERDGFPPQQ	563

LUPHJ	GITGAFMKNCTLNYSYRNIFPIWALGEYRRHVLLA	756
LUP_macrogen	GITGAFMKNCTLNYSYRNIFQLALMA-----	590
	***** : ;.	

Figura 3.13. Alineamiento entre el producto de traducción del transcrito de *LUP* del transcriptoma de hoja joven (LUPHJ), contra la secuencia obtenida tras la secuenciación (LUP_macrogen). **Morado:** residuos polares básicos, con carga positiva. **Azul:** residuos polares ácidos, con carga negativa. **Verde:** residuos polares neutros. **Rojo:** residuos apolares.

Con el fin de identificar los homólogos más parecidos a las secuencias de los productos de traducción de los transcritos *SQS* y *LUP*, se realizó una búsqueda con la herramienta BLASTp desde el servidor del NCBI. Los resultados revelaron que para *SQS* se obtuvieron las mejores homologías con las escualeno sintasas de *Vitis vinifera* y *Betula platyphylla*, mostrando una similitud entre 84% y 85%, y de identidad entre 84% y 85% con una cobertura de 99%, respectivamente (Tabla 3.12, A). En el caso de *LUP*, la mejor homología, basándose en el mejor resultado de “E-value” de 0 fue con una β -amirina sintasa de *Vitis vinifera*, con 57% de similitud, sin embargo, la mayor similitud (80%) e identidad (86%) fueron con una secuencia de lupeol sintasa de *Vitis vinifera* (Tabla 3.12, B). Adicionalmente, se realizó un alineamiento múltiple con el fin de observar las similitudes y diferencias para cada aminoácido. En el caso de *SQS*, mostró una longitud similar a las proteínas *SQS* caracterizadas de otras plantas (393-413 aa) y la diferencia se presentó en el extremo inicial, donde no se pudieron secuenciar 21 aa (Figura 3.14). De igual manera, el alineamiento de *LUP* contra las proteínas caracterizadas, mostró fragmentos incompletos debido a problemas de secuenciación, sin embargo, se identificaron los dominios que caracterizan a las OCS, destacando el motivo MLCYCR de lupeol sintasa (Figura 3.15).

Tabla 3.4. Secuencias que producen alineamientos significativos contra los resultados de secuenciación mediante BLASTp. A) Secuencias alineadas con *SQS*. B) Secuencias alineadas con *LUP*.

Descripción	Organismo	Cobertura	E-value	Similitud (%)	Identidad (%)	Accesión
A. Query: <i>SQS</i>_macrogen						
Squalene synthase	<i>Vitis vinifera</i>	99%	0.0	91%	85%	XP_002266150
Squalene synthase	<i>Betula platyphylla</i>	99%	0.0	84%	84%	AKR76253
B. Query: <i>LUP</i>_macrogen						
Beta-Amyrin Synthase 1	<i>Vitis vinifera</i>	98%	0.0	57%	45%	RVW17245
Lupeol synthase isoform X2	<i>Vitis vinifera</i>	98%	2e-170	80%	86%	XP_019078376

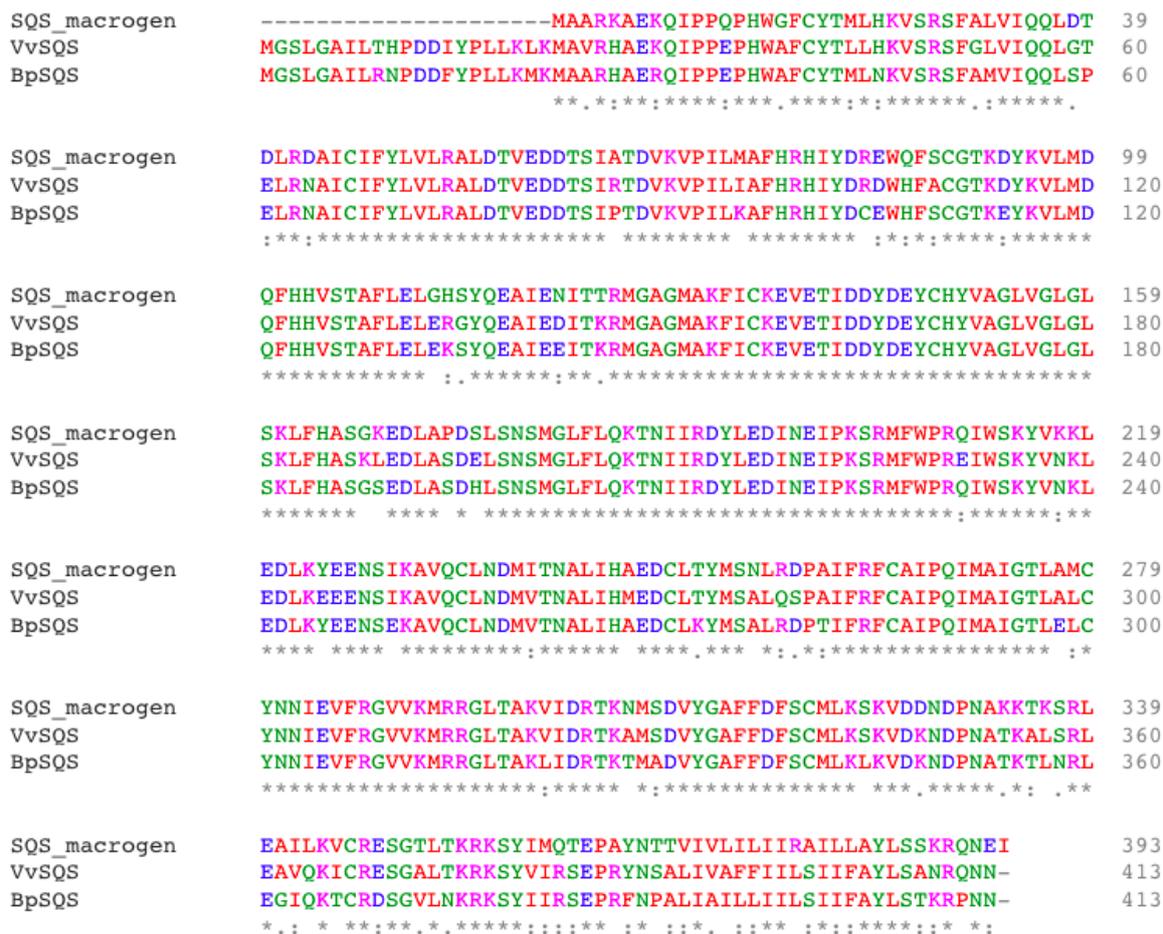


Figura 3.14. Alineamiento entre la secuencia SQS de *Vitis vinifera* (VvSQS) y *Betulla platyphylla* (BpSQS) contra la SQS recuperada (SQS_macrogen). **Morado:** residuos polares básicos, con carga positiva. **Azul:** residuos polares ácidos, con carga negativa. **Verde:** residuos polares neutros. **Rojo:** residuos apolares.



Figura 3.15. Alineamiento entre las OSC de *Vitis vinifera* (Vv-bAS y VvLUS) contra la LUP recuperada (LUP_macrogen). El recuadro rojo indica el motivo conservado de las OSC que es importante para la determinación del producto final. **Morado:** residuos polares básicos, con carga positiva. **Azul:** residuos polares ácidos, con carga negativa. **Verde:** residuos polares neutros. **Rojo:** residuos apolares.

CAPÍTULO IV

DISCUSIÓN GENERAL

El desarrollo de nuevas tecnologías de secuenciación masiva, como la secuenciación de ARN (RNA-seq), se han convertido en una poderosa herramienta para comprender los mecanismos moleculares asociados a procesos biológicos, asimismo, para estudiar transcriptomas completos en diversas condiciones y tejidos (Wang *et al.*, 2009; Hölzer y Marz, 2019). En la actualidad, se han secuenciado los genomas y transcriptomas de alrededor de 46 especies de plantas medicinales, lo que constituye una forma eficiente de descifrar nuevas funciones genéticas involucradas en vías metabólicas específicas presentes en plantas medicinales (Misra, 2014). Este enfoque ha permitido identificar nuevos genes que codifican enzimas involucradas de una vía, las secuencias que codifican las enzimas para la biosíntesis de triterpenos podrán eventualmente utilizarse para aplicaciones moleculares, como la ingeniería metabólica.

Aunque el papel biológico de los triterpenoides en las plantas sigue siendo poco claro, se ha demostrado que estos compuestos tienen varias propiedades medicinales (Furtado *et al.*, 2017). Los triterpenoides son uno de los metabolitos potencialmente responsables de las bioactividades anteriormente mencionadas en *P. andrieuxii*; sin embargo, sus genes para la biosíntesis de estos compuestos, son en gran parte desconocidos. En este estudio, el seguimiento del análisis transcriptómico se realizó mediante la identificación de los genes relacionados con la biosíntesis del ácido betulínico y evaluación de sus niveles de expresión en diferentes condiciones. Además, se aisló y clonó el ORF del gen SQS, que es el precursor de todos los isoprenoides.

En primer lugar, los alineamientos realizados de las secuencias obtenidas de SQS, SQE, LUP y CYP en BLASTn y BLASTp y comparadas a los cuatro transcriptoma de los tejidos de hoja y raíz de plantas jóvenes y adultas de *P. andrieuxii*, confirman la identificación de genes ortólogos. Si bien los homólogos identificados en BLAST dieron las mejores concordancias, con porcentajes de identidad mayores a 69% y porcentajes de similitud mayores a 77% con respecto a las secuencias obtenidas del NCBI, el análisis filogenético generó aún más certeza sobre esta información, ya que al utilizar secuencias de enzimas participantes en la biosíntesis del ácido betulínico previamente caracterizadas y al establecer relaciones evolutivas con ellas, se corroboró un origen común y se obtuvo mayor certeza sobre su posible función catalítica y la presencia de los motivos distintivos.

Se ha reportado que *Catharanthus roseus*, una planta medicinal importante de la familia Apocynaceae, produce triterpenoides, como el ácido ursólico y el ácido oleanólico, en cantidades considerables en la capa de cera cuticular de la hoja. De tal manera, Huang *et al.*, en el 2012 caracterizaron sus genes de OSC multifuncionales, de los cuales comparten 74, 75 y 76% de identidad con el gen *CYP716A12* de *Medicago truncatula* y *CYP716A15* y *CYP716A17* de *Vitis vinifera* a nivel de aminoácidos, respectivamente, lo que sugiere que la subfamilia CYP716A está altamente conservada en las oxidaciones de triterpenos en C-28. Al expresar estos genes en levadura se logró la conversión del lupeol a ácido betulínico. Los porcentajes de identidad entre los transcritos identificados de *P. andrieuxii* y el gen *CYP716A15* de *Vitis vinifera* son similares con los genes de *C. roseus*.

Por otro lado, el estudio de los componentes estructurales de una secuencia, particularmente la identificación de motivos, es otra herramienta que amplía el conocimiento sobre la evolución molecular y permite construir modelos de los posibles procesos moleculares y celulares en los que estas secuencias participan (Bailey *et al.*, 2015). Las alineaciones de las secuencias PaSQS con otras SQS conocidas, mostraron regiones altamente conservadas, presentando un dominio transmembrana que están estrechamente relacionadas con secuencias de SQS de *G. glabra*, *M. truncatula*, *L. japonicas*, *E. tirucalli* y *P. ginseng* (Rong *et al.*, 2016).

Para comprender la relación de las PaLUP con las OSC de otras plantas con perfiles de productos conocidos, se contruyó el árbol filogenético, en el cual, la secuencia PaLUP de hoja joven mostró una preferencia por lupeol como sustrato, mientras que las secuencias de los transcriptomas de HA, HJ y RA, se agruparon con las secuencias de β -amirina. Estas observaciones sugirieron la probable participación de las secuencias de *P. andrieuxii* en la biosíntesis de ácido betulínico y ácido oleanólico. Estos resultados coinciden con Sandeep *et al.*, 2019, donde una OSC de *Lagerstroemia speciosa* convertía el precursor 2,3-oxidoescualeno en lupeol. Esta OSC en su motivo MLCYCR tiene un cambio de leucina por triptofano, otras tres OSC mantienen el motivo con triptofano (MWCYCR), por lo que en el árbol filogenético se agrupan con OSC tipo β -amirina, de los cuales, el perfil de producto de estas oxidoscualenos ciclasas se comprobó mediante su expresión heteróloga en levadura. Por otra parte, en estudios con experimentos de mutagénesis dirigida, revelaron que la diversidad funcional de las OSC está determinada por ciertos residuos de aminoácidos clave, que son esenciales para la posterior especificidad del proceso de ciclación (Xue *et al.*, 2018).

Por otro lado, en el árbol de las CYP450, las cuatro secuencias del transcriptoma de *P. andrieuxii* se agrupan en un clado, relacionadas con el grupo del clan CYP85, al que

pertenece la subfamilia CYP716A; el árbol reveló que las CYP de *P. andieruxii* también tienen actividad de oxidación C-28, y se agruparon con oxidasas C-28 previamente caracterizadas, como *CYP716A154* de *C. roseus* (Huang *et al.*, 2012), *CYP716A12* de *M. truncatula* (Carelli *et al.*, 2011; Fukushima *et al.*, 2011), *CYP716* de *Betula platyphylla* (Zhou *et al.*, 2016) y *CYP716A15* y *CYP716A17* de *Vitis vinifera* (Fukushima *et al.*, 2011). Sin embargo, Suzuki *et al.*, en el 2018 evaluaron enzimas CYP716A que pertenecían a diferentes familias de plantas, las cuales mostraron más del 70% de identidad a nivel de aminoácidos, mientras que exhiben diferencias significativas en la actividad catalítica y la especificidad del sustrato. De esto se puede inferir que cualquier predicción de los aminoácidos clave que determinan la actividad catalítica y la especificidad del sustrato es especulativa, la estructura cristalina de la subfamilia CYP716A aun no está disponible actualmente y la única manera de identificar el producto final en la biosíntesis de diferentes ácidos triterpénicos es la expresión heteróloga.

El resultado del análisis de expresión *in silico* mostró que el perfil de los genes triterpénicos varió considerablemente entre los tejidos. Entre todos los genes *PaSQS*, *PaSQE* y *PaLUP*, se expresaron mayormente en la hojas jóvenes con un rango de expresión de 11.11 a 408.02 TPM, mientras que *PaCYP* mostró el nivel de expresión más alto en raíz adulta con un valor de 38.40 TPM. Por su parte, la única OSC que se agrupó en el árbol filogenético con las secuencias de lupeol sintasa que fue derivada del transcriptoma de hoja joven fue la que obtuvo la mayor expresión entre los cuatro transcriptomas con un valor de expresión de 11.11 TPM.

Curiosamente, el nivel de expresión de *PaCYP* fue menor en los transcriptomas de hoja que en los de raíz. Además, en todos los transcriptomas *PaSQS* tuvo mayor expresión en comparación con los otros genes, posiblemente se deba a que la enzima SQS cataliza el primer paso de la biosíntesis de todos los triterpenoides y de los esteroides. También se observó que los patrones de expresión de *PaSQS* coincidieron con el análisis del transcriptoma de plantas de la familia Lamiaceae, con niveles más altos de transcripción en hojas (Aminfar *et al.*, 2019), de igual forma en el transcriptoma de *Platycodon grandiflorum* (Ma *et al.*, 2016). En general, en las plantas se encuentran de uno a tres genes SQS. El genoma de *A. thaliana* contiene dos secuencias anotadas como SQS, At4g34640 (*SQS1*) y At4g34650 (*SQS2*), de los cuales, el gen *SQS1* se expresa en todos los tejidos y durante todo el desarrollo de la planta, mientras que *SQS2* se expresa principalmente en el tejido vascular de los pecíolos de las hojas y cotiledones y el hipocótilo de las plántulas (Rong *et al.*, 2016).

En el caso del gen *SQE*, los transcriptomas de hoja mostraron los nivel de expresión más altos, coincidiendo una vez más con tres especies de plantas de la familia Lamiaceae (*Rosmarinus officinalis*, *Salvia officinalis* y *Thymus persicus*) (Aminfar *et al.*, 2019).

Como se muestra en la amplificación de los genes por PCR tiempo final, todos los cebadores diseñados sirvieron para amplificar un fragmento del tamaño esperado, en relación al gen blanco, la excepción con el cebador L1 no logró amplificar el gen *LUP* en tejido de raíz joven. Por su parte, los cebadores CYP (C1 y C2) mostraron el patrón de amplificación similar en las cuatro condiciones diferentes de la planta. Tomando en cuenta estas amplificaciones para un análisis semicuantitativo, los cebadores para *SQS* mostraron una mayor expresión, lo que coincide con el análisis de expresión *in silico*. A partir de estos resultados, se sugiere que la expresión de los niveles de transcripción varía claramente entre los órganos, aunque debe tenerse en cuenta que la naturaleza de la actividad transcripcional es muy dinámica y puede cambiar rápidamente con las condiciones climáticas, las etapas de desarrollo y el estrés ambiental (Aminfar *et al.*, 2019).

Los datos corroboran los resultados de Hiebert *et al.*, 2016, quienes muestran que el contenido de ácido betulínico en las hojas, no tiene relación aparente con la ontogenia de la planta, es decir, su acumulación no inicia en una determinada etapa de desarrollo. Por otro lado, indican que el almacenamiento del ácido betulínico se presenta en las hojas, pero no en las raíces, sin embargo, los resultados establecen la expresión del gen de la lupeol sintasa en el transcriptoma de hoja joven de *P. andrieuxii*, pero no en tejidos de raíces, ya que las secuencias de OSC encontradas en dichos transcriptomas se agruparon con la β -amirina.

Por otra parte, aunque las predicciones *in silico* indicaron que los genes *CYP* de *P. andrieuxii* codifican probablemente enzimas funcionales que oxidan el lupeol para la producción del ácido betulínico, solo su caracterización funcional permitiría confirmar la capacidad de cada gen.

Estos resultados no sólo proporcionan información valiosa para la mejor comprensión de la vía de biosíntesis de ácidos triterpénicos en *P. andrieuxii*, sino que también brindan oportunidades para la producción de ingredientes activos mediante la ingeniería de microorganismos. Dado que los metabolitos secundarios de alto valor son producidos por plantas en pequeñas cantidades, es difícil estudiar y profundizar su mecanismo biosintético. Para aumentar la producción de estos compuestos, que a menudo tienen propiedades farmacéuticas, la ingeniería metabólica requiere un conocimiento integral de las vías y flujos metabólicos, de hecho, la sobreexpresión de *SQE* ya se ha introducido como una nueva estrategia para impulsar la producción de triterpenos en las plantas (Aminfar *et al.*, 2019).

CAPITULO V

CONCLUSIONES Y PERSPECTIVAS

CONCLUSIONES

Con el uso de las herramientas bioinformáticas fue posible identificar por homología los transcritos de los genes involucrados en la biosíntesis del ácido betulínico en cuatro transcriptomas de *P. andrieuxii*.

El análisis filogenético de estos genes, demostró las relaciones evolutivas de los transcritos con secuencias ya caracterizadas, de las cuales: *PaSQS* y *PaSQE* de los cuatro transcriptomas, se agruparon en los clados correspondientes; un transcrito de *PaLUP* del transcriptoma de hoja joven se agrupó al clado con secuencias de lupeol sintasa, mientras que los transcritos de *PaLUP* de HA, RA y RJ se relacionaron más en secuencias de β -amirina; por su parte, las cuatro secuencias de *PaCYP* se agruparon en el clado en el que pertenece el clan correspondiente.

Utilizando la estrategia de alineamientos de secuencias, se pudo demostrar que todas las secuencias recuperadas del transcriptoma se conservan entre especies, mostrando los motivos conservados que las caracterizan.

Con relación a la expresión de genes *in silico*, en general, los genes biosintéticos en *P. andrieuxii* muestran un patrón de expresión espacio-temporal (tejido particular /etapa de desarrollo), siendo el gen *SQS* el mayormente expresado en el transcriptoma de hoja joven.

En general, los genes *SQS*, *SQE* y *CYP* se expresan en los cuatro transcriptomas de *P. andrieuxii*, sin embargo, el gen *OSC*, considerado como una lupeol sintasa, sólo se encontró en el transcriptoma de hoja joven. De modo que, la hipótesis planteada se cumplió, debido a que la acumulación del ácido betulínico en plantas de *P. andrieuxii*, posiblemente sólo se presente en hojas y no en raíces.

La secuenciación del transcrito del gen de la enzima lupeol sintasa y el plásmido recombinante pGEM-SQS, permitió validar el ensamblado de los transcriptomas y la identificación de genes por homología.

PERSPECTIVAS

Realizar un análisis de los perfiles de expresión por qRT-PCR en el tejido de hoja de plantas adultas y jóvenes, para determinar el patrón de expresión de los genes de biosíntesis y corroborar los resultados del análisis de expresión *in silico*.

Caracterizar los genes por medio de expresión heteróloga, con el fin de determinar la actividad catalítica y la especificidad del sustrato de las enzimas de la subfamilia CYP716A. Cuantificar la acumulación de los triterpenos específicos por HPLC, para evaluar los niveles de acumulación de dichos compuestos y para mostrar que las enzimas desarrolladas pueden ser útiles en la producción heteróloga de triterpenoides importantes desde el punto de vista farmacológico.

Establecer la estructura tridimensional de las enzimas identificadas por homología, con base en estructuras conocidas de proteínas relacionadas, de tal manera que sea posible determinar los sitios funcionales y tener mayor evidencia de la identidad de las enzimas. Con relación a esto, se realizaron los primeros avances de predicción de la estructura tridimensional de las cuatro enzimas involucradas en la biosíntesis del ácido betulínico de los transcriptomas de *P. andrieuxii* utilizando el modelado por homología, incluidos en el Anexo No. 3.

BIBLIOGRAFÍA

- Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., y Goecks, J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*, 44(W1), W3–W10.
- Ai-Xia, C., Yong-Gen, L., Ying-Bo, M., Shan, L., Ling-Jian, W. y Xiao-Ya, C. (2007). Plant terpenoids: biosynthesis and ecological functions. *Journal of Integrative Plant Biology*, 49(2), 179-186.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., y Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.
- Almeyda, A. I. (2017). Estudio de la acumulación de ácido betulínico y urechitol A durante el desarrollo de *Pentalinon andrieuxii* y su relación con la metilación de ADN. Tesis de Maestría. Centro de Investigación Científica de Yucatán. México. pp. 57.
- Aminfar, Z., Rabiei, B., Tohidfar, M., y Hossein M. (2019). Identification of key genes involved in the biosynthesis of triterpenic acids in the mint family. *Scientific Reports*, 9(15826).
- Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Recuperado en: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, 43(W1), 39-49.
- Bolger, A. M., Lohse, M. y Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
- Buckingham, J., Cooper, C. M., y Purchase, R. (2015). *Natural products desk reference*. Boca Ratón, FL: CRC Press, pp. 253.
- Carelli M., Biazzì E., Panara F., Tava A., Scaramelli L., Porceddu A., et al. (2011). *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell*, 23, 3070–3081.

- Chan, M. J., Balanza, E., Deharo, E., Muñoz, V., García, R. D. y Peña, L. M. (2003). Variation of leishmanicidal activity in four populations of *Urechites andrieuxii*. *Journal of Ethnopharmacology*, 86(2-3):135-268.
- Chen, C., Zheng, Y., Zhong, Y. Wu, Y., Li, Z., Xu, L. A. y Xu, M. (2018). Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora*. *BMC Genomics*, 19(550), 1-15.
- De los santos, J. E. (2019). Evaluación de la producción de ácido betulínico en raíces transformadas de *Pentalinon andrieuxii* (Müll. Arg.) Hansen & Wunderlin. Tesis de Maestría. Centro de Investigación Científica de Yucatán. México. pp. 93.
- Domínguez, D. S., Escalante, F., García, K., Ruiz G., Gutierrez, D., Chan, M. J., Giménez, A., y Peña, L. M. (2010). Antiprotozoal activity of betulinic acid derivatives. *Phytomedicine* 17, 379-382.
- Dubey, V.S., Bhalla, R. y Luthra, R. J. (2003). An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *Journal of Biosciences*, 28(5), 637-646.
- Einzhammer, D. A. y Xu, Z. Q. (2004). Betulinic acid: a promising anticancer candidate. *IDrugs* 7(4), 359-373.
- Engelhardt, B. E., Jordan, M. I., Repo, S. T. y Brenner, S. E. (2009). Phylogenetic molecular function annotation. *Journal of Physics: Conference Series*, 180, 14-18.
- Fang, Z., Martin, J. y Wang, Z. (2012). Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. *Cell Biosci*, 2(26), 1-8.
- Fukushima, E. O., Seki, H., Ohyama, K., Ono E., Umemoto, N., Mizutani, M., Saito, K. y Muranaka, T. (2011). CYP716A Subfamily members are multifunctional oxidases in triterpenoid biosynthesis, *Plant and Cell Physiology*, 52(2), 2050–2061.
- Furtado, N. A., Pirson, L., Edelberg, H., Miranda, L., Loira-Pastoriza, C., Preat, V., Larondelle, Y. y André, C. M. (2017). Pentacyclic triterpene bioavailability: an overview of in vitro and *in vivo* studies. *Molecules*, 22(3), 400.
- Gao, Y., Honzatko, R. B., y Peters, R. J. (2012). Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Natural Product Reports*, 29(10), 1153–1175.

-
- Góngora, E., Childs, K. L., Fedewa, G., Hamilton, J. P., Liscombe, D. K., *et al.* (2012). Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLOS ONE*, 7(12), e52506.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... *et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644-652.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. y Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-321.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., ... Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8(8), 1494–1512.
- Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*, 30(5), 1229-1235.
- Hata, S., Sanmiya, K., Kouchi, H., Matsuoka, M., Yamamoto, N. e Izui, I. (1997). cDNA cloning of squalene synthase genes from mono- and dicotyledonous plants, and expression of the gene in rice. *Plant and Cell Physiology*, 38(12), 1409-1413.
- Hayashi, H., Huang, P., Takada, S., Obinata, M., Inoue, K., Shibuya, M, y Ebizuka, Y. (2004). Differential expression of three oxidosqualene cyclase mRNAs in *Glycyrrhiza glabra*. *Biol Pharm Bull*, 27(7), 1086-92.
- Hiebert, M. R., Escalante, F., García, K., Dzib, G. R., Calvo, L. M., Luis, M. y Peña, L. M. (2016). Spatio-temporal variation of terpenoids in wild plants of *Pentalinon andrieuxii*. *Chemistry and Biodiversity*, 13(11), 1521-1526.
- Hiebert, M., Godoy, G. y Peña, L. M. (2020). Genetic transformation of the tropical vine *Pentalinon andrieuxii* via *Agrobacterium rhizogenes* produces plants with an increased capacity of terpenoid production. *In Vitro cellular & Developmental Biology-Plant*.

- Hill, R. A. y Connolly, J. D. (2015). Triterpenoids. *Natural Product Reports*, 32, 273-327.
- Hölzer, M. & Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *Gigascience*, 8(5), 1-16.
- Huang L., Li J., Ye H., Li C., Wang H., Liu B., et al. (2012). Molecular characterization of the pentacyclic triterpenoid biosynthetic pathway in *Catharanthus roseus*. *Planta*, 236, 1571–1581.
- Huang, J., Zha, W., An, T. et al. (2019). Identification of RoCYP01 (CYP716A155) enables construction of engineered yeast for high-yield production of betulinic acid. *Appl Microbiol Biotechnol* 103, 7029–7039.
- Jin, C., Zhang, J., Song, H. y Cao, Y. X. (2019). Boosting the biosynthesis of betulinic acid and related triterpenoids in *Yarrowia lipolytica* via multimodular metabolic engineering. *Microbial Cell Factories*, 18(77), 1- 18.
- Juárez, V., Alvarado, L. y Villaseñor, J. (2007). La familia Apocynaceae sensu lato en México: diversidad y distribución. *Revista Mexicana de Biodiversidad*, 78, 459- 482.
- Kerfeld, C. A. y Scott, K. M. (2011). Using BLAST to Teach “E-value-tionary” Concepts. *PLOS Biology*, 9(2), e1001014.
- Kirby, J. y Keasling, J. D. (2009). Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annual Review of Plant Biology*, 60(1), 335-355.
- Lezama, C. M., Isaac, A. P., Zamora, P., Uc, M. R., Justiniano, S. Y., Angel, R., Satoskar, A. y Hernández, L. (2007). Leishmanicidal activity of *Pentalinon andrieuxii*. *Fitoterapia*, 78, 255-257.
- Lezama, C. M., Pan, L., Isaac, A. P., Terrazas, C., Oghumu, S., Isaac, R. y Satoska, A. R. (2014). *Pentalinon andrieuxii* root extract is effective in the topical treatment of cutaneous Leishmaniasis caused by *Leishmania Mexicana*. *Phytotherapy Research*, 28(6), 909-916.
- Li, W., Jaroszewski, L., y Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics*, 17(3), 282-283.

- Liu, L., Han, R., Yu, N., Zhang, W., Xing, L. y Xie, D. (2018). A method for extracting highquality total RNA from plant rich in polysaccharides and polyphenols using *Dendrobium huoshanense*. *PLoS ONE*, 13(5), 1-9.
- Lu, H. Y., Liu, J. M., Zhang, H. C., Yin, T. y Gao, S. L. (2008). Ri-mediated transformation of *Glycyrrhiza uralensis* with a squalene synthase gene (GuSQS1) for production of glycyrrhizin. *Plant Mol. Biol. Rep.*, 26, 1-11.
- Ma, C. H., Gao, Z. J., Zhang, J. J., Zhang, W., Shao, J. H., Hai, M. R., Chen, J. W., Yang, S. C. y Zhang, G. H. (2016). Candidate genes involved in the biosynthesis of triterpenoid saponins in *Platycodon grandiflorum* identified by transcriptome analysis. *Frontiers in Plant Science*, 7, 673.
- Ma, Y., Yuan, L., Wu, B., Li, X., Chen, S. y Lu, S. (2012). Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *Journal of Experimental Botany*, 63, 2809-2823.
- McGinnis, S., y Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, 32, W20 – W25.
- Miettinen, K., Pollier, J., Buyst, D. et al. (2017). The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nature Communication*, 8(14153).
- Misra, B. B. (2014). An updated snapshot of recent advances in transcriptomics and genomics of phytomedicinals. *J. Postdoc. Research*, 2, 1–15.
- Morales, J. F. (2005). Estudios en las Apocynaceae neotropicales XIV: Nuevas lectotipificaciones en los géneros *Hylaea* J.F. Morales y *Pentalinon* Voigt (Apocynoideae, Echiteae). *Lankesteriana*, 5(2), 159-160.
- Morales, J. F. (2006). Estudios en las Apocynaceae neotropicales XXVIII: la familia apocynaceae (apocynoideae, rauvolfioideae) de el salvador, Centroamérica. *Darwiniana*, 44(2), 453-489.
- Morrison, D. A., Morgan, M. J. y Kelchner, S. A. (2015). Molecular homology and multiple-sequence alignment: an analysis of concepts and practice. *Australian Systematic Botany*, 28, 46-62.

- Moses, T., Pollier, J., Thevelein, J. M. y Goossens, A. (2013). Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology *in vivo* and *in vitro*. *New Phytologist*, 200(1), 27- 43.
- Nakamura, M., Linh, T. M., Lien, L. Q., Suzuki, H., Mai, N. C., Giang, V. H., Tamura, K., Thanh, N. V., Suzuki, H., Misaki, R., Muranaka, T., Ban, N. K., Fujiyama, K., & Seki, H. (2018). Transcriptome sequencing and identification of cytochrome P450 monooxygenases involved in the biosynthesis of maslinic acid and corosolic acid in *Avicennia marina*. *Plant Biotechnology*, 35(4), 341–348.
- Nelson, D. y Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *The Plant Journal*, 66, 194-211.
- Noguera, G. J., Fabian, L. E., Lombardo, E., y Finkielstein, L. M. (2018). Studies of 4-arylthiazolylhydrazones derived from 1-indanones as *Trypanosoma cruzi* squalene epoxidase inhibitors by molecular simulations. *Organic & Biomolecular Chemistry*, (16), 8525-8536.
- Oghumu, S., Varikuti, S., Saljoughian, N., Terrazas, C., Huntsman, A. C., Parinandi, N. L., Fuchs, J. R., Kinghorn, A. D. y Satoskar, A. R. (2017). Pentalinonsterol, a constituent of *Pentalinon andrieuxii*, possesses potent immunomodulatory activity and primes T cell immune responses. *Journal of Natural Products*, 80(9), 2515-2523.
- Oliveira, B. H. D., Santos C., y Espíndola, A. (2002). Determination of the triterpenoid, betulinic acid, in *Dolicarpus schottianus* by HPLC. *Phytochemical Analysis*, 13(2), 95-98.
- Palo, R.T. (1987). Distribution of birch (*Betula* SPP.), willow (*Salix* SPP.), and poplar (*Populus* SPP.) secondary metabolites and their potential role as chemical defense against herbivores. *Journal of Chemical Ecology*, 10, 499-520
- Pan, L., Lezama-Davila, C. M., Isaac-Marquez, A. P., Calomeni, E. P., Fuchs, J. R., Satoskar, A. R., y Kinghorn, A. D. (2012). Sterols with antileishmanial activity isolated from the roots of *Pentalinon andrieuxii*. *Phytochemistry*, (82), 128-135.
- Patočka, J. (2003). Biologically active pentacyclic triterpenes and their current medicine signification. *Journal of Applied Biomedicine*, 1, 7-12.

- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., y Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417-419.
- Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current protocols in Bioinformatics*, 3, Unit3.1.
- Peña, L. M., Yam, A., Knispel, N., Schramek, N., Huber, C. Graßberger, C., Ramírez, F. G., Escalante, F., García, K., Hiebert, M. K., Bacab, M. J., Godoy, G., Bacher, A. y Eisenreich, W. (2014). Isotopologue profiling of triterpene formation under physiological conditions. Biosynthesis of lupeol-3-(3'-R-hydroxy)-stearate in *Pentalinon andrieuxii*. *Journal of Organic Chemistry*, 79(7), 2864-2873.
- Pertsemliadis, A. y Fondon, J.W. (2001). Having a BLAST with bioinformatics (and avoiding BLASTphemy). *Genome Biology*, 2(10), 2002.1-2002.10.
- Qian, K., Kim, S. Y., Hung, H. Y., Huang, L., Chen, C. H. y Lee, K. H. (2011). New betulinic acid derivatives as potent proteasome inhibitors. *Bioorganic & Medicinal Chemistry Letters*. 21, 5944-5947.
- Rong, Q., Jiang, D., Chen, Y., Shen, Y., Yuan, Q., Lin, H., Zha, L., Zhang, Y. y Huang, L. (2016). Molecular cloning and functional analysis of squalene synthase 2 (SQS2) in *Salvia miltiorrhiza* Bunge. *Front Plant Science*, 7, 1274.
- Rzedowski, J. y Calderón, G. (1998). Apocynaceae. Flora del bajío y de regiones Adyacentes. Fascículo 70. Instituto de Ecología, Pátzcuaro, Michoacán. Recuperado en: <http://www1.inecol.edu.mx/publicaciones/resumeness/FLOBA/Vitaceae131.pdf>
- Sandeep, Rajesh, M., Chandan, C., Pradipto, M. y Sumit, G. (2019). Oxidosqualene cyclase and CYP716 enzymes contribute to triterpene structural diversity in the medicinal tree banaba. *New Phytologist*, 222(1), 408-424.
- Sandoval, C. A. (2014). Producción de ácido betulínico en cultivos de raíces transformadas de *Pentalinon andrieuxii* (Müll. Arg.) Hansen & Wunderlin. Tesis de Maestría. Centro de Investigación Científica de Yucatán. México. pp. 92.
- Sawai, S. y Saito, K. (2011). Triterpenoid biosynthesis and engineering in plants. *Frontiers in Plant Science*, 2: 1-8.

- Shashi, M., Ashoke, N. y Gita, R. (1992). Triterpenoids. *Phytochemistry*, 31(1): 2199- 2249.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. y Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7: 539
- Soetaert, S. S., Van Neste, C. M., Vandewoestyne, M. L., Head, S. R., Goossens, A., Van Nieuwerburgh, F. C. y Deforce, D. L. (2013). Differential transcriptome analysis of glandular and filamentous trichomes in *Artemisia annua*. *BMC Plant Biol* (13), 220-234.
- Souza, T. M., Alves, T. B., Pinheiro, K. A., Felipe, L. G., De Lima, G. M., Watanabe, T. F., Barbosa, C. C., Santos, V. A., Lopes, N. P., Valentini, S. R., Guido, R. V., Furlan, M., y Zanelli, C. F. (2016). Friedelin synthase from *Maytenus ilicifolia*: Leucine 482 plays an essential role in the production of the most rearranged pentacyclic triterpene. *Scientific Reports*, (6), 36858.
- Steele, J., Warhurst, D., Kirby, G. y Simmonds M. (1999). *In vitro* and *In vivo* evaluation of betulinic acid as an antimalarial. *Phytotherapy Research*, 13(2), 115-119.
- Suzuki, H, Fukushima, E. O., Shimizu, Y., Seki, H., Fujisawa, Y., Ishimoto, M., Osakabe, K., Osakabe, Y. y Muranaka, T. (2019). *Lotus japonicus* triterpenoid profile and characterization of the CYP716A51 and LjCYP93E1 genes involved in their biosynthesis in planta. *Plant Cell Physiol*, 60(11), 2496-2509.
- Tamura, K., Seki, H., Suzuki, H. *et al.* (2017). CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured stolons of *Glycyrrhiza uralensis*. *Plant Cell Rep*, 36, 437–445.
- Tang, Y., Zhong, L., Wang, X., Zheng, H., y Chen, L. (2019). Molecular identification and expression of sesquiterpene pathway genes responsible for patchoulol biosynthesis and regulation in *Pogostemon cablin*. *Botanical studies*, 60(1), 11.
- Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P. y Osbourn, A. (2014). Triterpene biosynthesis in plants. *Annual Review of Plant Biology*, 65(1), 225-257.

-
- Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. biotechnology of isoprenoids. *Advances in Biochemical Engineering/Biotechnology, Springer, Cham*, 148, 63-106.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. y Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7, 562–578.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., y Leunissen, J. A. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research*, 35, W71–W74.
- Vranová, E., Coman, D. y Gruissem, W. (2012). Structure and dynamics of the isoprenoid pathway network. *Molecular Plant*, 5(2), 318-333.
- Vranová, E., Coman, D. y Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology*, 64(1), 665-700.
- Wang, Z., Gerstein, M. y Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63.
- WFO (2021). *Pentalinon andrieuxii* (Müll.Arg.) BFHansen & Wunderlin. Publicado en Internet: <http://www.worldfloraonline.org/taxon/wfo-0000266800> (Acceso 19 de diciembre de 2021).
- Xu, Z., Peters, R. J., Weirather, J., Luo, H., Liao, B., Zhang, X., Zhu, Y., Ji, A., Zhang, B., Hu, S., Au, K. F., Song, J. y Chen, S. (2015). Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *The Plant Journal*, 82(6), 951- 961.
- Xue, Z., Tan, Z., Huang, A., Zhou, Y., Sun, J., Wang, X., Thimmappa, M. J., Osbourn, A. y Qi, X. (2018). Identification of key amino acid residues determining product specificity of 2, 3-oxidosqualene cyclase in *Oryza* species. *New Phytologist*, 218, 1076-1088.
- Yam-Puc, A, Chee, L., Escalante, F., Arunachalampillai, A., Wendt, O.F., Sterner, O., Peña, L. M. y Godoy, G. (2012). Steroids from the root extract of *Pentalinon andrieuxii*. *Phytochemistry Letters*, 5(1), 45-48.

- Yam-Puc, A. Escalante, F., Pech, M., Chan, M. J., Arunachalampillai, A., Wend, O. F., Sterner, O. y Peña, L. M. (2009). Trinorsesquiterpenoids from the root extract of *Pentalinon andrieuxii*. *Journal of Natural Products*, 72, 745-74
- Yasumoto, S., Fukushima, E. O., Seki, H. y Muranaka, T. (2016). Novel triterpene oxidizing activity of *Arabidopsis thaliana* CYP716A subfamily enzymes. *Febs Letters*, 590(4), 533-540.
- Yogeeswari, P., Sriram, D., (2005). Betulinic acid and its derivatives: a review on their biological properties. *Current Medicinal Chemistry*, 12(6), 657–666.
- Young, N. D., Debellé, F., Oldroyd, G. E., Geurts, R., Cannon, S. B., Udvardi, M. K., Benedito, V. A., Mayer, K. F., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D. R., Meyers, B. C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., ... Roe, B. A. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*, 480(7378), 520–524.
- Zhang, S., Wu, Y., Jin, J., Hu, B., Zeng, W., Zhu, W. Zheng, Y. y Chen, P. (2015). *De novo* characterization of *Panax japonicus* C. A. Mey transcriptome and genes related to triterpenoid saponin biosynthesis. *Biochemical and Biophysical Research Communications*, 466(3), 450-455.
- Zhao, S., Park, C. H., Li, X., Kim, Y. B., Yang, J., Sung, G. B., Park, N., Kim, S. y Park, S. U. (2015). Accumulation of rutin and betulinic acid and expression of phenylpropanoid and triterpenoid biosynthetic genes in Mulberry (*Morus alba* L.). *Journal of Agricultural and Food Chemistry*, 63(38), 8622-8630.
- Zhou, C., Li, J., Li, C. y Zhang, Y. (2016). Improvement of betulinic acid biosynthesis in yeast employing multiple strategies. *BMC Biotechnology*, 16(59).

1.4. Transcritos de la enzima CYP416A

CYP-HA

>HA_TRINITY_DN23331_c0_g1_i1

ATGGAGTTCTTCTATGCTCACTTCTTCTAGCTTAGTCTGATATCTGATCACTCTCTTTCTCTCTATTCTTTAAAGAAAAGAAAGGGTTGCCGGAAGTTGCCGCCGGCAAGACCGGATATCAATAGT
CGGGGAAAGCCTAGAGTTCTCTCTCCGGGATGGAAGGGTCAACCTGAAAAATTCATATTTGATAGGATGTCAAAGTATTCATCCCAAGTGTAGGACTCACCTTTAGTGAAAAAGCGGCTGTTTTCT
GTGGAGCAACCGGAAACAAATCTTGTCTCAAATGAGAACAGCTTGTCCAGGCATGGTGGCCGAACCTCCGTAACAAAGTGTTCGGTCTTCAACACAGACTCATCAAAGAAAGAACATCAAGT
GAGAAAAATGCTCCGAATTTCTTAAACCCGAGCTTTACAAAGATATATTGGCATGATGACACCAATGCTGGAAGACATTTGCCTCCAGCTGGGAAAAACAAATCAAGTGGAGTTTTCCCTCTCA
CGAAATATTTTACCTTCTGGTTAGCATGAGATTGTTTGTGAGTGTGAAGATCCCAAGCATGCGATAAGTTCGCTGATCCTTTCAATCTTTGGCTTCTGGACTTATTTCTCTCCCATAGACTTGGCTG
GAACCCCTTTTAAACAGGCCATCAAGGCATCAATTTTATCAGGAAAGAGCTTTTGGCCATTATCAAAGGAAAGAAAAATGATTTGGCGGAGGAAAAAGCTTCCCAACACAGGATATTTTGTCTCACATG
CTCTTGTACTAGTATGAAAAAGTTCATGAATGAATAGATATTGCTGATAAAAATCTTGGTTTTATTGATCGGAGGACACGACACTGCAAGCTCTGCATGCGCTTTCATTGTCAAATATCTTGTCTGAG
TTACCGGAAATCTATGAGGGAGTCTCAAGGAAACAAATAGAGATTGCAAGTCAAAGGCACCTGGAGAACTACTAACTGGGAAAGACATCAGAAAGATGAAGTATTCTGGAATGTAGCTTGTGAAGTCC
TTAGACTTGCACCACTCCAGGGAGCTTTAGAGAAAGCCATCACTGATTTTATGATAAATGATTTCAATCCCAAAAGGCTGGAAGCTATCTGGAGTCAAACTCAACCCAAAAACCCCTGAAATTT
TTCCAGAGCCGCAAGAAATTTGACCCCTCAAGATTGAGGGATCTGGACCTGCTCTTATACATTTTGGCCCTTTGGTGGAGGACCAAGAAATGTCCTGGAAAAAGAAATGCAAGAAATGGAAATTTG
TTTTATGATGATCACTTAGTCAAAGGTTTCAAGTGGGAGAAAAATCATTCTGATGAGAAAAATTTGTTGATGCAATGCCAATCTCTGCTAAGGGTCTGCCAGTGGCCCTCTTCTCATAAAGCA^{TAA}

CYP-HJ

>HY_TRINITY_DN38732_c3_g2_i1

ATGGTCGGGGAGAGTCTTCAAGTCTCTCAGCCGGCTGAAAGGCCATCTGAAAGTTCATTTTTGACCGTATGGCTAAGTACTCATCAAAGCTTTAGTGGCCATCTTCTTGGTGGAGCCAGCCGCTG
TGTTTTGTGGTCTGCTTCTAACAAATCTTGTCTCAAATGAAACAAACTTGTTCAGCCGCTGGTGGCCGATTTCTGTAACAAAGTTTTCCCATCTCAACACAACTTCATCAAAGAAAGAGGCTATAA
AGATGAGAAAAATGCTCTCAACTCTTAAACCCGAGCTTTGACGCGATATATTGGTATCATGGACCAAAATCGCTCAAAACACACTTTGCTCCGGGTGGGAAAAACAGGATCAAATCGAAGTTTTCCCT
CTTGGCCAAAGAAATATACCTTTTGGTTAGCCGCTAGGTTGTTTGTGAGCGTTGAGGATCCTGCTGAAAGTGAAGAAATTAAGTCTCAACCAATCAACGTTTTGGCTCTCTGTTCCATCGAATTTG
CCCGGAAACCCCTTTCAATCGTGCATCAAGGCCTCGAAACAAATTAGGAAAAATGCTTTTCCATCATTAAACAAAGAAAAATGATTTGGCGGAGGAAAAAGGATCACCACCCCAAGATATATTGTCTCA
CATGCTTTTGTACTAGTATGAAAAAGTTCATGACCTGAAATAGAAATAGCTGATAAGATTTGGTCTTATTAATTTGGCGGATGACACTGCAAGCTCTGTGCACTTTTCTGTCAGAGTTTCTTCTG
TGAAGTCTCCGAGATGTACGAGGAGTCTACAAAGAAACAAATGAGATTGCCAATTCAAAGCACCTGCTGAGTTTCTGCAACTGGGAAAGACATCCAGAAATGATATTCATGGAAGTGGCATGTGA
AGTATTGAGACTTGCACCACTCTTCAAGGCTCTTCAAGAAAGCCCTCAGGATTTATGTTCAATGATTTTCAATGATTTTCCATTCATTCGAAAGGATGGAAGATATTTGAGTGCAAATCAACCCAGAAATCTG
ATTTTTCCCTGAAACCTTAAATTTGACCCCTCAAGATTGAGGGATCTGGTCTGCTCGTATACATTTGTTCCATTTGGTGGAGGACCAAGAAATGTCCTGGAAAAAGAAATGCAAGAAATGGAAAT
CTAGTTTTATGATCAATCTGGTGAAGAGATTCAGATGGGAGAAAAATTCACCCGATGAAAGATTGTTGATGCAATGCCAATCTCTGCAAAAGGACTGCCGCTCCGACTTATCCCAAGCAAGCA^{TG}

CYP-RA

>RA_TRINITY_DN27887_c0_g1_i1

ATGGTCGGGGAGAGTCTTCAAGTCTCTCAGCCGGCTGAAAGGCCATCTGAAAGTTCATTTTTGACCGTATGGCTAAGTACTCATCAAAGCTTTAGTGGCCATCTTCTTGGTGGAGCCAGCCGCTG
TGTTTTGTGGTCTGCTTCTAACAAATCTTGTCTCAAATGAAACAAACTTGTTCAGCCGCTGGTGGCCGATTTCTGTAACAAAGTTTTCCCATCTCAACACAACTTCATCAAAGAAAGAGGCTATAA
AGATGAGAAAAATGCTCTCAACTCTTAAACCCGAGCTTTGACGCGATATATTGGTATCATGGACCAAAATCGCTCAAAACACACTTTGCTCCGGGTGGGAAAAACAGGATCAAATCGAAGTTTTCCCT
CTTGGCCAAAGAAATATACCTTTTGGTTAGCCGCTAGGTTGTTTGTGAGCGTTGAGGATCCTGCTGAAAGTGAAGAAATTAAGTCTCAACCAATCAACGTTTTGGCTCTCTGTTCCATCGAATTTG
CCCGGAAACCCCTTTCAATCGTGCATCAAGGCCTCGAAACAAATTAGGAAAAATGCTTTTCCATCATTAAACAAAGAAAAATGATTTGGCGGAGGAAAAAGGATCACCACCCCAAGATATATTGTCTCA
CATGCTTTTGTACTAGTATGAAAAAGTTCATGACCTGAAATAGAAATAGCTGATAAGATTTGGTCTTATTAATTTGGCGGATGACACTGCAAGCTCTGCATGCACTTTTCTGTCAGAGTTTCTTCTG
TGAAGTCTCCGAGATGTACGAGGAGTCTACAAAGAAACAAATGAGATTGCCAATTCAAAGCACCTGCTGAGTTTCTGCAACTGGGAAAGACATCCAGAAATGATATTCATGGAAGTGGCATGTGA
AGTATTGAGACTTGCACCACTCTTCAAGGCTCTTCAAGAAAGCCCTCAGGATTTATGTTCAATGATTTTCCATTCATTCGAAAGGATGGAAGATATTTGAGTGCAAATCAACCCAGAAATCTCTG
ATTTTTCCCTGAAACCTTAAATTTGACCCCTCAAGATTGAGGGATCTGGTCTGCTCGTATACATTTGTTCCATTTGGTGGAGGACCAAGAAATGTCCTGGAAAAAGAAATGCAAGAAATGGAAAT

CYP-RJ

>RY_TRINITY_DN48275_c0_g3_i2

ATGTCAAAGTATTCATCCCAAGTGTAGGACTCACCTTTAGTGAAGGCGGCTTTTTCTGTGGAGCAACCGGAAACAAATCTTGTCTCAAATGAGAACAGCTTTCAGGCGATGGTGGCCGA
ACTCCGTGAACAAAGTGTTCCTCTCAACACAGACTTCATCAAAGAAAGAGCCATCAAGATGAGAAAAATGCTCCGAATTTCTTAAACCCGAGCTTTACAAAGATATATTGGCATCATGACCCAC
ATTGCTGCAAGACATTTTGGCTCCAGCTGGGAAACAAATCAAGATTGAGTTTTCCCTCTCAGGAAATTTTACCTTCTGGTTAGCATGCAAGATTGTTTGGAGTGTGGAAGATCCCAAGCATGTCGA
TAAGTTCGCTGATCCTTTCAATCTTTTGGCTTCTGGACTTATTTCTCTCCCATAGACTTGGCTGGAACCCCTTTTAAACAGGCCATCAAGGCATCAAAATTTTCAATGGAAGAGCTTTGGCCATTATCA
AGGAAAGAAAAATGATTTGGCGGAGGAAAAAGCTTCCCAACACAGGATATTTGTCTCACATGCTTTGACTAGTATGATGAAAAATGGAAGTTCATGAATGAATTAGATATTGCTGATAAAAATCTTGGTTT
ATTGATCGGAGGACCGACACTGCAAGCTCTGCATGCGCTTTCATTTGCAAAATATCTGCTGAGTACCGGAAATCTATGAGGGAGTCTCAAGGAAACAAATAGAGATTTGCAAGTCAAAGGCACCTGGGA
GAACACTAAACTGGGAGACATACAGAAAGTGAAGTATTCTGGAATGTAGCTTGTGAAGTCTTGAAGTCTTGAAGTCTTCCAGGAGCTTTTACAGAAAGCCATCACTGATTTTCAATGCAATGGATT
CTCAATCCCAAAAGGCTGGAAGCTTATCTGGAGTGAACCTCAACCCCAAAAACCCCTGAAATTTTTCCAGAGCCGCAAGAAATTTGACCCCTCAAGATTGAGGGATCTGACCTGCTCTTATACATTTG
TCCCTTTGGTGGAGGACCAAGAAATGTCCTGGAAAAAGAAATGCAAGAAATGGAATTTCTGTTTTATGATGATCACTTAGTCAAAGGTTCAAGTGGGAGAAAAATCTTCTGATGAGAAAAATTTGTTG
GATCCAATGCCAATCTGCTAAGGCTGCTGCCAGTGGCCCTCTTCTCATAAAGCA^{TAA}

2. Números de accesiones de las secuencias empleadas para el árbol filogenético depositadas en el Genbank del NCBI.

AFK29283.1 [*Salvia miltiorrhiza*]
 ACV88718.1 [*Panax ginseng*]
 AJK30632.1 [*Panax ginseng*]
 AEA41713.1 [*Eleutherococcus senticosus*]
 AAV58897.1 [*Centella asiatica*]
 ADC32654.1 [*Aralia elata*]
 ACN69082.1 [*Diospyros kaki*]
 AXC43359.1 [*Camellia sinensis*]
 AGB05603.1 [*Camellia oleifera*]
 XP_022853448.1 [*Olea europaea* var. *sylvestris*]
 AWM98357.1 [*Osmanthus fragrans*]
 ADX01171.1 [*Bacopa monnieri*]
 XP_011092839.1 [*Sesamum indicum*]
 AFK29284.1 [*Salvia miltiorrhiza*]
 XP_022869491.1 [*Olea europaea* var. *sylvestris*]
 XP_011092466.1 [*Sesamum indicum*]
 AQY54416.1 [*Salvia officinalis*]

EOX97273.1 [*Theobroma cacao*]
AXK50227.1 [*Cucurbita pepo*]
AOV62783.1 [*Morus alba*]
XP_015885090.1 [*Ziziphus jujuba*]
XP_022855568.1 [*Olea europaea* var. *sylvestris*]
AQY54371.1 [*Salvia officinalis*]
AAK63847.1 [*Mentha x piperita*]
ABV08819.1 [*Salvia miltiorrhiza*]
AGQ04160.1 [*Lavandula angustifolia*]
ADV03080.1 [*Bacopa monnieri*]
XP_022879277.1 [*Olea europaea* var. *sylvestris*]
AIY24422.1 [*Jasminum sambac*]
AKQ62666.1 [*Betula platyphylla*]
AAV58896.1 [*Centella asiatica*]
ADK12004.1 [*Aralia elata*]
AAV87903.1 [*Panax ginseng*]
BAL41371.1 [*Glycyrrhiza uralensis*]
XP_024641792.1 [*Medicago truncatula*]
RDY06173.1 [*Mucuna pruriens*]
XP_006606308.1 [*Glycine max*]
BAE53430.1 [*Lotus japonicus*]
XP_023893584.1 [*Quercus suber*]
XP_024020591.1 [*Morus notabilis*]
PWA38718.1 [*Artemisia annua*]
BAA86932.1 [*Taraxacum officinale*]
AFJ19235.1 [*Catharanthus roseus*]
BAF63702.1 [*Olea europaea*]
AFH53506.1 [*Ocimum basilicum*]
AIS39793.1 [*Ilex asprella* var. *asprella*]
AHX97777.1 [*Gentiana straminea*]
ASB17950.1 [*Platycodon grandiflorus*]
ALS20401.1 [*Panax quinquefolius*]
ADK12003.1 [*Aralia elata*]
XP_022895811.1 [*Olea europaea* var. *sylvestris*]
XP_011093795.1 [*Sesamum indicum*]
KZV14753.1 [*Dorcoceras hygrometricum*]
AFH53223.1 [*Ocimum basilicum*]

Monooxygenasas CYP450

NP_199073.1 [*Arabidopsis thaliana*]
AHB62239.1 [*Lotus japonicus*]
BAE94181.1 [*Glycine max*]
ABC59085.1 [*Medicago truncatula*]
BAG68930.1 [*Glycyrrhiza uralensis*]
NP_193268.3 [*Arabidopsis thaliana*]
BAL45199.1 [*Medicago truncatula*]
BAL45200.1 [*Medicago truncatula*]
BAL45207.1 [*Glycyrrhiza uralensis*]
ABC59075.1 [*Medicago truncatula*]
BAL45204.1 [*Medicago truncatula*]
ABG88961.1 [*Avena strigosa*]
ABG88965.1 [*Avena strigosa*]
BAG68929.1 [*Glycyrrhiza uralensis*]
AEY75217.1 [*Panax ginseng*]
AHF45909.1 [*Bupleurum falcatum*]
BAX04008.1 [*Platycodon grandiflorus*]
AFO63031.1 [*Panax ginseng*]
APG38190.1 [*Aquilegia coerulea*]

ALR73782.1 [*Barbarea vulgaris* subsp. *arcuata*]
ALR73781.1 [*Barbarea vulgaris* subsp. *arcuata*]
AED94045.1 [*Arabidopsis thaliana*]
AED94048.1 [*Arabidopsis thaliana*]
ABC59076.1 [*Medicago truncatula*]
BAW34647.1 [*Glycyrrhiza uralensis*]
AHF22083.1 [*Artemisia annua*]
BAX04007.1 [*Platycodon grandiflorus*]
AHF22088.1 [*Maesa lanceolata*]
AFO63032.1 [*Panax ginseng*]
APZ88353.1 [*Eleutherococcus senticosus*]
AEX07773.1 [*Catharanthus roseus*]
ANY30853.1 [*Chenopodium quinoa*]
ANY30854.1 [*Chenopodium quinoa*]
BAJ84106.1 [*Vitis vinifera*]
BAJ84107.1 [*Vitis vinifera*]
AZS32332.1 [*Lagerstroemia speciosa*]
AZS32333.1 [*Lagerstroemia speciosa*]

3. Predicción de la estructura tridimensional de las enzimas de la biosíntesis de ácidos triterpénicos de *P. andrieuxii*

La estructura 3D de las enzimas de la biosíntesis de triterpenos se predijo usando un modelo de estructura por homología en el servidor SWISS-MODEL (<http://swissmodel.expasy.org/>). Se empleó una secuencia por cada gen de los cuatro transcriptomas, eligiendo la secuencia más larga. En el caso de las OSC identificadas se eligió la supuesta lupeol sintasa para predecir su estructura tridimensional.

Se obtuvo la estructura 3D de la enzima SQS (Figura 4.1, A) utilizando como molde una escualeno sintasa humana (No. de acceso 3weg.1.A). Esta plantilla mostró una identidad de 47.43% con la supuesta PaSQS. Para la secuencia SQE se utilizó como molde una escualeno epoxidasa humana (No. de acceso 6c6n.1.A), mostrando una identidad de 49.55% con PaSQE (Figura 4.1, B). Con respecto a PaLUP, se utilizó una OSC humana (Lanosterol sintasa) con número de acceso 1w6k.1.A, con 44.92% de identidad con PaLUP (Figura 4.1, C). Por último, se empleó como molde un citocromo P450: CYP120A1 de unión al ácido retinoico de *Synechocystis* sp. (No. de acceso 2ve3.1.A) con una identidad de 34.35% con PaCYP (Figura 4.1, D).

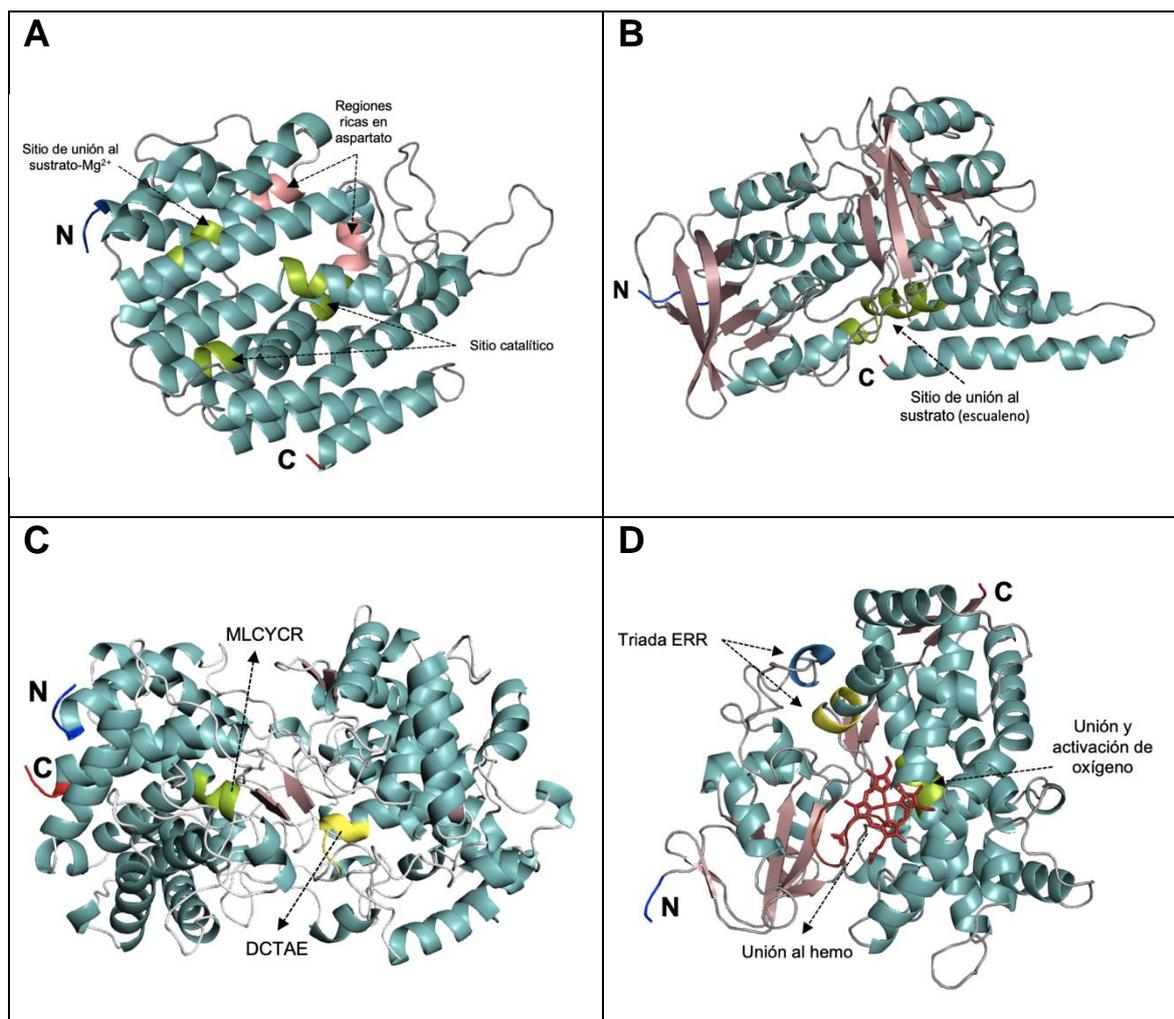


Figura 4.1. Predicción de las estructuras 3D de los productos de la traducción de los transcritos SQS, SQE, LUP y CYP de *P. andrieuxii*. Las estructuras **A.** *PaSQS*, **B.** *PaSQE*, **C.** *PaLUP* y **D.** *PaCYP*, se generaron utilizando el software SWISS-MODEL (<http://swissmodel.expasy.org/>). Los extremos amino (NH₂) y carboxilo (COOH) se presentan en azul y rojo, respectivamente. Además, las hélices α y las láminas β se muestran en verde marino y rosa claro, respectivamente.