

Identificación *In-silico* e *In-vitro* de genes *housekeeping* en raíces *in-vitro* de *Cecropia angustifolia*.

Laura Nathalia Bacca Jaramillo

UNIVERSIDAD ICESI

FACULTAD DE INGENIERÍA, DISEÑO Y CIENCIAS APLICADAS

PROGRAMA DE QUIMICA FARMACÉUTICA

SANTIAGO DE CALI

2024

Identificación *In-silico* e *In-vitro* de genes *housekeeping* en raíces de *Cecropia angustifolia*.

Laura Nathalia Bacca Jaramillo

TRABAJO DE GRADO PARA OPTAR POR EL TÍTULO DE PREGRADO EN QUÍMICA
FARMACEUTICA

Tutor:

Juan Esteban Vivas Moncayo. cMsC

Co-tutor:

Guillermo León Montoya Peláez.PhD

UNIVERSIDAD ICESI

FACULTAD DE INGENIERÍA, DISEÑO Y CIENCIAS APLICADAS

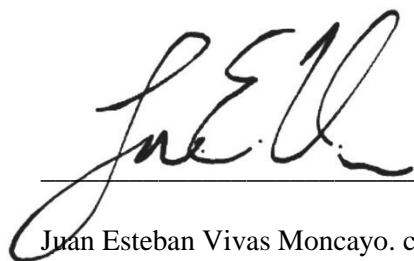
PROGRAMA DE BIOLOGÍA

SANTIAGO DE CALI

2024

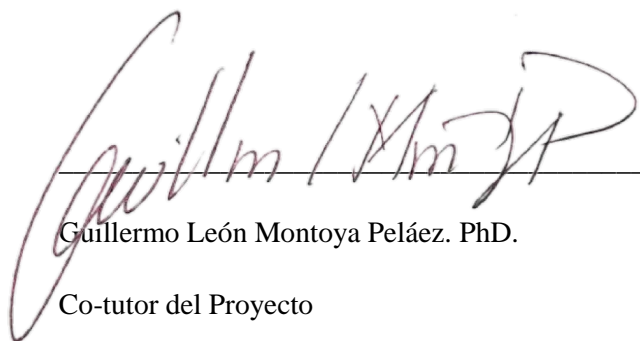


AVAL PARA LA ENTREGA DE PDG II:



Juan Esteban Vivas Moncayo. cMSc.

Tutor del Proyecto



Guillermo León Montoya Peláez. PhD.

Co-tutor del Proyecto



Laura Nathalia Bacca Jaramillo

Estudiante

Tabla de contenido

1. RESUMEN EJECUTIVO.....	6
2. INTRODUCCIÓN.....	6
2.1. Gliceraldehído-3-fosfato deshidrogenasa (GAPDH).....	8
2.2. Actina (ACT).....	9
2.3. β -tubulina (TUB).....	9
3. METODOLOGÍA PROPUESTA.....	10
3.1. Identificación de genes <i>housekeeping</i> y dominios conservados.....	10
3.2. Diseño de primers.....	12
3.3. Extracción de ARN.....	12
3.4. PCR convencional.....	14
3.5. Verificación de la identidad de los productos de PCR.....	15
4. RESULTADOS Y DISCUSIÓN.....	15
4.1. Identificación de los mensajeros codificantes para <i>C. angustifolia</i>	15
4.2. Anotación funcional de genes.....	16
4.3. Descripción estructural de los mensajeros.....	18
4.4. Diseño de primers mRNA específicos para los genes GAPDH, ACT y TUB.....	20
4.5. Extracción de mRNA.....	22
4.6. Verificación <i>in-vitro</i>	23
4.7. Secuenciación Sanger.....	25
5. CONCLUSIONES.....	29
6. AGRADECIMIENTOS.....	29
7. REFERENCIAS BIBLIOGRÁFICAS.....	30

Índice de tablas y figuras

Tabla 1. Protocolo de síntesis de cDNA empleando la enzima M-MuLV RT.....	13
Tabla 2. Termoperfil empleado para la síntesis de cDNA con la enzima M-MuLV RT.	13
Tabla 3. Protocolo de PCR empleando la enzima GreenTaq.	14
Tabla 4. Termoperfil empleado para la amplificación de mRNA.	14
Figura 1. Especies vegetales empleadas para el diseño de la base de datos para la obtención de genes mensajeros en <i>Cecropia angustifolia</i> . Los rombos de colores representan los genes reportados para cada especie.....	16
Tablas 5. Resultados del BlastP para los genes <i>housekeeping</i> en <i>Cecropia angustifolia</i>	17
Tabla 6. Primers diseñados para cada gen y sus características específicas.	21
Figura 3. Alineamientos de los mRNA con su respectivo set de primers empleados para su amplificación del dominio conservado para A. ACT. B. GAPDH. C. TUB. Los alineamientos se realizaron con Clustal O.....	22
Figura 7. Electroferogramas de las secuencias de los genes A) GAPDH y B) TUB.	28

Lista de abreviaturas

Cecropia angustifolia – **C. angustifolia**

Triterpenos pentacíclicos ácidos - **TPAs**

Housekeeping – **HK**

Gliceraldehído-3-fosfato deshidrogenasa - **GAPDH**

Actina – **ACT**

1. RESUMEN EJECUTIVO

Con el fin de realizar aportes al aprovechamiento del potencial químico de la diversidad en la región tropical de los andes colombianos, el laboratorio de investigación en productos naturales de la Universidad Icesi ha venido trabajando con *Cecropia angustifolia* y enfocando especial atención sobre sus triterpenos pentacíclicos ácidos (TPAs), los cuales son reconocidos como sus marcadores quimiotaxonómicos¹. No obstante, las rutas metabólicas en este género en particular no han sido estudiadas a profundidad, por lo que los análisis de los procesos enzimáticos para la producción de sus metabolitos deben basarse en modelos correspondientes a especies distintas. Estos modelos son inherentemente variables dadas las diferencias entre cada especie y *C. angustifolia*. Por lo cual, un proceso de reconocimiento de los procesos metabólicos de la planta debe iniciar con la identificación de genes *housekeeping* (HK) que puedan servir como marcadores biológicos y proporcionar una referencia confiable para comparar la expresión génica en diferentes condiciones ambientales y/o tejidos. De este modo, el objetivo de este proyecto fue evaluar, los genes que codifican las proteínas GAPDH, ACT y TUB como posibles genes *housekeeping*, haciendo uso de herramientas bioinformáticas, técnicas moleculares y cultivos *in vitro* de tejidos radiculares de *C. angustifolia*. Para cumplir con este objetivo, se realizó una revisión bibliográfica de los genes GAPDH, ACT y TUB presentes en especies taxonómicamente cercanas a *C. angustifolia*, luego, empleando diferentes herramientas bioinformáticas se diseñaron primers específicos para cada gen para después realizar una validación *in-vitro* empleando mRNA extraído de tejidos de *C. angustifolia*. Los resultados obtenidos se verificaron mediante secuenciación Sanger, en la cual se encontró la presencia de proteína de biosíntesis de molidopterinas y del gen TUB (β -tubulina) en la muestra de tejido empleada. Los resultados reportados en la presente investigación requieren mayor profundización y sirven de base para futuras investigaciones sobre caracterización de genes HK en *C. angustifolia*.

Palabras claves: *Cecropia angustifolia*, *housekeeping*, expresión génica, GAPDH, ACT.

2. INTRODUCCIÓN

Colombia, catalogado como el segundo país más megadiverso del planeta, atraviesa un momento trascendental de su desarrollo. Con una cifra de aproximadamente 30.736 especies de plantas y un sinnúmero de especies endémicas², su biodiversidad presenta una oportunidad única para implementar enfoques innovadores y sostenibles en la transformación de su base biológica³. Estos enfoques no solo garantizan el desarrollo económico sostenible, sino que también sientan las bases de la bioeconomía colombiana, teniendo en cuenta la importancia de la utilización responsable de las plantas como recursos renovables. Además, se aboga por el cultivo *in-vitro* para disminuir la dependencia de los recursos silvestres y facilitar la identificación de materias primas y compuestos

bioactivos^{4,5} que, tras los procesos de conversión, podrían ser empleados en diferentes sectores de la industria; como el farmacéutico, nutracéutico y cosmético⁶

Una de las especies endémicas de mayor interés en la actualidad es *Cecropia angustifolia*, comúnmente conocida como yarumo negro. Esta especie arbórea perteneciente a la familia Urticaceae en el orden de los Rosales⁷, crece en la región andina del continente americano, y su distribución abarca varios países como Colombia, Venezuela, Bolivia, Ecuador y Perú. *C. angustifolia* muestra una longevidad de alrededor de 35 años y puede alcanzar una altura de hasta 25 metros con un diámetro del tronco de unos 30 centímetros. Su tallo, de tonalidad negruzca y escasamente ramificado, se caracteriza por cicatrices de hojas muy notables; además, desempeña un papel fundamental en la regeneración de la cobertura vegetal⁴.

Esta planta en particular ha recibido una atención significativa en los círculos de investigación científica debido a que se ha reportado que, este tipo de árbol, presenta una acumulación abundante de diferentes componentes bioactivos como los triterpenos pentacíclicos ácidos (TPAs), principalmente en sus órganos radiculares⁸. Estos compuestos pueden tener potenciales aplicaciones como material de referencia, para evaluar la calidad de los extractos u otros componentes que contengan estos compuestos. Esta característica es crucial, ya que facilita la estandarización de perfiles químicos y contribuye al desarrollo de la bioeconomía⁹.

No obstante, la obtención de estos metabolitos secundarios a partir de especímenes silvestres de *C. angustifolia* puede ser un desafío debido a las concentraciones en los que se producen y la variabilidad de factores ambientales que pueden afectar su producción, como las condiciones de alta temperatura, alcalinidad, salinidad y humedad, entre otros factores¹⁰. Por lo anterior, surge la necesidad de emplear los cultivos *in-vitro* como una alternativa para la obtención de estos compuestos bajo condiciones bióticas y abióticas controladas.

Ahora bien, la descripción de la ruta metabólica para la producción de los TPAs en la especie de este género es desconocida, por lo que se deben tomar rutas ya descritas para algunas especies vegetales tales como *Largestroemia speciosa*, *Malus doméstica*, o *Vitis vinífera*⁸. Sin embargo, la inherente variabilidad entre estas rutas, sumada a la falta de estudios que demuestren la equivalencia funcional entre perfiles silvestres y de cultivo *in vitro* limita la aplicación directa de esta información en la investigación y el desarrollo biotecnológico. Estas condicionantes no solo afectan la precisión de la investigación, sino que también pueden representar barreras para maximizar el aprovechamiento del potencial bioeconómico. Por lo tanto, surge la necesidad de caracterizar y evaluar genes *housekeeping* (HK) o genes constitutivos en *C. angustifolia* para establecer una base molecular que permita

normalizar la expresión génica y así, avanzar en la comprensión de la biosíntesis de TPAs en esta especie^{11,12}

Según Joshi, et al. (2022), los HK se definen como aquellos “esenciales para la existencia o mantenimiento de funciones celular básicas¹¹. Se caracterizan por expresarse en todas o la mayoría de los conjuntos de organismos que presenten características o rasgos comunes; independiente del tipo de tejidos, etapa de desarrollo, etapa del ciclo celular o señal externa y condiciones de cultivo¹¹. Debido a que estos genes expresan la interacción entre el ARN polimerasa y el promotor sin regulación adicional¹³, es decir, que su actividad no depende de señales externas o internas que modifiquen el nivel de expresión en un momento dado, desempeñan un punto de referencia para la expresión génica en el área de la biología molecular, contribuyendo al estudio de la variabilidad génica¹⁴.

Sin embargo, la caracterización y estabilidad de genes HK para *C. angustifolia* sigue siendo desconocida, lo que limita considerablemente el análisis subsiguiente para la comprensión y el establecimiento de sus bases moleculares. Por lo tanto, se propone evaluar tres proteínas, cuyos genes puedan desempeñar la función de genes constitutivos: Gliceraldehído-3-fosfato deshidrogenasa (GAPDH), Tubulina (TUB) y Actina (ACT). Se seleccionaron estos genes debido a que, cuentan con la mayor cantidad de información bibliográfica disponible a través de diversas especies vegetales. Además, han sido ampliamente utilizados como genes de referencia en estudios de expresión génica¹⁵. Estos genes están altamente conservados a lo largo de la evolución de diferentes especies, mostrando gran semejanza en su estructura y función filogenética^{16,17}. Su expresión génica puede emplearse como referencia en técnicas cuantitativas como la PCR en tiempo real (qPCR), permitiendo evaluar la expresión de genes asociados con la producción de otros genes de interés en *C. angustifolia*.

2.1. Gliceraldehído-3-fosfato deshidrogenasa (GAPDH)

GAPDH es un polipéptido de 335 aminoácidos muy conservado en toda la escala filogenética. Contiene dos dominios principales: el dominio de unión a NAD⁺ (aminoácidos 1→150) y el dominio catalítico o de gliceraldehído-3-fosfato (aminoácidos 151→335) que contiene una cisteína en el sitio activo¹⁸. Esta preservación filogenética indica que existen semejanzas tanto en la estructura como en la función entre varias especies vegetales¹⁹. Esta es una enzima clave en la vía glucolítica, pues cataliza la conversión de gliceraldehído-3-fosfato (G3P) en 1,3-bifosfoglicerato en presencia de NAD⁺ y fosfato inorgánico¹⁹. Tradicionalmente es reconocida como un gen constitutivo, al ser altamente conservado en diversas especies vegetales a través del tiempo, siendo empleada como control en investigaciones de expresión génica y proteínas¹⁶. Diversos estudios han demostrado su papel en diferentes procesos celulares fundamentales como la regulación del citoesqueleto, la fusión

y el transporte de membranas²⁰, la unión a proteínas G de bajo peso molecular, la apoptosis, la reparación y replicación del ADN y la regulación de la transcripción y la traducción^{21,22}.

2.2. Actina (ACT)

Actina es una proteína de 42 kDa altamente conservada y muy abundante en eucariotas. Está codificada por ocho genes funcionales en las plantas¹⁷, los cuales pueden clasificarse en función de su secuencia y patrón de expresión temporal y espacial, en dos grandes grupos: reproductivos (ACT1, ACT3, ACT4, ACT11 y ACT12) y vegetativos (ACT2, ACT7 y ACT8)²³. De las actinas vegetativas, ACT7 parece ser la más implicada en el crecimiento de las raíces primarias, mientras que ACT2 y ACT8 tienen un papel más relevante en la formación de pelos radiculares²⁴ y en la elongación de células radiculares²⁵. Por ello, en plantas, actina desempeña un papel crucial en el crecimiento y desarrollo debido a su intervención en procesos fisiológicos celulares como el crecimiento y la división celular, la citocinesis y varios procesos de tráfico intracelular^{26,27}. Lo anterior se debe a que esta proteína forma el citoesqueleto de actina, el cual está compuesto por filamentos constituidos por polímeros de G-actina de diferentes isoformas, dependiendo del tipo de célula, tejido o función celular específica²⁸. Dada la importancia de esta proteína en organismos vegetales y su alta conservación entre diferentes especies de plantas, el gen ACT ha sido uno de los más empleados como genes control en estudios de expresión génica²⁹.

2.3. β -tubulina (TUB)

La proteína Tubulina está asociada a la producción de microtúbulos, los cuales son estructuras dinámicas multiproteicas que intervienen en muchas funciones celulares básicas y están formados por una subunidad α -tubulina (TUA) y una subunidad β -tubulina (TUB)^{30,31}. Los genomas de las plantas terrestres superiores y los metazoos codifican estas dos proteínas centrales en grandes familias multigénicas^{32,33}. Los eucariotas unicelulares, como las levaduras y las algas verdes, codifican una o dos subunidades diferentes de α y β -tubulina. La complejidad de las redes de genes de tubulina aumenta con la multicelularidad: los eucariotas superiores codifican de seis a nueve isotipos de tubulina de cada subunidad³². Los isotipos comparten hasta un 99% de identidad, y la mayoría de las diferencias residen en los extremos carboxi-terminales³².

Dada su expresión ubicua, estabilidad y alto grado de conservación entre sus diferentes isoformas, la β -tubulina se emplea comúnmente en estudios de expresión génica en organismos vegetales como un gen constitutivo que permite establecer un parámetro base de expresión para la evaluación de diversos genes de interés³⁴⁻³⁶.

Finalmente, los genes HK identificados pueden emplearse en técnicas como la reacción en cadena de la polimerasa con transcripción inversa cuantitativa (RT-qPCR), la cual se emplea frecuentemente para cuantificar la abundancia del ARN que se encuentra en una muestra de interés¹⁵. Esta técnica es exclusiva para conocer la expresión génica¹⁵, en vista de su especificidad, precisión, sensibilidad, reproducibilidad y amplio empleo en investigación a partir de los genes HK seleccionados anteriormente³⁷. No obstante, antes de poder realizar una medición de dicha expresión, se debe verificar que el producto amplificado mediante los primers diseñados, corresponde a la región del gen de interés cuyos niveles de expresión se desean medir. Para ello, es necesario emplear un método de secuenciación como Sanger, que permite conocer la secuencia de una muestra de cDNA amplificada³⁸. Los resultados de la secuenciación Sanger se comparan, mediante herramientas bioinformáticas, con secuencias de referencia encontradas en bases de datos como NCBI³⁹. Una vez realizado este análisis, se espera que las secuencias del gen amplificado y de las referencias obtenidas de bases de datos, sean muy similares, lo cual indica una alta precisión de los primers diseñados inicialmente. La información obtenida anteriormente será la base para futuros experimentos de medición de expresión génica de diferentes genes de interés en cultivos *in-vitro* de *Cecropia angustifolia*.

Por lo anterior, y con el fin de proporcionar información valiosa para el futuro control molecular, que permitan comparar la expresión de otros genes en diferentes condiciones de cultivo, el proyecto tiene como propósito, para el año 2024, evaluar si los genes que codifican las proteínas GAPDH, TUB y ACT cumplen la función de *housekeeping* en las raíces de *C. angustifolia*. Por tal motivo se plantea el objetivo general, que fue, evaluar, para el año 2024, los genes que codifican las proteínas GAPDH, ACT y TUB como posibles genes *housekeeping* en las raíces de *Cecropia angustifolia*.

3. METODOLOGÍA PROPUESTA

3.1. Identificación de genes *housekeeping* y dominios conservados.

Con el objetivo de obtener las secuencias de genes de referencia (HK) para *C. angustifolia*, se llevó a cabo una búsqueda bibliográfica para identificar los posibles mensajeros que codifican las proteínas GAPDH, TUB y ACT en las raíces de esta planta. Inicialmente, se realizó una búsqueda bibliográfica centrada en determinar y recopilar secuencias de los HK de interés en especies pertenecientes a los grupos taxonómicos más cercanos al orden de *pentapétalas*, *rosids* y *rosales*.

La elección de estos genes se realizó con base en artículos científicos publicados en bases de datos indexadas como ScienceDirect⁴⁰, Nature Portfolio⁴¹, Pubmed⁴², entre otros, enfocándose en artículos

publicados entre enero de 2005 y enero de 2024. Para maximizar la efectividad de la búsqueda, se exploraron temas, títulos, resúmenes y palabras clave pertinentes tales como: “*Housekeeping gen*”, “Conserved domains”, entre otros. Estos términos se combinaron empleando operadores lógicos como “AND”, “OR” y “NOT” para mejorar la precisión de los resultados de la búsqueda. Posteriormente, la información extraída de los documentos seleccionados fue organizada mediante el uso del gestor bibliográfico Mendeley.

A través del repositorio NCBI (<https://www.ncbi.nlm.nih.gov/>), se llevó a cabo la búsqueda de secuencias de ARNm, empleando de igual manera aquellas reportadas como "PROVISIONAL" debido a la carencia de información asociada. Es relevante señalar que, aunque estas secuencias se encuentren con dicha etiqueta, su utilización resulta ser necesaria en ausencia de secuencias confirmadas o completas para llevar a cabo determinadas investigaciones o análisis, como es este caso. Posteriormente, se recopilaron los números de acceso de la base de datos eligiendo especies taxonómicamente cercanas. Los resultados de esta búsqueda se organizaron en tres archivos de texto separados: uno para los números de acceso relacionados con GAPDH y los otros dos para los asociados con TUB y ACT, respectivamente en conjunto con el organismo donde fue reportado. Seguidamente, se empleó la herramienta Batch Entrez (<https://www.ncbi.nlm.nih.gov/sites/batchentrez>) para realizar búsquedas simultáneas y obtener archivos FASTA que contengan las secuencias de nucleótidos y proteínas correspondientes a los genes identificados.

A continuación, se utilizó el software Galaxy⁴³ para realizar un análisis BLASTn⁴⁴ y un BLASTp⁴⁵ empleando las secuencias obtenidas. Se empleó dos conjuntos de datos: el primero para el análisis de codificación de secuencias (TransDecoder CDS) y el segundo para el análisis de las secuencias de proteínas (TransDecoder PEP). Este paso permitió identificar las regiones y/o secuencias codificantes de los genes y los posibles productos proteicos. El archivo TransDecoder consiste en secuencias de ARNm codificantes, el cual se elaboró en el marco teórico de la tesis “Análisis de la expresión génica y metabólica asociada con la acumulación de metabolitos secundarios derivados de triterpenos pentacíclicos ácidos en raíces de *Cecropia angustifolia*” desarrollada por Juan Esteban Vivas en el año 2022 (datos no publicados)⁴⁶. Estas secuencias se derivan del procesamiento de la secuencia de *Cecropia obtusifolia*, obtenida de uno de los ensayos control del artículo “Integrated Analysis of the Transcriptome and Metabolome of *Cecropia obtusifolia*: A Plant with High Chlorogenic Acid Content Traditionally Used to Treat Diabetes Mellitus”.⁴⁷ Esta nueva lista se filtró en función de tres parámetros: Porcentaje de identidad >60%, E-value < 10⁻¹⁰ y longitud de la secuencia, el cual se consideró como tamaño óptimo la longitud mayor a 100 aa.

Posteriormente, se llevó a cabo un alineamiento de las secuencias obtenidas con los eventos únicos de los contigs resultantes empleando la herramienta Unipro UGENE⁴⁸. Esto permitió visualizar y comparar similitudes y diferencias entre las secuencias de referencia y las secuencias identificadas en *C. angustifolia*, proporcionando una perspectiva más detallada sobre los dominios conservados de los genes y las posibles variantes presentes en las raíces de la planta.

Finalmente, se realizó una búsqueda bibliográfica de los posibles dominios conservados para GAPDH, TUB y ACT, los cuales se resaltarán en los alineamientos. Este enfoque garantiza un diseño preciso de primers, evitando problemas como amplificaciones inespecíficas, asegurando la sensibilidad y especificidad necesarias para obtener el resultados control para la realización del procedimiento de RT-PCR.

3.2. Diseño de primers.

Para verificar la presencia de los genes en las muestras vegetales *in-vitro*, se emplearon herramientas bioinformáticas como Benchling (<https://www.benchling.com/>) para diseñar los primers destinados a la amplificación de los genes identificados previamente, basándose en los dominios conservados encontrados para GAPDH, TUB y ACT. Además, se hará uso de la plataforma web Oligo Analyzer de Eurofins Genomics (<https://eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/>) para evaluar la calidad de los primers en términos de su porcentaje de guanina – citocina (GC) entre 55 – 65%, temperatura de fusión (Tm) 60 -65 °C con una diferencia <5°C, longitud del amplicon máxima 22° pb y mínima de 146 pb, energía libre de heterodímeros y homodímeros > -10.0 kcal/mol, estructura secundaria (Hairpin) < -10 y longitud del primer entre 18-23 pb.

Una vez generados los primers, se procedió a verificar la amplificación del amplicón deseado utilizando la plataforma Primer Design Tools(<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) y realizando un análisis BLAST. Se seleccionó *Vitis vinifera* como organismo de interés para este propósito, dada su amplia disponibilidad de datos y su relevancia como referencia para *C. angustifolia* en estudios previos.

3.3. Extracción de ARN

Para la identificación de los genes a estudiar por PCR, se procedió a la extracción del ARN de los cultivos *in-vitro* de *C. angustifolia*. Para ello, se extrajeron las raíces del medio de cultivo en una cabina de flujo laminar procurando eliminar la mayor cantidad de medio/humedad superficial con toallas estériles. Una vez realizado este paso, el material vegetal se sometió a un proceso de maceración mediante nitrógeno líquido con el fin de homogeneizarlo, logrando un peso entre 70 – 100 mg. Por último, el material molido se agregó en los Power Bead Tubes del kit de extracción

RNeasy Power Plant Kit (50) (@Cat. No. 13500-50 QIAGEN). Posteriormente, se procedió a la extracción siguiendo las especificaciones del fabricante. Este procedimiento se realizó acorde con la metodología propuesta por Juan Esteban Vivas (2022) (datos no publicados) ⁴³

El RNA extraído se almacenó a -20°C, y su concentración se cuantificó con la ayuda del equipo NanoDrop-2000 (Thermo Fisher Scientific), en el cual se evaluó la calidad del ARN, considerando la relación A_{260}/A_{280} (presencia de proteínas) y A_{260}/A_{230} (presencia de fenoles), para asegurar que se encuentre próximo a 2.0, cumpliendo con la calidad requerida.

Las librerías de cDNA, se emplearon para el envío de la enzima transcriptoma englan bik M-MuLV RT (Cat No. M0253L, New England Biolabs). Teniendo en cuenta las concentraciones obtenidas se realizó ajustes a las cantidades de los componentes del termoperfil y el mix⁴³, los cuales se pueden observar en las **Tablas 1 y 2** respectivamente. Asimismo, el cDNA obtenido se cuantificó empleando el NanoDrop.

Tabla 1. Protocolo de síntesis de cDNA empleando la enzima M-MuLV RT.

Componentes	Volumen
Muestra de RNA	1000 ng/reacción
d(T)18	2,5 µM
DNTPs	0.25 µM
10 X M-MuLV Buffer	0.5 X
M-MuLV RT (200)	5 UI/µL
Volumen final	40 µL

Tabla 1. Termoperfil empleado para la síntesis de cDNA con la enzima M-MuLV RT.

Termoperfil		
Paso	Temperatura	Tiempo
Desnaturalización inicial	65 °C	5 minutos
Síntesis de cDNA	42 °C	1 hora y 20 minutos
Inactivación de la enzima	65 °C	20 minutos
Almacenamiento	4 °C	INFINITO

El cDNA obtenido anteriormente, se empleó para realizar la PCR y la electroforesis en geles de agarosa HyAgarose™ (Cat No. R9012 HR-500g, HydraGene) al 2% en buffer TAE.

3.4. PCR convencional

Para la amplificación de las secuencias mensajeras diana, se utilizaron los cebadores formulados en la sección 5.2 junto con la enzima Green Taq® DNA Polymerase (Cat No. E00043, GenScript. Se llevó a cabo una estandarización de los cebadores mediante un gradiente de temperatura que oscilaba entre 55 - 60 °C para GAPDH y ACT, y otro gradiente de temperatura entre 60 - 65 °C para TUB. El objetivo principal de esta fase era determinar la temperatura de alineación óptima para los cebadores. El protocolo de PCR y su termoperfil, se encuentran en las **Tablas 3 y 4**.

Tabla 2. Protocolo de PCR empleando la enzima GreenTaq.

Componentes	VCX volumen
Buffer GoTaq 5X	1 X
Cloruro de Magnesio	1.25 mM
dNTPs	0.2 Mm
Primer Forward	0.5 mM
Primer Reverse	0.5 mM
GreenTaq® Polimerasa	5 UI/uL
Agua DEPC	c.s.p
Muestra de cDNA	300 ng/reacción
Volumen final	50 µL

Tabla 3. Termoperfil empleado para la amplificación de mRNA.

Paso	Termoperfil		Tiempo
	Temperatura		
Desnaturalización inicial	95 °C		3 minutos
N° Ciclos	Ciclado de 3 pasos		
Desnaturalización	95 °C		1 minutos
Alineamiento	GAPDH – ACT	TUB	1 minutos
	50-60 °C	60 – 65 °C	
Extensión	72 °C		30 segundos
N° Ciclos	40 ciclos		
Extensión final	72 °C		5 minutos

Almacenamiento	4 °C	INFINITO
----------------	------	----------

3.5. Verificación de la identidad de los productos de PCR.

Una vez obtenidos los productos de PCR para cada uno de los genes de interés en el presente estudio, se validaron los resultados mediante Secuenciación Sanger, la cual fue realizada por el Laboratorio de Medicina Genómica de la Universidad Icesi (LMG) en el Analizador Genético 3500 de Applied Biosystems (Thermo Fisher Scientific). Para esto, los productos de PCR fueron purificados empleando el kit QIACquick® de QIAGEN (n.º de catálogo 28104). Luego, la reacción de secuenciación fue realizada con el kit BidDye Terminator V3.1 de Applied Biosystems™ (n.º de catálogo 4337455) siguiendo las instrucciones del fabricante para dos reacciones independientes empleando el primer Forward y Reverse, respectivamente. Una vez terminado el proceso de secuenciación, los productos secuenciados fueron purificados con el kit BidDye® Xterminator™ Applied Biosystems™ (número de catálogo 4376486) siguiendo las instrucciones del fabricante y realizando una modificación en el tiempo de agitación de 40 a 50 minutos por 2800 rpm.

4. RESULTADOS Y DISCUSIÓN

4.1. Identificación de los mensajeros codificantes para *C. angustifolia*.

Inicialmente, se realizó una búsqueda bibliográfica para identificar los genes mensajeros de interés que cumplen funciones de *housekeeping* en la especie *Cecropia*. Debido a la falta de información específica para esta especie, se procuró emplear datos de otras especies taxonómicamente cercanas. Sin embargo, no se encontró suficiente información relevante en estas plantas, lo cual puede deberse a la controvertida clasificación intra-familiar de Urticaceae, debatida durante más de un siglo⁴⁹.

Además, Cecropiaceae no es un grupo monofilético, es decir, sus miembros no descienden de un único ancestro común según el análisis filogenético molecular. Esto cuestiona la validez de usar únicamente rasgos morfológicos para definir esta familia. Los datos moleculares indican que Cecropiaceae es bifilético, con ambos linajes anidados dentro de Urticaceae^{49,50}. Esto sugiere que las distinciones morfológicas entre Cecropiaceae y Urticaceae no están respaldadas por evidencia genética. La incongruencia entre los rasgos morfológicos y la filogenia molecular subraya la necesidad de reevaluar la clasificación de Cecropiaceae e integrar datos moleculares para refinar los límites taxonómicos dentro de Urticaceae⁴⁹.

Esta incorrecta clasificación taxonómica puede llevar a errores significativos al usar datos genéticos de especies aparentemente cercanas. Si Cecropiaceae no está correctamente clasificada, se podrían estar comparando especies que no son tan cercanas filogenéticamente como se pensaba, lo que podría

explicar la variabilidad y los errores en los resultados. Por tanto, la búsqueda bibliográfica realizada permitió recopilar números de accesos a genes reportados en las bases de datos del NCBI, aunque muchos de estos genes estaban etiquetados como "PROVISIONAL". Esto se debe a que, este tipo de información no cuenta con verificación *in-vitro* o *in-vivo*, lo que puede resultar en variaciones no determinadas por eventos de llamado incorrecto de bases. Esta verificación es esencial para garantizar la certeza y fiabilidad de la expresión génica en la especie anotada, así como la reproducibilidad de su detección en múltiples especímenes de la misma especie⁵¹. Sin embargo, no se disponía de información verificada suficiente para muchos HK en *Cecropia*.

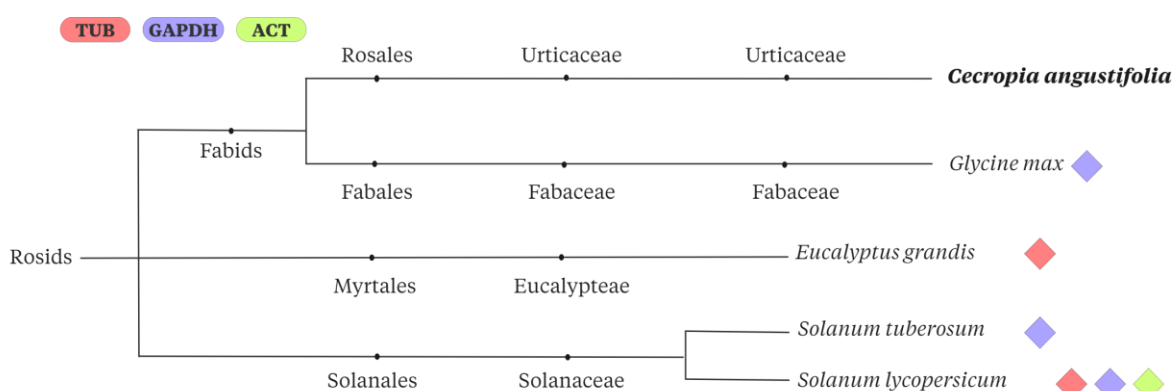


Figura 1. Especies vegetales empleadas para el diseño de la base de datos para la obtención de genes mensajeros en *Cecropia angustifolia*. Los rombos de colores representan los genes reportados para cada especie.

En la Figura 1, se pueden observar las especies taxonómicamente cercanas que se utilizaron para recopilar información sobre los genes mensajeros de *Cecropia*. Aunque estas especies pertenecen a la subclase Rosids, están filogenéticamente distantes entre sí⁵². Esto se refleja en el hecho de que, aunque *Solanum lycopersicum* contiene los tres genes de interés, su comparación con *Cecropia* es problemática debido a las marcadas diferencias morfológicas y ecológicas entre ambas especies.

4.2. Anotación funcional de genes

Una vez obtenidos los datos a partir de la búsqueda bibliográfica, se generó un listado de números de acceso, a los cuales se les realizó una anotación funcional utilizando la herramienta BlastP⁴⁵, empleando los contigs obtenidos del archivo TransDecoder⁴⁶ (ver sección 4.1).

En este archivo se llevó a cabo un ensamblaje *de novo* para la especie *C. angustifolia* cuyo genoma no ha sido secuenciado hasta la fecha⁴⁶. Debido a la falta de especies de referencia, no se pueden asumir los codones de inicio o final y por lo tanto se desconocen los marcos abiertos de lectura, lo que puede resultar en una variedad de proteínas no funcionales (y sus isoformas) a partir de la secuencia nucleotídica obtenida del TransDecoder^{46,53}.

El objetivo de este proceso es obtener una lista de contigs, correspondientes a la tabla 5, y alinearlos para contrastar su conservación con las proteínas de referencia. De esta manera, se asegura que los dominios conservados sean característicos de ese tipo de proteínas. De lo contrario, la anotación funcional podría generar falsos positivos o anotaciones inespecíficas^{54,55}.

Los falsos positivos pueden surgir si las secuencias anotadas como funcionales en realidad no corresponden a proteínas específicas debido a la falta de precisión en la identificación de los marcos de lectura⁵⁵. Sin una secuencia de referencia adecuada, los programas de anotación pueden asignar funciones incorrectas a secuencias que no están verdaderamente conservadas. Esto ocurre porque las herramientas como BlastP pueden identificar similitudes con proteínas conocidas basadas en alineaciones parciales o regiones no conservadas, llevando a errores en la identificación de las proteínas^{55,56,45}.

Tablas 5. Resultados del BlastP para los genes *housekeeping* en *Cecropia angustifolia*.

Identificación del contig	Porcentaje de identidad (%)	Longitud de alineación	E- value
ACT			
TRINITY_DN247_c0_g1_i5.p1	98,143	377	0
TRINITY_DN247_c0_g1_i7.p1	97,878	377	0
TRINITY_DN247_c0_g1_i1.p1	96,817	377	0
TRINITY_DN247_c0_g1_i3.p1	96,817	377	0
GAPDH			
TRINITY_DN444_c3_g1_i1.p1	90,691	333	0
TRINITY_DN444_c3_g2_i1.p1	90,419	334	0
TRINITY_DN444_c3_g1_i1.p1	90,149	335	0
TRINITY_DN444_c3_g1_i1.p1	90,691	333	0
TUB			
TRINITY_DN2487_c0_g1_i11.p1	95,937	443	0
TRINITY_DN2487_c0_g1_i1.p1	95,701	442	0

TRINITY_DN2487_c0_g1_i13.p1	95,516	446	0
TRINITY_DN2487_c0_g1_i12.p2	95,455	440	0
TRINITY_DN2487_c0_g1_i10.p1	93,995	433	0

La **Tabla 5** muestra los contigs (secuencias continuas de ADN ensambladas) obtenidos para cada uno de los genes, junto con sus respectivos porcentajes de identidad, longitudes y E-value. Dado que los porcentajes de identidad son muy elevados dentro de cada subgrupo, es posible identificar las secuencias más conservadas para cada gen. Estas secuencias permiten la identificación de los marcos abiertos de lectura de cada gen y sus dominios peptídicos más conservados.

4.3. Descripción estructural de los mensajeros.

Para la identificación de los dominios conservados, se optó por utilizar secuencias de proteínas (PEP) en lugar de secuencias de nucleótidos. Esta decisión se fundamenta en tres razones: En primer lugar, las secuencias de proteínas facilitan la lectura y análisis en comparación con las secuencias de nucleótidos. Mientras que las secuencias de nucleótidos representan la información genética base por base, las secuencias peptídicas muestran los aminoácidos codificados. Esto simplifica el análisis y facilita la identificación de patrones y dominios conservados en las proteínas.

En segundo lugar, el uso de secuencias de proteínas permite tener en cuenta las mutaciones silenciosas. Una mutación silenciosa es un cambio en el nucleótido que no altera el aminoácido codificado debido a la redundancia del código genético⁵⁴. Al analizar las secuencias de proteínas, se puede observar que los aminoácidos conservados permanecen sin cambios a pesar de estas mutaciones, lo que permite conocer la conservación de los dominios proteicos^{53,56}.

Finalmente, la base de datos utilizada para el análisis en TransDecoder⁴⁶ fue alimentada con proteínas de la familia Urticaceae. Dado que esta familia puede contener proteínas diferentes entre especies y no hay una única especie representativa, las secuencias de proteínas permiten una comparación más directa y significativa. Además, la clasificación intrafamiliar de Urticaceae, especialmente para *C. angustifolia*, puede ser controvertida, lo que hace que el uso de secuencias de proteínas sea más adecuado para identificar dominios conservados entre diferentes miembros de la familia.

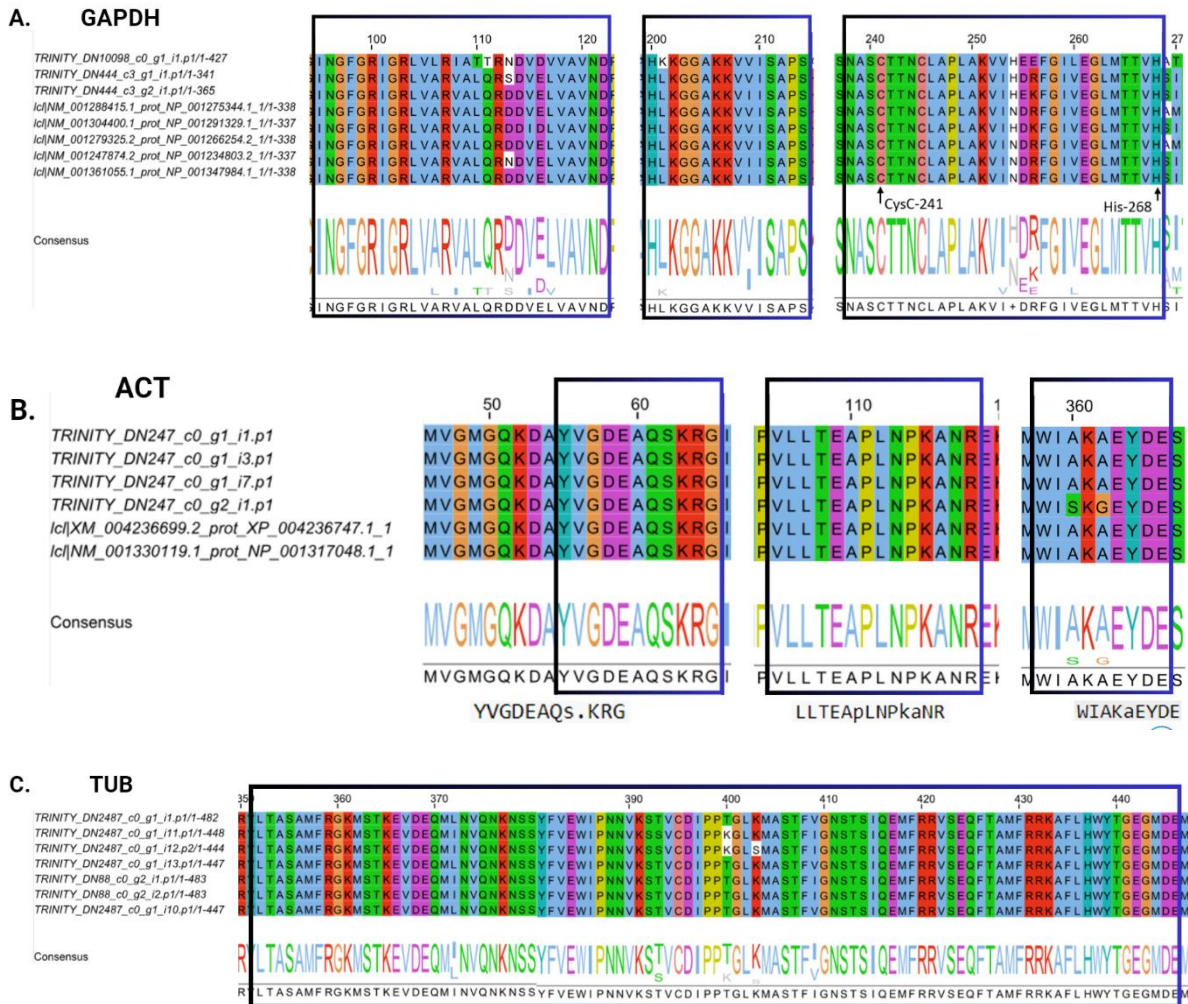


Figura 2. Dominios conservados para set de primers. **A.** ACT. **B.** GAPDH y **C.** TUB.

En la figura 2 se muestra el resultado de los alineamientos para los dominios funcionales de las proteínas GAPDH, ACT y TUB, respectivamente. Se identificaron regiones altamente conservadas en los dominios de cada una de estas proteínas. En la sección A, la alineación se centra en dos regiones altamente conservadas de GAPDH: una corresponde al dominio de unión a la coenzima y la otra al dominio catalítico. En la primera región (N-terminal), se encuentra la secuencia estrictamente conservada NGFGRIGR⁵⁷, junto con otros residuos cruciales para la unión del cofactor de nucleótido de piridina (Asp-122). La segunda región, ubicada en el centro de las secuencias GAPDH, incluye varios residuos que participan en la unión del sustrato, como la cisteína catalítica (Cys-241), una segunda cisteína cercana al sitio activo (Cys-245) y His-181, fundamental para activar Cys-241⁵⁷. Estas características conforman un dominio conservado. Sin embargo, dentro del recuadro verde, se

observan degeneraciones de los péptidos, que podría afectar la funcionalidad de la proteína GAPDH. Lo que puede dar lugar a cambios en los aminoácidos que se traducen a partir de ellos, alterando las propiedades estructurales o funcionales de la proteína. Esto podría resultar en variaciones en la capacidad de unión a sustratos, en la eficiencia catalítica o en la estabilidad de la proteína.

En las secciones B y C, aunque para actina y tubulina hay similitud entre los clústers de secuencias anotadas para la misma proteína y sus referencias, al observar el dominio de referencia, podemos notar discrepancias en las letras del histograma. Estas variaciones pueden señalar diferencias sutiles, pero potencialmente significativas en la secuencia de aminoácidos dentro del dominio⁵⁸. Aunque la variabilidad general pueda ser baja, incluso pequeños cambios en la secuencia pueden tener repercusiones en la estructura y función de la proteína⁵⁶. Un ejemplo de esto es la secuencia peptídica del dominio de unión de actina de referencia WIAKaEYDE, donde se encuentra una variante en la letra "a", resultando en WIAKYEYDE.

Además, la variabilidad presente en la secuencia peptídica representa un factor vital para el diseño de los primers para cada gen, ya que si estos no cuentan con una secuencia conservada a la cuál anillarse en la PCR, es posible que no se obtengan productos de PCR o que se obtengan inespecificidades; para solventar este inconveniente, es importante garantizar la conservación de la secuencia complementaria al extremo 3' OH con el fin de que se sinteticen nuevas cadenas de ADN a partir de regiones con un mayor grado de conservación⁵³.

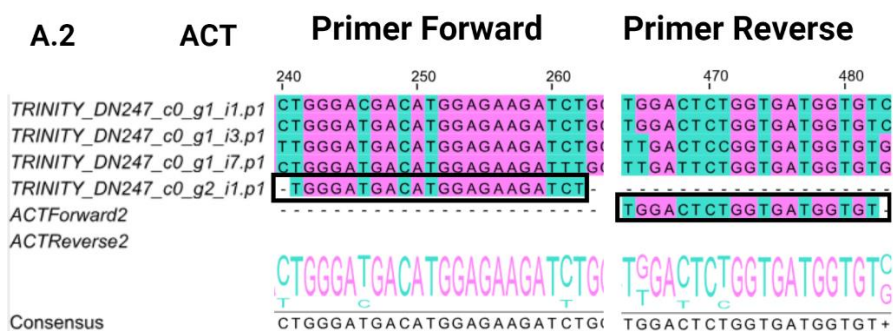
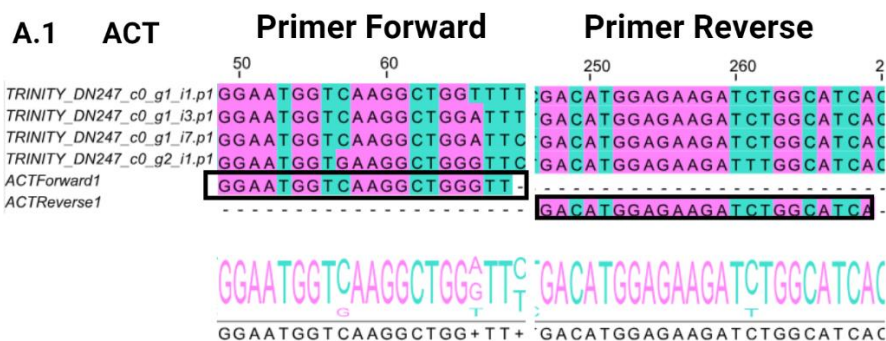
4.4. Diseño de primers mRNA específicos para los genes GAPDH, ACT y TUB.

Una vez identificados los dominios funcionales de interés, se realizó el diseño de primers empleando como referencias el dominio de unión a la coenzima y el dominio catalítico para GAPDH. Para actina y tubulina, se consideró que el diseño de cebadores abarcara la mayor parte del dominio. Cabe resaltar que, debido a limitaciones de tiempo y recursos, no se tuvo en cuenta la degeneración de los codones en el diseño, lo que ocasionó que los primers no fueran completamente específicos. En la Tabla 6 se presentan los primers empleados.

En la Figura 3, se muestra la ubicación de los primers diseñados en las regiones de los mensajeros que se amplificaron. Es importante señalar que, únicamente se presentan los alineamientos de los primers que lograron amplificación, con la excepción de los primers ACTF1R1 y ACTF2R2, los cuales no generaron ningún producto.

Tabla 6. Primers diseñados para cada gen y sus características específicas.

No.	Name	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Product length
1	ACTF1	GGAATGGTCAAGGCTGGGTT	Plus	20	49	68	58,8	55	221
	ACTR1	GACATGGAGAAGATCTGGCAT CA	Minus	23	247	268	57,5	47,8	
2	ACTF2	TGGGATGACATGGAGAAGATCT	Plus	22	241	262	56,2	45,5	242
	ACTR2	TGGACTCTGGTGTGTTGTT	Minus	19	446	482	56,6	50	
3	GADPHF1	GATCAAGATCGGAATCAACGGATT	Plus	24	270	293	55,1	41,7	351
	GADPHR1	GAAGGGTGGTGCCAAGAA	Minus	18	603	620	55,2	55,6	
4	GADPHF2	GTTTGTGTGGGTGTCAATGAG	Plus	22	657	678	54,6	45,5	214
	GADPHR2	TGGAGAGGTGGAAGAGCT	Minus	18	853	870	55,1	55,6	
5	TUBF1	TCTTCCGCCCGGACAACCTTCGT	Plus	21	364	384	66.37	59.09	142
	TUBR1	CCGAGAACTGCGACTGCTTGCA	Minus	22	484	505	65.79	59.09	



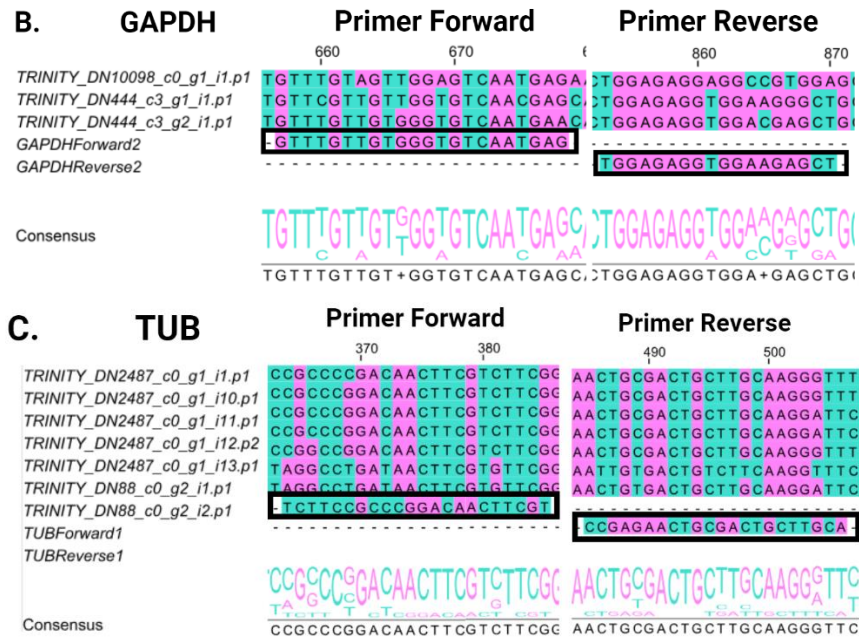


Figura 3. Alineamientos de los mRNA con su respectivo set de primers empleados para su amplificación del dominio conservado para A. ACT. B. GAPDH. C. TUB. Los alineamientos se realizaron con Clustal O.

Como se aprecia en la Figura 3 para ACT, no existe tanta variación en las secuencias codificantes, observándose que el primer reverse presenta solamente una variación. En contraste, GAPDH y TUB presentan algún grado de degeneración en relación con sus respectivas secuencias de referencia. Lo anterior se debe a que, si bien los contigs en la **Tabla 5** muestran un porcentaje de identidad muy elevado, una vez estos se comparan con otras secuencias homólogas, se pueden encontrar variaciones en la secuencia peptídica que indiquen variaciones significativas en la formación de sus codones. Estas diferencias pueden atribuirse a isoformas de las mismas proteínas^{58,59}. Además, estas presentan un E-value cercano a 0, lo que me indica que la coincidencia observada es menos probable que haya ocurrido al azar, lo que sugiere la similitud entre las secuencias de GAPDH y TUB es confiables.

4.5. Extracción de mRNA.

Para llevar a cabo las pruebas moleculares que permitan verificar el posible uso de los genes GAPDH, ACT y TUB como genes HK en *C. angustifolia*, se realizó la extracción de ácidos nucleicos a partir de una muestra de tejido vegetal de esta especie. Dada la sensibilidad de la técnica de PCR ante una posible contaminación que disminuya su eficiencia, es importante resaltar que la integridad del

mRNA extraído es indispensable para el correcto desarrollo de todas las pruebas y experimentos a realizar.

Si bien una célula bajo la influencia de factores internos y del entorno externo es una forma viva, dinámica y variable, se puede concluir que, en diversas muestras habrá una gran variabilidad, tanto en las cantidades absolutas como en el contenido relativo de cada ARNm. Los errores en el pipeteado y la transferencia del material pueden ser una de las razones de dicha variabilidad, lo que afecta significativamente la amplificación de productos de PCR de interés y las pruebas subsiguientes⁶⁰.

Ahora, no solo es importante considerar la calidad del ARN extraído del tejido vegetal recolectado, sino también la variabilidad en la expresión que presentan los genes bajo diferentes condiciones bióticas y abióticas. Varios genes, entre ellos ACT, TUB y GAPDH, son utilizados habitualmente como genes de referencia para ensayos de expresión génica en diferentes especies vegetales, pero se ha reportado que estos presentan una expresión variable entre diferentes especies, tejidos o condiciones ambientales⁶¹⁻⁶³. Por ejemplo, se ha reportado que la proteína GAPDH presenta una expresión estable en el pasto bufal (*Buchloe dactyloides*) bajo un tratamiento por calor⁶⁴, y en las hojas y raíces de *Lolium arundinaceum* bajo estrés por sequía⁶⁵; sin embargo, se ha comprobado que su expresión es inestable en especies como el arroz (*Oryza sativa*) en diversas condiciones ambientales⁶⁶.

En el caso de ACT y TUB, su uso como HKs ha sido recomendado para tratamientos con estrés por salinidad y sequía en *Vigna mungo*⁶⁷; no obstante, en la especie *Caragana intermedia* se encontró que su expresión es muy variable bajo cuatro condiciones de estrés abiótico diferentes⁶¹. En su investigación, Yang et al. 2015⁶⁵ encontraron que el gen TUB puede ser empleado como gen normalizador para experimentos de qRT-PCR en hojas y raíces de *Lolium arundinaceum* bajo estrés por salinidad y tratamientos con diferentes temperaturas.

4.6 Verificación *in-vitro*.

Una vez diseñados los primers y extraída la muestra a emplear para los experimentos subsiguientes, se procedió a la amplificación, mediante PCR convencional (punto final) los fragmentos de los genes de interés. Los productos de PCR para cada gen se elucidaron en geles de agarosa a una concentración del 2%. Los resultados obtenidos de las amplificaciones se muestran en la siguiente figura.

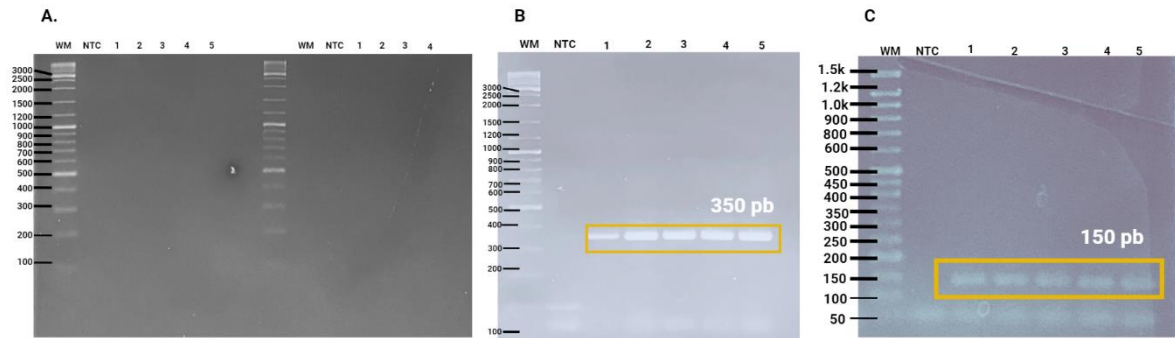


Figura 4. Electroforesis de los productos de PCR de la amplificación de los genes mRNA en geles de agarosa al 2%. **A.** Amplificación de ACT con primers ACTF1R1 y ACTF2R2. Carriles **WM**: marcador de peso molecular GeneRuler™ DNA Ladder 1kb (Cat No. SM0311, ThermoFisher Scientific), **NTC**: Control negativo. **1.** ACT F1R1. **2.** ACT F2R2. Carriles 1-5 Temperaturas para la estandarización de los primers. **1.** 60°C. **2.** 58,8 °C. **3.** 56,8 °C. **4.** 54,3 °C. **5.** 50,0 °C. **B.** Amplificación de GAPDH con primers GAPDHF2R2. Carriles **WM**: marcador de peso molecular GeneRuler™ DNA Ladder 1kb (Cat No. SM0311, ThermoFisher Scientific), **NTC**: Control negativo. Carriles: Temperaturas para la estandarización del primer a **1.** 60°C. **2.** 56,8 °C. **3.** 54,3 °C. **4.** 54.6 °C. **5.** 50,0 °C. **C.** Amplificación de TUB con primers TUBF1R1. Carriles **WM**: marcador de peso molecular GeneRuler™ DNA Ladder 1kb (Cat No. SM0311, ThermoFisher Scientific), **NTC**: Control negativo. Carriles: Temperaturas para la estandarización del primer a **1.** 60.0°C. **2.** 61.2 °C. **3.** 62.0 °C. **4.** 63.1 °C. **5.** 64.1 °C.

Como se muestra en la figura anterior, se determinaron productos para PCR con el peso molecular deseado para los genes GAPDH y TUB. Esto podría confirmar que las secuencias objetivas, correspondientes a los dominios funcionales de estos genes, están presentes en las muestras de ADNc utilizadas. Sin embargo, es importante tener en cuenta que la PCR no es una técnica que permita conocer la identidad exacta del producto obtenido, ya que su sensibilidad no es suficientemente alta⁶⁸. Además, la presencia de las bandas esperadas en el gel no proporciona información sobre la precisión de la secuencia nucleotídica de los productos amplificados. Por esta razón, los productos obtenidos fueron sometidos a secuenciación Sanger con el fin de validar los resultados.

Por otro lado, no se observaron bandas apreciables en los carriles 1–5 del gen ACT. Esta ausencia de amplificación puede atribuirse a diversos factores. Uno de ellos podría ser un diseño de primers poco efectivo. En este caso, no se tuvo en cuenta el principio de la degeneración de los cebadores para abarcar las posibles variaciones en el uso de codones. Sin embargo, si la degeneración no fue suficiente o si las variaciones en la secuencia del gen ACT son demasiado grandes, los primers pueden

no ser capaces de unirse de manera efectiva a la secuencia objetivo, lo que resulta en la ausencia de amplificación durante la PCR. Por tanto, cabe la posibilidad de que la secuencia del gen ACT en esta planta en particular difiera de la secuencia reportada. Esto podría deberse a la distancia taxonómica entre los genes de referencia⁴⁹, lo que podría haber conducido a cambios en el uso de codones.

Por otro lado, la variabilidad en la expresión del gen ACT en los diferentes estudios mencionados previamente (ver Extracción de mRNA) plantea dudas sobre su idoneidad como normalizador en estudios moleculares. Aunque un gen HK se utiliza como un estándar interno cuya expresión se asume constante entre distintos grupos experimentales⁶⁹, la inestabilidad en la expresión del gen ACT bajo ciertas condiciones ambientales resalta la necesidad de validar su uso como HK para futuras investigaciones. Esto no solo implica que la elección de un gen inadecuado pueda afectar significativamente el estudio, sino también que el gen seleccionado como normalizador podría no expresarse en el tejido vegetal que se desea investigar⁷⁰⁻⁷³. Este es un factor crucial, ya que podría ser una razón adicional para los resultados negativos obtenidos en las pruebas realizadas para el gen ACT.

4.7. Secuenciación Sanger

Como se mencionó anteriormente, se logró la amplificación de los productos de PCR con un peso molecular similar al establecido de forma *in-silico* para los genes GAPDH y TUB. Este resultado es significativo, ya que podría indicar la presencia de estos genes tanto en especímenes *in-vitro* como en especímenes silvestres. Es importante destacar que, los cultivos utilizados para la extracción de las muestras son clones obtenidos a partir de raíces de un individuo silvestre de *C. angustifolia*. Sin embargo, para asegurar la identidad de estos fragmentos, se realizó la verificación mediante secuenciación Sanger, seguida de un análisis en la herramienta Blastn para conocer la especificidad de estas secuencias.

Description	Total Score	Query Cover	E value	Per. Ident
Antiaris toxicaria isolate JCLG002 chromosome 1	107	35%	1e-18	86.52%
Antiaris toxicaria isolate JCLG002 chromosome 1	103	35%	6e-17	85.39%
PREDICTED: Cannabis sativa molybdopterin biosynthesis protein CNX1 (LOC115719472), mRNA	91.5	25%	1e-13	90.77%
PREDICTED: Cannabis sativa molybdopterin biosynthesis protein CNX1-like (LOC133032906), mRNA	91.5	25%	1e-13	90.77%
Euphorbia peplus chromosome 7	89.7	31%	4e-13	84.81%
Alnus glutinosa genome assembly, chromosome: 9	86.0	31%	5e-12	84.81%
Gnaphalium uliginosum genome assembly, chromosome: 3	83.3	26%	6e-11	86.76%
Quercus cerris genome assembly, chromosome: 6	83.3	30%	6e-11	83.33%
Quercus variabilis isolate HB chromosome 1a	83.3	30%	6e-11	83.33%
Quercus variabilis isolate HB chromosome 1b	83.3	30%	6e-11	83.33%
Quercus dentata isolate Qdleaves-2020 chromosome 1	83.3	30%	6e-11	83.33%
Quercus robur genome assembly, chromosome: 1	83.3	30%	6e-11	83.33%
PREDICTED: Humulus lupulus molybdopterin biosynthesis protein CNX1 (LOC133788780), mRNA	82.4	25%	6e-11	87.69%

Figura 5. Resultados obtenidos en la herramienta Blast Nucleotide a partir de la secuencia obtenida para el gen GAPDH.

En los resultados de secuenciación a partir de los productos de PCR para el gen GAPDH, se encontró que las secuencias obtenidas correspondían a la proteína de biosíntesis de molibdopterina en *Arabidopsis*, en lugar del gen GAPDH esperado. Por lo cual, es posible considerar la presencia de proteínas de biosíntesis de molibdopterinas en la muestra, sin embargo, se requieren análisis adicionales para validar esta hipótesis.

Tanto las molibdopterinas como la proteína GAPDH están involucradas en el metabolismo energético de la planta⁷⁴, aunque de maneras diferentes. Las molibdopterinas son el producto generado por la enzima Proteína de biosíntesis de molybdoterinas (PMB), que actúa como un aldehído deshidrogenasa, la cual requiere molibdeno en su estructura y desempeñan un papel crucial en varios procesos metabólicos en las plantas, como la fijación de nitrógeno, la asimilación de nitratos y la síntesis de fitohormonas⁷⁴. Por otro lado, GAPDH es una enzima clave en la vía glucolítica, implicada en la producción de energía en las células vegetales generando NADH como producto¹⁹. Esta interconexión metabólica sugiere que las actividades de estas enzimas pueden estar coordinadas para mantener el equilibrio energético en la célula vegetal⁷⁴.

Es importante considerar cómo los productos y subproductos de las molibdopterinas pueden afectar directa o indirectamente la ruta metabólica donde actúa GAPDH. Por ejemplo, la reducción de nitrato a amonio por la enzima nitrato reductasa, una molibdopterina, produce NADP⁺, lo que puede influir en la disponibilidad de NADP⁺/NADPH, cofactores esenciales para la actividad de GAPDH^{19,74}.

Además, la posible presencia de variantes de genes, tanto en PMB como en GAPDH, podría haber contribuido a la amplificación cruzada durante la PCR. Si los primers utilizados no eran específicos para GAPDH, es posible que también amplificaran secuencias homólogas de PBM, ya que esta proteína participa en una ruta similar a GAPDH y presenta dominios conservados muy parecidos, lo que explicaría el resultado inesperado obtenido en la secuenciación Sanger para el gen de interés.

La posible presencia de variantes génicas tanto en PBM como en GAPDH podría haber contribuido a la amplificación cruzada durante la PCR. Esto se debe a que, si los primers diseñados no eran específicos para amplificar GAPDH, es probable que también amplificaran secuencias homólogas de PBM. Ya que, esta participa en una ruta metabólica similar a GAPDH y presenta dominios conservados muy similares. Por lo tanto, la similitud entre ambas proteínas podría explicar el resultado inesperado obtenido en la secuenciación Sanger para el gen de interés

Como se observa en la figura 5, este resultado está respaldado por un valor de cobertura del 25%, un E-value (valor esperado) de 1×10^{-13} y un porcentaje de identidad del 90.77%. El valor de cobertura indica la proporción de la secuencia de consulta alineada con una secuencia en la base de datos; el resultado obtenido fue del 25%, relativamente bajo y no representativo de la cobertura completa de la secuencia sometida. En cuanto al E-value, este parámetro estima el número de alineaciones que se esperarían por azar con una puntuación igual o mayor. Un valor bajo indica una alineación significativa, por lo cual se puede afirmar que este resultado es significativo para esta secuencia. Por último, el porcentaje de identidad representa el porcentaje de nucleótidos idénticos en la región alineada; un valor del 90.77% indica una identidad significativamente alta entre la secuencia sometida y las PBM⁷⁵.

Description	Scientific Name	Query Cover	E value	Per. Ident
PREDICTED: Oryza brachyantha tubulin beta-1 chain-like (LOC102705253), mRNA	Oryza brachyantha	96%	1e-54	96.38%
Triticum aestivum mRNA, clone: tplb0033h20, cultivar Chinese Spring	Triticum aestivum	100%	5e-54	95.10%
Triticum aestivum mRNA, clone: tplb0062c05, cultivar Chinese Spring	Triticum aestivum	100%	5e-54	95.10%
PREDICTED: Syzygium oleosum tubulin beta chain-like (LOC115669818), mRNA	Syzygium oleosum	100%	5e-54	95.10%
PREDICTED: Triticum urartu tubulin beta-3 chain (LOC125516174), mRNA	Triticum urartu	100%	5e-54	95.10%
PREDICTED: Triticum aestivum tubulin beta-3 chain (LOC123132419), mRNA	Triticum aestivum	100%	5e-54	95.10%
PREDICTED: Triticum dicoccoides tubulin beta-3 chain (LOC119315850), mRNA	Triticum dicoccoides	100%	5e-54	95.10%
Bromus tectorum tubulin beta chain 4 isoform 4 (tub4) mRNA, complete cds	Bromus tectorum	96%	6e-53	95.65%
Bromus tectorum tubulin beta chain 4 isoform 3 (tub4) mRNA, complete cds	Bromus tectorum	96%	6e-53	95.65%
Bromus tectorum tubulin beta chain 4 isoform 2 (tub4) mRNA, complete cds	Bromus tectorum	96%	6e-53	95.65%

Figura 6. Resultados obtenidos en la herramienta Blast Nucleotide a partir de la secuencia obtenida para el gen TUB.

En cuanto al resultado obtenido por la herramienta Blastn para la secuencia del gen TUB, como se observa en la figura 6, la secuencia sometida corresponde en un 96.38% de identidad con el gen β -tubulina en la especie *Oryza brachyantha* reportada en el repositorio del NCBI. A su vez, se obtuvo un valor de cobertura del 96% y un E-value = 1×10^{-54} . Los parámetros anteriores indican, en primer lugar, que la secuencia de consulta está muy bien alineada con la secuencia anotada en la base de datos. Además, el E-value indica que hay un gran número de alineaciones que se esperarían por azar con una puntuación igual o mayor ya que se obtuvo un resultado significativamente bajo⁷⁵.

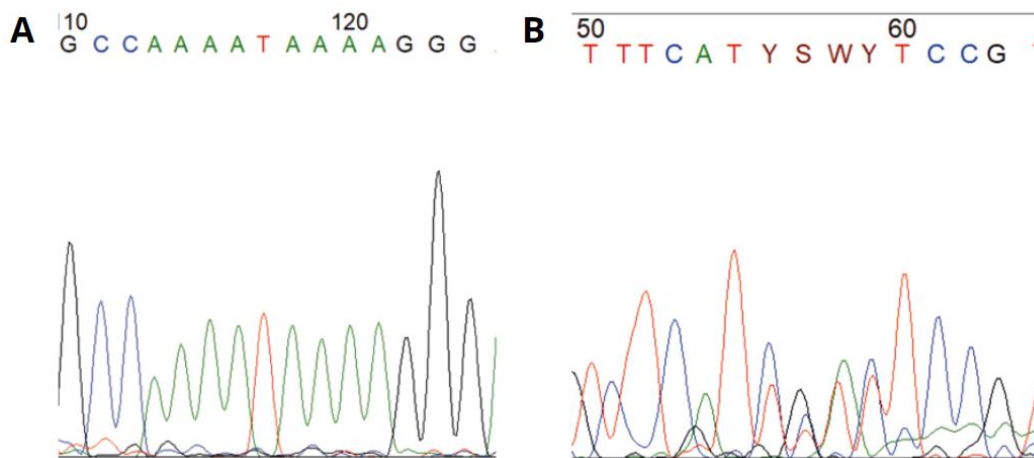


Figura 7. Electroferogramas de las secuencias de los genes A) GAPDH y B) TUB.

La Figura 7 presenta los electroferogramas que representan las secuencias obtenidas para los genes GAPDH y TUB. Se observan disparidades en la calidad de estos gráficos, lo que puede complicar la

interpretación de los resultados. Entre los errores más frecuentes que pueden explicar la falta de resolución en la **Figura 7**, se encuentran: (1) muy baja concentración de muestra empleada en la reacción de secuenciación, lo cual impide que se detecte señal fluorescente; y (2) un proceso ineficiente de purificación de la reacción de secuenciación, lo que ocasiona que no se eliminen correctamente los dNTPSs libres u otros residuos de la reacción y afecten la señal emitida por los amplicones a secuenciar^{76,77}.

5. CONCLUSIONES

Se logró identificar el gen TUB como un potencial gen constitutivo para *C. angustifolia*. Este resultado servirá de base para futuros estudios que permitan validar la hipótesis planteada y contribuirá a la normalización de la expresión génica en esta especie. Por otro lado, si bien no se obtuvieron resultados positivos en la amplificación del gen ACT y GAPDH, es necesario realizar investigaciones adicionales para evaluar su potencial uso como gen HK en *C. angustifolia* y validar la posible presencia de la proteína de biosíntesis de molibdopterinas en esta especie.

Estos hallazgos, junto con la identificación del gen TUB, servirán como base para futuras investigaciones sobre la caracterización de genes constitutivos en *C. angustifolia*, destacando la necesidad de un mayor enfoque en la investigación molecular de esta especie para evaluar con mayor precisión sus genes constitutivos.

6. AGRADECIMIENTOS

En primer lugar, deseo expresar mi más sincero agradecimiento a mi madre y mi madrina por proporcionarme esta valiosa oportunidad y por su constante apoyo y guía a lo largo de este proceso y etapa académica. Su orientación ha sido fundamental para mi progreso.

Quiero extender mi gratitud a Juan Esteban Vivas, mi tutor, por su paciencia, sabiduría y liderazgo en este recorrido. Su transmisión de conocimientos ha sido sobresaliente, despertando en mí una profunda pasión por la biología molecular y la biotecnología. Estoy segura de que será un docente excepcional, impartiendo su conocimiento con entusiasmo.

Agradezco también a Cristhian Castro por su invaluable contribución durante este proyecto. Ha sido un apoyo incondicional, guiándome pacientemente hacia mis metas y brindándome su afecto constante. Su arduo trabajo y las largas jornadas dedicadas a su trabajo y acompañarme han sido cruciales para superar los desafíos y lograr los resultados obtenidos. Gracias por ser mi punto de equilibrio.

Finalmente, quiero agradecer al Profesor Guillermo Montoya por permitirme formar parte del semillero de investigación, y a todas las personas que hacen parte de ello. Su apoyo no solo me ha permitido adquirir un vasto conocimiento, sino que también me ha brindado la oportunidad de iniciar mi trayectoria científica, contribuyendo significativamente a mi desarrollo profesional.

7. REFERENCIAS BIBLIOGRÁFICAS

- (1) Lodha, T. D.; Srinivas, A.; Sasikala, C.; Ramana, C. V. Hopanoid Inventory of Rhodoplanes Spp. *Arch Microbiol* **2015**, *197* (6), 861–867. <https://doi.org/10.1007/S00203-015-1112-5>.
- (2) Huddart, J. E. A.; Crawford, A. J.; Luna-Tapia, A. L.; Restrepo, S.; Di Palma, F. EBP-Colombia and the Bioeconomy: Genomics in the Service of Biodiversity Conservation and Sustainable Development. *Proc Natl Acad Sci U S A* **2022**, *119* (4). <https://doi.org/10.1073/PNAS.2115641119>.
- (3) Dietz, T.; Börner, J.; Förster, J. J.; von Braun, J. Governance of the Bioeconomy: A Global Comparative Study of National Bioeconomy Strategies. *Sustainability* **2018**, *Vol. 10*, Page 3190 **2018**, *10* (9), 3190. <https://doi.org/10.3390/SU10093190>.
- (4) Alzate, F.; Idárraga, A. Flora de Los Bosques Montanos de Medellin.
- (5) Fazili, M. A.; Bashir, I.; Ahmad, M.; Yaqoob, U.; Geelani, S. N. In Vitro Strategies for the Enhancement of Secondary Metabolite Production in Plants: A Review. *Bull Natl Res Cent* **2022**, *46* (1). <https://doi.org/10.1186/S42269-022-00717-Z>.
- (6) Guerriero, G.; Berni, R.; Muñoz-Sanchez, J. A.; Apone, F.; Abdel-Salam, E. M.; Qahtan, A. A.; Alatar, A. A.; Cantini, C.; Cai, G.; Hausman, J. F.; Siddiqui, K. S.; Hernández-Sotomayor, S. M. T.; Faisal, M. Production of Plant Secondary Metabolites: Examples, Tips and Suggestions for Biotechnologists. *Genes (Basel)* **2018**, *9* (6), 309. <https://doi.org/10.3390/GENES9060309>.
- (7) *yarumo negro, yarumo, (Cecropia angustifolia)*. <https://catalogoarbosurbanos.eia.edu.co/species/184> (accessed 2024-05-31).
- (8) Sandeep; Misra, R. C.; Chanotiya, C. S.; Mukhopadhyay, P.; Ghosh, S. Oxidosqualene Cyclase and CYP716 Enzymes Contribute to Triterpene Structural Diversity in the Medicinal Tree Banaba. *New Phytol* **2019**, *222* (1), 408–424. <https://doi.org/10.1111/NPH.15606>.
- (9) Krasteva, G.; Georgiev, V.; Pavlov, A. Recent Applications of Plant Cell Culture Technology in Cosmetics and Foods. *Eng Life Sci* **2021**, *21* (3–4), 68–76. <https://doi.org/10.1002/ELSC.202000078>.
- (10) Ramakrishna, A.; Ravishankar, G. A. Influence of Abiotic Stress Signals on Secondary Metabolites in Plants. *Plant Signal Behav* **2011**, *6* (11), 1720. <https://doi.org/10.4161/PSB.6.11.17613>.
- (11) Joshi, C. J.; Ke, W.; Drangowska-Way, A.; O'Rourke, E. J.; Lewis, N. E. What Are Housekeeping Genes? *PLoS Comput Biol* **2022**, *18* (7), e1010295. <https://doi.org/10.1371/JOURNAL.PCBI.1010295>.

- (12) Rhee, S. Y.; Dickerson, J.; Xu, D. Bioinformatics and Its Applications in Plant Biology. *Annu Rev Plant Biol* **2006**, *57*, 335–360. <https://doi.org/10.1146/ANNUREV.ARPLANT.56.032604.144103>.
- (13) Hounkpe, B. W.; Chenou, F.; de Lima, F.; de Paula, E. V. HRT Atlas v1.0 Database: Redefining Human and Mouse Housekeeping Genes and Candidate Reference Transcripts by Mining Massive RNA-Seq Datasets. *Nucleic Acids Res* **2021**, *49* (D1), D947–D955. <https://doi.org/10.1093/NAR/GKAA609>.
- (14) Eisenberg, E.; Levanon, E. Y. Human Housekeeping Genes, Revisited. *Trends Genet* **2013**, *29* (10), 569–574. <https://doi.org/10.1016/J.TIG.2013.05.010>.
- (15) Yu, Y.; Zhang, G.; Chen, Y.; Bai, Q.; Gao, C.; Zeng, L.; Li, Z.; Cheng, Y.; Chen, J.; Sun, X.; Guo, L.; Xu, J.; Yan, Z. Selection of Reference Genes for QPCR Analyses of Gene Expression in Ramie Leaves and Roots across Eleven Abiotic/Biotic Treatments. *Scientific Reports* **2019**, *9:1* **2019**, *9* (1), 1–13. <https://doi.org/10.1038/s41598-019-56640-3>.
- (16) Liang, W.; Zou, X.; Carballar-Lejarazú, R.; Wu, L.; Sun, W.; Yuan, X.; Wu, S.; Li, P.; Ding, H.; Ni, L.; Huang, W.; Zou, S. Selection and Evaluation of Reference Genes for QRT-PCR Analysis in *Euscaphis Konishii* Hayata Based on Transcriptome Data. *Plant Methods* **2018**, *14* (1), 42. <https://doi.org/10.1186/S13007-018-0311-X>.
- (17) Yu, Y.; Zhang, G.; Chen, Y.; Bai, Q.; Gao, C.; Zeng, L.; Li, Z.; Cheng, Y.; Chen, J.; Sun, X.; Guo, L.; Xu, J.; Yan, Z. Selection of Reference Genes for QPCR Analyses of Gene Expression in Ramie Leaves and Roots across Eleven Abiotic/Biotic Treatments. *Scientific Reports* **2019**, *9:1* **2019**, *9* (1), 1–13. <https://doi.org/10.1038/s41598-019-56640-3>.
- (18) Sirover, M. A. Structural Analysis of Glyceraldehyde-3-Phosphate Dehydrogenase Functional Diversity. *Int J Biochem Cell Biol* **2014**, *57*, 20–26. <https://doi.org/10.1016/J.BIOCEL.2014.09.026>.
- (19) Nicholls, C.; Li, H.; Liu, J. P. GAPDH: A Common Enzyme with Uncommon Functions. *Clin Exp Pharmacol Physiol* **2012**, *39* (8), 674–679. <https://doi.org/10.1111/J.1440-1681.2011.05599.X>.
- (20) Tisdale, E. J.; Kelly, C.; Artalejo, C. E. Glyceraldehyde-3-Phosphate Dehydrogenase Interacts with Rab2 and Plays an Essential Role in Endoplasmic Reticulum to Golgi Transport Exclusive of Its Glycolytic Activity. *J Biol Chem* **2004**, *279* (52), 54046–54052. <https://doi.org/10.1074/JBC.M409472200>.
- (21) Ikemoto, A.; Bole, D. G.; Ueda, T. Glycolysis and Glutamate Accumulation into Synaptic Vesicles. Role of Glyceraldehyde Phosphate Dehydrogenase and 3-Phosphoglycerate Kinase. *J Biol Chem* **2003**, *278* (8), 5929–5940. <https://doi.org/10.1074/JBC.M211617200>.
- (22) Hara, M. R.; Snyder, S. H. Nitric Oxide-GAPDH-Siah: A Novel Cell Death Cascade. *Cell Mol Neurobiol* **2006**, *26* (4–6), 527–538. <https://doi.org/10.1007/S10571-006-9011-6>.

- (23) Šlajcherová, K.; Fišerová, J.; Fischer, L.; Schwarzerová, K. Multiple Actin Isoforms in Plants: Diverse Genes for Diverse Roles? *Front Plant Sci* **2012**, *3* (OCT). <https://doi.org/10.3389/FPLS.2012.00226>.
- (24) Kato, T.; Morita, M. T.; Tasaka, M. Defects in Dynamics and Functions of Actin Filament in Arabidopsis Caused by the Dominant-Negative Actin Fz1-Induced Fragmentation of Actin Filament. *Plant Cell Physiol* **2010**, *51* (2), 333–338. <https://doi.org/10.1093/PCP/PCP189>.
- (25) Takatsuka, H.; Higaki, T.; Umeda, M. Actin Reorganization Triggers Rapid Cell Elongation in Roots. *Plant Physiol* **2018**, *178* (3), 1130. <https://doi.org/10.1104/PP.18.00557>.
- (26) Diao, M.; Huang, S. An Update on the Role of the Actin Cytoskeleton in Plasmodesmata: A Focus on Formins. *Front Plant Sci* **2021**, *12*, 647123. <https://doi.org/10.3389/FPLS.2021.647123>.
- (27) Szymanski, D.; Staiger, C. J. The Actin Cytoskeleton: Functional Arrays for Cytoplasmic Organization and Cell Shape Control. *Plant Physiol* **2018**, *176* (1), 106–118. <https://doi.org/10.1104/PP.17.01519>.
- (28) Paez-Garcia, A.; Sparks, J. A.; de Bang, L.; Blancaflor, E. B. Plant Actin Cytoskeleton: New Functions from Old Scaffold. *Plant Cell Monographs* **2018**, *23*, 103–137. https://doi.org/10.1007/978-3-319-69944-8_6.
- (29) Paez-Garcia, A.; Sparks, J. A.; de Bang, L.; Blancaflor, E. B. Plant Actin Cytoskeleton: New Functions from Old Scaffold. *Plant Cell Monographs* **2018**, *23*, 103–137. https://doi.org/10.1007/978-3-319-69944-8_6.
- (30) Radchuk, V. V. The Transcriptome Of The Tubulin Gene Family In Plants. **2008**, 219–241. https://doi.org/10.1007/978-1-4020-8843-8_11.
- (31) Gavazzi, F.; Pigna, G.; Braglia, L.; Gianì, S.; Breviario, D.; Morello, L. Evolutionary Characterization and Transcript Profiling of β -Tubulin Genes in Flax (*Linum Usitatissimum* L.) during Plant Development. *BMC Plant Biol* **2017**, *17* (1), 1–16. <https://doi.org/10.1186/S12870-017-1186-0/FIGURES/8>.
- (32) Gasic, I. Regulation of Tubulin Gene Expression: From Isoform Identity to Functional Specialization. *Front Cell Dev Biol* **2022**, *10*, 898076. <https://doi.org/10.3389/FCELL.2022.898076/BIBTEX>.
- (33) Gasic, I.; Mitchison, T. J. Autoregulation and Repair in Microtubule Homeostasis. *Curr Opin Cell Biol* **2019**, *56*, 80–87. <https://doi.org/10.1016/J.CEB.2018.10.003>.
- (34) Yu, Y.; Zhang, G.; Chen, Y.; Bai, Q.; Gao, C.; Zeng, L.; Li, Z.; Cheng, Y.; Chen, J.; Sun, X.; Guo, L.; Xu, J.; Yan, Z. Selection of Reference Genes for QPCR Analyses of Gene Expression in Ramie Leaves and Roots across Eleven Abiotic/Biotic Treatments. *Scientific Reports* **2019**, *9*:1 **2019**, *9* (1), 1–13. <https://doi.org/10.1038/s41598-019-56640-3>.
- (35) Ferguson, R. E.; Carroll, H. P.; Harris, A.; Maher, E. R.; Selby, P. J.; Banks, R. E. Housekeeping Proteins: A Preliminary Study Illustrating Some Limitations as Useful References in Protein

Expression Studies. *Proteomics* **2005**, 5 (2), 566–571.

<https://doi.org/10.1002/PMIC.200400941>.

- (36) Li, R.; Shen, Y. An Old Method Facing a New Challenge: Re-Visiting Housekeeping Proteins as Internal Reference Control for Neuroscience Research. *Life Sci* **2013**, 92 (13), 747. <https://doi.org/10.1016/J.LFS.2013.02.014>.
- (37) Hossain, M. S.; Ahmed, R.; Haque, M. S.; Alam, M. M.; Islam, M. S. Identification and Validation of Reference Genes for Real-Time Quantitative RT-PCR Analysis in Jute. *BMC Mol Biol* **2019**, 20 (1), 1–13. <https://doi.org/10.1186/S12867-019-0130-2/FIGURES/6>.
- (38) Crossley, B. M.; Bai, J.; Glaser, A.; Maes, R.; Porter, E.; Killian, M. L.; Clement, T.; Toohey-Kurth, K. Guidelines for Sanger Sequencing and Molecular Assay Monitoring. *J Vet Diagn Invest* **2020**, 32 (6), 767–775. <https://doi.org/10.1177/1040638720905833>.
- (39) Crossley, B. M.; Bai, J.; Glaser, A.; Maes, R.; Porter, E.; Killian, M. L.; Clement, T.; Toohey-Kurth, K. Guidelines for Sanger Sequencing and Molecular Assay Monitoring. *Journal of Veterinary Diagnostic Investigation* **2020**, 32 (6), 767–775. <https://doi.org/10.1177/1040638720905833>.
- (40) *ScienceDirect.com | Science, health and medical journals, full text articles and books.* <https://www.sciencedirect.com/> (accessed 2024-05-31).
- (41) *Nature.* <https://www.nature.com/> (accessed 2024-05-31).
- (42) *National Center for Biotechnology Information.* <https://www.ncbi.nlm.nih.gov/> (accessed 2024-05-31).
- (43) Batut, B.; van den Beek, M.; Doyle, M. A.; Soranzo, N. RNA-Seq Data Analysis in Galaxy. *Methods Mol Biol* **2021**, 2284, 367–392. https://doi.org/10.1007/978-1-0716-1307-8_20.
- (44) Chen, Y.; Ye, W.; Zhang, Y.; Xu, Y. High Speed BLASTN: An Accelerated MegaBLAST Search Tool. *Nucleic Acids Res* **2015**, 43 (16), 7762–7768. <https://doi.org/10.1093/NAR/GKV784>.
- (45) Jacob, A.; Lancaster, J.; Buhler, J.; Harris, B.; Chamberlain, R. D. Mercury BLASTP: Accelerating Protein Sequence Alignment. *ACM Trans Reconfigurable Technol Syst* **2008**, 1 (2), 1–44. <https://doi.org/10.1145/1371579.1371581>.
- (46) Vivas, J. E. Análisis de La Expresión Génica y Metabólica Asociada Con La Acumulación de Metabolitos Secundarios Derivados de Triterpenos Pentacíclicos Ácidos En Raíces de *Cecropia Angustifolia.*, Universidad ICESI, Santiago de Cali, 2022.
- (47) Cadena-Zamudio, J. D.; Nicasio-Torres, P.; Monribot-Villanueva, J. L.; Guerrero-Analco, J. A.; Ibarra-Laclette, E. Integrated Analysis of the Transcriptome and Metabolome of *Cecropia Obtusifolia*: A Plant with High Chlorogenic Acid Content Traditionally Used to Treat Diabetes Mellitus. *Int J Mol Sci* **2020**, 21 (20), 1–27. <https://doi.org/10.3390/IJMS21207572>.
- (48) Okonechnikov, K.; Golosova, O.; Fursov, M.; Varlamov, A.; Vaskin, Y.; Efremov, I.; German Grehov, O. G.; Kandrov, D.; Rasputin, K.; Syabro, M.; Tleukenov, T. Unipro UGENE: A Unified

- Bioinformatics Toolkit. *Bioinformatics* **2012**, *28* (8), 1166–1167.
<https://doi.org/10.1093/BIOINFORMATICS/BTS091>.
- (49) Wu, Z. Y.; Monro, A. K.; Milne, R. I.; Wang, H.; Yi, T. S.; Liu, J.; Li, D. Z. Molecular Phylogeny of the Nettle Family (Urticaceae) Inferred from Multiple Loci of Three Genomes and Extensive Generic Sampling. *Mol Phylogenet Evol* **2013**, *69* (3), 814–827.
<https://doi.org/10.1016/J.YMPEV.2013.06.022>.
- (50) Wilmot-Dear, C. M.; Friis, I. The Old World Species of *Pouzolzia* (Urticaceae, Tribus Boehmerieae). A Taxonomic Revision. *Nord J Bot* **2004**, *24* (1), 5–111.
<https://doi.org/10.1111/J.1756-1051.2004.TB00825.X>.
- (51) Wood, D. E.; Lin, H.; Levy-Moonshine, A.; Swaminathan, R.; Chang, Y. C.; Anton, B. P.; Osmani, L.; Steffen, M.; Kasif, S.; Salzberg, S. L. Thousands of Missed Genes Found in Bacterial Genomes and Their Analysis with COMBRES. *Biol Direct* **2012**, *7*, 37.
<https://doi.org/10.1186/1745-6150-7-37>.
- (52) *OneZoom: Pentapetalae*.
https://www.onezoom.org/life/@Pentapetalae=5316182?img=best_any&anim=flight#x764,y-68,w0.8156 (accessed 2024-05-31).
- (53) Geballe, A. P.; Morris, D. R. Initiation Codons within 5'-Leaders of MRNAs as Regulators of Translation. *Trends Biochem Sci* **1994**, *19* (4), 159–164. [https://doi.org/10.1016/0968-0004\(94\)90277-1](https://doi.org/10.1016/0968-0004(94)90277-1).
- (54) Babbitt, G. A.; Coppola, E. E.; Mortensen, J. S.; Ekeren, P. X.; Viola, C.; Goldblatt, D.; Hudson, A. O. Triplet-Based Codon Organization Optimizes the Impact of Synonymous Mutation on Nucleic Acid Molecular Dynamics. *J Mol Evol* **2018**, *86* (2), 91–102.
<https://doi.org/10.1007/S00239-018-9828-X>.
- (55) Kute, P. M.; Soukariéh, O.; Tjeldnes, H.; Trégouët, D. A.; Valen, E. Small Open Reading Frames, How to Find Them and Determine Their Function. *Front Genet* **2022**, *12*, 796060.
<https://doi.org/10.3389/FGENE.2021.796060/BIBTEX>.
- (56) Oelschlaeger, P. Molecular Mechanisms and the Significance of Synonymous Mutations. *Biomolecules* **2024**, *14* (1). <https://doi.org/10.3390/BIOM14010132>.
- (57) Zaffagnini, M.; Fermani, S.; Costa, A.; Lemaire, S. D.; Trost, P. Plant Cytoplasmic GAPDH: Redox Post-Translational Modifications and Moonlighting Properties. *Front Plant Sci* **2013**, *4* (NOV). <https://doi.org/10.3389/FPLS.2013.00450>.
- (58) Ghosh, S. Triterpene Structural Diversification by Plant Cytochrome P450 Enzymes. *Front Plant Sci* **2017**, *8*, 295540. <https://doi.org/10.3389/FPLS.2017.01886/BIBTEX>.
- (59) *Degenerate primers: Design and use*. <https://www.qiagen.com/us/knowledge-and-support/knowledge-hub/bench-guide/pcr/introduction/guidelines-for-degenerate-primer-design-and-use> (accessed 2024-05-31).
- (60) Kozera, B.; Rapacz, M. Reference Genes in Real-Time PCR. *J Appl Genet* **2013**, *54* (4), 391.
<https://doi.org/10.1007/S13353-013-0173-X>.

- (61) Zhu, J.; Zhang, L.; Li, W.; Han, S.; Yang, W.; Qi, L. Reference Gene Selection for Quantitative Real-Time PCR Normalization in Caragana Intermedia under Different Abiotic Stress Conditions. *PLoS One* **2013**, *8* (1). <https://doi.org/10.1371/JOURNAL.PONE.0053196>.
- (62) Gimeno, J.; Eattock, N.; Van Deynze, A.; Blumwald, E. Selection and Validation of Reference Genes for Gene Expression Analysis in Switchgrass (*Panicum Virgatum*) Using Quantitative Real-Time RT-PCR. *PLoS One* **2014**, *9* (3). <https://doi.org/10.1371/JOURNAL.PONE.0091474>.
- (63) Huang, L.; Yan, H.; Jiang, X.; Yin, G.; Zhang, X.; Qi, X.; Zhang, Y.; Yan, Y.; Ma, X.; Peng, Y. Identification of Candidate Reference Genes in Perennial Ryegrass for Quantitative RT-PCR under Various Abiotic Stress Conditions. *PLoS One* **2014**, *9* (4). <https://doi.org/10.1371/JOURNAL.PONE.0093724>.
- (64) Li, W.; Qian, Y. Q.; Han, L.; Liu, J. X.; Sun, Z. Y. Identification of Suitable Reference Genes in Buffalo Grass for Accurate Transcript Normalization under Various Abiotic Stress Conditions. *Gene* **2014**, *547* (1), 55–62. <https://doi.org/10.1016/J.GENE.2014.06.015>.
- (65) Yang, Z.; Chen, Y.; Hu, B.; Tan, Z.; Huang, B. Identification and Validation of Reference Genes for Quantification of Target Gene Expression with Quantitative Real-Time PCR for Tall Fescue under Four Abiotic Stresses. *PLoS One* **2015**, *10* (3). <https://doi.org/10.1371/JOURNAL.PONE.0119569>.
- (66) Jain, M.; Nijhawan, A.; Tyagi, A. K.; Khurana, J. P. Validation of Housekeeping Genes as Internal Control for Studying Gene Expression in Rice by Quantitative Real-Time PCR. *Biochem Biophys Res Commun* **2006**, *345* (2), 646–651. <https://doi.org/10.1016/J.BBRC.2006.04.140>.
- (67) Kundu, A.; Patel, A.; Pal, A. Defining Reference Genes for QPCR Normalization to Study Biotic and Abiotic Stress Responses in Vigna Mungo. *Plant Cell Rep* **2013**, *32* (10), 1647–1658. <https://doi.org/10.1007/S00299-013-1478-2>.
- (68) *What are the differences between PCR, RT-PCR, qPCR, and RT-qPCR? - Enzo.* <https://www.enzo.com/note/what-are-the-differences-between-pcr-rt-pcr-qpcr-and-rt-qpcr/> (accessed 2024-05-31).
- (69) Huggett, J.; Dheda, K.; Bustin, S.; Zumla, A. Real-Time RT-PCR Normalisation; Strategies and Considerations. *Genes Immun* **2005**, *6* (4), 279–284. <https://doi.org/10.1038/SJ.GENE.6364190>.
- (70) Narsai, R.; Ivanova, A.; Ng, S.; Whelan, J. Defining Reference Genes in Oryza Sativa Using Organ, Development, Biotic and Abiotic Transcriptome Datasets. *BMC Plant Biol* **2010**, *10*. <https://doi.org/10.1186/1471-2229-10-56>.
- (71) Xu, H.; Bao, J. D.; Dai, J. S.; Li, Y.; Zhu, Y. Genome-Wide Identification of New Reference Genes for QRT-PCR Normalization under High Temperature Stress in Rice Endosperm. *PLoS One* **2015**, *10* (11). <https://doi.org/10.1371/JOURNAL.PONE.0142015>.
- (72) Zhang, C.; Fu, J.; Wang, Y.; Bao, Z.; Zhao, H. Identification of Suitable Reference Genes for Gene Expression Normalization in the Quantitative Real-Time PCR Analysis of Sweet

- Osmanthus (*Osmanthus Fragrans* Lour.). *PLoS One* **2015**, *10* (8).
<https://doi.org/10.1371/JOURNAL.PONE.0136355>.
- (73) Zhou, Z.; Cong, P.; Tian, Y.; Zhu, Y. Using RNA-Seq Data to Select Reference Genes for Normalizing Gene Expression in Apple Roots. *PLoS One* **2017**, *12* (9).
<https://doi.org/10.1371/JOURNAL.PONE.0185288>.
- (74) Tejada-Jiménez, M.; Chamizo-Ampudia, A.; Galván, A.; Fernández, E.; Llamas, Á. Molybdenum Metabolism in Plants. *Metallomics* **2013**, *5* (9), 1191–1203.
<https://doi.org/10.1039/C3MT00078H>.
- (75) *Examining Your BLAST Results*. <https://fundis.org/component/sppagebuilder/41-examining-your-blast-results> (accessed 2024-05-31).
- (76) Crossley, B. M.; Bai, J.; Glaser, A.; Maes, R.; Porter, E.; Killian, M. L.; Clement, T.; Toohey-Kurth, K. Guidelines for Sanger Sequencing and Molecular Assay Monitoring. *J Vet Diagn Invest* **2020**, *32* (6), 767–775. <https://doi.org/10.1177/1040638720905833>.
- (77) ThermoFisher Scientific. *Sanger Sequencing by CE 1: Foundations*. <https://www.thermofisher.com/blog/behindthebench/sanger-sequencing-by-ce-1-foundations/> (accessed 2024-04-08).