

INSTITUTO TECNOLÓGICO DE COSTA RICA
VICERRECTORÍA DE INVESTIGACIÓN Y EXTENSIÓN
DIRECCIÓN DE PROYECTOS

INFORME FINAL DE PROYECTO

Código del Proyecto:

5401-1510-9301

Nombre del proyecto:

Identificación de factores de transcripción putativos en *Stevia rebaudiana* y *Tagetes patula* como herramienta para posterior uso en la descripción de rutas metabólicas de interés.

Departamento Académico responsable:

Escuela de Biología.

Investigadora responsable:

MSc. Silvana Alvarenga Vetunolo.

Otros investigadores:

MSc. Giovanni Garro Monge

Dr. Francisco José Torres Rojas

MGP. Adriana Álvarez Figueroa

Licda. Karol Jiménez Quesada

MBA. Karla Valerín Berrocal

Período 2013 - 2014

TABLA DE CONTENIDO

RESUMEN.....	3
INTRODUCCIÓN.....	4
MARCO TEÓRICO	6
METODOLOGÍA.....	9
RESULTADOS	12
Secuenciación y análisis de transcriptoma	12
Identificación de secuencias génicas reportadas.....	24
Alineamiento de genes contra el transcriptoma	28
DISCUSIÓN.....	30
RECOMENDACIONES.....	37
AGRADECIMIENTOS.....	38
REFERENCIAS	39
ANEXOS.....	43

RESUMEN

En años recientes, ha aumentado el interés por aprovechar y potencializar el uso de plantas medicinales y/o sus metabolitos secundarios, no obstante, esto ha desencadenado a su vez, una sobreexplotación de ciertas especies, amenazando su supervivencia. Actualmente la investigación básica y aplicada se ha combinado para lograr propagar especies con potencial medicinal y dilucidar las rutas biosintéticas pertinentes, con miras a su potencialización. Esta investigación tuvo por objetivo identificar secuencias génicas relacionadas con factores de transcripción putativos en *Stevia rebaudiana* y *Tagetes patula*. En el caso de *S. rebaudiana*, 15 genes de ruta metabólica de síntesis de glicósidos estaban presentes en la muestra de ARN secuenciada; para *T. patula* se obtuvo que los genes reportados para esta especie en el NCBI, contando un factor de transcripción, también se encontraron presentes en el material biológico trabajado. En el caso de *Tagetes patula*, se obtuvo un resultado favorable para la identificación de un factor de transcripción cuya función no se define satisfactoriamente en las bases de datos. En cuanto a los alineamientos contra factores de transcripción descritos en especies de la familia *Asteraceae*, los resultados fueron negativos para la identificación de alguno de ellos dentro del transcriptoma de *T. patula* y *S. rebaudiana*. Un futuro análisis de expresión génica mediante secuenciación de ADNc, sería más preciso mediante el ensamblaje de un transcriptoma *de novo* proveniente de una secuenciación completa.

PALABRAS CLAVE: factor de transcripción, glicósidos, metabolito secundario, ruta metabólica, tiofenos, transcriptoma.

ABSTRACT

In recent years, there has been increasing interest in exploiting and maximize use of medicinal plants and / or secondary metabolites, however, this has triggered turn overexploitation of certain species, threatening their survival. Currently the basic and applied research have combined to achieve propagate species with medicinal potential and to elucidate the relevant biosynthetic pathways, with a view to its potentiation. This research aimed to identify gene sequences related to putative transcription factors and *Tagetes patula* *Stevia rebaudiana*. In the case of *S. rebaudiana*, 15 genes metabolic synthesis pathway glycosides were present in the RNA sample sequenced; *T. patula* was obtained than those reported for this species in the NCBI, genes having a transcription factor also were present in the biological material worked. For *Tagetes patula*, favorable for the identification of a transcription factor which is not defined function satisfactorily in databases result was obtained. As alignments against transcription factors described species of the *Asteraceae* family, the results were negative for identifying any of them within the transcriptome of *T. patula* and *S. rebaudiana*. Future gene expression analysis by cDNA sequencing would be more accurate by assembling a transcriptome *de novo* from a complete sequencing.

Keywords: glycosides, methabolic pathway, secondary metabolite, tiophene, transcription factor, trancriptome.

INTRODUCCIÓN

Las plantas medicinales y remedios naturales forman parte de la cultura del mundo desde tiempos inmemorables y, aun en la actualidad, siguen empleándose en la medicina casera, a pesar de la disponibilidad en el mercado de todo tipo de fármacos (Tachjian *et al.*, 2010). En años recientes, ha aumentado el interés por aprovechar y potencializar el uso de plantas medicinales y/o sus metabolitos secundarios, no obstante, esto ha desencadenado a su vez, una sobre explotación de ciertas especies, amenazando su supervivencia (Crispin y Wurtele, 2013).

Una alternativa a la problemática anterior, ha sido el planteamiento de nuevas técnicas de propagación, pero también es necesario idear propuestas para el incremento en los niveles de producción de los compuestos bioactivos; por tanto, la investigación básica y aplicada se ha combinado para lograr la propagación de especies con potencial medicinal y dilucidar las rutas biosintéticas pertinentes, con miras a su potencialización. Algunos ejemplos son las especies *Aloe vera* (Liao *et al.*, 2004), *Atropa belladonna* (Yang *et al.*, 2011), *Catharanthus roseus* (Murata *et al.*, 2008), *Digitalis purpurea* (Einbond *et al.*, 2010), *Azadirachta indica* (Prabhat, 2014) *Echinacea purpurea* (Perry *et al.*, 1997), y *Hypericum* sp. (Karppinen *et al.*, 2006), *Jatropha curcas* (Mbakwe y Harry-Asobara, 2014) *Stevia rebaudiana* (Chaturvedi *et al.*, 2014), *Uncaria tomentosa* (Goncalves *et al.*, 2005)

Los numerosos estudios sobre las especies medicinales aún son pocos en comparación con todo lo que falta dilucidar acerca de los procesos de síntesis de metabolitos secundarios, tanto en términos de las rutas metabólicas y enzimas involucradas como de los factores internos y externos que afectan su acumulación, así como la metodología adecuada para la elaboración de extractos para corroboración de la citotoxicidad y confirmación validada del efecto curativo sobre los humanos. En respuesta a la inquietud anterior, la evolución de las tecnologías asociadas a la transcriptómica y metabolómica, han ganado credibilidad como opciones biológicamente más certeras, para la caracterización del material vegetal, incluyendo estudios muy específicos en términos de expresión génica. Con estas estrategias, ha sido posible ir, paulatinamente, dando respuesta a los temas más incomprendidos asociados con la acumulación y bioactividad de los metabolitos secundarios de las plantas medicinales (Crispin y Wurtele, 2013).

La naturaleza modular de los factores de transcripción ha llevado a la idea de que se puede ejercer un control en la expresión de estos a través de la ingeniería de proteínas, para regular a su vez la expresión del gen o genes deseados. La capacidad de generar factores de transcripción diseñados, tiene potencial para ayudar en la comprensión de la regulación génica y es de interés biotecnológico (Sitakanta *et al.*, 2006). En consecuencia, recientemente, se ha adicionado la secuenciación y análisis de transcriptomas, ya sea por mapeo contra un genoma o bien, mediante el ensamblaje *de novo*, como una estrategia innovadora para generar información acerca de la expresión, tanto de genes de ruta como de regulación.

Por todo lo expuesto anteriormente, el presente trabajo planteó el objetivo de identificar *in silico* secuencias genéticas relacionadas con factores de transcripción participantes en la ruta metabólica de

síntesis de glicósidos en *Stevia rebaudiana* y tiofenos en *Tagetes patula*, para su correspondiente mapeo contra el transcriptoma de cada especie.

Si bien existe un extenso número de trabajos enfocados en el estudio de las propiedades y beneficios de los edulcorantes naturales provenientes de plantas como *S. rebaudiana*, así como en las cualidades nematocidas de *T. patula*, aún es necesario profundizar en las características genéticas de ambas plantas, a través de técnicas moleculares, que faciliten la comprensión y análisis de los genes que intervienen en los procesos de síntesis de dichos compuestos de interés, de manera que en un futuro pueda incrementarse su productividad con miras a la comercialización de productos.

En el caso particular de compuestos bioactivos en plantas medicinales, muchos ya han sido identificados o descritos con amplitud; en este punto, la Biología Molecular y la Bioinformática, surgen como herramientas para la identificación de los genes implicados en el metabolismo de compuestos bioactivos, ya sea de aquellos que participan en la conversión de un compuesto a otro (genes de biosíntesis), como de aquellos que participan en la regulación de la ruta metabólica (genes de regulación).

La gran ventaja de conocer las regiones y/o las condiciones en que se da la acumulación de compuestos bioactivos en una planta es que ofrece una base sólida para proceder con la transcriptómica comparada, en la cual se verifiquen altos y bajos niveles de síntesis según las condiciones dadas, para relacionar y dar seguimiento a los genes de biosíntesis y de regulación involucrados (Crispin y Wurtele, 2013).

MARCO TEÓRICO

La *Stevia rebaudiana* perteneciente a la familia Asteraceae, es una planta herbácea, perenne, que crece en climas tropicales y sub tropicales, y es originaria del norte de Paraguay y Sur América (Bourlakis y Weightman, 2006 y Shreeder *et al.*, 2008). En sus hojas se sintetizan edulcorantes sin calorías y termo estables, ampliamente utilizados en la industria alimentaria y farmacéutica, que al ser más dulces que la sacarosa común, pueden ser aprovechados aún en pequeños volúmenes.

La síntesis de estos compuestos se debe a la formación de glucósidos diterpenos con propiedades edulcorantes en las hojas, como los esteviósidos y el rebaudósido A; se conoce también la existencia de al menos ocho compuestos análogos más, como rebaudósido B, C, D y D, dulcósido A y esteviolbiósidos. Todos estos compuestos poseen una estructura de esteviol en común, y se diferencian entre sí, por el número y unión de los residuos de carbono, lo que además afecta la intensidad del dulzor; sin embargo, sólo los dos primeros compuestos son generados por la planta en niveles aprovechables industrialmente (Bourlakis y Weightman, 2006).

Fenotípicamente la *Stevia sp.* se describe como una hierba perenne herbácea, de tallo pardo y poco ramificada, que alcanza entre 30 y 80 cm de altura, elevándose desde una base corta y gruesa; posee hojas sésiles, opuestas, ovoidales, de nervadura reticulada, de 5 a 10 cm, aserradas sobre todo en su extremo final, que adquieren una textura resinosa al secarse. Presenta pequeñas flores hermafroditas blanquecinas de corona tubular que conforman inflorescencias tipo capítulo, y una raíz fibrosa que abarca una cuarta parte del tamaño total de la planta (Djendoel, 2002 y U.V.A, 2003).

La *Stevia sp.* natural sin refinar contiene más de 100 elementos y aceites volátiles identificados, siendo el esteviósido, el principal glucósido de este vegetal usado comercialmente, el cual no puede ser degradado y utilizado como fuente de energía; además la planta contiene carbohidratos, proteínas, vitaminas y minerales, principalmente en sus hojas (Martínez, 2002). Las hojas secas de *S. rebaudiana* presentan una concentración de alrededor de 10% m/m de esteviol, no obstante esta composición es muy variable, y los protocolos de extracción estudiados aún no son precisos en cuanto a la purificación del compuesto activo. Al menos se sabe que de acuerdo con el tipo de glucósidos existentes éstos son 100 a 450 veces más dulces que la sacarosa, por lo que se confirma su enorme potencial (Bourlakis y Weightman, 2006).

Los compuestos edulcorantes de *S. rebaudiana* pueden ser aprovechados en la industria alimentaria, al modificar los procesos de endulzado tradicionales, también puede dárseles uso en la industria farmacéutica al añadirse a medicamentos de ingestión oral (Bourlakis y Weightman, 2006), o bien, estos glucósidos además son reconocidos por su actividad antihipertensiva, antidiabética, e insulino trópica (Sharipova *et al.*, 2011).

La importancia comercial de *S. rebaudiana* se resume en que los esteviósidos que la caracterizan, aparte de resultar hasta 300 veces más dulce que la sacarosa, no pueden ser degradados y utilizados como fuente de energía por el cuerpo, ya que no se descompone químicamente durante el proceso digestivo, por lo que son aptos para consumo en personas diabéticas. En unión con lo anterior, contribuye con las personas con problemas de obesidad endulzando sin calorías, restringiendo la acumulación de grasas, y disminuyendo la presión arterial. Incluso se ha encontrado su aplicación en el tratamiento de caries (Das *et al.*, 2011 y Flores y Cortés, 2011).

Los compuestos glicosídicos pueden emplearse en la elaboración de gran cantidad de alimentos, como galletas, repostería, bebidas, jugos de fruta, bebidas procesadas, helados, goma de mascar, o bien añadirse en forma pulverizada a bebidas como café y té; actualmente se comercializan en gran variedad de productos sin resultar dañinos para el ser humano (Ahmed *et al.*, 2007 y Rafiq *et al.*, 2007). Actualmente en el mercado circulan productos como Herbal Coffee, Migra Spray, Natuvia, Stevia clear, Stevia extracto, Stevia Spoonable, Stevita, Sweet Leaf Products, a base de esta planta (Flores y Cortés, 2011).

El consumo de la planta o de los productos a base de la misma, al día de hoy no han generado reportes relacionados con efectos secundarios ni mutagénicos perjudiciales; por el contrario, los estudios realizados por diferentes investigadores en distintos años, determinaron que el esteviósido es bien tolerado y efectivo, pudiendo ser tomado en cuenta como alternativa para pacientes diabéticos, ya que actúa estimulando en forma directa las células beta del páncreas, generando así una secreción considerable de insulina (Alvarenga, 2005 y Flores y Cortés, 2011).

Por su parte, *Tagetes patula* (marigold) es una especie vegetal perteneciente a la familia Asteraceae y corresponde a una planta nativa de la región tropical de América, que se extiende desde México hasta Bolivia; aunque también actualmente se cultiva en Europa y Estados Unidos debido al atractivo visual que representan sus flores, su delicada fragancia y sus propiedades medicinales. Esta especie crece en suelos arenosos o arcillosos con buen drenaje, y aunque sobrevive a temperaturas bajas, no es capaz de resistir heladas o la ausencia de luz. Esta planta suele ser pequeña y de vivos colores; las flores pueden ser sencillas, con pétalos periféricos que rodean a un gran número de estambres y pistilos situados en el centro, o bien, completamente recubiertas de pétalos de silueta semiesférica. El mayor atractivo de esta especie radica en su variedad de colores que incluyen variedades jaspeadas (Vidal *et al.*, 2008).

Sus flores le permiten catalogarse como una especie ornamental importante, y a partir de las mismas se extraen colorantes para textiles. Durante su etapa de floración pueden obtenerse aceites esenciales para la elaboración de perfumes y fragancias, y en la última década, se ha prestado especial atención a los efectos antimicóticos que caracterizan a estos aceites, siendo aprovechados en el tratamiento de la candidiasis y para atacar las infecciones de hongos en las plantas. Su consumo generalmente consiste en la preparación de infusiones de las hojas que alivian la tos, o su decocción para aprovecharse como antiinflamatorio, desinfectante y purgante. En forma molida, los tallos y hojas se cuecen y se consumen dadas las facultades cicatrizantes que se le atribuyen (Vidal *et al.*, 2008).

El creciente interés comercial en *T. patula* en la actualidad se debe a que la raíz de esta especie, y en general la de todas en el género *Tagetes sp.* contienen unos compuestos denominados tiofenos, que presentan un efecto inhibitor sobre los nemátodos y aléurodos, y de la invasión de ciertas malezas, como las de los géneros *Elytrigia*, *Cynodon*, *Calystegia* y *Convolvulus*, además actúan como repelentes de áfidos y hormigas. Es por lo anterior que la especie suele sembrarse en asociación con otros cultivos vulnerables. Los tiofenos se definen como compuestos sulfurosos heterocíclicos con una fuerte actividad biocida; entre estos compuestos se encuentran el α -tertienilo y el 5-(3-buten-1-inil)-2,2'-bitienilo (BBT)7 (Gómez y Zavaleta, 2001; Pupo *et al.*; 2007).

Por otro lado, la utilidad de especies como *Tagetes patula* se relaciona directamente con la disponibilidad y uso de plantas con aplicaciones medicinales a manera de biocidas, que conlleva un bajo costo y fácil disponibilidad, accesibilidad y biodegradabilidad de los componentes, en relación con los productos químicos que se sintetizan para tales fines. En el mundo agrícola adquieren relevancia aquellas plantas potencialmente útiles para el manejo de poblaciones de insectos plagas, cuyo efecto biológico se debe a la acción de taninos, flavonoides, alcaloides, aceites esenciales entre otros (Spainhour, 2005; Vu *et al.*; 2006, Medappa, 2003).

En el caso de *T. patula*, una investigación concreta demuestra su acción antihelmíntica contra *Haemonchus contortus* lo que constituyen un aporte significativo en el control de artrópodos de importancia en salud pública. Sin embargo, el conocimiento de sus actividades biológicas, adecuado procesamiento de la muestra y creación de productos terminales conllevan a abrir una excelente alternativa a mediano plazo, con desplazamiento competitivo, de los insecticidas sintéticos (Varshneya & Telang, 2006).

Se han realizado muy pocos estudios enfocados en la definición de los procesos bioquímicos involucrados en la síntesis de compuestos glicosídicos en *Stevia*, uno de éstos, realizado por Richman y colaboradores en 1999, resulta ser clave para el inicio de los estudios moleculares en la planta medicinal. En su investigación los autores explican cómo pueden descifrarse poco a poco las rutas metabólicas en plantas mediante técnicas moleculares que empleen las etiquetas de secuencias expresadas (ESTs), que corresponden a pequeñas secuencias dentro de una secuencia nucleotídica transcrita, que puede ser codificante o no, y que se usan para identificar genes (Humphrey *et al.*, 2006). De esta forma es cómo ellos han establecido la ruta metabólica en *Stevia*, y han identificado cuatro enzimas participantes dentro de la misma.

Sin embargo, aparte de lo mencionado en el párrafo anterior, tanto para *S. rebaudiana* como para *T. patula* no se conocen estudios moleculares enfocados en la caracterización genética, por lo que aún faltan análisis moleculares para conocer con mayor certeza las secuencias de los genes directamente involucrados en la síntesis de los compuestos de interés, de manera que estos puedan ser estudiados con detenimiento, y se pueda comprender su expresión y afectación dadas ciertas condiciones ambientales, con el fin de que la síntesis de estas sustancias pueda ser incrementada y estimulada con miras al escalamiento industrial.

METODOLOGÍA

El presente trabajo se realizó en el Centro de Investigación en Biotecnología (CIB) del Instituto Tecnológico de Costa Rica (ITCR) en Cartago, Costa Rica.

Secuenciación y análisis de transcriptoma

Material Biológico

El material biológico utilizado en la extracción de ARN total, se obtuvo en el caso de *S. rebaudiana* de los invernaderos del CIB. El material biológico de *T. patula* fue comprado en el negocio comercial EPA debido a su tradicional uso como planta ornamental.

En la extracción se utilizaron hojas procedentes del segundo y tercer nudo de *S. rebaudiana*, así como raíces de *T. patula*; en ambos casos las muestras fueron tomadas de plantas seleccionadas al azar, mezcladas en una sola muestra, las cuales, una vez recolectadas, fueron manipuladas con guantes de nitrilo y trasladadas en hielo al laboratorio.

Extracción de ARN total

La extracción de ARN total se realizó a partir de 150 mg de la muestra (hoja o raíz), siguiendo el protocolo *Total RNA Isolation from Plant* del *NucleoSpin RNA Plant kit* marca *Macherey-Nage*[®] (Anexo 1). Se realizaron tres modificaciones al protocolo, la primera en el punto dos del protocolo cuando se eligió el uso del reactivo RAP con mercaptoetanol en lugar del buffer RA1 para la lisis celular; la segunda, en el punto 7, para la digestión de ADN la mezcla se dejó actuar por un período de 30 min con las muestras colocadas en hielo y no a temperatura ambiente durante 15 min como lo sugiere el fabricante; en la tercera modificación, se empleó agua calidad DEPC en el punto 9 en lugar del agua libre de nucleasas que provee el kit, para la elución del ARN de la columna de sílica. Las muestras, los reactivos y las columnas fueron manejados a una temperatura aproximada de 4°C durante todo el proceso de extracción, incluso durante la centrifugación, desde la recolección de las muestras hasta su precipitación.

Las muestras de ARN se cuantificaron con el equipo NanoDrop Lite[®] y se procedió a corroborar la presencia de ARN mediante electroforesis en gel de agarosa al 1,3% durante 45 min a 80 V, cargando en el gel 5 µl de muestra con 2 µl de buffer de carga *Mass Ruler* de Fermentas[®] y 2 µl de marcador de peso molecular *Gene Ruler* 1 kb de Fermentas[®].

Comprobada la presencia de bandas de ARN se continuó con la precipitación de las muestras agregando 0.1 V de acetato de sodio 3M a pH 7-8 mezclando generosamente; luego se añadieron 2 V de etanol absoluto y se volvió a mezclar generosamente. Las muestras se mantuvieron a una temperatura de -20°C hasta el momento de su secuenciación.

Secuenciación

El protocolo de secuenciación utilizado para las muestras de ARN es propiedad de la empresa *Axeq Technologies*, a quien se contrató el servicio. El equipo utilizado para este fin fue el *Illumina HiSeq 2000 Sequencer*. Las muestras se trasladaron a una temperatura de 0 a 4°C a Estados Unidos, en donde lo primero en evaluar fue el *RNA Integrity Number* (RIN) como parámetro estandarizado del control de la integridad de éste ácido nucleico. El software utilizado fue el *Illumina Pipeline (CASAVA) v1.8.2* y el lenguaje de codificación de calidad para la secuenciación fue el de Sanger, dentro del código ASCII. La codificación de la calidad dentro del rango Sanger se confirmó mediante la herramienta *File viewer* de acceso libre.

Análisis de calidad de la secuenciación

Una vez con acceso a los resultados de ambas secuenciaciones, se procedió a verificar la calidad de los fragmentos (*reads*) generados, utilizando las herramientas *FastX-Toolkit (Hannon Lab)*, *Fastqc (Babraham Institute)* y *RobiNA (Lohse et al., 2010)*, siendo todos ellos softwares de acceso libre disponibles en la red.

Se realizaron cortes al inicio de los fragmentos de la base 1 a la base 15 empleando la herramienta *FastX-Toolkit* (tipo *clipping*), para continuar con otro corte para eliminar al final de los fragmentos los segmentos a partir de la aparición de una base con un porcentaje de calidad inferior a 30% (tipo *trimming*). Concluidos los cortes de los *reads* se procedió a filtrar por tamaño, de manera que sólo se continuaran trabajando aquellos fragmentos con una longitud superior a 70 bases; tanto el *trimming* como la discriminación por longitud se realizaron con la herramienta *RobiNA*.

Identificación de secuencias génicas reportadas

La identificación de las secuencias génicas reportadas se hizo empleando la base de datos del *National Center for Biotechnology Information (NCBI, U.S. National Library of Medicine)*; en el caso de *S. rebaudiana* de acuerdo con las publicaciones científicas de Kumar *et al.*, 2012; Mohamed *et al.*, 2011; Richman *et al.*, 2005 y Richman *et al.*, 1999 y para *T. patula* se buscaron secuencias reportadas en el NCBI para esta especie sin previa publicación, así como factores de transcripción reportados en especies de la familia *Astereaceae*, específicamente dentro de las familia *MYC* y *MYB* de factores de transcripción.

La búsqueda de secuencias para *S. rebaudiana* se centró en la identificación de enzimas participantes en la ruta metabólica de síntesis de glicósidos con o sin regiones promotoras descritas, así como la ubicación de al menos un gen constitutivo reportado para la especie en estudio. Para *T. patula* se centró específicamente en la búsqueda de secuencias de factores de transcripción para esta u otras especies dentro de la misma familia. Se realizaron alineamientos empleando la herramienta *Basic Local Alignment Search Tool (BLAST del NCBI)* para cada secuencia encontrada, para evitar seleccionar secuencias repetidas.

Análisis de regiones promotoras, elementos en *cis* y factores de transcripción en *S. rebaudiana*

Las secuencias génicas reportadas como regiones promotoras en *S. rebaudiana*, fueron analizadas mediante la herramienta *Database of Plant Cis-acting Regulatory DNA Elements* (PLACE, *National Institute of Agrobiological Sciences, Higo et al., 1999*) para predecir los posibles sitios de unión de factores de transcripción al ADN, que se ubican en los genes de estevia seleccionados.

La información concerniente a todos los posibles sitios de activación en *cis* presentes en las regiones promotoras de genes de estevia, fue priorizada según su relación con actividades de síntesis de metabolitos primarios y secundarios reportados en la especie modelo *Arabidopsis thaliana*; posteriormente se empleó la herramienta *Plant Interactome Database* (*Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard Medical School*) creada para la especie *A. thaliana*, para relacionar las secuencias reportadas como codificantes de factores de transcripción de esta especie, que podrían interactuar con los sitios en *cis* de las regiones promotoras de *S. rebaudiana*.

Alineamiento de genes contra transcriptomas

Una vez que los fragmentos secuenciados se encontraba dentro de los parámetros de calidad establecidos para este estudio, se continuó con su alineación o mapeo contra los genes seleccionados, siendo éstos de síntesis de glicósidos, un gen constitutivo de estevia, así como factores de transcripción reportados en *T. patula*, *A. thaliana* y en especies de la familia Asteraceae.

Establecimiento de relaciones filogenéticas

Mediante un alineamiento múltiple con la herramienta *Clustalw* (*EMBL-EBI*) se compararon las secuencias nucleotídicas de los factores de transcripción seleccionados para los alineamientos contra los transcriptomas de estevia y tagetes, con la finalidad de generar un árbol filogenético capaz de relacionar las diferentes especies en términos de regulación de la expresión.

RESULTADOS

Secuenciación y análisis de transcriptoma

Extracción de ARN total

La extracción de ARN total a partir de muestras de hoja de *Stevia rebaudiana* se realizó utilizando dos kits, uno marca Macherey-Nagel y otro QIAGEN, no obstante, las dos concentraciones mayores del ácido nucleico se obtuvieron con el primer kit, como se muestra en el cuadro 1 para las muestra 1 y 2, además de generar valores de densidad óptica más cercanos a 2.

Cuadro 1. Concentración de ARN total obtenido a partir de tejido de hoja de *Stevia rebaudiana* utilizando dos kits de extracción comerciales.

Muestra	Kit de extracción utilizado	Concentración de ARN (ng/μl)	Densidad óptica (A260/280)
1	NucleoSpin RNA Plant kit marca Macherey-Nagel®	348.7	2.13
2	NucleoSpin RNA Plant kit marca Macherey-Nagel®	159.4	1.99
3	RNeasy Plant Mini Kit marca QIAGEN®	33.8	1.85
4	RNeasy Plant Mini Kit marca QIAGEN®	117.0	1.66

Fuente: CIB, 2014.

La extracción de ARN total a partir de muestras de raíz de *Tagetes patula* se realizó utilizando el kit Macherey-Nagel. En el cuadro 2 se presentan los valores de concentración y densidad óptica obtenidos.

Cuadro 2. Concentración de ARN total obtenido a partir de tejido de raíz de *Tagetes patula* utilizando un kit de extracción comercial.

Muestra	Kit de extracción utilizado	Concentración de ARN (ng/μl)	Densidad óptica (A260/280)
1	NucleoSpin RNA Plant kit marca Macherey-Nagel®	18.0	1.84
2	NucleoSpin RNA Plant kit marca Macherey-Nagel®	59.5	1.90
3	NucleoSpin RNA Plant kit marca Macherey-Nagel®	93.9	2.02
4	NucleoSpin RNA Plant kit marca Macherey-Nagel®	127.0	1.92

Fuente: CIB, 2014.

Para verificar la presencia del ácido nucleico y visualizar su integridad, las muestras se sometieron a una electroforesis que generó dos bandas para cada una de las muestras, las cuales tenían un tamaño de entre 400 y 200 bases, como se observa en las figura 1 y 2.

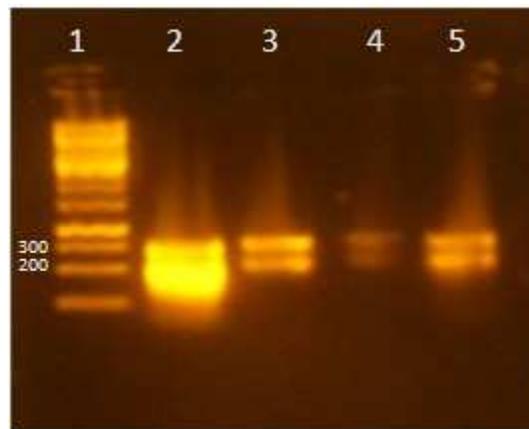


Figura 1. Presencia de bandas de ARN total de *Stevia rebaudiana* en gel de agarosa al 1.3%: 1. Marcador de peso molecular Gene Ruler 1 kb®, 2. Muestra 1 kit Macherey Nagel®, 3. Muestra 2 kit Macherey Nagel®, 4. Muestra 1 kit QIAGEN®, 5. Muestra 2 kit QIAGEN®.

Fuente: CIB, 2014.

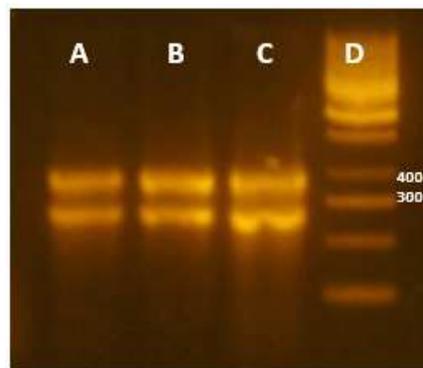


Figura 2. Presencia de bandas de ARN total de *Tagetes patula* en gel de agarosa al 1.3%: A. Muestra 1, B. Muestra 2, C. Muestra 3, D. Marcador de peso molecular Gene Ruler 1 kb®.

Fuente: CIB, 2014.

Secuenciación

Las muestras 2 de ARN total de cada especie fueron seleccionadas para su secuenciación; en los cuadros 3 y 4 se presentan los resultados del análisis de calidad al que fueron sometidas una vez resuspendidas.

Cuadro 3. Datos de calidad para la muestra 2 de ARN total de *Stevia rebaudiana* previo a su secuenciación.

Muestra	Concentración de ARN (ng/ µl)	Volumen (µl)	Cantidad de ARN (µg)	RIN	Densidad óptica (A260/280)
2	84.6	27	2.285	9.3	1.9

Fuente: AXEQ Technologies, 2013.

Cuadro 4. Datos de calidad para la muestra 2 de ARN total de *Tagetes patula* previo a su secuenciación.

Muestra	Concentración de ARN (ng/ µl)	Volumen (µl)	Cantidad de ARN (nM)	RIN	Densidad óptica (A260/280)
2	73.46	27	415.48	9.5	1.9

Fuente: AXEQ Technologies, 2013.

Los valores finales del proceso de secuenciación, incluyendo el número de fragmentos generados, se aprecia en los cuadros 5 y 6, en los cuales se hace notar el porcentaje de fragmentos que superaron el 30% de calidad Sanger, rango que se verificó con la herramienta *File Viewer*. Los datos se dividieron en dos librerías tipo *paired-end* para ambas especies.

Cuadro 5. Datos de calidad de la secuenciación de la muestra 2 de ARN total de *Stevia rebaudiana*.

Muestra	N° total de bases secuenciadas	N° de fragmentos	%N ^a	%GC ^b	%QC20 ^c	%QC30 ^d
2	5 728 786 458	56 720 658	0.003	52.70	95.50	88.81

a: Porcentaje de bases no identificadas; b: Porcentaje de bases guanina-citosina; c: Porcentaje de fragmentos con calidad Sanger superior a 20%; d: Porcentaje de fragmentos con calidad Sanger superior a 30%

Fuente: AXEQ Technologies, 2013.

Cuadro 6. Datos de calidad de la secuenciación de la muestra 2 de ARN total de *Tagetes patula*.

Muestra	N° total de bases secuenciadas	N° de fragmentos	%N ^a	%GC ^b	%QC20 ^c	%QC30 ^d
2	4 685 513 624	49 391 224	0.021	43.45	97.17	93.29

a: Porcentaje de bases no identificadas; b: Porcentaje de bases guanina-citosina; c: Porcentaje de fragmentos con calidad Sanger superior a 20%; d: Porcentaje de fragmentos con calidad Sanger superior a 30%

Fuente: AXEQ Technologies, 2013.

Análisis de calidad de la secuenciación

Se verificó que el valor de calidad asignado a cada una de las bases de los fragmentos, se encontraba en el rango de valores ASCII, correspondiente al porcentaje Sanger de calidad, ya que los caracteres correspondieron a los ubicados en el rango de 33 a 73 de dicho código, lo cual ubica el porcentaje de calidad de los fragmentos entre 0 y 40%, resultado que permitió continuar con la valoración de otros parámetros de calidad.

El primer dato verificado correspondió a la composición de nucleótidos de las librerías, lo cual se detalla en los cuadros 7 y 8.

Cuadro 7. Composición nucleotídica para dos librerías de ADN copia de *Stevia rebaudiana* con secuencias sin depurar.

Nucleótido	Representatividad en los fragmentos (%)	
	Librería 1	Librería 2
Adenina	23.8	23.6
Timina	23.6	23.6
Citosina	26.4	26.5
Guanina	26.2	26.3
Desconocido	0.0	0.0

Fuente: RobiNA, 2014

Cuadro 8. Composición nucleotídica para dos librerías de ADN copia de *Tagetes patula* con secuencias sin depurar.

Nucleótido	Representatividad en los fragmentos (%)	
	Librería 1	Librería 2
Adenina	28.3	28.3
Timina	28.2	28.2
Citosina	22.0	22.1
Guanina	21.4	21.4
Desconocido	0.0	0.0

Fuente: RobiNA, 2014

Se continuó con la valoración de la calidad promedio de las bases nucleotídicas para el total de fragmentos amplificados, como se presenta en las figuras 3 y 4, en las cuáles además se evidencia que la longitud promedio de los fragmentos fue de 101 bases.

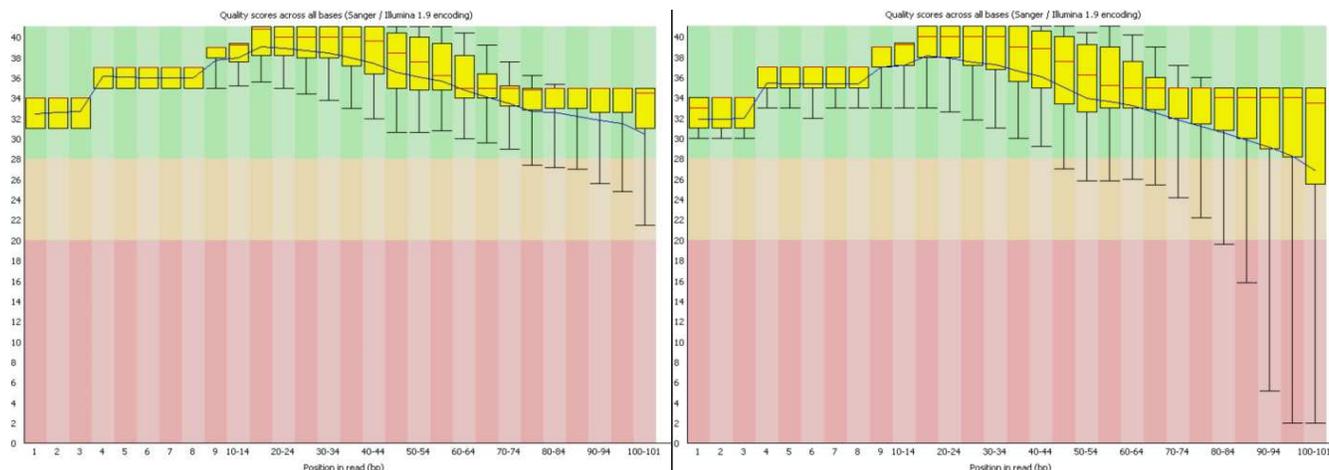


Figura 3. Calidad promedio por base para el total de fragmentos amplificados de ADN copia de *Stevia rebaudiana* con secuencias sin depurar: a. Librería 1, b. Librería 2.

Fuente: FASTQC, 2014.

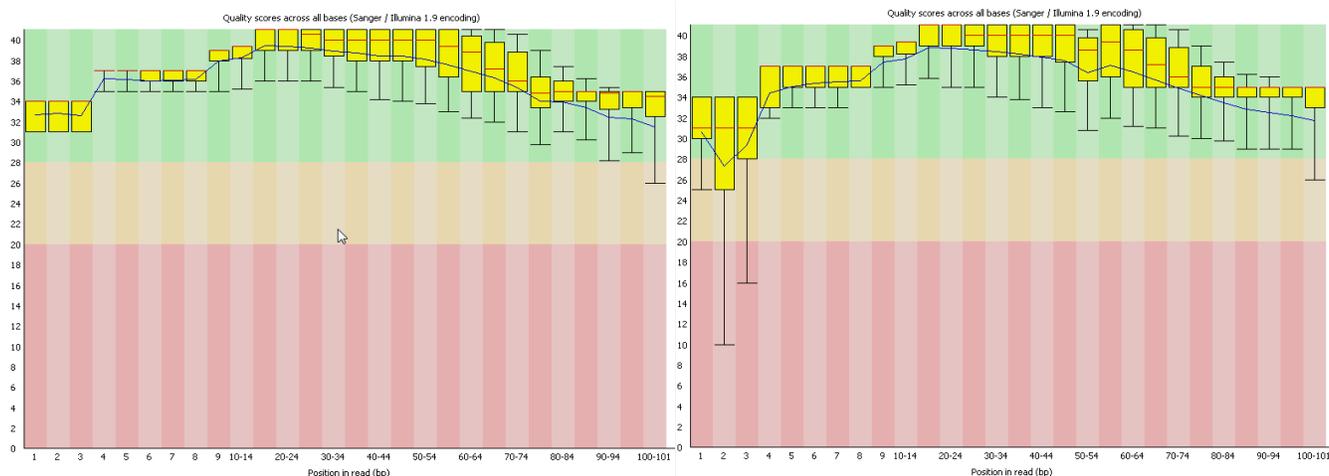


Figura 4. Calidad promedio por base para el total de fragmentos amplificados de ADN copia de *Tagetes patula* con secuencias sin depurar: a. Librería 1, b. Librería 2.

Fuente: FASTQC, 2014.

Mediante una gráfica de distribución de la longitud de las secuencias, la cual se presenta en la figura 5, se confirmó la uniformidad en el tamaño de los fragmentos amplificados durante la secuenciación, el cual se mantuvo constante generando fragmentos de 101 bases.

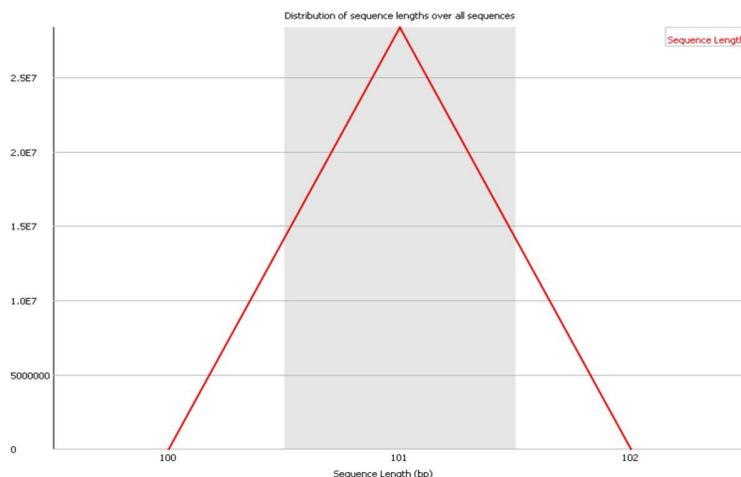


Figura 5. Distribución de la longitud de las frecuencias para las librerías 1 y 2 de ADN copia de *Stevia rebaudiana* y *Tagetes patula* con secuencias sin depurar.

Fuente: FASTQC, 2014.

En *Stevia rebaudiana* la calidad promedio de los fragmentos de la librería 1 fue de 35%, tal y como se muestra en la figura 6, para la librería 2 la calidad promedio fue de aproximadamente 33.5%.

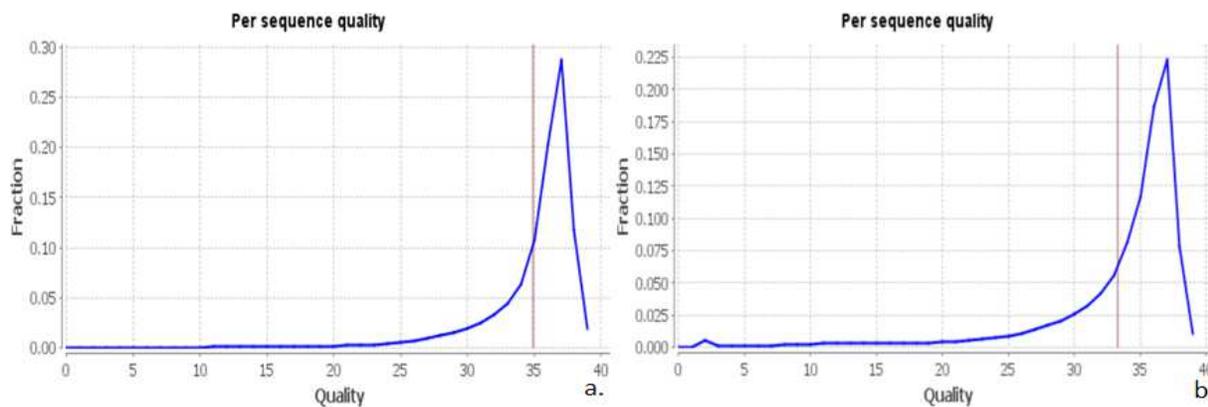


Figura 6. Distribución de la calidad a lo largo de los fragmentos amplificados de ADN copia de *Stevia rebaudiana* con secuencias sin depurar: a. Librería 1, b. Librería 2.

Fuente: RobiNA, 2014.

En *Tagetes patula* la calidad promedio de los fragmentos de la librería 1 fue de 36%, tal y como se muestra en la figura 7, para la librería 2 la calidad promedio fue de aproximadamente 35%.

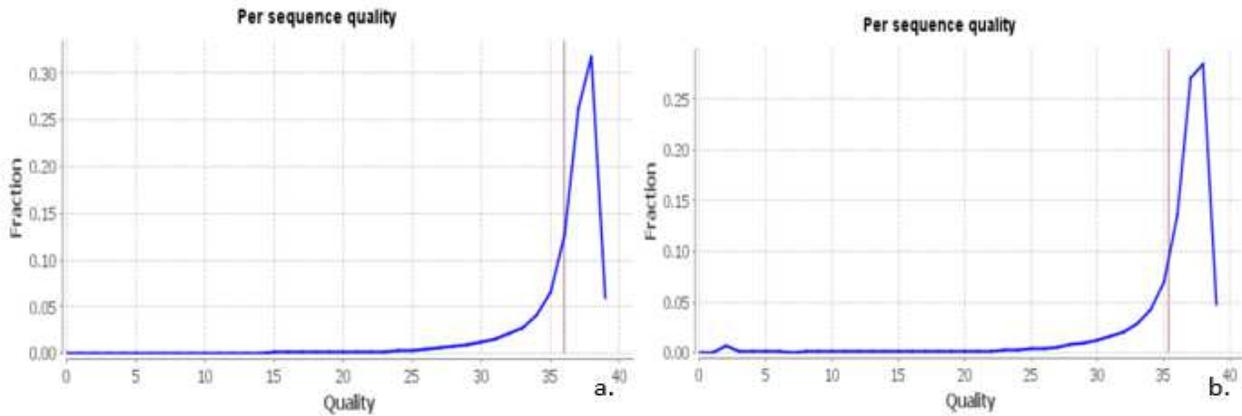


Figura 7. Distribución de la calidad a lo largo de los fragmentos amplificados de ADN copia de *Tagetes patula* con secuencias sin depurar: a. Librería 1, b. Librería 2.

Fuente: RobiNA, 2014.

La frecuencia de aparición de las bases a lo largo de los fragmentos para ambas especies se presenta en las figuras 8 y 9; la proporción de bases en los fragmentos parece equilibrarse en valores de entre 20 y 30% para todos los nucleótidos.

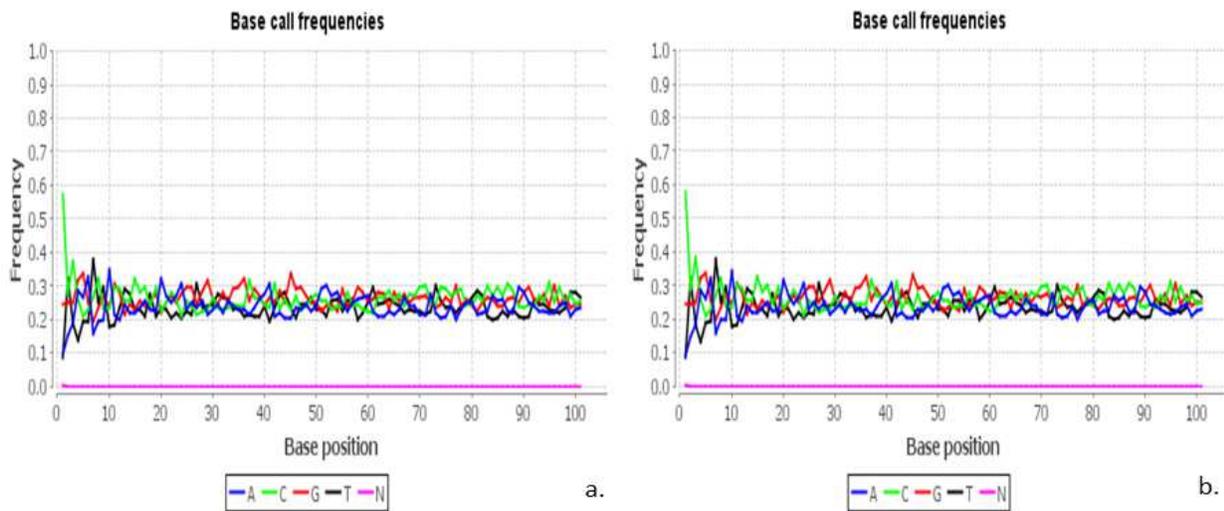


Figura 8. Distribución de la frecuencia de bases secuenciadas en los fragmentos amplificados de ADN copia de *Stevia rebaudiana* con secuencias sin depurar: a. Librería 1, b. Librería 2.

Fuente: RobiNA, 2014.

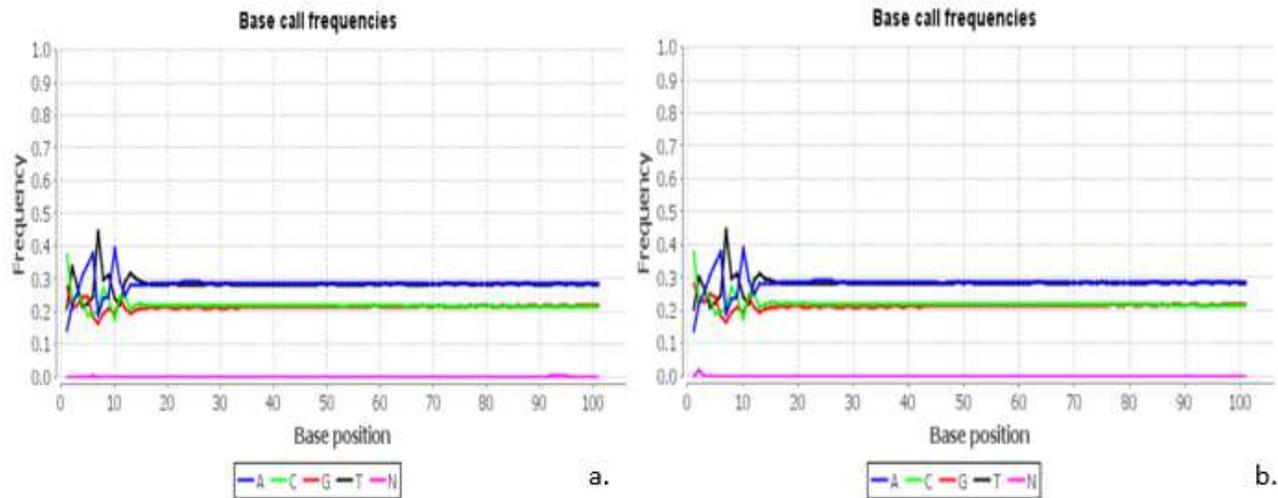


Figura 9. Distribución de la frecuencia de bases secuenciadas en los fragmentos amplificados de ADN copia de *Tagetes patula* con secuencias sin depurar: a. Librería 1, b. Librería 2.

Fuente: RobiNA, 2014.

Al evaluar los parámetros de calidad de los fragmentos generados con la secuenciación, se definió la realización de un corte al inicio de todas las secuencias en ambas librerías, el cual se conoce como *clipping*, eliminando de la base 1 a la 15. Posterior a este corte, se realizó otro, mediante el cual se eliminaron las bases con porcentajes inferiores al 30% de calidad Sanger hacia el final de los fragmentos, iniciando en el nucleótido 101 y devolviéndose en la secuencia hasta encontrar una base con calidad igual o superior al 30%, sitio en el que terminó el corte, conocido como *trimming*. Las secuencias que al final del proceso de corte tuvieron una longitud inferior a 70 bases fueron eliminadas.

Del total de 56 720 658 fragmentos generados con la secuenciación de *S. rebaudiana*, se mantuvieron 53 998 440 fragmentos después de los cortes, correspondiendo 27 627 790 a la librería 1 y 26 236 583 a la librería 2. En el caso de *T. patula*, del total de 49 391 224 se mantuvieron 45 126 895 fragmentos, separados 22 729 639 en la librería 1 y 22 397 256 en la librería 2.

Concluidos los cortes de las secuencias, para depurar la información por utilizar en los alineamientos, se evaluaron nuevamente los parámetros de calidad. Se inició con la evaluación de la composición de nucleótidos de las librerías, lo cual se detalla en los cuadros 9 y 10.

Cuadro 9. Composición nucleotídica para dos librerías de ADN copia de *Stevia rebaudiana* con secuencias depuradas.

Nucleótido	Representatividad en los fragmentos (%)	
	Librería 1	Librería 2
Adenina	24.2	24.2
Timina	23.8	23.9
Citosina	25.9	25.9
Guanina	26.1	26.0
Desconocido	0.0	0.0

Fuente: RobiNA, 2014

Cuadro 10. Composición nucleotídica para dos librerías de ADN copia de *Tagetes patula* con secuencias depuradas.

Nucleótido	Representatividad en los fragmentos (%)	
	Librería 1	Librería 2
Adenina	28.5	28.6
Timina	28.3	28.3
Citosina	21.8	21.9
Guanina	21.4	21.3
Desconocido	0.0	0.0

Fuente: RobiNA, 2014

Se continuó con la valoración de la calidad promedio de las bases nucleotídicas para el total de fragmentos amplificados, como se presenta en las figuras 10 y 11.

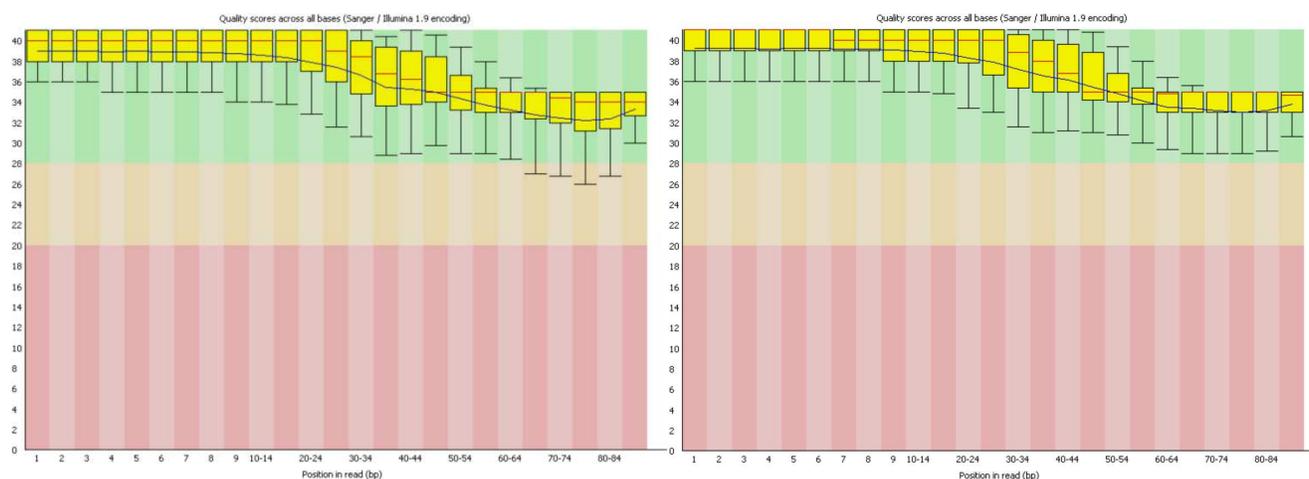


Figura 10. Calidad promedio por base para el total de fragmentos amplificados de ADN copia de *Stevia rebaudiana* con secuencias depuradas: a. Librería 1, b. Librería 2.

Fuente: FASTQC, 2014.

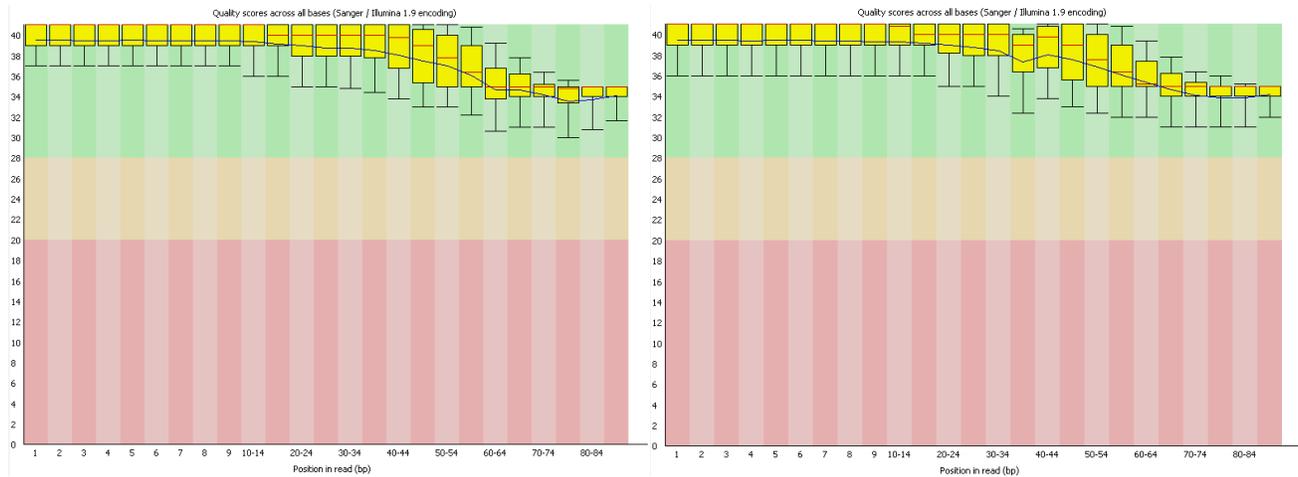


Figura 11. Calidad promedio por base para el total de fragmentos amplificados de ADN copia de *Tagetes patula* con secuencias depuradas: a. Librería 1, b. Librería 2.

Fuente: FASTQC, 2014.

Mediante la gráfica de distribución de la longitud de las secuencias que se presenta en la figura 12, se muestra que el tamaño de las secuencias que se conservaron de acuerdo con su calidad, fue superior a 80 bases y alcanzaron una longitud máxima de 87 bases para ambas especies.

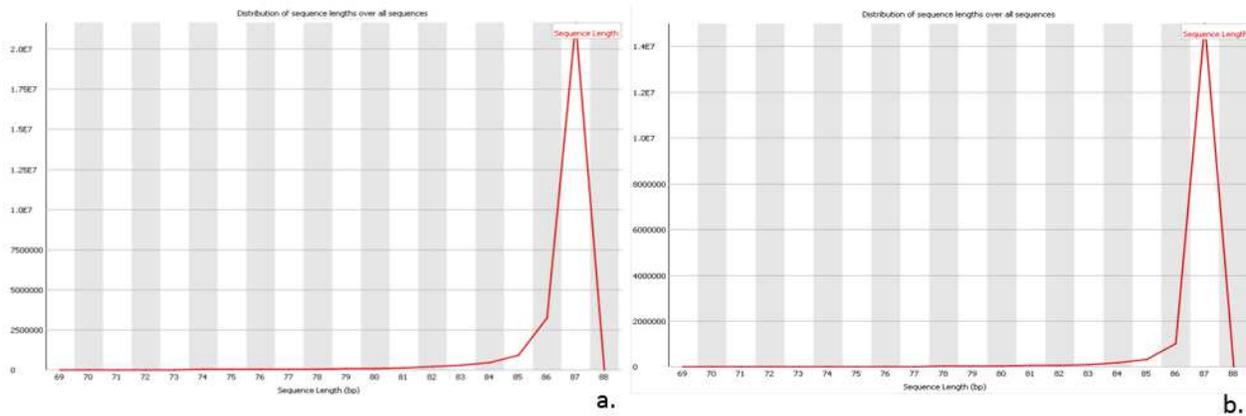


Figura 12. Distribución de la longitud de las frecuencias para las librerías 1 y 2 de ADN copia con secuencias depuradas: a. *Stevia rebaudiana* y b. *Tagetes patula*.

Fuente: FASTQC, 2014.

Además, se verificó la calidad promedio de los fragmentos para ambas especies; como se muestra en las figuras 13 y 14.

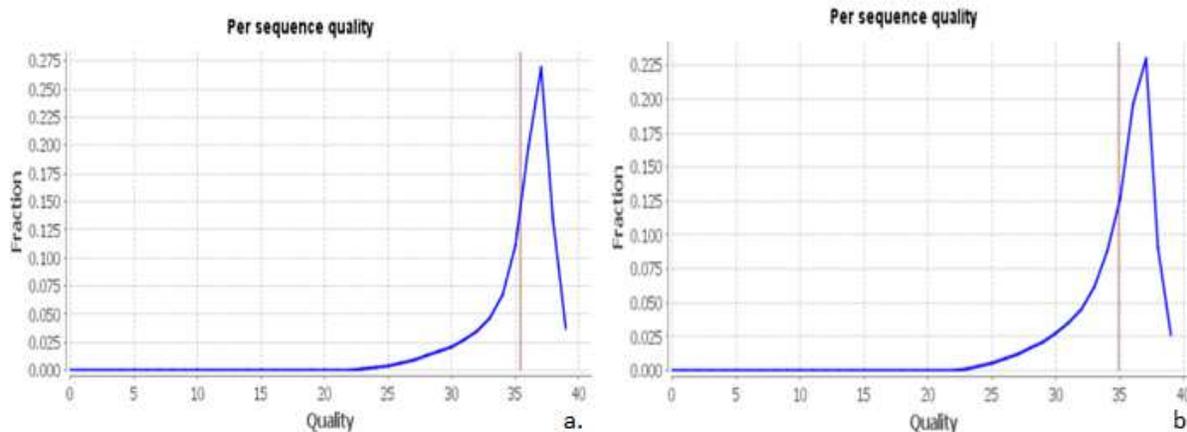


Figura 13. Distribución de la calidad a lo largo de los fragmentos amplificados de ADN copia de *Stevia rebaudiana* con secuencias depuradas: a. Librería 1, b. Librería 2.

Fuente: RobiNA, 2014.

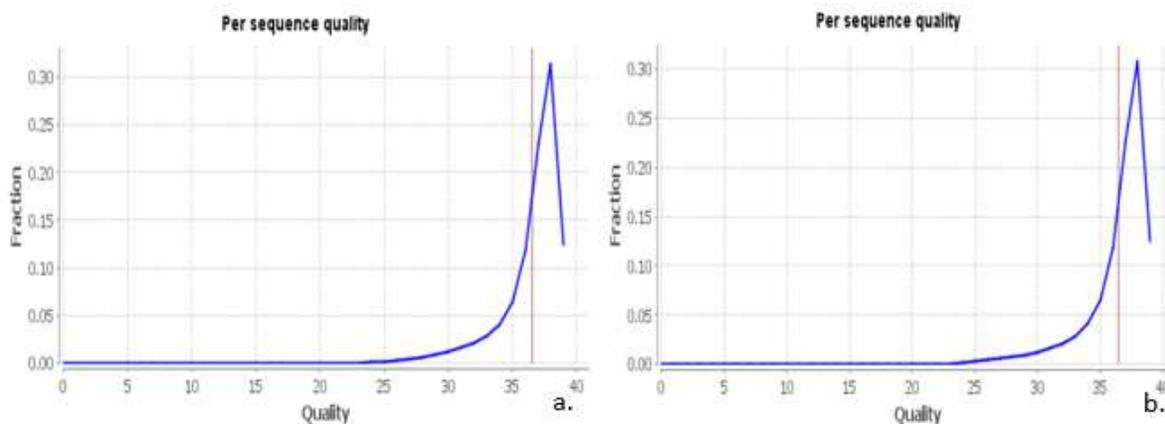


Figura 14. Distribución de la calidad a lo largo de los fragmentos amplificados de ADN copia de *Tagetes patula* con secuencias depuradas: a. Librería 1, b. Librería 2.

Fuente: RobiNA, 2014.

La frecuencia de aparición de las bases a lo largo de los fragmentos depurados para ambas librerías se presenta en las figuras 15 y 16. De acuerdo con las gráficas, la frecuencia de aparición de los diferentes nucleótidos en las secuencias, se mantuvo proporcional en un rango entre 20 y 30% en ambas especies.

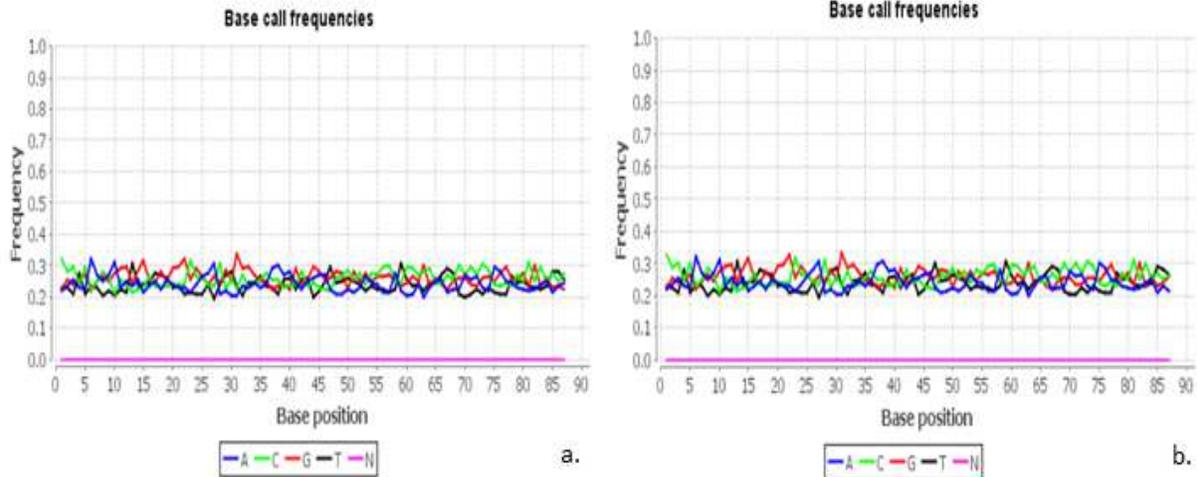


Figura 15. Distribución de la frecuencia de bases secuenciadas en los fragmentos amplificados de ADN copia de *Stevia rebaudiana* con secuencias depuradas: a. Librería 1, b. Librería 2.

Fuente: RobiNA, 2014.

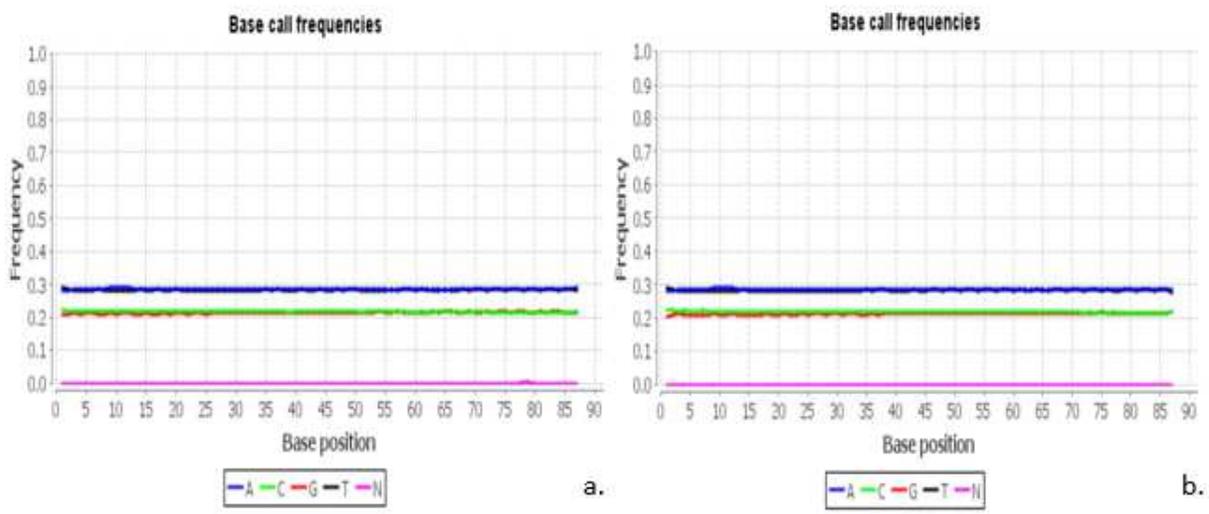


Figura 16. Distribución de la frecuencia de bases secuenciadas en los fragmentos amplificados de ADN copia de *Tagetes patula* con secuencias depuradas: a. Librería 1, b. Librería 2.

Fuente: RobiNA, 2014.

Identificación de secuencias génicas reportadas

Análisis de regiones promotoras, elementos en *cis* y factores de transcripción en *S. rebaudiana*

Mediante la búsqueda bioinformática de genes de *Stevia rebaudiana* en el NCBI, se logró determinar que existen al menos 20 genes diferentes descritos para esta especie, los cuáles participan dentro de la ruta de síntesis de glicósidos. La herramienta *Blast* permitió verificar que cada secuencia fuera única, es decir, que no se contemplaran diferentes accesiones referentes a una misma secuencia; los resultados de la búsqueda se resumen en el cuadro 11.

Cuadro 11. Genes de *Stevia rebaudiana* reportados en el *National Center for Biotechnology Information*.

N° de gen	Nombre del gen	Accesión	Promotor descrito	Accesión del promotor
1	UDP-glicosiltransferasa 79A2	AY345985.1	No	-
2	UDP- glicosiltransferasa 89B2	AY345983.1	No	-
3	UDP- glicosiltransferasa 88B1	AY345981.1	No	-
4	UDP- glicosiltransferasa 73E1	AY345979.1	No	-
5	UDP- glicosiltransferasa 76H1	AY345977.1	No	-
6	UDP- glicosiltransferasa 85A8	AY345975.1	No	-
7	UDP- glicosiltransferasa 85C1	AY345984.1	No	-
8	UDP- glicosiltransferasa 91D1	AY345980.1	No	-
9	UDP-glicosiltransferasa 71E1	AY345976.1	No	-
10	4-difosfocitidil-2-C-metil-D-eritritol kinasa	DQ269453.5	Si	FJ755687.1
11	IPP/DMAPP sintasa (ispH)	DQ269451.4	Si	FJ755690.1
12	geranilgeranil difosfato sintasa	DQ432013.3	Si	FJ755692.1
13	copalil pirofosfato sintasa (Cpps1)	AF034545.1	No	-
14	kaureno sintasa (KS22-1)	AF097311.1	No	-
15	kaureno sintasa (KS1-1)	AF097310.1	No	-
16	ácido ent-kaurenoic 13-hidroxilasa (KA13H)	DQ398871.3	Si	FJ755693.1
17	UDP-glicosiltransferasa 85C2	AY345978.1	No	-
18	UDP-glicosiltransferasa 74G1	AY345982.1	No	-
19	UDP-glicosiltransferasa 76G1	AY345974.1	No	-
20	NADPH citocromo P450 reductasa	DQ269454.4	Si	FJ755694.1
21	Actina	AF548026	No	-

Fuente: NCBI, 2014

La función asociada a los genes del cuadro 11 se presenta en el cuadro 12, así como su participación dentro de la ruta metabólica de síntesis de glicósidos (figura 17).

Cuadro 12. Función de 21 genes descritos en el *National Center for Biotechnology Information* para la especie *Stevia rebaudiana* y su relación con la síntesis de glicósidos.

Gen	Función	N° de reacción catalizada en la ruta metabólica
1 al 9	Síntesis de UDP-glicosiltransferasas	No definido
10	Síntesis de 4-(Citidina5'difosfo)-2-metil-D-eritritol 2-fosfato a partir de 4-(Citidina5'difosfo)-2-metil-D-eritritol	4
11	Isomerización de dimetilallil difosfato a isopentenil difosfato	8
12	Síntesis de geranilgeranil difosfato a partir de dimetilallil difosfato	9
13	Ciclación de geranil geranil pirofosfato (ggpp)en copalil pirofosfato (cpp)	10
14 y 15	Síntesis de (-)-kaureno por ionización dependiente de la ciclación del copalil difosfato	11
16	Síntesis de esteviol (ent-kaur-16-en-13-ol-19-oic ácido) por hidroxilación del ácido ent-kaurenoico (ent-kaur-16-en-19-oic ácido) con la participación de NADPH y oxígeno molecular	13
17	Síntesis de esteviolmonósido a partir de esteviol	14
18	Síntesis de esteviósido a partir de esteviobiósido	16
19	Síntesis de rebaudiósido A a partir de esteviósido	17
20	No definida	No definido
21	Síntesis de actina	-

Fuente: NCBI, 2014; Kumar *et al.*, 2012.

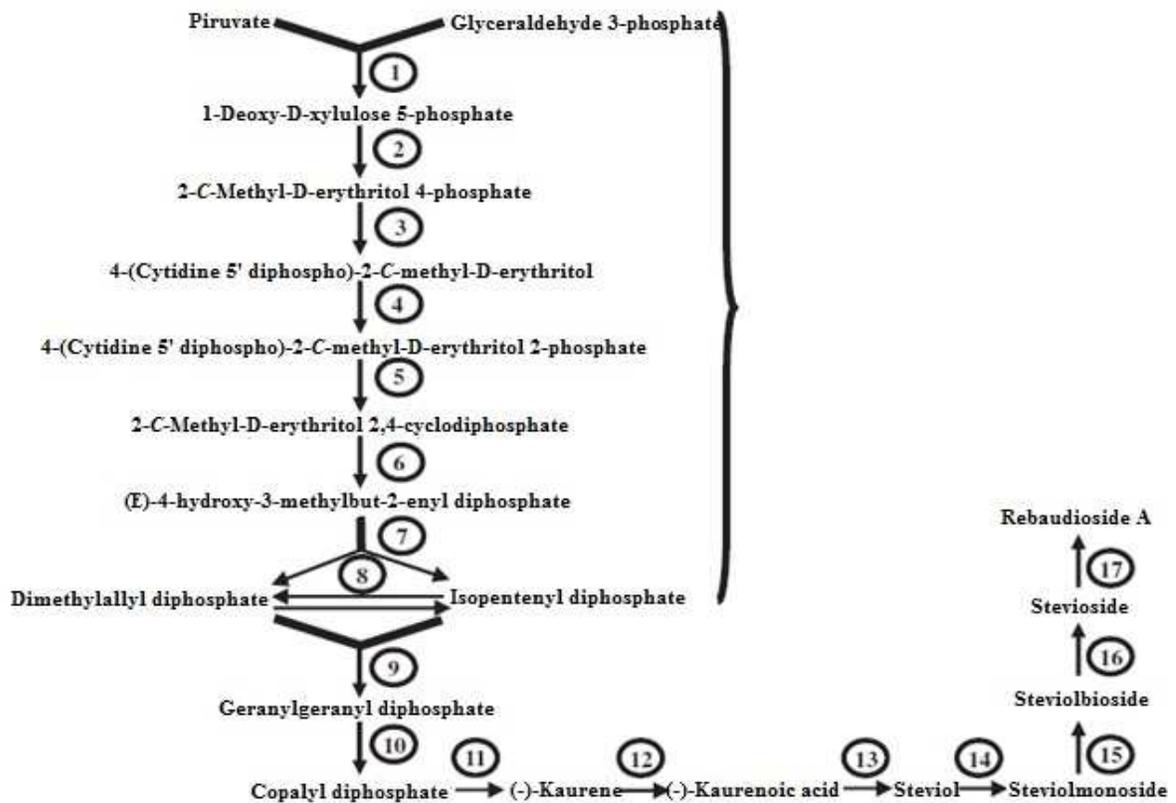


Figura 17. Esquema de la ruta metabólica de síntesis de compuestos glicosídicos en *S. rebaudiana* a partir de gliceraldehído-3- fosfato.

Fuente: Kumar *et al.*, 2012.

Con los resultados de la búsqueda de genes de estevia, se continuó con el análisis de promotores para los 5 genes que reportaban estas secuencias. La herramienta *PLACE* permitió verificar la presencia de las cajas CAAT, GATA y TATA corriente arriba de los genes respectivos, pero además los relacionó con la posible presencia de numerosos sitios de regulación y señalización de la transcripción.

Debido a que los sitios identificados por el programa como posibles sitios de anclaje de factores de regulación de la transcripción eran muchos y muy diversos, se seleccionaron únicamente aquellos descritos en la especie modelo *Arabidopsis thaliana*, específicamente con funciones en procesos biológicos en tejido de hoja, como se detalla en el cuadro 13.

Cuadro 13. Función de 11 sitios de regulación de la transcripción descritos en *Arabidopsis thaliana* y sus proteínas de anclaje.

Nombre de los sitios en <i>cis</i> seleccionados	Función de los sitios en <i>cis</i>	Factor de regulación relacionado	Accesión factor de transcripción
ABRELATERD1 (S000414)	Respuesta a la deshidratación por activación de una proteína quinasa dependiente de ácido fosfatídico.	<i>Arabidopsis thaliana</i> 3'-phosphoinositide-dependent protein kinase 1	NM_203001.3
ACGTATERD1 (S000415)	Respuesta temprana a la deshidratación por la Clp proteasa como subunidad reguladora dependiente de ATP	<i>Arabidopsis thaliana</i> chaperone protein ClpD	NM_124486.2 (3202 pb)
ARR1AT (S000454)	Regulador de respuesta por activación transcripcional mediante proteína reguladora dependiente de citoquinina	<i>Arabidopsis thaliana</i> AT3G16857	AK316883.1 (2346 pb)
DRECRTCOREAT (S000418)	Respuesta temprana a la deshidratación por acción de una histidina quinasa y un dominio de respuesta regulatoria	<i>Arabidopsis thaliana</i> ethylene receptor	NM_105305.4
EVENINGAT (S000385)	Control circadiano de la expresión	--	--
LTRECOREATCOR15 (S000153)	Respuesta a bajas temperaturas	--	--
MYB2CONSENSUSAT (S000409)	Sitios de reconocimiento del factor de transcripción MYB que posee un dominio de unión de ADN MYB R2R3 y es conocido por regular la expresión de genes de respuesta a deshidratación y salinidad.	<i>Arabidopsis thaliana</i> R2R3 MYB DNA binding domain transcription factor	NM_130287.2
MYBCORE (S000176)	Respuesta a estrés hídrico y estímulo a auxina, etileno, ácido jasmónico, ácido salicílico y sal.	<i>Arabidopsis thaliana</i> myb domain protein 14	NM_128674.3
MYCATERD1 (S000413)	Respuesta temprana a la deshidratación	-	-
MYCATRD22 (S000174)	Respuesta a la deshidratación, estrés oxidativo, ABA relacionado con síntesis de triptófano	<i>Arabidopsis thaliana</i> transcription factor MYC2	NM_102998.3
MYCONSENSUSAT (S000407)	Sitios de reconocimiento del factor de transcripción bHLH de desarrollo morfológico de hoja	<i>Arabidopsis thaliana</i> transcription factor EGL1 mRNA	NM_001198373.1

Fuente: Interactome, 2014; NCBI, 2014; PLACE; 2014

Selección de factores de transcripción para *T. patula* y especies de la familia Asteraceae

Mediante la búsqueda bioinformática de genes de *Tagetes patula* en el NCBI y Plant Transcription Factor Database 3.0, se logró determinar que existen al menos 5 genes diferentes descritos para esta especie. Además, se seleccionaron genes de las familias MYC y MYB de factores de transcripción,

dentro de especies de la familia Asteraceae. Los resultados de la búsqueda se resumen en los cuadros 14 y 15.

Cuadro 14. Genes de *Tagetes patula* reportados en el *National Center for Biotechnology Information*.

N° de gen	Nombre del gen	Accesión	Especie	Familia de factores de transcripción
1	Tagetes patula apetala 1-1 mRNA, complete cds	JX310277.1	<i>Tagetes patula</i>	Apetala
2	Tagetes patula chloroplast ndhF gene for NADH dehydrogenase subunit F, partial cds	AB530942.1	<i>Tagetes patula</i>	-
3	Tagetes patula cultivar guifei cyclophilin 2 (CYP2) mRNA, complete cds	HQ454979.1	<i>Tagetes patula</i>	-
4	Tagetes patula TpPAP1 mRNA for acid phosphatase precursor, complete cds	AB029086.2	<i>Tagetes patula</i>	-
5	Tagetes patula cultivar guifei cyclophilin 1 (CYP1) mRNA, partial cds	HQ454978.1	<i>Tagetes patula</i>	-

Fuente: NCBI, 2014.

Cuadro 15. Factores de transcripción en especies de la familia Asteraceae reportados en *Plant Transcription Factor Database*.

N° de gen	Transcription factor ID	Especie	Familia de factores de transcripción
6	Lsa003257	<i>Lactuca sativa</i>	MYB
7	Aan001331	<i>Artemisia annua</i>	ARR-B
8	Han002636	<i>Helianthus annuus</i>	MICK
9	Han008835	<i>Helianthus annuus</i>	M-Type
10	Han001541	<i>Helianthus annuus</i>	MYB

Fuente: PlantTFDB, 2014.

Alineamiento de genes contra el transcriptoma

Al cumplir los fragmentos con los parámetros de calidad determinados en esta investigación, se continuó con los alineamientos entre éstos y los genes seleccionados para ambas especies; se aceptó un no alineamiento de menos de 3 nucleótidos por fragmento.

Para *S. rebaudiana* se escogieron 8 genes de ruta metabólica de síntesis de glicósidos en *Stevia rebaudiana*, el gen constitutivo de actina y tres factores de transcripción descritos para la especie *Arabidopsis thaliana*. En el cuadro 16 se presentan los resultados.

Cuadro 16. Alineamiento de 11 genes descritos para las especies *Stevia rebaudiana* y *Arabidopsis thaliana* contra el transcriptoma de *Stevia rebaudiana*.

Especie	N° de gen o sitio de anclaje	N° de nucleótidos del gen (pb)	N° de fragmentos alineados	N° mínimo de nucleótidos alineados	% de alineamiento
<i>S. rebaudiana</i>	13	2 590	6 547	458 290	0.02
	14	2 792	170	11 900	0.00
	15	3 117	230	16 100	0.00
	16	1 678	384	26 880	0.00
	17	1 586	4 479	313 530	0.02
	18	1 555	521	36 470	0.00
	19	1 616	1 407	98 490	0.01
	21	1 396	1 347	94 290	0.01
<i>A. thaliana</i>	MYB2CONSESUSAT	1 138	0	0	0.00
	MYCATRD22	2 179	0	0	0.00
	MYCCONSENSUSAT	2 202	0	0	0.00

Fuente: RobiNA, 2014

Por su parte, para los alineamientos contra el transcriptoma de *T. patula* los genes trabajados fueron los 1, 3 y 4 de cuadro 14, reportados en el NCBI para *T. patula* específicamente, así como los genes 6,7,8,9 y 10, reportados en diferentes especies de la familia Asteraceae, tal y como se muestra en el cuadro 17.

Cuadro 17. Alineamiento de 3 genes descritos para la especie *T. patula* y 5 genes de especies de la familia Asteraceae contra el transcriptoma de *Tagetes patula*.

Especie	N° de gen	N° de nucleótidos del gen (pb)	N° de fragmentos alineados	N° mínimo de nucleótidos alineados	% de alineamiento
<i>T. patula</i>	1	919	11 040	772 800	0.05
	3	819	4 152	290 640	0.02
	4	1 843	774	54 180	0.00
Familia Asteraceae	6	1347	0	0	0.00
	7	1605	0	0	0.00
	8	744	0	0	0.00

9	305	0	0	0.00
10	1 445	0	0	0.00

Fuente: RobiNA, 2014

Establecimiento de relaciones filogenéticas

Para el análisis de las relaciones filogenéticas entre factores de transcripción en las especies *T. patula*, *L. sativa*, *A. thaliana*, *A. annua* y *H. annus*, se generó un árbol filogenético mediante alineamiento múltiple con los mismos genes comparados contra los transcriptomas de *S.rebaudiana* y *T. patula*. En la figura 18 se presenta el resultado.

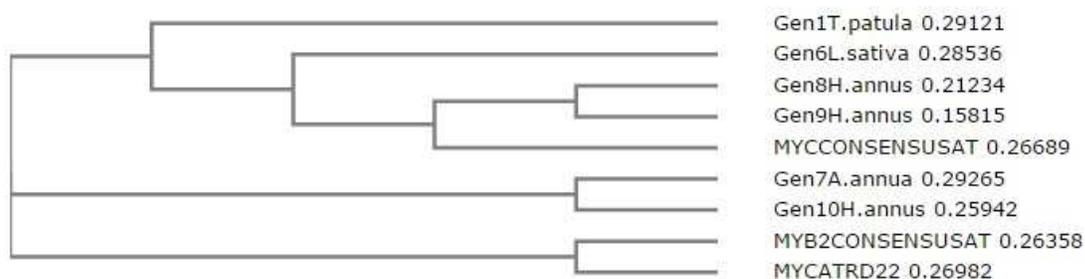


Figura 18. Árbol filogenético para secuencias de factores de transcripción en varias especies de la familia Asteraceae.

Fuente: Clustalw, 2015; NCBI, 2014; PlantTFDB, 2014.

DISCUSIÓN

Secuenciación y análisis de transcriptoma

Extracción de ARN total

En esta investigación se seleccionaron dos kits comerciales de las marcas QIAGEN y Macherey-Nagel para extracción de ARN de plantas. Se debe recordar que al concluir cualquier extracción de ARN es necesario recurrir a una evaluación de la calidad del mismo, para demostrar su integridad y pureza (Farrell, 2010). La primera técnica de control que se aplica usualmente es el perfil electroforético; por lo general para este fin se emplean geles de agarosa con algún compuesto fluorescente (bromuro de etidio, gel red) que evidencie la presencia de bandas en un perfil definido como resultado positivo o, la presencia de *smear* (material degradado), por lo general mediante exposición a luz UV (Karp, 2009).

Con ambos kits en este caso se obtuvo la presencia de dos bandas para ambas especies, con un peso molecular entre 200 bases 400 bases. Al visualizar la presencia de bandas de ARN en el gel de agarosa, se continuó con la cuantificación del mismo mediante el equipo NanoDrop lite (Thermo Scientific) para obtener su concentración.

La última prueba de calidad que se le realizó al ARN antes de seleccionar la muestra para su secuenciación fue la de densidad óptica a 260 y 280 nm ($OD_{260/280}$), la cual se relaciona con el grado de pureza de la molécula. Se considera que cuando esta relación de las lecturas genera un valor de 2 o muy cercano, la molécula de ARN tiene una pureza aceptable, contrarrestando la posibilidad de que la lectura se altere por la presencia de contaminación por proteínas, ADN o solventes orgánicos que absorben a longitudes de onda cercana (López 2007; Wilfinger *et al.*, 1997). En este caso, los valores de densidad óptica fueron adecuados en las muestras de ambas especies seleccionadas como candidatas para la secuenciación.

Secuenciación

Las técnicas de secuenciación de nueva generación permiten la obtención de millones de fragmentos secuenciados a partir del ADN copia (ADNc) de una muestra de ARNm; estos fragmentos se alinean contra un genoma de referencia y su abundancia es analizada en la muestra como medida cuantitativa de la expresión génica (Soneson y Delorenzi, 2013).

Con las secuenciaciones de los transcriptomas se generaron alrededor de 57 millones de fragmentos de ADNc de *S. rebaudiana* y aproximadamente 49 millones de *T. Patula*, que fueron analizados con las herramientas FASTQC, FastXTool kit y RobiNA.

En los últimos años se ha incrementado la secuenciación de ARN de especies no modelo, por lo que no existen genomas de referencia previamente descritos contra las cuales pueda realizarse una alineación de fragmentos, como sucede con *S. rebaudiana* y *T. patula*. Una de las estrategias propuestas para solventar este problema, es la comparación de los fragmentos o los transcriptomas ensamblados *de novo*, contra organismos modelo filogenéticamente cercanos (Hornett y Wheat, 2012).

La tecnología más empleada actualmente para la secuenciación es la plataforma Illumina (*Genome Analyser*) gracias a los excelentes resultados de calidad que genera, el relativo bajo costo por librería de ADNc y la posibilidad de identificación de SNP's (Hornett y Wheat, 2012; Lin, 2012); razones por las cuales ésta fue la plataforma seleccionada para las secuenciaciones.

Las secuenciaciones de las muestras generaron fragmentos de 101 bases, pero previamente, el ARN debió ser sometido a una evaluación de su integridad, para asegurar la calidad de secuenciación; en este proceso se empleó como referente el calor de RIN, el cual se ubica usualmente en una escala del 1 al 10, siendo 10 el valor máximo de calidad.

Las librerías que contienen los fragmentos de ADNc pueden ser single-end o paired-end, éstas últimas contemplan la secuenciación en ambas direcciones del ADNc, por lo que en algún ciclo se generarán traslapes de la información, obteniéndose así más información para un ensamblaje *de novo* (Lin, 2012). Este tipo de librería es la más recomendable para especies no modelo que carecen de un genoma descrito, razón por la cual fue la opción seleccionada para el almacenamiento de la información de transcripción de *Stevia* y *Tagetes*.

Análisis de calidad de la secuenciación

La calidad de los fragmentos generados con una secuenciación, es un valor determinado por la probabilidad de que la base indicada sea incorrecta. El porcentaje Sanger de calidad es la medida estándar para análisis biológicos; los valores ASCII que lo componen se ubican en el rango de 33 a 73, y otorgan una calidad máxima a los fragmentos de 40% y una mínima de 0% (Fass, 2012).

La literatura indica que el procesamiento recomendable de la información de un transcriptoma inicia con el diagnóstico de la calidad de los fragmentos, a los cuales debe eliminárseles los adaptadores, ausentes en este caso para la información generada por la empresa AXEQ (Britton, 2013).

Posteriormente, se procedió con los cortes tipo *clipping* y los *trimming* (Britton, 2012), tanto de la base 1 a la 15 al inicio de todos los fragmentos (*clipping*), como a partir de aquellas bases que hacia el final de los fragmentos tuvieran una calidad inferior al 30% (*trimming*). La eliminación de fragmentos con una longitud inferior a 70 bases se realizó como un paso extra para evitar la adhesión de fragmentos muy pequeños que pudieran generar falsos positivos en el alineamiento. La eliminación de fracciones de las secuencias con baja calidad, limita la contaminación de la información e incrementa la veracidad de los resultados de los alineamientos (Lin, 2012).

En general, los fragmentos de un transcriptoma se procesan para comparar sólo aquellos que cumplan con una calidad específica y/o una longitud determinada. Así que una vez que se ha depurado la información, debe evaluarse nuevamente la calidad, antes de proceder con los alineamientos o mapeos (Britton, 2012).

Identificación de secuencias génicas reportadas

Análisis de regiones promotoras, elementos en *cis* y factores de transcripción en *S. rebaudiana*

En total se identificaron 21 genes de *S. rebaudiana*, la mayoría de los cuales pudieron ser relacionados con la ruta metabólica de síntesis de glicósidos, de acuerdo con la función descrita en la base de datos del NCBI así como aquella descrita para algunos de éstos en la investigación que Kumar y colaboradores (2012) han desarrollado, documentando comportamientos fisiológicos y describiendo genes para esta especie.

Entre los genes descritos, se encontró la secuencia codificante para β -actina en *Stevia*, el cual se considera como un gen constitutivo de uso común como control interno de la expresión génica. Este tipo de genes se emplea en técnicas como microarreglos y qPCR como referentes para la cuantificación relativa de la expresión de los genes de interés, cuya expresión se está analizando en una determinada muestra. (Van Pelt-Verkuil, 2008).

De los 21 genes de estevia descritos en las bases de datos, únicamente de 5 de ellos se conocía la región promotora, la cual fue analizada con la herramienta bioinformática PLACE, para determinar los posibles sitios de anclaje a ADN por parte de proteínas de regulación. Esta herramienta fue utilizada en 2013 (Kumar y Kumar, 2013) para determinar los sitios en *cis* correspondientes a las cajas TATA y GATA, con la finalidad de evaluar la producción de glicósidos ante estímulos lumínicos en *Stevia*.

En relación con las regiones promotoras descritas en *S. rebaudiana*, la herramienta de análisis seleccionada generó predicciones acerca de la posible ubicación de las cajas TATA, GATA y CAAT de acuerdo con los genes descritos para especies como *Nicotiana tabacum*, *Oriza sativa*, *Pisum sativum*, *Phaseolus vulgaris*. Además también fue posible señalar la ubicación de posibles sitios de anclaje a otros factores de regulación como los factores de transcripción.

La selección de los sitios de interés para continuar con el mapeo de factores de transcripción contra el transcriptoma de *Stevia*, se realizó mediante una discriminación de acuerdo con su descripción para la especie modelo *A. thaliana*, específicamente para tejidos de hoja de esta especie; lo anterior se debe justamente a que la complejidad de las interacciones que ocurren entre el ADN y las proteínas reguladoras de la transcripción, depende de los diferentes tejidos del organismo, sus tipos celulares, estadios de desarrollo, estados fisiológicos y condiciones ambientales y/o hormonales (Carlberg y Molár, 2014; Ceunen y Geuns, 2013; Ceunen *et al.*, 2012; Latchman, 2003; Zhang, 2003).

Una vez seleccionados los sitios de unión al ADN que fueron predichos por el programa PLACE en las regiones promotoras de *Stevia* se recurrió al uso de la herramienta INTERACTOME de la Universidad de Harvard, la cual es sólo una de la gran cantidad de opciones que se han creado para este fin. Un interactoma es el complejo de trabajo entre redes de proteínas que se encargan de regular la expresión y metabolismo celular. Las interacciones proteína-proteína son un elemento fundamental de los sistemas biológicos, y su análisis puede proporcionar pistas valiosas acerca de las funciones una proteína desconocida (Broun, 2004).

En este estudio, una vez establecida la función concreta de los factores de regulación seleccionados de *A. thaliana*, se determinó que solamente tres de ellos correspondieron a factores de transcripción propiamente dichos, estos fueron proteínas con dominios MYC y MYB. Las proteínas MYB se relacionan con la respuesta de estrés de las plantas, como exposición a luz ultra-violeta, heridas, estrés anaeróbico y entrada de patógenos (Singh *et al.*, 2002). Las proteínas MYC se relacionan con la diferenciación de las células epidérmicas y patrones de tricomas actuando como cofactores con otras proteínas (Schwechheimer *et al.*, 1998).

Selección de factores de transcripción para *T. patula*

Una vez revisada la base de datos del NCBI, se encontró que existe un factor de transcripción de la familia APETALA descrito para *T. patula*; además, se han descrito dos genes de forma parcial y dos genes completos para esta misma especie, ninguno relacionado directamente con la síntesis de tiofenos.

También se trabajaron varios genes de las familias de factores de transcripción MYC y MYB reportados en especies de Asteraceas, principalmente porque durante los análisis en *S. rebaudiana*, se determinó que tuvieron una mayor incidencia en la síntesis de glicósidos; por lo que si se encontraban presentes en *T. patula* y eventualmente se podría ligarlos con la síntesis de tiofenos.

Alineamiento de genes contra el transcriptoma

Los niveles de expresión génica para cada unidad de ARN pueden aproximarse de acuerdo con el número de fragmentos secuenciados que se alinean con un genoma de referencia, relacionándolos directamente con niveles de abundancia. Para generar este tipo de comparaciones existen herramientas de fácil uso y libre acceso como Cuffdiff (Trapnell *et al.*, 2010), edgeR (Robinson, 2010), DESeq (Anders y Huber, 2010), PoissonSeq (Li *et al.*, 2012), baySeq (Hardcastle y Kelly, 2010) y limma. (Smyth, 2004).

El alineamiento es entonces el primer paso para la comparación y cuantificación de la expresión, y justamente muchas veces es el más lento, dada la dificultad computacional de determinar cuál es el punto de origen de un fragmento en el genoma; una de las más eficientes en la actualidad es el programa Bowtie, el cuál es rápido y bastante preciso y además permite encontrar zonas de no alineamiento o gaps, los cuales pueden relacionarse con inserciones o deleciones en un gen (Fass, 2012; Langmead y Salzberg, 2012).

En el análisis de alineación entre genes seleccionados se empleó la herramienta Bowtie. Los mapeos no se realizaron contra un genoma de referencia porque este no ha sido reportado para las especies en estudio.

Es importante recalcar que en un análisis de expresión en eucariotas se recomienda contar con más de 30 millones de fragmentos por muestra biológica secuenciada (Britton, 2013), por lo tanto, fue válido en este estudio realizar los mapeos entre los fragmentos del transcriptoma y los genes seleccionados.

En el caso de *S. rebaudiana*, todos los genes de ruta metabólica de síntesis de glicósidos estaban presentes en la muestra de ARN secuenciada; para *T. patula* se obtuvo que los genes reportados para esta especie en el NCBI, contando un factor de transcripción, también se encontraron presentes en el material biológico trabajado.

Se considera que un mapeo es válido si se genera más de un 80% de alineamiento contra un transcriptoma, y que, con menos de un 50% cuestiona si los fragmentos representan un transcriptoma incompleto o son contaminación de la muestra (Lin, 2013), esto aplica únicamente cuando el mapeo se realiza contra la totalidad de un genoma, pero en este caso no, porque se trabajó con genes de entre 2000 y 3000 pb; por tanto, los bajos porcentajes de alineación en ambos casos son esperables y justificados.

Es decir, el resultado obtenido en esta sección fue positivo para la comprobación de compuestos bioactivos (glicósidos) en la muestra de hoja trabajada. Ahora, basándose en el principio de que la expresión de un gen puede relacionarse con estos datos de alineación obtenidos, es importante notar que algunos de los genes seleccionados se encontraron menos representados en el transcriptoma, lo cual puede ser un indicador de lo que algunos autores describen como la producción controlada de glicósidos dada por ciertos genes.

Con respecto a lo expuesto con anterioridad, los genes 14, 15, 16, al alinear con un menor número de fragmentos, podría considerarse que su expresión en la muestra era menor en comparación con los demás genes; esto concuerda con los estudios correlativos entre las enzimas de ruta, que han puesto de manifiesto la importancia de los genes denominados como SrCPPS (13), SrKS (14 y 15), SrKO en la regulación del contenido de glicósidos (Kumar *et al.*, 2012). En esta investigación, el paso de copalil difosfato a kaureno, de kaureno a ácido kaureónico y de este último a esteviol, fueron los pasos más regulados.

En cuanto a la falta de alineamientos entre el transcriptoma de estevia y los factores de transcripción seleccionados se pueden desprender varias hipótesis. Una de ellas es que sólo alrededor del 7% del total de factores descritos de *Arabidopsis* sp., han sido genética y funcionalmente caracterizados, por lo que las regiones promotoras de los genes de estevia podrían estar interactuando con proteínas aún no caracterizadas para la especie modelo *Arabidopsis* o bien, los genes de *Stevia* poseen más de una región promotora de la transcripción que aún no se conoce.

La segunda hipótesis indica que la técnica de secuenciación requiere de varios ciclos, así como de la generación de la mayor cantidad de fragmentos posibles, aunque provengan de una única muestra, dado que algunos transcritos como los factores de transcripción, tienen un bajo porcentaje de

cobertura, es decir, se expresan en muy bajas cantidades, así que su representación dentro de un transcriptoma es muy baja, en comparación con aquellos transcritos que son expresados en grandes cantidades dentro de la muestra (Grabherr *et al.*, 2011). Por tanto, es probable también que las secuencias de este tipo de gen se perdieran por la sobreexpresión de otros genes.

En el caso de *Tagetes patula*, se obtuvo un resultado favorable para la identificación de un factor de transcripción cuya función no se define satisfactoriamente en las bases de datos; resultado favorable también se dio con el alineamiento de los genes 2 y 3 (cuadro 14) de esta especie, ya que sirvieron como control positivo para la detección de material biológico de la especie en estudio. Para efectos de esta investigación no se realizaron alineamientos de genes con secuencias codificantes parcialmente descritas.

Por otra parte, en cuanto a los alineamientos contra factores de transcripción descritos en especies de la familia Asteraceae, los resultados fueron negativos para la identificación de alguno de ellos dentro del transcriptoma de *T. patula* y las razones por las cuales esto sucedió son las mismas descritas previamente para los alineamientos entre *Stevia* y *A. thaliana*.

En futuros análisis se debe contemplar que la gran mayoría de estrategias implementadas para el análisis de la expresión diferencial de los datos de secuenciación de ARN exigen diseños experimentales con replicación de datos. Sin embargo, a pesar del abaratamiento de los costos de secuenciación, el generar muchas repeticiones de información a partir de réplicas de muestras eleva los costos, al punto que los pequeños proyectos con escaso financiamiento no pueden pagarlo. La carencia de replicadas disponibles, hace que la estimación de la prueba de hipótesis de la expresión diferencial no sea confiable (Rapaport *et al.*, 2013).

RECOMENDACIONES

Un futuro análisis de expresión génica mediante secuenciación de ADNc, sería más preciso mediante el ensamblaje de un transcriptoma *de novo* proveniente de una secuenciación completa de al menos 3 repeticiones de ARN proveniente de agrupaciones de muestras biológicas (*pool de genes*), con las cuales generar al menos 3 terabytes de información en cada repetición.

Pueden realizarse ensayos previos de expresión de genes de interés mediante la técnica de qPCR proponiendo un diseño experimental sólido que facilite la evaluación del cambio en la respuesta de la planta en términos de síntesis de esteviósidos, de acuerdo con diferentes condiciones ambientales.

AGRADECIMIENTOS

A la investigadora Ann Powell y la pasante Bárbara Blanco de la Universidad de California en Davis por el apoyo científico-técnico suministrado durante la realización de esta investigación.

Al programa E-Science, la Escuela de Computación y el Centro de Investigación en Computación del Instituto Tecnológico de Costa Rica por el apoyo tecnológico e informático suministrado para la conclusión del proyecto.

REFERENCIAS

- Ahmed, M.; Salahin, M.; Karim, R.; Razvy, M.; Hannan, M.; Sultana, R.; Hossain, M.; Islam, R. 2007. An efficient method or in vitro clonal propagation of a newly introduced sweetener plant (*Stevia rebaudiana* Bertoni) in Bangladesh. *American-Eurasian Journal of Scientific Research* 2 (2): 121-125.
- Alvarenga, S. 2005. Optimización del cultivo y procesamiento de *Stevia rebaudiana* para la obtención de un edulcorante natural. CONICIT. San José, Costa Rica. En: http://www.conicit.go.cr/boletin/boletin67/cultivo_salvarenga.html
- Anders, S.; Huber, W. 2010. Differential expression analysis for sequence count data. *Genome Biol*, 11(10): R106.
- Bourlakis, M.; Weightman, P. 2006. *Food Supply Chain Management*. 1era Edición. Estados Unidos: Blackwell Publishing. 413 p.
- Britton, M. 2012. Construction of Genes and Transcripts. Consultado en: http://training.bioinformatics.ucdavis.edu/docs/2012/05/RNA/_downloads/gene-construction.pdf
- Britton, M. 2013. Introduction to RNA-Seq. Consultado en: http://training.bioinformatics.ucdavis.edu/docs/2013/06/rnaseq/_downloads/MB_RNASeq_BC_Lecture_2_013.pdf
- Broun, P. 2004. Transcription factors as tools for metabolic engineering in plants. *Current Opinion in Plant Biology*, 7: 202-209.
- Carlberg, C.; Molnár, F. 2014. *Mechanisms of Gene Regulation*. Londres, Inglaterra: Springer Science 210 p.
- Ceunen, S.; Geuns, J. 2013. Spatio-temporal variation of the diterpene steviol in *Stevia rebaudiana* grown under different photoperiods. *Phytochemistry*, 89: 32-38.
- Ceunen, S.; Werbrouck, S.; Geuns, J. 2012. Stimulation of steviol glycoside accumulation in *Stevia rebaudiana* by red LED light. *Journal of Plant Physiology*, 169: 749-752.
- Chaturvedi, N.; Sharma, P.; Shukla, K. 2014. Pharmacological Facets of *Stevia rebaudiana* (Bert.) and its Secondary Metabolites. *Journal of Functional And Environmental Botany*, 4(1): 1-9.
- Crispin, M.; Wurtele, E. 2013. Use of Metabolomic and Transcriptomics to Gain Insights into the Regulation and Biosynthesis of Medicinal Compounds: *Hypericum* as a Model. En Chandra *et al.*, *Biotechnology for Medicinal Plants: Micropropagation and Improvement*. Berlín: Springer. 462 p.
- Das, A.; Gantait, S.; Mandal, M. 2011. Micropropagation of an Elite Medicinal Plant: *Stevia rebaudiana* Bert. *International Journal of Agricultural Research*. Mohanpur, India. 6 (1): 40-48
- Djendoel, D. 2002. *Botany of Stevia and Stevia rebaudiana*. *Stevia: The genus Stevia*. 1era Edición. Estados Unidos, Taylor and Fransis. p. 18-39.
- Einbond, L.; WuH, S.; Chang, T.; Panjikaran, M.; Wang, X.; Goldberry, S. 2010. Digitoxin activates EGR1 and synergizes with paclitaxel on human breast cancer cells. *J Carcinog*, 9:10–23.
- Farrell, R. 2010. *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*. 3era Edición. Londres, Inglaterra: Elsevier Academic Press 688 p.
- Fass, J. 2012. Read Alignment for RNA.Seq. Consultado en: http://training.bioinformatics.ucdavis.edu/docs/2012/05/RNA/_downloads/Alignment.pdf
- Flores, M; Cortés, R. 2011. *Stevia rebaudiana*: Propuesta para la Producción de Esteviósido en Biorreactor. ITCR. Cartago, Costa Rica. p 1-7.

- Gómez, O.; Zavaleta, E. 2001. La asociación de cultivos: Una estrategia más para el manejo de enfermedades, en particular *Tagetes sp.* *Revista Mexicana de Fitopatología*, 9 (1): 94-99.
- Goncalves, C.; Dinis, T.; Batista, M. 2005. Antioxidant properties of proanthocyanidins of *Uncaria tomentosa* bark decoction: a mechanism for anti-inflammatory activity. *Phytochemistry* 66: 89-98.
- Grabherr *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*: 29(7): 644-653.
- Hardcastle, T.; Kelly, K. 2010. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, 11:422.
- Higo, K.; Ugawa, Y.; Iwamoto, M.; Korenaga, T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.* 27 (1): 297-300.
- Hornett, E.; Wheat, C. 2012. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC Genomics*, 13: 361-377.
- Humphrey, T.; Richman, A.; Menassa, R.; Brandle, J. 2006. Spatial organization of four enzymes from *Stevia rebaudiana* that are involved in steviol glycoside synthesis. *Plant Molecular Biology*. Ontario, Canadá. 61: 47 – 62.
- Karp, G. 2009. Cell and Molecular Biology: Concepts and Experiments. 6ta Edición. Estados Unidos: Wiley. 832 p.
- Karppinen, K.; György Z.; Kauppinen, M.; Tolonen, A.; Jalonen, J.; Neubauer, P.; Hohtola, A.; Häggman, H. 2006. In vitro propagation of *Hypericum perforatum* L. and accumulation of hypericins pseudo hypericins, and phloroglucinols. *Propag Ornament Plants*, 6(4):170–179.
- Kumar, H.; Kaul, K.; Bajpai-Gupta, S.; Kumar, V.; Kumar, J. 2012. A comprehensive analysis of fifteen genes of steviol glycosides biosynthesis pathway in *Stevia rebaudiana* (Bertoni). *Gene*, 492: 276-284.
- Kumar, H.; Kumar, S. 2013. A functional (E)-4-hydroxy-3-methylbut-2-enyldiphosphate reductase exhibits diurnal regulation of expression in *Stevia rebaudiana* (Bertoni). *Gene* 527: 332–338.
- Langmead, B.; Salzberg, S. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4): 357-360.
- Latchman, D. 2003. *Eucaryotic Transcription Factors*. 4ta Edición. San Diego: Academic Press. 620 p.
- Li, J.; Witten, D.; Johnstone, I.; Tibshirani, R. 2012. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics*, 13(3):523-538.
- Liao, Z.; Chen, M.; Tan, F.; Sun, X.; Tang, K. 2004. Micropropagation of endangered Chinese aloe. *Plant Cell Tissue Org*, 76(1):83–86.
- Lin, 2012. NGS Impacts and Key Concepts. Consultado en: http://training.bioinformatics.ucdavis.edu/docs/2012/05/RNA/_downloads/RNAIntro.pdf
- Lohse, M.; Nunes-Nesi, A.; Krueger, P.; Nagel, A.; Hannemann, J.; Giorgi, F.; Childs, L.; Osorio, S.; Walther, D.; Selbig, J.; Sreenivasulu, N.; Stitt, M.; Fernie, A.; Usadel, B. 2010. Robin: An intuitive wizard application for R-based expression microarray quality assessment and analysis. *Plant Physiology*, 153: 642-51.
- López, G. 2007. Detección molecular de *Helicobacter pylori*. Informe de Trabajo Final de Graduación para optar por el grado de Bachiller en Ingeniería en Biotecnología. Instituto Tecnológico de Costa Rica, Cartago.
- Martínez, T. 2002. *La Hierba Dulce: Historia, Usos y Cultivo de la Stevia rebaudiana Bertoni*. 1era Edición. España, Colección Ciencias de la Salud. sp.

- Mbakwe, R.; Harry-Asobara, L. 2014. Efficacy of Extracts from *Jatropha curcas* (L.) and *Azadirachta indica* (A. Juss) on *Alstonia boonei* (De Wild) attacked by *Anobium punctatum* (De Geer). *Annals of Biological Research*, 5(5): 99-105.
- Medappa N. 2003. Prospects of using herbal products in the control of mosquito vectors. *Indian Council of Medical Research*. 33 (1): 1-10.
- Mohamed, A.; Ceunen, S.; Geuns, J.; Van den Ende, W.; De Ley, M. 2011. UDP-dependent glycosyl transferases involved in the biosynthesis of steviol glycosides. *Journal of Plant Physiology*, 168: 1136-1141.
- Murata, J.; Roepke, J.; Gordon, H.; De Luca, V. 2008. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell*, 20:524–542.
- Perry, N.; Van Klink, J.; Burgess, E.; Parmenter, G. 1997. Alkamide levels in *Echinacea purpurea*: a rapid analytical method revealing differences among roots, rhizomes, stems, leaves and flowers. *Planta Med*, 63(1):58–62.
- Prabhat, K.; Huang, L.; Srinivasan. 2014. Effect of three commercial biopesticides of neem (*Azadirachta indica*) and *Bacillus thuringiensis* on legume pod borer (*Maruca vitrata*) (Lepidoptera: Crambidae) in Thailand. *Internatinal journal of Tropical Insect Science*. 34(2): 80-87.
- Pupo, Y.; Herrera, L.; Vargas, B.; Marrero, Y.; Arévalo, R.; Jiménez, M. 2007. Efecto del extracto crudo de hojas de *Tagetes erecta* L. en el control de cuatro hongos patógenos de hortalizas *in vitro*. *Centro Agrícola*, 34 (4): 83-86.
- Rafiq, M.; Dahor, M; Mangrio, S.; Naqvi, H.; Qarshi, I. 2007. In vitro clonal propagation and biochemical analysis of field established *Stevia rebaudiana* Bertoni. *Pak. J. Bot.* 39 (7): 2476-2474.
- Rapaport, F.; Khanin, R.; Liang, Y.; Krek, A.; Zumbo, P.; Mason, C.; Socci, N.; Betel, D. 2013. Comprehensive evaluation of differential expression analysis methods for RNA-seq data. *Genome Biology*, 14: 2-21.
- Richman, A.; Gijzen, M.; Starratt, A.; Yang, Z.; Brandle, J. 1999. Diterpene synthesis in *Stevia rebaudiana*: recruitment and up-regulation of key enzymes from the gibberellin biosynthetic pathway. *The Plant Journal*, 19(4): 411-421.
- Richman, A.; Swanson, A.; Humphrey, T.; Chapman, R.; McGarvey, B.; Pocs, R.; Brandle, J. 2005. Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. *The Plant Journal*, 41: 56-67.
- Robinson, M.; McCarthy, D.; Smyth, G. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26:139-140.
- Schwechheimer, C.; Zourelidou, M.; Bevan, M. 1998. Plant Transcription Factor Studies. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 49: 127-150.
- Sharipova, R.; Strobukina, Y.; Mordovskoi, G.; Chestnova, R.; Mironov, F.; Kataev, V. 2011. Antituberculosis activity of glycosides from *Stevia rebaudiana* and hybrid compounds of steviolbioside and pyridinecarboxylic acid hydrazides. *Chemistry of Natural Compounds* 46 (6): 902-905.
- Shreeder, R.; Venkatachalam, L.; Thimmaraju, N.; Bhagyalakshmi, N.; Narayan, M.; Ravishankar, G. 2008. Direct organogenesis from leaf explants of *Stevia rebaudiana* and cultivation in bioreactor. *Biologia Plantarum*, 52 (2): 355-360.
- Singh, K.; Foley, R.; Oñate, L. 2002. Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology*, 5: 430-436.

- Sitakanta, P.; Claire, X.; Kong, Q.; Shen, K.; Yuan, L. 2006. Directed evolution of plant basic helix–loop–helix transcription factors for the improvement of transactivational properties. *Biochimica et Biophysica, Acta* 1759: 308-318.
- Smyth, K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, 3: Article 3.
- Soneson, C.; Delorenzi, M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14: 91-109.
- Spainhour Ch. 2005. *Natural Products*. Drug Discovery Handbook. 1era Edición. Wiley-Interscience.
- Tachjian, A.; Jahangir, A. 2010. Use of herbal products and potential interactions in patients with cardiovascular diseases. *J Am Coll Cardiol*, 55(6):515–525.
- Trapnell, C.; Williams, B.; Pertea, G.; Mortazavi, A.; Kwan G.; van Baren, M.; Salzberg S.; Wold, B.; Pachter, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoforms switching during cell differentiation. *Nat Biotechnol*, 28(5):511-515.
- Unión Vegetariana Argentina [U.V.A.]. 2003. *Stevia o yerba dulce*. Argentina: U.V.A. En: <http://www.uva.org.ar/stevia.html>
- Van Pelt-Verkuil, E.; Belkum, A.; Hays, J. 2008, *Principles and Technical Aspects of PCR Amplification*. Países Bajos: Springer Science. 342 p.
- Varshneya C. Telang, S. 2006. Utilization of plant biodiversity of North-Western Himalayan region of India for the development of herbal anthelmintics. *Journal of Veterinary Pharmacology and Therapeutics*, 29(1): 206-206.
- Vidal, J.; Carbajal, A.; Sisniegas, M.; Bobadilla, M. 2008. Efecto tóxico de *Argemone subfusiformis* Ownb. y *Tagetes patula* sobre larvas del IV estadio y pupas de *Aedes aegypti* L. *Revista Peruana de Biología*, 5 (2): 103-110.
- Vu, T.; Tran, T.; Biteau, F.; Mignard, B.; Fevre, J.; Guckert, A.; Bourgaud, F.; Gontier, E. 2006. Improvement of secondary metabolites production in hydroponic cultures by mechanical and biological processes. Proceedings of International Workshop on Biotechnology in Agriculture. Nong Lam University Ho Chi Minh City.
- Wilfinger, W.; Mackey, K.; Chomcynski, P. 1997. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *Biotechniques*, 22(3): 474-481.
- Yang, C.; Chen, M.; Zeng, L.; Zhang, L.; Liu, X.; Lan, X.; Tang, K.; Liao, Z. 2011. Improvement of tropane alkaloids production in hairyroot cultures of *Atropa belladonna* by overexpressing pmt and h6h genes. *Plant Omics*, 4(1):29–33.
- Zhang, J. 2003. Over expression analysis of plant transcription factors. *Current Opinion in Plant Biology*, 6: 430-440.

ANEXOS

Anexo 1. Total RNA Isolation from Plant Protocol

5 Protocols

5.1 RNA isolation from plant tissue or filamentous fungi

Before starting the preparation:

- Check if Wash Buffer RA3 and rDNase were prepared according to section 3.

1 Homogenize sample

Grind up to **100 mg tissue** under liquid N₂ (for handling and preparation methods see section 2.3).



Grind sample

2 Lyse cells

Add **350 µL Buffer RA1** and **3.5 µL β-mercaptoethanol (β-ME)** to 100 mg tissue and vortex vigorously.

If the lysate solidifies upon addition of Buffer RA1, use **350 µL Buffer RAP** instead.

Note: As alternative to β-ME the reducing agent DTT or TCEP may be used. Use a final concentration of 10–20 mM DTT or TCEP within the Lysis Buffer RA1 or RAP (e.g., add 7–14 µL of a 500 mM DTT or TCEP solution).



+ 350 µL RA1
+ 3.5 µL β-ME

or

+ 350 µL RAP
+ 3.5 µL β-ME

3 Filtrate lysate

Reduce viscosity and clear the lysate by filtration through **NucleoSpin® Filter (violet ring)**: Place NucleoSpin® Filter in a Collection Tube (2 mL), apply the mixture, and centrifuge for **1 min at 11,000 x g**. Transfer the filtrate to a new 1.5 mL microcentrifuge tube (not provided).

Important note: Do not disturb the pellet of cell debris at the bottom of the collecting tube, which may be visible after centrifugation.

In case of visible pellet formation (depending on sample amount and nature) transfer supernatant without any formed pellet to a new 1.5 mL microcentrifuge tube (not provided).



11,000 x g,
1 min

4 Adjust RNA binding conditions

Discard the NucleoSpin® Filter and add **350 µL ethanol (70 %)** to the homogenized lysate and mix by pipetting up and down (5 times).

Alternatively, transfer flow-through into a new 1.5 mL microcentrifuge tube (not provided), add **350 µL ethanol (70 %)**, and mix by vortexing (2 x 5 s).

After addition of ethanol a stringy precipitate may become visible which will not affect the RNA isolation. Be sure to disaggregate any precipitate by mixing and load all of the precipitate on the column as described in step 5. Do not centrifuge the ethanolic lysate before loading it onto the column in order to avoid pelleting the precipitate.



**+ 350 µL
70 % ethanol**

Mix

5 Bind DNA

For each preparation take one **NucleoSpin® RNA Plant Column (light blue ring)** placed in a Collection Tube and load the lysate. Centrifuge for **30 s** at **11,000 x g**. Place the column in a new Collection Tube (2 mL).

Maximum loading capacity of NucleoSpin® RNA Plant Columns is 750 µL. Repeat the procedure if larger volumes are to be processed.



Load lysate



**11,000 x g,
30 s**

6 Desalt silica membrane

Add **350 µL MDB** (Membrane Desalting Buffer) and centrifuge at **11,000 x g** for **1 min** to dry the membrane.

Salt removal will make the following rDNase digest much more effective. If the column outlet has come into contact with the flow-through for any reason, discard the flow-through and centrifuge again for 30 s at 11,000 x g.



+ 350 µL MDB



**11,000 x g,
1 min**

7 Digest DNA

Prepare **DNase reaction mixture** in a sterile 1.5 mL microcentrifuge tube (not provided): For each isolation, add **10 µL reconstituted rDNase** (see section 3) to **90 µL Reaction Buffer for rDNase**. Mix by flicking the tube.



**+ 95 µL
rDNase
reaction
mixture**

Apply **95 µL DNase reaction mixture** directly onto the center of the silica membrane of the column. Incubate at **room temperature for 15 min**.

RT, 15 min

8 Wash and dry silica membrane**1st wash**

Add **200 µL Buffer RAW2** to the NucleoSpin® RNA Plant Column. Centrifuge for **30 s** at **11,000 x g**. Place the column into a new Collection Tube (2 mL).

Buffer RAW2 will inactivate the rDNase.

+ 200 µL RAW2

**11,000 x g,
30 s**

2nd wash

Add **600 µL Buffer RA3** to the NucleoSpin® RNA Plant Column. Centrifuge for **30 s** at **11,000 x g**. Discard flow-through and place the column back into the Collection Tube.

Note: Make sure that residual buffer from the previous steps is washed away with Buffer RA3, especially if the lysate has been in contact with the inner rim of the column during loading of the lysate onto the column. For efficient washing of the inner rim flush it with Buffer RA3.



+ 600 µL RA3

**11,000 x g,
30 s**



+ 250 µL RA3

3rd wash

Add **250 µL Buffer RA3** to the NucleoSpin® RNA Plant Column. Centrifuge for **2 min** at **11,000 x g** to dry the membrane completely. Place the column into a nuclease-free Collection Tube (1.5 mL, supplied).

If for any reason, the liquid level in the Collection Tube has reached the NucleoSpin® RNA Plant Column after centrifugation, discard flow-through, and centrifuge again.

**11,000 x g,
2 min**

9 Elute RNA

Elute the RNA in **60 µL RNase-free H₂O**, (supplied) and centrifuge at **11,000 x g** for **1 min**.

If higher RNA concentrations are desired, elution can be done with 40 µL. Overall yield, however, will decrease when using smaller elution volumes.

For alternative elution procedures see section 2.4.



**+ 60 µL
RNase-free
H₂O**

**11,000 x g,
1 min**