



ARTÍCULO DE INVESTIGACIÓN/RESEARCH ARTICLE

ANÁLISIS *in silico* Y EXPRESIÓN GÉNICA DEL FACTOR DE TRANSCRIPCIÓN *TgNAC01* IMPLICADO EN XILOGÉNESIS Y ESTRÉS ABIÓTICO EN *Tectona grandis*.***In silico* Analysis and Gene Expression of *TgNAC01* Transcription Factor Involved in Xylogenesis and Abiotic Stress in *Tectona grandis*.**Vladimir CAMEL PAUCAR¹, Esteban GALEANO², Helaine CARRER².¹ Laboratorio de Biotecnología y Biología Molecular, Universidad Continental. Huancayo, Perú.² Departamento Ciências Biológicas, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo. São Paulo, Brasil.

For correspondence. vlad_camel@hotmail.com

Received: 24th January 2017, Returned for revision: 30th April 2017, Accepted: 15th August 2017.

Associate Editor: Geraldo Mäder.

Citation/Citar este artículo como: Camel Paucar V, Galeano E, Carrer H. Análisis *in silico* y expresión génica del factor de transcripción *TgNAC01* implicado en xilogénesis y estrés abiótico en *Tectona grandis*. Acta biol. Colomb. 2017;22(3):359-369. DOI: <http://dx.doi.org/10.15446/abc.v22n3.62164>**RESUMEN**

El xilema secundario es el componente más abundante de la biomasa vegetal. Por tanto, conocer los genes que regulan su formación ayudaría a diseñar estrategias para el mejoramiento genético de la madera. Así, el objetivo de este trabajo fue realizar el análisis computacional de la estructura primaria y secundaria del factor de transcripción (FT) *TgNAC01* de *Tectona grandis*, además de evaluar su historia evolutiva, dominios conservados y expresión génica en tejidos lignificados de árboles de 12 y 60 años. Para ello, se realizó una evaluación del potencial de interacción ion-electrón (PIIE), mediante el método del espectro de la información (MEI) utilizando la librería SFAPS de R-Project, seguido del modelamiento estructural utilizando el software MODELLER y visualizado mediante PyMol. Además, el análisis de alineamiento de secuencia múltiple y filogenia fue mediante el software Bioedit y MrBayes respectivamente. También se evaluó los niveles de síntesis del FT *TgNAC01* mediante qRT-PCR. Como resultados, se evidenció que el FT mantiene una estructura β -hoja antiparalela retorcida, que se compacta contra una α -hélice en la región N-terminal, teniendo así tres dominios α hélice y siete dominios β plegada. Asimismo, mediante el MEI se demostró que tiene alrededor de cinco funciones biológicas y mutaciones sobre los aminoácidos con mayor PIIE, lo que conlleva a evoluciones sobre las redes de regulación genética. Finalmente, el FT *TgNAC01* podría presentar un papel fundamental en la organización y desarrollo de las partes que componen la albura, como las células radiales de la zona cambial, los vasos, fibras y los anillos de crecimiento.

Palabras clave: estructura proteica, expresión génica, factor de transcripción, método de espectro de la información, xilema secundario.**ABSTRACT**

Secondary xylem is the most abundant component of plant biomass. Therefore, knowing the genes that regulate its formation would help to design strategies for wood genetic improvement. Thus, the objective of this work was to perform computational analysis of the primary and secondary structure of the *TgNAC01* transcription factor (FT) of *Tectona grandis*, and to evaluate its evolutionary history, conserved domains and gene expression in lignified tissues of trees with 12 and 60 years old. For this, an ion-electron interaction potential (IEP) was evaluated using the information-spectrum method (IEM) using the R-Project and SFAPS library, followed by structural modeling using the MODELLER software and visualized by PyMol program. In addition, the analysis of multiple sequence alignment and phylogeny was performed using Bioedit and MrBayes software, respectively. We also evaluated the qRT-PCR levels of *TgNAC01*. As results, it was found that *TgNAC01* maintains a twisted antiparallel β -sheet structure, which is compacted against an α -helix in the N-terminal region, having three α -helix domains and seven folded β -domains. Also, through the IEM, it was demonstrated that it has about five biological functions, and mutations on amino acids with higher IEP, which leads to evolutions on genetic regulation networks. Finally, the FT *TgNAC01* could play an essential role in the organization and development of the parts that make up the sapwood, such as the radial cells of the cambial zone, the vessels, fibers and the growth rings.

Keywords: gene expression, method of information spectrum, protein structure, secondary xylem, transcription factor.

INTRODUCCIÓN

Tectona grandis (teca) es un árbol tropical de importancia económica, ecológica y ambiental, debido a su madera fina, rápido crecimiento y reducción de emisiones de CO₂, siendo la tecla una de las especies forestales más plantada a nivel mundial (Nunifu y Murchison, 1999; White, 2016). Conocer los mecanismos de biosíntesis de su madera es un proceso molecular complejo que involucra a diversos genes y factores ambientales como sequías, calor, inundaciones, frío, sales y radiación, que influyen en la formación del xilema secundario (Zhong *et al.*, 2011; Lin *et al.*, 2014;). Por otro lado, investigaciones a nivel genómico, transcriptómico y proteómico en *T. grandis* buscan optimizar el incremento de la biomasa y la calidad de su madera (Diningrat *et al.*, 2015; Galeano *et al.*, 2015). Es por ello que se van desarrollando estudios genéticos en diferentes tejidos y etapas del desarrollo vegetativo y generativo, mostrando posibles vías transcripcionales implicadas en la xilogénesis en Teca (Galeano *et al.*, 2014; Mustari *et al.*, 2016; Camel *et al.*, 2017), información de importancia desde las perspectivas ecológicas, evolutivas e industriales (Camel *et al.*, 2017).

Consecuentemente, estudios realizados con RNAseq en *T. grandis* proponen familias de factores de transcripción diferencialmente expresados relacionados con la regulación de procesos de biosíntesis de la pared celular secundaria y deposición ectópica de celulosa, hemicelulosa y lignina (Galeano *et al.*, 2015; Mustari *et al.*, 2016), mientras que evaluaciones computacionales y experimentales sobre FTs NAC muestran que son específicos de plantas, y a su vez desempeñan un rol importante en las respuestas al estrés abiótico y a los procesos de formación de la pared celular secundaria (Yan *et al.*, 2017), siendo capaces de inducir la expresión génica de la ruta metabólica de fenilpropanoides y monolignoles (Yang *et al.*, 2011; Tang *et al.*, 2015;). Asimismo, investigaciones muestran que la especie primitiva *Klebsormidium flaccidum* mantiene 3 FTs NAC, mientras que *Panicum virgatum* presenta 425, siendo una de las plantas con mayor cantidad de factores de transcripción NAC (Hori *et al.*, 2014).

Por otro lado, la decodificación funcional de la estructura primaria, utilizando el método del espectro de la información (Veljkovic *et al.*, 1985; Deng y Huang, 2014), ayuda a identificar dominios implicados en la interacción proteína-proteína y proteína-ADN (Veljkovic *et al.*, 2009). Este método asigna a cada aminoácido un potencial de interacción ion-electrón, lo que determina la propiedad electrónica de la proteína mediante señales digitales de frecuencias y amplitudes (Perovic *et al.*, 2013; Veljkovic *et al.*, 2015), datos que ayudarían a explicar mejor los procesos de redes de regulación genética (Lin *et al.*, 2014). Asimismo, los modelamientos de la estructura secundaria mediante homologías, permite realizar posibles hipótesis acerca de la disposición espacial de los aminoácidos y como éstas

pueden interactuar con otras moléculas (Eswar *et al.*, 2006). Para ello, se requiere de proteínas homólogas modeladas a partir de experimentos. Así, la proteína 3ULX de *Arabidopsis thaliana* se presenta como ortóloga a *TgNAC01* (Chen *et al.*, 2011) y está conformada de una estructura β -hoja antiparalela retorcida, que se compacta contra una α -hélice en la región N-terminal. Consecuentemente, esta familia de FT NAC se deriva originalmente de tres proteínas (de las cuales toma la primera letra de cada una para formar sus siglas): NAM (*No Apical Meristem*), ATAF 1–2 (*Arabidopsis Transcription Activation Factor*), y CUC2 (*Cup-Shaped Cotyledon*); además, esta familia se caracteriza por mantener dominios conservados en la región N-terminal para la adhesión al ADN, mientras que su variabilidad de aminoácidos en la región C-terminal sirve para generar las diversas redes de interacción proteína – proteína y de ese modo generar nuevas propiedades emergentes de las células (Maugarny *et al.*, 2016).

También, evaluaciones de ontología génica muestran que el gen *VNI2* de *A. thaliana* es ortólogo a *TgNAC01*, FT que estaría implicado en la regulación negativa de la formación de vasos xilemáticos y una activación transcripcional frente a estrés osmótico, lo cual es una estrategia adaptativa que permite una respuesta rápida a los cambios ambientales (Yang *et al.*, 2011; Ko *et al.*, 2014). El gen *VNI2* también estaría relacionado en la senescencia foliar, regulación de los periodos de floración y morfogénesis, desarrollo de raíces laterales, movilización de nutrientes y maduración del fruto, además de regular la longevidad de la hoja a causa del ácido abscísico (ABA) (Seo y Park, 2011; Yang *et al.*, 2011). Así, evaluar la expresión génica de FTs en diferentes tejidos ayuda a conocer las posibles funciones en la cual está inmersa. Así, el método de qRT-PCR (reacción en cadena de la polimerasa de la transcripción reversa en tiempo real) es uno de los más difundidos y confiables para cuantificar la transcripción de determinados genes en tejidos u órganos, ya que es económico, rápido, reproducible y fiable. Sin embargo, la técnica requiere un procedimiento de normalización, siendo así que los genes enzimáticos *TgUbp*, *TgEf-1* y *TgCAD* son estables a nivel de su expresión, además de ser importantes en la ruta metabólica de fenilpropanoides en *T. grandis* (Galeano *et al.*, 2014). Así, el objetivo de este trabajo fue analizar la filogenia, propiedades y estructura proteica del factor de transcripción *TgNAC01* de *T. grandis* y cuantificar su expresión génica en tejidos lignificados.

MATERIALES Y MÉTODOS

Selección del gen Factor de Transcripción *TgNAC01*

Para la selección del FT *TgNAC01*, se utilizó la base de datos de genes diferencialmente expresados (2413 transcritos) que hacen parte del transcriptoma de *T. grandis*, la cual está depositada en el “Transcriptome Shotgun Assembly” (TSA número de acceso GDLT00000000), disponible en el NCBI

(<https://www.ncbi.nlm.nih.gov/>). Esta base de datos fue obtenida por análisis moleculares y de bioinformática, la expresión diferencial de los genes fue mediante el paquete de R-Project (Dseq), el cual permitió filtrar genes de bases de datos compuestas por cientos de miles de contigs, a decenas de genes (Galeano *et al.*, 2015).

Asimismo, estos transcritos fueron el punto de partida para realizar una búsqueda exhaustiva de factores de transcripción diferencialmente expresados de la familia *NAC* realizando anotaciones manuales en el NCBI y haciendo uso del software Blast2Go (<https://www.blast2go.com/>). Para las anotaciones funcionales, se realizó análisis de homología con las especies *Arabidopsis thaliana* y *Populus trichocarpa*, las cuales son especies de plantas que poseen buenas anotaciones funcionales; *A. thaliana* es una planta modelo y *P. trichocarpa* es un árbol modelo.

De ese modo, se encontró la secuencia del ARNm maduro de un FT *NAC*, al que se le codificó como *TgNAC01*. La identificación de la región CDS (secuencia codificadora, del inglés *coding sequence*) fue realizada utilizando la herramienta de análisis gráficos (Open Reading Frame Finder) (<https://www.ncbi.nlm.nih.gov/orffinder/>), y posteriormente se tradujo la región codificadora a una secuencia de aminoácidos mediante el traductor Expasy (<http://web.expasy.org/translate/>).

Selección de la plantilla y modelamiento de la estructura secundaria del FT *TgNAC01*

La estructura secundaria homóloga a la proteína *TgNAC01* se obtuvo a partir de la búsqueda de alineación local utilizando “protein blast” y la base de datos del PDB (Protein Data Bank). Además, se utilizó la proteína 3ULX como referencia para el modelado de *TgNAC01* y posteriormente mostrar algunas similitudes en la estructura, además de funciones bioquímicas y biológicas. El análisis de los motivos conservados entre las secuencias de aminoácidos fue mediante alineamiento de secuencias múltiples utilizando el software Bioedit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) y el algoritmo de alineamiento Muscle.

La obtención del modelamiento de la estructura secundaria de la proteína en formato “.pdb” fue mediante el software MODELLER (<https://salilab.org/modeller/>). Asimismo, para el modelado del FT *TgNAC01* se utilizó la plantilla (del inglés *template*) de la proteína 3ULX que tiene un ~59% de identidad aproximadamente, además posee <100 residuos de aminoácidos. Consecuentemente, MODELLER ubica las coordenadas atómicas del template, calculando automáticamente un modelo que contiene todos los átomos que no son de hidrógeno. Asimismo, se incluye la asignación de pliegues y alineación de dos secuencias de proteínas. Por otro lado, la visualización y edición de la estructura secundaria fue por medio del software de código abierto PyMol (<https://www.pymol.org/>).

Análisis del potencial de interacción ion-electrón y método del espectro de la información del FT *TgNAC01*

Los aminoácidos que conforman la secuencia del FT *TgNAC01* fueron evaluados mediante el uso del potencial de interacción ion-electrón (PIIE) (Tabla 1), con parámetros físicos esenciales de las moléculas biológicas, y que a su vez están relacionadas con la interacción proteína-proteína / proteína-ADN a más de 100 Å de distancia. El parámetro PIIE fue utilizado como base del método de espectro de información (MEI) para el análisis de estructura primaria y función de las proteínas. La región de aminoácidos biológicamente activos, fue evaluado con el paquete SFAPS (Deng y Huang, 2014), en el lenguaje de Programación R-Project. De ese modo, se realizó el análisis de MEI cruzado para los FTs *TgNAC01* y *1ULX*, además de evaluar la estructura primaria de los FTs ortólogos que poseen los códigos “*Ciclev10032865m*”, “*Potri.001G061200.2*” y “*VNI2*”, disponibles en el “Plant Transcription Factor Database” (<http://planttfdb.cbi.pku.edu.cn/>).

Alineamiento de secuencias múltiples y árbol filogenético del FT *TgNAC01*

El estudio de las divergencias evolutivas de los factores de transcripción *NAC* relacionados con los procesos de biosíntesis de xilema secundario y respuesta a estrés

Tabla 1. Valores del potencial de interacción ion-electron (PIIE) de aminoácidos, utilizado para estudiar el espectro de la información de la estructura primaria de proteínas.

Aminoácido	Abreviatura	Símbolo	EIIP
Leucina	Leu	L	0
Isoleucina	Ile	I	0
Asparagina	Asn	N	0,0036
Glicina	Gly	G	0,005
Ácido glutámico	Glu	E	0,0057
Valina	Val	V	0,0058
Prolina	Pro	P	0,0198
Histidina	His	H	0,0242
Lisina	Lys	K	0,0371
Alanina	Ala	A	0,0373
Tirosina	Tyr	Y	0,0516
Triptofano	Trp	W	0,0548
Glutamina	Gln	Q	0,0761
Metionina	Met	M	0,0823
Serina	Ser	S	0,0829
Cisteína	Cys	C	0,0829
Treonina	Thr	T	0,0941
Fenilalanina	Phe	F	0,0946
Arginina	Arg	R	0,0959
Ácido aspártico	Asp	D	0,1263

abiótico, puede ayudar a comprender mejor los procesos de evolución y adaptación de las plantas. En consecuencia, para analizar el alineamiento de secuencias múltiples y árbol filogenético del FT *TgNAC01*, se realizó una búsqueda de 18 genes ortólogos (que tuvieran como mínimo un porcentaje de similitud de 80 % con el FT *TgNAC01* de *T. grandis*, además de superar el umbral de $1e-95$ para el *e-value*). Para dicho proceso, se usó el buscador blast de la base de datos de Factores de transcripción en plantas (<http://planttfdb.cbi.pku.edu.cn/>). Con estos genes, fue realizada la evaluación de dominios conservados. Posteriormente, los gráficos de los alineamientos de secuencia se obtuvieron con ayuda del software Bioedit, mientras que el análisis filogenético fue mediante la aplicación del software MrBayes (<http://mrbayes.sourceforge.net/>), el cual implementa análisis bayesianos con el método de simulación Monte Carlo de cadena de Markov (MCMC).

Análisis de expresión génica del gen FT *TgNAC01* por PCR en tiempo real

Fue extraído el ARNm de seis muestras (seis réplicas biológicas) de los órganos hoja y raíz de plántulas de tres meses producidas in vitro, que crecieron a partir de semillas de plantaciones brasileñas de tecas de 12 años de edad. Además, se extrajo material genético (seis réplicas biológicas) de los tejidos xilema secundario de tallo de árboles de 12 y 60 años, xilema secundario de rama de árboles de 12 años. Asimismo, se extrajo ARNm de la albura de árboles de 12 y 60 años según métodos establecidos por (Galeano *et al.*, 2015). Además las plantaciones de teca de 12 y 60 años se encuentran a 500 metros de distancia y comparten las mismas condiciones de radiación solar, temperatura, precipitación y tipo de suelo. Posteriormente, tres muestras independientes para qRT-PCR de cada órgano/tejido fueron mezcladas para formar una muestra, obteniendo dos muestras por órgano/tejido. Las dos muestras de ARNm fueron sintetizadas a cDNA (usando un cebador oligo dT), con 1,0 ug de ARN. Para este proceso se utilizó el kit SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific) según las instrucciones del fabricante. La concentración de cDNA se determinó con el Ultraspec 2100 PRO espectrofotómetro (Amersham Biosciences, EE.UU.). Los cebadores (primers) para el tiempo real (también llamado qRT-PCR) se diseñaron flanqueando la región CDS del gen *TgNAC01* (primer forward: 5'-GGTCAAACAGGGCTACAGTTTC-3', primer reverse: 5'-ACCAATTCAAATCTCACTTGTGC-3'), seguido de la determinación de la curva estándar con varias diluciones de cDNA (Figura 6a) y la curva de *melting* (Figura 6b). La mezcla de qRT-PCR contenía 125 ng de cDNA de cada muestra, los cebadores a una concentración final de 50 mM cada uno, 12,5 μ l de la SYBR Green PCR Master Mix (Applied Biosystems, EE.UU.) para finalmente completar el volumen con agua millipore a 25 μ l. La reacción de cada gen se realizó por duplicado (repetición

técnica). Las reacciones de PCR sin cDNA también se realizaron como controles negativos para cada par de cebadores (Galeano *et al.*, 2014). Se realizaron las qRT-PCR empleando el Sistema TMStepOnePlus (Applied Biosystems, EE.UU.). Todas las reacciones de PCR se realizaron en las siguientes condiciones: 2 min a 50 ° C, 2 min a 95 ° C, y 45 ciclos de 15 s a 95 ° C y 1 min a 65 ° C en placas de reacción ópticas de 96 pocillos (Applied Biosystems, EE.UU.). La muestra de hoja se utilizó como calibrador para normalizar los valores entre diferentes platos y el gen *TgEF1 α* como control, por resultados satisfactorios previos al evaluar expresión génica en teca (Galeano *et al.*, 2015). Se llevaron a cabo todos los análisis estadísticos sobre las medias de los datos en el programa R-Project con un nivel de confianza del 95%, utilizando además la prueba F. El procedimiento de comparación de medias se llevó a cabo con la prueba LSD a un nivel de confianza del 95 % (Galeano *et al.*, 2015).

RESULTADOS

Selección del gen Factor de Transcripción *TgNAC01*

Realizando la búsqueda de factores de transcripción de la base de datos de genes diferencialmente expresados del transcriptoma de *T. grandis*, y utilizando las especies *A. thaliana* y *P. trichocarpa* para realizar las anotaciones funcionales en el NCBI, se encontró un transcrito (entre los 2413 disponibles) con el número de acceso GDLT01000898 y código "comp29050_c0_seq2", que se alineó con secuencias de la familia *NAC*. Así, a este gen se le dio el nombre de factor de transcripción *TgNAC01*.

Modelamiento de la estructura secundaria y análisis de método de espectro de la información del FT *TgNAC01*

Posterior a la traducción de la secuencia del gen regulador *TgNAC01* con la ayuda de Expasy, se identificaron, los dominios funcionales NAM, ATAF y CUC. Además, este FT tiene una secuencia de 164 aminoácidos y una estructura general β -hoja antiparalela muy retorcida, que se compacta contra una α -hélice en la región N-terminal. Asimismo, mantiene tres regiones α hélice y siete regiones β plegadas, y la parte restante de la estructura consiste principalmente en bucles (Fig. 1C).

Posteriormente, el análisis de espectro de la información consenso (EIC) entre los factores de transcripción *TgNAC01* y *3ULX*, mostró dos picos de frecuencia "F" de 0,0585 y 0.1875 (Fig. 1B), que representan el componente de largo alcance de la interacción proteína-proteína / proteína-ADN, implicados en la regulación de la expresión génica de la ruta metabólica de fenilpropanoides y respuesta a estrés abiótico en teca. Otro aspecto de las propiedades físico-químicas de la interacción ion-electrón es que los dominios funcionales **FRFHPTD** y **DPWDLP** (Fig. 1A) son esenciales en la regulación génica y tienen la característica de ser conservados dentro de la familia *NAC*, sub dominio NAM (Fig. 1A).

Por otra parte, el análisis de espectro de la información de tres FTs ortólogos de *TgNAC01* (*Ciclev10032865m*, *Potri.001G061200.2* y *VNI2*), muestra picos diferentes, debido a la diversidad de funciones biológicas que tienen los genes reguladores (Fig. 2). La decodificación de la información a partir del espectro de frecuencias y amplitud de las secuencias proteicas de esos tres factores de transcripción (Fig. 2), muestran divergencia evolutiva, siendo esto característico de los FTs y genes implicados en evolución adaptativa de diversas funciones metabólicas y la regulación de las mismas. Respecto al método del espectro de la información (MEI), se observaron mutaciones en el gen NAC que potencialmente alteran la amplitud del pico dominante en $F=0,0625$, e influirían en la afinidad de unión del factor de transcripción con el promotor de otros genes enzimáticos y otros FTs. Es importante señalar también que la inspección del espectro de información (EI) para el FT *VNI2* individualmente, no muestra una amplitud en la $F=0.0625$, pero sí en las frecuencias $F=0.0585$ y $F=0.0703$. De ese modo, se encontró que cambios en las secuencias de aminoácidos divergieron en función de esos picos de frecuencia (Fig. 2D). Para identificar los cambios que contribuyen a la evolución

del gen, se calculó la variación de los cuatro genes en función de espectro consenso, notando de ese modo que *TgNAC01* y *Ciclev10032665m* de *Citrus dementina* comparten dos picos de frecuencia, los cuales son $F=0.0625$ y $F=0.1875$ (Fig. 3). Además, este último también es compartido con el FT *VNI2* (Fig. 3). Por otro lado, se muestran dos picos dominantes, para los FTs *TgNAC01* y *Potri.001G061200.2*, con $F=0.0625$ y $F=0.2812$, respectivamente (Fig. 2).

Asimismo, el espectro de la información consenso (EIC) mostró que la preferencia de unión al receptor blanco del FT *TgNAC01* es $F=0.0625$ y lo comparten los factores de transcripción *Ciclev10032865m*, *Potri.001G061200.2* y *TgNAC01* (Fig. 3). La segunda frecuencia con mayor amplitud, que corresponde a $F=0.3984$, es compartido con los FTs *Ciclev10032865m* y *Potri.001G061200.2*, mientras que el pico de amplitud en la frecuencia $F=0.1875$ lo tienen los genes *Ciclev10032865m*, *Potri.001G061200.2* y *TgNAC01* (Fig. 3). Asimismo, mediante el alineamiento múltiple de secuencias (Fig. 5), se observa que las mutaciones más comunes en los FTs NAC, y subdominio NAM, se encuentran en los aminoácidos V29, V11, D50, S54, S63, I80, N86, L102, K130 y C132, lo que aumenta notablemente los picos

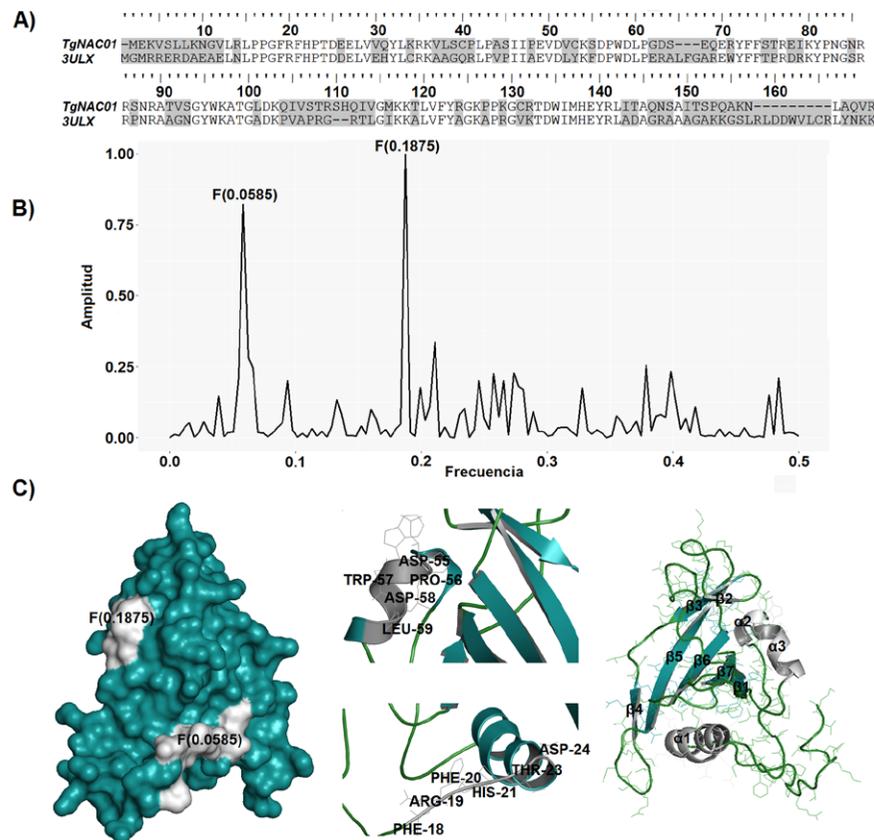


Figura 1. (A) Secuencia de residuos de aminoácidos de las proteínas *TgNAC01* y *3ULX*, resaltando los dominios conservados en común. (B) Análisis del espectro de la información consenso entre los factores de transcripción *TgNAC01* y *3ULX*, se muestra dos picos de frecuencia $F = 0.0585$ y $F=0.1875$. (C) Modelación proteica del factor de transcripción *TgNAC01*, donde se muestra la estructura secundaria del FT, además de la región de aminoácidos, con gran potencial de interacción ion-electron.

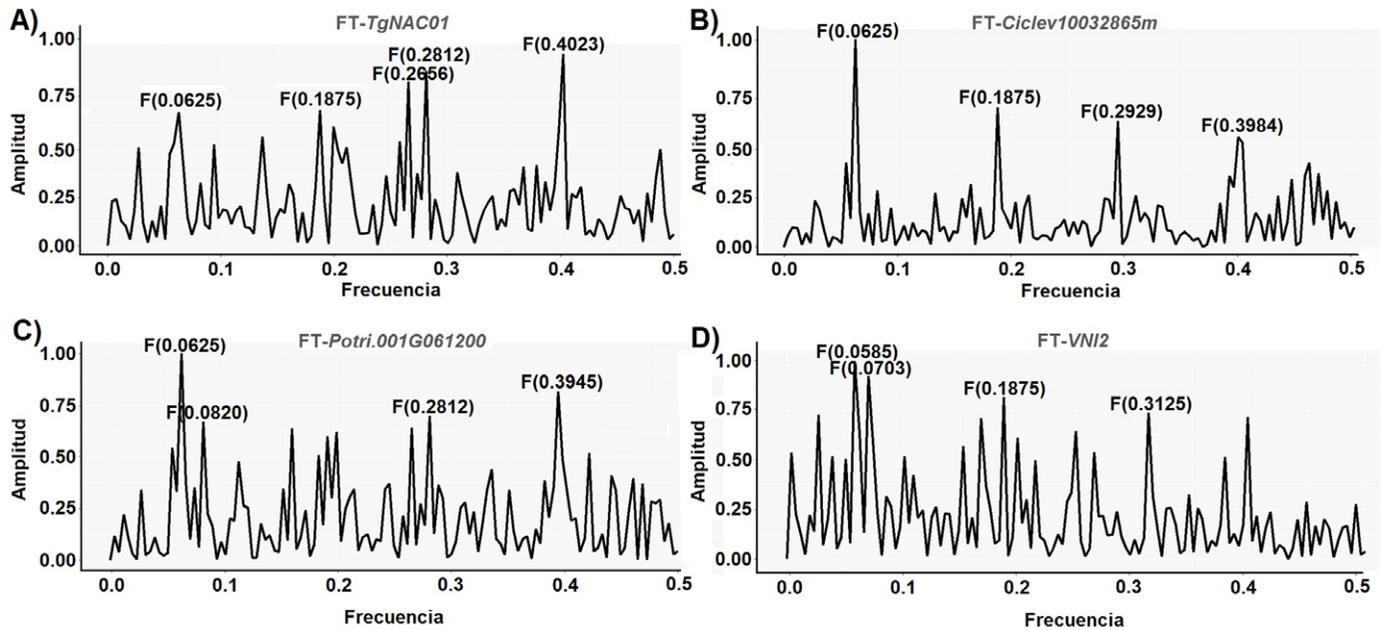


Figura 2. (A) Espectro de la información (EI) de la secuencia del FT *TgNAC01*, (B) correlación de amplitud y frecuencia de gen FT *Ciclev10032865* de *Citrus clementina*, (C) (EI) del FT de la especie *P. trichocarpa* (*Potri.001G081200.2*), (D) (EI) del gen master regulador *VNI2* de *Arabidopsis thaliana*.

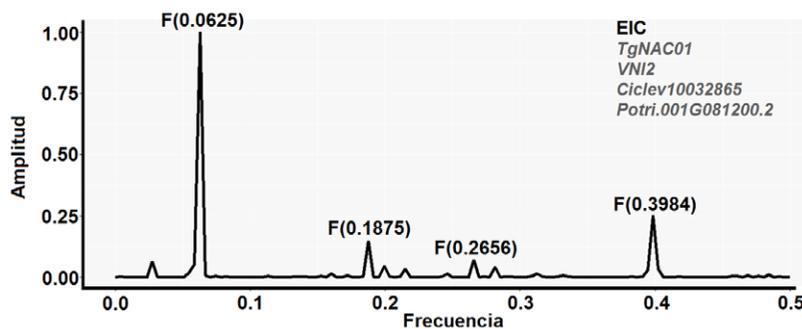


Figura 3. Espectro de la información consenso entre el FT *TgNAC01* y 3 ortólogos (*Ciclev10032865*, *Potri.001G081200.2* y *VNI2*), donde se muestra cuatro picos dominantes de frecuencia.

de amplitud, y pueden desempeñar un papel importante en la aparición de una nueva función en el gen. Además, las mutaciones en la región C-terminal podrían aumentar o disminuir la amplitud en la frecuencia dominante, lo que implicaría influir en la afinidad de unión al receptor.

Por estos resultados, se puede observar que las secuencias blanco de interacción (proteína-proteína / proteína-DNA) son diversos. Además, observando la ortología génica y los resultados del método del espectro de la información, se puede argumentar que los FTs comparten ciertas funciones bioquímicas y biológicas entre sí.

Alineamiento de secuencias múltiples y árbol filogenético del FT *TgNAC01*

Como resultados, los genes *SMIL_00015837-RA_Salv* de *Salvia miltiorrhiza* y *XP_011086848.1* de *Sesamum indicum* son los más próximos al gen *TgNAC01* de *T. grandis* (Fig. 4)

debido a su relación con el orden Lamiales. Por otro lado, el gen *Potri.001G061200.2* de *Populus trichocarpa* correspondió con el gen *NAC* de *Salix purpurea* (*SapurV1A.0173s0020.3.p*), *Gossypium raimondii* (*Gorai.012G125500.2*), *Theobroma cacao* (*Thecc1EG008477t1*), *Manihot esculenta* *Manes.09G185300.1.p* y *Juglans regia* (*WALNUT_00019442-RA*) (Fig. 4), todos ellos conservando los cinco dominios funcionales en la región N-terminal (Fig. 5).

Por otro lado, los genes *Ciclev10032865m* (*Citrus clementina*) y *orange1.1g031844m* (*Citrus sinensis*), también conservan cinco dominios funcionales al igual que *AT5G13180.1* (*Arabidopsis thaliana*) y *KHN03193.1* (*Glycine max*). Asimismo, el gen *Prupe.1G220400.1.p* está relacionado con *XP_008222225.1* (*Prunus mume*), *FANhyb_rscf00000097* (*Fragaria ananassa*), *Pyrus bretschneideri* (*Pbr004500.1*) y *XP_015881782.1* (*Ziziphus jujuba*).

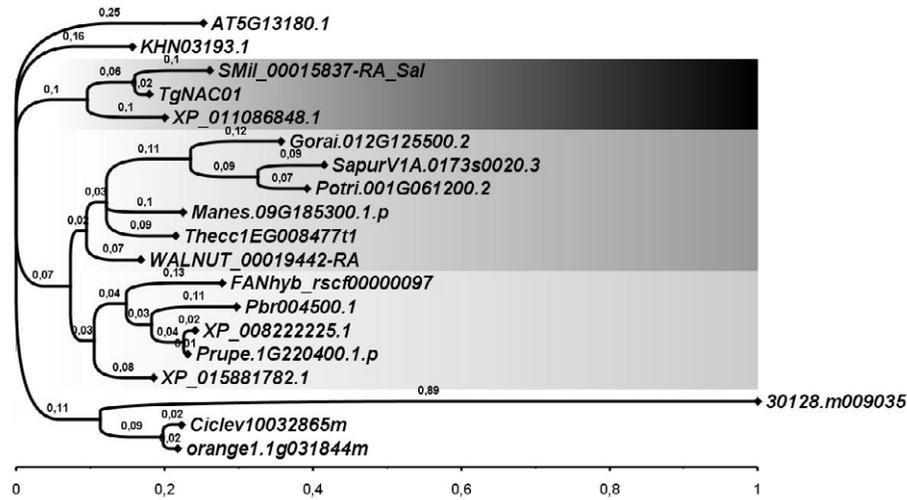


Figura 4. Árbol filogenético del FT *TgNAC01* de *T. grandis* y 18 genes ortólogos, que tienen un porcentaje de similitud < 80 %, además de superar el umbral de *e-value* 1e-95.

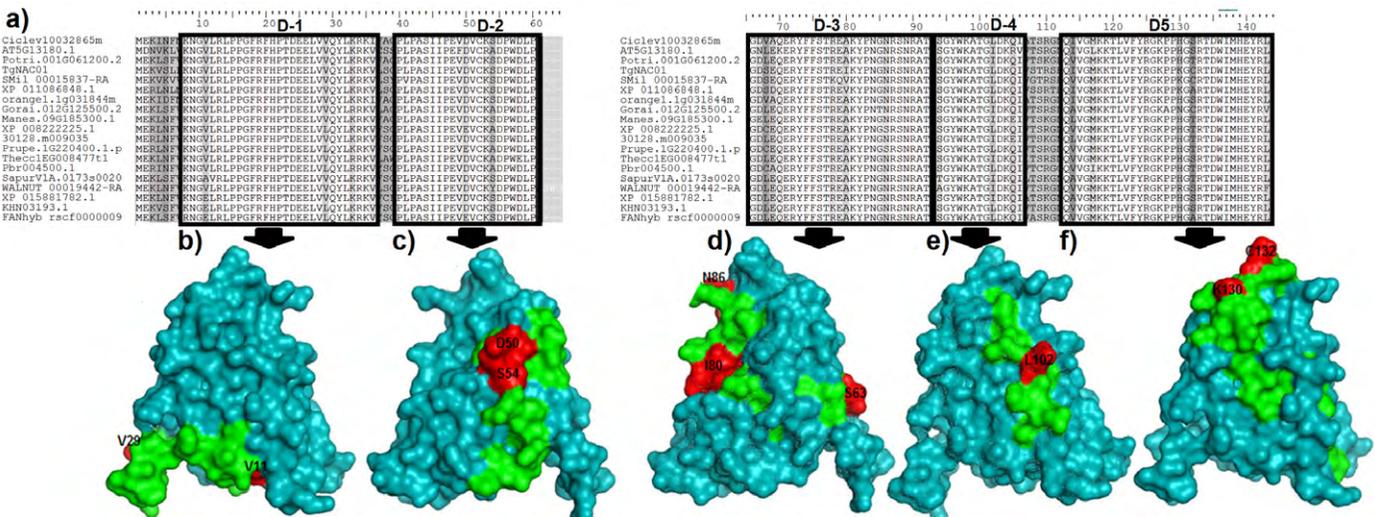


Figura 5. (A) Alineamiento de secuencias factores de transcripción NAC ortólogos al gen *TgNAC01* de Teca. (B) Identificación del dominio funcional D-1 y aminoácidos V20 y V11 que mutaron temporalmente. (C) Selección del dominio conservado D-2, presentando 2 aminoácidos D50 y S54 con mayor cambio evolutivo. (D) Secuencia conservada de aminoácidos D-3 y ubicación en la estructura proteica de las 3 mutaciones más frecuentes, situados en N86, I80 y S63. (E) Dominio D-4 y cambios en el aminoácido L102. (F) Secuencia conservada D-5 y cambios en la secuencia primaria del dominio en los K130 y C132.

Adicionalmente, analizando los dominios conservados de la familia de factor de transcripción NAC, se observa que los genes *Ciclev10032865m*, *AT5G13180.1*, *Potri.001G061200.2*, *TgNAC01*, *SMil_00015837-RA_Salv*, *XP_011086848.1*, *orange1.1g031844m*, *Gorai.012G125500.2*, *Manes.09G185300.1.p*, *XP_008222225.1*, *30128.m009035*, *Prupe.1G220400.1.p*, *Thecc1EG008477t1*, *Pbr004500.1*, *SapurV1A.0173s0020.3.p*, *WALNUT_00019442-RA*, *XP_015881782.1*, *KHN03193.1*, *MENK1F*, *KNGVLRPPGFRFHPTDEELVQYLKRV* y *FANhyb_rscf00000097.1.g00004.1* mantienen cinco dominios conservados, viendo así que el primer dominio D-1 tiene los

aminoácidos **KNGVLRPPGFRFHPTDEELVQYLKRV** sin diferencias marcadas entre las secuencias. Sin embargo, en la posición de los aminoácidos de Valina [V29-V11] se presentan más mutaciones entre las secuencias alineadas (Fig. 5b). Por otra parte, el dominio D-2 tiene conservada la secuencia de aminoácidos **PLPASIPEVDVCKSDPWDL**, donde los aminoácidos Ácido aspártico [D50] y Serina [S54] mutaron a través del tiempo, siendo claves, ya que tienen altos valores de potencial de interacción ion-electrón (Fig. 5c). Asimismo, resultados del dominio D-3 presentan a los aminoácidos **GDSEQERYFFSTREIKYPNGNRSNRAT** como los más

conservados, aunque existen mutaciones en Serina [S63], Isoleucina [I80] y Asparagina [N86] (Fig. 5d). El cuarto dominio D-4 conserva los trece aminoácidos **SGYWKATGLDKQI**, de los cuales Leucina [L102], presentó mayores modificaciones en los genes ortólogos a teca (Fig. 5e), mientras que en el quinto dominio existen 31 aminoácidos conservados (**QIVGMKKTLVYFRGKPPKGCRTDWIMHEYRL**), habiendo mutaciones significativas en Lisina [K130] y Cisteína [C132] (Fig. 5f).

Análisis de la expresión génica del factor de transcripción *TgNAC01*

Los primers diseñados para el gen *NAC* presentaron una adecuada curva estándar (Figura 6a) y de melting (Figura 6b). Así, los órganos hoja y raíz de teca (provenientes de material vegetal *in vitro* de 3 meses de edad), no presentan niveles de expresión génica significativos para el gen *NAC* cuando fueron comparados con los otros órganos y tejidos

(Fig. 6c). La porción más externa de xilema secundario de tallo de tecas de 60 y 12 años presentó aproximadamente 16 veces más expresión (Fig. 6c), mientras que en el xilema secundario de rama de teca de 12 años mostró 30 veces más expresión, ambos cuando son comparados con hoja (Tabla 2). Del análisis de ANOVA, se muestra que el valor *p* del test F es menor de 0,05, lo que representa una diferencia estadística significativa entre las medias de las siete variables, con un 95 % de nivel de confianza (Tabla 3).

Para determinar cuáles medias tienen diferencia significativa unas con otras, se realizó la prueba LSD (prueba de medias) al 95 % de confianza estadística. De ese modo, se encontraron 4 grupos, con una confianza estadística del 95 %. El primer grupo se compone de hoja y raíz, el segundo es de xilema secundario de tallo en árboles de 12 y 60 años, el tercero es de xilema secundario de rama, mientras que el cuarto grupo está conformado por la expresión génica del FT *TgNAC01* en albura de tallo de árboles de 12 y 60 años de edad (Tabla 4).

Tabla 2. Medias de las muestras (tres réplicas biológicas y dos réplicas técnicas) del FT *TgNAC01* en hoja, raíz, xilema secundario de rama de árboles de 12 años, xilema secundario de tallo de árbol de 60 años, xilema secundario de tallo de árbol de 12 años, albura de árboles de 60 años y albura de árboles de 12 años.

Tejido	Gen	Muestra1	Muestra2	Media	Desv. Est
Hoja	<i>NAC</i>	1	1	1	0
Raíz	<i>NAC</i>	2,41	2,00	2,21	0,29
Tallo Xil. Sec 60a	<i>NAC</i>	17,78	14,89	16,33	2,05
Tallo Xil. Sec 12a	<i>NAC</i>	19,80	15,89	17,84	2,77
Rama Xil. Sec 12a	<i>NAC</i>	29,40	33,32	31,36	2,77
Albura 60a	<i>NAC</i>	46,47	53,55	50,01	5,01
Albura 12a	<i>NAC</i>	55,74	43,34	49,54	8,77

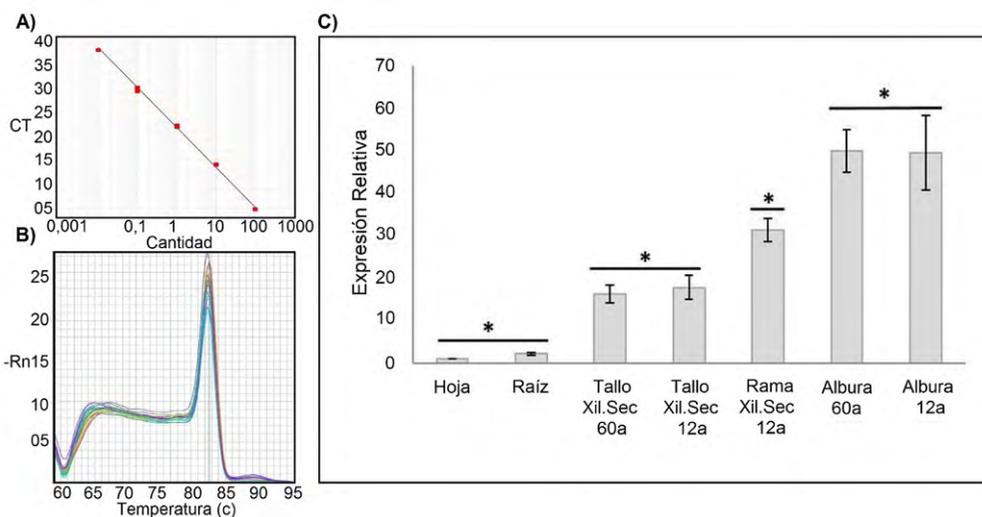


Figura 6. (A) Curva Estándar del gen *TgNAC01*. (B) Curva de melting del gen *TgNAC01*. (C) Histograma con desviación estándar de la expresión génica del factor de transcripción *TgNAC01* en los órganos hoja, raíz y tallo. Además, se presenta la diferencia estadística (95% confianza) en la expresión entre los diferentes tejidos y diferentes edades, con asterisco. Los valores del eje Y corresponden a la expresión relativa del gen.

Tabla 3. Análisis de varianza de expresión génica de gen factor de transcripción *TgNAC01* entre tejidos y muestras repetidas de hoja, raíz, xilema secundario de rama de 12 años, xilema secundario de tallo de árbol de 60 años, xilema secundario de tallo de árbol de 12 años, albura de árboles de 60 años y albura de árboles de 12 años.

Fuente	Suma de cuadrados	Grados de libertad	Cuadrado medio	Test F	Valor p
Entre grupos	4962,65	6	827,108	---	---
Dentro de grupos	121,536	7	17,36	47,64	0,00001

Tabla 4. Análisis de medias mediante la prueba LSD entre tejidos (hoja, raíz, xilema secundario de rama, xilema secundario de tallo y albura) que presentan expresión del gen *TgNAC01*.

Muestras	Prueba LSD 95%
Hoja	a
Raíz	a
Tallo60a	b
Tallo12a	b
Rama12a	c
Albura60a	d
Albura12a	d

Lo más interesante es que la albura del tallo de árboles de *T. grandis* a los 12 y 60 años presentó 50 veces más expresión del gen *NAC* cuando se compara con hoja y raíz, y casi 35 veces más expresión comparado con la porción externa de xilema secundario de tallo de ambas edades (Fig. 6). En consecuencia, los datos experimentales de expresión génica del gen *TgNAC01* en los tejidos lignificados como xilema secundaria de tallo, rama y albura de árboles de 12 y 60 años, comprueban que el gen *NAC* está relacionado directamente con la formación de lignina y acoplamiento de monolignoles. Además, como *NAC* es un gen factor de transcripción, es probable que esté asociado con la regulación transcripcional de los genes miembros de la vía fenilpropanoides. Por otro lado, como el gen *NAC* no muestra diferencias estadísticamente significativas entre las edades de 60 y 12 años, estos datos evidencian que este gen no tiene funciones en la maduración de la madera de teca a lo largo de la síntesis de lignina. En consecuencia, el FT *NAC* muestra mayor expresión en albura (tejido compuesto mayormente de xilema secundario), el cual juega un papel fundamental en la organización y desarrollo de las partes que componen este tejido, como las células radiales de la zona cambial, los vasos, fibras y los anillos de crecimiento.

DISCUSIÓN

Las plantas carecen de un sistema nervioso, y en reemplazo a ello mantiene óptimos sistemas bioquímicos que regulan espacial y temporalmente sus respuestas fisiológicas (Chen *et al.*, 2011; Kim *et al.*, 2016). Los estudios de la estructura primaria y secundaria de la proteína que configura el FT *TgNAC01* ayudan a entender los procesos de evolución

y adaptación de la teca al medio ambiente (Perovic *et al.*, 2013; Skj *et al.*, 2014), siendo este gen un regulador maestro de importantes rutas metabólicas como la biosíntesis de xilema secundario y la respuesta a estrés abiótico, como fue descrito recientemente (Camel *et al.*, 2017). En el genoma de las plantas, existen diversos genes *NAC* (Aude Maugarny, Beatriz Goncalves, Stefan Jouannic, Michael Melkonian, 2016; Chen *et al.*, 2011) que son considerados los principales implicados en la respuesta frente al estrés abiótico, y fueron caracterizados como esenciales en la transición de algas acuáticas a plantas terrestres (Aude Maugarny, Beatriz Goncalves, Stefan Jouannic, Michael Melkonian, 2016). Además, esta trayectoria de la evolución puede estar codificada en los aminoácidos con gran potencial de interacción ion-electrón. Según los resultados del espectro de la información (MEI) de este trabajo, los genes ortólogos *VNI2*, *Ciclev10032865m* y *Potri.001G061200.2* comparten ciertas funciones, sin embargo cambios en los aminoácidos V29, V11, D50, S54, S63, I80, N86, L102, K130 y C132 probablemente dieron nuevas propiedades biológicas a los FTs. Este fenómeno ha sido descrito previamente (V. Veljkovic, I. Cosic, B. Dimitrijevic, 1985; Veljkovic *et al.*, 2014). Asimismo, el mayor pico de frecuencia consenso $F=0.0625$ muestra alta capacidad de interacción molecular y mutaciones puntuales sobre esa secuencia blanco, que potencialmente conllevarían a nuevas redes de regulación genética.

En efecto, estudios recientes muestran que genes ortólogos a *TgNAC01*, como *VNI2* tienen múltiples funciones, como la respuesta frente a estrés salino, al ácido abscísico, desarrollo de xilema, senescencia y longevidad

de las hojas, (Kim *et al.*, 2016; Yang *et al.*, 2011), los cuales estarían en función de los diferentes picos de frecuencias del espectro de la información. Asimismo, mediante el uso de la proteína modelo seleccionada (*3ULX*), se identificó que los dominios conservados en la región N-Terminal de la proteína *NAC* pertenecen a la sub familia *NAM* (del inglés *No Apical Meristem*), y esta región está implicada en los procesos de desarrollo de la planta (Yang *et al.*, 2011; Aude Maugarny *et al.*, 2016). Además, esta región está relacionada con procesos catabólicos, metabolismo de carbohidratos, regulación génica frente al estrés abiótico, producción de biopolímeros y procesos de xilogénesis (Galeano *et al.*, 2014). Por otro lado, con respecto a la expresión génica del gen *TgNAC01* entre tejidos y edades de la teca, se observó que existe mayor expresión en el tejido albura (tanto en la edad de 12 como de 60 años), desempeñando un papel fundamental en la organización y desarrollo de las partes que componen este tejido, como las células radiales de la zona cambial, los vasos, fibras y los anillos de crecimiento (Diningrat *et al.*, 2015; Galeano *et al.*, 2015). Por otro lado, como el gen *NAC* no muestra diferencias estadísticas significativas entre las edades de 60 y 12 años, se evidencia así que este gen no tiene funciones en la maduración de la madera de teca a lo largo de la síntesis de lignina (Galeano *et al.*, 2015). Contrario a este gen, estudios previos encontraron una relación significativa de otros factores de transcripción, como el *TgMYB3* y *TgMYB4* de teca que están relacionados con la maduración de esta especie (Galeano *et al.*, 2015). Así, es probable que el gen *TgNAC01* tenga alguna regulación positiva o negativa en la expresión de otros factores de transcripción como los *MYB* de teca (Camel *et al.*, 2017), ya que genes de la familia *NAC* son tradicionalmente considerados como reguladores maestro en la biosíntesis de lignina (Galeano *et al.*, 2015b). Los resultados de este estudio son valiosos, ya que permiten generar hipótesis acerca de cómo los factores de transcripción encajan en los procesos de regulación genética.

CONCLUSIONES

Con la secuencia del cDNA del gen *TgNAC01* (el cual es un transcrito diferencialmente expresado, DERs, específico de tejido de xilema secundario de tallo), fue posible modelar la estructura secundaria de la proteína y analizar la secuencia de aminoácidos mediante el método de espectro de la información, con el apoyo de las bases de datos. Asimismo, el gen *FT TgNAC01*, se presume estar implicado en la formación de lignina y acoplamiento de monolignoles, por sus dominios proteicos, los picos del espectro de la información consenso (EIC), el dendrograma y los alineamientos. Además, es posible que este gen esté implicado en la regulación transcripcional de genes miembros de la vía fenilpropanoides, ya que su expresión génica es mayor en tejidos lignificados cuando es comparada con hoja de tecas

de tres meses de edad. Siendo así, este *FT* podría jugar un papel importante en la organización y desarrollo de las partes que componen la albura, como las células radiales de la zona cambial, los vasos, fibras y anillos de crecimiento. Finalmente, observando los resultados de los alineamientos que se complementan con la expresión génica, los motivos del *FT TgNAC01* se mantienen conservados dentro de la sub familia *NAM*, la cual es una información esencial que abre puertas a otras investigaciones, para así aumentar los estudios funcionales de genes de interés en teca, árbol de gran importancia biotecnológica.

AGRADECIMIENTOS

Los autores agradecen a la “Escola Superior de Agricultura Luiz de Queiroz” de la Universidade de São Paulo, por el soporte financiero y bioinformático.

CONFLICTO DE INTERESES

Los autores declaran que no existen conflictos de intereses.

REFERENCIAS

- Maugarny-Calès A, Gonçalves B, Jouannic S, Melkonian M, Wong GKS, Laufs P. Apparition of the *NAC* transcription factors predates the emergence of land plants letter to the editor. *Mol Plant*. 2016;1345-1348. Doi:10.1016/j.molp.2016.05.016
- Camel V, Galeano E, Carrer H. Red de coexpresión de 320 genes de *Tectona grandis* relacionados con procesos de estrés abiótico y xilogénesis. *TIP*. 2017;20(2):5-14. Doi:10.1016/j.recqb.2017.04.001
- Chen Q, Wang Q, Xiong L, Lou Z. A structural view of the conserved domain of rice stress-responsive *NAC1*. *Protein Cell*. 2011;2(1):55-63. Doi:10.1007/s13238-011-1010-9
- Deng S, Huang D. SFAPS: An R package for structure/function analysis of protein sequences based on informational spectrum method. *Methods*. 2014;69(3):207-212. Doi:10.1016/j.ymeth.2014.08.004
- Diningrat DS, Widiyanto SM, Pancoro A., I, Shim D, Panchangam B, *et al.* Transcriptome of teak (*Tectona grandis* L.f) in vegetative to generative stages development. *J Plant Sci*. 2015;10(1):1-14. Doi:10.3923/jps.2015.1.14
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, *et al.* Comparative protein structure modeling using Modeller. *Curr Protoc Bioinforma*. 2006. Doi:10.1002/0471250953.bi0506s15.Comparative
- Galeano E, Vasconcelos TS, Ramiro DA, De Martin VDF, Carrer H. Identification and validation of quantitative real-time reverse transcription PCR reference genes for gene expression analysis in teak (*Tectona grandis* L.f.). *BMC Res Notes*. 2014;7(1):464. Doi:10.1186/1756-0500-7-464
- Galeano E, Vasconcelos TS, Vidal M, Mejia-guerra MK, Carrer H. Large-scale transcriptional profiling of lignified tissues in *Tectona grandis*. *BMC Plant Biol*. 2015b;15:221. Doi:10.1186/s12870-015-0599-x

- Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, *et al.* Factors for plant terrestrial adaptation. *Nat Commun.* 2014;5:1-9. Doi:10.1038/ncomms4978
- Kim HJ, Nam HG, Lim PO. Regulatory network of NAC transcription factors in leaf senescence. *Curr Opin Plant Biol.* 2016;33:48-56. Doi:10.1016/j.pbi.2016.06.002
- Ko J, Jeon H, Kim W, Kim J, Han K. The MYB46 / MYB83-mediated transcriptional regulatory programme is a gatekeeper of secondary wall biosynthesis. *Ann Bot.* 2014;1099-107. Doi:10.1093/aob/mcu126
- Lin L, Lucas M De, Turco G, Toal TW, Gaudinier A, Young NF, *et al.* An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature.* 2014;517:571-575. Doi:10.1038/nature14099
- Mustari E, Diningrat DS, Ratnasih R, Widiyanto SM. APETALA2 and APETALA3 Genes expression profiling on floral development of teak (*Tectona grandis* Linn f.). *J Plant Sci.* 2016;11(4):61-8. Doi:10.3923/jps.2016.61.68
- Nunif TK, Murchison HG. Provisional yield models of Teak (*Tectona grandis* Linn F.) plantations in northern Ghana. *For Ecol Manag.* 1999;120:171-178.
- Perovic VR, Muller CP, Niman HL, Veljkovic N, Dietrich U, Tosic DD, *et al.* Novel phylogenetic algorithm to monitor human tropism in egyptian H5N1-HPAIV reveals evolution toward efficient human-to-human transmission. *PLoS ONE.* 2013;8(4). Doi:10.1371/journal.pone.0061572
- Seo PJ, Park C. Signaling linkage between environmental stress resistance and leaf senescence in *Arabidopsis*. *Plant Signal Behav.* 2011;6(10):1564-1566. Doi:10.4161/psb.6.10.17003
- Skj L, Yao X, Scarabelli G, Grant BJ. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics.* 2014:1-11. Doi:10.1186/s12859-014-0399-6
- Tang X, Zhuang Y, Qi G, Wang D, Liu H, Wang K. Poplar PdmYB221 is involved in the direct and indirect regulation of secondary wall biosynthesis during wood formation. *Nat Publ Group.* 2015:1-14. Doi:10.1038/srep12240
- Veljkovic V, Cosic I., Dimitrijevic B., Lavolic D. Is It possible to analyze DNA and protein sequences by the methods of digital signal processing? *Trans Biomed Eng.* 1985(5):337-341.
- Veljkovic V, Glisic S, Claude P, Scotch M, Branch DR, Perovic VR, *et al.* *In silico* analysis suggests interaction between Ebola virus and the extracellular matrix. *Front Microbiol.* 2015;6:1-11. Doi:10.3389/fmicb.2015.00135
- Veljkovic V, Glisic S, Veljkovic N, Bojic T, Dietrich U, Perovic VR, *et al.* Influenza vaccine as prevention for cardiovascular diseases: possible molecular mechanism. *Vaccine.* 2014. Doi:10.1016/j.vaccine.2014.07.007
- Veljkovic V, Veljkovic N, Muller CP, Müller S, Glisic S, Perovic V, *et al.* Characterization of conserved properties of hemagglutinin of H5N1 and human influenza viruses: possible consequences for therapy and infection control. *BCM Struct Biol.* 2009;10:1-10. Doi:10.1186/1472-6807-9-21
- White KJ. Tree breeding with teak (*Tectona grandis*). *Aust For.* 2016;9158. Doi:10.1080/00049158.1962.10675915
- Yan H, Zhang A, Ye Y, Xu B, Chen J, He X, *et al.* Genome-wide survey of switchgrass NACs family provides new insights into motif and structure arrangements and reveals stress-related and tissue-specific NACs. *Nat Publ Group.* 2017:1-15. Doi:10.1038/s41598-017-03435-z
- Yang S, Seo PJ, Yoon H, Park C. The *Arabidopsis* NAC Transcription factor VNI2 integrates abscisic acid signals into leaf senescence via the COR / RD Genes. *Plant Cell.* 2011;23:2155-2168. Doi:10.1105/tpc.111.084913
- Zhong R, Mccarthy RL, Lee C, Ye Z. Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation. *Plant Physiol.* 2011;157:1452-1468. Doi:10.1104/pp.111.181354