

© 2017 Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Zaragoza.  
Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).  
TIP Revista Especializada en Ciencias Químico-Biológicas, 20(2): 5-14, 2017.  
DOI: 10.1016/j.recqb.2017.04.001

## RED DE COEXPRESIÓN DE 320 GENES DE *Tectona grandis* RELACIONADOS CON PROCESOS DE ESTRÉS ABIÓTICO Y XILOGÉNESIS

Vladimir Camel<sup>1a</sup>, Esteban Galeano<sup>2b</sup> y Helaine Carrer<sup>2c</sup>

<sup>1</sup>Instituto de Biotecnología e Ingeniería Genética (IBIG), Facultad de Ciencias Forestales y del Ambiente, Universidad Nacional del Centro del Perú, Av. Mariscal Castilla #3909, Huancayo, Región Junín, Perú.

<sup>2</sup>Centro de Biotecnología Agrícola (CEBTEC), Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Av. Pádua Dias #11, Piracicaba, Edo. De São Paulo, Brazil.

E-mails: <sup>a</sup>vlad\_camel@hotmail.com, <sup>b</sup>estebangg18@hotmail.com, <sup>c</sup>hecarrer@usp.br

### RESUMEN

*Tectona grandis* es un árbol maderable de importancia económica en bosques tropicales y subtropicales. Mediante este estudio, se identificaron familias de factores de transcripción (FTs) y genes codificantes para enzima, diferencialmente expresados en el xilema del tallo, implicados en la regulación de la respuesta a estrés abiótico y xilogénesis en *T. grandis*. Así, fue analizada la distribución evolutiva de 19 genes codificantes para FTs de *T. grandis* mediante análisis filogenéticos. También, fue utilizada la minería de bases de datos y publicaciones para identificar 320 genes de *Arabidopsis thaliana* (ortólogos a *T. grandis*) como soporte experimental y predictivo. Como resultados, se encontraron FTs de las familias *bZIP*, *MYB*, *NAC*, *ER*, *bHLH*, *NuY* y genes que codifican enzimas. Así mismo, se logró analizar el interactoma de *T. grandis* encontrando correlaciones de Pearson significativas para genes que regulan vías metabólicas de fenilpropanoides y estrés abiótico. Además, la red de coexpresión reveló nodos y aristas entre los genes *TgRAP1*, *TgMyB1*, *TgHSF1*, *TgMyB3*, *TgNAC1*, *TgTsiid1*, *TgLieTFs1*, *TgNuy3*, *TgRAP2* y *TgNuy4*. En particular, los análisis de ontología génica mostraron 31 genes de respuesta a estrés abiótico, principalmente *TgHShT1*, *TgHSF1* y *TgHSF2* como correguladores. Además, se encontró que el regulador maestro *TgNAC1*, está involucrado en la corregulación de otros factores de transcripción.

**Palabras Clave:** biología de sistemas, enriquecimiento funcional, genes diferencialmente expresados, minería de literatura, ontología génica.

### Coexpression Network of 320 genes of *Tectona grandis* involved in abiotic stress and xylogenesis processes

### ABSTRACT

Teak (*Tectona grandis*) is a timber tree of economic importance in tropical and subtropical forests. The aim of this work was to identify families of transcription factors (TFs) and enzyme-coding genes differentially expressed (DREs) in stem xylem and their regulation involved in abiotic stress response and xylogenesis in *T. grandis*. Therefore, the evolutionary distribution of 19 TFs of *T. grandis* was derived using a phylogenetic analysis. Besides, specific data mining procedures of databases and publications were performed in order to identify 320 *Arabidopsis thaliana* genes (orthologous to *T. grandis*) as experimental and predictive support. As results, we found transcription factors of the *bZIP*, *MYB*, *NAC*, *ER*, *bHLH* families, and enzyme-coding genes. Furthermore, interactome analysis in *T. grandis* showed a significant Pearson correlation for genes regulating metabolic pathways of phenylpropanoids and abiotic stress. Also, the coexpression network revealed nodes and edges between *TgRAP1*, *TgMyB1*, *TgHSF1*, *TgMyB3*, *TgNAC1*, *TgTsiid1*, *TgLieTFs1*, *TgNuy3*, *TgRAP2* and *TgNuy4* genes. Gene ontology analyses showed that 31 genes respond to abiotic stress, mainly *TgHShT1*, *TgHSF1* and *TgHSF2*, as co-regulators. In addition, the TFs master regulator *TgNAC1* was found to be involved in the co-regulation of other TFs.

**Key Words:** systems biology, functional enrichment, differentially expressed genes, text mining, gene ontology.

## INTRODUCCIÓN

La Teca (*Tectona grandis*) es un árbol caducifolio de bosques tropicales y subtropicales, que posee un rápido crecimiento y es apreciado por su madera (Anish *et al.*, 2015). Además, esta especie tiene un gran potencial para captura de carbono e incrementar su biomasa para formación de energía renovable y sustentable (Krause *et al.*, 2006; Quiala *et al.*, 2012). A pesar de los numerosos estudios fenotípicos en *T. grandis*, poco se conoce de la especie a un nivel molecular y bioquímico (Diningrat *et al.*, 2015; Galeano *et al.*, 2014). Así, la tecnología de RNAseq (secuenciación de RNA) permite revelar genes diferencialmente expresados (DERs) (Wan *et al.*, 2012) y dentro de ellos, identificar algunos factores de transcripción (FTs) implicados en respuestas fisiológicas a diferentes factores ambientales o cambios metabólicos (Diningrat *et al.*, 2015; Galeano *et al.*, 2015). Algunos procesos que podrían ser regulados por este tipo de factores de transcripción son el depósito de celulosa, hemicelulosa y lignina en la pared celular. Este tipo de genes reguladores pueden activarse por variación en la concentración de sales en la célula, estrés hídrico, deficiencia de nutrientes, oxidación, alta osmolaridad o cambios de condiciones de luz y temperaturas extremas (Gill & Tuteja, 2010; Prash & Sonnewald, 2013).

Generalmente, los FTs reconocen a los promotores corriente arriba (del inglés, *upstream*) en las secuencias de genes que serán regulados y pueden ser activados por procesos biológicos y ambientales (Matsui *et al.*, 2008). Además, se han observado interacciones de algunas familias de FTs para dar respuesta fisiológica a diferentes tipos de estrés y a procesos metabólicos secundarios, como el depósito de lignina y posterior producción de madera (Lin *et al.*, 2015; Sundar *et al.*, 2008). Los factores de transcripción son proteínas que al unirse al DNA pueden interactuar con otros reguladores transcripcionales (como pueden ser los reguladores maestros), en procesos como remodelación de la cromatina, modificación de proteínas, reclutamiento o bloqueo de las RNA polimerasas que acceden a la cadena de DNA para el proceso de transcripción (Udvardi *et al.*, 2007).

Por otro lado, la tecnología de RNAseq es una técnica cuantitativa que ayuda a determinar niveles de expresión de RNA y su aplicación directa puede direccionarse a la construcción de bases de datos a gran escala para hacer estudios de redes de coexpresión (Dameron *et al.*, 2013; Mizrachi *et al.*, 2010). Dentro de las largas listas de genes detectadas por el RNAseq en diferentes condiciones ambientales, de edad o de tejidos, se pueden analizar por bioinformática patrones de expresión similares entre sí (también llamado coexpresión); genes con patrones y funciones biológicas similares pueden ser anotados por ontologías génicas y luego ser agrupados (Consortium, 2000; Dameron *et al.*, 2013). Además, la sistematización y agrupación de los genes coexpresados, pueden estar apoyados con resultados experimentales (Consortium, 2000; Dameron *et al.*, 2013), y de ese modo poder elucidar rutas moleculares que

rigen a los organismos vivos (Fröhlich *et al.*, 2007; Jin *et al.*, 2014). Actualmente, se conocen 58 familias de FTs en plantas (Jin *et al.*, 2014; Naika *et al.*, 2013), de las cuales algunas de ellas cumplen funciones importantes en la regulación de la expresión génica, especialmente al inicio de la transcripción. Por ejemplo, la familia *bZIP*, caracterizada por poseer un zíper de leucinas, y la familia *bHLH*, distinguida por tener dos hélices alfa unidos como dímeros al DNA, se encuentran en todos los organismos eucariontes, regulando procesos fisiológicos y de desarrollo central (Ariel *et al.*, 2007; Sauvé *et al.*, 2004). Así, el objetivo de este trabajo bioinformático fue identificar familias de factores de transcripción (FTs) y genes enzimáticos diferencialmente expresados (DERs) en el xilema del tallo de Teca y de los que deriven posibles implicados en la regulación a la respuesta de estrés abiótico. Finalmente, es importante analizar, sistematizar e integrar los diversos datos de las redes de coexpresión génica y la información generada de diferentes experimentos de expresión génica.

## MATERIALES Y MÉTODOS

### Selección de genes de interés diferencialmente expresados en el xilema secundario del tallo de *Tectona grandis*

La búsqueda y selección de genes enzimáticos de interés relacionados con la respuesta a estrés abiótico y la regulación de las vías metabólicas de síntesis del xilema secundario del tallo, fue realizado a partir del transcriptoma de *T. grandis*, el cual está depositado en el “Transcriptome Shotgun Assembly” (TSA número de acceso GDLT00000000), disponible en el NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), obtenido por Galeano y colaboradores en el año 2015 (Galeano *et al.*, 2015). De esta base de datos, fueron descargados 2,413 genes diferencialmente expresados. Las funciones de estos genes fueron posteriormente anotadas mediante el software Blast2Go ([https://www.blast2go.com/blast2go-pro](https://www.blast2go.com/)).

### Identificación de factores de transcripción en *T. grandis* e identificación de ortólogos en *Arabidopsis thaliana*

Un total de 19 factores de transcripción de tipo basales, activadores, coactivadores y reguladores *maestros*, fueron seleccionados manualmente a partir de las anotaciones realizadas con el Blast2Go a los genes diferencialmente expresados de Teca. Posteriormente, la secuencia codificante de los 19 FTs de Teca fue utilizada como referencia en las bases de datos “TAIR” (<http://www.arabidopsis.org/>) y “plant TFTDB” (<http://planttftdb.cbi.pku.edu.cn/>), para seleccionar los FTs ortólogos en *A. thaliana* que tuvieran un porcentaje de identidad mayor a un 70%.

### Minería de literatura (*Text mining*)

Fue realizada una búsqueda de artículos relacionados con factores de transcripción y genes enzimáticos de *A. thaliana* ortólogos a los genes de Teca, que estuvieran implicados experimentalmente en la regulación de las vías metabólicas relacionadas con defensa de plantas, respuestas a estrés y síntesis de xilema secundario. Para ello, se utilizó la herramienta basada

en internet para minería de literatura denominada PubTator ([www.PubTator/index.cgi](http://www.PubTator/index.cgi)) (Wei *et al.*, 2012). Así, en esta búsqueda avanzada se utilizaron como palabras clave los códigos de los genes y proteínas establecidos por el NCBI. Además, se utilizaron los descriptores biológicos “gen ortólogo”, “factor de transcripción” y “genética vegetal” como filtros para disminuir el número de publicaciones. Asimismo, fueron rastreadas bases de datos en *A. thaliana* con relación a redes de coexpresión (<http://atted.jp/>) y respuesta de estrés abiótico (<http://caps.ncbs.res.in/stifdb/browse.html#genename>). Se utilizaron las bases de datos de *Arabidopsis* por ser una planta modelo, con numerosos recursos bioinformáticos disponibles.

### Dendrogramas y dominios conservados de los factores de transcripción diferencialmente expresados de *T. grandis*

La identificación de las secuencias codificantes para proteína (del inglés *Coding sequence*, o CDS) fueron obtenidas con el uso del programa ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi>) y la traducción de nucleótidos a aminoácidos del CDS fue realizado con el programa ExPasy (<http://web.expasy.org/translate/>). Posteriormente, la búsqueda de similitud proteica y de dominios se realizó con la base de datos de pFam (<http://pfam.xfam.org/>). De los 19 factores de transcripción identificados entre los genes diferencialmente expresados, se seleccionaron de 4 a 8 genes ortólogos con porcentaje de identidad mayor a 70% correspondientes a cada FTs. Luego, se realizaron alineamientos múltiples para cada FTs con el programa Clustal Omega y edición de secuencias mediante el programa Bioedit. Los alineamientos fueron almacenados en formato “.aln” y posteriormente transformados al formato “.meg” con el uso del software MEGA6, versión 6.06 (<http://www.megasoftware.net/mega.php>). Finalmente, con este programa se obtuvo un dendrograma según el método de Neighbor Joining.

### Análisis de redes de interacción entre *A. thaliana* y *T. grandis*

Para obtener la red de interacciones (tipo coexpresión, interacciones físicas, predicciones y dominios de proteína compartidos) de los factores de transcripción de *T. grandis*, se utilizaron como referencia los 19 factores de transcripción de *A. thaliana* ortólogos a Teca. Para obtener el interactoma de Teca, se utilizó el complemento Genemania en el software Cytoscape (<http://www.cytoscape.org/>), el cual utiliza la información disponible sobre la regulación génica de *A. thaliana* (como especie modelo) para predecir y ajustar las gráficas de interacción de otras especies de plantas, de las cuales no se conoce su regulación. Genemania utiliza el principio de redes de peso (del inglés *weight networks*) para ponderar las predicciones y calcular el coeficiente de correlación de Pearson para cada par de genes.

### Enriquecimiento funcional basado en categorías de ontología génica

El análisis de enriquecimiento funcional basado en categorías de ontología génica (del inglés, *Gene Ontology*) de los 19 FTs

y 301 genes enzimáticos de *A. thaliana* fue realizado mediante la herramienta web de ontología génica atriGO (<http://bioinfo.cau.edu.cn/agriGO/>) y el complemento “Bingo” del software Cytoscape (<http://www.BiNGO/Home.html>). El sondeo de enriquecimiento por minería de datos se realizó a un valor  $p$  de corte  $\leq 0,01$  después de aplicar la corrección Benjamini - Hochberg.

## RESULTADOS

### Genes diferencialmente expresados obtenidos del xilema secundario del tallo de *T. grandis* y minería de literatura

Se seleccionaron 19 FTs de *T. grandis* que pertenecen a 13 familias, incluyendo *bHLH*, *MyB*, *HSF*, *NAC*, *Mad box*, *bZIP*, *ARF*, *ERF*, *NY*, *IIITFs*, *IITFs*, *gata zinc*, *Tsiid* (Archivos adicionales 1) y 301 genes enzimáticos (Archivos adicionales 2), en donde los genes ortólogos comparten regiones de dominio funcional por encima del 70% de similitud. Por otro lado, se encontraron 1,641 publicaciones relacionadas con investigaciones de respuesta a factores de estrés usando como modelo *A. thaliana* depositados en PubTator. Posteriormente, se obtuvieron 1,683 FTs distribuidos en 19 familias relacionados con estrés biótico y abiótico, regulando la expresión de 4,172 genes con funciones de señalización de ácido abscísico, estrés por frío y sequía, respuesta a la luz, variaciones en las concentraciones de sales celulares, estrés oxidativo y rehidratación, entre otros (Sowdhamini *et al.*, 2009). Trabajos recientes en árboles como *Populus trichocarpa* y *Pinus taeda* han aumentado la información sobre las familias de factores de transcripción inmersos en procesos de respuesta a estrés abiótico y de biosíntesis de madera.

### Dendrograma para la determinación de familias y motivos conservados de FTs de *T. grandis*

Los FTs de Teca con ortólogos en *A. thaliana* poseen dominios funcionales involucrados en la regulación de la expresión génica, y con procesos de estrés abiótico (Tabla I). Así, la familia de FTs *HSF*, clases A y B, es una de las familias encontradas más abundantes que posee el motivo de localización nuclear y mantiene la conformación de giro de hélice para lograr la interacción con el DNA (Harrison *et al.*, 1994). Este dominio sirve para regular la respuesta al calor y otros tipos de estrés ambiental, siendo la planta perenne *Mimulus guttatus* la más próxima a *T. grandis* en motivos conservados (Figura 1).

Dentro de las familias *HSF*, *RAP*, *ARF*, *bHLH*, *MyB*, *NF-Y* (FTs que regulan la síntesis de pared celular y estrés abiótico), los genes más relevantes fueron *TgMyB1*, *TgMyB2*, *TgMyB3*, *TgHSF1*, *TgERF1*, *TgERF2*, *TgNuy1*, *TgNuy2* y *TgNuy3*.

El gen *TgMyB1* correspondió con *Myb-like* de *M. guttatus* (mgv1a011009m) y *P. trichocarpa* (Potri.006G000800.5) (Figura 1), todos ellos conteniendo la secuencia conservada LHEQLE (Tabla I). El gen *TgMyB2* correspondió con *Myb-like* de *M. guttatus* (mgv1a000686m) y *P. trichocarpa*

FTs <i>Tectona grandis</i>	Motivos Conservados
<i>Tg MyB1</i>	QITEALKLQMEVQKRLHEQLEVQRQLQLRIDAQGKYLKKIIEEQQ
<i>Tg MyB2</i>	IVKGPWSKEED
<i>Tg MyB3</i>	DQEEDLIIRLHKLLGNRWSLIA
<i>Tg HSF1</i>	FKHNNFSSFVRQLNTYGFKK
<i>Tg HSF2</i>	FARDLLPKYFKHNNFSSFVRQLNTYGFRKVVPRWFEFSND
<i>Tg HShT1</i>	KRRLT
<i>Tg ERF1</i>	KEEFVHLRRQSTGFSRGSSEYKRGVTLHKCGRWEARMGQFLGKK
<i>Tg ERF2</i>	RKRKNQYRGIRQRPWGKWAAEIRD
<i>Tg bZIP1</i>	SNRESARRSRLRKQKHL
<i>Tg NAC1</i>	EWFFFCPRDRKYPNG
<i>Tg bHLH1</i>	LNHVEAERQRREKLNQRFYALRAVVPNISKMDKASLLGDA
<i>Tg Gata1</i>	LCNACGLMWANKG
<i>Tg Nuy1</i>	REQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFITGEASDKCQREKRKTINGDDLLWAMTTLGFE
<i>Tg Nuy2</i>	SRHLHAMRRPRGNGGRFL
<i>Tg Nuy3</i>	RAKAELEKK
<i>Tg Nuy4</i>	RKPYLHESRHLHA
<i>Tg IETFs1</i>	DEYFHCE
<i>Tg ARF1</i>	RGQPRRHLLTTGWSTFVTSKRL

**Tabla I. Motivos conservados evolutivamente de 18 factores de transcripción de *T. grandis*, pertenecientes a las familias MyB, HSF, HShT, ERF, bZIP, NAC, bHLH, Gata, Nuy, IETFs y ARF.**

(Potri.006G241700.3), *TgMyB3* con *Myb-like* de *M. guttatus* (mgv1a020950m), *Jatropha curcas* (Jcr4S00199.80) y *Prunus persica* (ppa024533m) (Figura 1). *TgHSF1*, el cual posee dominios con interacción al DNA (Tabla I) mostró homología con *M. guttatus* (mgv1a022013m), *P. trichocarpa* (Potri.004G042600.1) y *A. thaliana* (AT4G36990.1). También, el gen *TgERF1*, caracterizado por regiones de aminoácidos que reconocen la secuencia GCCGCC en plantas, está relacionado con *M. guttatus* (mgv1a021246m) y *T. aestivum* (Tae060207), y el FTs *TgERF2* con *M. guttatus* (mgv1a008553m) y *Nicotiana tabacum* (Nta002851). Además, la familia NF-Y presentó para el gen *TgNuy1*, homología con *Utricularia gibba* (Scf00915.g22704.t1) y *M. guttatus* (mgv1a018555m), *TgNuy2* con *M. guttatus* (mgv1a010361m) y *Vitis vinifera* (GSVIVT01022601001), *TgNuy3* con *N. tabacum* (Nta011673) y *M. guttatus* (mgv1a012967m).

Asimismo, analizando los dominios conservados de la familia de factores de transcripción MYB, se observa que los genes *TgMyB2* y *TgMyB3* mantienen el dominio de unión al DNA (Tabla I), teniendo patrones estructurales tipo hélice-giro-hélice y tres triptófanos para la formación de un núcleo hidrofóbico (Ogata *et al.*, 1995). Por otro lado, las regiones conservadas del gen *TgHSF1* (familia HSF, o *Heat shock*) poseen dominios de interacción proteína-proteína (Figura 2), también tipo hélice-giro-hélice (Kotak *et al.*, 2004). Además,

el gen *TgHSF2* posee un dominio de interacción al DNA, indicando una mayor especificidad al DNA (Tabla I). El gen *TgERF2* (de la familia *ethylene response factor*) posee un dominio rico en arginina relacionado con la unión al esqueleto azúcar-fosfato (Allen *et al.*, 1998). También fueron encontrados los FTs *TgbZIP1* (familia *bZIP*), *TgbHLH1* (familia *bHLH*), factores de transcripción basales que mantienen un zíper de leucinas y de reconocimiento de promotores corriente arriba y corriente abajo, donde el dominio de *TgbHLH1* se une a los promotores formando dímeros (Schumacher *et al.*, 2000). Asimismo, fueron identificados los genes *TgNuy1*, *TgNuy2*, *TgNuy3*, *TgNuy4* ricos en Gln, Ser, Thr y pertenecientes a la familia NF-YA, con dominios de interacción en el surco menor de DNA y el motivo CCAAT (Nardini *et al.*, 2013). Finalmente, los dominios conservados de la familia de factores de transcripción NAC se caracterizan por poseer un dominio conservado NAC N-terminal. En este estudio, fue encontrado el gen *TgNAC1* (Tabla I). Los genes de esta familia pueden interactuar y regular con otros FTs y al mismo tiempo unirse al DNA, no poseen motivos hélice-giro-hélice y generalmente la estructura del monómero de dominio es  $\beta$ -hoja antiparalela y doblada (Nole-Wilson & Krizek, 2000). Para los dominios conservados del gen *TgARF1* se ha observado que los dímeros del motivo ARF se unen como pinzas moleculares y de reconocimiento específico en los promotores corriente arriba o corriente abajo (Boer *et al.*, 2014).



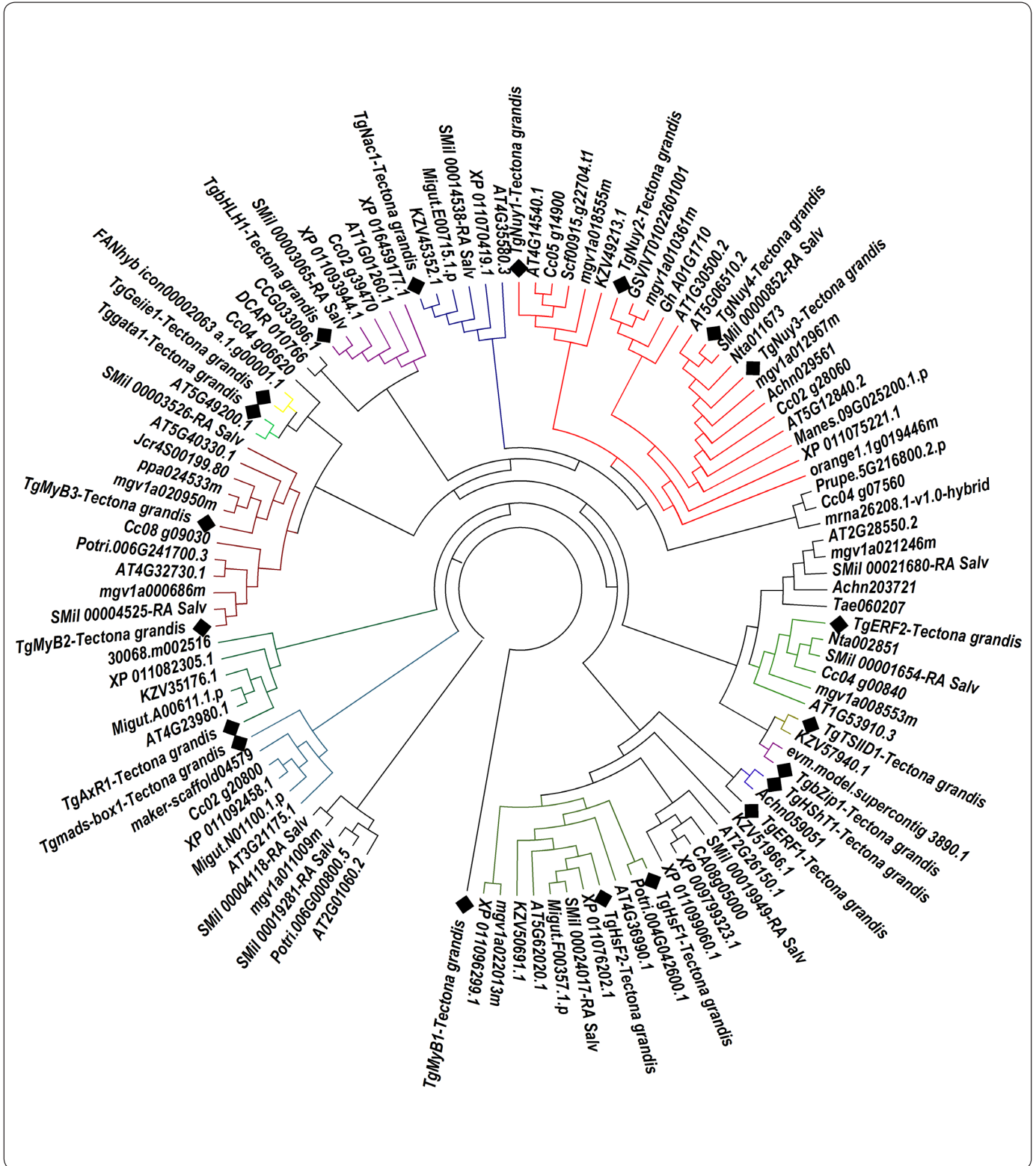


Figura 1. Árbol filogenético de las familias de los factores de transcripción diferencialmente expresados en xilema secundario de *T. grandis*, (bHLH, MyB, HSF, NAC, Mad box, bZIP, ARF, ERF, NY, IIBTFs, IIEtfs, gata zinc y Tsiid), son representados con los cuadrados de color negro. Los clúster representan las asociaciones de secuencias de aminoácidos de FTs con un porcentaje de similitud superior al 70%. El clado con más FTs es constituido por los FTs *TgNuy1*, *TgNuy2*, *TgNuy3* y *TgNuy4*, la familia MyB y HSF tienen 3 factores de transcripción, del mismo modo son detalladas las demás familias de genes reguladores en *Tectona grandis*.

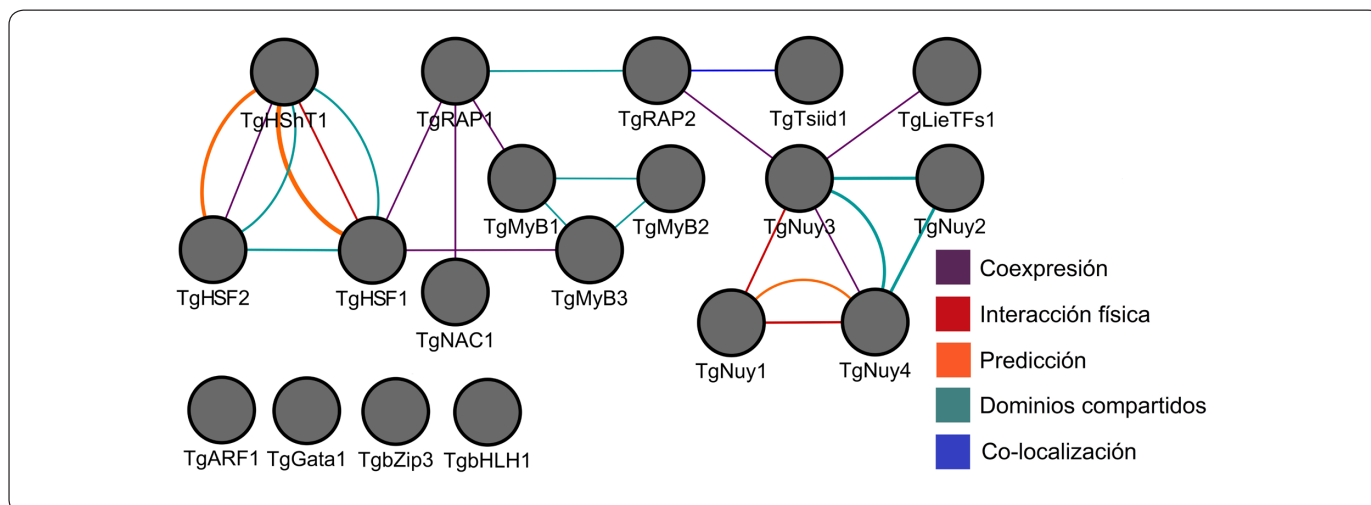


Figura 2. Interactoma de 19 FTs diferencialmente expresados en tejido xilemático de tallo en *T. grandis*. Se incluye la red de coexpresión, donde los FTs *TgRAP1*, *TgRAP2*, *TgMyB1*, *TgHsf1*, *TgHShT1*, *TgHsf2*, *TgMyB3*, *TgNAC1*, *TgLieTFs1*, *TgNuy3* y *TgNuy4* se interconectan para hacer emerger propiedades funcionales de las células. Asimismo, se muestran interacciones físicas entre los FTs *TgHShT1-TgHsf1* y *TgNuy3-TgNuy4-TgNuy1*. Además, se presentan diferentes redes como predicciones, dominios de proteína compartidos y otros tipos de interacción.

#### Análisis topológico de la red de coexpresión de FTs del xilema secundario asociado a la respuesta de estrés abiótico en *T. grandis*

Los genes de *A. thaliana* en la red de coexpresión se dividen en grupos de acuerdo a su función (Figura 3), (Tabla II). De este modo, se identificaron genes relacionados a la respuesta de estrés abiótico. Se seleccionaron 301 genes enzimáticos y 19 FTs con un porcentaje de similitud superior al 70 %, de los cuales 10 son de respuesta a estrés por calor, 16 responden a estrés por luz y temperatura, mientras que 108 genes están relacionados con procesos metabólicos celulares (Tabla II).

Por otro lado, los FTs *TgHsf1*, *TgHsf2* y *TgHShT1* responden a estrés de calor en *T. grandis*, existiendo interacciones físicas entre los nodos de *TgHShT1* con *TgHsf1* y *TgNuy4-TgNuy3* con *TgNuy1* (Figura 2). Extendiendo el modelo de predicción, se observa la interacción física de *TgHsf2* con *TgHShT1* y *TgHsf1*. La red de coexpresión entre nodos e interacción por aristas (Figura 2) están entrelazadas por los FTs *TgRAP1*, *TgMyB1*, *TgHsf1*, *TgMyB3*, *TgNAC1*, *TgTsiid1*, *TgLieTFs1*, *TgNuy3*, *TgRAP2* y *TgNuy4*, integrando la información parcial sobre 15 FTs. Entre las familias de FTs de Teca existen 4 agrupaciones que mantienen dominios funcionales compartidos entre los genes, teniendo como primer grupo *TgHsf2*, *TgHShT1*, *TgHsf1*, luego *TgMyB1*, *TgMyB2*, *TgMyB3*, siguiéndole *TgNuy2*, *TgNuy3*, *TgNuy4* y finalmente *TgRAP1*, *TgRAP2* (Figura 1) y (Figura 2).

#### Análisis de enriquecimiento funcional en genes de Teca

La obtención de las ontologías génicas se realizó con 19 factores de transcripción y 301 genes enzimáticos ortólogos a genes

de *A. thaliana* involucrados en la regulación de respuesta a estrés abiótico y en la formación de pared celular. De ellos, 108 de los genes diferencialmente expresados tienen relación directa con procesos celulares y metabólicos (Figura 3) y 32 genes dan respuesta a estrés abiótico, calor, estrés oxidativo y alta intensidad de luz. Los genes *TgHShT1*, *TgHsf1*, *TgHsf2* son posibles factores de transcripción de respuesta a estrés abiótico que regulan positivamente otros genes de *T. grandis* (Figura 2). Además, de los datos obtenidos, fueron detectados genes de FTs maestros (*master regulators*) que regulan el estrés abiótico en *T. grandis*. Entre ellos, fue encontrado el regulador *master TgNAC1*, involucrado en la regulación negativa de algunas funciones y ortólogo de *NTL9* de *A. thaliana* (Yoon *et al.*, 2008), quien responde a estrés osmótico y senescencia foliar. Asimismo, el FT *TgHShT1* es ortólogo de *AtHsfA2* y un regulador clave en la respuesta a calor inducida bajo estrés ambiental. El gen *TgHShT1* regula la transcripción de varios genes implicados en defensa frente al estrés térmico, incrementando su transcripción rápidamente bajo señales de estrés oxidativo, y su inhibición conlleva a la disfunción de diferentes organelos y muerte celular (Liu *et al.*, 2013; Nishizawa-Yokoi *et al.*, 2010) (Archivos adicionales 3). Además, el gen *TgHShT1* posee interacción con 14 genes ortólogos en Teca y de respuesta a estrés oxidativo (Figura 2). Asimismo, los genes *TgERF1* y *TgERF2* responden a procesos biológicos de crecimiento y desarrollo, así como a señales hormonales y respuesta a estrés biótico y abiótico. El gen *TgMyB3* es ortólogo de *MyB23* de *A. thaliana*, quien regula su propio promotor y está relacionado con especialización celular epidérmica y formación de tricomas (Kang *et al.* 2009; Kirik *et al.* 2001).



GO-ID	Descripción	Valor p	Valor p Corregido	Frecuencia del clúster	Porcentaje
15979	Fotosíntesis	3.16E-09	2.85E-06	12/301	5.00%
9628	Respuesta al estímulo abiótico	9.69E-07	3.13E-04	32/301	13.40%
8152	Procesos metabólicos	1.09E-06	3.13E-04	108/301	45.30%
9408	Respuesta al calor	1.48E-06	3.13E-04	10/301	4.20%
9266	Respuesta a estímulos de temperatura	1.73E-06	3.13E-04	16/301	6.70%
9416	Respuesta a estímulos de luz	3.35E-05	4.19E-03	16/301	6.70%
9314	Respuesta a radiación	5.05E-05	4.19E-03	16/301	6.70%
9987	Procesos celulares	5.36E-05	4.19E-03	108/301	45.30%
50896	Respuesta a estímulos	5.50E-05	4.19E-03	57/301	23.90%
19684	Fotosíntesis, reacción a la luz	5.67E-05	4.19E-03	6/301	2.50%
44271	Proceso de biosíntesis celular de compuesto de nitrógeno	6.00E-05	4.19E-03	14/301	5.80%
30091	Reparación de proteínas	6.46E-05	4.19E-03	3/301	1.20%
10206	Reparación de fotosistema	6.46E-05	4.19E-03	3/301	1.20%
9765	Fotosíntesis, captador de luz	6.56E-05	4.19E-03	4/301	1.60%
44281	Pequeña molécula de proceso metabólico	6.95E-05	4.19E-03	29/301	12.10%
44237	Proceso metabólico celular	8.00E-05	4.51E-03	84/301	35.20%
51188	Proceso de biosíntesis del cofactor	1.03E-04	5.49E-03	8/301	3.30%
44238	Proceso metabólico primario	2.06E-04	1.03E-02	86/301	36.10%
5975	Proceso metabólico de carbohidratos	2.93E-04	1.39E-02	20/301	4%
44283	Pequeña molécula de proceso biosintético	3.97E-04	1.79E-02	17/301	7.10%
5982	Metabolismo del almidón	6.33E-04	2.72E-02	4/301	1.60%
9108	Proceso de biosíntesis de coenzima	9.13E-04	3.75E-02	5/301	2.10%
9644	Respuesta a la alta intensidad de la luz	1.13E-03	4.42E-02	4/301	1.60%

Tabla II. Nodos de ontologías génicas más significativos. Son incluidos sus valores p y frecuencias de clúster, donde las descripciones con mayor porcentaje son representados por las actividades metabólicas y procesos celulares.

Por otro lado, dentro de los genes enzimáticos, se identificó el transcrito de  $\beta$ -galactosidasa fundamental en la síntesis de carbohidratos, glicoproteínas y glicolípidos (Roach *et al.*, 2011). También se encontraron genes enzimáticos como quinasas, esenciales en las vías de señalización para el crecimiento, desarrollo y percepción a las diversas respuestas de estrés ambiental en plantas (Ihnatowicz *et al.*, 2008). Además, se encontraron genes que codifican enzimas relacionados con la actividad fotosintética y en procesos de fosforilación y activación de FTs a nivel de la membrana intracelular, regulados por reacciones químicas hormonales como el etileno y el ácido abscísico (Abe *et al.*, 2003; Yoon *et al.*, 2008; Zhou *et al.*, 2014). La respuesta fisiológica de la planta a los cambios ambientales es dinámica a nivel de la expresión génica. En ese punto, los FTs *maestros* son seleccionados evolutivamente por las plantas para tener funciones esenciales en la corregulación de procesos metabólicos como la xilogénesis y adaptación a estrés ambiental (Banti *et al.*, 2010; Lin *et al.*, 2015), fenómenos que se evidencian analizando las redes de coexpresión, para describir los patrones de correlación entre los genes y lograr identificar biomarcadores (Consortium, 2000).

Así, para activar los procesos metabólicos es necesario que los motivos funcionales de los FTs interactúen físicamente con los promotores *upstream* para su regulación y generalmente interactúan físicamente entre ellos para regular la expresión génica (Lin *et al.*, 2015; Obata & Fernie, 2012). Finalmente, es posible que existan diferencias en los motivos funcionales entre genes ortólogos de familias de factores de transcripción, que podrían explicar divergencias funcionales y en algunos casos se expresarían en variaciones fenotípicas (Carretero-Paulet *et al.*, 2010; Jin *et al.*, 2014).

## CONCLUSIONES

Con la información del transcriptoma de *T. grandis* fue posible establecer y visualizar redes de interacción de genes diferencialmente expresados en el xilema secundario del tallo, apoyados por la búsqueda de minería y bases de datos con enriquecimiento ontológico. Así, con la información disponible y utilizando el software Cytoscape (herramienta para analizar, modelar y predecir funciones génicas), se identificaron FTs implicados en procesos lignocelulósicos y respuesta a estrés abiótico.



Los patrones evolutivos de los 19 FTs analizados aquí se mantienen conservados dentro de las familias y sub familias. Además, son importantes en la regulación génica, ya que fueron identificados a partir de los genes diferencialmente expresados (DERs) en el tejido del xilema secundario del tallo. Debido a que las funciones particulares de una proteína se determinan analizando su secuencia de residuos de aminoácidos, la evaluación de dominios conservados funcionales mediante genes ortólogos de *T. grandis* en *A. thaliana* es útil para identificar posibles funciones del gen (Consortium, 2000; Dameron et al., 2013), siendo fundamental hacer uso de esta especie modelo, ya que tiene abundante recurso bioinformático disponible. También, la base de datos PubMed fue útil para obtener publicaciones que documentan evidencia experimental encontradas con el programa pubtator para filtrar la información y optimizar la sistematización de datos, la predicción de genes y el soporte experimental. Finalmente, este estudio bioinformático aumenta los estudios funcionales de genes de interés en *T. grandis*, 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.

## REFERENCIAS

- Abe, H., Urao, T., Ito, T., Seki, M., & Shinozaki, K. (2003). Transcriptional Activators in Abscisic Acid Signaling. *Society*, 15(January), 63–78. <https://doi.org/10.1105/tpc.006130.salt>
- Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., & Suzuki, M. (1998). A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *The EMBO Journal*, 17(18), 5484–96. <https://doi.org/10.1093/emboj/17.18.5484>
- Anish, M. C., Anoop, E. V., Vishnu, R., Sreejith, B., & Jijeesh, C. M. (2015). Effect of growth rate on wood quality of teak (*Tectona grandis* L. f.): a comparative study of teak grown under differing site quality conditions. *Journal of the Indian Academy of Wood Science*, 12(1), 81–88. <https://doi.org/10.1007/s13196-015-0147-1>
- Ariel, F. D., Manavella, P. A., Dezar, C. A., & Chan, R. L. (2007). The true story of the HD-Zip family. *Trends in Plant Science*, 12(9), 419–426. <https://doi.org/10.1016/j.tplants.2007.08.003>
- Banti, V., Mafessoni, F., Loreti, E., Alpi, A., & Perata, P. (2010). The Heat-Inducible Transcription Factor HsfA2 Enhances Anoxia Tolerance in *Arabidopsis*. *Plant Physiology*, 152(3), 1471–1483. <https://doi.org/10.1104/pp.109.149815>
- Boer, D. R., Freire-Ríos, A., Van Den Berg, W. A. M., Saaki, T., Manfield, I. W., Kepinski, S., Coll, M. (2014). Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell*, 156(3), 577–589. <https://doi.org/10.1016/j.cell.2013.12.027>
- Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martínez-García, J. F., Bilbao-Castro, J. R., & Robertson, D. L. (2010). Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, poplar, rice, moss, and algae. *Plant Physiology*, 153(3), 1398–412. <https://doi.org/10.1104/pp.110.153593>
- Consortium, T. G. O. (2000). Gene ontologie: Tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556.Gene>
- Dameron, O., Bettembourg, C., & Le Meur, N. (2013). Measuring the Evolution of Ontology Complexity: The Gene Ontology Case Study. *PLoS ONE*, 8(10), 1–18. <https://doi.org/10.1371/journal.pone.0075993>
- Dharmawardhana, P., Brunner, A. M., & Strauss, S. H. (2010). Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. *BMC Genomics*, 11, 150. <https://doi.org/10.1186/1471-2164-11-150>
- Diningrat, D. S., Widiyanto, S. M., Pancoro, a., . I., Shim, D., Panchangam, B., Carlson, J. E. (2015). Transcriptome of Teak (*Tectona grandis*, L.f) in Vegetative to Generative Stages Development. *Journal of Plant Sciences*, 10(1), 1–14. <https://doi.org/10.3923/jps.2015.1.14>
- Fröhlich, H., Speer, N., Poustka, A., & Beissbarth, T. (2007). GOSim—an R-package for computation of information theoretic GO similarities between terms and gene products. *BMC Bioinformatics*, 8(1), 166. <https://doi.org/10.1186/1471-2105-8-166>
- Galeano, E., Vasconcelos, T. S., Ramiro, D. A., De Martin, V. D. F., & Carrer, H. (2014). Identification and validation of quantitative real-time reverse transcription PCR reference genes for gene expression analysis in teak (*Tectona grandis* L.f.). *BMC Research Notes*, 7(1), 464. <https://doi.org/10.1186/1756-0500-7-464>
- Galeano, E., Vasconcelos, T. S., Vidal, M., Mejía-Guerra, M. K., & Carrer, H. (2015). Large-scale transcriptional profiling of lignified tissues in *Tectona grandis*. *BMC Plant Biology*, 15(1), 221. <https://doi.org/10.1186/s12870-015-0599-x>
- Gill, S. S., & Tuteja, N. (2010). Polyamines and abiotic stress tolerance in plants. *Plant Signaling & Behavior*, 5(1), 26–33. <https://doi.org/10.4161/psb.5.1.10291>
- Harrison, C. J., Bohm, A. A., & Nelson, H. C. (1994). Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science*, 263(5144), 224–227. <https://doi.org/10.1126/science.8284672>
- Ihnatowicz, A., Pesaresi, P., Lohrig, K., Wolters, D., Müller, B., & Leister, D. (2008). Impaired photosystem I oxidation induces STN7-dependent phosphorylation of the light-harvesting complex I protein Lhca4 in *Arabidopsis thaliana*. *Planta*, 227(3), 717–722. <https://doi.org/10.1007/s00425-007-0650-8>
- Jin, J., Zhang, H., Kong, L., Gao, G., & Luo, J. (2014). PlantTFDB 3.0: A portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research*, 42(D1), 1182–1187. <https://doi.org/10.1093/nar/gkt1016>
- Kang, Y. H., Kirik, V., Hulskamp, M., Nam, K. H., Hagely, K., Lee, M. M., & Schiefelbein, J. (2009). The MYB23 gene provides a positive feedback loop for cell fate specification in the *Arabidopsis* root epidermis. *The Plant Cell*, 21(4), 1080–1094. <https://doi.org/10.1105/tpc.108.063180>
- Kirik, V., Schnittger, A., Radchuk, V., Adler, K., Hulskamp, M., & Baumlein, H. (2001). Ectopic expression of the *Arabidopsis* AtMYB23 gene induces differentiation of trichome cells. *Dev Biol*, 235(2), 366–377. <https://doi.org/10.1006/dbio.2001.0287>
- Kotak, S., Port, M., Ganguli, A., Bicker, F., & Von Koskull-Döring, P. (2004). Characterization of C-terminal domains of *Arabidopsis* heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. *Plant Journal*, 39(1), 98–112. <https://doi.org/10.1111/j.1365-313X.2004.02111.x>
- Krause, G. H., Gallé, A., Virgo, A., García, M., Bucic, P., Jahns, P., & Winter, K. (2006). High-light stress does not impair

- biomass accumulation of sun-acclimated tropical tree seedlings (*Calophyllum longifolium* Willd. and *Tectona grandis* L. f.). *Plant Biology*, 8(1), 31–41. <https://doi.org/10.1055/s-2005-872901>
- Lin, L., Lucas, M. De, Turco, G., Toal, T. W., Gaudinier, A., Young, N. F., Brady, S. M. (2015). An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature*, 517(7536), 571–575. <https://doi.org/10.1038/nature14099>
- Liu, J., Sun, N., Liu, M., Liu, J., Du, B., Wang, X., & Qi, X. (2013). An Autoregulatory Loop Controlling *Arabidopsis* HsfA2 Expression: Role of Heat Shock-Induced Alternative Splicing. *Plant Physiology*, 162(1), 512–521. <https://doi.org/10.1104/pp.112.205864>
- Matsui, A., Ishida, J., Morosawa, T., Mochizuki, Y., Kaminuma, E., Endo, T. A., Seki, M. (2008). *Arabidopsis* transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant and Cell Physiology*, 49(8), 1135–1149. <https://doi.org/10.1093/pcp/pcn101>
- Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F., & Myburg, A. a. (2010). De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. *BMC Genomics*, 11(1), 681. <https://doi.org/10.1186/1471-2164-11-681> [pii]
- Naika, M., Shameer, K., Mathew, O. K., Gowda, R., & Sowdhamini, R. (2013). STIFDB2: An updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in *Arabidopsis* and rice. *Plant and Cell Physiology*, 54(2), 1–15. <https://doi.org/10.1093/pcp/pcs185>
- Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., ... Mantovani, R. (2013). Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell*, 152(1–2), 132–143. <https://doi.org/10.1016/j.cell.2012.11.047>
- Nishizawa-Yokoi, A., Tainaka, H., Yoshida, E., Tamoi, M., Yabuta, Y., & Shigeoka, S. (2010). The 26S proteasome function and Hsp90 activity involved in the regulation of HsfA2 expression in response to oxidative stress. *Plant and Cell Physiology*, 51(3), 486–496. <https://doi.org/10.1093/pcp/pcq015>
- Nole-Wilson, S., & Krizek, B. A. (2000). DNA binding properties of the *Arabidopsis* floral development protein AINTEGUMENTA. *Nucleic Acids Res*, 28(21), 4076–4082. <https://doi.org/DOI10.1093/nar/28.21.4076>
- Obata, T., & Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences*, 69(19), 3225–3243. <https://doi.org/10.1007/s00018-012-1091-5>
- Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Sarai, A. (1995). Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Nature Structural Biology*, 2(4), 309–20. <https://doi.org/10.1038/nsb0495-309>
- Prasch, C. M., & Sonnewald, U. (2013). Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. *Plant Physiology*, 162(4), 1849–66. <https://doi.org/10.1104/pp.113.221044>
- Quiala, E., Cañal, M. J., Rodríguez, R., Yagüe, N., Chávez, M., Barbón, R., & Villedor, L. (2012). Proteomic profiling of *Tectona grandis* L. leaf. *Proteomics*, 12(7), 1039–1044. <https://doi.org/10.1002/pmic.201100183>
- Riechmann, J. L. (2000). *Arabidopsis* Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science*, 290(5499), 2105–2110. <https://doi.org/10.1126/science.290.5499.2105>
- Roach, M. J., Mokshina, N. Y., Badhan, A., Snegireva, A. V., Hobson, N., Deyholos, M. K., & Gorshkova, T. A. (2011). Development of Cellulosic Secondary Walls in Flax Fibers Requires -Galactosidase. *Plant Physiology*, 156(3), 1351–1363. <https://doi.org/10.1104/pp.111.172676>
- Sauvé, S., Tremblay, L., & Lavigne, P. (2004). The NMR solution structure of a mutant of the max b/HLH/LZ free of DNA: Insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. *Journal of Molecular Biology*, 342(3), 813–832. <https://doi.org/10.1016/j.jmb.2004.07.058>
- Schumacher, M. A., Goodman, R. H., & Brennan, R. G. (2000). The structure of a CREB bZIP somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *Journal of Biological Chemistry*, 275(45), 35242–35247. <https://doi.org/10.1074/jbc.M007293200>
- Scoville, A. G., Barnett, L. L., Bodbyl-roels, S., Kelly, J. K., & Lena, C. (2012). NIH Public Access, 191(1), 251–263. <https://doi.org/10.1111/j.1469-8137.2011.03656.x>. Differential
- Sowdhamini, R., Shameer, K., Ambika, S., Varghese, S. M., Karaba, N., & Udayakumar, M. (2009). STIFDB *Arabidopsis* stress responsive transcription factor dataBase. *International Journal of Plant Genomics*, 2009. <https://doi.org/10.1155/2009/583429>
- Sundar, A. S., Varghese, S. M., Shameer, K., Karaba, N., Udayakumar, M., & Sowdhamini, R. (2008). STIF: Identification of stress-upregulated transcription factor binding sites in *Arabidopsis thaliana*. *Bioinformatics*, 2(10), 431–437. <https://doi.org/10.6026/97320630002431>
- Tominaga-Wada, R., Nukumizu, Y., Sato, S., Kato, T., Tabata, S., & Wada, T. (2012). Functional Divergence of MYB-Related Genes, *WEREWOLF* and *AtMYB23* in *Arabidopsis*. *Bioscience, Biotechnology, and Biochemistry*, 76(5), 883–887. <https://doi.org/10.1271/bbb.110811>
- Udvardi, M. K., Kakar, K., Wandrey, M., Montanari, O., Murray, J., Andrianakaja, A., Town, C. D. (2007). Legume Transcription Factors: Global Regulators of Plant Development and Response to the Environment. *Plant Physiology*, 144(2), 538–549. <https://doi.org/10.1104/pp.107.098061>
- Wan, L., Yan, X., Chen, T., & Sun, F. (2012). Modeling RNA degradation for RNA-Seq with applications. *Biostatistics*, 13(4), 734–747. <https://doi.org/10.1093/biostatistics/kxs001>
- Wei, C.-H., Harris, B. R., Li, D., Berardini, T. Z., Huala, E., Kao, H.-Y., & Lu, Z. (2012). Accelerating literature curation with text-mining tools: a case study of using PubTator to curate genes in PubMed abstracts. *Database : The Journal of Biological Databases and Curation*, 2012, bas041. <https://doi.org/10.1093/database/bas041>
- Yoon, H. K., Kim, S. G., Kim, S. Y., & Park, C. M. (2008). Regulation of leaf senescence by NTL9-mediated osmotic stress signaling in *Arabidopsis*. *Mol. Cells*, 25(3), 438–445.
- Zhou, J., Yang, Y., Yu, J., Wang, L., Yu, X., Ohtani, M., Zhuge, Q. (2014). Responses of *Populus trichocarpa* galactinol synthase genes to abiotic stresses. *Journal of Plant Research*, 127(2), 347–358. <https://doi.org/10.1007/s10265-013-0597-8>