



Centro de Investigación Científica de Yucatán, A. C.

Posgrado en Ciencias Biológicas

**ANÁLISIS FUNCIONAL DE LOS GENES TMBIM DE
PAPAYA (*CARICA PAPAYA* L.) EN LA
TOLERANCIA AL ESTRÉS ABIÓTICO**

Tesis que presenta

SAMUEL DAVID GAMBOA TUZ

En opción al título de

DOCTOR EN CIENCIAS

(Ciencias Biológicas: Opción Biotecnología)

Mérida, Yucatán, México

2020

CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A.C.

POSGRADO EN CIENCIAS BIOLÓGICAS



RECONOCIMIENTO

Por medio de la presente hago constar que el trabajo de tesis de **Samuel David Gamboa Tuz** titulado “**Análisis funcional de los genes TMBIM de papaya (*Carica papaya* L.) en la tolerancia al estrés abiótico**” fue realizado en el Laboratorio de Fisiología Molecular del Estrés Abiótico y Transformación Genética en la Unidad de Biotecnología, en la Línea de Agrobiotecnología, del Centro de Investigación Científica de Yucatán, A.C. bajo la dirección del Dr. Luis Carlos Rodríguez Zapata, dentro de la opción de Biotecnología, perteneciente al Programa de Doctorado del Posgrado en Ciencias Biológicas de este Centro.

Atentamente:



Dra. Cecilia Hernández Zepeda
Directora de Docencia

Mérida, Yucatán, México, a 19 de mayo de 2020.

DECLARACIÓN DE PROPIEDAD

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



Samuel David Gamboa Tuz

Este trabajo se llevó a cabo en la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, en el que participé bajo la dirección del Dr. Luis Carlos Rodríguez Zapata.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por la beca otorgada (CVU 550706; Beca No.438058) para la realización de mis estudios de doctorado en este Centro. Además, gracias por el apoyo económico que me permitió asistir a un taller de bioinformática en el Centro de Ciencias Genómicas (CCG) de la Universidad Nacional Autónoma de México (UNAM) en Cuernavaca, Morelos en 2018; dicho apoyo me fue otorgado como parte del programa “Fortalecimiento e Internacionalización del Doctorado en Ciencias (Ciencias Biológicas) del CICY” (Clave FOMIX: 247355. Clave CICY: 603010001 FOMIX-D-CB FOMIX).

Al Centro de Investigación Científica de Yucatán, A.C. (CICY), gracias a todo el personal de la institución por su apoyo. Quiero agradecer especialmente a la coordinación de posgrado opción Biotecnología por el apoyo económico que me permitió asistir a un taller de Bioinformática en el CCG-UNAM en Cuernavaca, Morelos en 2019.

Al Dr. Luis Carlos Rodríguez Zapata, mi asesor de tesis, por permitirme ser parte de su equipo de trabajo en el Laboratorio de Fisiología Molecular del Estrés Abiótico y Transformación Genética de Plantas, por el conocimiento transmitido, por su tiempo y aprecio.

Al Dr. Alejandro Pereira Santana, por el conocimiento compartido y las colaboraciones.

A mi comité tutorial conformado por el Dr. Luis Carlos Rodríguez Zapata, el Dr. Manuel Martínez Estévez, el Dr. Raúl Tapia Tussell y el Dr. Lenin Sánchez Calderón. Gracias por sus observaciones y comentarios en cada evaluación.

A mi comité de evaluación predoctoral conformado por el Dr. Luis Carlos Rodríguez Zapata, el Dr. Manuel Martínez Estévez, el Dr. Raúl Tapia Tussell, el Dr. Lenin Sánchez Calderón, la Dra. Blondy Beatriz Canto Canché, la Dra. Ana Ly Arroyo Herrera y el Dr. Enrique Castaño de la Serna. Gracias por su retroalimentación.

A mi comité revisor de tesis y sínodo de examen doctoral conformado por el Dr. Luis Carlos Rodríguez Zapata, el Dr. Manuel Martínez Estévez, el Dr. Raúl Tapia Tussell, el Dr. Alejandro Pereira Santana, el Dr. José Ramón Pacheco Arjona, el Dr. Lorenzo Felipe Sánchez Teyer y el Dr. Enrique Castaño de la Serna. Gracias por sus observaciones y comentarios para mejorar el presente trabajo.

Al Q.F.B. Miguel Ángel Keb Llanes, técnico de laboratorio, por el apoyo brindado en el Laboratorio de Fisiología Molecular del Estrés Abiótico y Transformación Genética de Plantas.

A mis compañeros de laboratorio durante el doctorado: Jorge E., Gaby F., Antonio R., Ricardo O., Aarón G., Alejandro P., Alejandro Z., Sandi R., Marijo G., Roberto C., Carito S., Evelyn C., Ricardo O., Christian A., Osmar S., Jordán A., Karla R. Magdalena R., Ángela T. Gracias por su apoyo en el laboratorio, colaboraciones y mucho más. Fui afortunado de tenerlos como *labmates*.

A mis amigos de otros laboratorios y unidades del CICY: Gabi R., Areli R., Yessica B. y Rufino G. Gracias por los momentos que compartieron conmigo.

A mis padres Aida y David. Gracias por su apoyo.

ÍNDICE

RESUMEN.....	ii
ABSTRACT	iv
INTRODUCCIÓN.....	1
CAPÍTULO I	3
ANTECEDENTES	3
1.1 Estrés abiótico	3
1.1.1 Estrés salino	3
1.2 La superfamilia TMBIM	6
1.2.1 Apoptosis, la familia BCL-2 y el descubrimiento del gen BI-1 humano	6
1.2.2 Homólogos de BI-1: clasificación y evolución de la superfamilia TMBIM	11
1.2.3 Las proteínas TMBIM son canales de calcio dependientes de pH	15
1.2.5 Funciones de los genes BI-1 en plantas	18
1.2.6 Funciones de los genes LFG en plantas	23
1.2.7. Función dual de los genes TMBIM en animales, levaduras y plantas.....	23
1.2.8 Potencial biotecnológico de los genes TMBIM en plantas	24
1.3 <i>Carica</i> papaya L.	24
1.3.1 Biología de la planta de papaya.....	24
1.3.2 Fruto de papaya	25
1.3.3 Papaya y estrés abiótico	29
1.3.4 Genómica y transcriptómica de papaya.....	30
1.3.5 La superfamilia TMBIM en papaya.....	31
HIPÓTESIS	37
JUSTIFICACIÓN	39
OBJETIVO GENERAL	41
OBJETIVOS ESPECÍFICOS	43
ESTRATEGIA EXPERIMENTAL	45

CAPÍTULO II	47
NEW INSIGHTS INTO THE PHYLOGENY OF THE TMBIM SUPERFAMILY ACROSS THE TREE OF LIFE: COMPARATIVE GENOMICS AND SYNTENY NETWORKS REVEAL INDEPENDENT EVOLUTION OF THE BI AND LFG FAMILIES IN PLANTS	47
ABSTRACT	49
2.1 INTRODUCTION	53
2.2 MATERIALS AND METHODS.....	57
2.2.1 HMM search and retrieval of TMBIM protein sequences.....	57
2.2.2 Phylogenetic analysis pipeline	57
2.2.3 Motif discovery and HMMLogo analyses	58
2.2.4 Structural multiple sequence alignment	58
2.2.5 Determination of orthologous groups (OGs) in plants.....	59
2.2.6 Tandem gene analysis in plants	59
2.2.7 Synteny network approach in plants	60
2.2.8 Generation of species trees	61
2.3 RESULTS	63
2.3.1 Phylogeny overview of the TMBIM superfamily across the three domains of life	63
2.3.2 Sequence and motif conservation indicate a pH-regulated calcium channel activity in TMBIM proteins in eukaryotes and bacteria, but not in Archaea	67
2.3.3 Two major phylogenetic groups are found in the LFG family in plants	68
2.3.4 Two OGs of the TMBIM superfamily are shared by most plants	71
2.3.5 Syntenic network and tandem analyses of the TMBIM superfamily in plants ..	72
2.4 DISCUSSION	77
2.4.1 Eukaryotic BI and LFG families diverged early in prokaryotes	77
2.4.2 New insights into the TMBIM superfamily in fungi, and a possible archetype BI sequence	78
2.4.3 Two ancient transposition events of the BI family occurred in monocots	78

2.4.4 Expansion of the LFG family in plants and vertebrates	81
2.5 CONCLUSIONS	83
CAPÍTULO III	85
TRANSCRIPTOMICS AND CO-EXPRESSION NETWORKS REVEAL TISSUE-SPECIFIC RESPONSES AND REGULATORY HUBS UNDER MILD AND SEVERE DROUGHT IN PAPAYA (<i>CARICA PAPAYA</i> L.).....	85
ABSTRACT	87
3.1 INTRODUCTION	89
3.2 MATERIALS AND METHODS.....	95
3.2.1 Plant material and experimental design	95
3.2.2 Physiological measurements	95
3.2.3 RNA isolation and Illumina sequencing	95
3.2.4 De novo transcriptome assembly and annotation.....	96
3.2.5 Expression level quantification.....	96
3.2.6 Differential expression analysis.....	96
3.2.7 GO functional enrichment analysis of sets of differentially expressed genes .	97
3.2.8 Gene co-expression network analysis	97
3.2.9 Statistical analysis and data visualization	98
3.3 RESULTS	99
3.3.1 Effect of drought stress on papaya plants	99
3.3.2 RNA-Sequencing and expression level quantification	101
3.3.3 Differential Expression Analysis.....	104
3.3.4 Core sets of DDEGs identified in the reference-based transcriptomic analysis	105
3.3.5 GO functional enrichment analysis of the sets of differentially expressed genes under control condition.....	108
3.3.6 GO functional enrichment analysis of the sets of DDEGs	111
3.3.7 Co-expression networks	113

3.3.8 TFs as regulatory hub.....	116
3.4 DISCUSSION	117
3.4.1 Biological processes differentially modulated under CN condition.....	117
3.4.2 Shared core sets of stress regulated genes, and genes related to water deprivation and ABA as targets for genetic improvement in papaya	117
3.4.3 Biological processes differentially up-regulated under moderate drought stress	118
3.4.4 Biological processes commonly modulated under severe drought stress	119
3.4.5 Long-distance movement of mRNA under drought conditions.....	120
3.4.6 Gene co-expression networks and transcriptional hubs	121
3.5 CONCLUSIONS	123
CAPÍTULO IV	125
SOBRE-EXPRESIÓN DE LOS GENES CPBI-1, CPLFG1 y CPLFG3 DE PAPAYA (<i>CARICA PAPAYA</i> L.) EN PLANTAS TRANSGÉNICAS DE TABACO (<i>NICOTIANA TABACUM</i> L.) Y SU FUNCIÓN EN LA TOLERANCIA AL ESTRÉS	125
4.1 INTRODUCCIÓN.....	125
4.2 MATERIALES Y MÉTODOS	129
4.2.1 Clonación y vectores	129
4.2.2 Transformación genética de tabaco mediante <i>Agrobacterium</i>	130
4.2.3 Confirmación de la transformación por PCR	131
4.2.4 Selección de las generaciones T1 y T2 de plantas transgénicas de tabaco...	133
4.2.5 Tratamientos de estrés.....	133
4.2.6 Medición de parámetros fisiológicos	134
4.2.7 Medición de contenido de prolina	135
4.2.8 Visualización de datos y análisis estadísticos	135
4.2.9 Expresión de los genes TMBIM de papaya en el transcriptoma de papaya....	135
4.3 RESULTADOS	137

4.3.1 Plantas transgénicas de tabaco transformadas con los genes CpBI-1, CpLFG1 y CpLFG3 de papaya (<i>Carica papaya</i> L.)	137
4.3.2 Germinación de semillas de tabaco transgénicas en medio con NaCl	139
4.3.3 Tolerancia de plantas transgénicas de tabaco en diferentes concentraciones de NaCl	139
4.3.4 Recuperación de tabacos transgénicos después de choque salino.....	143
4.3.5 Expresión de los genes TMBIM de papaya en sequía	147
4.4 DISCUSIÓN.....	149
4.4.1 Aclaración de las relaciones evolutivas de las proteínas LFG de Papaya	149
4.4.2 Los genes TMBIM de papaya regulan positivamente la tolerancia contra el estrés salino	150
4.4.3 Función dual del gen CpLFG1 en cuanto a la tolerancia al estrés salino	154
4.4.4 Los genes CpBI-1, CpLFG1 y CpLFG3 no están involucrados en la respuesta general de papaya al estrés por sequía	156
4.5 CONCLUSIONES	157
CAPÍTULO V	159
DISCUSIÓN GENERAL, CONCLUSIONES Y PERSPECTIVAS.....	159
5.1 DISCUSIÓN GENERAL.....	159
5.2 CONCLUSIONES GENERALES	167
5.3 PERSPECTIVAS	169
BIBLIOGRAFÍA.....	171

LISTADO DE FIGURAS

Figura 1.1. Principales canales y transportadores involucrados en el transporte de Na ⁺ en plantas.....	4
Figura 1.2. Vías extrínseca e intrínseca de la apoptosis. Tomado de Li & Sheng, 2012 (licencia: https://creativecommons.org/licenses/by/2.0/legalcode).....	8
Figura 1.3. La familia BCL-2. Representación de la estructura de los dominios BH1-4 de las proteínas de la familia BCL-2.....	9
Figura 1.4. Estructura predicha de la proteína HsBI-1.....	10
Figura 1.5. La superfamilia TMBIM del ser humano.	13
Figura 1.6. Filogenia de la superfamilia TMBIM representada por los genes TMBIM humanos.....	14
Figura 1.7. Estructura tridimensional de la proteína BsYetJ de <i>Bacillus subtilis</i>	16
Figura 1.8. Regulación de la proteína BsYetJ por pH.....	17
Figura 1.9. Interacciones proteína-proteína de los genes BI-1 en plantas.....	18
Figura 1.10. Posible mecanismo de acción de BI-1 en plantas.	20
Figura 1.11. Planta de papaya.	25
Figura 1.12. Fruto de papaya. El fruto del lado derecho proviene de una planta hermafrodita, y la del lado izquierdo de una planta femenina	26
Figura 1.13. Principales países productores de papaya en el mundo entre 2000 y 2017. 27	
Figura 1.14. Principales estados productores de papaya en México entre 2003 y 2017. . 28	
Figure 1.15. Análisis de la secuencia la proteína CpBI-1 de papaya.....	32
Figura 1.16. Expresión del en CpBI-1 de papaya en condiciones de estrés abiótico. a) Tratamiento de desecación de las raíces en papel filtro. b) Tratamiento de choque salino (300 mM). Datos tomados de Gamboa-Tuz, SD, 2015.	33

Figura. 1.17. Análisis de las secuencias de proteína de CpLFG1, CpLFG2 y CpLFG3 de papaya.....	34
Figura 1.18. Expresión de los genes CpLFG1 y CpLFG2 en condiciones de estrés abiótico.	35
Figura 1.19. Estrategia experimental.	45
Figure 2.1. Classification and conserved sequence regions of the TMBIM superfamily of proteins.....	64
Figure 2.2. Phylogenetic relationships of the 48 analyzed plant species.	66
Figure 2.3. Phylogenetic relationships of BI and LFG families in plants.	70
Figure 2.4. Synteny Network Analysis of the TMBIM superfamily in plants.	75
Figure 2.5. Proposed model of evolution and expansion of BI and LFG families in plants.	80
Figure 3.1. Effect of drought stress on papaya plants.	100
Figure 3.2. Reference-based and de novo transcriptomic analyses.	103
Figure 3.3. Shared core sets of up- and down- regulated DDEGs identified in the reference-based transcriptomic analysis.	108
Figure 3.4. Functional GO enrichment analysis of the sets of up-regulated genes in leaf, sap, and roots under control condition identified in the reference-based transcriptomic analysis.	111
Figure 3.5. Functional GO enrichment analysis of the sets of DDEGs identified in the reference-based transcriptomic analysis.	112
Figure 3.6. Gene co-expression networks (GCN) of leaf and root transcriptomes of papaya plant during drought treatments.	114
Figura 4.1. Mapa del vector Binario pK7FWG2,0	130
Figura 4.2 Líneas de plantas transgénicas de tabaco generación T1.....	138

Figura 4.3. Ensayo de germinación en 175 mM NaCl.	139
Figura 4.4. Plantas transgénicas de tabaco sometidas a diferentes concentraciones de NaCl.	141
Figura 4.5. Datos de fotosíntesis de las plantas transgénicas de tabaco sometidas a diferentes concentraciones de NaCl.	142
Figura 4.6. Datos de contenido de prolina de las plantas transgénicas de tabaco sometidas a diferentes concentraciones de NaCl.	143
Figura 4.7. Ensayo de recuperación de plantas transgénicas de tabaco sometidas a 175 mM NaCl.....	145
Figura 4.8. Datos de fotosíntesis de las plantas transgénicas de tabaco en ensayo de recuperación después de choque salino a 175 mM.	146
Figura 4.9. Expresión de los genes TMBIM de papaya en condiciones de sequía.	147
Figura 4.10. Filogenia de las proteínas LFG de Arabidopsis, papaya y cebada.	149
Figura 4.11. Función de los genes TMBIM de papaya en la tolerancia al estrés abiótico en plantas transgénicas de tabaco.	150
Figura 4.12. Posibles mecanismos controlados por los genes TMBIM de papaya en la tolerancia al estrés salino.....	153
Figura 4.13 Función dual del gen CpLFG1.	155

LISTADO DE CUADROS

Cuadro 1.1. Parálogos de HsBI-1 en el ser humano.	12
Cuadro 3.1. Samples and libraries used for de novo and reference-based transcriptomic analyses.	102
Cuadro 4.1 Cebadores utilizados para amplificar los genes CpBI-1, CpLFG1 y CpLFG3.	129
Cuadro 4.2. Cebadores usados para amplificar los genes CpBI-1, CpLFG1 y CpLFG2, y nptII.	132
Cuadro 4.3 Datos de expresión de los genes TMBIM de papaya en tratamiento de sequía.	148
Cuadro 4.4. Porcentaje de identidad y similitud entre las proteínas TMBIM de papaya y arabidopsis.	156

ABREVIATURAS

ADN	Ácido desoxirribonucleico
AIF	<i>Apoptosis inducing factor</i>
APAF-1	<i>Apoptosis protease activating factor-1</i>
BAK	<i>BCL-2 antagonist killer</i>
BAX	<i>BCL-2-associated x protein</i>
BCL-2	<i>B cell CLL/lymphoma-2</i>
Bcl-xL	<i>Bcl-2 long isoform</i>
BI	<i>Bax inhibitor</i>
BI-1	<i>Bax inhibitor-1</i>
BR	Brasinoesteroides
CCC	<i>Chloride cation co-transporter</i>
cDNA	<i>complementary DNA</i>
cDNA-AFLP	<i>cDNA Amplified Fragments Long Polymorphism</i>
DISC	<i>Death-inducing signaling complex</i>
FAS	<i>FAS receptor</i>
FASL	<i>FAS ligand</i>
GAAP	<i>Golgi anti-apoptotic protein</i>
GHITM	<i>Growth hormone-inducible transmembrane</i>
GRINA	<i>Glutamate receptor, ionotropic N-Methyl-d-Aspartate associated protein</i>
HKT	<i>High affinity potassium transporter</i>

LCT1	<i>Low affinity cation transporter 1</i>
LFG	<i>Lifeguard</i>
<i>nptII</i>	Neomicina fosfotransferasa
NSCC	<i>Non-selective cation channel</i>
PCD	<i>Programmed cell death</i>
PCR	<i>Polymerase chain reaction</i>
PRSV	<i>Papaya ring-spot virus</i>
RECS1	<i>Responsive to centrifugal force and shear 1</i>
RNA-seq	<i>Ribonucleic acid sequencing</i>
SOS	<i>Salt overly sensitive</i>
TaHsfA6f	<i>Heat shock activating factor A6f</i>
TMBIM	<i>Transmembrane Bax Inhibitor Motif containing</i>
TNF	<i>Tumor necrosis factor</i>
TNFR	<i>Tumor necrosis factor receptor</i>
UPR	<i>Unfolded protein response</i>

RESUMEN

Los cultivos vegetales están expuestos a diversos tipos de estrés abiótico, causando serias pérdidas económicas. Se necesitan estudios funcionales de genes vegetales específicos para poder desarrollar variedades más tolerantes contra el estrés abiótico. La superfamilia de genes/proteínas *Transmembrane Bax Inhibitor Motif containing* (TMBIM) se encuentra conservada en procariontes y eucariontes. Las proteínas TMBIM son canales iónicos con un dominio de siete regiones transmembranales que regulan las concentraciones de calcio celular e inhiben la muerte celular programada; tienen otras varias funciones, incluyendo tolerancia al estrés. En eucariontes, la superfamilia TMBIM se divide en las familias *Bax Inhibitor* (BI) y *Lifeguard* (LFG). En plantas, los genes BI pueden conferir tolerancia contra varios tipos de estrés abiótico, y posiblemente también los genes LFG. En el presente trabajo, se llevó a cabo un estudio filogenómico de las proteínas TMBIM en arqueas, bacterias y eucariontes, con énfasis en las proteínas TMBIM de plantas. También, se generaron datos transcriptómicos de plantas de papaya (*Carica papaya* L. 'Maradol') —un importante cultivo tropical— bajo condiciones de sequía mediante RNA-seq; estos datos fueron analizados para determinar si los genes TMBIM están involucrados en la respuesta de papaya a sequía. Además, se utilizaron plantas transgénicas de tabaco (*Nicotiana tabacum* L. 'Petite Havana SR1') expresando constitutivamente los genes CpBI-1, CpLFG1 y CpLFG3 de papaya para evaluar si dichos genes pueden conferir tolerancia al estrés, específicamente a salinidad. Se identificaron un total de 685 proteínas TMBIM en 171 organismos a través de los tres dominios de la vida; se reportan por primera vez las proteínas TMBIM en arqueas. En plantas, la familia BI permaneció como un solo grupo filogenético con pocos duplicados y dos transposiciones sinténicas en monocotiledóneas; la familia LFG se dividió en los clados LFG I y LFG II (A-B) con varios duplicados y grupos ortólogos. Los genes TMBIM de papaya no parecen estar involucrados en la respuesta transcripcional general de papaya al estrés por sequía. Los genes CpBI-1, CpLFG1 y CpLFG3 mejoraron la germinación de semillas transgénicas de tabaco en 175 mM NaCl. El gen LFG1 presentó una función dual: redujo la tolerancia de plantas transgénicas de tabaco crecidas en hidroponía a 150 mM NaCl, pero mejoró su supervivencia después de un choque salino a 175 mM NaCl. Las proteínas TMBIM de papaya tienen diferentes funciones en cuanto a la tolerancia al estrés abiótico dependiendo del estado del desarrollo y tipo de estrés, y esto se puede explicar en parte debido a las diferentes historias evolutivas de las familias BI y LFG en plantas.

ABSTRACT

Plant crops are exposed to diverse types of abiotic stress, which causes economic losses worldwide. Functional studies of plant genes under abiotic stress conditions are necessary if crop varieties with enhanced tolerance against abiotic stress are to be developed. The Transmembrane Bax Inhibitor Motif containing (TMBIM) superfamily of genes/proteins is widespread across prokaryotic and eukaryotic species. The TMBIM proteins are ion channels with a conserved domain of seven transmembrane regions that regulate cellular Ca^{2+} concentrations and inhibit programmed cell death; they have other several functions, including stress tolerance. In eukaryotes, the TMBIM superfamily is divided into the Bax Inhibitor (BI) and Lifeguard (LFG) families. Plant BI genes can confer tolerance against several types of abiotic stress. Plant LFG genes have been less studied, but they might confer abiotic stress tolerance as well. In the present work, a phylogenomic study of the TMBIM proteins across Archaea, Bacteria, and Eukarya was carried out with emphasis on plant TMBIM proteins. Also, transcriptomic data of papaya plant (*Carica papaya* L. 'Maradol')—an economically important tropical fruit crop—under drought stress was generated through RNA-seq; these data were analyzed to find out if the TMBIM genes were involved in the papaya drought response. Furthermore, transgenic tobacco plants (*Nicotiana tabacum* L. 'Petite Havana SR1') constitutively expressing the CpBI-1, CpLFG1, and CpLFG3 genes from papaya were used to evaluate if such genes can confer abiotic stress tolerance, specifically to NaCl stress. A total of 685 TMBIM proteins were identified in 171 organisms across the three domains of life; the presence of TMBIM proteins in Archaea is reported for the first time. In plants, the BI family remained as a single phylogenetic group with few duplicates and two major syntenic transpositions in monocots, while the LFG family was divided into the LFG I and LFG II (A-B) clades with several duplicates and orthogroups. The TMBIM genes of papaya were not found in the general transcriptional responses of papaya plant to drought stress. The CpBI-1, CpLFG1, and CpLFG3 genes improved germination rates of transgenic tobacco seeds under 175 mM NaCl treatment *in vitro*. The LFG1 gene had a dual role in salt tolerance of hydroponically-grown tobacco plants: it reduced the salt tolerance of transgenic tobacco plants at 150 mM NaCl, but improved their survival rate after a salt shock of 175 mM NaCl. The results of this work indicate that the TMBIM proteins of papaya have different functions on abiotic stress tolerance depending on plant developmental stage and type of abiotic stress, and this can be partially explained due to different evolutionary histories of the BI and LFG families in land plants.

INTRODUCCIÓN

Las plantas se ven expuestas constantemente a uno o más tipos de estrés abiótico, es decir condiciones climáticas no favorables que afectan su rendimiento, crecimiento y supervivencia. Esto es causa de serias pérdidas económicas (Mehrabi y Ramankutty, 2017; Qadir *et al.*, 2014). Además, el cambio climático amenaza con aumentar la distribución, frecuencia y severidad de diferentes tipos de estrés abiótico en el mundo. Para hacer frente a esta problemática, se requieren análisis funcionales de genes de diferentes especies vegetales que puedan conferir tolerancia a uno o más tipos de estrés abiótico. Esto con el fin de generar conocimiento básico acerca de las respuestas moleculares de las plantas, el cual puede contribuir a largo plazo a la generación de variedades modificadas genéticamente más tolerantes.

Los genes de la superfamilia TMBIM (del inglés *Transmembrane Bax Inhibitor motif containing*), codifican proteínas integrales de membrana con un único dominio compuesto por siete regiones transmembranales (hélices alfa). Evidencia creciente en bacterias, humanos y plantas ha demostrado que estas proteínas son canales iónicos que regulan las concentraciones de Ca^{2+} intracelular (Sierla *et al.*, 2019; Chang *et al.*, 2014; Bultynck *et al.*, 2012). Los genes TMBIM cumplen diferentes funciones biológicas en diferentes organismos y contextos. La función principal por la que se conoce a toda la superfamilia de genes TMBIM es la de inhibición de la muerte celular programada, anti-apoptosis o cito-protección. Sin embargo, algunos genes TMBIM también pueden inducir la muerte celular bajo determinadas condiciones.

Las proteínas de la superfamilia TMBIM se encuentran codificadas en el genoma de virus, bacterias y eucariontes. En eucariontes esta superfamilia se ha dividido en las familias BI (del inglés *Bax Inhibitor*) y LFG (del inglés *Lifeguard*), con varias duplicaciones dependiendo del linaje y la especie. A pesar de que varios estudios filogenéticos han ofrecido un panorama general sobre la evolución de las proteínas TMBIM (Henke *et al.*, 2011; Hu *et al.*, 2009), no se ha llevado a cabo un estudio extenso y sistemático de su evolución a través de los tres dominios de la vida. En especial, no se ha estudiado las duplicaciones y relaciones de ortología y sintenia de las familias BI y LFG en plantas, información necesaria para comprender la función de estos genes.

Los genes de la familia BI en plantas regulan la muerte celular (Kawai *et al.*, 1999), el metabolismo de lípidos (Nagano *et al.*, 2009), la autofagia (Xu *et al.*, 2017), la senescencia (Yue *et al.*, 2012), el desarrollo del embrión (Reza *et al.*, 2018), las interacciones planta-microbio (Hernández-López *et al.*, 2018; Hückelhoven *et al.*, 2003) y la tolerancia al estrés abiótico (Watanabe y Lam, 2009). Por otra parte, algunos miembros de la familia LFG en plantas regulan las interacciones planta-patógeno (Weis *et al.*, 2013), la vía de los brasinoesteroides (YAMAGAMI *et al.*, 2009) y la tolerancia al estrés del retículo endoplásmico (Guo *et al.*, 2018). Debido a sus múltiples funciones en plantas, en particular la tolerancia a diversos tipos de estrés, los genes de la superfamilia TMBIM son atractivos desde un punto de vista biotecnológico.

La planta de papaya (*Carica papaya* L.) es cultivada en varios países en zonas tropicales y subtropicales. México ocupa el primer lugar en exportación y tercero en producción a nivel mundial (FAOSTAT, 2016). Se cuenta con el genoma completo de la variedad ‘SunUp’ (Ming *et al.*, 2008) y varios estudios transcriptómicos, pero no bajo condiciones de estrés abiótico. Este recurso sería útil para explorar la participación de genes específicos en procesos generales de la respuesta al estrés abiótico en dicho cultivo.

Previamente, se había aislado un gen de la familia BI, CpBI-1, y dos genes de la familia LFG, CpLFG1 y CpLFG3, de papaya cuya expresión a nivel transcripcional es inducida bajo condiciones de desecación y salinidad (Gamboa-Tuz, 2015). Sin embargo, es necesario comprobar su función biológica mediante genética reversa (plantas transgénicas) para explorar su potencial biotecnológico.

En el presente trabajo se realizó un análisis filogenómico de la superfamilia de proteínas TMBIM en los tres dominios de la vida, con énfasis en la evolución de las familias BI y LFG en plantas. Además, se estudió el transcriptoma de papaya bajo condiciones de sequía y se analizó si los genes TMBIM de papaya se encuentran en la respuesta general de esta planta ante dicho estrés. Finalmente, se analizó la función de los genes CpB-1, CpLFG1 y CpLFG3 de papaya en cuanto a la tolerancia al estrés abiótico, particularmente salinidad, mediante su expresión constitutiva en plantas transgénicas de tabaco.

CAPÍTULO I

ANTECEDENTES

1.1 Estrés abiótico

El estrés abiótico se refiere a cualquier condición ambiental que reduce el crecimiento, rendimiento o supervivencia de una planta por debajo de sus niveles óptimos (Cramer *et al.*, 2011). Los cultivos vegetales se encuentran expuestos a diversos tipos de estrés abiótico, lo cual tiene un impacto negativo en su supervivencia y productividad, y ocasiona severas pérdidas económicas.

1.1.1 Estrés salino

El estrés salino es uno de los tipos de estrés abiótico más limitante de la productividad agrícola en el mundo, pues causa severos efectos adversos sobre la germinación, el vigor de las plantas y el rendimiento de los cultivos (Munns y Tester, 2008). La salinidad es la condición del suelo caracterizada por altas concentraciones de sales solubles, siendo el NaCl la más soluble y abundante. Se considera que un suelo es salino cuando presenta una conductividad eléctrica (EC) de 4 dS/m o superior, lo cual es equivalente a aproximadamente 40 mM de NaCl y genera una presión osmótica de aproximadamente 0.2 MPa (Munns y Tester, 2008). A nivel global, se ha estimado que más de 800 millones de hectáreas se ven afectadas por la salinidad. Alrededor de 32 millones (2%) de hectáreas destinadas a la agricultura de secano y 45 millones (20%) de hectáreas de tierra para agricultura de regadío se ven afectadas por la salinidad en todo el mundo (Munns y Tester, 2008). La salinidad afecta a las plantas de dos maneras principales: 1) las altas concentraciones de sales en el suelo reducen el potencial osmótico, lo que conlleva a la reducción de la expansión de las hojas, cierre de estomas, reducción en la fotosíntesis y pérdida de biomasa; 2) las altas concentraciones Na^+ dentro de la planta pueden llegar a niveles tóxicos resultando en la inhibición de procesos importantes (como la fotosíntesis) y desordenes nutrimentales debido a una deficiencia de K^+ (Munns y Tester, 2008; Bartels y Sunkar, 2005).

El Na^+ entra a las células epidérmicas de las raíces de manera pasiva dependiendo del gradiente de concentración y la diferencia de potencial a través de la membrana plasmática.

A la fecha, no se han identificado canales iónicos selectivos para Na^+ , pero se ha propuesto que la principal vía de entrada de Na^+ a las raíces de las plantas es mediante canales iónicos no selectivos (NSCC). Sin embargo, también podrían existir otras vías responsables de la entrada de Na^+ como transportadores catiónicos de baja afinidad (LCT1), transportadores de K^+ de alta afinidad (HKT), transportadores de alta afinidad por el K^+ (KUP/HAK/KT), canales de K^+ tipo Shaker (AKT1) y co-transportadores cloro-cación (CCC; Figura 1.1) (Adams y Shin, 2014; Deinlein *et al.*, 2014).

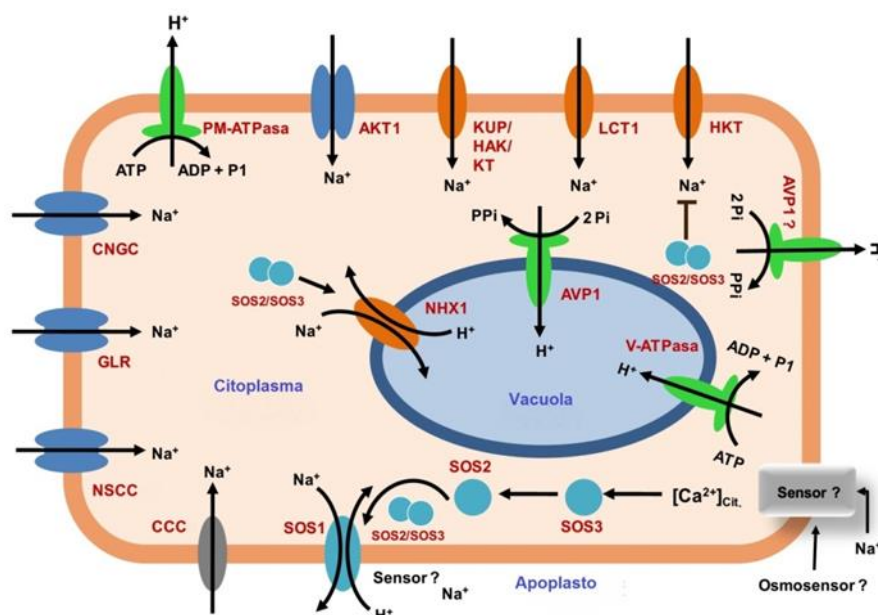


Figura 1.1. Principales canales y transportadores involucrados en el transporte de Na^+ en plantas. NSCC: canales iónicos no selectivos; GLR: canales similares a receptores de glutamato; CNGC: canal iónico nucleotídico cerrado; LCT1: transportadores catiónicos de baja afinidad; HKT: transportadores de K^+ de alta afinidad; KUP/HAK/KT: transportadores de alta afinidad por el K^+ ; AKT1: canales de K^+ tipo Shaker; CCC: co-transportadores cloro-cación; NHX: antiportador Na^+/H^+ ; SOS: vía altamente sensible a la sal; PM-ATPasa: H^+ ATPasa de la membrana plasmática; AVP1: H^+ Pirofosfatasa de vacuola; V-ATPasa: H^+ ATPasa de vacuola. Esta figura fue modificada de Quintal *et al.* (2012). Licencia: <https://creativecommons.org/licenses/by/4.0/>.

El exceso de la concentración de Na^+ en las células de la raíz posiblemente es percibido por sensores (osmóticos y/o iónicos), aún no identificados, los cuales inducen un cambio en segundos mensajeros como las especies reactivas de oxígeno (ROS) y las concentraciones de Ca^{2+} citosólico. Estos cambios en segundos mensajeros dan lugar a la activación de una compleja red de señalización —la cual involucra sensores y cinasas

dependientes de calcio (CBL, CIPK y CDPK), factores de transcripción (CAMTA, NAC, WRKY, bHLH, MYB, AP2/ERF y bZIP) y hormonas (ABA)— que culmina en la síntesis de osmolitos compatibles (prolina, glicina betaína), de proteínas detoxificadoras de ROS (catalasas y peroxidasas) y la activación de vías de exclusión de Na^+ a nivel celular y de planta completa (Deinlein *et al.*, 2014). Las principales vías de exclusión de Na^+ del citosol identificadas son la vía HKT, SOS y NHX, las cuales se explican brevemente en los siguientes párrafos (Zhang y Shi, 2013).

Los transportadores de membrana HKT. Desde la identificación del gen TaHKT2;1 de trigo como mediador del transporte Na^+/K^+ , se han identificado y caracterizado muchos genes HKT de diferentes especies de plantas. Existen al menos dos grupos de transportadores HKT: la clase I, que transporta de manera más selectiva el Na^+ ; y la clase II, que actúa como un co-transportador Na^+/K^+ (Hamamoto *et al.*, 2015). AtHKT1;1 es un transportador clase I que provee un mecanismo clave de protección contra la sobre-acumulación de Na^+ en las hojas: mutaciones en el gen AtHKT1;1 de arabidopsis, causan la sobreacumulación de Na^+ en las hojas; en contraparte una sobre-expresión de AtHKT1;1 aumenta la tolerancia al estrés salino (Deinlein *et al.*, 2014). El modelo actual de la función de AtHKT1 sugiere que este transportador descarga el Na^+ del xylema hacia las células del parénquima, reduciendo así el transporte de Na^+ de las raíces hacia la parte aérea y hojas (Zhang y Shi, 2013).

El antiporter vacuolar NHX1. AtNHX1 es un transportador Na^+/H^+ que se encarga de la compartimentación del Na^+ en la vacuola. Este gen puede complementar funcionalmente a ScNHX1, un antiporter Na^+/H^+ de levadura. Además, confiere tolerancia al estrés salino en varias especies de plantas como arabidopsis, tomate y canola. Mediante la compartimentación del Na^+ en la vacuola se reducen los efectos tóxicos del Na^+ en el citosol al mismo tiempo que se mantiene la presión de turgencia de la célula, proveyendo tolerancia contra los efectos osmóticos y iónicos de la salinidad a la vez (Zhang y Shi, 2013). Por mencionar un ejemplo, la concentración de Na^+ en la vacuola de plantas de tabaco adaptadas a salinidad es de hasta 780 mM, mientras que en el citoplasma es de 96 mM (Binzel *et al.*, 1988). Sin embargo, las vacuolas tienen un límite para acumular el Na^+ por lo que es necesaria la coordinación de esta vía con otras (Zhang y Shi, 2013).

La vía de señalización SOS. La vía de señalización SOS se compone de tres genes claves involucrados en la expulsión del Na^+ : SOS1, SOS2 y SOS3. SOS3 codifica una proteína

que percibe el incremento del Ca^{2+} citosólico activado por el exceso de Na^+ en el citoplasma. Después de unirse al Ca^{2+} , SOS3 es capaz de activar a SOS2, una proteína Serina/Treonina cinasa perteneciente a la familia SnRK3. La interacción SOS3-SOS2 recluta a SOS2 hacia la membrana plasmática, lo cual lleva a la activación de SOS1, un antiporter Na^+/H^+ que se encarga de la expulsión del exceso de Na^+ hacia el medio externo. La pérdida de función de cualquiera de estos genes resulta en diferentes niveles de hipersensibilidad hacia el estrés salino, siendo las plantas mutantes *sos1* las más sensibles identificadas hasta ahora. Otras funciones atribuidas a la vía SOS son los cambios en la arquitectura de las raíces, transporte de auxinas y remodelación del cito-esqueleto, entre otras (Ji *et al.*, 2013; Zhang y Shi, 2013). El complejo SOS3-SOS2 también regula a NHX y HKT para reducir la concentración de Na^+ en el citoplasma (Quintal *et al.*, 2012).

1.2 La superfamilia TMBIM

Los genes de la superfamilia TMBIM (del inglés *Transmembrane Bax Inhibitor Motif containing*) codifican proteínas con un único dominio conservado de siete regiones transmembranales (hélices alfa) que actúan como canales iónicos que regulan las concentraciones de Ca^{2+} intracelular. La superfamilia TMBIM tomó su nombre del primero de sus miembros que fue identificado y caracterizado: el gen Bax Inhibitor-1 del ser humano, abreviado como hBI-1 o HsBI-1, o simplemente como BI-1 (Xu y Reed, 1998). Este gen fue identificado mediante el cribado funcional de una biblioteca de cDNA en levadura debido a su capacidad de inhibir la apoptosis inducida por la expresión ectópica del gen Bax de mamífero, de ahí su nombre (Xu y Reed, 1998).

Los genes de la superfamilia TMBIM son mayormente conocidos por su función de inhibidores de la PCD debido al contexto en el que fueron descubiertos (Xu y Reed, 1998). Sin embargo, algunos genes TMBIM pueden inducir la muerte celular bajo determinadas condiciones. Además, las proteínas codificadas por estos genes forman complejos con otras proteínas en animales y plantas, poseen múltiples funciones y participan en la regulación de varios procesos celulares.

1.2.1 Apoptosis, la familia BCL-2 y el descubrimiento del gen BI-1 humano

Apoptosis y la familia BCL-2

La muerte celular programada (PCD, por sus siglas en inglés) es un proceso de suicidio celular regulado genéticamente. Sucede en animales y plantas en respuesta a señales internas o externas como parte del desarrollo o en respuesta a estímulos ambientales. Se han descrito varios tipos de PCD en animales y plantas.

En animales, particularmente en mamíferos, el tipo de PCD más estudiado es el de la apoptosis (PCD tipo I), la cual se distingue por un conjunto de características morfológicas y bioquímicas particulares: condensación de la cromatina, aparición de invaginaciones en la membrana plasmática (*belb* o *belbbing*, en inglés), ‘escalaramiento’ del ADN genómico, y formación y lisis de cuerpos apoptóticos. Un tipo de muerte celular parecido a la apoptosis ha sido descrito también en plantas (Lord y Gunawardena, 2012; Reape y McCabe, 2010).

La apoptosis puede ocurrir por dos vías, la extrínseca y la intrínseca (Li y Sheng, 2012) (Figura 1.2). La vía extrínseca de la apoptosis está regulada por los receptores de la membrana plasmática FAS y TNFR1 que reconocen a los ligandos FASL y TNFa, respectivamente. Al unirse ligando y receptor se forma el complejo de activación de muerte celular, DISC. Este complejo activa a la caspasa 8 que a su vez activa a la caspasa 3, la cual se encarga de degradar y dismantelar los componentes de la célula. Por otra parte, la vía intrínseca se encuentra regulada por los miembros de la familia BCL-2, la cual contiene miembros pro- y anti-apoptóticos que interaccionan físicamente entre sí (a nivel proteína-proteína). Dependiendo del resultado de esta interacción la célula ejecuta o no la apoptosis. De ejecutarse, la vía intrínseca de la apoptosis inicia con la despolarización de la membrana mitocondrial, debido a la formación de un poro formado por la proteína codificada por el gen Bax y la liberación de los factores pro-apoptóticos que residen en la matriz mitocondrial, citocromo c, AIF y Smac/Diablo. Estos factores pro-apoptóticos se ensamblan junto con las proteínas caspasa 9 y APAF-1 para dar lugar al apoptosoma, el cual se encarga de activar a la caspasa 3 para degradar todos los componentes celulares (Figura 1.2).

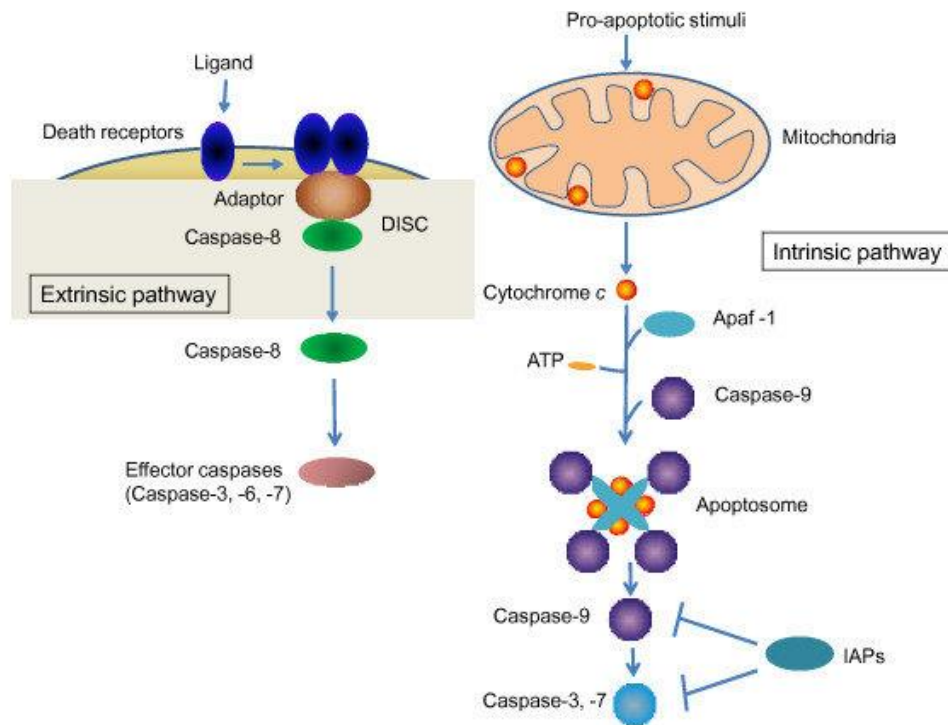


Figura 1.2. Vías extrínseca e intrínseca de la apoptosis. Tomado de Li & Sheng, 2012 (licencia: <https://creativecommons.org/licenses/by/2.0/legalcode>).

La familia de genes/proteínas BCL-2 son los principales reguladores de la vía intrínseca de la apoptosis en mamíferos y se dividen en tres subtipos o subfamilias: BCL-2, Bax, y BH3-only (Chipuk *et al.*, 2010). Estas subfamilias se caracterizan por la arquitectura de sus dominios BH1, BH2, BH3 y BH4 y sus funciones pro- o anti-apoptóticas (Figura 1.3). Los genes de la subfamilia BCL-2 codifican proteínas anti-apoptóticas con los cuatro dominios BH (1-4) y una región transmembranal (TM). Los genes de la subfamilia BAX codifican proteínas pro-apoptóticas con únicamente tres de los dominios BH (1-3) y una región transmembranal. Los genes de la subfamilia BH3-only codifican proteínas pro-apoptóticas únicamente con el dominio BH3, algunas contienen un dominio transmembranal (Figura 1.3).

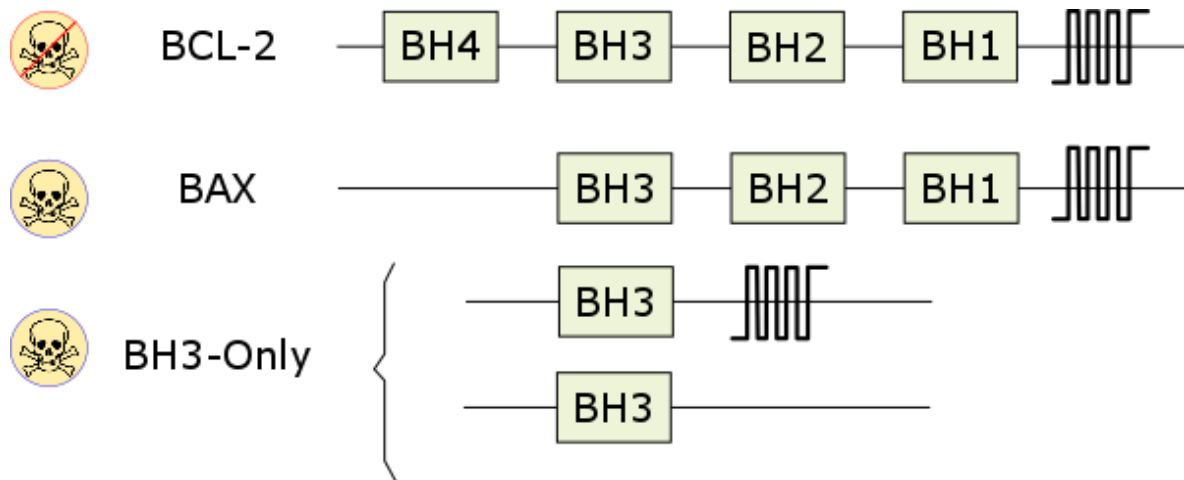


Figura 1.3. La familia BCL-2. Representación de la estructura de los dominios BH1-4 de las proteínas de la familia BCL-2.

La familia BCL-2 está conservada en mamíferos y otros vertebrados, pero no se encuentra en otros reinos, incluyendo levaduras y plantas. Sin embargo, los genes BCL-2, tanto anti-apoptóticos como pro-apoptóticos, siguen cumpliendo sus funciones celulares cuando son expresados ectópicamente en levaduras y plantas. Por ejemplo, el gen BCL-2 puede inhibir la muerte celular cuando es expresado ectópicamente en plantas e incluso conferir tolerancia al estrés abiótico (Xu *et al.*, 2004), y el gen Bax puede inducir la muerte celular en levaduras y plantas (Kawai-Yamada *et al.*, 2001; Xu y Reed, 1998).

El gen Bax Inhibitor-1 humano, HsBI-1

El gen Bax Inhibitor-1 (BI-1) humano (HsBI-1) fue identificado mediante un cribado funcional de una biblioteca de cDNA en levadura (Xu y Reed, 1998). La expresión ectópica del gen Bax de mamífero en levaduras y plantas produce un fenotipo letal, es decir induce la muerte celular en estos organismos (como se mencionó en la sección anterior). Se han generado líneas celulares de levaduras que expresen el gen Bax de manera inducible solo cuando son crecidas en medio con galactosa (bajo el control del promotor GAL10). Así, estas líneas crecen sin problema en medio sin galactosa, pero mueren cuando son transferidas a un medio con galactosa debido a la expresión y acumulación de la proteína Bax. Entonces, cuando una biblioteca de cDNA es expresada en estas líneas y se les crece en medio con galactosa, es posible identificar genes “inhibidores de Bax” funcionales mediante el cribado

de células de levadura capaces de crecer en dicho medio. Usando esta herramienta, Xu and Reed (1998) identificaron al gen BI-1 humano, abreviado como hBI-1 o HsBI-1.

Además de inhibir la PCD inducida por Bax, el gen BI-1 puede inhibir la muerte celular causada por inductores de la vía intrínseca de la apoptosis, como estaurosporina y etopósido, pero no inhibe la muerte celular inducida por factores de la vía extrínseca, como el ligando FasL (Xu y Reed, 1998).

El ORF del gen BI-1 humano aislado por Xu and Reed (1998) codifica una proteína de 237 aminoácidos con un dominio transmembrana conformado por 6-7 regiones altamente hidrofóbicas—hélices alfa (Figura 1.4). Esta proteína se localiza principalmente en la membrana del retículo endoplásmico, e interactúa con las proteínas codificadas por los genes anti-apoptóticos Bcl-2 y Bcl-xl, pero no con la proteína codificada por el gen anti-apoptótico Bax (Xu y Reed, 1998).

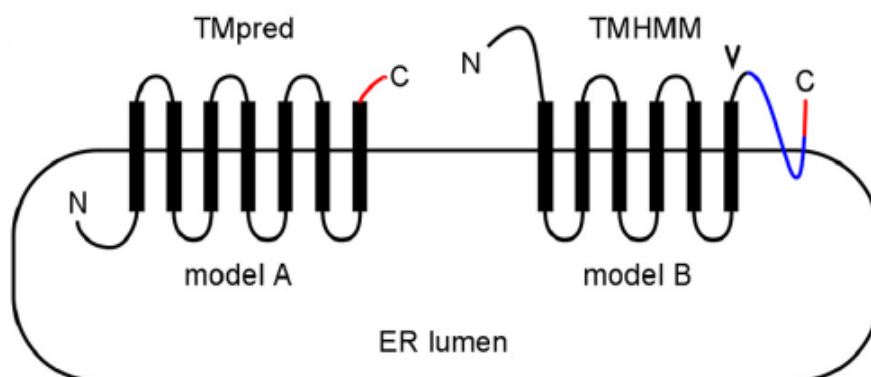


Figura 1.4. Estructura predicha de la proteína HsBI-1. Modelo A: siete regiones transmembranales con el amino terminal en el lumen del citosol y el carboxilo terminal en el citoplasma; Modelo B: seis regiones transmembranales con el amino terminal y el carboxilo terminal en el citoplasma. La mayoría de los datos estructurales favorecen el modelo A. Tomado de Bulynk *et al.* (2012). Permiso para publicar en tesis de *American Society for Biochemistry and Molecular Biology*: (<https://www.copyright.com/openurl.action?rwr=26&issn=1083-351&WT.mc.id=American%20Society%20for%20Biochemistry%20and%20Molecular%20Biology>)

Además del dominio de siete regiones transmembranales, la proteína BI-1 del ser humano posee en el extremo del carboxilo terminal un motivo corto rico en residuos de lisina (EKDKKKEKK), el cual forma una hélice super-enrollada y ha sido sugerido como sensor

de pH (Kim *et al.*, 2008), aunque otros estudios sugieren que no es así (Bultynck *et al.*, 2014). Lo que sí está claro, es que este motivo rico en lisinas está conservado en plantas y es importante para algunas funciones de BI-1 en plantas como la regulación del calcio citosólico y la interacción con otras proteínas (Ihara-Ohori *et al.*, 2007; Kawai-Yamada *et al.*, 2004).

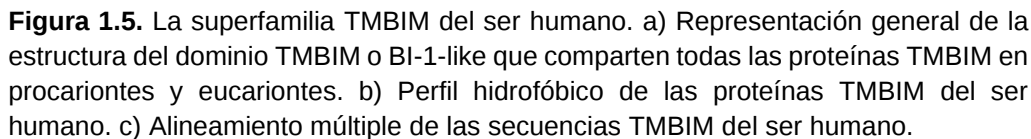
En el ser humano, BI-1 actúa como un canal de fuga pasiva de calcio que reduce las concentraciones basales de Ca^{2+} en el retículo endoplásmico, de esta manera reduce la carga de calcio que es liberada al citosol y a la mitocondria durante un estímulo de muerte celular o estresante, evitando así la muerte celular (Bultynck *et al.*, 2014). Además de la regulación del calcio celular mediante la fuga pasiva de calcio, la proteína BI-1 también podría regular las concentraciones de calcio mediante su interacción con IP3 (Kiviluoto *et al.*, 2012). La proteína BI-1 también interactúa con IRE1a para regular la repuesta a proteínas no plegadas y la autofagia (Castillo *et al.*, 2011; Lisbona *et al.*, 2009). También, BI-1 interactúa con G-actina y regula la polimerización de actina y adhesión celular (Lee *et al.*, 2010). El gen BI-1 del ser humano tiene muchas otras funciones en cáncer y enfermedades degenerativas o autoinmunes (Li *et al.*, 2014; Robinson *et al.*, 2011; Reimers *et al.*, 2008)

1.2.2 Homólogos de BI-1: clasificación y evolución de la superfamilia TMBIM

Además del gen HsBI-1 (también denominado TMBIM6), en el genoma humano se encuentran otras cinco proteínas parálogas que contienen un dominio similar al BI-1: GHITM (TMBIM5), RECS1 (TMBIM1), LFG (TMBIM2), GRINA (TMBIM3) y GAAP (TMBIM4) (Cuadro 1.1). Se han identificado homólogos de estas proteínas en varias especies de bacterias, protozoarios, hongos, invertebrados, vertebrados y plantas (Henke *et al.*, 2011). En conjunto, todas estas proteínas son consideradas como una sola superfamilia denominada la superfamilia TMBIM, y su característica en común es el dominio TMBIM o *BI-1-like*, el cual está compuesto de 7 regiones transmembranales (Figura 1.5).

Cuadro 1.1. Parálogos de HsBI-1 en el ser humano.

Nombres	Descripción
RECS1	- Se localiza principalmente en el aparato de Golgi, endosoma y lisosoma. - Inhibe la vía extrínseca de la apoptosis inducida por el ligando Fas.
TMBIM1	Se expresa en todos los tejidos, particularmente en el tejido muscular, con excepción del timo, testículos y bazo. Su expresión aumenta en respuesta a estrés mecánico.
LFG3	En mamíferos euterios este gen presenta un homólogo denominado Tmbim1b/LFG5, cuya función aún no ha sido caracterizada.
LFG	Se localiza principalmente en el aparato de Golgi y las balsas lipídicas de la membrana plasmática.
TMBIM2	Atenúa la vía extrínseca de la apoptosis inducida por el ligando Fas, pero no TNFa; se une al receptor Fas e interfiere con la activación de la caspasa 8.
LFG2	- Inhibe la liberación del Ca^{2+} del retículo endoplásmico en respuesta a FasL. - Reducción de su expresión reduce la adhesión de la célula incrementando la migración celular y metástasis. - Se expresa principalmente en el sistema nervioso central.
GRINA	Se localiza principalmente en el aparato de Golgi y la membrana plasmática.
TMBIM3	Presenta una función anti-apoptótica bajo condiciones de estrés del retículo endoplásmico, y la respuesta de proteínas desplegadas (UPR). Aumenta sus niveles de expresión mediante los factores de transcripción PERK y ATF4.
LFG1	- Reduce la liberación del Ca^{2+} del Retículo endoplásmico mediada por IP3R. - Presenta actividad anti-apoptótica sinérgica con BI-1. - Se expresa en todos los tejidos del ser humano.
GAAP	Se localiza principalmente en el aparato de Golgi, aunque también se le puede encontrar en el retículo endoplásmico.
TMBIM4	- Es requerido para que la célula sea viable (en humanos). - Inhibe las vías intrínseca y extrínseca de la apoptosis
LFG4	- Reduce el contenido basal de Ca^{2+} en el retículo endoplásmico y aparato de Golgi; y también reduce la liberación de calcio mediada por IP3R en respuesta a un estímulo apoptótico. - Forma un canal iónico y se oligomeriza dependiendo del pH, pero sus monómeros aún mantienen su función de canal. - Se expresa en todos los tejidos (expresión constitutiva).
GHITM	Se localiza en la membrana interna de la mitocondria.
TMBIM5	- Inhibe la fragmentación mitocondrial y la liberación del citocromo C y Smac/Diablo durante la ejecución de la vía intrínseca de la apoptosis. - No se encuentra en plantas.



La superfamilia TMBIM en procariontes

El número de homólogos por genoma y las relaciones filogenéticas de la superfamilia TMBIM en bacterias no están claras (si pertenecen a la familia BI o LFG, o tiene sus propias agrupaciones) y nunca se ha explorado el genoma de las arqueas.

La superfamilia TMBIM se divide en BI y LFG en eucariontes

Aunque el dominio de las proteínas TMBIM se encuentra conservado a nivel de estructura (Siete regiones transmembranas), tienen poca conservación a nivel de secuencia de aminoácidos. Debido a esto, la superfamilia de proteínas TMBIM ha sido clasificada en dos familias en eucariontes: *Bax Inhibitor* (BI), conteniendo a las proteínas HsBI-1/TMBIM6 y GHITM/TMBIM5, y *Lifeguard* (LFG), conteniendo a RECS1/TMBIM1, LFG/TMBIM2, GRINA/TMBIM3 y GAAP/TMBIM4 (Figura 1.6). Éstas últimas han sido también denominadas como LFG3, LFG2, LFG1 y LFG4, respectivamente (Cuadro 1.1 y Figura 1.6).

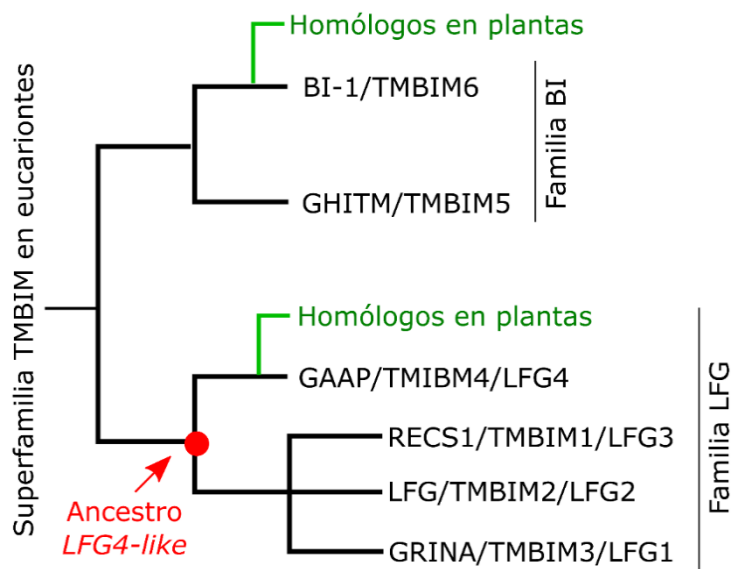


Figura 1.6. Filogenia de la superfamilia TMBIM representada por los genes TMBIM humanos.

En cuanto a la familia BI en plantas, únicamente se encuentran homólogos de HsBI-1, pero no de GHITM/TMBIM5 (Figura 1.6). La evolución de la familia BI en otros organismos, como los hongos, no ha sido estudiada.

De acuerdo con Hu y colaboradores (2009), las proteínas de la familia LFG tuvieron un origen común (y diferente al HsBI-1 del ser humano) a partir de una proteína similar a LFG4 (GAAP) hace aproximadamente 2,000 millones de años (Figura 1.6). Sin embargo, no se descarta la posibilidad de que las familias BI y LFG en eucariontes hayan tenido un origen en común remoto en procariontes.

En vertebrados, las proteínas LFG tuvieron dos duplicaciones principales: LFG4 y LFG1. Las proteínas LFG4 permanecieron como un solo grupo, mientras que las proteínas LFG1 tuvieron más duplicaciones (LFG2 y LFG3, LFG5 en mamíferos euterios), dando lugar a varios grupos específicos en vertebrados (Hu *et al.*, 2009).

En plantas, las proteínas LFG también se derivaron de un antepasado de tipo LFG4 y se han identificado en briofitas, gimnospermas, *Arabidopsis thaliana* y *Hordeum vulgare* (Weis *et al.*, 2013). Al parecer, las proteínas LFG en plantas han sufrido varios eventos de duplicación; por ejemplo, cebada y *Arabidopsis* poseen cinco proteínas LFG denominadas cada una como HvLFGa-e y AtLFG1-5, respectivamente (Weis *et al.*, 2013). Según análisis filogenéticos previos, las proteínas HvLFGe y AtLFG5 se agrupan juntas mientras que las proteínas HvLFGa-d y AtLFG1-4 están más estrechamente relacionadas entre sí (Weis *et al.*, 2013). Además, en un análisis preliminar se han identificado numerosas proteínas LFG en 23 especies de plantas, las cuales se agruparon en tres clados (Gamboa-Tuz, 2015). Sin embargo, no se ha profundizado más en la presencia y aparición de clados específicos de estas proteínas a través de la evolución de las plantas, ni se han estudiado las relaciones de ortología y sintenia, por lo que su historia evolutiva aún no está clara en plantas.

1.2.3 Las proteínas TMBIM son canales de calcio dependientes de pH

Varios estudios bioquímicos de las proteínas TMBIM en modelos animales y estudios de cristalografía de rayos X de la proteína BsYetJ de la bacteria *Bacillus subtilis* (Chang *et al.*, 2014) ha provisto un modelo sobre la estructura tridimensional de las proteínas TMBIM y su función como canales iónicos de calcio (Figuras 1.7 y 1.8).

La estructura tridimensional de la proteína BsYetJ de *Bacillus subtilis* corresponde a la de un canal formado por siete regiones transmembrana (TM1-7) con el extremo N-terminal

hacia el citoplasma y el extremo C-terminal hacia la región periplásmica (Chang *et al.*, 2014). A pH 8, BsYetJ presenta una conformación cerrada con 56 Å de longitud (de los cuales 31 Å se encuentran incrustados en la membrana) y 34 Å de diámetro (Figura 1.7A). A pH 6, BsYetJ adopta una conformación abierta donde la TM2 se desplaza hasta 13.5 Å formando un poro (junto con TM5-7 y la bicapa lipídica), el cual tiene un ancho de hasta 11 Å en la región periplásmica y se estrecha hasta 5 Å en la región del citoplasma (Figura 1.7B). La estructura general de la proteína consta de tres componentes: TM1-3, TM4-6 y TM7 (Figura 1.7C). TM1-3 y TM4-6 forman una estructura en sándwich de triple hélice cada uno; TM7 se encuentra en el centro del canal, paralelo a TM3 y anti-paralelo a TM6 (Figura 1.7C).

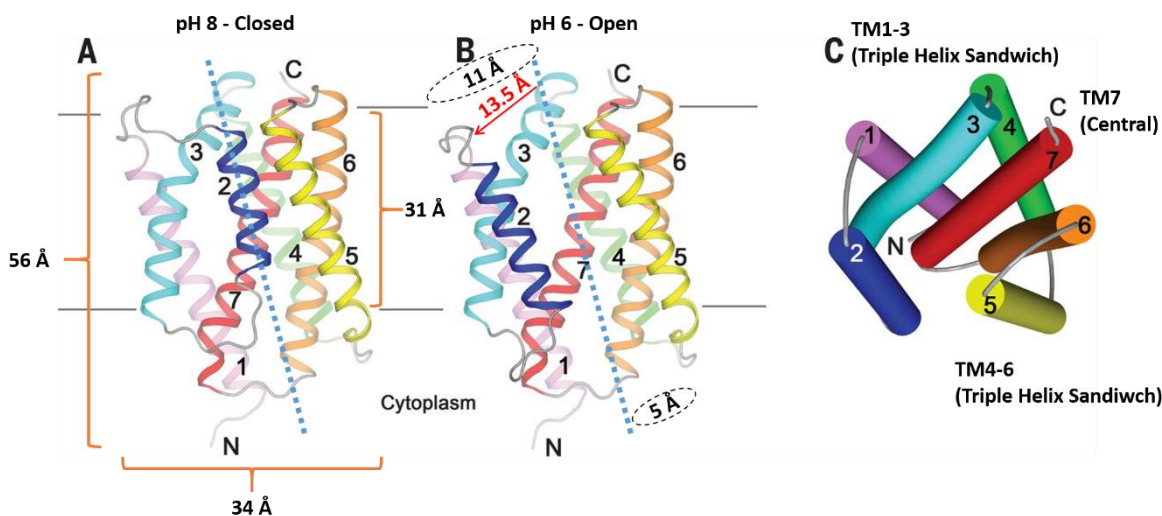


Figura 1.7. Estructura tridimensional de la proteína BsYetJ de *Bacillus subtilis*. A) Conformación cerrada de BsYetJ a pH 8. B) Conformación abierta de BsYetJ a pH 6. C) Estructura general representada con cilindros de la conformación abierta de BsYetJ. En A-C, las hélices se indican con colores, y los 'loops' que las conectan y los extremos N- y C-terminal se indican en gris. Esta figura fue tomada y modificada de (Chang *et al.*, 2014), permiso para uso en tesis (<https://s100.copyright.com/clients/aaas/tc/aaasgeneraltc.pdf>).

BsYetJ regula el paso del Ca^{2+} a través de la membrana dependiendo del pH (Chang *et al.*, 2014). La afluencia de calcio a través de proteo-liposomas formados a partir de la proteína

purificada de BsYetJ es mayor a un pH neutro o fisiológico (7.0-7.4) que a un pH más bajo (6.5) o más elevado (7.9) (Chang et al., 2014). Es decir, únicamente transporta el calcio a un pH neutro (~7.4; Figura 1.8). Esta actividad de pH está regulada por un sensor de pH conformado por dos residuos de aspartato (D171 y D195) conservados en todas las proteínas TMBIM en bacterias y eucariontes (Chang et al., 2014).

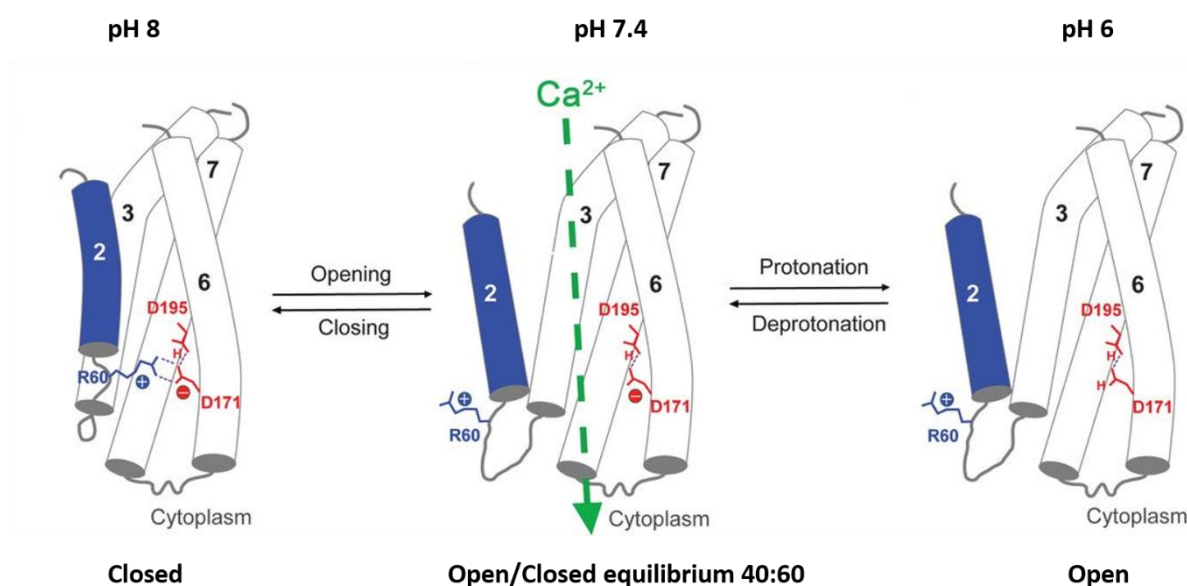


Figura 1.8. Regulación de la proteína BsYetJ por pH. Esta figura fue tomada y modificada de (Chang et al., 2014) (<https://s100.copyright.com/clients/aaas/tc/aaasgeneraltc.pdf>).

1.2.4 Funciones de los genes TMBIM en hongos

Se han caracterizado funcionalmente los genes TMBIM de tres especies dentro de Ascomycota. El homólogo de BI-1 en la levadura *S. cerevisiae*, conocido como YNL305Cp, Ybh3p o BXI1p, inhibe la PCD inducida por la proteína BAX y protege a la levadura contra choque térmico y estrés del ER (Cebulski et al., 2011). De manera similar, el gen MrBI-1 de *Metarhizium robertsii* brinda protección contra la PCD inducida por BAX y choque térmico (Chen et al., 2015). Por último, el gen SsBI-1 de *Sclerotinia sclerotiorum* también brinda

protección contra BAX, choque térmico y H_2O_2 , y es necesario para la virulencia de este patógeno (Yu *et al.*, 2015).

1.2.5 Funciones de los genes BI-1 en plantas

Los genes BI-1 han sido estudiados extensivamente en plantas, lo que ha revelado su participación en numerosos procesos celulares mediante su interacción con diversas proteínas (Figura 1.9).

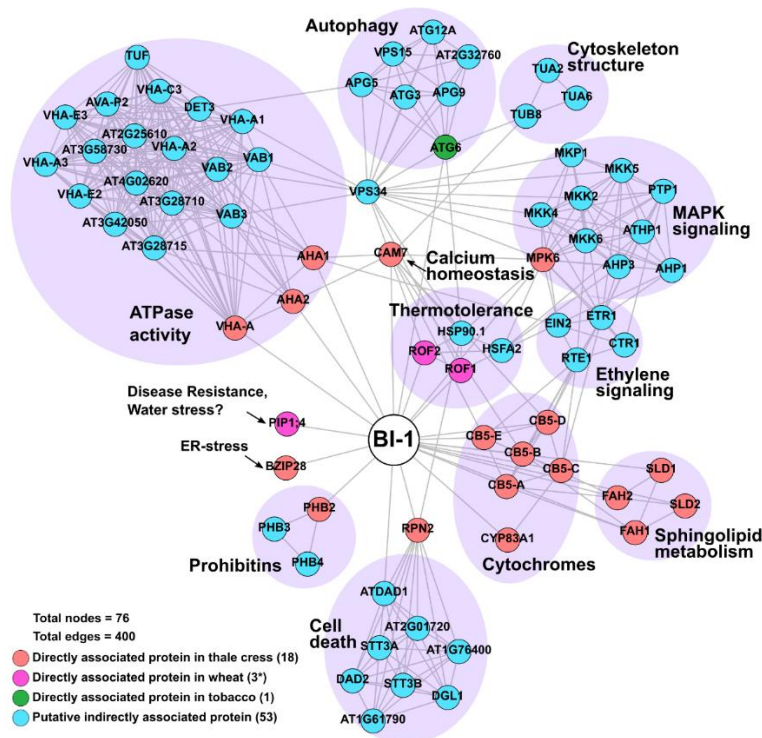


Figura 1.9. Interacciones proteína-proteína de los genes BI-1 en plantas.

Muerte celular

Similar al gen BI-1 del ser humano, los genes BI-1 de plantas pueden inhibir la muerte celular inducida por el gen Bax de mamífero. Por ejemplo, el gen OsBI-1 de arroz puede suprimir la muerte celular inducida por la expresión ectópica del gen Bax en levadura (Kawai *et al.*, 1999), y el gen AtBI-1 de Arabidopsis puede inhibir la muerte celular inducida por la expresión ectópica del gen Bax *in planta* (Kawai-Yamada *et al.*, 2001). Además, AtBI-1

puede inhibir la apoptosis inducida por Bax en cultivo de células humanas HEK293 (Bolduc *et al.*, 2003).

Calcio celular y metabolismo de esfingolípidos, modo de acción de las proteínas BI-1 en plantas

El calcio intracelular es un segundo mensajero que regula muchos procesos de la célula vegetal mediante signaturas específicas, incluyendo las respuestas a estreses ambientales (Perochon *et al.*, 2011; Luan, 2009). Los esfingolípidos son componentes de la bicapa lipídica de la membrana plasmática y participan en la señalización de vías de la muerte celular y respuesta estrés (Michaelson *et al.*, 2016). La proteína AtBI-1 de *arabidopsis* interactúa con proteínas relacionadas a la señalización de calcio y metabolismo de esfingolípidos, lo cual brinda una idea del modo de acción de AtBI-1 en plantas (Figura 1.10).

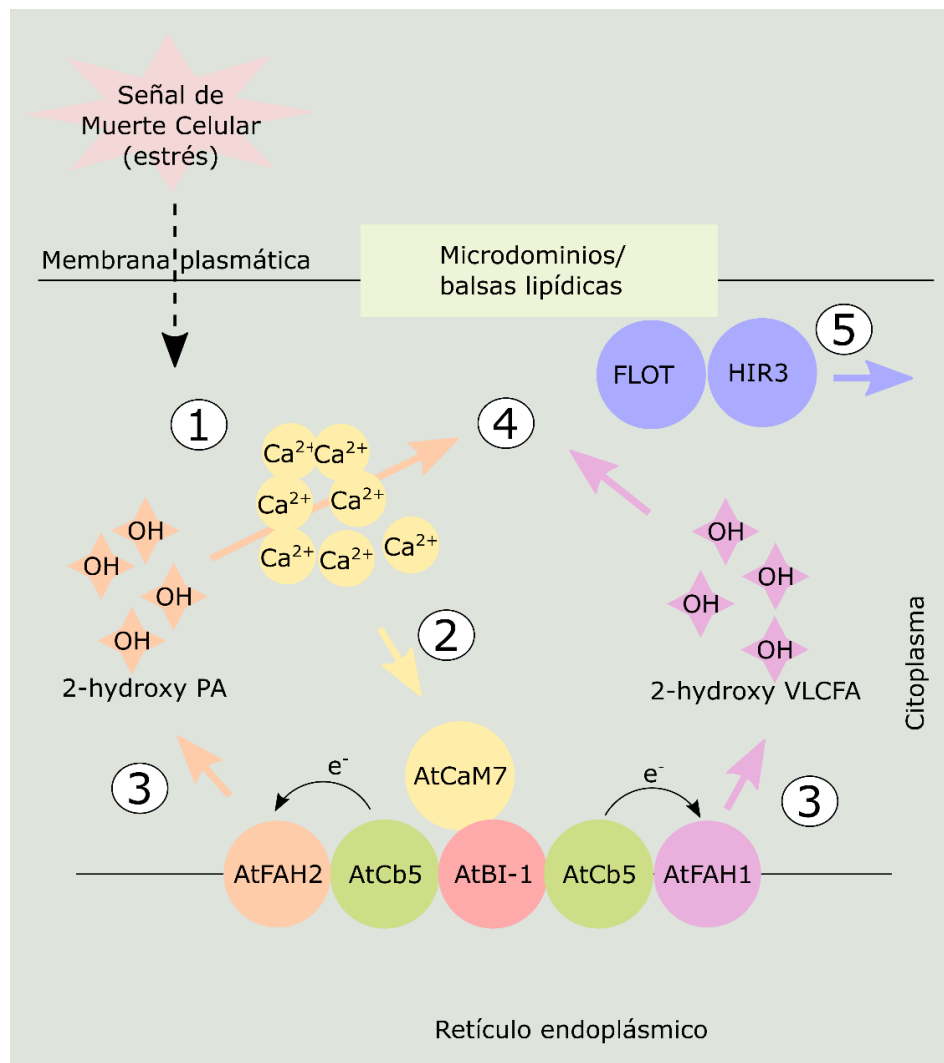


Figura 1.10. Posible mecanismo de acción de BI-1 en plantas. Cuando la célula percibe una señal de muerte celular (puede ser un estrés), activa la señalización de calcio (1). Las señales de calcio son percibidas por AtCaM7 causando su unión a la proteína AtBI-1 (2). Mediante la interacción con AtCb5, AtBI-1 promueve la activación de las hidroxilasas de esfingolípidos FAH1 (hidroxila VLCFA) y FAH2 (hidroxila PA) (3). Esta hidroxilación de esfingolípidos puede interrumpir la señal de muerte celular o contribuir al reforzamiento de las balsas lipídicas en la membrana plasmática (4). Finalmente, BI-1 reduce la acumulación de las proteínas FLOT y HIR3 —inductores de la muerte celular— en los micro-dominios de la membrana plasmática por un mecanismo aún desconocido (5). En conjunto, estas actividades llevan a la inhibición de la muerte celular mediada por BI-1 en las células vegetales. Modelo general basado en (Ishikawa *et al.*, 2015; Nagano *et al.*, 2009, 2012, 2014; Ihara-Ohori *et al.*, 2007)

AtBI-1 interactúa con una proteína calmodulina, AtCAM7, la cual es un sensor de calcio (Figura 1.10, 2) (Ihara-Ohori *et al.*, 2007). Mutaciones en la región C-terminal de AtBI-1

impiden la interacción entre AtBI-1 y AtCAM7 y la supresión de la muerte celular, por lo que dicha interacción es indispensable para la función de AtBI-1 (Ihara-Ohori *et al.*, 2007). Además, en experimentos con levadura se encontró que las bombas homólogas de SERCA (en animales) Pmr1 y Spf1, localizadas en el aparato de Golgi y retículo endoplásmico respectivamente, son indispensables para que AtBI-1 inhiba la muerte celular inducida por la expresión de BAX (Ihara-Ohori *et al.*, 2007). En ese mismo estudio se demostró que AtBI-1 en cultivos celulares BY-2 de tabaco modula las oscilaciones de calcio citosólico inducidas por tratamiento con ácido ciclopiazónico y H₂O₂. Estos estudios sugieren que el gen AtBI-1 de *Arabidopsis* realiza su función mediante la regulación del calcio citosólico a través de bombas localizadas en el retículo endoplásmico (Ihara-Ohori *et al.*, 2007).

La proteína codificada por AtBI-1 interactúa con el citocromo B5 de *Arabidopsis* (AtCb5; una proteína transportadora de electrones) en el retículo endoplásmico, y a su vez el citocromo B5 interactúa con las hidroxilasas AtFAH1 y AtFAH2 (Nagano *et al.*, 2009). La sobreexpresión de AtBI-1 incrementa los niveles de 2-hidroxilación de ácidos grasos en *Arabidopsis*, lo que sugiere que AtBI-1 regula la 2-hidroxilación de ácidos grasos por medio de la interacción con el complejo AtCb5-AtFAH1/2 (Figura 1.10, 3). AtFAH1 hidroxila preferentemente ácidos grasos de cadena muy larga (VLCFA, por sus siglas en inglés), mientras que AtFAH2 hidroxila preferentemente el ácido palmítico (PA, por sus siglas en inglés). También se ha demostrado la interacción de AtBI-1 con un complejo de proteínas formado por AtELO2, KCR1, PAS2, CER10 y AtCb5-D, las cuales también se encargan de la 2-hidroxilación de VLCFAs (Nagano *et al.*, 2019). El efecto de esta hidroxilación de ácidos grasos podría estar ligada a la interrupción de alguna señal de muerte celular llevada a cabo por esfingolípidos no hidroxilados o reforzamiento de las balsas lipídicas y micro-dominios en la membrana plasmática para proteger a las plantas contra el estrés (Figura 1.10, 4). Además, la sobreexpresión de BI-1 en arroz causa una reducción significativa de la acumulación de las proteínas FLOT (del inglés *Flotillin Homolog*) y HIR3 (del inglés, *Hypersensitive Reaction Protein3*) en los micro-dominios y balsas lipídicas de la célula (Figura 1.10, 5), dichas proteínas son reguladores positivos de la muerte celular por estrés oxidativo y ácido salicílico (Ishikawa *et al.*, 2015).

BI-1 en plantas y el estrés abiótico

En plantas los genes BI-1 regula la tolerancia a varios tipos de estrés abiótico (Watanabe y Lam, 2009). La expresión del gen AtBI-1 de *Arabidopsis thaliana* L.) aumenta en respuesta a choque térmico (Watanabe y Lam, 2006), tunicamicina (Watanabe y Lam, 2008), PEG-8000 (Duan *et al.*, 2010), H₂O₂ (Nagano *et al.*, 2012; Kawai-Yamada *et al.*, 2004) y frío (Nagano *et al.*, 2014). Mutantes de pérdida de función del gen AtBI-1 son más sensibles a tunicamicina, choque térmico, estrés hídrico y estrés oxidativo (Nagano *et al.*, 2019; Duan *et al.*, 2010; Watanabe y Lam, 2006, 2008). Sobre-expresión de AtBI-1 confiere tolerancia contra estrés oxidativo en células BY-2 de tabaco y en células en suspensión de arroz (Ishikawa *et al.*, 2010; Kawai-Yamada *et al.*, 2004), y también confiere tolerancia a la sequía en plantas de caña de azúcar (Ramiro *et al.*, 2016). La sobre-expresión del gen SIBI-1 de tomate (*Solanum lycopersicum* L.) confiere tolerancia contra choque térmico y frío en discos de hoja de tabaco (Chae *et al.*, 2003). La expresión del gen CpBI-1 de chile (*Capsicum annuum* L.) aumenta en respuesta a altas y bajas temperaturas, desecación, salinidad e inundación, su sobre-expresión confiere tolerancia contra calor, sequía y salinidad en plantas transgénicas de tabaco (Isbat *et al.*, 2009). La expresión del gen TaBI-1 de trigo (*Triticum aestivum* L.) aumenta en respuesta a estrés por calor, y su sobre-expresión ectópica rescata el fenotipo sensible a calor de las mutantes de pérdida de función de *Arabidopsis* (Lu, Zheng, *et al.*, 2018). Silenciamiento del gen GhBI-1 en algodón (*Gossypium hirsutum* L.) resulta en una tasa de muerte celular acelerada bajo condiciones de estrés salino (Zhang *et al.*, 2018). A pesar de estas evidencias que indica que los homólogos de BI-1 en plantas confieren tolerancia contra varios tipos de estrés abiótico, la sobre-expresión de AtBI-1 o TaBI-1 en *Arabidopsis* no mejora su tolerancia contra frío o estrés salino (Lu, Yu, *et al.*, 2018; Nagano *et al.*, 2014).

Otras funciones

Además de las funciones ya mencionadas (muerte celular, señalización de calcio, metabolismo de esfingolípidos y estrés abiótico) el gen BI-1 en plantas participa en muchas otras funciones como la senescencia mediada por jasmonato (Yue *et al.*, 2012), desarrollo del embrión (Reza *et al.*, 2018), la respuesta de proteínas no plegadas (Ruberti *et al.*, 2018) e interacciones planta-microbio (Hernández-López *et al.*, 2018; Eichmann *et al.*, 2010; Babaeizad *et al.*, 2009; Imani *et al.*, 2006; Hükelhoven *et al.*, 2003) (Figura 1.9).

1.2.6 Funciones de los genes LFG en plantas

La sobreexpresión de HvLFGa en cebada y de AtLFG1 y AtLFG2 en arabidopsis promueve la infección por el moho polvoriento, un hongo biótrofo; silenciamiento, o mutaciones de pérdida de función tienen el efecto contrario (Weis *et al.*, 2013). Además, AtLFG2, también conocido como BIL4, regula positivamente la vía de señalización de brasinoesteroides (BR) (YAMAGAMI *et al.*, 2009). La expresión de los genes AtLFG1 y AtLFG2 aumenta en respuesta a tunicamicina y salinidad, y tienen una función dual en la regulación de la respuesta de proteínas no plegadas (UPR por sus siglas en inglés) y la muerte celular para proteger a las plantas contra el estrés del retículo endoplásmico (Guo *et al.*, 2018). La expresión de TaGAAP, una proteína LFG de trigo, aumenta en respuesta a estrés por calor y es activado directamente por el factor de transcripción TaHsfA6f (Xue *et al.*, 2015).

No se tienen más datos acerca de la función de los genes LFG en plantas. No se ha estudiado su función en algún tipo de estrés abiótico como la salinidad o sequía.

1.2.7. Función dual de los genes TMBIM en animales, levaduras y plantas

A pesar de los genes TMBIM por lo general ha sido considerados cito-protectores, existen también evidencias de que los mismos genes pueden inducir la muerte, o tienen funciones duales en la regulación de varios procesos.

El gen BI-1 de humano, el cual protege a las células de la muerte celular, también puede inducir la muerte celular en células HT1080 bajo condiciones ácidas (Kim *et al.*, 2008). Interesantemente, realiza esta actividad mediante la vía intrínseca de la mitocondria. También, el gen AtBI-1 de arabidopsis, el cual protege a las células contra la muerte celular y condiciones estresantes, puede inducir la muerte celular en células HT1080, por una vía similar a la vía intrínseca de la apoptosis (Li-Hua *et al.*, 2002). El gen BXI1p de levadura, que protege a las células de levadura contra choque térmico, también puede inducir la muerte celular cuando las células son sometidas a H₂O₂ y ácido acético, también por una vía mitocondrial (Büttner *et al.*, 2011).

El gen BI-1 de tabaco interactúa con ATG6, un regulador de la autofagia en plantas. El silenciamiento de BI-1 reduce la actividad de autofagia mediada por la vía mediata por el gen N del virus de mosaico de tabaco y viológeno, pero indujo la muerte celular programada.

Además, la sobre-expresión de BI-1 aumenta la autofagia, pero también aumentó la muerte celular. Los autores de dicho estudio concluyeron que el gen BI-1 tiene una función doble: inducir o inhibir la muerte celular dependiendo de las condiciones fisiológicas (Xu *et al.*, 2017). También, el gen AtBI-1 atenúa la función protectora de bzip28 —un regulador positivo de la tolerancia a estrés por retículo endoplásmico mediante la activación de la respuesta de proteínas no plegadas— por lo tanto, AtBI-1 reduce la tolerancia de la planta ante el estrés en lugar de mejorarla (Ruberti *et al.*, 2018).

El gen AtLFG1 de arabidopsis puede inhibir la muerte celular inducida por Bax en levadura, pero también puede inducir la muerte celular cuando es sobre-expresado en tabaco y arabidopsis (Sierla *et al.*, 2019).

1.2.8 Potencial biotecnológico de los genes TMBIM en plantas

Las funciones de los genes AtBI-1 y AtLFG1-5 de Arabidopsis han sido ampliamente estudiadas, demostrando su potencial biotecnológico para el mejoramiento genético y hacer frente al estrés abiótico. Sin embargo, es necesario caracterizar las funciones de estos genes en otras especies que sean de interés agrícola. En el presente trabajo se estudiará la función de los genes BI y LFG de *Carica papaya*.

1.3 *Carica papaya* L.

La planta de papaya (*Carica papaya* L.) pertenece a la familia Caricaceae, dentro del orden Brassicales. Probablemente se originó en el sur de México y Centroamérica, pero actualmente se cultiva en zonas tropicales y subtropicales de varias partes del mundo.

1.3.1 Biología de la planta de papaya

La planta de papaya es una herbácea semi-leñosa que puede alcanzar hasta 10 m de altura (Figura 1.11), aunque con las técnicas de cultivo modernas no sobrepasa los 6 m, esto para facilitar la recolección de los frutos. Generalmente presenta un único tallo hueco sin ramas laterales y con cicatrices, resultantes de la abscisión de las hojas. Las hojas son grandes (aproximadamente 0.6 m²), se encuentran agrupadas en la sección superior de los individuos adultos (arregladas en forma de hélice super-enrollada), duran de 3-6 meses

bajo condiciones tropicales adecuadas y continuamente emergen nuevas en el ápice conforme se caen las de la parte inferior (Jiménez *et al.*, 2014).



Figura 1.11. Planta de papaya.

La temperatura óptima para el crecimiento de la planta de papaya se encuentra entre 21-33°C, en estas condiciones la papaya puede producir hasta 2 hojas por semana y de 8-16 frutos por mes. Las plantas son sensibles al frío, por lo general no toleran temperaturas menores a 10° C. Requieren de grandes cantidades de riego, entre 1000 y 1800 mm pluviales al año (Jiménez *et al.*, 2014; Silva *et al.*, 2007).

La expresión sexual está regulada por un sistema incipiente de cromosomas X-Y y las flores pueden presentar 3 formas sexuales: femenina, masculina y hermafrodita (Jiménez *et al.*, 2014; Silva *et al.*, 2007). El desarrollo de las plantas de papaya es bastante rápido, la floración ocurre aproximadamente entre 3 y 8 meses después de la germinación de las semillas y a los 9-15 meses ya pueden ser cosechados sus frutos (Jiménez *et al.*, 2014).

1.3.2 Fruto de papaya

Los frutos de papaya son bayas climatéricas que presentan una gran diversidad de formas y tamaños. Los frutos de las plantas hermafroditas tienden a ser alargados y tienen una

forma cilíndrica, mientras que los frutos de las plantas femeninas tienden a ser redondos (Figura 1.12). El peso del fruto puede variar de menos de 100 g en algunas plantas silvestres hasta más de 10 kg en algunas variedades (Jiménez *et al.*, 2014).



Figura 1.12. Fruto de papaya. El fruto del lado derecho proviene de una planta hermafrodita, y la del lado izquierdo de una planta femenina.

El fruto de papaya es muy nutritivo: consiste principalmente de agua y carbohidratos, es bajo en calorías (5 calorías por cada 100 g) y rico en vitaminas y minerales, particularmente en las vitaminas A y C. La papaya se puede consumir en fresco, en mermeladas, conservados, postres y en algunas recetas de cocina. Por otra parte, del fruto de papaya se obtiene una cisteína proteasa importante para la industria llamada papaína, la cual se utiliza ampliamente para ablandar carne, tratamientos de heridas y quemaduras, y también se utiliza como producto cosmético (Silva *et al.*, 2007).

México se ha mantenido entre los principales productores de papaya entre 2000 y 2017 (Figura 1.13). En 2017, por tercer año consecutivo, México ocupó el tercer lugar mundial en producción total de fruto de papaya por año con 961,768 toneladas, precedido por Brasil con 1,057,101 toneladas e India con 5,940,000 toneladas (FAOSTAT, 2019). Dentro de nuestro país, los estados con mayor producción durante 2003 y 2018 han sido Oaxaca, Colima, Chiapas, Veracruz, Michoacán, Guerrero, Yucatán y Tabasco (SIAP, 2019) (Figura 1.14).

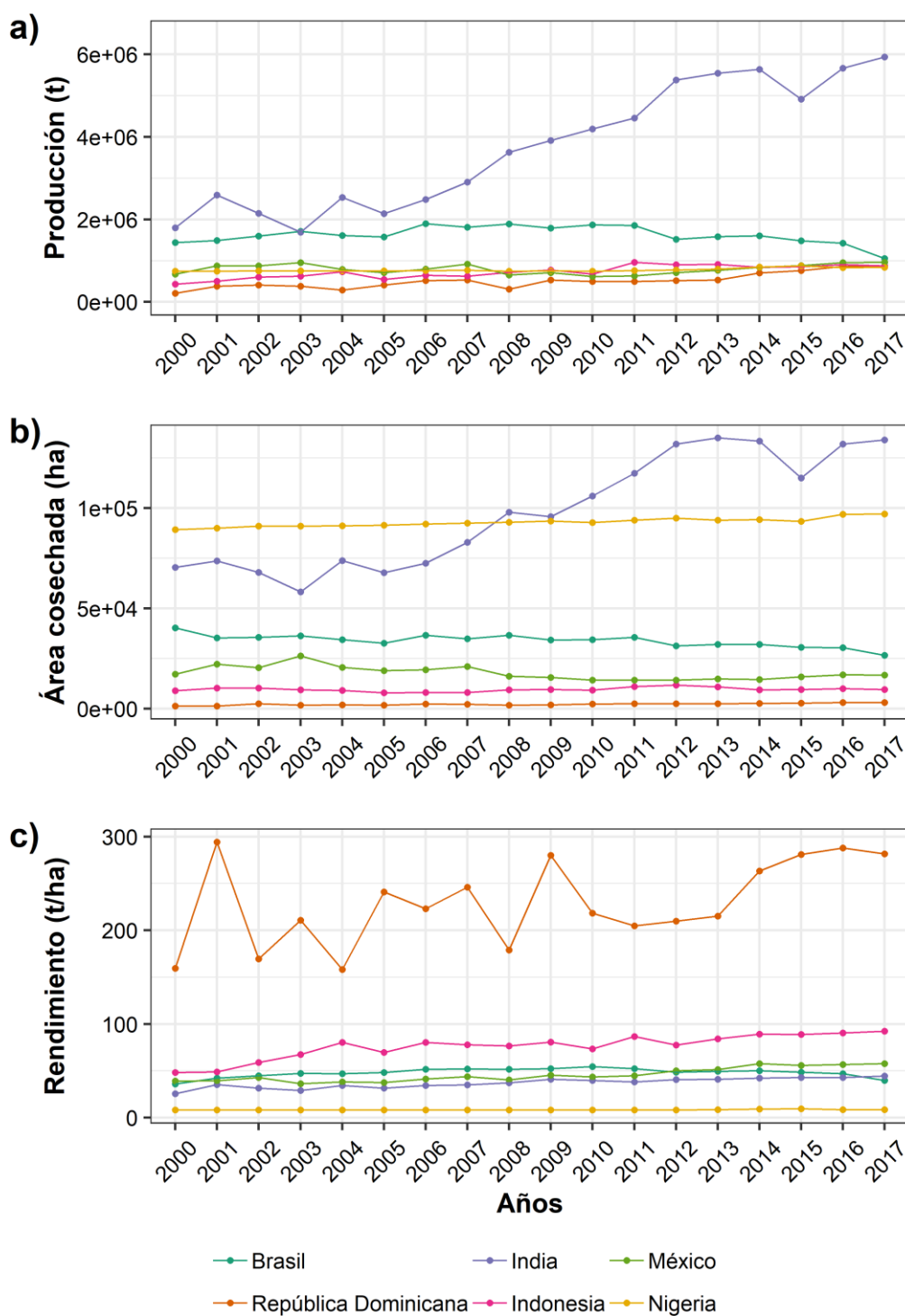


Figura 1.13. Principales países productores de papaya en el mundo entre 2000 y 2017. Datos tomados de FAOSTAT, 2019.

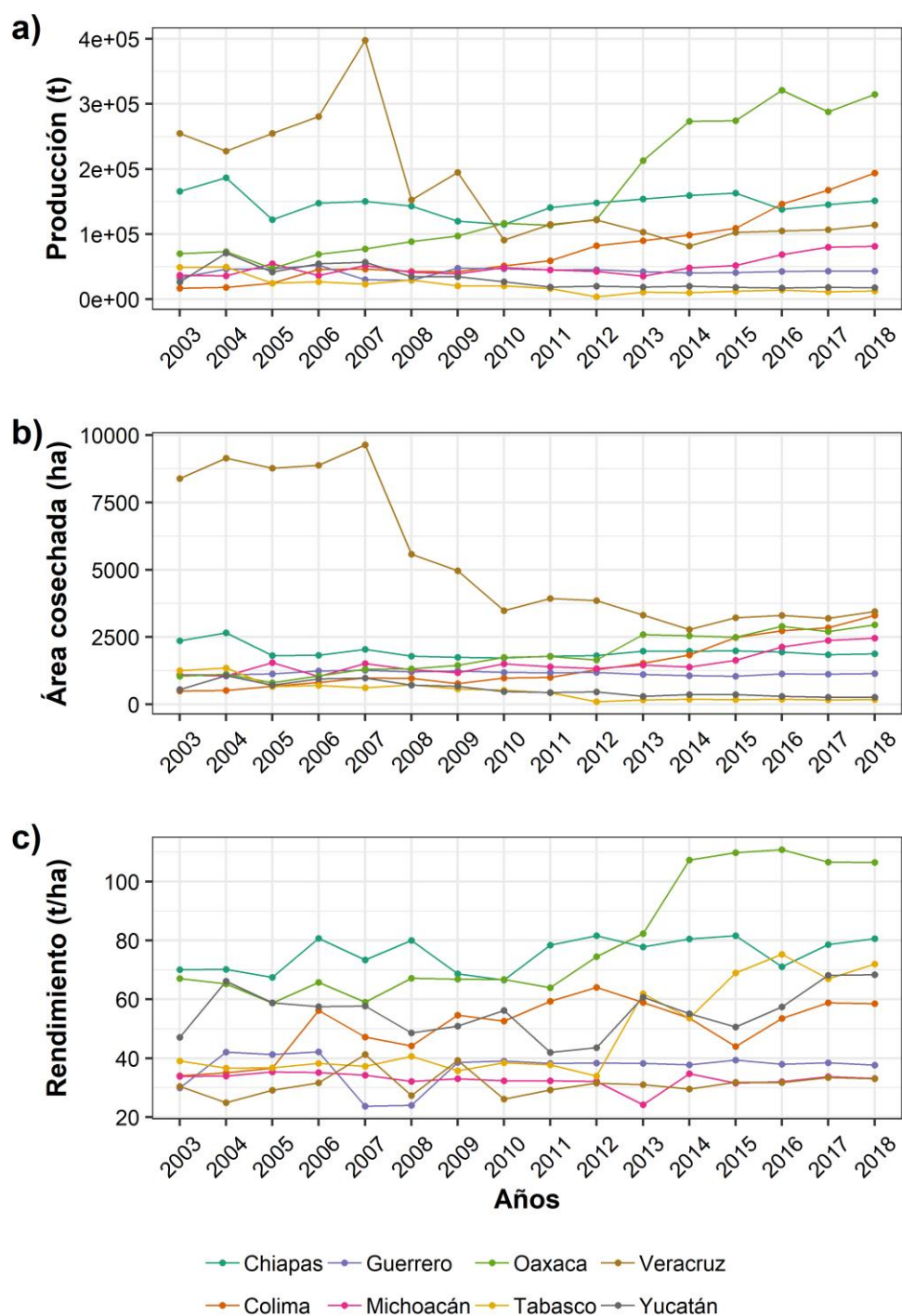


Figura 1.14. Principales estados productores de papaya en México entre 2003 y 2017. Datos tomados de SIAP, 2019.

1.3.3 Papaya y estrés abiótico

Carica papaya está clasificada como una planta con metabolismo C3 (Campostrini y Glenn, 2007). Bajo condiciones de 2000 $\mu\text{mol m}^{-2} \text{s}^{-1}$ de densidad de flujo de fotones fotosintéticos (PPFD), las plantas pueden alcanzar valores de asimilación neta de carbono (A) de hasta 25-20 $\mu\text{mol CO}_2 \mu\text{mol m}^{-2} \text{s}^{-1}$. La asimilación de CO_2 puede ser limitada por factores ambientales como la luz, temperatura, disponibilidad de agua y concentración de sales.

Salinidad

En cuanto a la sensibilidad de los cultivos de papaya a la salinidad, los resultados son variados. La germinación de las semillas se ve inhibida por concentraciones muy bajas de NaCl; sin embargo, el crecimiento de las plántulas se ve estimulada por concentraciones de 3200 ppm de sólidos totales disueltos (Kottenmeier *et al.*, 1983).

Plantas de la variedad 'Maradol' regadas con agua salina de conductividad eléctrica superior a 4 dS m^{-1} , sufren una reducción significativa en el índice estomático, densidad estomática, longitud del tallo, acumulación de biomasa, área foliar, cantidad de clorofila y niveles de K^+ , Zn^{2+} , Cu^{2+} y Fe^{2+} . A su vez, las plantas experimentan un aumento significativo en el grosor de la lámina foliar y en los niveles de N y Na^+ (Parés y Basso, 2013; PARÉS *et al.*, 2008).

En condiciones de choque salino (300 mM) en hidroponía, las plantas de papaya reducen su tasa fotosintética, aumentan la concentración de prolina y sufren daño oxidativo en las hojas (Gamboa-Tuz, 2015).

Sequía

Durante el estrés por sequía, el crecimiento de las plantas de papaya se detiene, se induce la abscisión foliar y disminuyen la tasa fotosintética, conductancia estomática y tasa de transpiración (Mahouachi *et al.*, 2006, 2007).

Las plantas de papaya son consideradas relativamente resistentes a la sequía y han sido clasificadas como una especie que responde a dicho estrés mediante el aplazamiento de la desecación. El potencial hídrico de las hojas se ve poco afectado por la sequía; esto

probablemente se deba a un ajuste osmótico provocado por la acumulación de los iones inorgánicos Na^+ , K^+ y Cl^- (Mahouachi *et al.*, 2006). También se ha reportado un aumento significativo en las concentraciones de ácido abscísico (Mahouachi *et al.*, 2007) y contenido de prolina (Mahouachi *et al.*, 2012). Además, la fluorescencia de clorofila no se ve afectada por el estrés hídrico (Marler y Mickelbart, 1998).

Por otra parte, la rehidratación estimula de nuevo el crecimiento y la emergencia de hojas nuevas, se reactiva la fotosíntesis y se reduce la concentración de iones (Na^+ , K^+ y Cl^-), ácido abscísico y prolina a sus valores normales, indicando que el estrés por sequía no tiene efectos irreversibles sobre estos parámetros (Mahouachi *et al.*, 2006, 2007, 2012).

1.3.4 Genómica y transcriptómica de papaya

Genoma

Papaya es una planta diploide ($2n=18$), posee un genoma relativamente pequeño de 370 megabases distribuidas en 9 cromosomas. Se ha secuenciado el genoma completo de la variedad 'SunUp', una variedad transgénica resistente al virus de la mancha anular de la papaya (PRSV), la cual proviene de la variedad 'Sunrise'. Papaya tiene una reducción de entre el 11 y el 20% de sus genes con respecto a Arabidopsis, con quien compartió un ancestro común hace 72 millones de años. El genoma de papaya contiene menos genes de resistencia, pero también contiene más genes relacionados a la producción de almidón y volátiles (Ming *et al.*, 2008, 2012).

Transcriptoma

Se han reportado varios estudios del transcriptoma de papaya de diferentes tejidos y condiciones, y utilizando diferentes técnicas. Se ha reportado el transcriptoma de las raíces mediante una biblioteca de cDNA (Porter *et al.*, 2008); el transcriptoma del fruto en proceso de maduración mediante cDNA-AFLP (Fabi *et al.*, 2010); el transcriptoma de flores masculinas, femeninas y hermafroditas para identificar los genes determinantes del sexo usando superSAGE (Urasaki *et al.*, 2012); el transcriptoma del proceso de reversión del sexo de flor masculina a hermafrodita inducida por frío usando RNA-seq (Lin *et al.*, 2016); el transcriptoma de la variedad 'Sunup' vs 'Sunset' para identificar los mecanismos de

tolerancia a PRSV usando RNA-seq (Fang *et al.*, 2016); el transcriptoma de callos embriogénicos usando RNA-seq (Jamaluddin *et al.*, 2017); y el transcriptoma de flores infectadas con el virus de Meleira para conocer los mecanismos de respuesta ante la enfermedad pegajosa de la papaya usando RNA-seq (Madroñero *et al.*, 2018).

Hasta la fecha no se ha llevado a cabo un estudio dirigido al estudio del transcriptoma de papaya para conocer sus respuestas a algún tipo de estrés abiótico, como choque de calor, sequía o salinidad.

1.3.5 La superfamilia TMBIM en papaya

En el genoma de papaya se encuentran codificadas cuatro proteínas TMBIM: una de la familia BI (CpBI1— y tres de la familia LFG—CpLFG1, CpLFG2 y CpLFG3 (Gamboa-Tuz, 2015).

La proteína codificada por el gen CpB-1 tiene 246 aminoácidos de longitud (27.36 kDa). Esta proteína es 67% idéntica a la proteína OsBI-1 de arroz y 79% idéntica a la proteína AtBI-1 de Arabidopsis. La proteína CpBI-1 es más lejana filogenéticamente de las proteínas BI-1 que han sido analizadas funcionalmente en plantas (Figura 1.15a). Análisis predictivos indican que esta proteína solo cuenta con seis regiones transmembrana (Figura 1.15b); sin embargo, la comparación de la secuencia de la proteína CpBI-1 con otros homólogos en plantas demuestra una alta conservación de su secuencia a lo largo de toda la proteína, incluyendo las siete regiones transmembranales (Figura 1.15c). Los dos residuos de aspartato (D) que son necesarios para regular su actividad de canal iónico también se encuentran conservados en esta proteína (Figura 1.15c). Además, la proteína CpBI-1 tiene una región rica en lisina (K+) en su extremo carboxilo, la cual es probable que forme una estructura súper-enrollada (característica de las proteínas de la familia BI en eucariontes; Figura 1.15b,c).

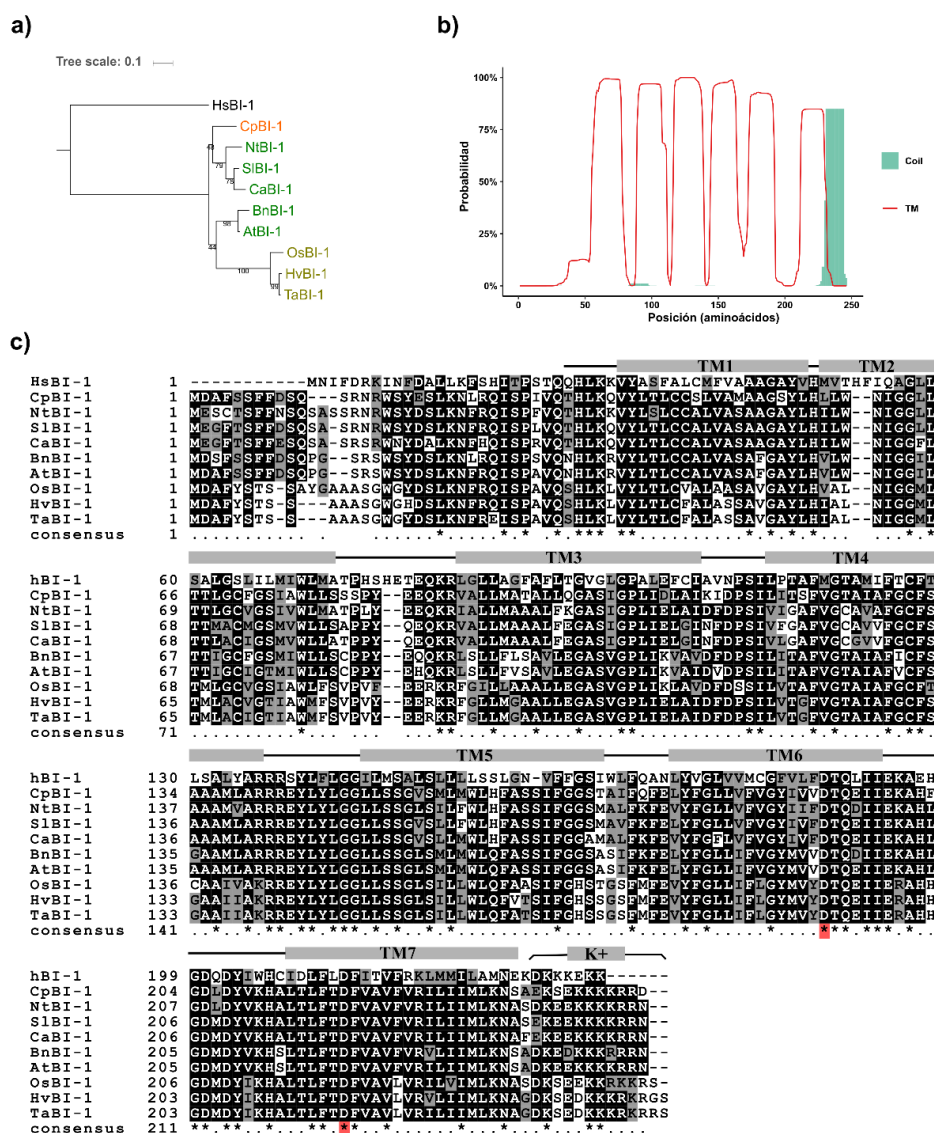


Figure 1.15. Análisis de la secuencia la proteína CpBI-1 de papaya. a) Árbol filogenético de las proteínas BI-1 en plantas. b) Probabilidad por posición de aminoácidos de formar una hélice transmembranal (TMhelix) y estructura superenrollada (Coil) de la proteína CpBI-1 de papaya. c) Alineamiento múltiple de las secuencias BI-1 de humano (Hs), arroz (Os), cebada (Hv), Trigo (Ta), Arabidopsis (At), canola (Bn), papaya (Cp), tabaco (Nt), chile (Ca) y tomate (Sl). Los aminoácidos idénticos y los similares están marcados en negro y gris, respectivamente (50% por posición).

El gen CpBI-1 aumenta su expresión en hojas en tratamiento de desecación en papel filtro (Figura 1.16ba). También, aumenta su expresión en respuesta a un choque salino de 300

mM NaCl en hojas plantas de papaya crecidas en condiciones de hidroponía (Figura 1.16a). Por el contrario, su expresión en la raíz baja en ambas condiciones de estrés (Figura 1.16a,b) (Gamboa-Tuz, 2015).

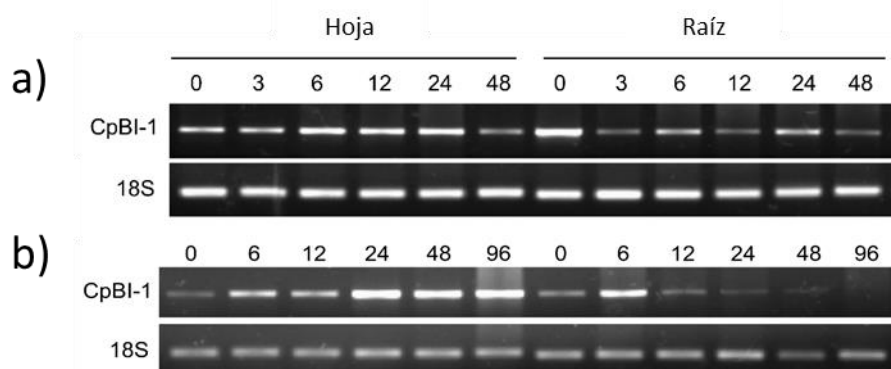


Figura 1.16. Expresión del en CpBI-1 de papaya en condiciones de estrés abiótico. a) Tratamiento de desecación de las raíces en papel filtro. b) Tratamiento de choque salino (300 mM). Datos tomados de Gamboa-Tuz, SD, 2015.

El gen CpLFG1 codifica una proteína de 234 aminoácidos (26.15 kDa), la cual tiene una identidad del 63% con la proteína AtLFG4 y una similitud de 78% con las proteínas AtLFG1 y AtLFG2 de Arabidopsis. El gen CpLFG2 codifica una proteína de 242 aminoácidos (27.16 kDa) y tiene una identidad de 82% con AtLFG2 y de 78 y 79% con AtLFG3 y AtLFG4, respectivamente. El gen CpLFG3 codifica una proteína de 257 aminoácidos (28.14 kDa) y tiene 79% de identidad con AtLFG5. Las proteínas CpLFG1 y CpLFG2 pertenecen a un solo clado junto con las proteínas AtLFG1-3; la proteína CpLFG3 pertenece a un clado diferente junto con la proteína AtLFG5 (Figura 1.17a). Los tres genes LFG de papaya contienen siete regiones transmembrana (Figura 1.17b). La región del dominio está muy conservada en estas proteínas, pero son variables en extremo del amino-terminal (Figura 1.17c). Los dos residuos de aspartato necesarios para cumplir con su función de canal iónico también están conservados en las proteínas LFG de papaya.

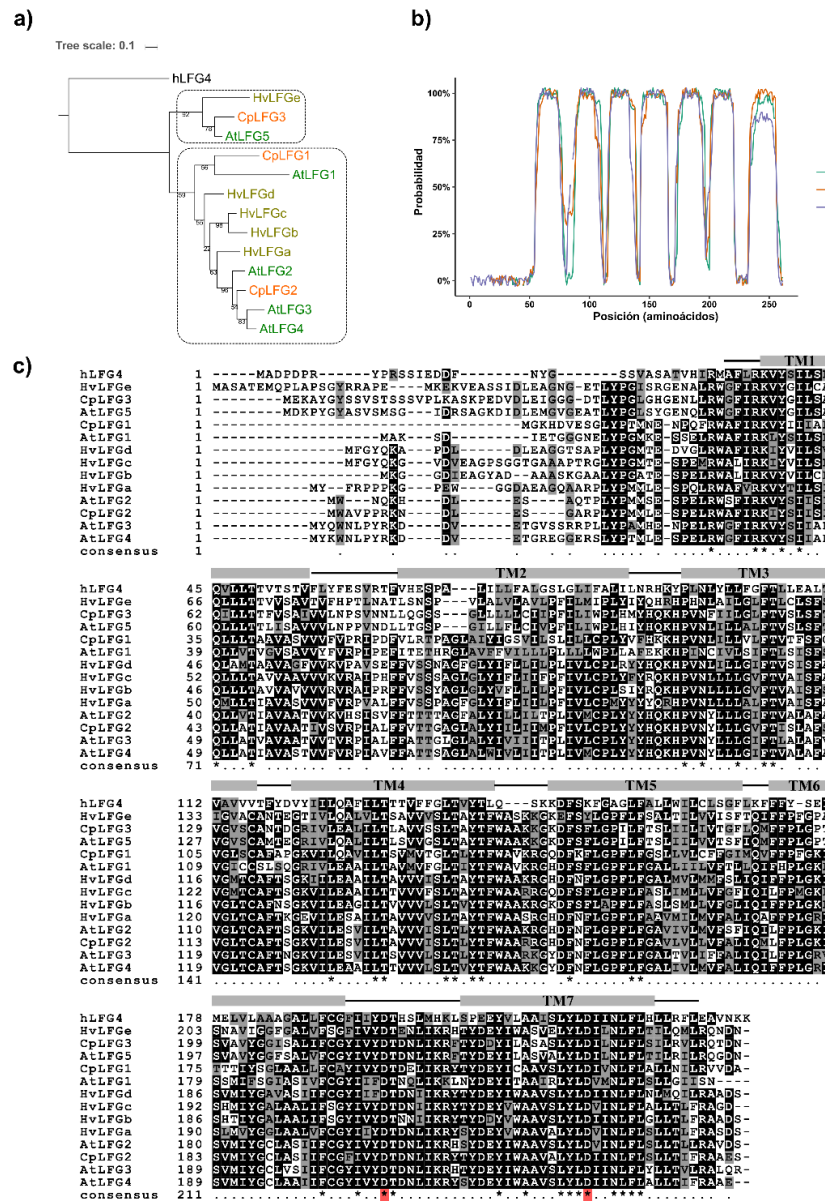


Figura. 1.17. Análisis de las secuencias de proteína de CpLFG1, CpLFG2 y CpLFG3 de papaya. a) Árbol filogenético de las proteínas LFG en plantas. b) Probabilidad por posición de aminoácidos de formar una hélice transmembranal (TMhelix) de las proteínas CpLFG1, CpLFG2 y CpLFG3 de papaya. c) Alineamiento múltiple de las secuencias LFG de humano (Hs), cebada (Hv), Arabidopsis (At) y papaya (Cp). Los aminoácidos idénticos y los similares están marcados en negro y gris, respectivamente (50% por posición).

En condiciones de desecación el gen CpLFG1 presentó un patrón irregular de expresión, esto probablemente se deba a que no está involucrado en el estrés osmótico o debido a la baja expresión de este gen (Figura 1.18). En condiciones de choque salino por 300 mM NaCl, el gen CpLFG1 no se expresó sino hasta el final del tratamiento, posiblemente actuando en una situación de estrés extrema (Figura 1.18). El gen CpLFG3 aumentó su expresión en hojas tanto en el estrés por desecación como en el estrés por salinidad; su expresión bajó en las raíces en ambas condiciones de estrés (Gamboa-Tuz, 2015).

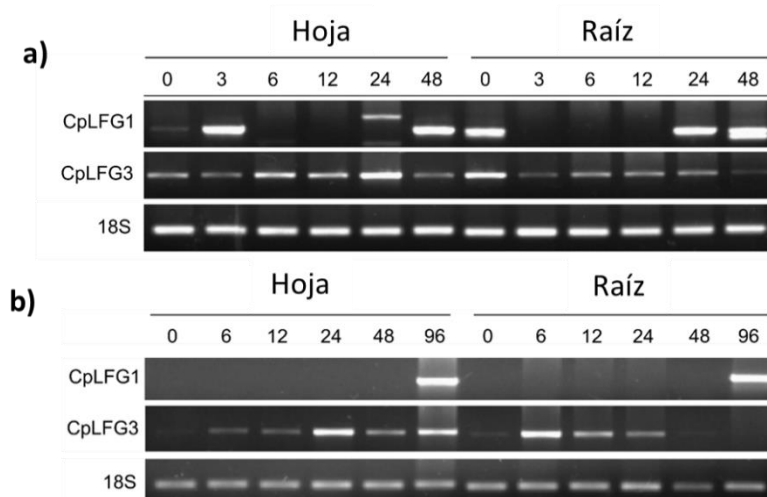


Figura 1.18. Expresión de los genes CpLFG1 y CpLFG2 en condiciones de estrés abiótico. a) Tratamiento de desecación de las raíces en papel filtro. a) Tratamiento de choque salino (300 mM NaCl). Datos y figuras obtenidas de Gamboa-Tuz, SD, 2015.

A pesar de la respuesta de los genes TMBIM de papaya (CpBI-1, CpLFG1 y CpLFG3) a desecación y sequía, se desconoce si desempeñan alguna función en la tolerancia a algún tipo de estrés abiótico.

HIPÓTESIS

Alguno de los genes CpBI-1, CpLFG1 o CpLFG3 confieren tolerancia al estrés abiótico.

JUSTIFICACIÓN

El estrés abiótico puede resultar devastador para los cultivos agrícolas. Es necesaria la caracterización funcional de genes vegetales con potencial biotecnológico para generar conocimiento básico que, a la larga, permita la generación de variedades agrícolas más tolerantes a uno o varios tipos de estrés abiótico.

La superfamilia de genes TMBIM son reguladores de la muerte celular conservados en procariontes y eucariontes, y en eucariontes se dividen en las familias BI y LFG. Sin embargo, no se ha estudiado la superfamilia completa en los tres dominios de la vida—arqueas, bacterias y eucariontes. Tampoco se ha caracterizado la evolución de las familias BI y LFG en plantas tomando en cuenta las relaciones de ortología y sintenia. Esta información es necesaria para explicar las posibles funciones de estos genes.

La planta de papaya (*Carica papaya* L.) es cultivada en regiones tropicales y subtropicales. A nivel mundial, México ocupa el tercer lugar en producción y primero en exportación. A pesar de que se cuenta con el genoma de la variedad ‘Sunup’, no se cuenta con estudios transcriptómicos acerca de su respuesta a estreses abióticos, como la sequía.

Análisis funcionales de genes de las familias BI y LFG de varias especies vegetales han demostrado que tienen el potencial de conferir tolerancia a varios tipos de estrés abiótico. Se han aislado tres genes TMBIM de papaya: uno de la familia BI, CpBI-1, y dos de la familia LFG, CpLFG1 y CpLFG3, los cuales responden a estrés salino y desecación. Sin embargo, no se tienen datos de su papel en la respuesta de papaya al estrés por sequía, y tampoco se ha probado que puedan conferir tolerancia a algún tipo de estrés.

En el presente trabajo se llevó a cabo un análisis filogenómico de toda la familia TMBIM en arqueas, bacterias y eucariontes, con énfasis en la evolución independiente de las familias BI y LFG en plantas. Se estudió el transcriptoma de papaya en tratamiento de sequía mediante RNA-seq, y se analizó si los genes TMBIM de papaya se encuentran involucrados en la respuesta general transcripcional de papaya. Finalmente, se evaluó si los genes CpBI-1, CpLFG1 y CpLFG3 son capaces de conferir tolerancia a un estrés por choque salino.

OBJETIVO GENERAL

Determinar la función de los genes CpBI-1, CpLFG1 y CpLFG3 de papaya (*Carica papaya* L). en la tolerancia al estrés abiótico.

OBJETIVOS ESPECÍFICOS

1. Reconstruir la filogenia de la superfamilia TMBIM en arqueas, bacterias y eucariontes.
2. Reconstruir la historia evolutiva de las familias BI y LFG en plantas.
3. Generar un transcriptoma de referencia de la planta de *Carica papaya* L. bajo condiciones de sequía.
4. Evaluar la tolerancia de plantas transgénicas de tabaco (*Nicotiana tabacum* L.) transformadas con los genes CpBI-1, CpLFG1 y CpLFG3 bajo condiciones de estrés abiótico por choque salino (NaCl).
5. Medir la expresión de los genes CpBI-1, CpLFG1 y CpLFG3 en el transcriptoma de *Carica papaya* L. bajo condiciones de sequía.

ESTRATEGIA EXPERIMENTAL

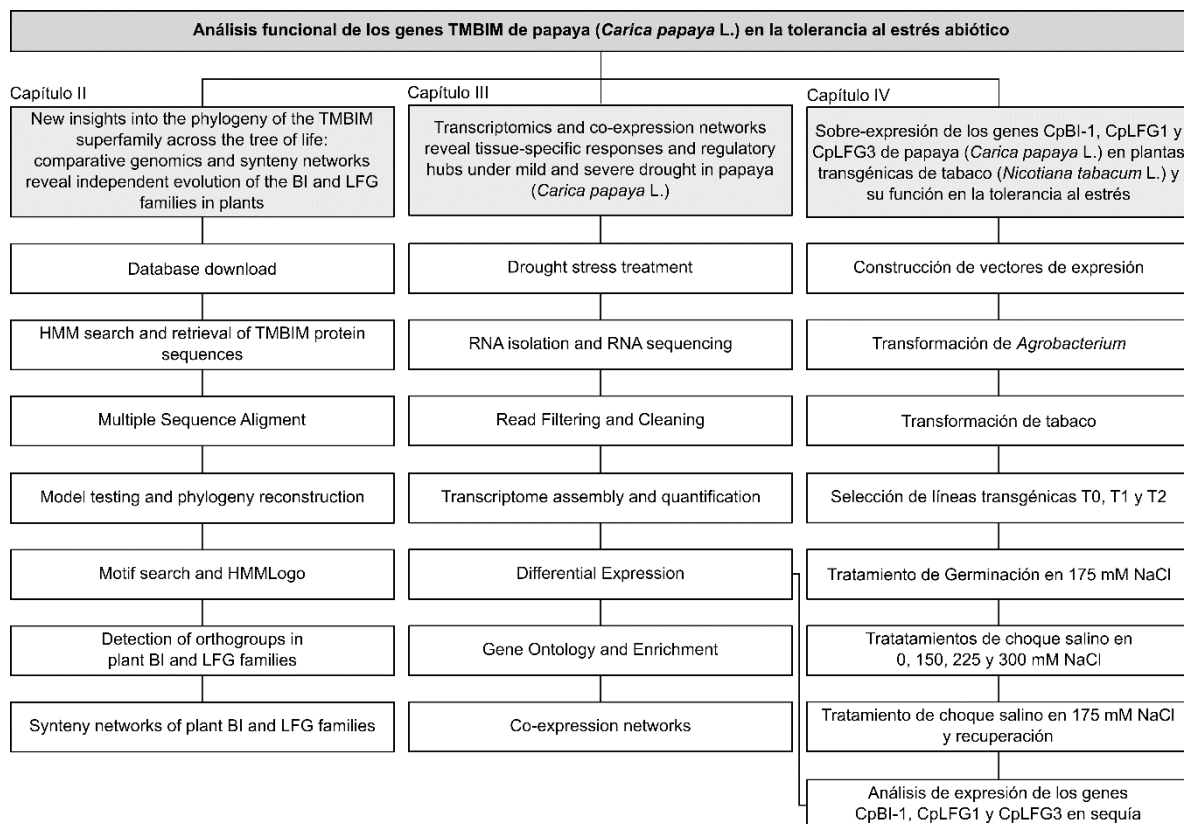


Figura 1.19. Estrategia experimental.

CAPÍTULO II**NEW INSIGHTS INTO THE PHYLOGENY OF THE TMBIM SUPERFAMILY
ACROSS THE TREE OF LIFE: COMPARATIVE GENOMICS AND SYNTENY
NETWORKS REVEAL INDEPENDENT EVOLUTION OF THE BI AND LFG
FAMILIES IN PLANTS****Authors:**

Samuel D. Gamboa-Tuz^a, Alejandro Pereira-Santana^b, Tao Zhao^b, M. Eric Schranz^b, Enrique Castano^c, Luis C. Rodriguez-Zapata^a

Affiliations:

^a Biotechnology Unit, Centro de Investigacion Cientifica de Yucatan, 97205 Yucatan, Mexico.

^b Biosystematics Group, Wageningen University and Research, 6708 PB Wageningen, The Netherlands.

^c Biochemistry and Molecular Biology Unit, Centro de Investigacion Cientifica de Yucatan, 97205 Yucatan, Mexico.

Journal:

Molecular Phylogenetics and Evolution.

Volume 126, September 2018, Pages 266-278.

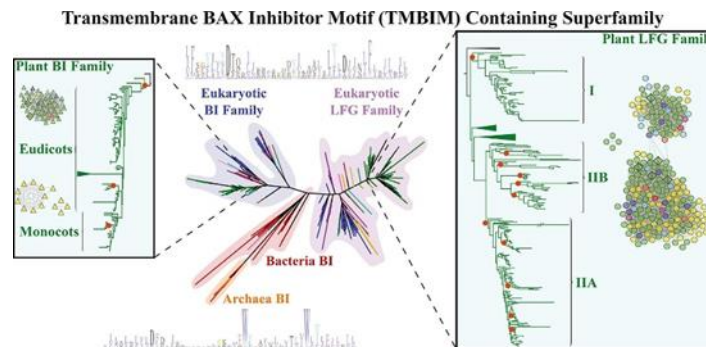
<https://doi.org/10.1016/j.ympev.2018.04.032>.

Status:

Received 1 December 2017; Received in revised form 19 April 2018; Accepted 20 April 2018
Available online 25 April 2018.

ABSTRACT

The Transmembrane BAX Inhibitor Motif containing (TMBIM) superfamily, divided into BAX Inhibitor (BI) and Lifeguard (LFG) families, comprises a group of cytoprotective cell death regulators conserved in prokaryotes and eukaryotes. However, no research has focused on the evolution of this superfamily in plants. We identified 685 TMBIM proteins in 171 organisms from Archaea, Bacteria, and Eukarya, and provided a phylogenetic overview of the whole TMBIM superfamily. Then, we used orthology and synteny network analyses to further investigate the evolution and expansion of the BI and LFG families in 48 plants from diverse taxa. Plant BI family forms a single monophyletic group; however, monocot BI sequences transposed to another genomic context during evolution. Plant LFG family, which expanded through whole genome and tandem duplications, is subdivided in LFG I, LFG IIA, and LFG IIB major phylogenetic groups, and retains synteny in angiosperms. Moreover, two orthologous groups (OGs) are shared between bryophytes and seed plants. Other several lineage-specific OGs are present in plants. This work clarifies the phylogenetic classification of the TMBIM superfamily across the three domains of life. Furthermore, it sheds new light on the evolution of the BI and LFG families in plants providing a benchmark for future research.



Keywords:

Bax inhibitor 1; Gene family evolution; Lifeguard; Programmed cell death; Synteny network; TMBIM

RESUMEN

La superfamilia de proteínas *Transmembrane BAX Inhibitor Motif containing* (TMBIM), la cual se divide en las familias *Bax Inhibitor* (BI) y *Lifeguard* (LFG), comprende un grupo de reguladores de la muerte celular que protegen a las células en procariontes y eucariontes. Sin embargo, hasta ahora no se ha investigado la evolución de esta superfamilia en plantas. En este trabajo, se identificaron 685 proteínas TMBIM en 171 especies de arqueas, bacterias y eucariontes, y se llevó a cabo un análisis filogenético de toda la superfamilia TMBIM. Además, se llevaron a cabo análisis de ortología y redes de sintenia para profundizar en la evolución y expansión de las familias BI y LFG en 48 especies de plantas. En plantas, la familia BI forma un solo grupo monofilético; sin embargo, las secuencias BI de las plantas monocotiledóneas se trasladaron a otro contexto genómico durante su historia evolutiva. La familia LFG en plantas, la cual se expandió a través de duplicaciones de genoma completo y en tándem, está dividida en los grupos LFG I, LFG IIA y LFG IIB, y conserva su sintenia en las angiospermas. Dos grupos ortólogos (OGs) están compartidos entre briofitas y plantas con semilla. Se encontraron otros OGs específicos de ciertos linajes de plantas. Este trabajo clarifica la clasificación filogenética de la superfamilia TMBIM a través de los tres dominios de la vida. Además, brinda nuevo conocimiento sobre la evolución de las familias BI y LFG en plantas y provee un marco de referencia para futuras investigaciones.

2.1 INTRODUCTION

Programmed Cell Death (PCD) is essential for cellular homeostasis, development, and environmental responses of multicellular organisms (Lord y Gunawardena, 2012); its importance in unicellular organisms is also recognized (Jiménez-Ruiz *et al.*, 2010). Most research regarding PCD has focused on apoptosis and the animal B-Cell Lymphoma 2 (BCL-2) gene family, which contains both pro- and anti-apoptotic members (Chipuk *et al.*, 2010). No BCL-2 homologues have been identified outside the animal kingdom. However, biological functions of the BCL-2 family remain conserved in fungi and plants. For example, heterologous expression of the human pro-apoptotic BCL-2 Associated X (BAX) gene/protein activates cell death in budding yeast (*Saccharomyces cerevisiae*) and Arabidopsis (*Arabidopsis thaliana*) (Baek *et al.*, 2004; Priault *et al.*, 2003). Another group of PCD regulators, linked to BCL-2, is the Transmembrane BAX Inhibitor Motif containing (TMBIM) superfamily, also referred to as BAX Inhibitor-1 (BI-1). The name of this superfamily derives from the first characterized human BI-1 gene (HsBI-1/TMBIM6), which inhibits BAX-induced PCD (Xu y Reed, 1998). This HsBI-1 protein is 237 amino acids (aa) long, contains a domain composed of 6–7 transmembrane regions, and locates in the endoplasmic reticulum (ER) (Xu y Reed, 1998). In mammals, this protein participates in the regulation of cytosolic calcium concentrations, protection against ER stress, and cancer development, among other functions (Henke *et al.*, 2011; Robinson *et al.*, 2011). Other five TMBIM proteins (some with similar anti-PCD functions) are coded in the human genome: Responsive to Centrifugal Force and Shear Stress 1 (RECS1)/TMBIM1, Lifeguard (LFG)/TMBIM2, Glutamate Receptor Ionotropic N-Methyl-D-Aspartate Associated (GRINA)/TMBIM3, Golgi Anti-Apoptotic Protein (GAAP)/TMBIM4, and Growth-Hormone Inducible Transmembrane (GHITM)/TMBIM5 (Rojas-Rivera y Hetz, 2014). In contrast to the BCL-2 family, the TMBIM superfamily is widespread across both evolutionarily-distant and closely-related organisms and is highly conserved in function and structure (Rojas-Rivera y Hetz, 2014; Henke *et al.*, 2011). Similar to HsBI-1, BXI1p protein from budding yeast provides protection against the heterologous expression of BAX and other stressful stimuli such as ER-stress, heat shock, and ethanol and glucose-induced PCD (Cebulski *et al.*, 2011; Chae *et al.*, 2003). In prokaryotes, BsYetJ protein from the bacteria *Bacillus subtilis* works as a pH-regulated calcium channel (Chang *et al.*, 2014). In plants, BI-1 homologues are key regulators of PCD, cytosolic calcium concentrations (Ihara-Ohori *et al.*, 2007),

sphingolipid metabolism (Nagano *et al.*, 2012), autophagy (Xu *et al.*, 2017), and Methyl Jasmonate (MeJA)-induced senescence (Yue *et al.*, 2012). Furthermore, plant BI-1 proteins determine the outcome of plant-pathogen interactions with biotrophic and necrotrophic fungi (Babaeizad *et al.*, 2009); and provide protection against several types of abiotic stresses such as heat, drought, oxidative stress, and salt stress (Duan *et al.*, 2010; Ishikawa *et al.*, 2010; Isbat *et al.*, 2009; Watanabe y Lam, 2006; Kawai-Yamada *et al.*, 2004). In eukaryotes, the TMBIM superfamily is further divided into the BI and Lifeguard (LFG) families (Henke *et al.*, 2011; Hu *et al.*, 2009). The BI family is composed of homologues of the GHITM/TMBIM5 and BI-1/TMBIM6 proteins. TMBIM5 orthologues are present in the choanoflagellate *Monosiga brevicollis*, the hemichordate *Saccoglossus kowleskii*, the nematode *Caenorhabditis elegans*, and vertebrates (Henke *et al.*, 2011). No TMBIM5 orthologues are present in plants. TMBIM6 orthologues are present in protists, algae, animals, and plants. Moreover, the LFG family is composed of homologues of the human TMBIM1-4 proteins plus the Tmbim1b protein from cow (*Bos taurus*) (Zhou *et al.*, 2008). These proteins have been renamed as LFG1-5 (Hu *et al.*, 2009). The LFG family expanded from a single LFG4-like ancestor before the divergence of major eukaryotic groups 2,000 million years ago (Ma) (Hu *et al.*, 2009). In animals, this ancestor was duplicated and gave rise to LFG4 (GAAP/TMBIM4) and the precursor of LFG1 (GRINA/TMBIM3). Subsequently, this LFG1 precursor underwent additional duplications leading to the vertebrate proteins LFG2 (LFG/TMBIM2) and LFG3 (RECS1/TMBIM1). LFG5 (Tmbim1b) protein possibly derived from an LFG2- or LFG3-like precursor (Hu *et al.*, 2009). In plants, LFG proteins have been reported in bryophytes, gymnosperms, and a few angiosperms (Weis *et al.*, 2013; Hu *et al.*, 2009). Apparently, LFG proteins in plants have undergone several rounds of duplications similarly as seen in animals. Gene and genome duplications have been major players in the acquisition of novel traits during plant evolution (Flagel y Wendel, 2009). Gene duplication occurs by means of small duplication events (tandem, segmental, transposon-mediated duplications), and whole genome duplications (WGDs) and triplications (WGTs) (Panchy *et al.*, 2016). The latter two, also known as polyploidizations, are drastic events that lead to the abrupt increment of both genome size and gene content followed by gene loss (fractionation) (Fawcett *et al.*, 2013), and are considered a common mode of speciation (Van de Peer *et al.*, 2017). Several ancient polyploidization events (paleopolyploidy) occurred during plant evolution. A WGD (ζ) occurred in the common ancestor of seed plants 319 Ma, and another (ϵ WGD) occurred in the common ancestor of angiosperms 192 Ma (Jiao *et al.*,

2011). A WGT (At- γ) occurred in the common ancestor of most eudicots, and two additional WGD (At- α and At- β) occurred in the common ancestor of Brassicaceae (Jaillon, 2007). In monocots, several rounds of WGDs (τ , σ , ρ) occurred in the common ancestor of grass species, such as rice (*Oryza sativa*) and wheat (*Triticum aestivum*) (Jiao *et al.*, 2014; Tang *et al.*, 2010). Other WGDs have been detected in other economically and ecologically important plant species. The TMBIM superfamily in plants was also expected to take place according to these different duplications events. Despite efforts to understand the molecular and biological functions of the TMBIM proteins in plants, little attention has been paid to their evolution considering paleopolyploidy events. As more sequenced genomes become available, comparative genomics allows us to get a deeper understanding about the evolution and duplication of the TMBIM superfamily in plants and other organisms. In the present study, we provide a phylogenetic overview and general motif analysis of the TMBIM superfamily of proteins across a wide range of genomes from Archaea, Bacteria, and Eukarya. Then, to get new insights about the evolution of the TMBIM superfamily in plants, we deepened in the individual analysis of the BI and LFG families in 48 plants through the integration of the available WGD/T information, phylogenetic analysis, detection of orthologous groups, and synteny network analysis to propose distinct models of evolution. Additionally, we briefly discuss new findings regarding the classification of the TMBIM proteins in prokaryotes, fungi, and animals. Our main goal was to determine the evolutionary history of the TMBIM superfamily in plants, information needed for future functional studies.

2.2 MATERIALS AND METHODS

2.2.1 HMM search and retrieval of TMBIM protein sequences

We conducted BLAST and PSI-BLAST (Altschul *et al.*, 1990, 1997) searches against the non-redundant protein database of the National Center of Biotechnological Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>) using as query the protein sequences of TMBIM1-6 (human), AtBI-1 and AtLFG1-5 (Arabidopsis), BIX1p (budding yeast), EYccA (*Escherichia coli*), and BsYetJ (*B. subtilis*). Then, we used the BLAST-Hits results of these searches (954 sequences) to construct a custom TMBIMHMM profile and searched on the proteomes of 256 selected species from Archaea, Bacteria, and Eukarya (Table S1). HMM profile construction (hmmbuild) and searches (hmmsearch) were done with HMMER 3.1b2 (<http://hmmer.org/>; (Eddy, 1998)), and identified sequences were retrieved with Seqret from EMBOSS 3.0 (Rice *et al.*, 2000).

2.2.2 Phylogenetic analysis pipeline

Since no large-scale analysis on the evolution of the TMBIM superfamily had been performed, we first conducted a phylogenetic overview across Archaea, Bacteria, and Eukarya (Figure S1a). The TMBIM sequences (see Section 3.2.1) were aligned in blocks using the [-profile] option of MUSCLE v3.8.31 (Edgar, 2004). The resulting Multiple Sequence Alignments (MSA) were edited in UGENE v1.25.0 (Okonechnikov *et al.*, 2012) as follows: N- and C-terminal regions were trimmed, and positions with more than 20% of gaps were removed, leaving only the TMBIM domain. Evolutionary Model testing was performed in ProtTest v3.4 (Darriba *et al.*, 2011). All phylogenetic trees were based on the LG substitution model (the best that fitted our data) and were inferred through the Maximum Likelihood method in RaxML 8.2.9 (Stamatakis, 2014). Whenever indicated in figure captions, the number of rapid bootstrap replicates was determined by the bootstopping criterion using the [-autoMRE] option (Pattengale *et al.*, 2009). Since we analyzed numerous sequences from evolutionary distant organisms we had to remove several positions, which may contain evolutionary relevant information, from the MSA in order to diminish the amount of gaps. Hence, we conducted specific phylogenetic analyses of the BI and LFG families in plants (Figure S1e). In addition, we also performed specific phylogenetic analyses of the

TMBIM proteins of prokaryotes, fungi, and animals in order to discuss particular findings about their classification (Figure S1b–d). Phylogenetic trees were visualized with FigTree 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) and iTOL v3.4 (Letunic y Bork, 2016).

2.2.3 Motif discovery and HMMLLogo analyses

As part of our overview of the TMBIM superfamily (Figure S1a), the complete set of retrieved TMBIM sequences (Section 3.2.1) from all organisms (Prokaryotes and Eukaryotes) was subjected to a motif discovery analysis using a stand-alone version of MEME v4.10.2 (Bailey *et al.*, 1994, 2009). Parameters were set to find 60 motifs with a minimum and maximum width of 10 to 25 amino acids, and minimum E-value threshold with no limit. Motif architecture of every sequence is depicted in the phylogenetic tree of Figure S2. For easier interpretation, the motifs were clustered based on their frequency among the main phylogenetic groups of TMBIM sequences. To do so, we divided the TMBIM sequences in 13 groups according to our phylogenetic analyses: Archaea BI, Bacteria BI, Fungi GHITM, Metazoa GHITM, Fungi BI, Metazoa BI, Plant BI, Fungi LFG, Metazoa LFG4, Metazoa LFG1-like, Plant LFG I, Plant LFG IIA, and Plant LFG IIB. The number of sequences in each group that contained any of the 60 motifs were counted, and a percentage matrix was generated. Based on this matrix, the 60 motifs were hierarchically clustered using the euclidean distance (`dist()`) and complete clustering (`hclust()`) methods, and visualized through ComplexHeatmap v3.5 (Gu *et al.*, 2016) in R v3.4.1 (R Development Core Team y R Core Team, 2017). The 60 motifs are briefly described in Text S1 and Figure S6. HMMLogos were generated by means of Skyline using default settings (<http://skylign.org/>; (Wheeler *et al.*, 2014)).

2.2.4 Structural multiple sequence alignment

To highlight the most conserved positions, regions, and motifs of the entire TMBIM superfamily, a Structural Multiple Sequence Alignment (SMSA) of representative TMBIM sequences was performed with PROMALS3D (Pei *et al.*, 2008) and colored with BOXSHADE 3.21 (Hofmann y Baron, 1996); protein coordinates of the 3D model of BsYetJ were included (Fig S1a).

2.2.5 Determination of orthologous groups (OGs) in plants

We detected TMBIM OGs in 48 plants with ProteinOrtho v5.15 (Lechner *et al.*, 2011) using default options. ProteinOrtho implements an extended version of the reciprocal best alignment heuristic (based on BLAST searches) approach to detect co-orthologous proteins/genes among multiple species, and employs spectral partitioning to cluster them in OGs (Lechner *et al.*, 2011). Methods based on constructing OGs from heuristic pairwise comparisons may be too inclusive and create mixed groups that not accurately represent evolutionary relationships (Kristensen *et al.*, 2011); hence, we mapped the detected (co-)orthologous TMBIM proteins of each OG to specific phylogenetic trees of the plant BI and LFG families. We only retained those OGs whose sequences were contained within single clades (monophyletic) in the phylogenetic trees (Figure S1f–g). On the other hand, the OGs were also mapped to the species tree in Figure 2.2 to easily point shared species and infer the possible moments of origin of each OG during plant evolution.

2.2.6 Tandem gene analysis in plants

To detect tandem duplicates of the TMBIM sequences in the 48 analyzed plants, each genome was compared against itself in SynMap (Lyons *et al.*, 2008) in CoGe (Lyons y Freeling, 2008). SynMap implements the algorithm LAST (Kielbasa *et al.*, 2011) to find homologous genes or regions between two genomes, and identify syntenic pairs by finding collinear series of putative homologous sequences with DAGChainer (Haas *et al.*, 2004). SynMap also implements the blast2raw algorithm to detect tandem duplicates. This analysis was performed with the online tool of SynMap (<https://genomevolution.org/coge/SynMap.pl>) with DAGChainer option settings as follows: relative gene order, maximum distance (-D) 30, minimum number of aligned pairs (-A) 5, tandem duplication distance of 15, and C-score 0.1 (Figure S1h).

2.2.7 Synteny network approach in plants

We performed a synteny analysis of the TMBIM sequences of 47 plant species; tobacco (*Nicotiana tabacum*) was not included because of its fragmented genome assembly. Parallel coordinated plots and pairwise dot plots are commonly used to analyze synteny between two or among a few genomes; however, as the number of genes and genomes to be analyzed increases, such approaches become impractical. Hence, we opted to follow the recently developed Synteny Network approach described by (Zhao y Schranz, 2017) and (Zhao *et al.*, 2017) that allows the simultaneous analysis and easy visualization of syntenic blocks among a greater number of genes and genomes. In this approach all synteny relationships are depicted as a network, where “nodes” represent sequences and “edges” or “connections” represent pairwise synteny relationships. This syntenic network approach can lead to the detection of transposed genes to another genomic region (separate subnetworks) to discover specific evolutionary patterns on gene families. Briefly, we performed a reciprocal all-against-all pairwise comparison between the entire proteomes of 47 plants by means of the protein similarity search tool RAPSearch2 (Zhao *et al.*, 2012). Similarity results and gene position information (GFF files) were used to calculate synteny blocks between all pairwise compared genomes with MCScanX (Wang *et al.*, 2012) using default parameters, creating a score matrix of all synteny relationships among all gene families (i.e. a network). Synteny information of the TMBIM superfamily were retrieved from the matrix, and used to construct dense connected network clusters or communities (k-clique=4) by using the Clique percolation method implemented in CFinder (Derényi *et al.*, 2005; Palla *et al.*, 2005). The resulting network communities were visualized with Cytoscapev3.5.1 (Shannon *et al.*, 2003) and Gephi v0.9.1 (Bastian *et al.*, 2009). In addition, to know the relationship between the detected synteny networks and the phylogeny of the TMBIM proteins of plants, we performed a phylogenetic profiling of the network communities by mapping them to a phylogenetic tree with all of the TMBIM sequences detected in the 48 plant species (including those missing from the network). For this specific analysis, a phylogenetic tree was built from the alignment of all of the TMBIM sequences from plants made with hmalign [-trim] in HMMER 3.1b2 (Eddy, 1998) and computed in RaxML8.2.9 (Stamatakis, 2014) by using the LG substitution model. For detailed information about the synteny network approach please see (Zhao y Schranz, 2017) and (Zhao *et al.*, 2017) (Figure S1h).

2.2.8 Generation of species trees

Species trees were generated with PhyloT (Letunic, 2015) and visualized with iTOL v3.4 (Letunic y Bork, 2016). Information about WGD events was taken from CoGepedia (https://genomeevolution.org/wiki/index.php/Plant_paleopolyploidy, last accessed on 1/12/2017) and (Renny-Byfield y Wendel, 2014). Information about ploidy level, genome size, and chromosome number was obtained from the Plant DNA C-values database (<http://data.kew.org/cvalues/CvalServlet?querytype=1>) and available published genomes (Table S1).

2.3 RESULTS

2.3.1 Phylogeny overview of the TMBIM superfamily across the three domains of life

We identified a total of 685 TMBIM proteins in 171 species distributed among archaea, bacteria, protists, fungi, animals, algae, and plants (Tables S1 and S2). Prokaryotic TMBIM proteins were divided into the Archaea BI and Bacteria BI clades (Figure 2.1a and b). Archaea BI clade contained proteins from six methanogenic archaeal species within Euryarchaeota, and seven bacterial species distributed among Acidobacteria, Spirochaetes, Chlamydie, Planctomycetes, γ proteobacteria, and Formicutes (Clostridia) (Figure S3a and b). Bacteria BI clade contained proteins from 33 bacterial species distributed in Fusobacteria, Spirochaetes, Bacteroidetes, Chlamydie, Planctomycetes, Proteobacteria (α , β , γ , δ , and ϵ), and Formicutes (Bacilli); and the sequence from the protist *Paulinella chromatophora* (Rhizaria) (Figure S3a and b). This *P. chromatophora* sequence (Pchromatophora1) clustered together with the TMBIM proteins from the two cyanobacteria *Prochlorococcus marinus* (Promarinus1) and *Trichodesmium erythraeum* (Terythraeum1) (Figure S3b). *P. chromatophora* contains chromatophores, photosynthetic entities of α -cyanobacterial origin (Marin *et al.*, 2007). Hence, it is possible that the TMBIM sequence from *P. chromatophora* was acquired by Horizontal Gene Transfer (HGT) from a cyanobacterial symbiont. Furthermore, the phylogenetic placement of prokaryotic TMBIM proteins was rather scattered among different phyla also suggesting HGT, a widely recognized mechanism in prokaryotes (Soucy *et al.*, 2015). YccA (Ecoli1) and BsYetJ (Bsubtilis1) proteins from *E. coli* and *B. subtilis*, respectively, were also included in the Bacteria BI clade (Figure S3b). YccA inhibits the activity of the Filamentous Temperature-Sensitive (FtsH) protease in *E. coli* and acts as an inhibitor of PCD in budding yeast (van Stelten *et al.*, 2009); BsYetJ is a pH-regulated calcium channel in *B. subtilis* (Chang *et al.*, 2014). On the other hand, Eukaryotic TMBIM proteins were unmistakably divided into the BI and LFG families, which is in agreement with literature (Hu *et al.*, 2009) (Figure 2.1a and b). The eukaryotic BI family was further divided into the GHITM and BI major clades (Figure 2.1a and b). The GHITM major clade included proteins from animals (Metazoa GHITM), fungal species from Dykaria (Fungi GHITM) (Figures. 2.1b and S2), and species from the protist lineages Cryptophyta, Rhizaria, Amoebozoa, and Choanomonada (Ophisthokonta) (Table S3). The BI major clade included proteins from all analyzed protist lineages (Table S3), animals (Metazoa BI), earlybranching fungi (Fungi BI), red and green algae, and land

plants (Plant BI) (Figures. 2.1b and S2). The eukaryotic LFG family was mainly divided in a lineage specific manner (Figure 2.1b). This family contained proteins from all the analyzed protist lineages (except for Excavata) (Table S3), as well as fungi (Fungi LFG), animals (Metazoa LFG4 and Metazoa LFG1-like), algae, and plants (Plant LFGs) (Figures 2.1b and S2).

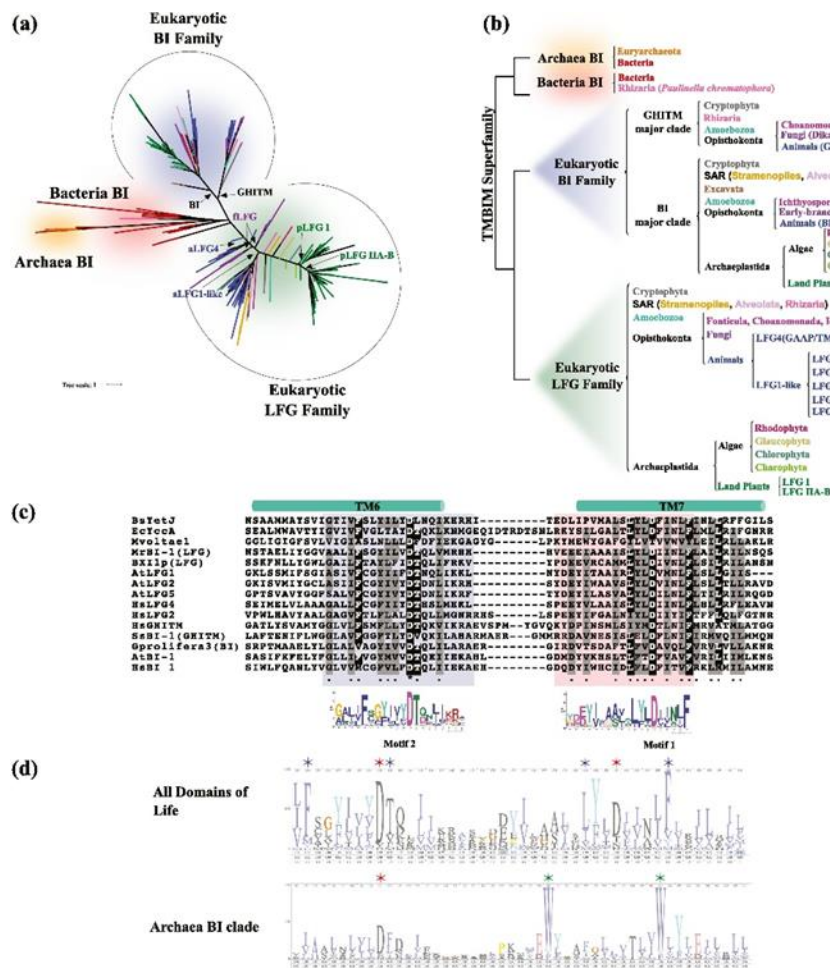


Figure 2.1. Classification and conserved sequence regions of the TMBIM superfamily of proteins. (a) Maximum likelihood unrooted tree inferred from the aligned domains of 685 TMBIM proteins from 171 species. Branch colors represent different groups of organisms as indicated in (b). Branches leading to BI and GHITM major clades, and branches leading to animal (aLFG4 and aLFG1-like), fungi (fLFG) and plant (pLFG I and Plant IIA-B) LFG groups are indicated with arrows. The complete tree is shown in Figure S2. (b) Diagram depicting the proposed general classification of the TMBIM proteins on the studied groups of organisms. (c) Multiple sequence alignment of the

TM6 and TM7 transmembrane regions of representative TMBIM proteins. TM6 and TM7 of the BsYetJ protein from *Bacillus subtilis* are indicated with bars. Motifs M1 and M2 are blue and red shaded, respectively, and their sequence LOGOs are shown below each one. Full alignment is shown in Figure S4. (d) HMMLogo of the C terminal region of the aligned domains of the total 685 identified TMBIM proteins, and the Archaea BI group. Full HMMLogos are depicted in Figure S7. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

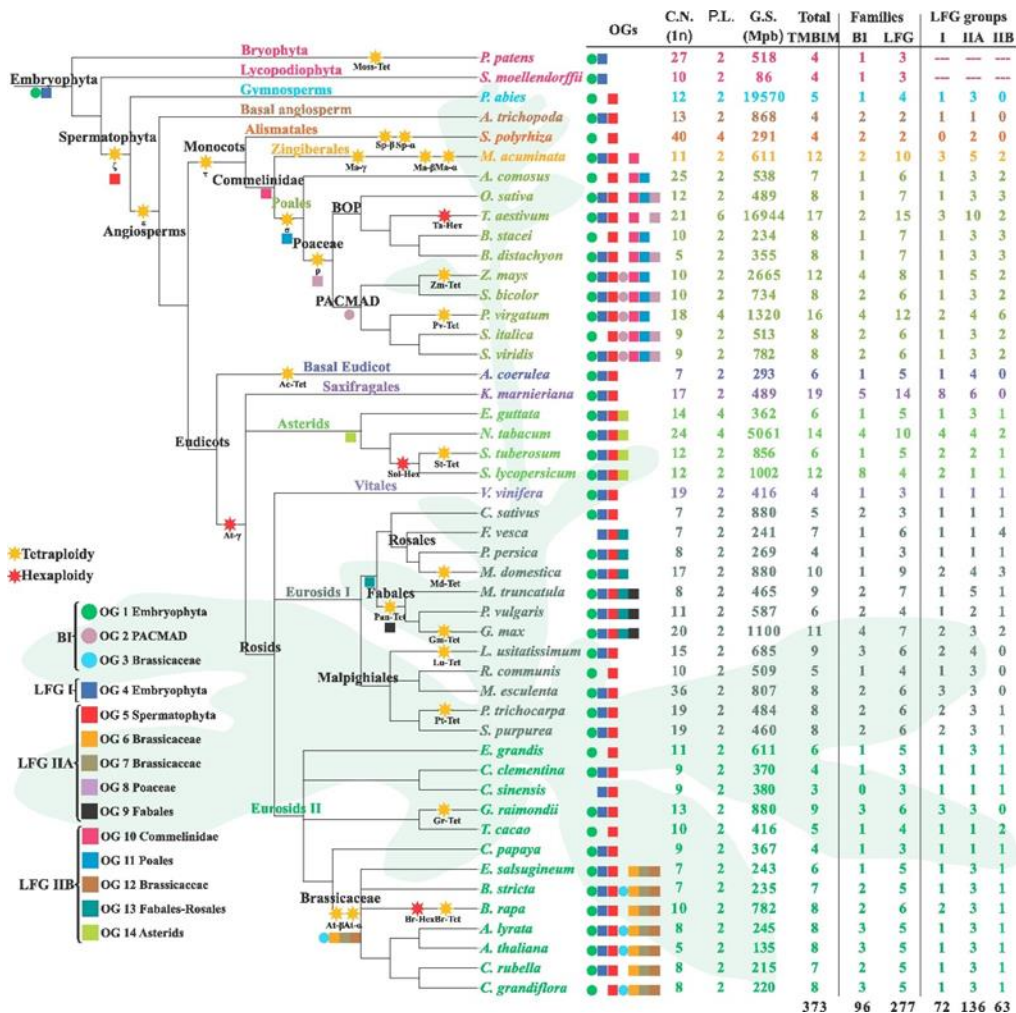


Figure 2.2. Phylogenetic relationships of the 48 analyzed plant species. Phylogenetic relationships among plant species were based on the NCBI taxonomy by using PhyloT program (Letunic, 2015) and modified according to the APG taxonomy (<http://www.mobot.org/MOBOT/research/APweb/>). Suggested positions of orthologous groups (OGs; colored circles and squares) and occurrence of ancient whole genome duplications (WGDs; yellow stars) and triplications (WGT; red stars) are indicated in the branches of the tree. Names of WGDs and WGTs are indicated. On the table: OGs=Presence/absence of OGs in each species, C.N.=1n chromosome number, P.L.=ploidy level, G.S.=1C genome size in megabase pairs (Mbp). Total number of identified TMBIM proteins per species, family, and groups are also indicated in the table. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3.2 Sequence and motif conservation indicate a pH-regulated calcium channel activity in TMBIM proteins in eukaryotes and bacteria, but not in Archaea

Transmembrane proteins evolve to a faster rate than water-soluble proteins due to their exposure to stronger adaptive forces in the boundaries of organelles and cells, but their inner transmembrane regions evolve at a slower rate than their aqueous portions (Sojo *et al.*, 2016). Accordingly, our SMSA indicates little overall sequence conservation of the TMBIM superfamily of proteins (as a whole) except for some highly conserved residues within the transmembrane regions, particularly TM4, TM6, and TM7, which contained M6, M2 and M1 motifs, respectively (Figure S4). Motifs M1 and M2 were the most abundant in the whole family being present in both the BI and LFG families, and in the Bacteria BI clade, but not in Archaea BI clade (Figures. S5, S6, Table S4, and Text S1). TM6 (M2) contains a Phenylalanine (F), an Aspartate (D), and a Threonine (T) highly conserved residues; and TM7 (M1) contains a Leucine (L), an D, and a F highly conserved residues (Figure 2.1c and d). These M1 and M2 motifs together are similar to the motifs 7, 8, and 10 reported by (Hu *et al.*, 2009) in LFG proteins of animals and plants. However, we found that M1 and M2 motifs were also present in LFG proteins from yeast, several protists, and bacteria (Figures. S2 and S6). Moreover, the conserved D residues in Motifs M1 and M2 (together) conform a Di-aspartyl pH sensor that regulates the calcium channel properties of the BsYetJ protein in *B. subtilis* (Chang *et al.*, 2014). Our data highlights the evolutionary conservation of this region in most proteins of the TMBIM superfamily suggesting a possible conserved function of the pH-regulated calcium channel in Bacteria and Eukarya. On the other hand, Archaea BI clade presents a distinct sequence composition in TM6 and TM7 containing motifs M51 and M37 instead of M2 and M1, respectively (Figure S2). Motif 37 (in TM7) in the Archaea BI clade contains two well conserved Tryptophan (W) residues (instead of an D), suggesting a distinct function or mechanism of the TMBIM proteins in this Archaea BI clade (Figures. 2.1d and S7); further sampling and functional analyses of these archaeal proteins are necessary.

2.3.3 Two major phylogenetic groups are found in the LFG family in plants

We analyzed the proteomes of 57 organisms from Archaeplastida covering 9 algae and 48 plants. We found a total of 19 TMBIM proteins in seven algal species (Table S3). No TMBIM proteins were found in the Chlorophytes *Micromonas pusilla* and *Ostreococcus lucimarinus* possibly due to independent gene losses. Algal TMBIM proteins were present in both the BI (only within the BI major clade) and LFG families (Figure 2.1b and Table S3). Interestingly, and like the results of (Hu *et al.*, 2009), LFG proteins from Chlorophytes were more closely related to the Metazoa LFG1-like clade. The exception was *C. subellipsoidea* that possessed LFG proteins close to the Metazoa LFG1-like clade and LFG clade from plants (Figure S2 and Table S3). This is not surprising since many homologous sequences of Chlamydomonas were present in the common ancestor of both animals and plants, and some genes shared by Chlamydomonas and animals have been lost in plants (Merchant *et al.*, 2007). In plants we found a total of 373 TMBIM proteins in 48 species; 96 and 277 sequences belonged to the BI and LFG families, respectively (Figure 2.2). Regarding the BI family, plants possess an average of two sequences per species, with a minimum of one in several species, and a maximum of eight in tomato (*Solanum lycopersicum*) (Figure 2.2). This plant BI family formed a single monophyletic group that was divided in a lineage-specific manner, mainly between monocots and eudicots (Figures. 2.3a and S8). However, seven sequences from tomato (*Solanum lycopersicum*) belonged to a specific long-branching clade (Slyopersicum- only clade) formed out of the Slycopersucum7 sequence and the tandem duplicated array Slycopersicum6, 8–12 (Figures. 2.3a and S8). The single remaining BI sequence from tomato (Slycopersicum4) was clustered inside an asterid specific clade next to tobacco, potato (*Solanum tuberosum*), and seep monkeyflower (*Erythranthe guttata*; synonym: *Mimulus guttatus*) (Figure S8). Regarding the LFG family, plants possess an average of 5.7 sequences per species (about 2.8-fold more than in plant BI family), with a minimum of two in Amborella (*Amborella trichopoda*) and common duckweed (*Spirodela polyrhiza*), and a maximum of 19 sequences in *Kalanchoe marnieriana* (Figure 2.2). In monocots (Commelinidae), the number of LFG proteins seems to be related to particular paleopolyploidy events: banana (*Musa acuminata*), wheat, maize (*Zea mays*) and switchgrass (*P. virgatum*) contain more LFG proteins in their genomes, and they have also experienced particular polyploidy events apart of the τ , σ , and ρ WGDs shared by most monocots (Figure 2.2). Our phylogenetic analysis revealed that the LFG family in plants was

firstly separated into Bryophytes-Lycophytes (the moss *Physcomitrella patens* and the clubmoss *Selaginella moellendorffii*) and seed plants (gymnosperms and angiosperms) (Figure 2.3b). LFG proteins in seed plants were divided in two major groups supported by high bootstrap values: >60 for LFG I group, and >80 for LFG II group (Figure S9). This result is similar to that found by (Weis *et al.*, 2013) when they examined the LFG proteins from *Arabidopsis* and barley (*Hordeum vulgare*). LFG I group contained 72 sequences from 45 species (average of 1.6 sequences per species) (Figure 2.2); common duckweed was the only absent. LFG II group was subsequently divided in two subgroups: IIA and IIB (Figure 2.3b). LFG IIA subgroup contained 136 sequences from 46 species (average of 2.9 sequences per species) (Figure 2.2); and LFG IIB subgroup contained 63 sequences from 37 species (average number of 1.7 sequences per species) (Figure 2.2).

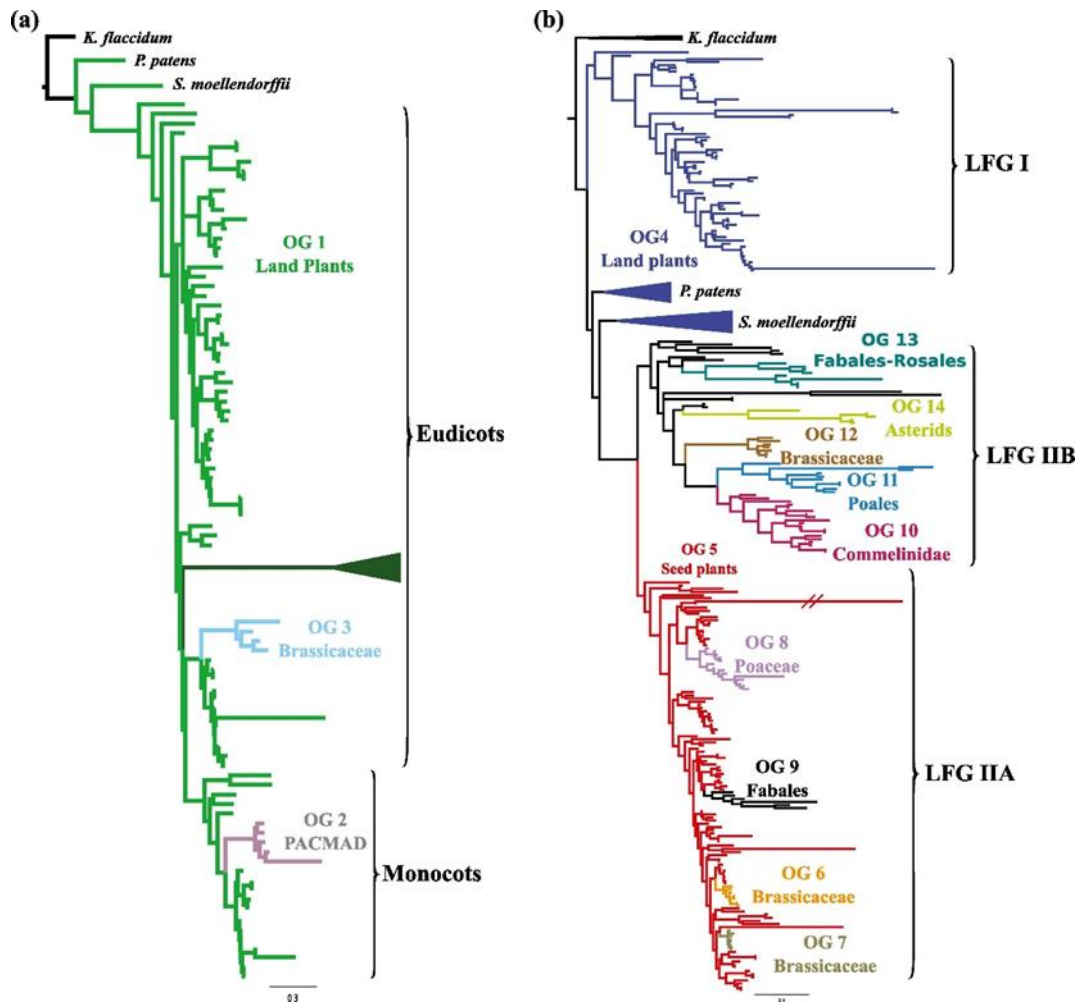


Figure 2.3. Phylogenetic relationships of BI and LFG families in plants. Maximum likelihood trees of 96 and 373 proteins of BI (a) and LFG (b) families, respectively, from 48 land plants. Sequences of the filamentous terrestrial alga *Klebsormidium flaccidum* were included as outgroups. In (a) the “*Slycopersicum*-only” branch of the BI family is shown in dark-green color. Name and number of OGs are indicated in the trees branches with different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3.4 Two OGs of the TMBIM superfamily are shared by most plants

Detection of OGs is interesting to comparative genomics and functional analysis, since orthologues commonly perform equivalent biological functions in different species. We found 14 OGs of TMBIM proteins in plants (Table S5) and these OGs were mapped to the plant species tree (Figure 2.2) and specific phylogenetic trees of the BI and LFG families in plants (Figures. 2.3, S8 and S9). Two of them (OGs 1 and 4) are shared by mosses, lycophytes, gymnosperms (OG1 only) and angiosperms; so, these OGs are probable shared by most land plants taxa (Embryophyte) (Figure 2.2). OG1 corresponded to the BI family (Figure 2.3a) and was shared by 46 species; only missing strawberry (*Fragaria vesca*) and sweet orange (*Citrus sinensis*) (Figure 2.2). OG4 corresponded to the LFG I group (Figure 2.3b) and was shared by 39 species; missing Norway spruce, common duckweed, pineapple, *Brachypodium stacei*, *Setaria italica*, castor bean (*Ricinus communis*), rose gum (*Eucalyptus grandis*), cacao (*Theobroma cacao*), and *Capsella grandiflora* (Figure 2.2). Interestingly, Arabidopsis AtB-1 and AtLFG5 from OGs 1 and 4, respectively, are highly expressed in most tissues (Figure S10). OG5, which corresponded to the LFG IIA subgroup (Figure 2.3b), was shared between the gymnosperm Norway spruce and all sampled angiosperms; thus this OG5 might have appeared due to the duplication of an ancestral protein of OG4 in the common ancestor of seed plants prior their diversification after ζ -WGD (Figure 2.2). Arabidopsis representative of OG5, AtLFG4, is highly expressed in the flower stamen and pollen (Figure S10) probably having some particular function in reproduction. Other two (2–3), four (6–9), and five (10–14) lineage-specific OGs are present in the BI family, the LFG IIA subgroup, and the LFG IIB subgroup, respectively (Figure 2.3a and b). OGs 3, 6, 7, and 12 were shared by Brassicaceae, each OG containing one sequence from Arabidopsis: AtBI-2, AtLFG3, AtLFG2, and AtLFG1, respectively (Figures. S8 and S9). Brassicaceae has experimented two additional WGDs (α and β) in addition to the WGT shared by all eudicots (γ) (Figure 2.2); so, one or both of these WGDs might have contributed to the appearance of the ancestor sequences of these lineage-specific OGs during Brassicaceae speciation. AtLFG2 (OG7, LFG IIA) presents higher expression in roots and seeds, while AtLFG3 (OG6, LFG IIA) is highly expressed in seeds only (Figure S10). On the contrary, AtBI-2 (OG3, BI) is expressed at low levels in all tissues (Figure S10). In monocots, OGs 10, 11, 8, and 2 are shared by Commelinidae, Poales, Poaceae, and PACMAD

grasses, respectively (Figure 2.2); σ and ρ WGDs might have contributed to the appearance of OGs 11 and 8 during Poales and Poaceae speciation, respectively (Figure 2.2). OG13 (LFG IIB; Figure 2.3a) might have appeared during the speciation of nitrogen-fixing fabids and was lost in Cucurbitales, but retained in Fabales and Rosales (Figure 2.2). Lastly, OG 14 was shared by Asterids (Figure 2.2). OGs are convenient when describing evolutionary relationships across species (Gabaldón *et al.*, 2013) since orthologous genes typically conserve similar biological functions in different organisms (Koonin, 2005). Ancestor genes of OG1 (BI) and OG4 (LFG I) were probably present since the last common ancestor of all plants, probably playing a conserved function in cell maintenance and survival, which might be the reason why they are highly expressed in all tissues of *Arabidopsis* (Figure S10). Duplication events gave rise to lineage-specific OGs, which may represent specific functions and expression patterns in Brassicaceae, legumes, asterids, and monocots. Further functional and expression analyses are required to confirm this.

2.3.5 Syntenic network and tandem analyses of the TMBIM superfamily in plants

The plant TMBIM synteny network contained 270 nodes (sequences) linked by 5,029 edges (pairwise synteny relationship) (Table S6). No sequences from the moss *P. patens*, the clubmoss *S. moellendorffii*, and the gymnosperm Norway spruce (*Picea abies*) were found to be connected to the network due to their phylogenetic distance from angiosperms and the large genome size of Norway spruce. We used clique percolation clustering at size $k=4$ (implemented in CFinder) to define denser and connected clusters or communities numbered from 1 to 5 (Figures 2.4a, S11a, and Table S7). These five communities were used for phylogenetic profiling (Figure 2.4b). Syntenic clusters 1, 3, and 4 were specific for the BI family, while clusters 2 and 5 corresponded to the LFG family (Figure 2.4b). No syntenic connections were found between BI and LFG families (Figure 2.4b), confirming that these families splitted apart before the evolutionary emergence of plants from a more ancestral organism, probably a prokaryote. Cluster 1 comprises 48 nodes connected by 634 edges from 29 species (Tables S7 and S8), mostly eudicots; 22.5% of these sequences (only eudicots) came from tandem duplication events (Figure 2.4b). Only one monocot sequence from common duckweed (*Spolyrhiza4*), which is one of the most ancient monocotyledonous clade (Alimastales), was part of this syntenic cluster and it was well

connected with eudicots (36 connections; Figure S11a). Nevertheless, Cluster 1 lacked syntenic sequences from grapevine (*Vitis vinifera*) and Amborella, but contained a sequence (Acoerulea6) from the basal eudicot columbine (*Aquilegia caerulea*). Regarding asterids, only one representative from seep monkeyflower, tomato, and potato were found in this BI syntenic cluster (Table S8). AtBI-1 and AtBI-2 sequences from Arabidopsis were clustered inside this group and were syntenic to each other (but not AtBI-3, a tandem of AtBI-1) (diamonds in Figure S11); both sequences belong to a syntenic block emerged from the At- α WGD (Y. Wang *et al.*, 2016). BI sequences from monocots were clustered apart in the lineage-specific syntenic clusters 3 and 4, showing evidences of ancient transposition events (Figures 2.4a, b and S11). Cluster 3 was formed by 10 nodes connected by 38 edges from eight species within Poales (Figure S11a), and Cluster 4 was formed by 4 nodes from three species within PACMAD group and pineapple (*Ananas comosus*), a Bromeliaceae (Figure S11a). No tandem duplicated sequences were found in clusters 3 and 4 (Figure 2.4b). The early-diverging monocot common duckweed and banana were absent in these clusters (Figure S11a and Table S8). Nevertheless, as mentioned above, a sequence from common duckweed (Alimastales) was syntenic to the rest of eudicot plants and grouped in BI Cluster 1 (Figure S11a). Since cluster 3 contained sequences from both PACMAD and BOP (OG1) but not the Bromeliaceae pineapple (Figure S11a and b), and cluster 4 contained sequences from Bromeliaceae (OG1) and the PACMAD (OG2) but not the BOP group of grasses (Figure S11a and b); we infer that the common ancestor of Poales (most recent common ancestor of Bromeliaceae and Poaceae) experienced a replicative transposition (Monocot transposition 2 in Figure S11a and b) and cluster 3 was only retained by PACMAD and BOP, and cluster 4 was only retained in Bromeliaceae and PACMAD. These ancient transpositions could be the reason why we found that BI sequences from monocots and eudicots were clustered apart in our phylogenetic analysis (Figure 2.3a). The LFG family was formed by two syntenic clusters: Cluster 2 contained all considered angiosperm species (44), and Cluster 5 contained the three sampled Fabales species (Tables S7 and S8). Cluster 2 was formed of 198 nodes connected by 4,342 edges (Table S7). This cluster was divided in two communities corresponding to the LFG I (cluster 2b) and LFG II (cluster 2a) gene groups (Figure 2.4a and b). Even when these two communities are poorly connected to each other, they shared syntenic relationship since both contain sequences from the pivotal species grapevine, columbine, and Amborella (Figure S11 and Table S8). LFG I community (cluster 2b; red inner stripe in the tree in Figure 2.4b) was formed by 57 nodes

connected by 771 edges from 39 species (Table S8); only 10.5% of its sequences came from tandem duplication events. This LFG I (cluster 2b) community contained sequences from both monocots and eudicots, including AtLFG5 from *Arabidopsis* (diamond in Figure S11), a sequence from *Amborella* (Atrichopoda1) and one from banana (clique k=4). The LFG II community (cluster 2b; orange inner stripe in the tree of Figure 2.4b) was formed by 141 nodes connected by 3,567 edges from all of the 44 analyzed angiosperm species (Table S8) suggesting that this community has retained its synteny since the last common ancestor of angiosperms. LFG II community contains up to 32.6% of tandem duplicate genes, most of them in monocots, suggesting that it has also expanded through small-scale duplication events (black dots at the end of branches in Figure 2.4b and see (Table S2). Only one sequence from *Amborella* was found inside this community (Atrichopoda2). AtLFG1-4 from *Arabidopsis* were grouped in this LFG II community, but only AtLFG3 and AtLFG4 were directly syntenic to each other (diamonds in Figure S11). These sequences came from a syntenic block that emerged from At- α WGD (Wang et al., 2016). AtLFG1 and AtLFG2 (but not AtLFG3 and AtLFG4) were syntenic to each of the two sequences from columbine and grapevine, suggesting that these genes derived from an ancient WGD event (At- γ). AtLFG2 is one of the few sequences that links the LFG II community to the LFG I community, and according to (Wang et al., 2016) this sequence is a relocated γ gene, i.e. a gene that relocated or transposed to another genomic context after the At- γ WGT. The lineage-specific syntenic cluster 5 was formed by 4 nodes connected by 6 edges from three Fabales species belonging to OG13 (Figure S11b). These data indicate a probable transposition of some OG13 members of the Fabales ancestor to another genomic context after the legume-specific WGD, the Panilionoid- Tetraploidy (Pan-Tet).

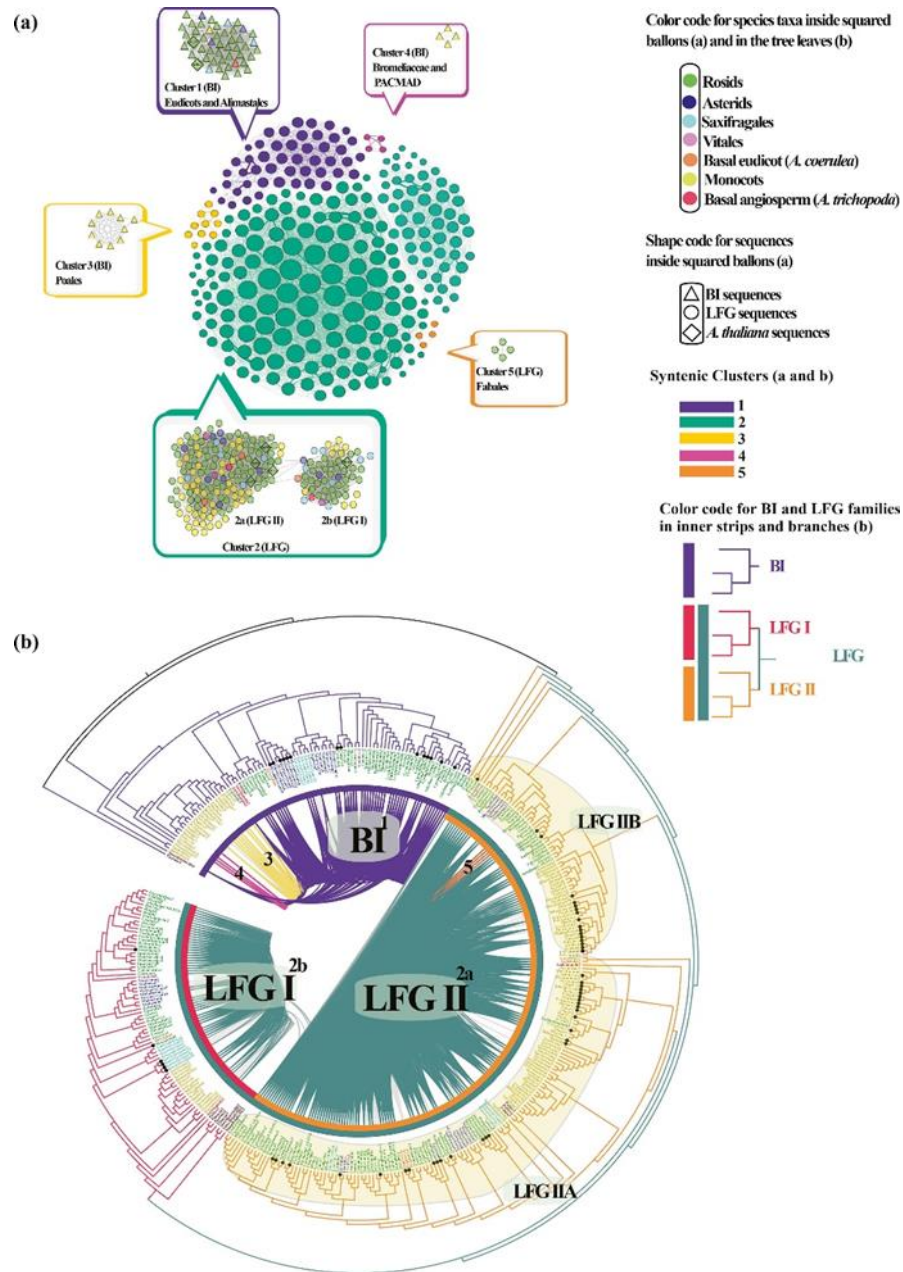


Figure 2.4. Synteny Network Analysis of the TMBIM superfamily in plants. (a) Complete synteny network of the TMBIM gene superfamily based on clique percolation method at

k=4. Node size indicates the number of connections with other nodes. The complete cluster network is shown next to each community inside squared balloons. (b) Phylogenetic reconstruction of the TMBIM gene superfamily of land plants and syntenic relationship of the five clusters found in the network. The BI family is marked with purple inner stripe and also purple branches. The LFG family is marked with dark turquoise inner stripe and then split into LFG I group (red inner stripe and branches) and the LFG II group (orange inner stripe and branches). Branches for subgroups in LFG II (LFG IIA and LFG IIB) are enclosed on the tree with light-orange background color. Tandem duplicated genes are marked with black dots at the end of the branches. Clique size k=4 was used to depict the five communities. Clique connections below 4 ($k < 4$) are shown in gray colored lines. Communities are coded from 1 to 5 for easy tracking (Table S7). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.4 DISCUSSION

2.4.1 Eukaryotic BI and LFG families diverged early in prokaryotes

Previous studies on the evolution of the TMBIM superfamily covered a limited number of sequences and/or organisms, mainly animals (Henke *et al.*, 2011; Hu *et al.*, 2009). Hence, detailed knowledge about the evolution of this superfamily in other lineages like prokaryotes, protists, fungi, algae, and plants is still lacking. Here we examined the evolution of the TMBIM superfamily across a wide range of species from the three domains of life, with greater detail in plants. We report for the first time the presence of TMBIM proteins in Archaea and several bacterial species (Figure S3). Previous work suggested that the BI and LFG proteins make two different families in eukaryotes with a probable very distant ancestor (Hu *et al.*, 2009). Our results support this hypothesis since the BI and LFG families were splitted apart from each other and from the prokaryotic TMBIM proteins (Figure 2.1a). Furthermore, within the Bacteria BI clade we found a smaller cluster of four bacterial sequences comprising the proteins from *Gimesia maris*, *Lentisphaera araneosa*, *Capnocytophaga gingivalis* and *Riemerella anatipestifer* that seemed to be closer to the Eukaryotic LFG family (green square in Figure S2); these proteins might be closer representatives of the common ancestor of Eukaryotic LFG family and Prokaryotic TMBIM proteins. Sampling of more bacterial genomes and metagenomes are still necessary to clarify this. On the other hand, similarly to our results (Figure S3), a scarce and scattered occurrence pattern has also been reported for other PCD related proteins in bacteria and archaea. For instance, (Asplund-Samuelsson *et al.*, 2012) found that only 262 of 1463 prokaryotic genomes contained metacaspases (caspase homologs); and only five of those species belonged to Euryarchaeota. Some hypotheses have suggested that eukaryotic PCD molecules and pathways had their origins in prokaryotes as several homologs of eukaryotic PCD-related genes, such as Caspases, Apoptosis-Inducing Factor, and Cytochrome C have been found in archaeal and bacterial species (Taylor-Brown y Hurd, 2013; Koonin y Aravind, 2002). Our findings also add up to the hypothesis that the eukaryotic PCD core machinery comes from a prokaryotic origin.

2.4.2 New insights into the TMBIM superfamily in fungi, and a possible archetype BI sequence

We found no TMBIM proteins in Microsporidia and Neocallimastigomycota, similarly to the results of (Chen *et al.*, 2015). Microsporidia are obligated intracellular parasites that have experienced many gene losses and genome reductions during evolution (Corradi, 2015), possibly explaining the lack of TMBIM proteins in these fungi. Previous studies have identified a single BI-I like protein in fungi: BXIp in budding yeast (Cebulski *et al.*, 2011), MrBI-1 in *Metarhizium robertsii* (Chen *et al.*, 2015), and SsBI-1 in *Sclerotinia sclerotiorum* (Yu *et al.*, 2015). Our phylogenetic analysis indicates that both BI and LFG families are present in most fungi (Figure S12). Moreover, in the BI family, we found that early-diverging fungal species and Dykaria only retained homologues of the BI-1/TMBIM6 and GHITM/TMBIM5 proteins, respectively (Figure S12). Considering that GHITM/TMBIM5 homologues are absent in algae, plants, and many protists but present in animals (Figure 2.1b), and that BI-1/TMBIM6 was present in almost all eukaryotes, we infer that an archetype BI-1/TMBIM6 protein ancestor was present in the common ancestor of Eukaryotes. Then, a GHITM/ TMBIM5 homologue appeared as a duplicate of BI-1/TMBIM6 during Eukaryote evolution, which was retained in Dykaria and Animals but lost in early-diverging fungi, algae, plants, and many protists; early-diverging fungi retained this archetype BI-1/TMBIM6 protein. Therefore, MrBI-1 and BXI1p are LFG homologues, and SsBI-1 is a GHITM/ TMBIM5 homolog. These insights shed new light on the classification of the TMBIM proteins in Fungi, and will help guiding and interpreting future functional analyses.

2.4.3 Two ancient transposition events of the BI family occurred in monocots

The Plant BI family experienced few duplication events and remained as a single monophyletic group containing a single major OG(1) (Figures. 2.3a and S11b). In monocots, after the split of Alismatales from the rest of monocot lineages, the BI genes from OG1 transposed to another genomic context in the common ancestor of Commelinidae (Figure 2.5a). In duckweed this BI protein from OG1 seems to have been duplicated, and transposed to another genomic context (Figure 2.5a). This explains why a BI-1 sequence (not from OG1) from common duckweed still conserved the same synteny as eudicot plants (Figure S11a

and b). A subsequent replicative transposition occurred in the common ancestor of Poales during σ WGD (Figures. 2.5a and S11b). This transposition was lost in the common ancestor of the BOP clade but retained in the common ancestors of Bromeliaceae and PACMAD grasses, ultimately giving rise to the appearance of OG2 in the latter. This second transposition could be the reason why a sequence from pineapple remains syntenic to the sequences of the OG2 in the PACMAD group (cluster 4 in Figure S11b). Further sampling of more monocot genomes is necessary to clarify this. A duplication during the At- α/β WGD in eudicots (Figure 2.5a) possibly gave rise to OG 3 in Brassicaceae.

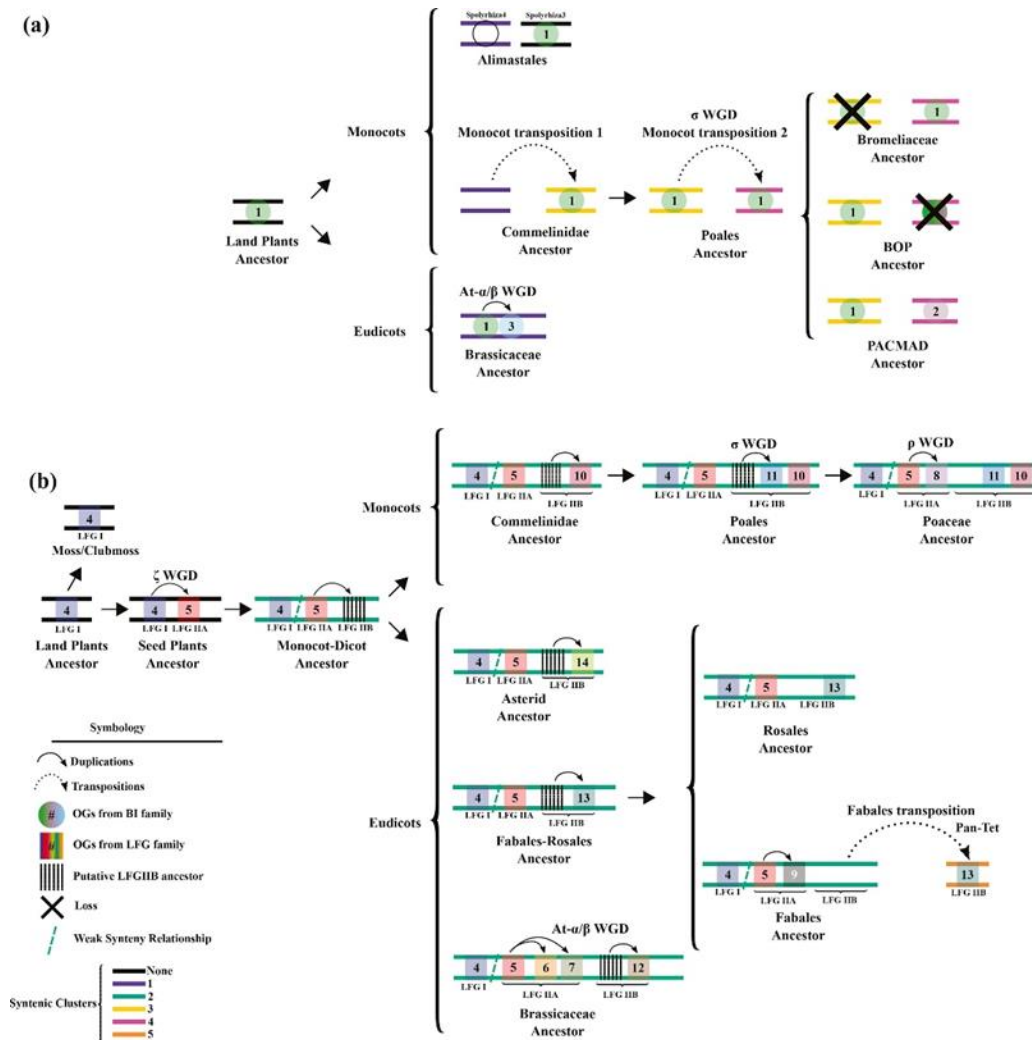


Figure 2.5. Proposed model of evolution and expansion of BI and LFG families in plants. (a) Evolution and expansion of the BI family in plants. A BI gene was present in the last common ancestor of land plants, which represents the ancestor of OG1 (green circle). In monocots, after the split of Alismatales, the BI genes transposed to another genomic context in the common ancestor of Commelinidae. In duckweed this BI protein from OG1 seems to have been duplicated, and transposed to another genomic context. Therefore, a BI-1 sequence (not from OG1) from common duckweed still conserves the same synteny as eudicot plants. Then, this transposed OG 1 (in monocots) experienced a subsequent replicative transposition to another genomic context in the ancestor Poales. This duplication was lost in the common ancestor of the BOP clade, but was retained in the common ancestor of Bromeliaceae and the common ancestor of the

PACMAD group of grasses, leading to the appearance of OG2 in the latter (purple circle). This second transposition could be the reason why a sequence from pineapple remains syntenic to the sequences within the PACMAD group. On the other hand, in Eudicots a duplication during the At- α/β WGD possibly gave rise to OG 3 in Brassicaceae (blue circle). (b) Evolution and expansion of the LFG family in plants. An LFG gene was present in the last common ancestor of land plants, which represents the ancestor of OG4 (blue square). This gene represents the ancestor of the LFG I group and the entire LFG family in land Plants. Possibly, during ζ WGD a duplication gave rise to the LFG IIA group in seed plants (OG 5, red square). A subsequent duplication event in the common ancestor of monocots and dicots gave rise to the putative ancestor gene of LFG IIB (stripped box). Subsequent duplication events of LFG IIA (OG5) and IIB, but not LFG I (OG4), gave rise to several lineage-specific OGs in angiosperms, some of which coincide with WGDs in Brassicaceae, Poales and Poaceae. OG13 (LFG IIB) in the ancestor of Fabales was transposed to another genomic context, possibly during the Panilionoid tetraploidy (Pan-Tet). Symbology: Parallel lines represent the plant genomes and are colored according to the syntenic clusters in Figure 2.4. Dashed “/” line represents the poor syntenic relationship between clusters 2a and 2b. Number of the OGs are indicated inside squares and circles. Solid and dotted arrows represent duplications and transpositions, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.4.4 Expansion of the LFG family in plants and vertebrates

An LFG I protein, ancestor of OG4, was present in the common ancestor of plants. During ζ WGD this protein was duplicated in the common ancestor of seed plants (but not mosses and clubmosses) giving rise to the ancestor of OG5 and the LFG IIA group (Figure 2.5b). LFG I and LFG II groups belong to two separated syntenic communities, therefore evolved under different genomic contexts; however, evidence sustains that both groups were syntenic in the common ancestor of monocots and dicots (Figure 2.4a). LFG I (OG4) group remained mostly unexpanded during angiosperm diversification, similar to the case of Metazoa LFG4 clade in animals (Figure S13). On the contrary, the LFG IIA group (OG5) expanded during angiosperm evolution through WGDs and tandem duplications giving rise to LFG IIB subgroup and several lineage-specific OGs (Figure 2.5b), similar to the expansion of Metazoa LFG1-like clade in animals. Metazoa LFG1-like clade experienced several gene duplications (Hu et al., 2009), giving place to 4 small subclades corresponding to LFG1, LFG2, LFG3, and LFG5 proteins in vertebrates (Figure S13). Even when WGDs are rare in animals (in contrast to plants) is worth noting their important influence on the expansion of gene families. Vertebrate genomes have experienced two rounds of ancient WGD (1R-WGD

and 2R-WGD) and a lineage-specific duplication event in teleost fishes (3RWGD) (Figure S13). These WGDs have shaped the diversity of organisms through preferential retention of duplicated genes (Nakatani *et al.*, 2007). For example, in the teleost zebrafish (*Danio rerio*), which experienced the 3R-WGD (Inoue *et al.*, 2015; Glasauer y Neuhauss, 2014), TMBIM genes appear to be preferentially expanded only in the LFG1-like clade. On the other hand, in plants, OGs 10–13 (within LFG IIB) probably originated from the duplication of a putative LFG IIB ancestor (which originated from OG5). The ancestor of OG13 in Fabales species (but not Rosales) was transposed to another genomic context, possibly during the Pan-Tet (Figure 2.5b). The possible functional consequences of this transposition needs to be addressed through experimental studies. OGs 6–9 (within LFG IIA) originated from duplications of OG5 proteins. In humans, the duplicate members of Metazoa LFG1-like clade diverged to perform both overlapping and specific functions (Carrara *et al.*, 2017; Rojas-Rivera y Hetz, 2014; Hu *et al.*, 2009); a similar scenario could take place in plants. Overexpression of AtLFG1-2 genes in Arabidopsis supports the infection by the biotrophic fungus *Erysiphe cruciferarum*; conversely, mutant plants of these genes are less susceptible. Similar phenotypes were found in epidermal cells of barley overexpressing or silencing the HvLFGa gene, respectively (Weis *et al.*, 2013). AtLFG2 (BIL4) mediates cell elongation under the Brassinosteroid signaling pathway (YAMAGAMI *et al.*, 2009). Similar to our results, (Weis *et al.*, 2013) found that LFG proteins from Arabidopsis and Barley were divided in two main groups: AtLFG5 and HvLFGe were more closely related to each other, and HvLFGa-d and AtLFG1-4 were clustered together in a different group. By expanding the sampling of analyzed plant genomes, we found that AtLFG1-3 (and its orthologues) are specific to Brassicaceae. However, as shown by (Weis *et al.*, 2013) these proteins present overlapping functions with barley. Other lineagespecific functions for AtLFG1-2 probably exist. According to our data, functional analysis of AtLFG4 and AtLFG5 would exemplify more representative functions of the LFG family in plants.

2.5 CONCLUSIONS

We identified TMBIM proteins in previously unexplored organisms from all the three domains of life. Our findings indicate that the Eukaryotic BI and LFG families emerged from independent distant relatives in bacteria. However, in order to elucidate the origin of each family in eukaryotes, further search across sequenced bacterial genomes and metagenomes is necessary. We also provide a wider scope of the phylogenetic relationships of these proteins and clarify their relationships among some taxa such as fungi. The major contribution of this work is that we provide a deep analysis about the evolutionarily history of the TMBIM superfamily in land plants, and that we detected OGs that arose across plant evolution due to specific WGD events. Brassicaceae and monocots, with four specific OG each, represent a good example of the crucial role of WGD on gene family expansions, creating new material for gene functional diversification. Further functionality and expression data from several species remain necessary to understand the TMBIM superfamily. Our results provide a benchmark to carry out research in this regard and comprehend the role of these proteins, not only in the PCD regulation but in other functions as well.

Appendix A. Supplementary material Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ympev.2018.04.032>.

CAPÍTULO III**TRANSCRIPTOMICS AND CO-EXPRESSION NETWORKS REVEAL TISSUE-SPECIFIC RESPONSES AND REGULATORY HUBS UNDER MILD AND SEVERE DROUGHT IN PAPAYA (*CARICA PAPAYA* L.)****Authors:**

Samuel David Gamboa-Tuz^a, Alejandro Pereira-Santana^b, Jesus Alejandro Zamora-Briseño^a, Enrique Castano^b, Francisco Espadas-Gil^a, Jorge Tonatíuh Ayala-Summano^{cd}, Miguel Angel Keb-Llanes^a, Felipe Sanchez-Teyer^a, Luis Carlos Rodriguez-Zapata^a

Affiliations:

^a Biotechnology Unit, Yucatan Center for Scientific Research (CICY), 97205, Merida, Yucatan, Mexico.

^b Plant Biochemistry and Molecular Biology Unit, Yucatan Center for Scientific Research (CICY), 97205, Merida, Yucatan, Mexico.

^c IDIX S.A. de C.V., Av. Sonterra 3035 int. 26, Querétaro, Mexico.

^d Polytechnic University of Huatusco, 94100, Veracruz, Mexico.

Journal:

Scientific Reports 8, 14539 (2018)

<https://doi.org/10.1038/s41598-018-32904-2>.

Status:

Received 02 July 2018; Accepted 18 September 2018; Published 28 September 2018.

ABSTRACT

Plants respond to drought stress through the ABA dependent and independent pathways, which in turn modulate transcriptional regulatory hubs. Here, we employed Illumina RNA-Seq to analyze a total of 18 cDNA libraries from leaves, sap, and roots of papaya plants under drought stress. Reference and de novo transcriptomic analyses identified 8,549 and 6,089 drought-responsive genes and unigenes, respectively. Core sets of 6 and 34 genes were simultaneously up- or down-regulated, respectively, in all stressed samples. Moreover, GO enrichment analysis revealed that under moderate drought stress, processes related to cell cycle and DNA repair were up-regulated in leaves and sap; while responses to abiotic stress, hormone signaling, sucrose metabolism, and suberin biosynthesis were up-regulated in roots. Under severe drought stress, biological processes related to abiotic stress, hormone signaling, and oxidation-reduction were up-regulated in all tissues. Moreover, similar biological processes were commonly down-regulated in all stressed samples. Furthermore, co-expression network analysis revealed three and eight transcriptionally regulated modules in leaves and roots, respectively. Seventeen stress-related TFs were identified, potentially serving as main regulatory hubs in leaves and roots. Our findings provide insight into the molecular responses of papaya plant to drought, which could contribute to the improvement of this important tropical crop.

RESUMEN

Las plantas responden al estrés por sequía a través de las vías dependiente e independiente de ABA, las cuales, a su vez, modulan ejes de regulación transcripcional. En el presente trabajo, se utilizó secuenciación de RNA (RNA-seq) por Illumina para analizar 18 bibliotecas de cDNA de hojas, savia y raíz de plantas de papaya bajo condiciones de estrés por sequía. Análisis transcriptómicos de referencia y de novo identificaron 8,549 genes y 6,089 unigenes, respectivamente, responsivos a sequía. Conjuntos de 6 y 34 genes fueron regulados hacia arriba y hacia abajo, respectivamente, en todas las muestras estresadas. Además, análisis de enriquecimiento de ontología génica (GO) revelaron que, bajo un estrés por sequía moderado, procesos relacionados al ciclo celular y reparación del ADN son regulados hacia arriba en hojas y savia, mientras que genes de respuesta a estrés abiótico, señalización hormonal, metabolismo de sacarosa y biosíntesis de suberina fueron regulados hacia arriba en raíces. Bajo condiciones severas de estrés por sequía, procesos biológicos relacionados a estrés abiótico, señalización hormonal y óxido-reducción fueron regulados hacia arriba en todos los tejidos. Procesos biológicos similares fueron regulados hacia abajo en todas las muestras estresadas. Además, análisis de redes de co-expresión revelaron tres y ocho módulos de regulación transcripcional en hojas y raíces, respectivamente. Diecisiete factores de transcripción relacionados al estrés fueron identificados, que potencialmente actúan como reguladores principales en hojas y raíces. Los resultados de este estudio proveen mayor conocimiento acerca de las respuestas moleculares de la planta de papaya a la sequía, lo cual podría contribuir al mejoramiento de este importante cultivo tropical.

3.1 INTRODUCTION

Drought threatens the productivity and survival of agricultural crops worldwide. To cope with drought stress, different plant tissues, such as leaves, roots, and sap, implement general and specific responses. Leaves are the main photosynthetic organs, and regulate water loss through transpiration. Upon drought stress photosynthesis decays (Chaves *et al.*, 2009), stomata are closed (Arve, 2011), and cuticle waxes are deposited (Xue *et al.*, 2017; Tom y D., 2006) in order to reduce excessive water loss. Roots regulate water and nutrient uptake, they are the first organs to detect water deficit, and additionally they transmit signals to the aerial parts (Shabala *et al.*, 2016; Schachtman y Goodger, 2008). Under drought stress, roots modify their architecture to increase water uptake (Comas *et al.*, 2013), and they increase the biosynthesis of suberin to regulate water traffic between plant and soil (Franke y Schreiber, 2007). Upon drought stress many protein-coding genes, some of which are involved in abiotic stress tolerance, are regulated by the Absciscic Acid (ABA)-dependent and the ABA-independent signaling pathways (Yoshida *et al.*, 2014). These genes can be classified into functional and regulatory genes (Singh y Laxmi, 2015). Functional genes include those that perform specific cellular functions such as: late embryogenesis abundant (LEA) proteins, heat shock proteins (HSP), reactive oxygen species (ROS), scavenging enzymes, osmoprotectant synthesizing enzymes, among many others. While regulatory genes control the expression and/or activity of other genes, and they include: transcription factors (TF), kinases, phosphatases, among others. The main characterized TF families regulating abiotic stress responses in plants include: AP2/ERF (APETALA2/ethylene-responsive element-binding factor), DREB (dehydration-responsive element-binding), bZIP (basic leucine zipper), AREB/ABF (ABA-responsive element-binding protein/ABA-binding factor), NAC (NAM, ATAF1/2, CUC), MYB (myeloblastosis oncogene), bHLH (basic helix-loop-helix proteins), and WRKY (Baldoni *et al.*, 2015; Singh y Laxmi, 2015; Nakashima *et al.*, 2014; Pan y Jiang, 2014; Kazuo y Kazuko, 2005).

Efficient communication of detected external signals between distal tissues is of great importance for coordinated plant development and to generate rapid responses against unfavorable conditions. The vascular system regulates the long-distance trafficking of several molecules (water, nutrients, photoassimilates, among others) between distal tissue (Spiegelman *et al.*, 2013). The phloem sap is responsible for the movement, distribution and trafficking of these and other macromolecules (Kehr y Buhtz, 2008), such as proteins and

RNAs. However, the molecular participation of the phloem in response to drought is less studied.

Due to the complex molecular responses of plants to drought and other types of stress, omics approaches have been implemented to unravel their intricate mechanisms (Yoshida *et al.*, 2015; Urano *et al.*, 2010). The development of Next Generation Sequencing (NGS) technologies, or High-Throughput Sequencing, has permitted the analysis of several plants transcriptomes through the sequencing of their RNA (RNA-Seq) species, e.g mRNA, microRNA, etc. (Unamba *et al.*, 2015; Martin *et al.*, 2013; Barrera-Figueroa *et al.*, 2012). This has generated substantial data sets of hundreds or thousands of regulated genes in response to drought. Bioinformatic tools have been developed to analyze the enormous quantity of information generated, such as enrichment analyses, to determine meaningful regulated biological processes (Huang *et al.*, 2009). Also, co-expression networks have been applied for the identification of putative regulatory transcriptional hubs in plants (Schaefer *et al.*, 2017; Serin *et al.*, 2016). These advances in technology and in silico analysis have permitted the elucidation of molecular mechanisms in model and non-model plants, including several economically important agricultural crops.

Papaya plant (*Carica papaya* L.) is a fruit crop grown world-wide in tropical and sub-tropical regions. Papaya fruit is a rich source of nutrients and papain, a digestive enzyme with several industrial applications. In 2016, Mexico was the third largest producer of papaya fruit, having produced about 951,922 metric tons (FAOSTAT, 2018). Additionally, Mexico has been the leading exporter worldwide for several years (FAOSTAT, 2018). Papaya plants have been considered relatively resistant to drought, responding through a desiccation postponement mechanism (Mahouachi *et al.*, 2006). Upon drought stress papaya plants accumulate proline²⁷ and ions, such as K⁺, Na⁺, and Cl⁻² probably contributing to osmotic adjustment. Levels of ABA and Jasmonic acid (JA) hormones have also been found to increase in response to drought in papaya (Mahouachi *et al.*, 2007, 2012). However, water scarcity may limit papaya physiological performance (Mahouachi *et al.*, 2006, 2007, 2012). Previous studies on papaya transcriptomes have focused on the analysis of root specific gene expression (Porter *et al.*, 2008), fruit ripening (Fabi *et al.*, 2010), sex determination (Urasaki *et al.*, 2012), cold-induced sex reversal (Lin *et al.*, 2016), expression changes in the papaya ringspot virus (PRSV)-resistant transgenic 'Sunup' (Fang *et al.*, 2016), somatic

embryogenesis (Jamaluddin *et al.*, 2017), and sticky disease responses (Madroño *et al.*, 2018). However, transcriptomic analyses for the elucidation of papaya plant molecular responses to drought remain scarce.

In the present study we employed Illumina RNA-Seq to analyze the transcriptome of leaves, sap, and roots of papaya plants under well-watered (control) condition, and after 10 and 20 days of drought stress. We identified tissue-specific sets of differentially expressed genes (DEGs) through reference and *de novo* assembly approaches. Functional enrichment analysis of these sets of DEGs revealed specific biological processes regulated among tissues under the control condition, and in response to drought. Furthermore, through co-expression network analysis, we identified several abiotic stress related TFs, which may act as putative regulatory hubs in leaves and roots under drought stress. Our findings provide a profound understanding of the molecular responses of papaya plant to drought stress, and provide critical baseline information for future genetic improvement and breeding programs of this important tropical fruit crop.

3.2 MATERIALS AND METHODS

3.2.1 Plant material and experimental design

Seeds of papaya (*Carica papaya* L.) 'Maradol roja', acquired from Semillas del Caribe® (Guadalajara, Mexico), were sown in separate substrate-filled pots. After germination, the papaya plants were grown under greenhouse conditions and equivalently watered. Drought stress was imposed on three-month-old plants by withholding watering; healthy well-watered plants of same age were used as controls. Visual assessment of phenotype, physiological measurements, and sample collection for RNA isolation, were performed on plants under control condition (CN), and at 10 and 20 days after stress imposition (DASI).

3.2.2 Physiological measurements

Physiological measurements were performed on the second fully expanded leaf of three independent papaya plants. Photosynthetic (A) and transpiration (T) rates were determined by means of a portable Li-6400 photosynthetic system (Li-Cor, Lincoln, NE, USA). Leaf water potential (Ψ) was determined by means of a Wescor thermocouple psychrometer sample chamber C-52 connected to a Wescor HR-33T Dew point microvoltmeter (Wescor Inc., Logan, UT, USA).

3.2.3 RNA isolation and Illumina sequencing

Samples from leaf, sap, and root tissues were collected from two independent plants, immediately frozen in liquid nitrogen, and stored at -80°C until use. Total RNA from these samples was isolated using TRIzol® reagent (Invitrogen®). Quality and concentration of purified RNA were assessed by 1% agarose gel electrophoresis and a 2100 Bioanalyzer (Agilent Technologies®). The cDNA libraries were sequenced at the Genomic Services Laboratory, Advanced Genomic Unit (UGA, Cinvestav-Langebio), Mexico. A total of 18 cDNA Libraries (Table 3.1) were prepared using the Illumina® TruSeq® RNA Sample Prep Kit v2, and paired-end sequenced in a 2×75 High Output configuration on the NextSeq500 Illumina platform. Quality check of RNA-seq raw reads was performed by means of FastQC (Andrews y others, 2010) software. Trimmomatic v0.36 (Bolger *et al.*, 2014) was used for adapter clipping and filtering, and only reads with a score $>Q30$ and a minimum length of 54 pb were kept (clean reads) for further analyses.

3.2.4 De novo transcriptome assembly and annotation

The clean reads of the 18 sequenced cDNA libraries were used to construct a de novo transcriptome assembly by means of Trinity v2.2.0 (Grabherr *et al.*, 2011) with default options. The longest isoform of each assembled unigene was annotated with Blast2GO (Götz *et al.*, 2008). Annotations were based on BLASTX (Altschul *et al.*, 1997) similarity searches against the Plant ref-seq protein database of the NCBI (ftp://ftp.ncbi.nlm.nih.gov/refseq/release/plant/*protein.faa.gz). BLASTX search parameters were: HSP cut-off length 33, report 20 hits, maximum E-value 1e-3. Blast2GO Mapping and annotation parameters were: E-value 1e-6, annotation cut-off 55, GO weight 5, HSP-hit coverage cut-off 20. Additionally, the unigenes were blasted against the gene models of the reference genome of the transgenic ‘SunUp’ papaya (Ming *et al.*, 2008), to find the correspondence between them. Furthermore, putative coding regions (>100 bp) from the unigenes were detected by means of Transdecoder v3.0.0 (<https://github.com/TransDecoder/TransDecoder/wiki>) with default options.

3.2.5 Expression level quantification

Expression level quantification was performed using both de novo and reference assembly approaches. For the de novo transcriptome analysis, expression quantification was estimated by mapping the 18 cDNA libraries (clean reads) to the assembled transcriptome by means of Bowtie2 (Langmead y Salzberg, 2012) and RNA-Seq by Expectation-Maximization (RSEM) software v1.2.27 (Li y Dewey, 2011), using the scripts included in the Trinity package (Haas *et al.*, 2013). For the reference-based transcriptome analysis, the 18 cDNA libraries (clean reads) were mapped to the reference genome of the transgenic ‘SunUp’ papaya (Ming *et al.*, 2008), downloaded from Phytozome v12.1 (Goodstein *et al.*, 2012), by means of TopHat (Trapnell *et al.*, 2009) and HTSeq (Anders *et al.*, 2015). Read counts resulting from the de novo (unigenes) and reference-based (genes) transcriptomes were normalized to transcripts per million (TPM).

3.2.6 Differential expression analysis

Two series of differential expression analyses were performed for both the reference and de novo assembly approaches. Firstly, leaves, sap, and roots samples were compared under control condition, i.e L-CN = L-CN_vs_S-CN and L-CN_vs_R-CN; S-CN = S-CN_vs_L-CN

and S-CN_vs_R-CN; R-CN = R-CN_vs_L-CN and R-CN_vs_S-CN (Figure 3.2 and Supplementary Figures S3 and S6). Secondly, control samples were compared against the stressed samples, i.e L-10 = L-CN_vs_L-10; L-20 = L-CN_vs_L-20; S-10 = S-CN_vs_S-10; S-20 = S-CN_vs_S-20; R-10 = R-CN_vs_R-10; R-20 = R-CN_vs_R-20 (Figure 3.2 and Supplementary Figures 4 and 7). Only read counts of genes and unigenes with TPM values of ≥ 1 and ≥ 16 (parameters established based on linear regression analysis), respectively, were used for differential expression analysis by means of EdgeR (Robinson *et al.*, 2010) in R v3.0 (R Development Core Team y R Core Team, 2017). Differentially expressed genes (DEGs) or unigenes (DEUs) were defined as those presenting an absolute fold change (FC) ≥ 2 and an adjusted P-value (FDR) ≤ 0.001 in any pairwise comparison.

3.2.7 GO functional enrichment analysis of sets of differentially expressed genes

For each set of DEGs obtained from the reference-based transcriptomic analysis, significant enriched GO terms (in the Biological Process category), based on Plaza 4.0 (Van Bel *et al.*, 2018) annotations, were detected by means of the Fisher's exact test (FDR < 0.05) implemented in Blast2GO (Götz *et al.*, 2008). Specific sets of DEGs under control condition (Supplementary Figure S3d) were compared against each other. Sets of up- and down-regulated DDEGs, identified in the reference-based transcriptomic analysis, were separately submitted to GO enrichment analysis by comparing them to the total expressed genes (TPM ≥ 1) in leaf, sap, and roots under CN condition. Then, enriched GO terms and genes (Arabidopsis identifiers) were clustered and visualized as networks by means of ClueGO v2.5.0 (Bindea *et al.*, 2009) in Cytoscape v3.6.0 (Shannon *et al.*, 2003).

3.2.8 Gene co-expression network analysis

TPM values of the sets of up-regulated genes (obtained from the reference-based transcriptomic analysis) found in leaves and roots (but not sap) at 10 and 20 DASI were used to build two tissue-specific pairwise co-expression matrices by means of the GENIE3 Bioconductor package (Aibar *et al.*, 2017; Huynh-Thu *et al.*, 2010) which is based in Random Forest machine learning model (Breiman, 2001), using default parameters. A total of 95 papaya gene models corresponding to transcription factors (TFs), 39 in leaves (Supplementary Table S7) and 56 in roots (Supplementary Table S8), from the principal families implied in responses to abiotic stress (AP2/ERF, AREB/ABF, NAC, MYB, bHLH,

WRKY, and HSF [Heat shock factor]) were set as “Regulator nodes”, and all of the up-regulated DDEGs (leaves and roots) were set as “Target nodes” for the building of these “TF-stress responsive genes” networks. Identified expression patterns were taken as an indication of putative regulatory links. All connections (links) between genes (nodes) were exported as tables and loaded into Cytoscape v.3.6.0 (Shannon *et al.*, 2003), and they can be found in Supplementary Tables S9 and S10. Calculation of network indices and parameters were performed using the NetworkAnalyzer built-in app in Cytoscape and can be inspected in the Supplementary Tables S7 and S8. The networks were taken as undirected but weighted. Network clusters or communities were determined based on topological edge connections using the GLay (Su *et al.*, 2010) network clustering algorithm plug-in in Cytoscape. Final networks were displayed using the yGraph Organic layout and the “Degree” parameter was used to depict the size of the nodes.

3.2.9 Statistical analysis and data visualization

All data analyses, graphics, and heatmaps were made in R 3.4.4 (R Development Core Team y R Core Team, 2017). Significant statistical differences of physiological parameters among treatments were determined with one-way ANOVA test followed by Tukey’s test ($P < 0.001$). Graphics were made with ggplot2 (Wickham, 2016). Venn diagrams were plotted with Vennerable package (<https://github.com/js229/Vennerable>). Heatmaps were drawn with either Heatmap3 (Zhao *et al.*, 2014) or ComplexHeatmap (Gu *et al.*, 2016) packages.

3.3 RESULTS

3.3.1 Effect of drought stress on papaya plants

We imposed drought stress by stopping watering on three-month-old ‘Maradol’ papaya plants. Visual examination of the plant phenotypes and physiological measurements were performed under control (CN) condition, and at 10 and 20 days after stress imposition (DASI). Under CN condition the plants appeared healthy and presented dark-green colored leaves (Figure 3.1a). At 10 DASI the plants still retained most of their leaves, however some leaves were curled and wilted (indicated with white arrows in Figure 3.1a). At 20 DASI the plants had lost most of their leaves and the remaining leaves were very curled and wilted (Figure 3.1a). The CO₂ assimilation rate (A), transpiration rate (T), and leaf water potential (Ψ) of the plants under CN condition presented mean values of 4.76 ± 0.07 s.d. $\mu\text{mol CO}_2 \text{ m}^{-2}\text{s}^{-1}$, 2.48 ± 0.33 s.d. $\text{mmol H}_2\text{O m}^{-2}\text{s}^{-1}$, and -0.33 ± 0.02 s.d. MPa, respectively (Figure 3.1b–d). Compared to CN plants, A mean value of the drought-stressed plants significantly decreased (Tukey $p < 0.001$) by 49.5 and 71.3% at 10 and 20 DASI, respectively (Figure 3.1b). T mean value decreased by 31.5% at 10 DASI, but was significantly reduced (Tukey $p < 0.001$) by 80.9% until 20 DASI (Figure 3.1c). Ψ mean value significantly decreased (Tukey $p < 0.001$) by 172 and 180% at 10 and 20 DASI, respectively (Figure 3.1d). We isolated RNA samples from leaves (L), sap (S), and roots (R) of these same plants and performed both reference and de novo RNA-seq transcriptomic analyses.

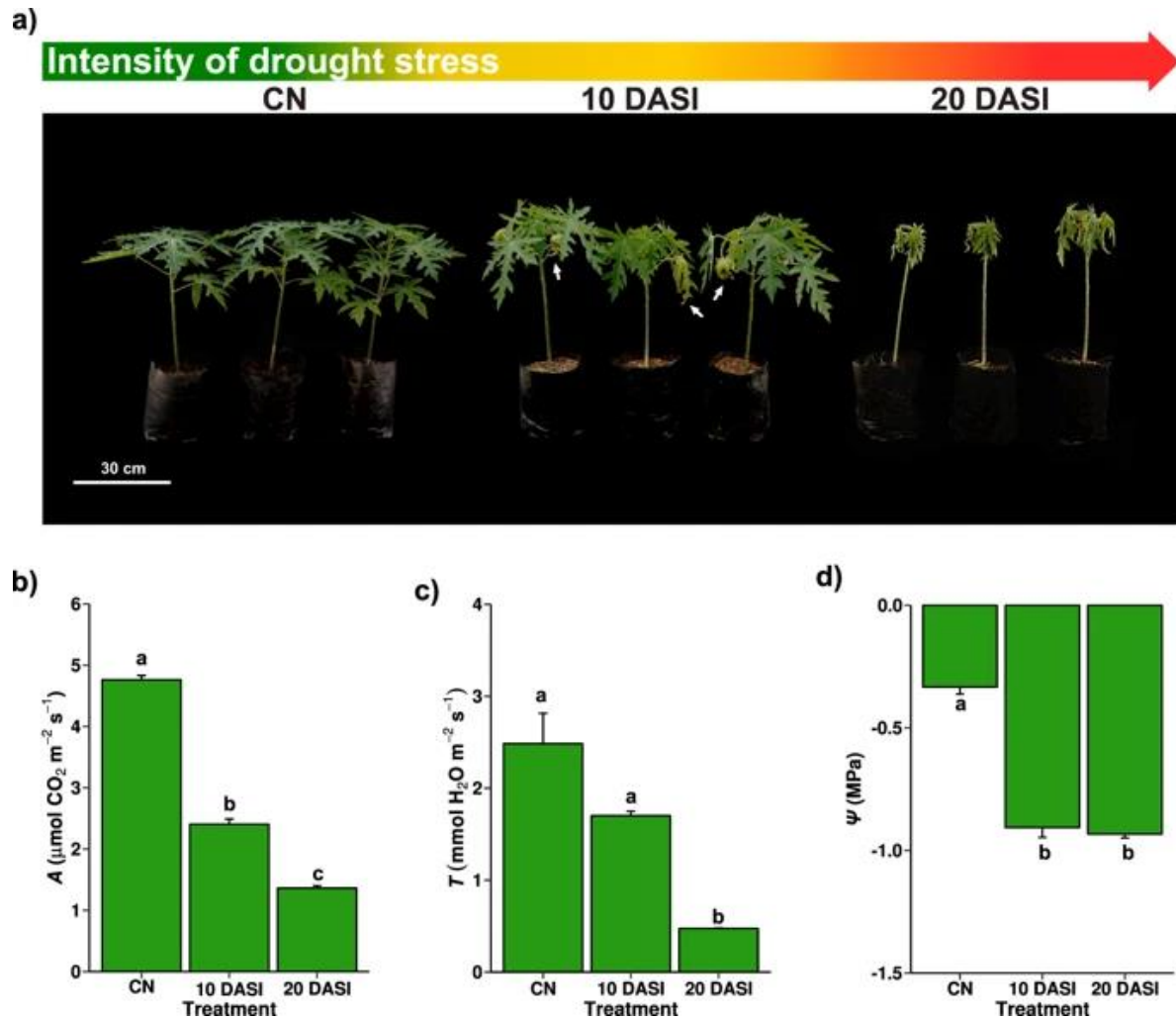


Figure 3.1. Effect of drought stress on papaya plants. (a) Three-month-old 'Maradol' papaya plants under control condition (CN) and at 10 and 20 days after stress imposition (DASI). The gradient-colored arrow indicates the intensity of the imposed stress: green = well-watered or no stress, yellow = mild stress, and red = severe stress. White arrows in the plants at 10 DASI indicate leaf wilting. (b–d) Effect of drought stress on papaya plant physiology: (b) CO_2 assimilation rate (A), (c) transpiration rate (T), and (d) leaf water potential (Ψ) of the papaya plant depicted in a). Bars and error bars represent means and standard deviation (s.d.), respectively, of three independent plants ($n = 3$). Different letters indicate statistically significant differences (ANOVA, Tukey $p < 0.001$).

3.3.2 RNA-Sequencing and expression level quantification

We sequenced a total of 18 cDNA libraries from 9 samples (in duplicates): S-CN, R-CN, L-CN, S-10, R-10, L-10, S-20, R-20, and L-20 (Table 3.1). We obtained a total of 621,077,480 raw reads, but we only kept 617,929,151 (99.49%) clean reads ($Q > 30$) (Table 3.1), which were used to de novo assemble the papaya transcriptome by means of Trinity software (Grabherr *et al.*, 2011) (for details please see Supplementary Figure S1 and Table S1). Expression levels were quantified by mapping the clean reads of each sample to both the extant reference genome of the transgenic ‘SunUp’ papaya (Ming *et al.*, 2008) and to our de novo assembled transcriptome. Hereafter, we will refer to as “genes” to the gene models from the reference genome, and “unigenes” to the features assembled in our de novo transcriptome. We only considered genes and unigenes as “expressed” if they had TPM values ≥ 1 and ≥ 16 , respectively. Based on this delimitation, a total of 21,360 genes (77% of the total 27,699 genes present in the ‘SunUp’ papaya genome) were expressed in the reference-based transcriptomic analysis (Figure 3.2a and Supplementary Table S2). Conversely, a total of 18,500 unigenes were expressed ($\text{TPM} \geq 16$) in our de novo transcriptome assembly (Figure 3.2b and Supplementary Table S3), and of those, 11,864 unigenes represented 43% of the total gene models from the reference genome (27,699 genes). In both reference and de novo transcriptomic analyses, Pearson’s correlation coefficient based on expression values of each library indicated high correlation among sample replicates (Supplementary Figure 2a,b). Hierarchical clustering and Principal Component Analysis (PCA) indicated major grouping of the samples according to tissue type, rather than stress treatment (Supplementary Figure 2a–d). We performed differential expression analysis on the sets of 21,360 and 18,500 expressed genes and unigenes.

Cuadro 3.1. Samples and libraries used for de novo and reference-based transcriptomic analyses.

Sample ^a	Treatment ^b	Library ^c	Raw Reads	GC%	Clean Reads (Q > 30)	GC%
S-CN	Control	S-CNa	27,805,197	42	27,640,601	42
		S-CNb	37,912,534	43	37,736,959	43
R-CN	Control	R-CNa	47,387,690	44	47,091,639	43
		R-CNb	36,666,468	43	36,503,924	43
L-CN	Control	L-CNa	42,292,905	44	42,143,433	44
		L-CNb	43,896,258	44	43,743,975	44
S-10	10 DASI	S-10a	32,528,720	43	32,077,543	43
		S-10b	39,861,192	44	39,687,165	44
R-10	10 DASI	R-10a	32,723,799	45	32,550,260	45
		R-10b	31,347,259	44	31,214,117	44
L-10	10 DASI	L-10a	22,994,548	44	22,902,184	44
		L-10b	39,835,409	44	39,702,331	43
S-20	20 DASI	S-20a	30,694,925	43	30,495,357	43
		S-20b	31,872,909	43	31,711,157	42
R-20	20 DASI	R-20a	33,133,317	43	32,986,618	43
		R-20b	26,405,792	44	26,292,893	43
L-20	20 DASI	L-20a	35,290,183	43	35,144,767	43
		L-20b	28,428,375	44	28,304,228	43
Totals=			621,077,480		617,929,151	

^aL = leaves, S = sap, R = roots; CN = Control, 10 = 10 DASI, 20 = 20 DASI. ^bDASI = days after stress imposition. ^c“a” and “b” indicate sample replicates.

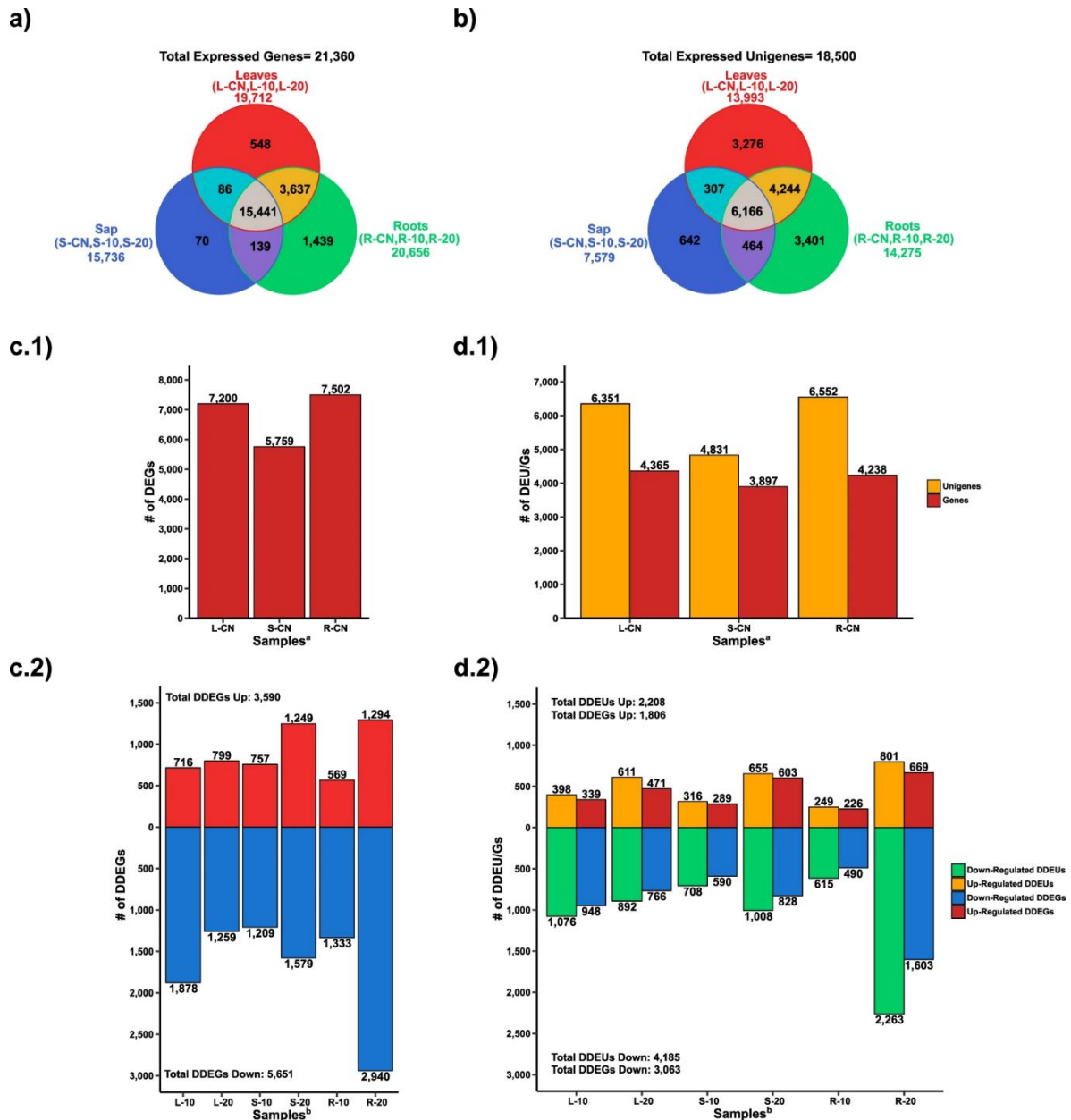


Figure 3.2. Reference-based and de novo transcriptomic analyses. (a) Total number of expressed genes (TPM ≥ 1) per tissue identified in the reference-based transcriptomic approach. (b) Total number of expressed unigenes (TPM ≥ 16) per tissue identified in the de novo transcriptomic approach. (c.1) Number of up-regulated genes among tissues under control (CN) condition identified in the reference-based transcriptomic analysis. (d.1) Number of up-regulated unigenes (yellow) and corresponding gene models (red) among tissues under CN condition identified in the de novo transcriptomic analysis. (c.2.) Number of up-regulated (red) and down-regulated (blue) DDEGs in stressed tissues identified in the reference-based transcriptomic analysis. (d.2) Number of up-regulated DDEUs (yellow) and corresponding gene models (red), and down-

regulated DDEUs (turquoise) and corresponding gene models (blue) in stressed tissues identified in the de novo transcriptomic analysis. DEU/G = Differentially expressed unigenes/genes; DDEU/G = Drought-responsive differentially expressed unigenes/genes. Sample names are described in Table 3.1. aComparisons to obtain the number of up-regulated genes by tissue type under control condition: L-CN = L-CN_vs_S-CN + L-CN_vs_R-CN; S-CN = S-CN_vs_L-CN + S-CN_vs_R-CN; R-CN = R-CN_vs_L-CN + R-CN_vs_S-CN. bComparisons to obtain the number of up- and down-regulated genes under drought stress: L-10 = L-CN_vs_L-10; L-20 = L-CN_vs_L-20; S-10 = S-CN_vs_S-10; S-20 = S-CN_vs_S-20; R-10 = R-CN_vs_R-10; R-20 = R-CN_vs_R-20.

3.3.3 Differential Expression Analysis

In the reference-based transcriptomic analysis, under control condition we found totals of 7,200, 5,759, and 7,502 up-regulated genes in L-CN, S-CN, and R-CN, respectively (Figure 3.2c.1, Supplementary Figure S3, and Supplementary Table S4). Moreover, under drought-stress conditions we identified a total of 8,549 Drought-responsive Differentially Expressed Genes (DDEGs) of which 3,590 were up-regulated and 5,651 were down-regulated (Figure 3.2c.2 Supplementary Figure S4, and Supplementary Table S4). Hierarchical clustering of the expression values of the total DDEGs across samples revealed 7 major clusters; the samples were grouped, firstly, according to tissue type, and secondly, according to stress treatments (Supplementary Figure S5a,b).

In the de novo transcriptomic analysis, under CN condition we found totals of 6,351, 4,831, and 6,552 up-regulated unigenes in L-CN, S-CN, and R-CN, respectively (Figure 3.2d.1, Supplementary Figure S6, and Supplementary Table S5). These unigenes corresponded to 4,365, 3,897, and 4,238 gene models from the reference genome in L-CN, S-CN, and R-CN, respectively (Figure 3.2d.1). Moreover, under drought-stress conditions we found a total of 6,089 Drought-responsive Differentially Expressed Unigenes (DDEUs) of which 2,208 were up-regulated and 4,185 were down-regulated (Figure 3.2d.2, Supplementary Figure S7, and Supplementary Table S5). These unigenes corresponded to 1,806 up-regulated gene models, and 3,063 down-regulated gene models, from the reference genome of the transgenic 'SunUp' papaya (Figure 3.2d.2). Clustering analysis of the total DDEUs revealed 6 major expression patterns; the samples were grouped, firstly, according to tissue type, and secondly, according to stress treatments (Supplementary Figure S5c,d).

Between 58 and 85% of the DDEUs identified by means of the de novo approach, were shared with the DDEGs identified in the reference-based approach (Supplementary Figure S8a–c). This indicates that both transcriptomic approaches recovered similar sets of differentially expressed gene models from the reference genome. We only utilized the genes from the reference-based transcriptomic analysis for further enrichment and co-expression analyses.

3.3.4 Core sets of DDEGs identified in the reference-based transcriptomic analysis

We compared the intersection of all sets of up- and down-regulated DDEGs (identified in our reference-based transcriptomic analysis) across all tissues under stress treatments to determine shared core sets. We found 6 genes that were inside of a shared Core set of up-regulated genes (CUG) (Figure 3.3a) and 34 genes that were inside a shared Core set of down-regulated genes (CDG), across all samples under stress treatments (Figure 3.3b). The CUG set included *evm.TU.model_supercontig_232.12* (dihydroflavonol 4-reductase-like1), *evm.TU.model_supercontig_81.90* (highly ABA-induced PP2C gene 2), *evm.TU.model_supercontig_9.214* (Rubber elongation factor protein (REF)), and *evm.TU.model_supercontig_217.19* (sucrose phosphate synthase 2 F) (clade marked with a red circle in Figure 3.3c). Additionally, twelve genes within the CDG set had relatively high expression in leaves, sap, and roots under CN condition (clade marked with a green circle in Figure 3.3c). These genes included *evm.TU.model_supercontig_19.124* (PYR1-like 6), *evm.TU.model_supercontig_109.28* (tonoplast intrinsic protein 2;2), and *evm.TU.model_supercontig_21.12* (Gibberellin-regulated family protein) (Figure 3.3c). Twenty-two genes within the CDG set had relatively high expression in roots and leaves, but not sap, under CN condition (clade marked with a blue circle in Figure 3.3c). These genes included *evm.TU.model_supercontig_50.19* (myb domain protein 14), *evm.TU.model_supercontig_131.3* (Leucin-rich repeat protein kinase family protein), and *evm.TU.model_supercontig_6.303* (Peroxidase superfamily protein) (Figure 3.3c). Other core sets of DDEGs were identified for 10 DASI and 20 DASI treatments for each plant tissue. For example, 270 (CUL), 368 (CUS), and 395 (CUR) DDEGs were determined as up-regulated cores; and 812 (CDL), 348 (CDS), and 1,175 (CDR) DDEGs were determined as down-regulated cores in leaves, sap, and roots, respectively (Figure 3.3a,b,

Supplementary Figure S9a, and Supplementary Table S2). Twelve (CU10) and 79 (CU20) DDEGs were up-regulated in all three tissues, and 77 (CD10) and 171 (CD20) DDEGs were down-regulated in all three tissues, at 10 DASl and 20 DASl, respectively (Figure 3.3a,b, Supplementary Figure S9b, and Supplementary Table S2).

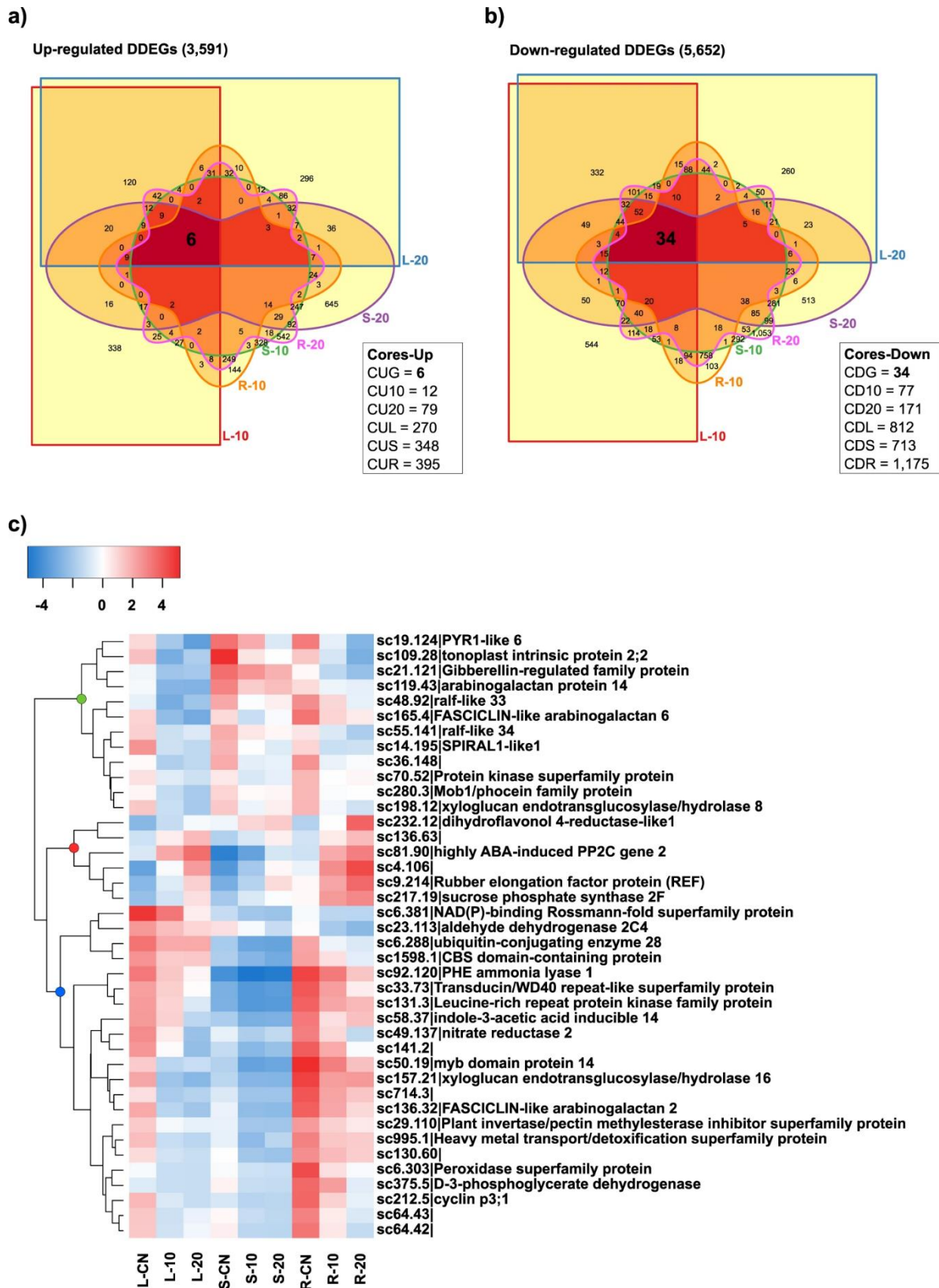


Figure 3.3. Shared core sets of up- and down- regulated DDEGs identified in the reference-based transcriptomic analysis. (a) Venn diagram depicting the intersections of all sets of up-regulated DDEGs. (b) Venn diagram depicting the intersections of all sets of down-regulated DDEGs. (c) Heatmap depicting the expression of the shared core sets of up- or down-regulated DDEGs (CUG and CDG respectively) in all tissues and treatments. In a) and (b): CUG = core of up-regulated genes in all tissues during stress treatments; CU10 = core of up-regulated genes in all tissues at 10 DAS; CU20 = core of up-regulated genes in all tissues at 20 DAS; CUL = core of up-regulated genes in L-10 and L-20; CUS = core of up-regulated genes in S-10 and S-20; CUR = core of up-regulated genes in R-10 and R-20; CDG = core of down-regulated genes in all tissues during stress treatments; CD10 = core of down-regulated genes in all tissues at 10 DAS; CD20 = core of down-regulated genes in all tissues at 20 DAS; CDL = core of down-regulated genes in L-10 and L-20; CDS = core of down-regulated genes in S-10 and S-20; CDR = core of down-regulated genes in R-10 and R-20. In (c) green circle = clade of genes relatively highly expressed under CN condition in leaves, sap, and roots; red circle = clade of the core set of 6 up-regulated DDEGs in all tissues and treatments (CUG); blue circle = clade of genes relatively highly expressed in leaves and roots, but lowly expressed in sap, under CN condition. In gene annotations in (c): “sc” = “evm.TU.supercontig_”. Heatmap color key indicates the mean-centered $\log_2(\text{TPM} + 1)$ values of the mean of TPMs per sample duplicates. DDEG = Drought-responsive differentially expressed genes. Separate shared core sets are depicted in Supplementary Figure S9. Samples names are described in Table 3.1.

3.3.5 GO functional enrichment analysis of the sets of differentially expressed genes under control condition

We identified enriched GO terms in tissue-specific sets of DEGs under CN condition identified by our reference-based transcriptomic approach, i.e. 2,641, 3,527, and 2,703 DEGs for L-CN, S-CN, and R-CN, respectively (Supplementary Figure S3d). Totals of 35, 13, and 36 enriched GO terms were found in L-CN, S-CN, and R-CN, respectively (Supplementary Table S6). As expected, in L-CN several enriched GO terms (17) were related to photosynthesis and response to light, for example: “chlorophyll biosynthetic process”, “photosynthetic electron transport in photosystem I”, “photosystem II assembly”, and “response to red light” (Figure 3.4a, and Supplementary Table S6). Other GO terms were found that related to development (ovule development), rRNA and tRNA metabolism, pigment biosynthesis (chlorophyll), and seven biosynthetic pathways, among other biological processes (Supplementary Figure S10a, and Supplementary Table S6). The S-CN sample was enriched in GO terms related to transport (“intracellular protein transport”, “ER to Golgi vesicle-mediated transport”, “vacuolar transport”), RNA and protein

modification (“mRNA splicing, via spliceosome”, “protein glycosylation”, “protein modification by small protein removal”, “proteasome-mediated ubiquitin-dependent protein catabolic process”, “protein processing”), translation (“mature ribosome assembly”, “translational initiation”) and stress response (“response to endoplasmic reticulum stress”), among others (“small GTPase mediated signal transduction” and “vacuole organization”) (Figure 3.4b, Supplementary Figure S10b, and Supplementary Table S6). In R-CN the 10 most enriched GO terms were: “hydrogen peroxide catabolic process”, “protein phosphorylation”, “response to chitin”, “defense response to fungus”, “oxidation-reduction process”, “positive regulation of transcription, DNA-templated”, “microtubule-based movement”, “hormone-mediated signaling pathway”, “mitotic cell cycle process”, and “response to salicylic acid” (Figure 3.4c, and Supplementary Table S6). Other GO terms related to the regulation of cellular process, cell wall metabolism, development, and secondary metabolism were also represented (Supplementary Figure S10c, and Supplementary Table S6).

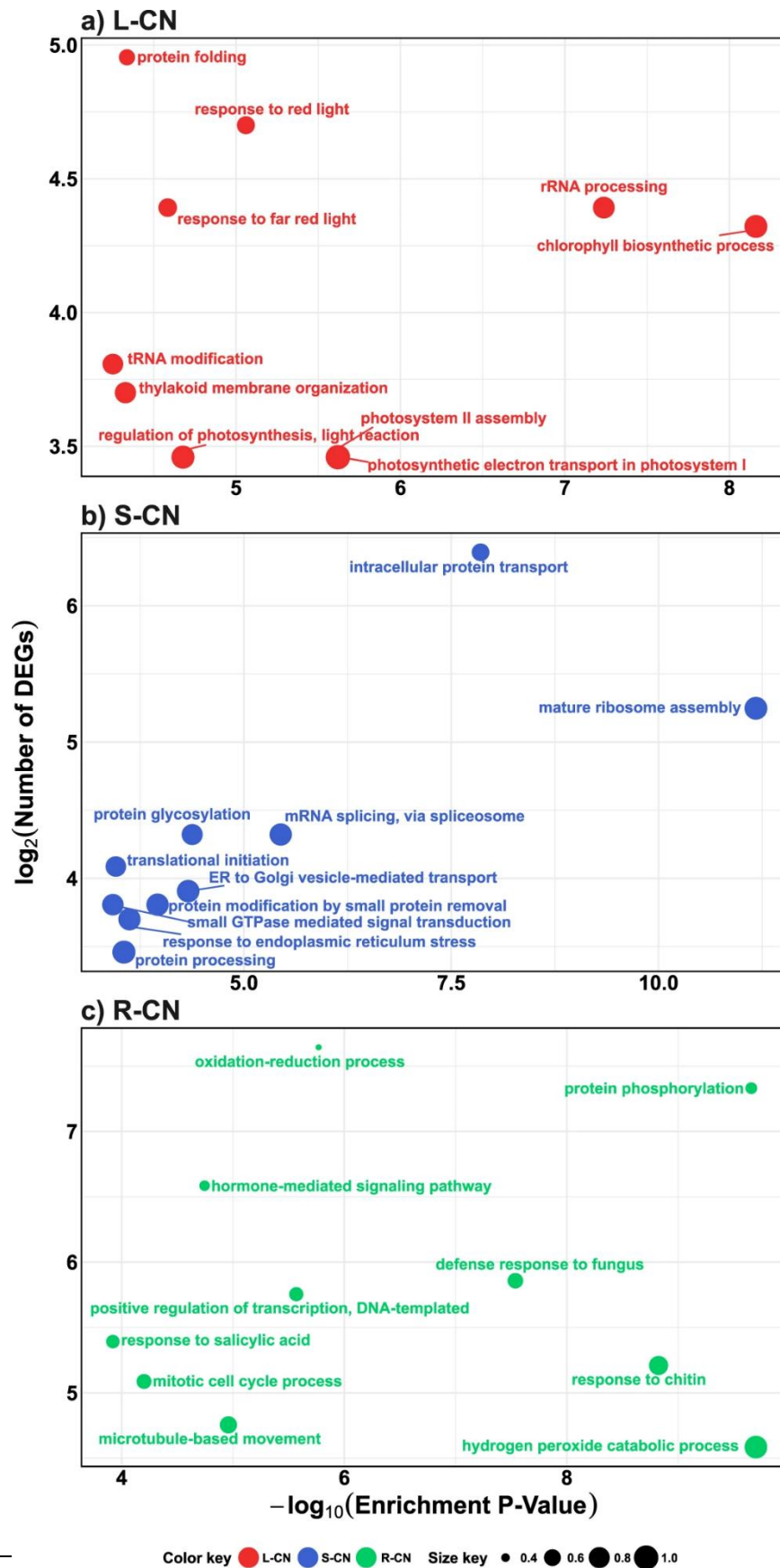


Figure 3.4. Functional GO enrichment analysis of the sets of up-regulated genes in leaf, sap, and roots under control condition identified in the reference-based transcriptomic analysis. Top 10 enriched GO terms (Biological Process category) identified in (a) L-CN, (b) S-CN, and (c) R-CN. Only the top 10 (based on enrichment P-Value) enriched GO terms are plotted in function of their enrichment P-value (x axis) and number of genes (y axis). Size key indicates the ratio of the number of genes in the test-set/total number of genes, for any given GO term. The complete lists of enriched GO terms are presented in Supplementary Table S6.

3.3.6 GO functional enrichment analysis of the sets of DDEGs

We found enriched GO terms for each tissue (leaf, sap, and root) at a specific time-point stress treatment (10 and 20 DASI). All sets of up-regulated and down-regulated GO terms can be inspected in the Supplementary Table S6. At 10 DASI, five of the top 10 enriched GO terms in the up-regulated gene set in leaves (L-10) were related to cell cycle, three were related to DNA molecule, one to sexual reproduction, and one to micro-tubule based movement. (Figure 3.5a, Supplementary Figure S11a and Supplementary Table S6). In sap (S-10) “DNA metabolic process”, “double-strand break repair”, and “meiotic cell cycle process” were enriched in the up-regulated gene set (Figure 3.5b, Supplementary Figure S11b, and Supplementary Table S6). In roots (R-10), four terms related to cuticle or suberin formation, two to abiotic stress response (water and salt), two to hormones (ABA and gibberellin), two to carbohydrate metabolism, among others, were enriched in the up-regulated gene set (Figure 3.5c, Supplementary Figure S11c, and Supplementary Table S6). At 20 DASI response to water deprivation, ABA-related responses, and carbohydrate metabolism were enriched in the up-regulated gene sets of all three samples (L-20, S-20, and R-20). Other GO terms related to heat, salt, and oxidative stresses were also enriched in the up-regulated gene sets at 20 DASI (Figure 3.5d–f, Supplementary Figure S11d–f, and Supplementary Table S6). At both 10 and 20 DASI (and in all tissues), we found enriched GO terms in the down-regulated gene sets that related to: abiotic stress responses (such as water and osmotic stresses), defense related hormones (such as jasmonic acid, salicylic acid, and abscisic acid), defense against biotic stresses, cell wall metabolism, oxidation-reduction, ion transport, and development (Figure 3.5a–f, Supplementary Figure S11g–l, and Supplementary Table S6). “Response to karrikin” was also an enriched GO term in all down-regulated sets of DDEGs (Figure 3.5a–f, and Supplementary Table S6).

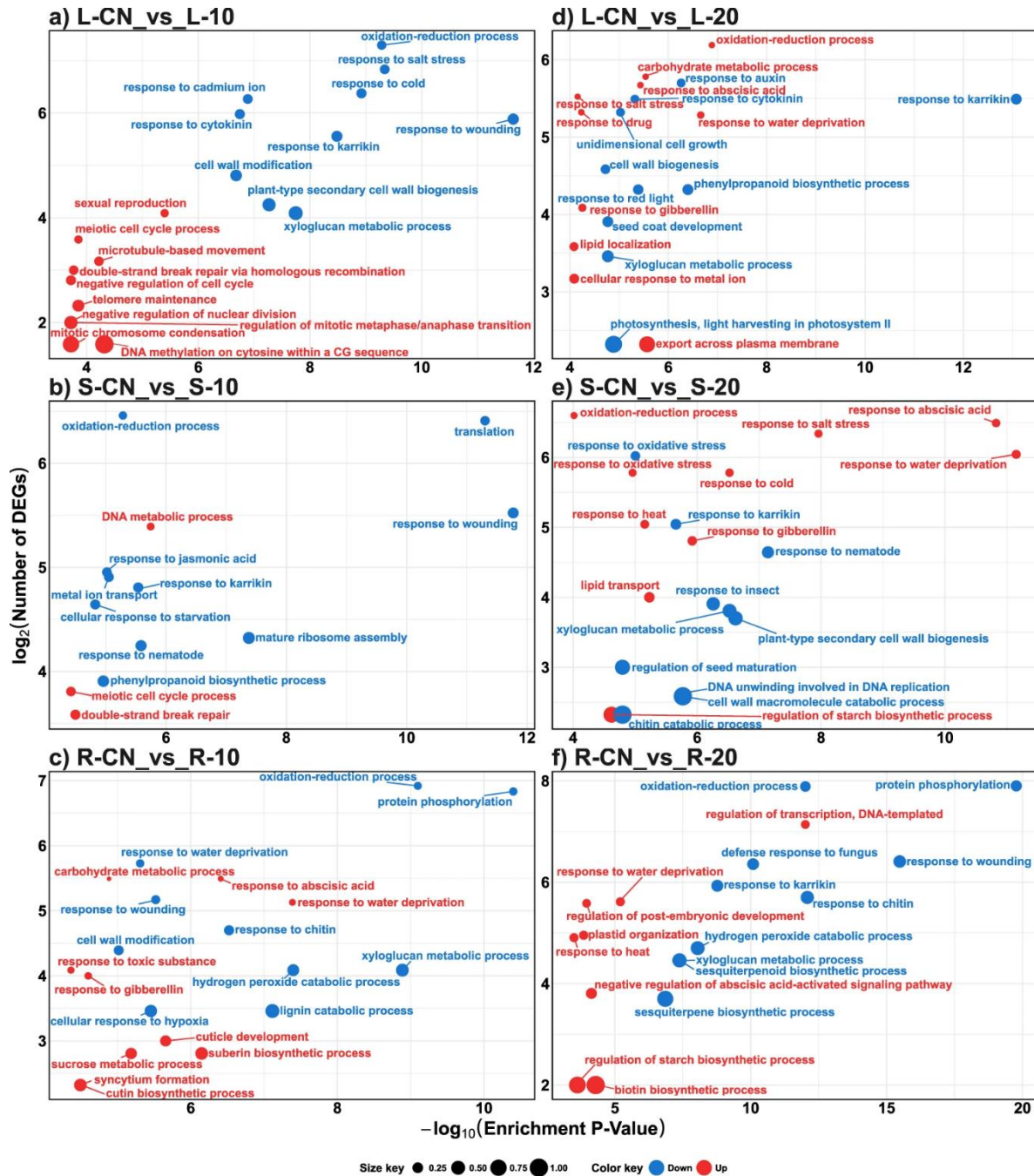


Figure 3.5. Functional GO enrichment analysis of the sets of DDEGs identified in the reference-based transcriptomic analysis. Top 10 enriched GO terms (Biological Process category) identified in the sets of up-regulated (red) and down-regulated (blue) DDEGs in (a) L-CN_vs_L-10, (b) S-CN_vs_S-10, (c) R-CN_vs_R-10, (d) L-CN_vs_L-20, (e) S-CN_vs_S-20, and (f) R-CN_vs_R-20. Only the top 10 (based on enrichment P-Value) enriched GO terms are plotted in function of their enrichment P-value (x axis) and number of genes (y axis). GO terms in sets of up-regulated and down-regulated genes

are indicated in red and blue, respectively. Size key indicates the ratio of the number of genes in the test-set/total number of genes, for any given GO term. DDEG = Drought-responsive differentially expressed genes. The complete lists of enriched GO terms are presented in Supplementary Table S6.

3.3.7 Co-expression networks

Based on TPM values of up-regulated DDEGs only, we built two independent gene co-expression networks (GCNs), one for leaves and another for roots, and detected natural gene co-expression communities (Figure 3.6 and Supplementary Figures S12 and S13). In such networks the up-regulated DDEGs are represented by nodes, and pairwise co-expression relationships between DDEGs are represented by edges. Selected TFs were set as “regulator nodes” of the network in order to detect regulatory hubs (listed in the “Regulator nodes” column in Supplementary Tables S7 and S8). For the leaf GCN we determined three communities (I-a–III-a) composed of a total of 921 DDEGs (nodes) connected to 37 (of 39) TFs set as “regulator nodes” (Figure 3.6a and Supplementary Table S7). Moreover, 42.9, 39.8 and 17.1% of these 921 DDEGs were up-regulated at 10 DASI, at 20 DASI, and at both stress treatments, respectively (see node color key in Figure 3.6a). These DDEGs were clearly clustered according to time-point treatments: Communities I-a and II-a presented a high proportion (88 and 69% respectively) of up-regulated DDEGs at 10 DASI (blue nodes); community III-a presented a high proportion (74%) of up-regulated DDEGs at 20 DASI (orange nodes) and contained most of the genes (10.9%) that were up-regulated under both stress treatments in the leaf GNC (purple nodes, Figure 3.6a). Furthermore, 69 genes belonging to enriched GO terms related to cell cycle process (nodes in different red tones, orange, and pink) were clustered at 10 DASI in the I-a community (Figure 3.6b), and 36, 44, and 22 genes belonging to enriched GO terms related to abiotic stress stimulus (nodes in different green tones), oxidation-reduction (cyan nodes), and carbohydrate metabolic processes (dark blue nodes) were clustered at 20 DASI in the III-a community (Figure 3.6b).

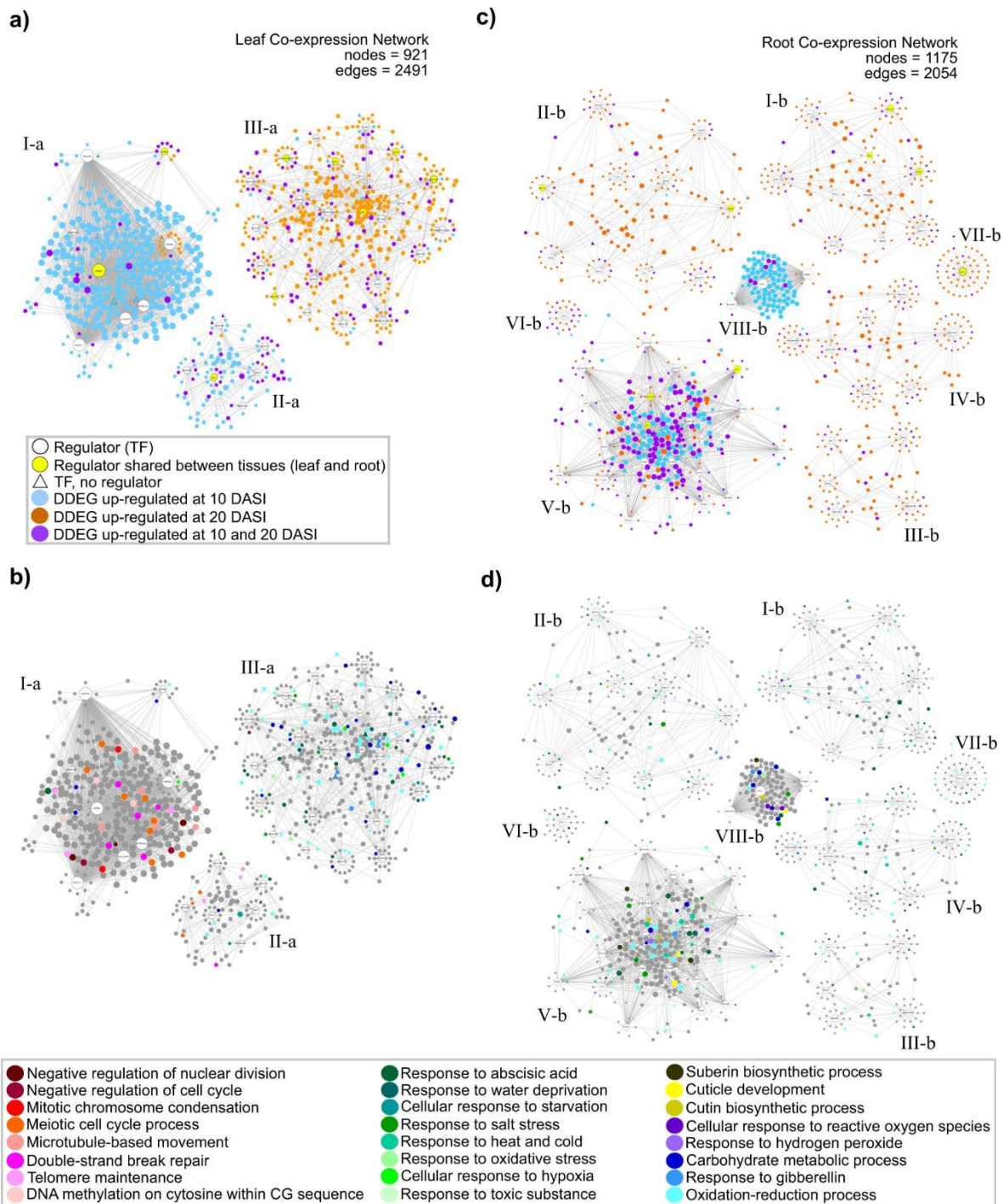


Figure 3.6. Gene co-expression networks (GCN) of leaf and root transcriptomes of papaya plant during drought treatments. Up-regulated TFs, previously reported in response to abiotic stress and found in the stress datasets were set as “Regulators” for the construction of the networks. (a) GCN for leaf tissues showing time-point stress responses of up-regulated DDEGs. The three communities are labeled as I-a, II-a, and

III-a, and the 37 most connected TFs are depicted in the network as white circles. (b) Same GCN as in a), but showing the location in the network of genes belonging to different enriched GO groups by Biological Process (BP). (c) GCN for root tissues showing time-point stress responses of up-regulated DDEGs. The eight communities are labeled as I-b–VIII-b, and the 56 most connected TFs are depicted in the network as white circles. (d) Same GCN as in b), but showing the location in the network of genes belonging to different enriched GO groups by BP. Nodes in (a) and (c) are colored according the time-point stress treatments, and nodes in (b) and (d) are colored according to enriched GO groups as shown in the Figure legends. DASI = days after stress imposition. Detailed networks are depicted in Supplementary Figures S12 and S13.

For the root GCN we determined eight communities (I-b–VIII-b) composed of a total of 1,175 DDEGs (nodes) connected to the total 56 TFs set as “regulator nodes” (Figure 3.6c and Supplementary Table S8). In this root GCN the nodes (DDEGs) were more scattered and less connected than in leaf GCN (compare Figure 3.6a,c). Different from the leaf GCN, in the root GCN most of the DDEGs (59.3%) were up-regulated at 20 DASI, followed by those shared (elements) by both stress treatments (26%), and finally those DDEGs up-regulated at 10 DASI (14.6%) (see node color key in Figure 3.6c). Five communities (I-b–IV-b, and VII-b) contained a high proportion (between 79 and 88% per community) of up-regulated DDEGs at 20 DASI (orange nodes) and a low proportion (between 11 and 19%) of shared elements (purple circles). Community V-b was composed by a mix DDEGS up-regulated at 10 or 20 DASI (24 and 25% respectively; blue and orange nodes) and a high percentage (50.1%) of shared elements between both stress treatments (purple circles). Community VI-b contained high proportions of DDEGs up-regulated at 20 DASI (40%) and under both stress treatments (43%), and a low proportion of DDEGs at 10 DASI (16%). Community VIII-b consisted primarily (84%) of DDEGs up-regulated at 10 DASI (blue nodes) (Figure 3.6c). Furthermore, we found that the enriched GO terms related to cellular responses, to abiotic stress stimulus (nodes in different green tones), oxidation-reduction process (cyan nodes), responses to reactive oxygen species (purple nodes), and cutin and cuticle development (yellow nodes) were mainly found in communities V-b and VIII-b from the 10 DASI (Figure 3.6d). The remaining communities showed a minimal and scattered distribution of GO terms.

3.3.8 TFs as regulatory hub

In both leaf and root GCNs, we identified the regulator nodes (i.e. selected TFs) with highest connectivity because of their relevance as regulatory hubs in response to drought (highlighted in red in the Degree column in Supplementary Tables S7 and S8). In the leaf GCN, the six TFs with the highest degree distribution (from 299 to 266) were WRKY70 (evm.TU.supercontig_19.44), MYB94 (evm.TU.supercontig_111.6), RAP2.11 (evm.TU.supercontig_51.136), bHLH (STP; evm.TU.supercontig_55.132), HSFB-2A (evm.TU.supercontig_107.31), and AP2/ERF (evm.TU.supercontig_2.268) homologues (Supplementary Table S7). These six TFs were clustered in the community I-a and linked with other DDEGs at 10 DAS (Supplementary Figure S12). In contrast to the leaf GNC, in the root GNC the TFs regulators were less connected and with lower degree distribution values. The three regulators in the roots with the highest degree distribution (from 91 to 86) were MYB63 (evm.TU.supercontig_34.3), bHLH (ICE1; evm.TU.supercontig_70.77), and bHLH (evm.TU.contig_26556.1) homologues (Supplementary Table S8). These three regulators were clustered together in the community VIII-b, composed mainly of genes up-regulated at 10 DAS (Supplementary Figure S13). Furthermore, nine TFs regulators were shared between the leaf and root GCNs, which are homologues of ABI5 (evm.TU.supercontig_5.28), RAP2.6 (evm.TU.supercontig_38.79), bHLH (evm.TU.supercontig_20.63), ANAC072 (RD26; evm.TU.supercontig_80.93), ANAC074 (evm.TU.supercontig_165.12), MYB48 (evm.TU.supercontig_190.35), MYB94 (evm.TU.supercontig_111.6), bZIP1 (evm.TU.supercontig_9.75), and WRKY75 (evm.TU.supercontig_807.4) (yellow nodes in Figure 3.6a,b, Supplementary Figure S12 and S13, and Supplementary Tables S7 and S8). In the leaf GCN, six of these regulators were clustered at 20 DAS in the III-a community; only MYB94 was among the highest degree distribution TFs mentioned above (Supplementary Figure S12). In the root GCN, the nine TFs were included in the I-b, II-b, V-b, and VII-b communities (Supplementary Figure S13); however, none of them was among the highest degree distribution TFs grouped in community VIII-b.

3.4 DISCUSSION

3.4.1 Biological processes differentially modulated under CN condition

Papaya tissues clearly presented different expression patterns (Supplementary Figures S2 and S5). Under the control condition, root tissues were already enriched with GO terms that might relate to stress response and defense, such as hydrogen peroxide catabolic process (Figure 3.4c). Similar results were obtained from the transcriptomic analysis of papaya roots in a previous study (Porter *et al.*, 2008). Evidently, the roots maintain these processes for a rapid response upon water deficit. The sap tissue presented a unique transcriptomic profile enriched with several biological processes related to mRNA splicing via spliceosome, RNA and protein transport, protein post-translational modifications (for example glycosylation, deneddylation, and proteolysis), regulation of localization, SNARE interactions, and stress (even under CN condition) (Supplementary Figure S10b). Similar results were identified in the transcriptome of sap in melon fruit (Omid *et al.*, 2007). Grafting experiments on papaya plants could provide a further understanding of how transcripts move long-distance through the phloem sap. The transcriptome profile in leaves was enriched with GO terms related to photosynthesis and chloroplast metabolic processes as expected (Figure 3.4a). Promoter regions of these tissue-specific genes (from leaves, sap, and roots) could be useful for the regulation of cis-genes in papaya plant.

3.4.2 Shared core sets of stress regulated genes, and genes related to water deprivation and ABA as targets for genetic improvement in papaya

We found shared core sets of 6 and 34 DDEGs that were up- or down-regulated, respectively, in all stressed samples (Figure 3.3a,b). Interestingly, in the shared core set of up-regulated genes, the Rubber elongation factor protein (REF) showed an important transcriptional regulation in the three studied tissues (Figure 3.3c). This protein takes part in the biosynthesis of natural rubber (a component of latex) which has been implied as playing a key role in defense against pathogens, preventing their entry into wounded tissues (El Moussaoui *et al.*, 2001). This finding is interesting because papaya is a lactiferous species that shows a dense network of articulated and anastomosing laticifer vessels in leaves, stems (El Moussaoui *et al.*, 2001) and root tissues (Rao *et al.*, 2013). Homologs of REF

protein in non-rubber producing species such as hot pepper demonstrate its involvement in drought tolerance and other stress conditions (Kim *et al.*, 2010, 2016). So, this drought responsive REF homolog in papaya could also be involved in mechanisms that both prevent pathogen attacks under unfavorable conditions and cope with abiotic stress. Together with other candidate genes, these sets of shared core genes are promising targets for further functional studies addressing genetic improvement due to their consistent response and expression across all samples. On the other hand, we found 142 and 192 only-up- and only-down-regulated genes under any treatment, which were enriched in terms related to water deprivation or ABA (Supplementary Figure S14 and Table S2). These gene sets included several genes involved in ABA metabolism and signaling in response to stress, such as ABI1, ABI5, PYR1-like, and Protein phosphatase 2 C (Supplementary Figure S14a,b). We also found several TFs related to MYB, HD, bZIP, bHLH, NAC, and WRKY families; chaperones; and redox-related enzymes. The functional analysis of these genes in planta could lead to the development of crop varieties with higher tolerance to drought. For example, TFs are attractive targets for the generation of plants with higher tolerance to abiotic stresses (H. Wang *et al.*, 2016), since TFs are able to simultaneously control the expression of many stress related genes.

3.4.3 Biological processes differentially up-regulated under moderate drought stress

At 10 DASI the papaya plants were already under the effects of the imposed drought stress as demonstrated by the reduction in physiological performance (Figure 3.1a); however, the plants still retained several leaves, so we designated this time-point as a moderate drought stress. Like other types of stress, drought increases the production of ROS, which can damage different types of molecules, including DNA. This damage can result in reduced protein synthesis and genomic instability, which can ultimately contribute to a reduction in plant performance (Tuteja *et al.*, 2009). In the present study we found that double-strand break-repair processes were enriched in the up-regulated gene sets in leaves and sap, under moderate drought stress condition (10 DASI) (Figure 3.5a,b). In both, leaf and sap, homologues of RPA32B (evm.TU.supercontig_233.2) and RAD54 (evm.TU.supercontig_62.19) were up-regulated (Supplementary Table S4). Moreover, in

roots we found enriched GO terms related to common abiotic stress responses such as water deprivation, salt stress, ABA, and oxidation-reduction process (Supplementary Table S6). We also found that genes related to suberin biosynthesis were enriched in roots at 10 DASI. Suberin is an extracellular biopolymer found in the cell walls of aerial and underground tissues of plants, which modulates water movement and solute uptake, and is thought to play an important role in plant tolerance against drought, salinity, and pathogen attack (Vishwanath *et al.*, 2015; Schachtman y Goodger, 2008). In this study several genes involved in the synthesis and transport of suberin monomers were up-regulated in roots at 10 DASI, such as homologues of 3-ketoacyl-CoA synthase 2 (evm.TU.supercontig_1103.1), cytochrome P450 (evm.TU.supercontig_112.48), and ABC-2 type transporter (evm.TU.supercontig_114.19). These results suggest that under moderate drought stress papaya plants preferentially regulate different biological processes in roots, as opposed to the sap and leaves, conceivably because roots experience drought stress first.

3.4.4 Biological processes commonly modulated under severe drought stress

At 20 DASI, the papaya plants reached their lowest performance and clearly presented observable foliar damage (Figure 3.1a), so we considered this treatment as severe drought stress. At this time-point, genes involved in water deprivation or salt stress, which are closely related stresses, were enriched in the up-regulated DDEG sets. Additionally, up-regulated DDEGs involved in cold, heat, starvation, and hypoxia stress were enriched, (Supplementary Table S6) indicating a cross-talk between different stress pathways. Furthermore, response to ABA, and oxidation-reduction processes were also up-regulated. On the other hand, several terms related to abiotic stress, ABA (and several hormones), and oxidation-reduction were enriched in the down-regulated gene sets. GO terms related to biotic stress were also down-regulated in all samples. Moreover, plant cell wall composition and elasticity are modulated in response to drought and other stresses (Gall *et al.*, 2015). We found “plant cell metabolism” among the enriched GO terms in the down-regulated gene set, which may indicate that roots are adapting to drought stress.

3.4.5 Long-distance movement of mRNA under drought conditions

The study of the long-distant movement of mRNA through vascular system, most precisely in sap phloem, has attracted more attention in recent years focusing mainly on plant development and response to pathogens attack (Julia y Friedrich, 2018; Thieme *et al.*, 2015; Spiegelman *et al.*, 2013). However, little attention has been paid under abiotic stress. In this work, papaya sap transcriptome presented a unique profile in comparison to leaves and roots (Supplementary Figure S5a,c), showing a large number of genes (3,527) that were specifically regulated in this vascular tissue, even in non-stress conditions (Supplementary Figure S3d). Sap profile was enriched in GO terms involved in RNA and protein trafficking, mRNA splicing, mature ribosome assembly, translation initiation, post-translational modifications, and several proteins that resemble a nuclear environment (Figure 3.4). This enrichment results are in agreement with previous work by Figueroa and cols. (Figueroa-Yañez *et al.*, 2016), which demonstrated by in vitro translation assay, that protein translation could occur in sap.

PlaMoM database (Guan *et al.*, 2017) reports 11,440 experimentally confirmed mobile genes in Arabidopsis, which represent 5,234 loci. Through homology comparison (Blast searches) against these Arabidopsis loci, we identified 4,408 papaya gene models (from the total 27,769) that could act as putative mobile genes (mRNAs) (Supplementary Table S2). According to ClueGo analysis, this set of putative mobile mRNAs were enriched in biological processes related to photosynthesis, response to hormone, response to temperature stimulus, response to inorganic substance, nucleic acid metabolic process, chromosome organization, cellular macromolecule localization, among others (Supplementary Figure S15a). According to their Molecular Function (MF), 74 and 88 of these mobile papaya mRNAs were clustered in DNA binding and mRNA binding GO terms, respectively (Supplementary Figure S15b), which is in accordance to previous studies (Julia y Friedrich, 2018).

From the totals of DDEGs per tissue, including both up- and down-regulated, (3,566, 3,721, and 4,564 in leaf, sap, and root, respectively), we detected 846, 929, and 995 putative mobile DDEGs in leaf, sap, and root tissues, respectively, (Supplementary Figure S15c). Interestingly, the tissue-specific set of 435 mobile DDEGs in sap were only enriched in GO terms within the BP category related to response to high/low light intensity and

photosynthesis processes (Supplementary Figure S15c). It is worth mentioning that only 24.9% of the total (3,721) of sap DDEGs found in our transcriptome analysis corresponded with those reported as mobile in Arabidopsis PlaMoM database. We also detected a shared core set of mobile mRNAs in response to drought (Supplementary Figure S15d). This intersection consisting in 182 mobile DDEGs were enriched in GO terms within the BP category related to heat response, secondary metabolic processes, and regulation of signal transduction (Supplementary Figure S15d,e). As proposed by Thieme and cols. (Thieme *et al.*, 2015), long-distance mRNA trafficking, could be a rapid alert to distant tissues to achieve a systemic adaptation of an upcoming adverse condition. Mobile elements represent a source of genetic material that could have important implications for the improvement of drought-tolerance in species, as demonstrated by previous studies (Figueroa-Yañez *et al.*, 2016).

3.4.6 Gene co-expression networks and transcriptional hubs

The advent of transcriptome profiling experiments have increased the complexity of biological datasets and their interpretation. Unlocking the immense potential of transcriptome data requires the use of new system-level analysis to reveal meaningful relationships between genes and biological processes, as well as the regulatory mechanisms which control specific responses (Serin *et al.*, 2016). Currently, the use of biological networks has become a popular and useful approach for depicting the complex organization of biological systems and deciphering the intricate relationships among genes. This network approach has been applied to gene co-expression matrices from transcriptome data to provide early insights into the functional regulation in a spatiotemporal manner for specific phenomena (Oldham *et al.*, 2006). Some efforts have been made to optimize these network methods and build optimal GCN from transcriptomes data (Huang *et al.*, 2017; Marbach *et al.*, 2012). We constructed two independently GCNs for leaves and roots of papaya plant to infer gene responses and relationships during drought, and gain insights into their intricate regulation during stress. The “TF-stress responsive genes” co-expression networks were built on the basis of hub genes (Allen *et al.*, 2012), which control abiotic stress responses. A similar strategy was successfully applied on *Xerophyta viscosa* transcriptome to identify key genes in drought response (Costa *et al.*, 2017). Through such analysis, the

authors found that orthologues of the seed maturation regulators ABI3 and ABI5 played a key role in drought tolerance in vegetative tissues. Furthermore, they suggest that this desiccation tolerance trait in the vegetative tissues of *X. viscosa* comes from a desiccation-tolerance seed character (Costa *et al.*, 2017). Our results regarding the GCNs in leaves and roots of papaya plant, demonstrate the potential of this strategy to detect gene modules during drought conditions and to identify gene regulators that play a key role during stress conditions (Supplementary Figures S12 and S13). Besides this, these networks can contain genes of interest for studies about drought responses in tropical plants. We propose a list of 17 candidate TFs that can be used in future work (CRISPR/Cas9 gene edition, knock-out studies, and mutant plant phenotypes analysis) to evaluate their participation in conferring stress tolerance in papaya plant as well as other agro-economically important crops.

3.5 CONCLUSIONS

This study provided pivotal insight into the biological processes regulated under moderate and severe drought stress in papaya plants. Tissue-specific genes under CN condition and in response to drought were identified, and the analysis of their up-stream genomic regions could lead to the development of tissue and drought specific promoters. This study also yielded a list of drought responsive genes and TFs as candidates for future functional analyses in papaya. Furthermore, the analysis of papaya plant transcriptomes by GCNs provided a clear concept of the modules, or gene communities, that interact during drought within a specific tissue and at a specific point in time, and suggest the regulatory mechanisms that the papaya plant utilizes to cope with abiotic stress.

CAPÍTULO IV

SOBRE-EXPRESIÓN DE LOS GENES CPBI-1, CPLFG1 y CPLFG3 DE PAPAYA (*CARICA PAPAYA* L.) EN PLANTAS TRANSGÉNICAS DE TABACO (*NICOTIANA TABACUM* L.) Y SU FUNCIÓN EN LA TOLERANCIA AL ESTRÉS

4.1 INTRODUCCIÓN

Los cultivos agrícolas se ven expuestos a diversos tipos de estrés abiótico, lo cual tiene un impacto negativo en su supervivencia y causa severas pérdidas económicas. Por ejemplo, en 2013, la salinidad afectó a más del 20% de la tierra irrigada globalmente, causando la pérdida de aproximadamente \$27.3 miles de millones de dólares americanos (Qadir *et al.*, 2014); entre 1961 y 2014, el calor y la sequía causaron la pérdida de unos 237 miles de millones de dólares americanos mundialmente (Mehrabi y Ramankutty, 2017). Para hacer frente a esta problemática es necesario primeramente estudiar la función de los genes involucrados en la tolerancia a diversos tipos de estrés abiótico.

La muerte celular programada (PCD por su nombre en inglés) es un proceso de suicidio celular regulado genéticamente que se encarga de la eliminación de células dañadas o no requeridas, manteniendo así la homeostasis. La PCD ocurre tanto en plantas como animales en respuesta a estímulos internos y ambientales. En las plantas, la PCD se lleva a cabo durante diferentes etapas del desarrollo (Daneva *et al.*, 2016), interacciones con patógenos (Coll *et al.*, 2011) y en respuesta a estreses ambientales (Petrov *et al.*, 2015), tales como déficit hídrico (Duan *et al.*, 2010), estrés salino (Huh *et al.*, 2002), choque térmico (Li *et al.*, 2012), exposición a ozono (Overmyer *et al.*, 2005) y radiación UV (Danon y Gallois, 1998). Sin embargo, poco se conoce acerca de los mecanismos y genes reguladores de la PCD en plantas, menos aún en el contexto de estrés abiótico.

Por otra parte, se han caracterizado extensamente los tipos y genes reguladores de la PCD en animales, en especial de la apoptosis o PCD tipo I. En vertebrados, los genes de la familia BCL-2 regulan la vía mitocondrial de la apoptosis (también conocida como vía intrínseca de la apoptosis) mediante la interacción entre sus miembros anti-apoptóticos, como Bcl-2 y Bcl-xL, y pro-apoptóticos, como Bax y Bak (Chipuk *et al.*, 2010). Las funciones biológicas de los genes de la familia BCL-2 están conservadas a través de diferentes reinos.

Por ejemplo, la expresión ectópica de genes anti-apoptóticos BCL-2 de mamífero puede inhibir la PCD en levaduras y plantas, y conferir tolerancia al estrés abiótico en las últimas (Kim *et al.*, 2014; Li y Dickman, 2004; Xu *et al.*, 2004). No obstante, no se ha encontrado homólogos de los genes BCL-2 en plantas.

Un análisis funcional de una biblioteca de cDNA de humano en células de levadura condujo a la identificación del gen Bax Inhibitor-1 (HsBI-1), nombrado así debido a su capacidad de inhibir la apoptosis inducida por el gen BAX (Xu y Reed, 1998). La proteína codificada por hBI-1 está compuesta por 237 aminoácidos, contiene 6-7 regiones transmembranales, se localiza en el retículo endoplásmico, e interactúa físicamente con las proteínas BCL-2 y Bcl-xL (Xu y Reed, 1998). A la larga, el gen hBI-1 se convirtió en el miembro fundador de la superfamilia TMBIM (del inglés *Transmembranae Bax Inhibitor Motif containing*) de reguladores de la PCD, la cual está conservada en arqueas, bacterias, hongos, animales, protistas y plantas (Samuel D Gamboa-Tuz *et al.*, 2018). Los miembros de la superfamilia TMBIM contienen un dominio conservado de 6-7 regiones transmembranales (anotado en Pfam como PF01027), y algunos de ellos han sido considerados como anti-PCD, anti-apoptóticos o cito-protectores (Rojas-Rivera y Hetz, 2014; Henke *et al.*, 2011; Gubser *et al.*, 2007). Evidencia creciente sugiere que las proteínas codificadas por los genes TMBIM son canales iónicos que regulan la homeostasis intracelular de calcio en bacterias y eucariontes (Carrara *et al.*, 2015; Chang *et al.*, 2014; de Mattia *et al.*, 2009), incluyendo plantas (Sierla *et al.*, 2019). En eucariontes, las proteínas TMBIM se dividen en las familias *Bax Inhibitor* (BI) y *Lifeguard* (LFG) (Samuel D Gamboa-Tuz *et al.*, 2018).

En humanos, los genes de la familia BI están representados por los genes GHITM/TMBIM5 y el ya mencionado BI-1/TMBIM6 (Samuel D Gamboa-Tuz *et al.*, 2018). Homólogos de HsBI-1, pero no de GHITM, han sido identificados en plantas. Los homólogos de BI-1 en plantas regulan la muerte celular (Kawai-Yamada *et al.*, 2001, 2004; Sanchez *et al.*, 2000; Kawai *et al.*, 1999), concentraciones de calcio (Ihara-Ohori *et al.*, 2007), el metabolismo de lípidos (Nagano *et al.*, 2009, 2012, 2014, 2019; Ishikawa *et al.*, 2015), la autofagia (Xu *et al.*, 2017), la senescencia (Yue *et al.*, 2012), el desarrollo del embrión (Reza *et al.*, 2018), las interacciones planta-microbio (Hernández-López *et al.*, 2018; Eichmann *et al.*, 2010; Babaeizad *et al.*, 2009; Imani *et al.*, 2006; Hückelhoven *et al.*, 2003) y la tolerancia al estrés abiótico (Watanabe y Lam, 2009). La expresión del gen AtBI-1 de *Arabidopsis*

thaliana L.) aumenta en respuesta a choque térmico (Watanabe y Lam, 2006), tunicamicina (Watanabe y Lam, 2008), PEG-8000 (Duan *et al.*, 2010), H₂O₂ (Nagano *et al.*, 2012; Kawai-Yamada *et al.*, 2004) y frío (Nagano *et al.*, 2014). Mutantes de pérdida de función del gen AtBI-1 son más sensibles a tunicamicina, choque térmico, estrés hídrico y estrés oxidativo (Nagano *et al.*, 2019; Duan *et al.*, 2010; Watanabe y Lam, 2006, 2008). La sobre-expresión de AtBI-1 confiere tolerancia contra estrés oxidativo en células BY-2 de tabaco y en células en suspensión de arroz (Ishikawa *et al.*, 2010; Kawai-Yamada *et al.*, 2004), y también confiere tolerancia a la sequía en plantas de caña de azúcar (Ramiro *et al.*, 2016). La sobre-expresión del gen SIBI-1 de tomate (*Solanum lycopersicum* L.) confiere tolerancia contra choque térmico y frío en discos de hoja de tabaco (Chae *et al.*, 2003). La expresión del gen CpBI-1 de chile (*Capsicum annuum* L.) aumenta en respuesta a altas y bajas temperaturas, desecación, salinidad, e inundación; su sobre-expresión confiere tolerancia contra calor, sequía y salinidad en plantas transgénicas de tabaco (Isbat *et al.*, 2009). La expresión del gen TaBI-1 de trigo (*Triticum aestivum* L.) aumenta en respuesta a estrés por calor, y su sobre-expresión ectópica rescata el fenotipo sensible a calor de las mutantes de pérdida de función de Arabidopsis (Lu, Zheng, *et al.*, 2018). Silenciamiento del gen GhBI-1 en algodón (*Gossypium hirsutum* L.) resulta en una tasa de muerte celular acelerada bajo condiciones de estrés salino (Zhang *et al.*, 2018). A pesar de estas evidencias que indica que los homólogos de BI-1 en plantas confieren tolerancia contra varios tipos de estrés abiótico, la sobre-expresión de AtBI-1 o TaBI-1 en Arabidopsis no mejora su tolerancia contra frío o estrés salino (Lu, Yu, *et al.*, 2018; Nagano *et al.*, 2014).

La familia LFG se expandió en eucariontes a partir de una proteína ancestral hace 2,000 millones de años dando lugar a varios duplicados en animales y plantas (Hu *et al.*, 2009). Los humanos poseen cuatro genes codificantes de proteínas LFG: LFG1/TMBIM3, LFG2/TMBIM2, LFG3/TMBIM1 y LFG4/TMBIM4. El gen hLFG4 es necesario para la viabilidad celular e inhibe tanto la vía intrínseca como la vía extrínseca de la apoptosis (Gubser *et al.*, 2007). La proteína hLFG4 se localiza principalmente en el aparato de Golgi, por lo que también se le conoce como GAAP (del inglés *Golgi-antiapoptotic protein*) (Gubser *et al.*, 2007). En plantas, la familia LFG se ha dividido en dos clados principales, LFG I y LFG II (Samuel D Gamboa-Tuz *et al.*, 2018). Arabidopsis y cebada (*Hordeum vulgare* L.) poseen cinco proteínas LFG cada una, AtLFG1-5 (también referidas como AtGAAP1-5) y HvLFGa-e, respectivamente (Weis *et al.*, 2013). La sobreexpresión de HvLFGa en cebada

y de AtLFG1 y AtLFG2 en *arabidopsis* promueve la infección por el moho polvoriento, un hongo biótrofo; silenciamiento, o mutaciones de pérdida de función tienen el efecto contrario (Weis *et al.*, 2013). Además, AtLFG2, también conocido como BIL4, regula positivamente la vía de señalización de brasinoesteroides (BR) (YAMAGAMI *et al.*, 2009). La expresión de los genes AtLFG1 y AtLFG2 aumenta en respuesta a tunicamicina y salinidad, y tienen una función dual en la regulación de la respuesta de proteínas no plegadas (UPR por sus siglas en inglés) y la muerte celular para proteger a las plantas contra el estrés del retículo endoplásmico (Guo *et al.*, 2018). La expresión de TaGAAP, una proteína LFG de trigo, aumenta en respuesta a estrés por calor y es activado directamente por el factor de transcripción TaHsfA6f (Xue *et al.*, 2015). Se requiere mayor información acerca del papel de las proteínas LFG en la tolerancia al estrés en plantas.

Papaya (*Carica papaya* L.) es un cultivo agrícola de importancia comercial. Existen cuatro genes TMBIM en papaya: uno de la familia BI, CpBI-1, y tres de la familia LFG, CpLFG1-3 (Samuel D Gamboa-Tuz *et al.*, 2018; Gamboa-Tuz, 2015). La expresión de los genes CpBI-1, CpLFG1 y CpLFG3 responde a estrés por desecación y choque salino (Gamboa-Tuz, 2015), por lo que es posible que participen en el mecanismo de tolerancia contra estos estreses. En este capítulo (IV) se analiza la función de los genes CpBI-1, CpLFG1 y CpLFG3 en la tolerancia contra el estrés abiótico, particularmente salinidad, mediante su expresión en plantas transgénicas de tabaco. También, se explora brevemente su regulación transcripcional en sequía.

4.2 MATERIALES Y MÉTODOS

4.2.1 Clonación y vectores

Los genes CpBI-1, CpLFG1 y CpLFG3 de papaya (*Carica papaya* L. 'Maradol') fueron clonados y secuenciados previamente en el vector P-GEM (Promega, USA) (Gamboa-Tuz, 2015). A partir de estas clonas, los genes CpBI-1, CpLFG1 y CpLFG3 fueron amplificados por Reacción en Cadena de la Polimerasa (PCR por sus siglas en inglés) con cebadores específicos que además contienen la primera mitad de los sitios de recombinación attB1 y attB2 (Cuadro 4.1). Estos productos de PCR fueron purificados y re-amplificados con cebadores conteniendo la segunda mitad de los adaptadores attB1 y attB2 para obtener los sitios de recombinación completos (Cuadro 4.1). Posteriormente, fueron purificados y clonados mediante la reacción de recombinación BP en el vector donador pDONR221 (Invitrogen, USA), generando así clonas de entrada compatibles con la tecnología Gateway. Finalmente, estos genes fueron sub-clonados mediante la reacción de recombinación LR en el vector binario de destino pK7FWG2,0 (Figura 4.1) (Karimi *et al.*, 2002), generando así las clonas de (sobre-) expresión con las construcciones: 35S::CpBI-1-GFP, 35S::CpLFG1-GFP y 35S::CpLFG3-GFP.

Cuadro 4.1. Cebadores utilizados para amplificar los genes CpBI-1, CpLFG1 y CpLFG3.

Gen	Cebador	5' – 3' Secuencia*	Temperatura (°C)
CpBI-1	attbCpBI-1Fw	AAAAAGCAGGCTTCACCATGGATGCTTTCTCTTCTTC	59
	attbCpBI-1Rv	AGAAAGCTGGGTATCTCTTCTTTCTTCTTCTCTC	
CpLFG1	attbCpLFG1Fw	AAAAAGCAGGCTTCACCATGTGGGCTGTACCGCCTA	63
	attbCpLFG1Rv	AGAAAGCTGGGTGCTTTCAGCAGCTCTGAAATATTG	
CpLFG3	attbCpLFG3Fw	AAAAAGCAGGCTTCACCATGAGAAGGCGTACGGAT	62
	attbCpLFG3Rv	AGAAAGCTGGGTATTGTCCGTCTGCCTTAGC	
Adaptadores	Adapter attB1	GGGGACAAGTTTGTACAAAAAAGCAC	60
	Adapter attB2	GGGGACCACTTTGTACAAGAAAGTTC	

*En negrita, los sitios de recombinación attB1 y attB2; en azul, la secuencia Kozak; en rojo, el codón de inicio.

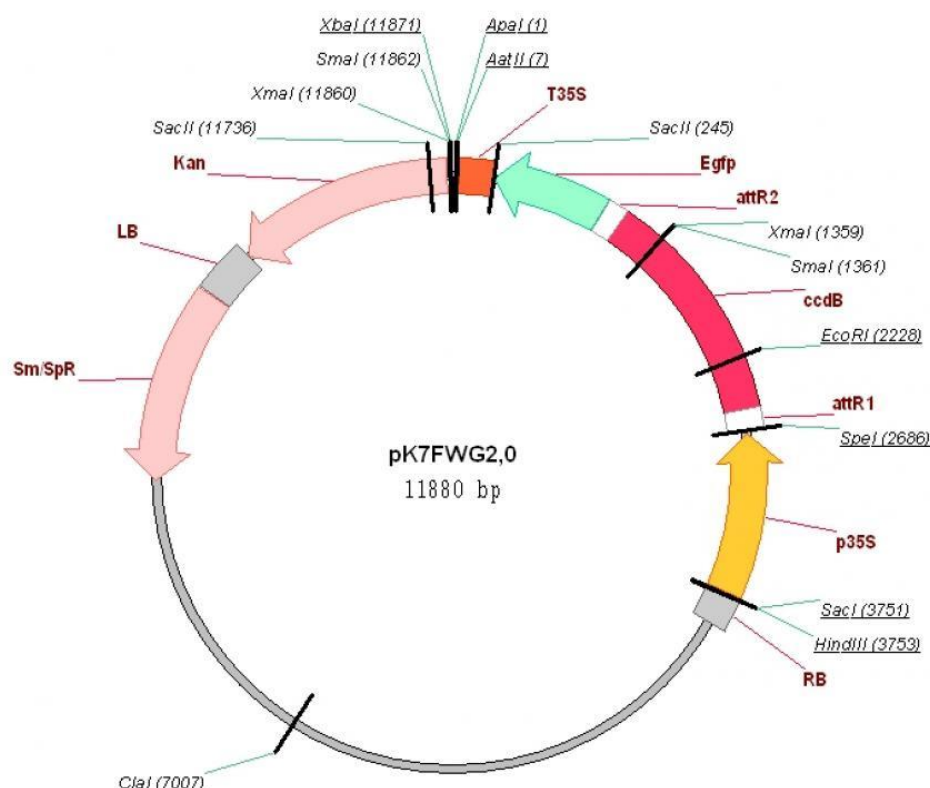


Figura 4.1. Mapa del vector Binario pK7FWG2,0. Sm/SpR: gen de resistencia a estreptomicina y espectinomicina para selección de bacterias transformadas; LB: borde izquierdo (en inglés, *left border*); Kan: gen de resistencia a kanamicina para selección de plantas transformadas; T35S: terminador 35S; Egfp: Proteína verde fluorescente; attR2: sitio de recombinación 2 para la reacción LR; ccdB: Gen que codifica para la toxina ccdB; attR1: sitio de recombinación 1 para la reacción LR; p35S: promotor 35S; RB: borde derecho (en inglés, *right border*).

4.2.2 Transformación genética de tabaco mediante *Agrobacterium*

Las construcciones de expresión en el vector pK7FWG2,0 (descritas en la sección 4.2.1) fueron introducidas en células de *Agrobacterium tumefaciens* cepa EHA105 empleando un protocolo estándar de transformación genética de *Agrobacterium* mediante CaCl_2 .

Para la transformación genética de tabaco por *Agrobacterium* se empleó la metodología descrita por (Clemente, 2006) con algunas modificaciones. Se describe el protocolo de manera general (de requerirse, se puede referir a la cita original para detalles y composición de medios de cultivo empleados):

- Preparación de explantes: Se obtuvieron explantes a partir de hojas de plantas de tabaco (*Nicotiana tabacum* L. 'Petite Havana SR1') crecidas *in vitro* en medio de mantenimiento. Estos explantes fueron adaptados previamente durante 24 horas en medio de pre-cultivo
- Preparación de inóculo de *Agrobacterium*: Las células de *Agrobacterium* transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP fueron crecidas en medio YEP líquido con antibióticos (100 mg/l rifampicina y 50 mg/l espectinomicina) hasta alcanzar una OD₆₀₀ entre 0.5 y 1.0. Las células fueron cosechadas por centrifugación y re-suspendidas en medio de co-cultivo líquido suplementado con 100 µM de acetosiringona (a la misma OD₆₀₀).
- Transformación de tabaco: Los explantes fueron transferidos a medio de co-cultivo semisólido, donde fueron inoculados con las células de *Agrobacterium* transformadas con alguna de las clonas de expresión (descritas en el párrafo anterior). La infección fue complementada con dos rondas de tres minutos de infiltración al vacío a 40 cm Hg con un descanso de 5 minutos entre rondas. Se eliminó el exceso de bacterias y los explantes fueron incubados en el mismo medio de co-cultivo semi-sólido durante tres días para que se llevara a cabo el proceso de infección (transformación genética).
- Selección y regeneración de explantes transformados: Los explantes fueron lavados con una solución salina (0.9% NaCl) con antibióticos (200 mg/l timentina y 200 mg/L cefotaxima) y enjuagados tres veces con agua destilada. Después, los explantes fueron colocados en medio de selección con 150 mg/l de kanamicina. Los explantes que generaron brotes (aproximadamente 1 mes después) fueron transferidos a un medio de enraizamiento. Las plantas regeneradas totalmente fueron plantadas en sustrato Sunshine 3/ Peat Moss/Vermiculita/Hortipearl 2:1:1:1 w/w/w/w. Estas plantas regeneradas son consideradas la generación T0.

4.2.3 Confirmación de la transformación por PCR

Para corroborar la presencia de las construcciones en las plantas, se extrajo el ADN de las plantas de tabaco y se confirmó mediante PCR la presencia de los genes TMBIM de papaya

(CpBI-1, CpLFG1 o CpLFG3) y del gen nptII (que confiere resistencia a kanamicina; representado en el mapa del vector en la Figura 4.1):

- **Extracción de ADN:** 200 mg de tejido de hoja de tabaco fueron macerados en nitrógeno líquido y re-suspendidos en 600 µl de búfer TELT (50 mM Tris, pH 8; 62.5 mM EDTA, pH 8; 2.5 M LiCl; 4% Tritón x 100) en un tubo eppendorf de 1.5 ml. A la muestra se le añadieron 600 µl de fenol:cloroformo:isoamílico (25:24:1); se mezcló vigorosamente con vortex por 15 segundos y se centrifugó a 13,000 rpm por 10 minutos. La fase acuosa (sobrenadante) se transfirió a un tubo eppendorf de 1.5 ml nuevo y se le añadió 1 volumen de isopropanol para precipitar el ADN (en forma de pastilla); se dejó incubando durante 20 minutos a temperatura ambiente y se centrifugó a 13,000 rpm por 15 minutos. Se descartó el sobrenadante y la pastilla se lavó con 1 ml de etanol al 70%; se centrifugó a 7,500 rpm por 5 minutos. Se descartó el sobrenadante y, después de secar por 15 minutos, la pastilla se re-suspendió en 100 µl de agua libre de nucleasas. La integridad del ADN extraído se analizó mediante electroforesis en gel de agarosa al 1% (corrida a 95 mV por 40 minutos) y visualizados con bromuro de etidio.
- **PCR:** Condiciones de reacción: 1x PCR buffer, 2.5 mM MgCl₂, 10 mM dNTP's, 0.2 µM de cada cebador (Cuadro 4.2), 2 µl de ADN (~ 100 ng), 1 unidad de GoTaq Polimerasa (Promega). Condiciones del termociclador: 1 ciclo de desnaturalización a 95°C por 10 minutos; 30 ciclos de: desnaturalización a 95°C por 30 segundos, hibridación (ver Cuadro 4.2) por 30 segundos, extensión a 72°C por 1 minuto; 1 ciclo de extensión a 72°C por 5 minutos. Los productos amplificados fueron analizados por electroforesis en gel de agarosa al 1% (corrida a 95 mV por 40 minutos) y visualizados con bromuro de etidio.

Cuadro 4.2. Cebadores usados para amplificar los genes CpBI-1, CpLFG1 y CpLFG2, y nptII.

Target gene	Primer name	5' – 3' sequence	Temp (°C)
CpBI-1	CpBI-1FW	ATGGATGCTTTCTTCTTTTC	47
	CpBI-1Rv	ATCTCTTCTTTCTTCTTCTCT	

CpLFG1	CpLFG2Fw	ATGGGAAAACACGATGTG	54
	CpLFG2Rv	CGCATCAACTACCCTTAG	
CpLFG3	CpLFG3Fw	ATGGAGAAGGCGTACGGAT	52
	CpLFG3Rv	ATTGTCCGTCTGCCTTAGC	
NptII	NptIIFw	ATGATTGAACAAGATGGATTGC	56
	NptIIRv	TCAGAAGAAGCTCGTCAAGAAGG	

4.2.4 Selección de las generaciones T1 y T2 de plantas transgénicas de tabaco

Para obtener semillas de la generación T1, las plantas regeneradas que fueron positivas a la confirmación por PCR (descrita en sección 4.2.3) fueron auto-polinizadas cubriendo las inflorescencias con bolsas de plástico o papel. Cuando estuvieron maduras, las cápsulas fueron colectadas y se dejaron secar en una incubadora a 36°C por al menos 3 semanas. Las semillas fueron colectadas de las cápsulas secas en tubos eppendorf (1 cápsula por tubo) y almacenadas a 4°C hasta su uso.

Lavado de semillas: Las semillas (colectadas en tubos eppendorf) fueron lavadas con 1 ml de etanol al 95% por 1 minuto. Se descartó el etanol y las semillas fueron lavadas con 1 ml de cloro comercial al 50% con 50 µl/50 ml de tween-80 por 15 minutos. Se descartó el cloro y las semillas fueron lavadas tres veces con agua destilada estéril. Las semillas se dejaron en agua a 4°C y oscuridad por cuatro días.

Después del lavado, las semillas fueron germinadas en medio MS (Murashige and Skooge) suplementado con 150 mg/l de kanamicina para seleccionar a las plantas transgénicas de la siguiente generación, T1.

El procedimiento de extracción de ADN y confirmación del transgen por PCR (descrito en la sección 4.2.3) y el germinado de semillas en medio MS con kanamicina (descrito en esta sección, 4.2.4) se repitió para obtener semillas de la generación T2.

4.2.5 Tratamientos de estrés

Semillas WT y de la generación T2 de las líneas transgénicas 35S:CpBI-1-GFP (07), 35S:CpLFG1-GFP (14) y 35S:CpLFG3-GFP (02) fueron lavadas con el mismo procedimiento descrito en la sección anterior (4.2.4).

Germinación en NaCl: Las semillas fueron germinadas en medio MS o MS + 175 mM NaCl. Se tomaron fotografías 45 días después de iniciado el tratamiento.

Aclimatación de plantas a medio hidropónico: Las semillas de tabaco WT fueron germinadas en medio MS y las semillas de plantas de tabaco transgénicas (35S:CpBI-1-GFP, 35S:CpLFG1 y 35S:CpLFG3) fueron germinadas en medio MS + 150 mg/l de kanamicina (para garantizar que sean transgénicas). Después de la germinación, las plantas fueron transferidas en un sustrato compuesto de Sunshine 3, Peat Moss, Hortipearl y Vermiculita en proporción 2:1:1:1 w/w/w/w, donde fueron regadas de manera regular y fertilizadas con ½ Hoagland cada 15 días. Se permitió a las plantas crecer durante uno o dos meses. Estas plantas fueron transferidas a un medio hidropónico con ½ Hoagland y aeración constante, y se les permitió adaptarse durante 10 días.

Estas plantas se utilizaron para dos ensayos:

1. Tratamiento de choque salino a varias concentraciones de NaCl: Plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP de un mes de edad (crecidas en medio hidropónico) fueron transferidas a un medio hidropónico con ½ Hoagland + diferentes concentraciones de NaCl—0, 150, 225 y 300 mM—durante diez días. Se tomaron parámetros fisiológicos y fotografías.
2. Tratamiento de recuperación después de choque salino de 175 mM NaCl: Plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP de dos meses de edad (crecidas en medio hidropónico) fueron transferidas a un medio hidropónico con ½ Hoagland + 175 mM NaCl durante ocho días (D8). Terminado este tiempo, las plantas fueron transferidas a ½ Hoagland (sin NaCl) para su recuperación durante 6 (R6) y 12 (R12) días. Se tomaron parámetros fisiológicos y fotografías.

4.2.6 Medición de parámetros fisiológicos

La tasa fotosintética neta se midió con un analizador infrarrojo de gases (IRGA) LI-6400/XT (LI-COR). La fluorescencia de la clorofila (F_v/F_m) y el índice de rendimiento absoluto (PI_{abs}) se midieron con un equipo M-PEA (Hansatech Instruments).

4.2.7 Medición de contenido de prolina

Se maceraron 250 mg de tejido de hoja con nitrógeno líquido hasta obtener un polvo fino. El macerado se re-suspendió en 5 ml de ácido sulfosalicílico al 3%. La muestra se filtró con papel filtro Whatmann No. 1. Un ml del filtrado se puso a reaccionar con 1 ml de ninhidrina y 1 ml de ácido acético glacial en baño maría a 95°C durante una hora. La reacción se detuvo en agua fría. La muestra se mezcló con 2 ml de tolueno y se agitó vigorosamente en vortex; se dejó reposar la muestra hasta que se separó en dos fases (superior e inferior). Se tomó 1 ml de la fase superior y se midió su absorbancia a 520 nm con luz visible. El contenido de prolina se determinó con una curva de calibración y se expresó en nanomoles por gramo de peso fresco (nmol g⁻¹ FW).

4.2.8 Visualización de datos y análisis estadísticos

Todos los datos fueron visualizados y analizados en R (R Development Core Team y R Core Team, 2017). Para encontrar si existían diferencias entre medias de las plantas de tabaco WT y transgénicas por tratamiento, se utilizaron las pruebas no paramétricas de Kruskal-Wallis (significancia de $p \leq 0.05$) y Dunn (significancia de $p \leq 0.05$).

4.2.9 Expresión de los genes TMBIM de papaya en el transcriptoma de papaya

Se obtuvieron los datos de expresión—TPM (transcritos por millón) y valores de fold change, P-value y FDR—de los cuatro genes TMBIM de papaya—CpBI-1, CpLFG1, CpLFG2 y CpFLFG3—a partir de los datos de transcriptoma de papaya (capítulo III) (Samuel David Gamboa-Tuz *et al.*, 2018). Los valores de TPM fueron visualizados en un mapa de calor usando el software ComplexHeatmap v2.2.0 (Gu *et al.*, 2016).

4.3 RESULTADOS

4.3.1 Plantas transgénicas de tabaco transformadas con los genes CpBI-1, CpLFG1 y CpLFG3 de papaya (*Carica papaya* L.)

Para estudiar la función de los genes TMBIM de papaya (*Carica papaya* L. 'Maradol') —CpBI-1, CpLFG1 y CpLFG3— se generaron plantas transgénicas de tabaco transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP mediante *Agrobacterium*. Las plantas de tabaco no transformadas se denominaron como WT (del inglés *wild type*). En cada generación de plantas transgénicas —T0, T1 y T2— se extrajo el ADN y se confirmó la presencia del gen nptII (que confiere resistencia a kanamicina) y de los genes TMBIM de papaya mediante PCR con cebadores específicos (se muestran ejemplos de plantas positivas en la Figura 4.2), y se confirmó que las plantas fueran resistentes a kanamicina (150 mg/ml).

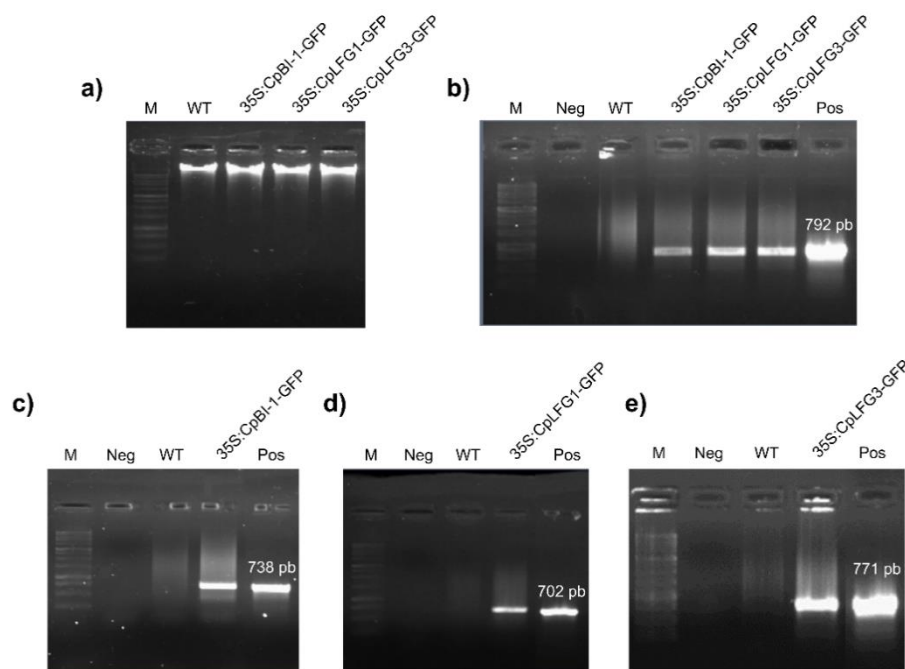


Figura 4.2. Confirmación de la transformación genética de tabaco con las construcciones conteniendo los genes TMBIM de papaya. Se purificó el ADN genómico de plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP (a), y se confirmó la

presencia del gen *nptII* (b) y de los genes CpBI-1 (c), CpLFG1 (d) y CpLFG3 (e) de papaya (*Carica papaya* L.) mediante PCR con cebadores específicos. Carriles: M = Marcador 1 kb ladder (Promega); Neg = Control negativo; Pos = Control positivo; WT = Planta de tabaco WT; 35S:CpBI-1-GFP = Planta de tabaco transgénica transformada con el gen CpBI-1; 35S:CpLFG1-GFP = Planta de tabaco transgénica transformada con el gen CpLFG1-GFP, 35S:CpLFG3-GFP = Planta de tabaco transgénica transformada con el gen CpLFG3. El tamaño del amplicón esperado es indicado arriba del control positivo (b-e). Imagen representativa.

Se obtuvo un total de cuatro (07, 08, 09, 11), cinco (03, 04, 06, 11, 14) y cinco (02, 15, 21A, 20, 21B) líneas de plantas de tabaco transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP, respectivamente, que fueron positivas a PCR (tanto para el gen *nptII* como para el gen TMBM específico) y resistentes a kanamicina (Figura 4.3).

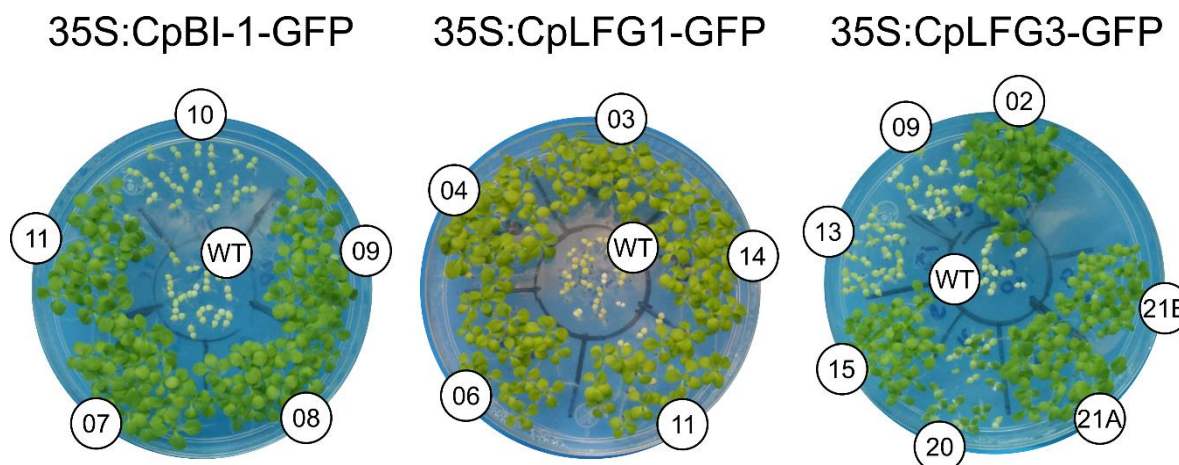


Figura 4.3. Líneas de plantas transgénicas de tabaco generación T1. Plantas de tabaco WT y diferentes líneas transgénicas de la generación T1 transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP germinadas en medio MS + 150 mg/l kanamicina. Diferentes números indican diferentes líneas.

En las siguientes secciones (4.3.2, 4.3.3, 4.3.4; pruebas de tolerancia al estrés salino) solo se reportan los resultados de plantas de la generación T2 de las líneas 35S:CpBI-1-GFP 07, 35S:CpLFG1-GFP 14 y 35S:CpLFG3-GFP 02, por ser las líneas que mostraron los fenotipos más constantes y marcados.

4.3.2 Germinación de semillas de tabaco transgénicas en medio con NaCl

Se probó la tolerancia de las plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP mediante un ensayo de germinación en medio salino (MS + 175 mM NaCl). Todas las plantas—tanto WT como transgénicas—fueron capaces de germinar en 175 mM NaCl, aunque la germinación estuvo retardada (datos no mostrados). Después de 45 días de tratamiento, las plantas WT no fueron capaces de mantener el crecimiento de sus raíces (Figura 4.4). Por otro lado, las tres líneas de tabaco transgénicas transformadas con los genes CpBI-1, CpLFG1 y CpLFG3 presentaron un claro crecimiento de raíces primarias a los 45 días en medio salino (Figura 4.4). Además, presentaron mayor crecimiento en la parte aérea de la planta (Figura 4.4).

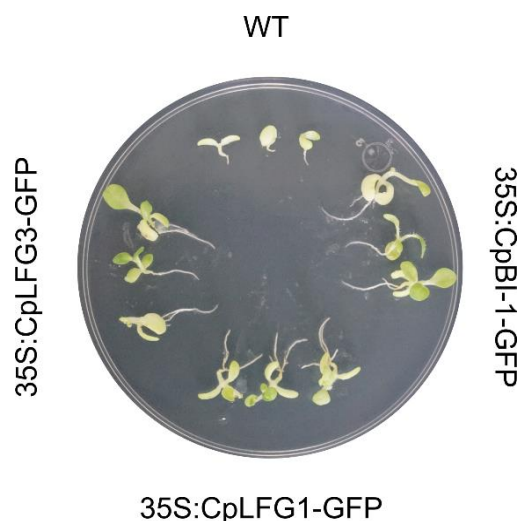


Figura 4.4. Ensayo de germinación en 175 mM NaCl. Plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1 y 35S:CpLFG3 germinadas en medio MS + 175 mM NaCl. Fotografía tomada a los 45 días después de la germinación. Imagen representativa.

4.3.3 Tolerancia de plantas transgénicas de tabaco en diferentes concentraciones de NaCl

Las plantas de tabaco WT y transgénicas (35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP) de un mes de edad fueron adaptadas en un medio hidropónico por una

semana, y sometidas a diferentes concentraciones de NaCl—0, 150, 225 y 300 mM—durante 10 días (Figura 4.5).

En medio con 0 mM NaCl, todas las plantas, tanto WT como las tres líneas transgénicas, presentaron un crecimiento similar (Figuras 4.5a). Los valores de fotosíntesis fueron $2.5 \pm 0.4SE$, $1.7 \pm 0.7SE$, $3.0 \pm 0.3SE$ y $3.4 \pm 0.5SE$ (Figura 4.6), y los valores de prolina fueron $0.4 \pm 0.0SE$, $3.9 \pm 1.1SE$, $0.65 \pm 0.0SE$ y $0.59 \pm 0.1SE$ (Figura 4.7) para las plantas WT, 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP, respectivamente.

En medio con 150 mM NaCl, las plantas WT y las líneas 35S:CpBI-1-GFP y 35S:CpLFG3-GFP presentaron un daño visible similar, caracterizado por la pérdida de turgencia en las hojas (Figura 4.5B). Sin embargo, las plantas transgénicas 35S:CpBI-1-GFP y 35S:CpLFG3-GFP mantuvieron valores más altos de tasa fotosintética, $1.3 \pm 0.0SE$ y $0.9 \pm 0.1SE$, que las plantas WT, $0.3 \pm 0.1SE$ (Figura 4.6). Por otro lado, las plantas transgénicas 35S:CpLFG1-GFP presentaron daño más severo que las otras líneas, demostrado por síntomas más severos de daño (Figura 4.5B), una marcada reducción de la tasa fotosintética, $0.02 \pm 0.0SE$ (Figura 4.6) y aumento de contenido de prolina, $29.58 \pm 4.7SE$ (Figura 4.7).

En medio con 225 y 300 mM, todas las plantas—WT y las tres líneas transgénicas—presentaron severos daños, caracterizados por pérdida de turgencia y clorosis (Figura 4.5C). Las plantas sufrieron una marcada reducción de la fotosíntesis que fluctuó entre -0.1 y 0.1 (Figura 4.6).

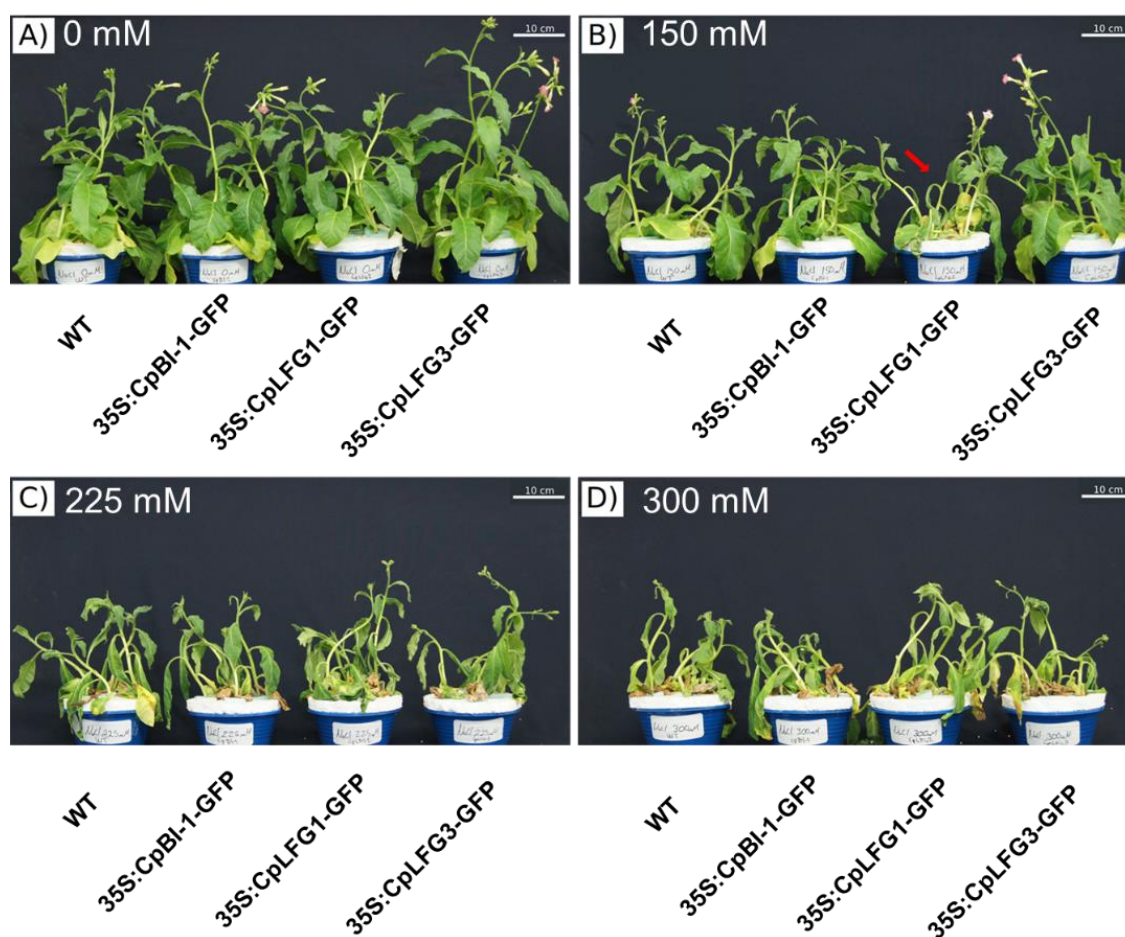


Figura 4.5. Plantas transgénicas de tabaco sometidas a diferentes concentraciones de NaCl. Plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP sometidas a A) 0 mM, B) 150 mM, C) 225 mM y D) 300 mM NaCl durante 10 días. La flecha roja indica el mayor daño en la línea 35S:CpLFG1-GFP a 150 mM NaCl después de 10 días de tratamiento.

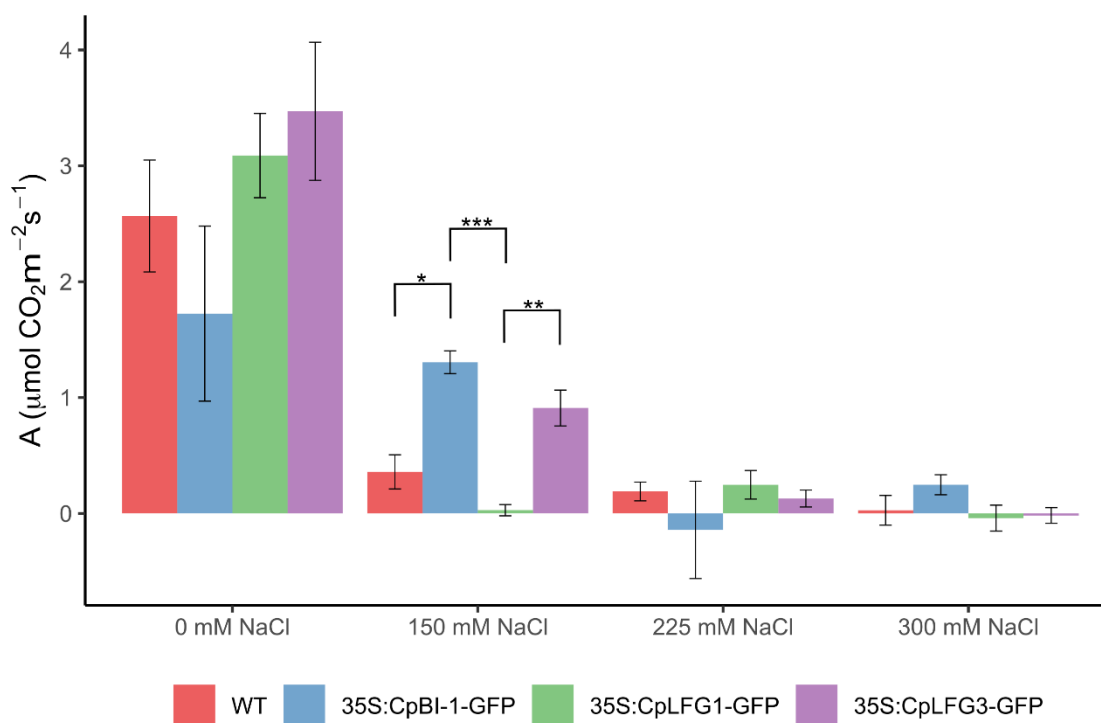


Figura 4.6. Datos de fotosíntesis de las plantas transgénicas de tabaco sometidas a diferentes concentraciones de NaCl. Plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP. Las barras indican la media y las barras de error, el error estándar (SE). Significancia (P-value): * ≤ 0.05 , ** ≤ 0.01 , *** ≤ 0.001 .

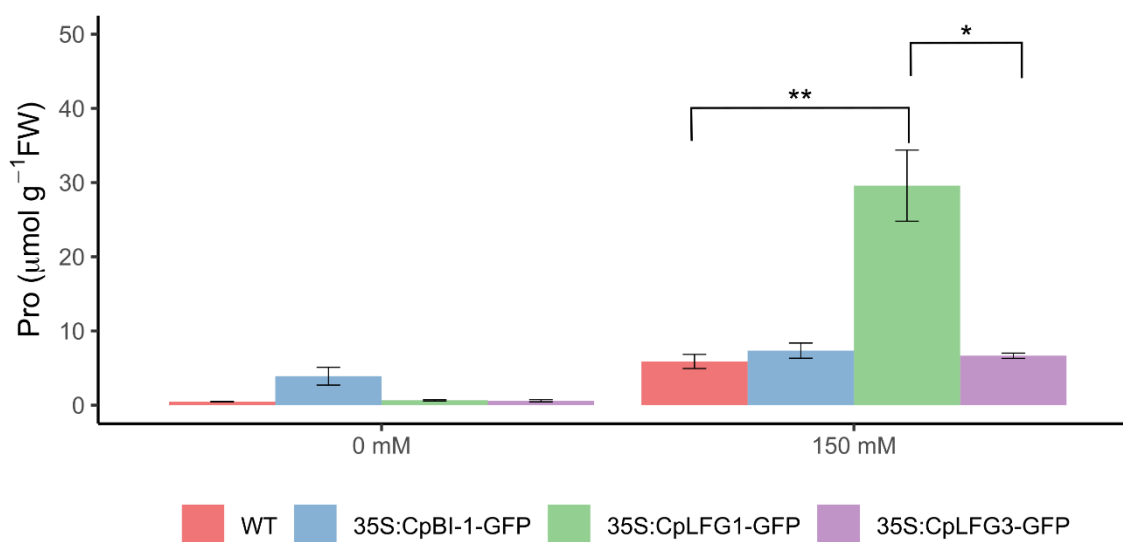


Figura 4.7. Datos de contenido de prolina de las plantas transgénicas de tabaco sometidas a diferentes concentraciones de NaCl. Plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP. Pro = contenido de prolina. Las barras indican la media y las barras de error, el error estándar (SE). Significancia (P-value): * ≤ 0.05 , ** ≤ 0.01 , *** ≤ 0.001 .

4.3.4 Recuperación de tabacos transgénicos después de choque salino

Plantas de tabaco WT y transgénicas (líneas 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP) de dos meses de edad fueron adaptadas en un medio hidropónico por 10 días y sometidas a un choque salino de 175 mM por ocho días (D8). Después del choque salino, las plantas (WT y transgénicas) fueron transferidas a un medio hidropónico sin NaCl para que se recuperaran por seis (R6) y doce (R12) días (Figura 4.8).

Al inicio del tratamiento (D0), las plantas WT y transgénicas (35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP) presentaron crecimiento y fenotipo similares (Figura 4.8a). La tasa fotosintética (A) de las plantas WT y las líneas transgénicas 35S:CpBI-1-GFP y 35S:CpLFG3-GFP fueron similares: $4.5 \pm 0.5\text{SE}$, $5.0 \pm 0.7\text{SE}$ y $4.1 \pm 0.4\text{SE}$, respectivamente (Figura 4.9a), pero la tasa fotosintética de la línea 35S:CpLFG1-GFP fue menor: $2.0 \pm 0.7\text{SE}$ (Figura 4.9a). La eficiencia de la fotosíntesis (F_v/F_m) fue similar entre todas las plantas, aproximadamente 0.8 (Figura 4.9b). El rendimiento de la fotosíntesis (PI abs) se mantuvo entre 4.3 y 5.9 entre todas las plantas WT y transgénicas (Figura 4.9c).

A los ocho días de choque salino (D8), todas las plantas presentaron daños visibles, pérdida de turgencia y clorosis, pero las plantas de la línea 35S:CpLFG1-GFP presentaban mejor apariencia que todas las demás plantas (Figura 4.8a). La línea 35S:CpLFG3-GFP presentaba mejor fenotipo que las plantas WT y que la línea 35S:CpBI-1-GFP; estas dos últimas fueron las más dañadas (Figura 4.8a). No se tomaron mediciones fisiológicas en este punto (D8) del experimento.

A los seis días de recuperación (R6), las plantas WT y la línea transgénica 35S:CpBI1-GFP presentaron severos daños en comparación con las líneas 35S:CpLFG1-GFP y 35S:CpLFG3-GFP; la línea 35S:CpLFG1-GFP fue la que presentó menor daño visible (Figura 4.8a). Las tasas de fotosintética (A) de las líneas 35S:CpBI-1-GFP y CpLFG1-GFP, $-1.89 \pm 0.5SE$ y $-1.74 \pm 0.4SE$, fueron menores que las de las plantas WT y de la línea 35S:CpLFG3-GFP, $-0.5 \pm 0.1SE$ y $-0.3 \pm 0.0SE$, respectivamente (Figura 4.9a). Las plantas WT y las tres líneas transgénicas no presentaron diferencias en la eficiencia del fotosistema II (Fv/Fm), con valores entre 0.67 y 0.8 (Figura 4.9b). El rendimiento fotosintético (PI abs) de la línea 35S:CpLFG1-GFP fue mayor, $3.59 \pm 0.7SE$, que el de las otras líneas. Aunque no fue estadísticamente significativo—WT, $2.22 \pm 0.5SE$; 35S:CpBI-1-GFP, $0.92 \pm 0.2SE$; y 35S:CpLFG3-GFP, $1.32 \pm 0.2SE$ (Figura 4.9c). Interesantemente, los brotes florales de las plantas WT y de la línea 35S:CpBI-1-GFP sufrieron mayor daño en comparación con las líneas 35S:CpLFG1-GFP y 35S:CpLFG3-GFP; algunas plantas de la línea 35S:CpLFG1-GFP incluso presentaron floración (Figura 4.8a,b).

A los doce días de recuperación (R12), las plantas WT y de la línea 35S:CpBI-1-GFP presentaron una tasa de supervivencia del 0%; las plantas de la línea 35S:CpLFG3-GFP, del 25%; y las plantas de la línea 35S:CpLFG1-GFP, del 62% (Figura 4.8c). La tasa fotosintética (A) y el rendimiento fotosintético (PI abs) de las plantas de la línea 35S:CpLFG1-GFP que sobrevivieron en R12 continuaron mejorando en comparación con R6—valores de $-0.01 \pm 0.0SE$ en R12 vs $-1.74 \pm 0.4SE$ en R6 para tasa de fotosíntesis (A), y de $5.08 \pm 1SE$ en R12 vs $3.59 \pm 0.7SE$ en R6 para rendimiento (PI abs; Figura 4.9a,c).

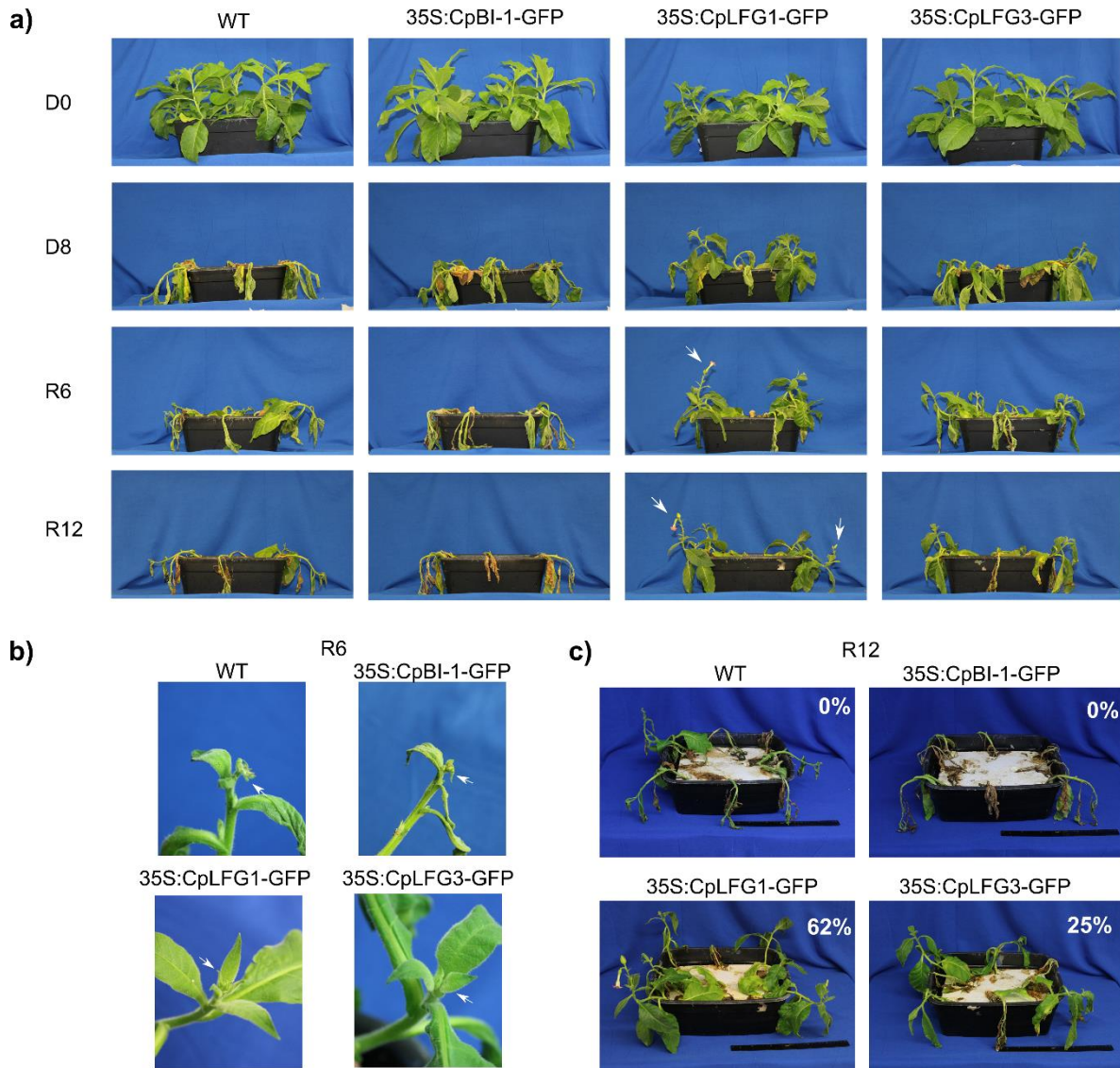


Figura 4.8. Ensayo de recuperación de plantas transgénicas de tabaco sometidas a 175 mM NaCl. a) Plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP sometidas a un choque salino de 175 mM durante 8 días (D8) y dejadas en recuperación durante 6 (R6) y 12 (R12) días. D0 = día inicial del tratamiento. b) Meristemo apical de las plantas mostradas en a) a los seis días de recuperación (R6). c) Foto aérea de las plantas de tabaco mostradas en a) a los 12 días de recuperación (R12); las tasas de supervivencia están indicadas en cada panel.

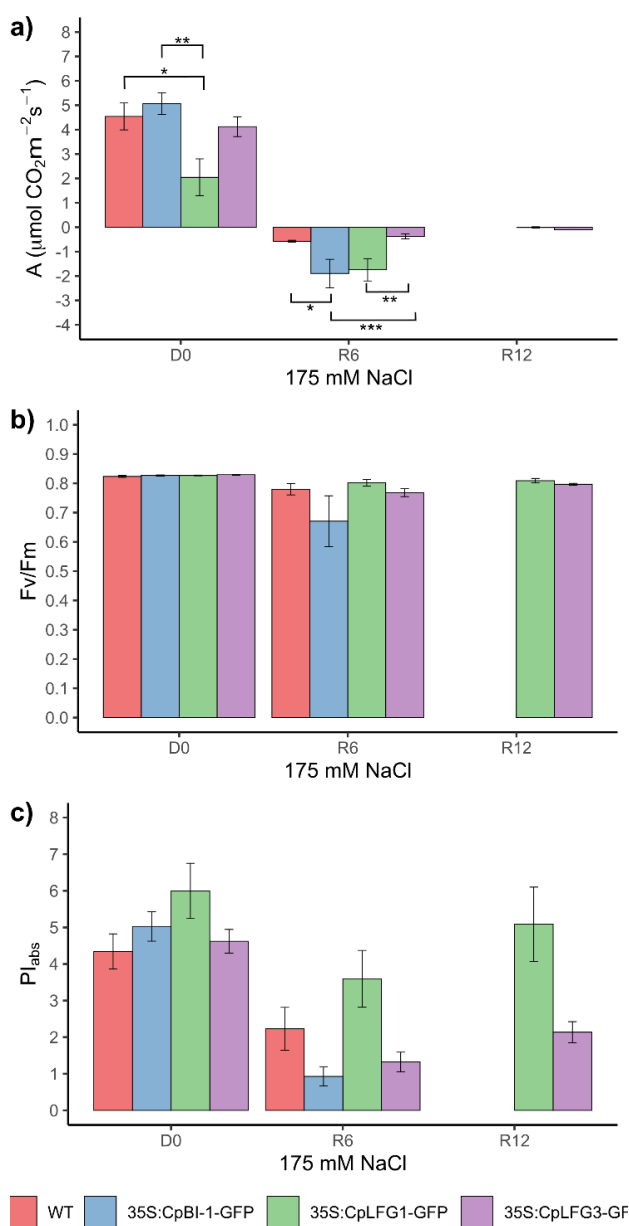


Figura 4.9. Datos de fotosíntesis de las plantas transgénicas de tabaco en ensayo de recuperación después de choque salino a 175 mM. Plantas de tabaco WT y transgénicas transformadas con las construcciones 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP sometidas a un choque salino de 175 mM durante 8 días (D8) y dejadas en recuperación durante 6 (R6) y 12 (R12) días. D0 = día inicial del tratamiento. Las barras indican la media y las barras de error, el error estándar (SE). Significancia (P-value): * ≤ 0.05 , ** ≤ 0.01 , *** ≤ 0.001 .

4.3.5 Expresión de los genes TMBIM de papaya en sequía

Se obtuvieron los datos de expresión de los genes TMBIM del transcriptoma descrito en la sección anterior (Capítulo III). No se encontraron datos de expresión del gen CpLFG1 en ningún tratamiento (Figura 4.10). El único gen expresado diferencialmente (Fold change ≥ 2 , FDR ≤ 0.05), fue el gen CpLFG2 a los 20 días de estrés por sequía en hoja (Cuadro. 4.1). Sin embargo, el gen CpLFG2 no fue analizado mediante plantas transgénicas en el presente trabajo.

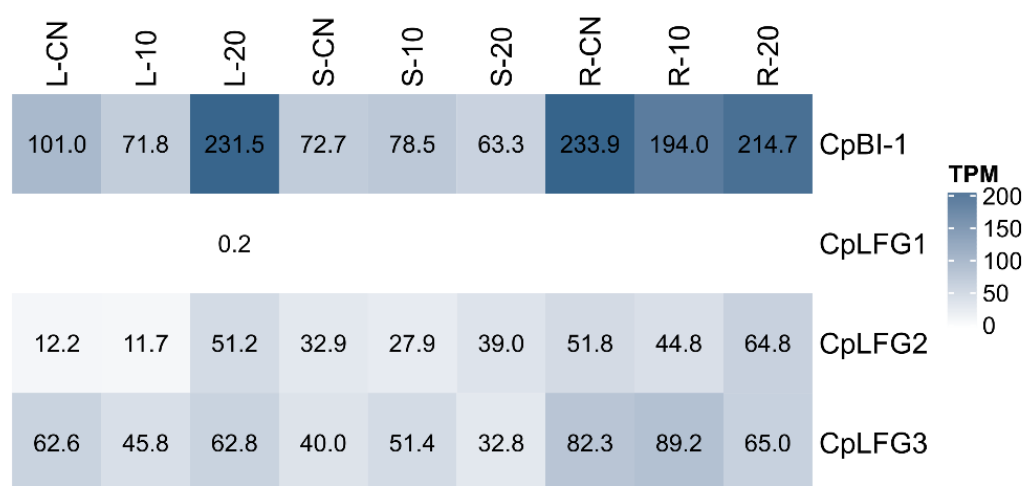


Figura 4.10. Expresión de los genes TMBIM de papaya en condiciones de sequía. L-CN: Control, Hoja; L-10: Sequía 10 días, Hoja; L-20: Sequía 20 días, Hoja; S-CN: Control, Savia; S-10: Sequía 10 días, Savia; S-20: Sequía 20 días, Savia; R-CN: Control, Raíz; R-10: Sequía 10 días, Raíz; R-20: Sequía 20 días, Raíz. Datos tomados de (Samuel David Gamboa-Tuz *et al.*, 2018).

Cuadro 4.3. Datos de expresión de los genes TMBIM de papaya en tratamiento de sequía. No se obtuvieron datos de expresión del gen CpLFG1 (ver figura 4.9). En gris, genes expresados diferencialmente (fold change absoluto ≥ 2 , FDR ≤ 0.05). Datos tomados de (Samuel David Gamboa-Tuz *et al.*, 2018).

Gen	Muestra	Fold Change	P-Value	FDR
CpBI-1	L-10	-1.77301273	1.69E-05	0.00013433
CpBI-1	L-20	1.67046329	0.06314777	0.18418466
CpBI-1	R-10	-1.14561741	0.48250139	0.69133357
CpBI-1	R-20	-1.23008753	0.20975151	0.33986577
CpBI-1	S-10	-1.164706	0.37435616	0.56116742
CpBI-1	S-20	-1.66576147	2.64E-07	1.61E-06
CpLFG2	L-10	-1.31416369	0.07287709	0.16464326
CpLFG2	L-20	3.06310686	0.00026853	0.0023792
CpLFG2	R-10	-1.10152491	0.57531905	0.76457219
CpLFG2	R-20	1.10555109	0.49458716	0.63833307
CpLFG2	S-10	-1.48208353	0.0368442	0.10172008
CpLFG2	S-20	-1.23459264	0.08807006	0.15306353
CpLFG3	L-10	-1.72381563	9.71E-05	0.00064156
CpLFG3	L-20	-1.36555449	0.26626532	0.4864701
CpLFG3	R-10	1.13549605	0.49320305	0.70059021
CpLFG3	R-20	-1.43247582	0.01725251	0.04606395
CpLFG3	S-10	1.01968062	0.90558871	0.98180932
CpLFG3	S-20	-1.76873588	9.68E-09	7.29E-08

4.4 DISCUSIÓN

4.4.1 Aclaración de las relaciones evolutivas de las proteínas LFG de Papaya

De acuerdo con los resultados del capítulo II (Samuel D Gamboa-Tuz *et al.*, 2018), el gen CpBI-1 de papaya pertenece al mismo grupo ortólogo que el gen AtBI-1 de Arabidopsis—OG1. El gen CpLFG1 no pertenece a un grupo ortólogo específico, pero se encuentra en el mismo clado, LFG IIB, que el gen AtLFG1 (el cual pertenece al OG12, específico de Brassicaceae); el gen CpLFG2 —no caracterizado en el presente trabajo— es ortólogo de AtLFG4, ambos en el OG5 dentro del clado LFG IIA; CpLFG3 es ortólogo de AtLFG5—OG4 en el clado LFG I (Figura 4.11).

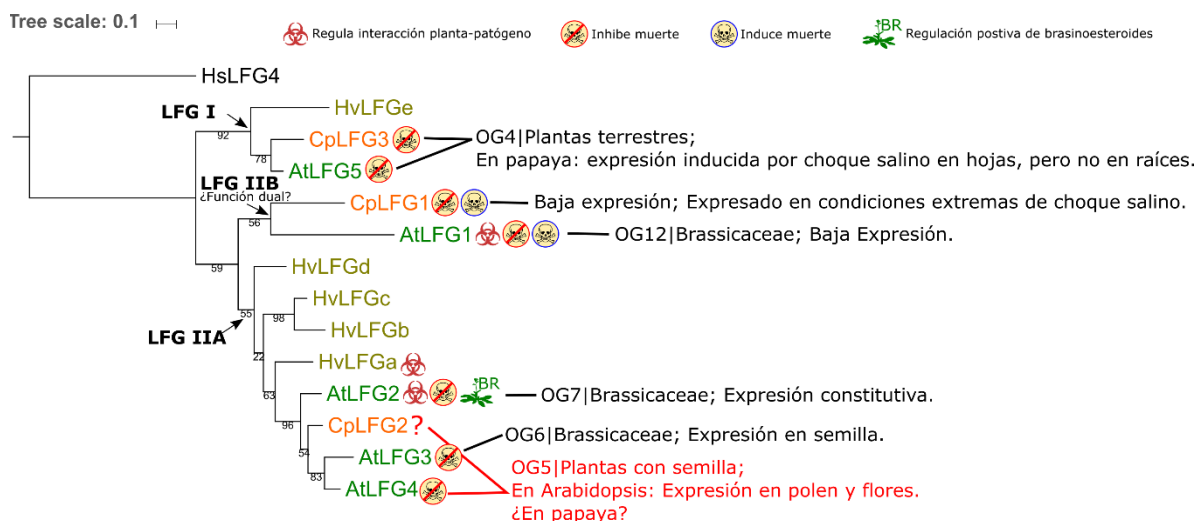


Figura 4.11. Filogenia de las proteínas LFG de Arabidopsis, papaya y cebada. Hs = *Homo sapiens* (humano), Cp = *Carica papaya* (papaya), Hv = *Hordeum vulgare* (cebada), At = *Arabidopsis thaliana* (arabidopsis).

4.4.2 Los genes TMBIM de papaya regulan positivamente la tolerancia contra el estrés salino

En el presente trabajo se caracterizó la función del gen CpBI-1 de la familia BI y de los genes CpLFG1 y CpLFG3 de la familia LFG de papaya en cuanto a su tolerancia al estrés abiótico, particularmente salinidad por NaCl, mediante el uso de plantas transgénicas de tabaco en dos diferentes etapas de crecimiento —germinación y plantas de 1-2 meses de edad— y bajo diferentes condiciones de estrés. Los resultados están resumidos en la Figura 4.12.

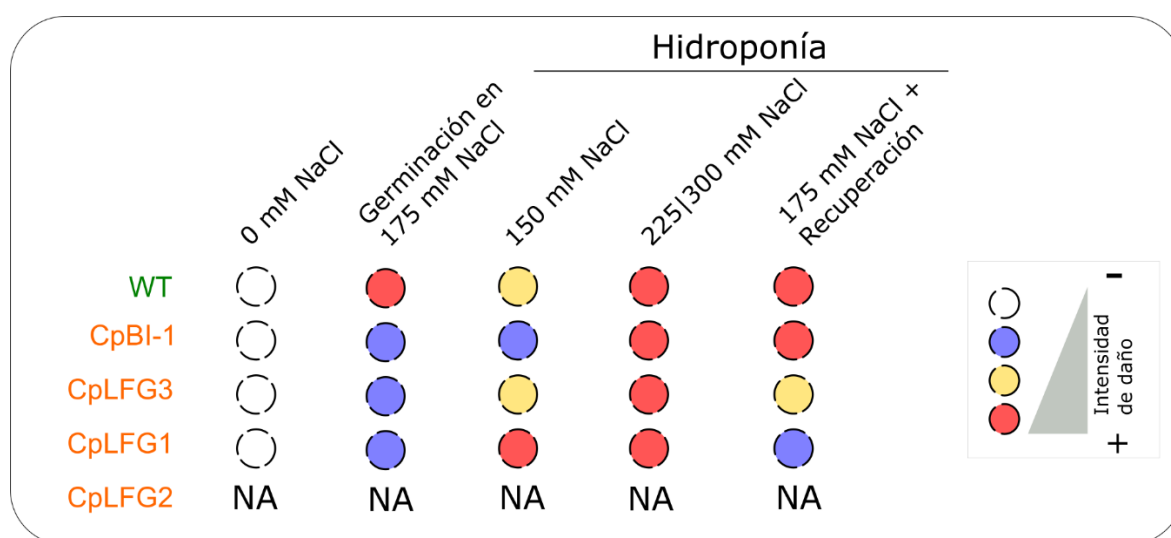


Figura 4.12. Función de los genes TMBIM de papaya en la tolerancia al estrés abiótico en plantas transgénicas de tabaco. a) Esquema resumiendo los resultados de tolerancia de las plantas de tabaco WT y transgénicas 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP. b) La función dual del gen CpLFG1 de papaya en la tolerancia a NaCl podría deberse a la regulación de pH.

El control -sin aplicar algún tratamiento- no presenta diferencias fenotípicas evidentes comparado con las plantas transgénicas (35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP) y las plantas WT en cuanto a germinación (no mostrado) o crecimiento (Figuras 4.5a y 4.6a; Figura 4.12). En otros estudios también se ha mencionado que mutaciones o la sobre-expresión de homólogos de BI-1 en plantas no tiene un efecto sobre su fenotipo en condiciones normales. Por otro lado, Yamagami y cols. (2009) reportaron que la sobre expresión del gen AtLFG2 de Arabidopsis —nombrado como BIL4 en dicho

estudio, el cual es regulador positivo de la vía de los brasinoesteroides— induce un fenotipo enano en condiciones normales (YAMAGAMI *et al.*, 2009). El gen AtLFG2 pertenece al grupo ortólogo OG7, el cual es específico de la familia Brassicaceae dentro del clado LFG IIA; por lo tanto, AtLFG2 no tiene un ortólogo en papaya (Figura 4.11) (Samuel D Gamboa-Tuz *et al.*, 2018). El homólogo más cercano a AtLFG2 en papaya sería CpLFG2, el cual pertenece al grupo ortólogo OG4 AtLFG4 y no fue estudiado en el presente estudio (Figura 4.11). Posiblemente, la sobre-expresión del gen CpLFG2 tendría un efecto similar a AtLFG2 sobre la ruta de brasinoesteroides y estatura de la planta; o tal vez no, ya que, como se mencionó, AtLFG2 pertenece a un grupo específico de la familia Brassicaceae, no Caricaceae. Análisis funcionales de CpLFG2 son requeridos para aclarar esto.

En otros estudios, se ha analizado la función de cuatro genes TMBIM (todos de la familia BI) de plantas en cuanto a su tolerancia al estrés salino—CaBI-1 de chile, GhBI-1 de algodón, TaBI-1 de trigo y AtBI-1 de *Arabidopsis*—con resultados contrastantes. Isbat y cols. (2009) encontraron que plántulas de tabaco transgénicas (de 15 días de edad) transformadas con el gen CaBI1 de chile y germinadas en diferentes concentraciones de NaCl (0, 50, 100, y 200 mM) presentaron mayor longitud de raíz y peso fresco después de 10 días de tratamiento (Isbat *et al.*, 2009). Zhang y cols. (2018) encontraron que el silenciamiento del gen GhBI-1 de algodón reduce la tolerancia de plántulas de esta especie a un tratamiento de 300 mM de NaCl, causando un aumento de muerte celular en hojas y raíces—lo cual indica que el gen GhBI-1 es necesario para la tolerancia contra NaCl en algodón (Zhang *et al.*, 2018). Lu *et al.* (2018) estudiaron el efecto de la sobre-expresión heteróloga del gen TaBI-1 de trigo en el crecimiento de plántulas de *Arabidopsis* (de cuatro días de edad) en condiciones de 100 y 120 mM NaCl; además, estudiaron el efecto del NaCl sobre el crecimiento de la mutante *atbi-1-2*, la cual tiene una mutación no funcional en la región del carboxilo terminal de AtBI-1. En contraste con la función reportada para los genes CaBI-1 (Isbat *et al.*, 2009) y GhBI-1 (Zhang *et al.*, 2018), Lu y cols. (2018) no encontraron diferencias significativas en el crecimiento de hojas y raíces entre las plantas WT, transgénicas (TaBI-1) y mutantes (AtBI-1) a los 13 días después de iniciado el tratamiento (Lu, Yu, *et al.*, 2018). A la fecha, no se ha llevado ningún análisis funcional de los genes LFG (o GAAP) en cuanto a la tolerancia al estrés salino, u otro tipo de estrés abiótico.

Similar a lo encontrado por Isbat y cols. (2009) y Zhang y cols. (2018), en el presente estudio encontramos que la sobre-expresión de los genes CpBI-1, CpLFG1 y CpLFG3 de papaya pueden mejorar la tolerancia de las plantas transgénicas contra NaCl: las plantas de tabaco transgénicas 35S:CpBI-1-GFP, 35S:CpLFG1-GFP y 35S:CpLFG3-GFP presentaron raíces más largas y mayor desarrollo en la parte aérea en comparación con las plantas de tabaco WT después de ser germinadas en 175 mM NaCl (Figura 4.4 y Figura 4.12). Sin embargo, este resultado no fue evidente sino hasta 45 días después del tratamiento.

En humanos los miembros de la superfamilia TMBIM han sido considerados como “integradores de las vías del estrés” debido a que regulan muchas respuestas adaptativas a cambios ambientales, incluyendo PCD, señalización de calcio, regulación de la respuesta no plegada (UPR), bioenergética de la mitocondria, actividad en lisosomas, etc (Rojas-Rivera y Hetz, 2014). Además, son considerados “centinelas celulares” ya que, dependiendo del contexto, pueden tener efectos opuestos en la regulación de la apoptosis y la supervivencia de la célula en respuesta a las mismas perturbaciones biológicas, lo cual refleja su impacto en muchas vías interconectadas (Rojas-Rivera y Hetz, 2014). En plantas, los genes BI-1 han sido los más estudiados de la familia TMBIM, y al parecer también tienen una función de integradores de diversas vías de señalización y centinelas celulares. El estrés salino es un tipo de estrés complejo que involucra un componente osmótico y uno salino (Munns y Tester, 2008), y estreses secundarios como el estrés oxidativo (Abogadallah, 2010) y el estrés del retículo endoplásmico (Liu *et al.*, 2007). Además, está regulado por múltiples genes. Es posible que los genes CBI-1, CpLFG1 y CpLFG3 actúen de manera similar a los genes TMBIM del ser humano, integrando la regulación de varias vías de para proteger a las células contra el estrés salino (Figura 4.13).

A pesar de lo anterior, los genes TMBIM pueden tener un efecto opuesto a la protección, como se discute en la siguiente sección (4.4.3).

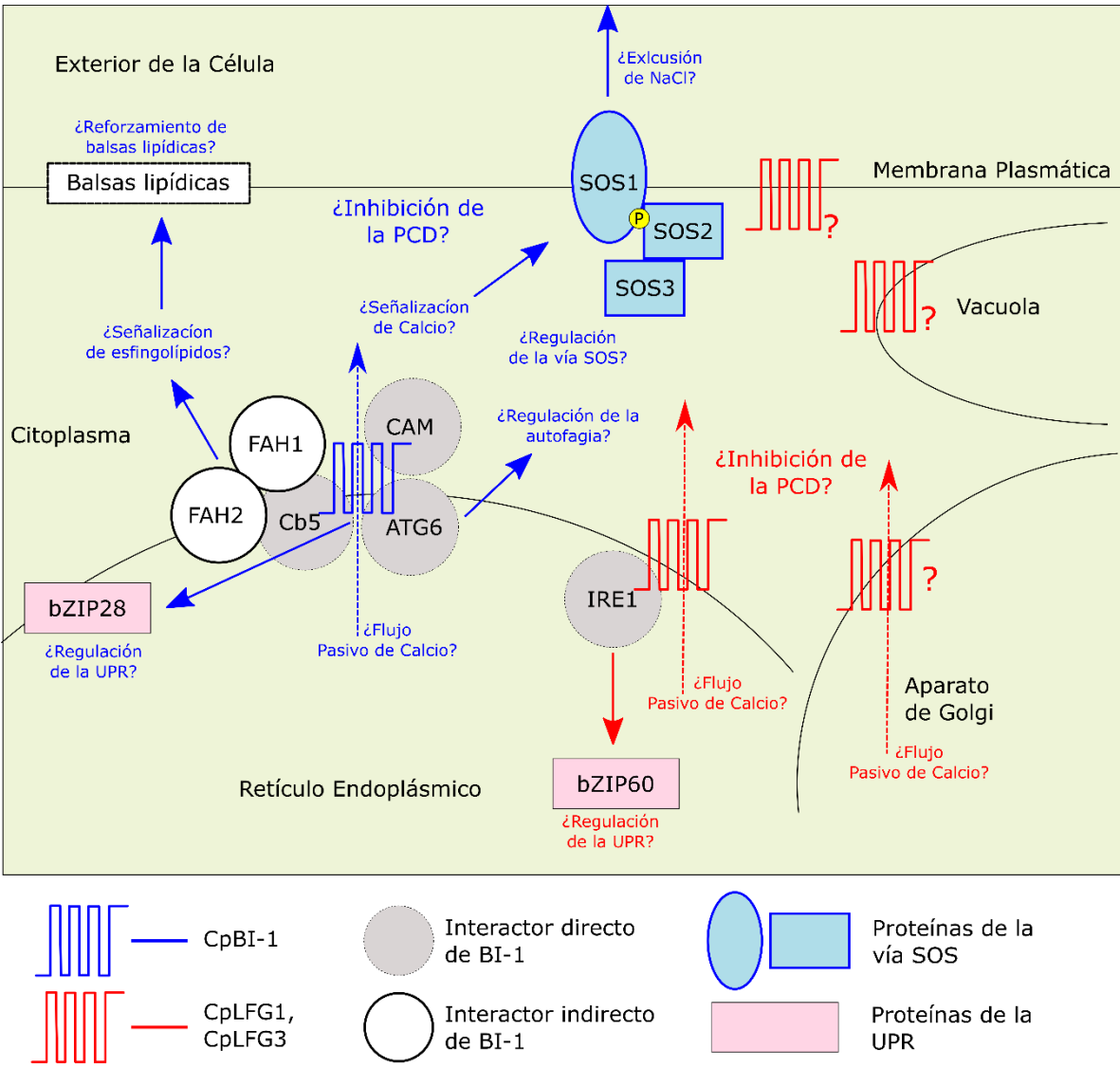


Figura 4.13. Posibles mecanismos controlados por los genes TMBIM de papaya en la tolerancia al estrés salino.

4.4.3 Función dual del gen CpLFG1 en cuanto a la tolerancia al estrés salino

Un resultado no esperado que se obtuvo en el presente trabajo es la función dual del gen CpLFG1 de papaya (Figuras 4.11 y 4.12). En el tratamiento de 150 mM NaCl por 10 días, las plantas transgénicas 35S:CpLFG1-GFP tuvieron síntomas de daño más severos y una marcada reducción en su asimilación de fotosíntesis con respecto a las plantas WT (a pesar de no ser estadísticamente significativa; Figuras 4.5 y 4.6). Además, las plantas 35S:CpLFG1-GFP tuvieron un aumento en su acumulación de prolina con respecto a las plantas WT y las otras líneas transgénicas, soportado por una significancia estadística (Figura 4.7). Por otro lado, en el tratamiento de 175 mM las plantas transgénicas 35S:CpLFG1-GFP tuvieron menos daños visibles, mayor tasa de supervivencia y mejor rendimiento absoluto de fotosíntesis que las plantas WT y las otras líneas transgénicas (Figuras 4.8 y 4.9).

Los genes de la superfamilia TMBIM en eucariontes (tanto de la familia BI como de la familia LFG) han sido considerados como inhibidores de la muerte celular o anti-PCD desde que fueron descubiertos, y se les conoce por conferir protección contra diversos agentes dañinos para la célula (Rojas-Rivera y Hetz, 2014). La función cito-protectora de estos genes está conservada en animales, levaduras y plantas. A pesar de lo anterior, también se ha reportado que estos genes pueden tener el efecto contrario a la protección celular, es decir, pueden inducir la muerte celular en determinadas circunstancias. Esta función parece estar relacionada a un pH bajo: en condiciones ácidas, la sobre-expresión del gen BI-1 humano en células HT1080 incrementa la muerte celular mediante la translocación de la proteína Bax a la mitocondria y liberación citocromo C (Kim *et al.*, 2008). Sobre-expresión del gen Ybh3p de *Saccharomyces cerevisiae* (de la familia LFG, también designado como BXI1p o YNL305Cp), aumenta la muerte celular por la vía mitocondrial inducida por H₂O₂ y ácido acético en células levadura (Büttner *et al.*, 2011). La expresión ectópica del gen AtBI-1 de Arabidopsis en células humanas HT1080 induce la muerte celular de manera similar a Bax, es decir, por la vía mitocondrial de la apoptosis (Li-Hua *et al.*, 2002). Recientemente, se ha reportado que el gen AtLFG1, el cual se encuentra en el mismo clado que el gen CpLFG1 descrito en el presente estudio (Figura 4.11), puede inhibir la muerte celular inducida por Bax en levadura, pero induce la muerte celular cuando es sobre-expresado en tabaco y arabidopsis (Sierla *et al.*, 2019). Esto significa que otros genes TMBIM, tanto de la

familia BI como de la familia LFG, tienen una función dual en la inhibición/inducción de la muerte celular, posiblemente debido al pH.

La proteína BsYetJ de la bacteria *Bacillus subtilis* (la única proteína TMBIM cuya estructura ha sido determinada por cristalografía de rayos X) es un canal de calcio que cambian su conformación y actividad sobre el flujo de calcio de acuerdo al pH (Chang *et al.*, 2014). Esta actividad está regulada por un sensor de pH conformado por dos residuos de aspartato (D) localizados en las regiones transmembranales TM6 y TM7, un residuo de aspartato en cada región TM (Figura 4.14). Posiblemente la actividad de CpLFG1 de papaya también esté regulada de manera similar por el pH. Existe una correlación positiva entre las concentraciones de NaCl (estrés salino) y la alcalinización del apoplasto (GEILFUS y MÜHLING, 2012), es decir aumento de pH bajo un estrés salino más severo. De ser así, es posible que la proteína CpLFG1 induzca la muerte celular a una concentración de 150 mM NaCl cuando el pH es ácido—como sus homólogos remotos HsBI-1 de humano y Ybh3p de levadura—y, por otra parte, proteja a las plantas a una concentración mayor de NaCl, 175 mM, cuando el pH es más alcalino. A los 225 y 300 mM el estrés es demasiado intenso como para que la planta fuera protegida por CpLFG1 (Figura 4.12). Esto explicaría la función dual observada del gen CpLFG1 de papaya en el presente estudio (Figura 4.14).

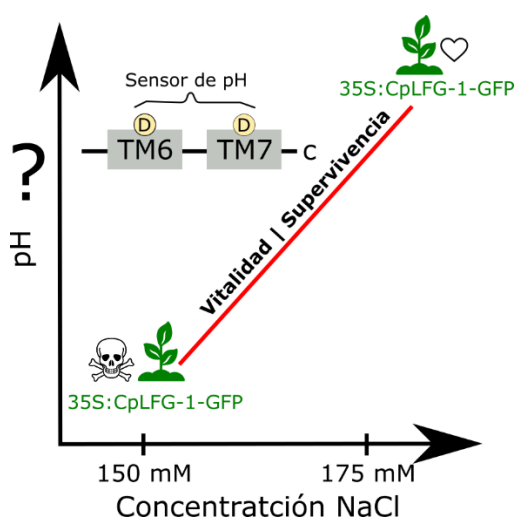


Figura 4.14. Función dual del gen CpLFG1.

Los genes AtLFG1 de arábidopsis y CpLFG1 de papaya (estudiado en el presente estudio), son los únicos genes caracterizados del clado LFG IIB. A pesar de no ser ortólogos y de

que sus secuencias han divergido bastante (Cuadro 4.4), ambos presentan características similares: 1) tienen una función dual en la inhibición e inducción de muerte celular (Sierla *et al.*, 2019) y 2) presentan una muy baja expresión en hojas y raíces (Weis *et al.*, 2013) (Figuras 4.10 y 4.11). Además, el gen CpLFG1 solamente se expresa en condiciones muy estresantes de salinidad (Gamboa-Tuz, 2015). Es posible que todo el clado LFG IIB este compuesto de secuencias que han divergido para obtener nuevas funciones en condiciones extremas de estrés (Figura 4.11). Más análisis de genómica funcional en otras especies son requeridos para aclarar esto.

Cuadro 4.5. Porcentaje de identidad y similitud entre las proteínas TMBIM de papaya y arabidopsis. En amarillo se indica a los ortólogos.

Papaya	Arabidopsis	Identidad (%)	Similitud (%)
CpBI-1	AtBI-1	79	91
CpLFG1	AtLFG1	54	78
CpLFG2	AtLFG4	79	92
CpLFG3	AtLFG5	79	89

4.4.4 Los genes CpBI-1, CpLFG1 y CpLFG3 no están involucrados en la respuesta general de papaya al estrés por sequía

Los genes no actúan solos, sino que interactúan con muchos otros a diferentes niveles. En el capítulo III (Samuel David Gamboa-Tuz *et al.*, 2018) se reportaron varios genes expresados diferencialmente y *hubs* de regulación transcripcional de papaya bajo condiciones de sequía. Sin embargo, los genes TMBIM de papaya analizados en el presente trabajo—CpBI1, CpLFG1 y CpLFG3—no se expresaron diferencialmente. El gen CpLFG2 fue el único que se expresó diferencialmente en hoja a los 20 días de sequía (Figura 4.10 y Cuadro 4.4). Resulta interesante aislar el gen CpLFG2 y analizar su función mediante plantas transgénicas de tabaco en condiciones de sequía. Posiblemente también cumpla una función en el estrés salino, ya que el estrés salino tiene un componente osmótico.

4.5 CONCLUSIONES

1. Los genes CpBI-1, CpLFG1 y CpLFG3 de papaya (*Carica papaya* L. 'Maradol') confieren mayor tolerancia al estrés por NaCl durante la germinación y primeros días de desarrollo.
2. El gen CpLFG2 tiene una función dual: regula negativamente la tolerancia al estrés salino a concentraciones moderadas de NaCl (150 mM), y regula positivamente la tolerancia al estrés salino a concentraciones altas (175 mM). Posiblemente esté en función del pH celular o del medio.
3. Los genes CpBI-1, CpLFG1 y CpLFG3 de papaya no forman parte de la respuesta de papaya a sequía.

CAPÍTULO V

DISCUSIÓN GENERAL, CONCLUSIONES Y PERSPECTIVAS

5.1 DISCUSIÓN GENERAL

Genómica funcional

Gracias al continuo desarrollo y mejoramiento de las tecnologías de secuenciación masiva, las cuales son cada vez más accesibles, al día de hoy se encuentran disponibles los genomas secuenciados completos de diversas especies a través de los tres dominios de la vida: arqueas, bacterias y procariontes. Esto ha dado lugar a gigantescas bases de datos generales y especializadas—NCBI, Phytozome, Ensembl, entre muchas otras—donde se cuenta con las secuencias de miles de genes y otros elementos genéticos. Estas bases de datos continúan creciendo día tras día gracias a proyectos globales ambiciosos como el *Earth BioGenome Project*, el cual persigue la secuenciación del genoma completo de todas las especies en el mundo (Lewin *et al.*, 2018), y a otros numerosos proyectos de secuenciación individuales.

Después de contar con tal colección de información, el siguiente reto es conocer la función biológica de los genes que se encuentran anotados (o por anotar) en dichas bases datos—papel de la genómica funcional. En biología y genética, la palabra “función” suele utilizarse de dos maneras: 1) “efecto seleccionado” y 2) “papel causal”. El efecto seleccionado implica conservación de la función biológica en un contexto evolutivo, i.e. la función y el gen han sido seleccionados por selección natural y evolución. El papel causal hace referencia a asignar una función a un gen por tener una propiedad determinada. El primer caso, efecto seleccionado, es el más pertinente desde un punto de vista biológico (Doolittle y Brunet, 2017; Graur *et al.*, 2013).

Entonces, para conocer la función de un gen o [súper]familia de genes se requiere la integración de análisis bioinformáticos (*dry lab*) y de laboratorio (*wet lab*). Estos análisis bioinformáticos son multidisciplinarios e incluyen, entre muchos otros, el uso de la filogenómica y la transcriptómica para predecir y comprender la evolución y función de diferentes familias génicas y sus interacciones. Además, el análisis bioinformático debe complementarse con estudios de genética reversa, empleando mutantes y/o organismos transgénicos.

En el presente trabajo se utilizaron herramientas bioinformáticas—filogenómica y transcriptómica—y de genética reversa—plantas transgénicas de tabaco—para estudiar la función de la superfamilia TMBIM en la tolerancia al estrés abiótico en plantas, particularmente los genes CpBI-1, CpLFG1 y CpLFG3 de *Carica papaya* L. ‘Maradol’.

Desde la identificación del gen BI-1 del ser humano en 1998 por Xu y Reed, se han reportado numerosas proteínas similares (BI-1-like) en procariontes y eucariontes, conocidas en conjunto como la superfamilia TMBIM. Varios estudios sobre la filogenia de estas proteínas han brindado un panorama general sobre su distribución y evolución (Mariotti *et al.*, 2014; Henke *et al.*, 2011; Hu *et al.*, 2009). Sin embargo, dichos estudios se han enfocado en grupos particulares o con bajo número de muestreos. En el presente trabajo se identificaron 685 proteínas en 171 especies distribuidas en arqueas, bacterias y eucariontes. Los resultados del presente estudio han proporcionado la base para extender el estudio de las proteínas TMBIM en varias direcciones: papel de las TMBIM en arqueas, identificación de motivos cuya función espera ser estudiada, entender la clasificación de las proteínas TMBIM en hongos, y, la aportación más importante del presente trabajo, un detallado análisis de la evolución de las proteínas TMBIM en plantas, lo cual servirá como marco de referencia para comprender las funciones descritas de estas proteínas en plantas.

TMBIM en arqueas, homólogos remotos con diferente tipo de regulación

Este es el primer estudio donde se reportan las proteínas TMBIM en arqueas (Samuel D Gamboa-Tuz *et al.*, 2018). En las proteínas TMBIM de arqueas no se encuentra conservado el sensor di-aspartato (ubicado en las regiones TM6 y TM7) que regula la actividad de canal de calcio de las proteínas TMBIM en bacterias y eucariontes en función de pH (Chang *et al.*, 2014). En las proteínas TMBIM de arqueas el primer aspartato se encuentra conservado (en la TM6), pero no el segundo. En su lugar se encuentran dos triptófanos conservados en todas las secuencias TMBIM de arqueas identificadas (Samuel D Gamboa-Tuz *et al.*, 2018). Este dato sugiere que la actividad de canal de calcio de las proteínas TMBIM en arqueas es regulada de una manera diferente a la actividad de las proteínas TMBIM en bacterias y eucariontes—las cuales se regulan por pH. Análisis de mutaciones puntuales ayudarían a elucidar las posibles funciones de estos triptófanos en la regulación de la actividad de canal de calcio de las proteínas TMBIM en arqueas

El motivo conteniendo el sensor di-aspartato de pH, característica unificadora de las proteínas TMBIM en bacterias y eucariontes

Hu y colaboradores (2009), reportaron 10 motivos característicos para la familia LFG en eucariontes. Al incluir a la familia BI en el presente estudio, se encontraron 60 motivos y 12 arquitecturas que sirven para identificar diferentes clados en toda la superfamilia TMBIM (Samuel D Gamboa-Tuz *et al.*, 2018). Los motivos 7, 8 y 10 reportados por Hu y colaboradores (2009) corresponden a los motivos 1 y 2 encontrados en el presente estudio (Samuel D Gamboa-Tuz *et al.*, 2018). Se encontró que estos motivos no son exclusivos de la familia LFG, sino que igual se encuentran conservados en la familia BI de eucariontes y en bacterias (Samuel D Gamboa-Tuz *et al.*, 2018). Estos motivos contienen al sensor di-aspartato de pH característico de la superfamilia TMBIM en bacterias y eucariontes (Cuadro 5.1).

Cuadro 5.1 Motivos de la familia LFG y de la superfamilia TMBIM.

Motivo 7 LFG	D-T-H-X-[IL]-M-[HKR]-[KR]-[IV]-S-P-E-[DE]-Y-X(6)-L-Y-X-D-[ILV](2)-N-L-F	(Hu <i>et al.</i> , 2009)
Motivo 8 LFG	D-T-Q-X(3)-G-X-[KR]-X(4)-[DNS]-P-E-[DE]-Y-X(6)-[IL]-Y-X-D-[IV](2)-X-[IL]-F	(Hu <i>et al.</i> , 2009)
Motivo 10 LFG	D-T-X(2)-[IL](2)-[KR](2)-X(2)-Y-[DN]-[EQ]-Y-X(6)-L-Y-L-D-X(2)-N-L-F	(Hu <i>et al.</i> , 2009)
Motivo 1 TMBIM	YD[ED]Y[IV]WA[AS][VLI]SL[YF]LD[IFV][IV]N[LI]F	Presente estudio; (Samuel D Gamboa-Tuz <i>et al.</i> , 2018)
Motivo 2 TMBIM	[GA][AL][LI][IVL]F[SC]G[YF]I[VI]YDT[QD]N[LI]I[KE]R[AY]	Presente estudio; (Samuel D Gamboa-Tuz <i>et al.</i> , 2018)

Reclasificación de las proteínas TMBIM en hongos

Las proteínas TMBIM de hongos no han sido analizadas desde un punto de vista evolutivo. Esto ha dado lugar a que se les conozca como homólogos de BI-1, cuando en realidad pertenecen a otros clados y familias de la superfamilia TMBIM. Por ejemplo, el gen MrBI-1 de *Metarhizium robertsii*, el cual brinda protección contra la PCD inducida por BAX y choque térmico (Chen, Duan, Chen, Shang, & Wang, 2015), en realidad pertenece a la familia GHITM; y el SsBI-1 de *Sclerotinia sclerotiorum* (hongo patógeno de plantas), el cual brinda protección contra BAX, choque térmico y H₂O₂ y es necesario para la virulencia de este patógeno (Yu *et al.*, 2015), en realidad pertenece a la familia LFG. Sería interesante encontrar las diferencias en función biológica de estos diferentes clados de proteínas TMBIM en hongos.

Duplicación de las proteínas LFG en plantas, fuente de sub-funcionalización

En vertebrados, la familia LFG se ha dividido en dos grupos: LFG4 y LFG1; el clado LFG1 se ha subdividido dando lugar a los subclados LFG1, LFG2, LFG3 y LFG5 (Hu *et al.*, 2009). Similar a lo ocurrido con la familia LFG en vertebrados, la familia LFG en plantas, incluyendo *Carica papaya*, también han experimentado varias duplicaciones dando lugar a dos grupos principales LFG I y LFG II (Samuel D Gamboa-Tuz *et al.*, 2018). El clado LFG II se subdivide en los clados LFG IIA y IIB (Samuel D Gamboa-Tuz *et al.*, 2018). Arabidopsis y cebada poseen cinco proteínas LFG en sus genomas, AtLFG1-5 y HvLFGa-e (Weis *et al.*, 2013), respectivamente, mientras que papaya posee tres, CpLFG1-3 (Samuel D Gamboa-Tuz *et al.*, 2018; Gamboa-Tuz, 2015). HvLFGe, AtLFG5 y CpLFG3 pertenecen al clado LFG I y al grupo ortólogo OG4, el cual es compartido por todas las plantas terrestres (Samuel D Gamboa-Tuz *et al.*, 2018). CpLFG1 y AtLFG1 pertenecen al subclado LFG IIB, pero CpLFG1 pertenece a un grupo ortólogo presente solamente en Brassicaceae (OG12). CpLFG2 y AtLFG1-3 pertenecen al clado LFG IIA: CpLFG2 y AtLFG4 se encuentran en el mismo grupo ortólogo (OG5) junto con todas las plantas con flores, mientras que AtLFG2 (OG7) y AtLFG3 (OG6) pertenecen a grupos ortólogos específicos de Brassicaceae

La alta conservación de las proteínas BI y LFG en plantas indica que tienen una conservación de función, En el capítulo IV se encontró que los genes CpBI-1, CpLFG1 y CpLFG3 pueden conferir tolerancia a estrés salino en plántulas de transgénicas de tabaco después de la germinación. Sin embargo, las diferentes duplicaciones de la familia LFG (descritas en el párrafo anterior) pueden haber llevado a la sub-funcionalización de las proteínas LFG en plantas, similar a lo sucedido en animales.

CpLFG2 (Clado LFG IIA), regulador de la floración y estrés por sequía

En mamíferos, el gen LFG5 (presente únicamente en mamíferos euterios) se expresa solamente en testículos, indicando que podría tener una función relacionada a la fertilidad y reproducción (Mariotti *et al.*, 2014). De manera similar, el gen AtLFG3 de Arabidopsis se expresa mayormente en semillas, y el gen AtLFG4 se expresa mayormente en flores y polen (Samuel D Gamboa-Tuz *et al.*, 2018). El gen CpLFG2—el cual no fue analizado en el presente estudio—es ortólogo de AtLFG4, así que es posible que también se encuentre expresado mayormente en las flores. Además, el gen CpLFG2 fue el único gen TMBIM de

papaya que fue expresado diferencialmente en respuesta a sequía (Capítulo IV). Sería interesante conocer el papel que este gen CpLFG2 juega en la floración y sequía, y sus posibles aplicaciones biotecnológicas.

CpLFG1 (clado LFG IIB), función dual en la tolerancia al estrés salino

Los únicos genes de plantas del clado LFG IIB que han sido estudiados son AtLFG1 y CpLFG1. Ambos genes presentan niveles de expresión basal muy bajos (Samuel D Gamboa-Tuz *et al.*, 2018; Gamboa-Tuz, 2015; Weis *et al.*, 2013). Esto podría indicar que la expresión basal de los genes LFG del clado LFG IIB en plantas es muy baja (al menos en hoja y raíz), aunque posiblemente se expresen en situaciones de extremas. Por ejemplo, el gen CpLFG1 de papaya se expresó fuertemente en hojas y raíces después de 96 horas en un tratamiento de 300 mM de NaCl (Gamboa-Tuz, 2015). El gen AtLFG1 se expresa en plántulas de Arabidopsis, mayormente en los órganos sexuales (tejido que aún no se analiza en papaya), y en respuesta a tratamiento de tunicamicina y NaCl (Guo *et al.*, 2018). Sin embargo, en condiciones de sequía moderada y severa, el gen CpLFG1 de papaya no pudo ser detectado por el análisis transcriptómico del presente trabajo, así que posiblemente no está implicado en la tolerancia a sequía (Capítulo IV) (Samuel David Gamboa-Tuz *et al.*, 2018). Se requiere más investigación al respecto. Además, mediante el uso de plantas transgénicas de tabaco, se demostró en el presente estudio una función dual del gen CpLFG1 de papaya, y posiblemente de todo el clado LFG IIB (Capítulo IV). El gen CpLFG1 protege contra el estrés salino en condiciones severas de estrés (175 mM), pero induce la muerte de la planta en condiciones moderadas (150 mM). Esta función dual en el estrés salino no se observó en los otros genes TMBIM de papaya estudiados, indicando que es una posible sub-funcionalización del clado LFG IIB en plantas.

CpLFG3 (Clado LFG I)

AtLFG5 y HvLFG6, del clado LFG I, no han sido estudiadas, pero análisis preliminares indican que AtLFG5 puede inhibir la PCD en levadura, aunque en menor medida en comparación con los demás genes LFG de Arabidopsis (Sierla *et al.*, 2019). En el presente estudio se analizó la función del gen CpLFG3 de papaya del clado LFG I. De acuerdo con los resultados del presente trabajo, este gen no parece estar involucrado en la tolerancia al

estrés salino en plantas adultas cultivadas en hidroponía. Se necesitan más análisis funcionales sobre el clado LFG I en plantas para estudiar su función en otros tipos de estrés.

Transcriptómica de papaya en sequía para el estudio de genes de la familia TMBIM

La sequía, causa de déficit hídrico en la planta, es el principal tipo de estrés abiótico que amenaza la producción agrícola en el mundo. El estrés hídrico puede ocasionar la activación de una vía de muerte celular dependiente del retículo endoplásmico en raíces de *arabidopsis*, y el gen AtBI-1 es responsivo a este tipo de estrés (Duan *et al.*, 2010). La sequía también puede inducir vías de muerte celular en las hojas de la planta, ocasionando la fragmentación del ADN nuclear y estrés oxidativo (Hameed *et al.*, 2013). Reza y cols. 2018 analizaron el transcriptoma del embrión de una conífera (*Picea abies*), proceso que involucra la muerte celular programada, y encontraron al gen PaBI-1 entre los genes expresados diferencialmente; análisis de silenciamiento demostraron su papel indispensable en la formación del embrión (Reza *et al.*, 2018). Debido a que la sequía puede inducir la muerte celular en plantas, se esperaba encontrar a los genes TMBIM entre los genes expresados diferencialmente en el transcriptoma de papaya en condiciones de sequía (Capítulo III), y como parte de alguno de los *hubs* de regulación transcripcional identificados o entre los grupos de ontología enriquecidos. Sin embargo, solamente el gen CpLFG2—el cual no fue analizado mediante plantas transgénicas en el presente trabajo—estuvo regulado diferencialmente a los 20 días de estrés (Capítulo IV) y no se encontró en algún *hub* de regulación. Más análisis, con plantas transgénicas, sobre la función del gen CpLFG2 en condiciones de sequía son requeridos.

Relevancia del presente trabajo y perspectivas general

En el presente trabajo se abordó el estudio de la función de las proteínas TMBIM de manera integrativa, haciendo uso de la filogenómica, transcriptómica e ingeniería genética. La principal contribución de este trabajo ha sido proveer un marco de referencia para entender la historia (evolución) de las proteínas TMBIM y entender sus funciones biológicas conservadas (Capítulo II). También, se aportan datos de transcriptoma de papaya en condiciones de sequía, el cual es el primer transcriptoma de esta planta en condiciones de estrés abiótico y sería útil para estudiar no solo las proteínas TMBIM de papaya, sino muchas otras familias de genes que se expresaron diferencialmente en los *hubs* de

regulación encontrados (Capítulo III). Por último, el presente trabajo hace énfasis en la función dual del gen CpLFG1 de papaya, y su posible conservación en el clado LFG IIB (Capítulo IV). Este clado ha divergido más que los otros, por lo que podría dar lugar a muchos tipos de subfuncionalización en diferentes especies.

Ya que los genes de la familia TMBIM son integradores de varias vías de señalización, se necesitan muchos más análisis de los genes TMBIM de papaya para determinar su participación en la autofagia, muerte celular programada, estrés de retículo endoplásmico, respuesta a proteínas no plegadas, regulación de las especies reactivas de oxígeno, señalización de calcio, exclusión de sodio, señalización de esfingolípidos, reforzamiento de las balsas lipídicas, senescencia, y muchos procesos más, y cómo la integración de estas vías contribuye a su función de protección o de inducción de muerte en condiciones de estrés abiótico.

5.2 CONCLUSIONES GENERALES

1. Las arqueas contienen proteínas TMBIM en sus genomas, y es posible que la regulación de su actividad como canales iónicos sea regulada de una manera diferente a las presentes en bacterias y eucariontes.
2. En hongos, las especies de las ramas basales conservan genes de la subfamilia BI, mientras que las especies del subreino Dicaria conservan genes de la subfamilia GHITM. Esto no se había considerado en análisis funcionales de las proteínas TMBIM en hongos.
3. En plantas, las proteínas de la familia BI-1 permanecieron como un solo grupo filogenético y pocas duplicaciones, pero sufrió dos transposiciones mayores en monocotiledóneas.
4. Dos transposiciones mayores ocurrieron en la familia LFG en plantas, pero fue antes de la división de monocotiledóneas y dicotiledóneas. Esta sintenia se ha mantenido conservada a lo largo de la historia evolutiva de las plantas.
5. Múltiples eventos de duplicación han dado lugar a la subfuncionalización de los genes LFG en plantas.
6. Los genes CpBI-1, CpLFG1 y CpLFG3 mejoran el crecimiento de raíces después de la germinación en plantas transgénicas de tabaco.
7. El gen CpLFG1 tiene una función dual en la tolerancia contra choque salino en plantas transgénicas de tabaco adultas: CpLFG1 hace a las plantas más sensibles a un estrés moderado (150 mM), pero aumenta su tolerancia en condiciones más severas (175 mM).
8. Además de tolerancia al estrés, el gen CpLFG1 posiblemente se encuentre involucrado en la regulación de la floración o fertilidad de papaya, lo cual sería de gran interés biotecnológico.

5.3 PERSPECTIVAS

1. Realizar más análisis filogenómicos de la superfamilia TMBIM en procariontes (arqueas y bacterias).
2. Analizar la red de sintenias de la superfamilia TMBIM en invertebrados y vertebrados.
3. Aislar y caracterizar los genes TMBIM de arqueas.
4. Aislar y caracterizar el gen CpLFG2 de papaya.
5. Transformar genéticamente plantas de papaya (intragenesis) para corroborar la función de los genes CpBI-1, CpLFG1 y CpLFG3 en la papaya misma.
6. Identificación de los mecanismos de acción de CpLFG1 mediante más análisis bioquímicos y moleculares (mutaciones puntuales, interacciones, bloqueadores de canales, medición de contenido de sales, etc).
7. Aislar y caracterizar los *hubs* de regulación encontrados en el análisis de las redes de co-expresión a partir de los datos transcriptómicos de papaya bajo condiciones de sequía.
8. Análisis transcriptómicos de papaya bajo otras condiciones de estrés (salinidad, choque de calor, etc.).

BIBLIOGRAFÍA

- Abogadallah, G. M. (2010) Insights into the significance of antioxidative defense under salt stress, *Plant Signaling & Behavior*. Taylor & Francis, 5(4), 369–374.
- Adams, E. y R. Shin (2014) Transport, signaling, and homeostasis of potassium and sodium in plants, *Journal of Integrative Plant Biology*. John Wiley & Sons, Ltd, 56(3), 231–249.
- Aibar, S., C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters y S. Aerts (2017) SCENIC: single-cell regulatory network inference and clustering, *Nature Methods*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 14, p. 1083.
- Allen, J. D., Y. Xie, M. Chen, L. Girard y G. Xiao (2012) Comparing Statistical Methods for Constructing Large Scale Gene Networks, *PLOS ONE*. Public Library of Science, 7(1), p. e29348.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers y D. J. Lipman (1990) Basic local alignment search tool, *Journal of molecular biology*, 215.
- Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller y D. J. Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Research*, 25(17), 3389–3402.
- Anders, S., P. T. Pyl y W. Huber (2015) HTSeq—a Python framework to work with high-throughput sequencing data, *Bioinformatics*, 31(2), 166–169.
- Andrews, S. y others (2010) FastQC: a quality control tool for high throughput sequence data.
- Arve, L. (2011) Stomatal Responses to Drought Stress and Air Humidity, en Torre, S. y Shanker, A. (eds.). Rijeka: InTech, p. Ch. 12.
- Asplund-Samuelsson, J., B. Bergman y J. Larsson (2012) Prokaryotic Caspase Homologs:

-
- Phylogenetic Patterns and Functional Characteristics Reveal Considerable Diversity, PLOS ONE. Public Library of Science, 7(11), p. e49888.
- Babaeizad, V., J. Imani, K. H. Kogel, R. Eichmann y R. Hückelhoven (2009) Over-expression of the cell death regulator BAX inhibitor-1 in barley confers reduced or enhanced susceptibility to distinct fungal pathogens, Theoretical and Applied Genetics, 118(3), 455–463.
- Baek, D., J. Nam, Y. D. koo, D. H. kim, J. Lee, J. C. jeong, S. Kwak, W. S. chung, C. O. lim, J. D. bahk, J. C. hong, S. Y. lee, M. Kawai-yamada, H. Uchimiya y D. Yun (2004) Bax-induced cell death of Arabidopsis mediated through reactive oxygen-dependent and -independent processes, Plant Molecular Biology, 56(1), 15–27.
- Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li y W. S. Noble (2009) MEME SUITE: tools for motif discovery and searching, Nucleic acids research. Oxford Univ Press, p. gkp335.
- Bailey, T. L., C. Elkan y others (1994) Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Department of Computer Science and Engineering, University of California, San Diego.
- Baldoni, E., A. Genga y E. Cominelli (2015) Plant MYB Transcription Factors: Their Role in Drought Response Mechanisms, International Journal of Molecular Sciences, 16(7), 15811–15851.
- Barrera-Figueroa, B. E., L. Gao, Z. Wu, X. Zhou, J. Zhu, H. Jin, R. Liu y J.-K. Zhu (2012) High throughput sequencing reveals novel and abiotic stress-regulated microRNAs in the inflorescences of rice, BMC Plant Biology, 12(1), p. 132.
- Bartels, D. y R. Sunkar (2005) Drought and Salt Tolerance in Plants, Critical Reviews in Plant Sciences. Taylor & Francis, 24(1), 23–58.
- Bastian, M., S. Heymann, M. Jacomy y others (2009) Gephi: an open source software for exploring and manipulating networks., Icwsm, 8, 361–362.
- Bindea, G., B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman,

- F. Pagès, Z. Trajanoski y J. Galon (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, *Bioinformatics*, 25(8), 1091–1093.
- Binzel, M. L., F. D. Hess, R. A. Bressan y P. M. Hasegawa (1988) Intracellular Compartmentation of Ions in Salt Adapted Tobacco Cells, *Plant Physiology*, 86(2), 607 LP – 614.
- Bolduc, N., M. Ouellet, F. Pitre y L. F. Brisson (2003) Molecular characterization of two plant BI-1 homologues which suppress Bax-induced apoptosis in human 293 cells, *Planta*, 216(3), 377–386.
- Bolger, A. M., M. Lohse y B. Usadel (2014) Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, 30(15), 2114–2120.
- Breiman, L. (2001) Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author), *Statist. Sci. The Institute of Mathematical Statistics*, 16(3), 199–231.
- Bultynck, G., S. Kiviluoto, N. Henke, H. Ivanova, L. Schneider, V. Rybalchenko, T. Luyten, K. Nuyts, W. De Borggraeve, I. Bezprozvanny, J. B. Parys, H. De Smedt, L. Missiaen y A. Methner (2012) The C terminus of bax inhibitor-1 forms a Ca²⁺-permeable channel pore, *Journal of Biological Chemistry*, 287(4), 2544–2557.
- Bultynck, G., S. Kiviluoto y A. Methner (2014) Bax Inhibitor-1 Is Likely a pH-Sensitive Calcium Leak Channel, Not a H⁺/Ca²⁺ Exchanger, *Science Signaling*, 7(343), p. pe22 LP-pe22.
- Büttner, S., D. Ruli, F.-N. Vögtle, L. Galluzzi, B. Moitzi, T. Eisenberg, O. Kepp, L. Habernig, D. Carmona-Gutierrez, P. Rockenfeller, P. Laun, M. Breitenbach, C. Khoury, K.-U. Fröhlich, G. Rechberger, C. Meisinger, G. Kroemer y F. Madeo (2011) A yeast BH3-only protein mediates the mitochondrial pathway of apoptosis, *The EMBO Journal*. Nature Publishing Group, 30(14), 2779–2792.
- Campostrini, E. y D. M. Glenn (2007) Ecophysiology of papaya: a review, *Brazilian Journal of Plant Physiology*. scielo, 413–424.

- Carrara, G., N. Saraiva, M. Parsons, B. Byrne, D. L. Prole, C. W. Taylor y G. L. Smith (2015) Golgi Anti-apoptotic Proteins Are Highly Conserved Ion Channels That Affect Apoptosis and Cell Migration, *Journal of Biological Chemistry*, 290(18), 11785–11801.
- Carrara, G., M. Parsons, N. Saraiva y G. L. Smith (2017) Golgi anti-apoptotic protein: a tale of camels, calcium, channels and cancer, *Open Biology*. The Royal Society, 7(5), p. 170045.
- Castillo, K., D. Rojas-Rivera, F. Lisbona, B. Caballero, M. Nassif, F. a Court, S. Schuck, C. Ibar, P. Walter, J. Sierralta, A. Glavic y C. Hetz (2011) BAX inhibitor-1 regulates autophagy by controlling the IRE1 α branch of the unfolded protein response, *The EMBO Journal*, 30(21), 4465–4478.
- Cebulski, J., J. Malouin, N. Pinches, V. Cascio y N. Austriaco (2011) Yeast bax inhibitor, Bxi1p, is an ER-localized protein that links the unfolded protein response and programmed cell death in *Saccharomyces cerevisiae*, *PLoS ONE*, 6(6), 1–7.
- Chae, H. J., N. Ke, H. R. Kim, S. Chen, A. Godzik, M. Dickman y J. C. Reed (2003) Evolutionarily conserved cytoprotection provided by Bax Inhibitor-1 homologs from animals, plants, and yeast, *Gene*, 323(1–2), 101–113.
- Chang, Y., R. Bruni, B. Kloss, Z. Assur, E. Kloppmann, B. Rost, W. a Hendrickson y Q. Liu (2014) Structural basis for a pH-sensitive calcium leak across membranes., *Science* (New York, N.Y.), 344(6188), 1131–5.
- Chaves, M. M., J. Flexas y C. Pinheiro (2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell, *Annals of Botany*, 103(4), 551–560.
- Chen, Y., Z. Duan, P. Chen, Y. Shang y C. Wang (2015) The Bax inhibitor MrBI-1 regulates heat tolerance, apoptotic-like cell death, and virulence in *Metarhizium robertsii*, *Scientific Reports*. Nature Publishing Group, 5(April), p. 10625.
- Chipuk, J. E., T. Moldoveanu, F. Llambi, M. J. Parsons y D. R. Green (2010) The BCL-2 Family Reunion, *Molecular Cell*, 37(3), 299–310.

- Clemente, T. (2006) *Nicotiana* (*Nicotiana tobaccum*, *Nicotiana benthamiana*) BT - *Agrobacterium* Protocols, en Wang, K. (ed.). Totowa, NJ: Humana Press, 143–154.
- Coll, N. S., P. Epple y J. L. Dangi (2011) Programmed cell death in the plant immune system, *Cell Death And Differentiation*. Macmillan Publishers Limited, 18, p. 1247.
- Comas, L., S. Becker, V. M. Cruz, P. F. Byrne y D. A. Dierig (2013) Root traits contributing to plant productivity under drought , *Frontiers in Plant Science* , p. 442.
- Corradi, N. (2015) Microsporidia: eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers, *Annual review of microbiology*. Annual Reviews, 69, 167–183.
- Costa, M.-C. D., M. A. S. Artur, J. Maia, E. Jonkheer, M. F. L. Derks, H. Nijveen, B. Williams, S. G. Mundree, J. M. Jiménez-Gómez, T. Hesselink, E. G. W. M. Schijlen, W. Ligterink, M. J. Oliver, J. M. Farrant y H. W. M. Hilhorst (2017) A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*, *Nature Plants*. Macmillan Publishers Limited, 3, p. 17038.
- Cramer, G. R., K. Urano, S. Delrot, M. Pezzotti y K. Shinozaki (2011) Effects of abiotic stress on plants: a systems biology perspective, *BMC Plant Biology*, 11(1), p. 163.
- Daneva, A., Z. Gao, M. Van Durme y M. K. Nowack (2016) Functions and Regulation of Programmed Cell Death in Plant Development, *Annual Review of Cell and Developmental Biology*. Annual Reviews, 32(1), 441–468.
- Danon, A. y P. Gallois (1998) UV-C radiation induces apoptotic-like changes in *Arabidopsis thaliana*, *FEBS Letters*. John Wiley & Sons, Ltd, 437(1–2), 131–136.
- Darriba, D., G. L. Taboada, R. Doallo y D. Posada (2011) ProtTest 3: fast selection of best-fit models of protein evolution, *Bioinformatics*. Oxford Univ Press, 27(8), 1164–1165.
- Deinlein, U., A. B. Stephan, T. Horie, W. Luo, G. Xu y J. I. Schroeder (2014) Plant salt-tolerance mechanisms, *Trends in Plant Science*, 19(6), 371–379.
- Derényi, I., G. Palla y T. Vicsek (2005) Clique percolation in random networks, *Physical*

- review letters. APS, 94(16), p. 160202.
- Doolittle, W. F. y T. D. P. Brunet (2017) On causal roles and selected effects: our genome is mostly junk, BMC Biology, 15(1), p. 116.
- Duan, Y., W. Zhang, B. Li, Y. Wang, K. Li, Sodmergen, C. Han, Y. Zhang y X. Li (2010) An endoplasmic reticulum response pathway mediates programmed cell death of root tip induced by water stress in Arabidopsis, New Phytol, 186(3), 681–695.
- Eddy, S. R. (1998) Profile hidden Markov models., Bioinformatics, 14(9), p. 755.
- Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research , 32(5), 1792–1797.
- Eichmann, R., M. Bischof, C. Weis, J. Shaw, C. Lacomme, P. Schweizer, D. Duchkov, G. Hensel, J. Kumlehn y R. Hückelhoven (2010) BAX INHIBITOR-1 Is Required for Full Susceptibility of Barley to Powdery Mildew, Molecular Plant-Microbe Interactions. Scientific Societies, 23(9), 1217–1227.
- Fabi, J. P., L. R. B. C. Mendes, F. M. Lajolo y J. R. O. do Nascimento (2010) Transcript profiling of papaya fruit reveals differentially expressed genes associated with fruit ripening, Plant Science, 179(3), 225–233.
- Fang, J., A. Lin, W. Qiu, H. Cai, M. Umar, R. Chen y R. Ming (2016) Transcriptome Profiling Revealed Stress-Induced and Disease Resistance Genes Up-Regulated in PRSV Resistant Transgenic Papaya, Frontiers in Plant Science, 7, p. 855.
- FAOSTAT (2018) *No Title*, <http://www.fao.org/faostat/en/#data>.
- Fawcett, J. A., Y. de Peer y S. Maere (2013) Significance and Biological Consequences of Polyploidization in Land Plant Evolution, en Greilhuber, J., Dolezel, J., y Wendel, J. F. (eds.) Plant Genome Diversity Volume 2: Physical Structure, Behaviour and Evolution of Plant Genomes. Vienna: Springer Vienna, 277–293.
- Figuerola-Yañez, L., A. Pereira-Santana, A. Arroyo-Herrera, U. Rodríguez-Corona, F. Sanchez-Teyer, J. Espadas-Alcocer, F. Espadas-Gil, F. Barredo-Pool, E. Castaño y

- L. C. Rodriguez-Zapata (2016) RAP2.4a Is Transported through the Phloem to Regulate Cold and Heat Tolerance in Papaya Tree (*Carica papaya* cv. Maradol): Implications for Protection Against Abiotic Stress, PLOS ONE. Public Library of Science, 11(10), p. e0165030.
- Flagel, L. E. y J. F. Wendel (2009) Gene duplication and evolutionary novelty in plants, *New Phytologist*. Blackwell Publishing Ltd, 183(3), 557–564.
- Franke, R. y L. Schreiber (2007) Suberin — a biopolyester forming apoplastic plant interfaces, *Current Opinion in Plant Biology*, 10(3), 252–259.
- Gabaldón, T., E. V Koonin, T. Gabaldon y E. V Koonin (2013) Functional and evolutionary implications of gene orthology, *Nat Rev Genet*. Nature Publishing Group, 14(5), 360–366.
- Gall, H. Le, F. Philippe, J.-M. Domon, F. Gillet, J. Pelloux y C. Rayon (2015) Cell Wall Metabolism in Response to Abiotic Stress, *Plants*. Editado por P. Harris. MDPI, 4(1), 112–166.
- Gamboa-Tuz, S. D. (2015) Análisis in silico y expresión diferencial de genes que codifican proteínas con dominio similar la inhibidor de bax 1 de *Carica papaya* en respuesta a salinidad y sequía, Tesis de Doctorado. Centro de Investigación Científica de Yucatán. México. Tesis de Doctorado. Centro de Investigación Científica de Yucatán. México., p. 121.
- Gamboa-Tuz, Samuel D, A. Pereira-Santana, T. Zhao, M. E. Schranz, E. Castano y L. C. Rodriguez-Zapata (2018) New insights into the phylogeny of the TMBIM superfamily across the tree of life: Comparative genomics and syntenic networks reveal independent evolution of the BI and LFG families in plants, *Molecular Phylogenetics and Evolution*, 126, 266–278.
- Gamboa-Tuz, Samuel David, A. Pereira-Santana, J. A. Zamora-Briseño, E. Castano, F. Espadas-Gil, J. T. Ayala-Summano, M. Á. Keb-Llanes, F. Sanchez-Teyer y L. C. Rodríguez-Zapata (2018) Transcriptomics and co-expression networks reveal tissue-specific responses and regulatory hubs under mild and severe drought in

- papaya (*Carica papaya* L.), *Scientific Reports*, 8(1), p. 14539.
- GEILFUS, C.-M. y K. H. MÜHLING (2012) Transient alkalization in the leaf apoplast of *Vicia faba* L. depends on NaCl stress intensity: an in situ ratio imaging study, *Plant, Cell & Environment*. John Wiley & Sons, Ltd, 35(3), 578–587.
- Glasauer, S. M. K. y S. C. F. Neuhauss (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences, *Molecular Genetics and Genomics*, 289(6), 1045–1060.
- Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam y D. S. Rokhsar (2012) Phytozome: a comparative platform for green plant genomics, *Nucleic Acids Research*, 40(D1), p. D1178.
- Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo y A. Conesa (2008) High-throughput functional annotation and data mining with the Blast2GO suite, *Nucleic Acids Research*, 36(10), 3420–3435.
- Grabherr, M. G. *et al.* (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data, *Nature biotechnology*, 29(7), 644–652.
- Graur, D., Y. Zheng, N. Price, R. B. R. Azevedo, R. A. Zufall y E. Elhaik (2013) On the Immortality of Television Sets: “Function” in the Human Genome According to the Evolution-Free Gospel of ENCODE, *Genome Biology and Evolution*, 5(3), 578–590.
- Gu, Z., R. Eils y M. Schlesner (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data, *Bioinformatics*, 32(18), 2847–2849.
- Guan, D., B. Yan, C. Thieme, J. Hua, H. Zhu, K. R. Boheler, Z. Zhao, F. Kragler, Y. Xia y S. Zhang (2017) PlaMoM: a comprehensive database compiles plant mobile macromolecules, *Nucleic Acids Research*, 45(D1), D1021–D1028.
- Gubser, C., D. Bergamaschi, M. Hollinshead, X. Lu, F. J. M. Van Kuppeveld y G. L. Smith (2007) A new inhibitor of apoptosis from vaccinia virus and eukaryotes, *PLoS Pathogens*, 3(2), 0246–0259.

- Guo, K., W. Wang, W. Fan, Z. Wang, M. Zhu, X. Tang, W. Wu, X. Yang, X. Shao, Y. Sun, W. Zhang y X. Li (2018) Arabidopsis GAAP1 and GAAP3 Modulate the Unfolded Protein Response and the Onset of Cell Death in Response to ER Stress, *Frontiers in Plant Science*, 9, p. 348.
- Haas, B. J., A. L. Delcher, J. R. Wortman y S. L. Salzberg (2004) DAGchainer: a tool for mining segmental genome duplications and synteny, *Bioinformatics*, 20(18), 3643–3646.
- Haas, B. J. *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, *Nature Protocols*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 8, p. 1494.
- Hamamoto, S., T. Horie, F. Hauser, U. Deinlein, J. I. Schroeder y N. Uozumi (2015) HKT transporters mediate salt stress resistance in plants: from structure and function to the field, *Current Opinion in Biotechnology*, 32, 113–120.
- Hameed, A., M. Goher y N. Iqbal (2013) Drought induced programmed cell death and associated changes in antioxidants, proteases, and lipid peroxidation in wheat leaves, *Biologia Plantarum*, 57(2), 370–374.
- Henke, N., D. A. Lisak, L. Schneider, J. Habicht, M. Pergande y A. Methner (2011) The ancient cell death suppressor BAX inhibitor-1, *Cell Calcium*. Elsevier Ltd, 50(3), 251–260.
- Hernández-López, A., M. Díaz, J. Rodríguez-López, G. Guillén, F. Sánchez y C. Díaz-Camino (2018) Uncovering Bax inhibitor-1 dual role in the legume–rhizobia symbiosis in common bean roots, *Journal of Experimental Botany*, 70(3), 1049–1061.
- Hofmann, K. y M. D. Baron (1996) Boxshade 3.21, Pretty printing and shading of multiple-alignment files. Kay Hofmann ISREC Bioinformatics Group, Lausanne, Switzerland.
- Hu, L., T. F. Smith y G. Goldberger (2009) LFG: A candidate apoptosis regulatory gene family, *Apoptosis*, 14(11), 1255–1265.

- Huang, D. W., B. T. Sherman y R. A. Lempicki (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Research*, 37(1), 1–13.
- Huang, J., S. Vendramin Alegre, L. Shi, K. M. McGinnis, S. Vendramin, L. Shi y K. M. McGinnis (2017) Construction and Optimization of a Large Gene Coexpression Network in Maize Using RNA-Seq Data, *Plant Physiology*, 175(1), 568 LP – 583.
- Hückelhoven, R., C. Dechert y K.-H. Kogel (2003) Overexpression of barley BAX inhibitor 1 induces breakdown of mlo-mediated penetration resistance to *Blumeria graminis*., *Proceedings of the National Academy of Sciences of the United States of America*, 100(9), 5555–60.
- Huh, G.-H., B. Damsz, T. K. Matsumoto, M. P. Reddy, A. M. Rus, J. I. Ibeas, M. L. Narasimhan, R. A. Bressan y P. M. Hasegawa (2002) Salt causes ion disequilibrium-induced programmed cell death in yeast and plants, *The Plant Journal*. John Wiley & Sons, Ltd (10.1111), 29(5), 649–659.
- Huynh-Thu, V. A., A. Irrthum, L. Wehenkel y P. Geurts (2010) Inferring Regulatory Networks from Expression Data Using Tree-Based Methods, *PLOS ONE*. Public Library of Science, 5(9), p. e12776.
- Ihara-Ohori, Y., M. Nagano, S. Muto, H. Uchimiya y M. Kawai-Yamada (2007) Cell Death Suppressor Arabidopsis Bax Inhibitor-1 Is Associated with Calmodulin Binding and Ion Homeostasis, *Plant Physiology*, 143(2), 650 LP – 660.
- Imani, J., H. Baltruschat, E. Stein, G. Jia, J. Vogelsberg, K. H. Kogel y R. Hückelhoven (2006) Expression of barley BAX Inhibitor-1 in carrots confers resistance to *Botrytis cinerea*, *Molecular Plant Pathology*, 7(4), 279–284.
- Inoue, J., Y. Sato, R. Sinclair, K. Tsukamoto y M. Nishida (2015) Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling, *Proceedings of the National Academy of Sciences*, 112(48), 14918–14923.
- Isbat, M., N. Zeba, S. R. Kim y C. B. Hong (2009) A BAX inhibitor-1 gene in *Capsicum*

- anuum is induced under various abiotic stresses and endows multi-tolerance in transgenic tobacco, *Journal of Plant Physiology*. Elsevier, 166(15), 1685–1693.
- Ishikawa, T., K. Takahara, T. Hirabayashi, H. Matsumura, S. Fujisawa, R. Terauchi, H. Uchimiya y M. Kawai-Yamada (2010) Metabolome analysis of response to oxidative stress in rice suspension cells overexpressing cell death suppressor bax inhibitor-1, *Plant and Cell Physiology*, 51(1), 9–20.
- Ishikawa, T., T. Aki, S. Yanagisawa, H. Uchimiya y M. Kawai-Yamada (2015) Overexpression of BAX INHIBITOR-1 Links Plasma Membrane Microdomain Proteins to Stress, *Plant Physiology*, 169(2), 1333–1343.
- Jaillon, and others (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, *Nature*. Nature Publishing Group, 449(7161), 463–467.
- Jamaluddin, N. D., N. Mohd Noor y H.-H. Goh (2017) Genome-wide transcriptome profiling of *Carica papaya* L. embryogenic callus, *Physiology and Molecular Biology of Plants*. New Delhi: Springer India, 23(2), 357–368.
- Ji, H., J. M. Pardo, G. Batelli, M. J. Van Oosten, R. A. Bressan y X. Li (2013) The Salt Overly Sensitive (SOS) Pathway: Established and Emerging Roles, *Molecular Plant*, 6(2), 275–286.
- Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, Y. Hu, H. Liang, P. S. Soltis, D. E. Soltis, S. W. Clifton, S. E. Schlarbaum, S. C. Schuster, H. Ma, J. Leebens-Mack y C. W. dePamphilis (2011) Ancestral polyploidy in seed plants and angiosperms, *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 473(7345), 97–100.
- Jiao, Y., J. Li, H. Tang y A. H. Paterson (2014) Integrated Syntenic and Phylogenomic Analyses Reveal an Ancient Genome Duplication in Monocots, *The Plant Cell*. American Society of Plant Biologists, 26(7), 2792–2802.
- Jiménez-Ruiz, A., J. F. Alzate, E. T. MacLeod, C. G. K. Lüder, N. Fasel y H. Hurd (2010) Apoptotic markers in protozoan parasites, *Parasites & Vectors*, 3(1), p. 104.

- Jiménez, V. M., E. Mora-Newcomer y M. V Gutiérrez-Soto (2014) Biology of the Papaya Plant BT - Genetics and Genomics of Papaya, en Ming, R. y Moore, P. H. (eds.). New York, NY: Springer New York, 17–33.
- Julia, K. y K. Friedrich (2018) Long distance RNA movement, *New Phytologist*. Wiley/Blackwell (10.1111), 218(1), 29–40.
- Karimi, M., D. Inzé y A. Depicker (2002) GATEWAY™ vectors for Agrobacterium-mediated plant transformation, *Trends in Plant Science*, 7(5), 193–195.
- Kawai-Yamada, M., L. Jin, K. Yoshinaga, A. Hirata y H. Uchimiya (2001) Mammalian Bax-induced plant cell death can be down-regulated by overexpression of Arabidopsis Bax Inhibitor-1 (AtBI-1), *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 98(21), 12295–12300.
- Kawai-Yamada, M., Y. Ohori y H. Uchimiya (2004) Dissection of Arabidopsis Bax inhibitor-1 suppressing Bax-, hydrogen peroxide-, and salicylic acid-induced cell death., *The Plant cell*, 16(1), 21–32.
- Kawai, M., L. Pan, J. C. Reed y H. Uchimiya (1999) Evolutionally conserved plant homologue of the Bax Inhibitor-1 (BI-1) gene capable of suppressing Bax-induced cell death in yeast, *FEBS Letters*, 464(3), 143–147.
- Kazuo, N. y Y. Kazuko (2005) Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants, *Physiologia Plantarum*. Wiley/Blackwell (10.1111), 126(1), 62–71.
- Kehr, J. y A. Buhtz (2008) Long distance transport and movement of RNA through the phloem, *Journal of Experimental Botany*, 59(1), 85–92.
- Kielbasa, S. M., R. Wan, K. Sato, P. Horton y M. C. Frith (2011) Adaptive seeds tame genomic sequence comparison, *Genome Research*. Cold Spring Harbor Laboratory Press, 21(3), 487–493.
- Kim, E. Y., Y. S. Seo, H. Lee y W. T. Kim (2010) Constitutive expression of CaSRP1, a hot pepper small rubber particle protein homolog, resulted in fast growth and improved

- drought tolerance in transgenic Arabidopsis plants, *Planta*, 232(1), 71–83.
- Kim, E. Y., K. Y. Park, Y. S. Seo y W. T. Kim (2016) Arabidopsis Small Rubber Particle Protein Homolog SRPs Play Dual Roles as Positive Factors for Tissue Growth and Development and in Drought Stress Responses, *Plant Physiology*, 170(4), 2494 LP – 2510.
- Kim, H. R., G. H. Lee, K. C. Ha, T. Ahn, J. Y. Moon, B. J. Lee, S. G. Cho, S. Kim, Y. R. Seo, Y. J. Shin, S. W. Chae, J. C. Reed y H. J. Chae (2008) Bax inhibitor-1 is a pH-dependent regulator of Ca²⁺ channel activity in the endoplasmic reticulum, *Journal of Biological Chemistry*, 283(23), 15946–15955.
- Kim, Y., M. Wang, Y. Bai, Z. Zeng, F. Guo, N. Han, H. Bian, J. Wang, J. Pan y M. Zhu (2014) Bcl-2 suppresses activation of VPEs by inhibiting cytosolic Ca²⁺ level with elevated K⁺ efflux in NaCl-induced PCD in rice, *Plant Physiology and Biochemistry*, 80, 168–175.
- Kiviluoto, S., L. Schneider, T. Luyten, T. Vervliet, L. Missiaen, H. De Smedt, J. B. Parys, a Methner y G. Bultynck (2012) Bax Inhibitor-1 is a novel IP3 receptor-interacting and -sensitizing protein, *Cell Death and Disease. Nature Publishing Group*, 3(8), p. e367.
- Koonin, E. V (2005) Orthologs, paralogs, and evolutionary genomics, *Annu Rev Genet*, 39.
- Koonin, E. V y L. Aravind (2002) Origin and evolution of eukaryotic apoptosis: the bacterial connection, *Cell death and differentiation*, 9(4), 394–404.
- Kottenmeier, W., H. Chang, S. M. Siegel y B. Z. Siegel (1983) Stimulation of growth in papaya and other plants by dilute salt solutions, *Water, Air, & Soil Pollution*, 20(4), 447–450.
- Kristensen, D. M., Y. I. Wolf, A. R. Mushegian y E. V Koonin (2011) Computational methods for Gene Orthology inference, *Briefings in Bioinformatics*, 12(5), 379–391.
- Langmead, B. y S. L. Salzberg (2012) Fast gapped-read alignment with Bowtie 2, *Nature Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.*, 9, p. 357.

- Lechner, M., S. Findeiß, L. Steiner, M. Marz, P. F. Stadler y S. J. Prohaska (2011) Proteinortho: Detection of (Co-)orthologs in large-scale analysis, *BMC Bioinformatics*, 12(1), p. 124.
- Lee, G.-H., T. Ahn, D.-S. Kim, S. J. Park, Y. C. Lee, W. H. Yoo, S. J. Jung, J.-S. Yang, S. Kim, A. Muhlrads, Y.-R. Seo, S.-W. Chae, H.-R. Kim y H.-J. Chae (2010) Bax inhibitor 1 increases cell adhesion through actin polymerization: involvement of calcium and actin binding., *Molecular and cellular biology*, 30(7), 1800–1813.
- Letunic, I. (2015) phyloT: Phylogenetic Tree Generator. *Phylot.biobyte.de*, *Nucleic Acids Res. Nature Publishing Group*, 10(10), p. 725.
- Letunic, I. y P. Bork (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, *Nucleic acids research*. Oxford Univ Press, p. gkw290.
- Lewin, H. A. *et al.* (2018) Earth BioGenome Project: Sequencing life for the future of life, *Proceedings of the National Academy of Sciences*, 115(17), 4325 LP – 4333.
- Li-Hua, Y., K.-Y. Maki, N. Mikihiro, W. Kaori, R. J. C y U. Hirofumi (2002) Induction of mammalian cell death by a plant Bax inhibitor, *FEBS Letters*. Wiley-Blackwell, 512(1–3), 308–312.
- Li, B., R. K. Yadav, G. S. Jeong, H.-R. Kim y H.-J. Chae (2014) The Characteristics of Bax Inhibitor-1 and its Related Diseases, *Current Molecular Medicine*, 14(5), 603–615.
- Li, B. y C. N. Dewey (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinformatics*, 12(1), p. 323.
- Li, W. y M. B. Dickman (2004) Abiotic stress induces apoptotic-like features in tobacco that is inhibited by expression of human Bcl-2, *Biotechnology Letters*, 26(2), 87–95.
- Li, Z. y M. Sheng (2012) Caspases in synaptic plasticity, *Molecular Brain*, 5(1), p. 15.
- Li, Z., H. Yue y D. Xing (2012) MAP Kinase 6-mediated activation of vacuolar processing enzyme modulates heat shock-induced programmed cell death in Arabidopsis, *New*

- Phytologist. John Wiley & Sons, Ltd (10.1111), 195(1), 85–96.
- Lin, H., Z. Liao, L. Zhang y Q. Yu (2016) Transcriptome analysis of the male-to-hermaphrodite sex reversal induced by low temperature in papaya, *Tree Genetics & Genomes*, 12(5), p. 94.
- Lisbona, F., D. Rojas-Rivera, P. Thielen, S. Zamorano, D. Todd, F. Martinon, A. Glavic, C. Kress, J. H. Lin, P. Walter, J. C. Reed, L. H. Glimcher y C. Hetz (2009) BAX Inhibitor-1 Is a Negative Regulator of the ER Stress Sensor IRE1??, *Molecular Cell*. Elsevier Ltd, 33(6), 679–691.
- Liu, J.-X., R. Srivastava, P. Che y S. H. Howell (2007) Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling, *The Plant Journal*. John Wiley & Sons, Ltd, 51(5), 897–909.
- Lord, C. E. N. y A. H. L. A. N. Gunawardena (2012) Programmed cell death in *C. elegans*, mammals and plants, *European Journal of Cell Biology*, 91(8), 603–613.
- Lu, P.-P., T.-F. Yu, W.-J. Zheng, M. Chen, Y.-B. Zhou, J. Chen, Y.-Z. Ma, Y.-J. Xi y Z.-S. Xu (2018) The Wheat Bax Inhibitor-1 Protein Interacts with an Aquaporin TaPIP1 and Enhances Disease Resistance in Arabidopsis , *Frontiers in Plant Science* , p. 20.
- Lu, P.-P., W.-J. Zheng, C.-T. Wang, W.-Y. Shi, J.-D. Fu, M. Chen, J. Chen, Y.-B. Zhou, Y.-J. Xi y Z.-S. Xu (2018) Wheat Bax Inhibitor-1 interacts with TaFKBP62 and mediates response to heat stress, *BMC Plant Biology*, 18(1), p. 259.
- Luan, S. (2009) The CBL–CIPK network in plant calcium signaling, *Trends in Plant Science*, 14(1), 37–42.
- Lyons, E., B. Pedersen, J. Kane y M. Freeling (2008) The Value of Nonmodel Genomes and an Example Using SynMap Within CoGe to Dissect the Hexaploidy that Predates the Rosids, *Tropical Plant Biology*, 1(3), 181–190.
- Lyons, E. y M. Freeling (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences, *The Plant Journal*. Wiley Online Library, 53(4), 661–673.

- Madroñero, J., S. P. Rodrigues, T. F. S. Antunes, P. M. V Abreu, J. A. Ventura, A. A. R. Fernandes y P. M. B. Fernandes (2018) Transcriptome analysis provides insights into the delayed sticky disease symptoms in *Carica papaya*, *Plant Cell Reports*, 37(7), 967–980.
- Mahouachi, J., V. Arbona y A. Gómez-Cadenas (2007) Hormonal changes in papaya seedlings subjected to progressive water stress and re-watering, *Plant Growth Regulation*, 53(1), 43–51.
- Mahouachi, J., R. Argamasilla y A. Gómez-Cadenas (2012) Influence of Exogenous Glycine Betaine and Absciscic Acid on Papaya in Responses to Water-deficit Stress, *Journal of Plant Growth Regulation*, 31(1), 1–10.
- Mahouachi, J., A. R. Socorro y M. Talon (2006) Responses of Papaya Seedlings (*Carica papaya* L.) to Water Stress and Re-Hydration: Growth, Photosynthesis and Mineral Nutrient Imbalance, *Plant and Soil*, 281(1), 137–146.
- Marbach, D., J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, T. D. Consortium, M. Kellis, J. J. Collins y G. Stolovitzky (2012) Wisdom of crowds for robust gene network inference, *Nature Methods*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 9, p. 796.
- Marin, B., E. C. M. Nowack, G. Glöckner y M. Melkonian (2007) The ancestor of the *Paulinella* chromatophore obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like α -proteobacterium, *BMC Evolutionary Biology*, 7(1), p. 85.
- Mariotti, M., T. F. Smith, P. H. Sudmant y G. Goldberger (2014) Pseudogenization of testis-specific *Lfg5* predates human/Neanderthal divergence, *Journal of Human Genetics*, 59(5), 288–291.
- Marler, T. E. y M. V Mickelbart (1998) Drought, Leaf Gas Exchange, and Chlorophyll Fluorescence of Field-grown Papaya, *Journal of the American Society for Horticultural Science* jashs. Washington, DC: American Society for Horticultural Science, 123(4), 714–718.

- Martin, L., Z. Fei, J. Giovannoni y J. Rose (2013) Catalyzing plant science research with RNA-seq , *Frontiers in Plant Science* , p. 66.
- de Mattia, F., C. Gubser, M. M. T. van Dommelen, H.-J. Visch, F. Distelmaier, A. Postigo, T. Luyten, J. B. Parys, H. de Smedt, G. L. Smith, P. H. G. M. Willems y F. J. M. van Kuppeveld (2009) Human Golgi Antiapoptotic Protein Modulates Intracellular Calcium Fluxes, *Molecular Biology of the Cell. American Society for Cell Biology (mboc)*, 20(16), 3638–3645.
- Mehrabi, Z. y N. Ramankutty (2017) The cost of heat waves and droughts for global crop production, *bioRxiv*, p. 188151.
- Merchant, S. S. *et al.* (2007) The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions, *Science. American Association for the Advancement of Science*, 318(5848), 245–250.
- Michaelson, L. V, J. A. Napier, D. Molino y J.-D. Faure (2016) Plant sphingolipids: Their importance in cellular organization and adaption, *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1861(9, Part B), 1329–1335.
- Ming, R. *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus), *Nature*, 452(7190), 991–996.
- Ming, R., Q. Yu, P. H. Moore, R. E. Paull, N. J. Chen, M.-L. Wang, Y. J. Zhu, M. A. Schuler, J. Jiang y A. H. Paterson (2012) Genome of papaya, a fast growing tropical fruit tree, *Tree Genetics & Genomes*, 8(3), 445–462.
- El Moussaoui, A., M. Nijs, C. Paul, R. Wintjens, J. Vincentelli, M. Azarkan y Y. Looze (2001) Revisiting the enzymes stored in the laticifers of *Carica papaya* in the context of their possible participation in the plant defence mechanism, *Cellular and Molecular Life Sciences CMLS*, 58(4), 556–570.
- Munns, R. y M. Tester (2008) Mechanisms of Salinity Tolerance, *Annual Review of Plant Biology. Annual Reviews*, 59(1), 651–681.
- Nagano, M., Y. Ihara-Ohori, H. Imai, N. Inada, M. Fujimoto, N. Tsutsumi, H. Uchimiya y M.

- Kawai-Yamada (2009) Functional association of cell death suppressor, Arabidopsis Bax inhibitor-1, with fatty acid 2-hydroxylation through cytochrome b 5, *Plant Journal*, 58(1), 122–134.
- Nagano, M., K. Takahara, M. Fujimoto, N. Tsutsumi, H. Uchimiya y M. Kawai-Yamada (2012) Arabidopsis Sphingolipid Fatty Acid 2-Hydroxylases (AtFAH1 and AtFAH2) Are Functionally Differentiated in Fatty Acid 2-Hydroxylation and Stress Responses, *Plant Physiology*, 159(3), 1138–1148.
- Nagano, M., T. Ishikawa, Y. Ogawa, M. Iwabuchi, A. Nakasone, K. Shimamoto, H. Uchimiya y M. Kawai-Yamada (2014) Arabidopsis Bax inhibitor-1 promotes sphingolipid synthesis during cold stress by interacting with ceramide-modifying enzymes, *Planta*, 240(1), 77–89.
- Nagano, M., C. Kakuta, Y. Fukao, M. Fujiwara, H. Uchimiya y M. Kawai-Yamada (2019) Arabidopsis Bax inhibitor-1 interacts with enzymes related to very-long-chain fatty acid synthesis, *Journal of Plant Research*, 132(1), 131–143.
- Nakashima, K., K. Yamaguchi-Shinozaki y K. Shinozaki (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat, *Frontiers in Plant Science*, 5, p. 170.
- Nakatani, Y., H. Takeda, Y. Kohara y S. Morishita (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates, *Genome Research*. Cold Spring Harbor Laboratory Press, 17(9), 1254–1265.
- Okonechnikov, K., O. Golosova, M. Fursov y others (2012) Unipro UGENE: a unified bioinformatics toolkit, *Bioinformatics*. Oxford Univ Press, 28(8), 1166–1167.
- Oldham, M. C., S. Horvath y D. H. Geschwind (2006) Conservation and evolution of gene coexpression networks in human and chimpanzee brains, *Proceedings of the National Academy of Sciences*, 103(47), 17973 LP – 17978.
- Omid, A., T. Keilin, A. Glass, D. Leshkowitz y S. Wolf (2007) Characterization of phloem-sap transcription profile in melon plants, *Journal of Experimental Botany*, 58(13), 3645–3656.

- Overmyer, K., M. Brosché, R. Pellinen, T. Kuittinen, H. Tuominen, R. Ahlfors, M. Keinänen, M. Saarma, D. Scheel y J. Kangasjärvi (2005) Ozone-Induced Programmed Cell Death in the Arabidopsis radical-induced cell death1 Mutant, Plant Physiology, 137(3), 1092 LP – 1104.
- Palla, G., I. Derenyi, I. Farkas y T. Vicsek (2005) Uncovering the overlapping community structure of complex networks in nature and society, Nature, 435(7043), 814–818.
- Pan, L. y L. Jiang (2014) Identification and expression of the WRKY transcription factors of Carica papaya in response to abiotic and biotic stresses, Molecular Biology Reports, 41(3), 1215–1225.
- Panchy, N., M. Lehti-Shiu y S.-H. Shiu (2016) Evolution of Gene Duplication in Plants, Plant Physiology, 171(4), 2294–2316.
- PARÉS, J., M. ARIZALETA, M. E. SANABRIA y G. GARCÍA (2008) Efecto de los niveles de salinidad sobre la densidad estomática, índice estomático y grosor foliar en plantas de carica papaya L. , Acta Botánica Venezuelica . scielon , 27–34.
- Parés, J. y C. Basso (2013) Efecto del cloruro de sodio sobre el crecimiento y estado nutricional de plantas de papaya, Bioagro. Universidad Centroccidental Lisandro Alvarado, 25(2), 109–116.
- Pattengale, N. D., M. Alipour, O. R. P. Bininda-Emonds, B. M. E. Moret y A. Stamatakis (2009) How Many Bootstrap Replicates Are Necessary?, en Batzoglou, S. (ed.) Research in Computational Molecular Biology: 13th Annual International Conference, RECOMB 2009, Tucson, AZ, USA, May 18-21, 2009. Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg, 184–200.
- Van de Peer, Y., E. Mizrahi y K. Marchal (2017) The evolutionary significance of polyploidy, Nat Rev Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 18(7), 411–424.
- Pei, J., B.-H. Kim y N. V Grishin (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments, Nucleic Acids Research, 36(7), p. 2295.

- Perochon, A., D. Aldon, J.-P. Galaud y B. Ranty (2011) Calmodulin and calmodulin-like proteins in plant calcium signaling, *Biochimie*, 93(12), 2048–2053.
- Petrov, V., J. Hille, B. Mueller-Roeber y T. S. Gechev (2015) ROS-mediated abiotic stress-induced programmed cell death in plants , *Frontiers in Plant Science* , p. 69.
- Porter, B. W., K. S. Aizawa, Y. J. Zhu y D. A. Christopher (2008) Differentially expressed and new non-protein-coding genes from a *Carica papaya* root transcriptome survey, *Plant Science*, 174(1), 38–50.
- Priault, M., N. Camougrand, K. W. Kinnally, F. M. Vallette y S. Manon (2003) Yeast as a tool to study Bax/mitochondrial interactions in cell death, *FEMS Yeast Research*, 4(1), 15 LP – 27.
- Qadir, M., E. Quillérrou, V. Nangia, G. Murtaza, M. Singh, R. J. Thomas, P. Drechsel y A. D. Noble (2014) Economics of salt-induced land degradation and restoration, *Natural Resources Forum*. John Wiley & Sons, Ltd (10.1111), 38(4), 282–295.
- Quintal, E. B., D. E. L. A. C. E. M. ILEANA, D. E. F. M. L. MARIA y M. M. ESTEVEZ (2012) Plants' challenges in a salinized world: The case of *Capsicum*, *African Journal of Biotechnology*.
- R Development Core Team y R Core Team (2017) R: A Language and Environment for Statistical Computing. Vienna, Austria.
- Ramiro, D. A., D. M. Melotto-Passarín, M. de A. Barbosa, F. dos Santos, S. G. P. Gomez, N. S. Massola Júnior, E. Lam y H. Carrer (2016) Expression of *Arabidopsis Bax Inhibitor-1* in transgenic sugarcane confers drought tolerance, *Plant Biotechnology Journal*, p. n/a-n/a.
- Rao, K. S., K. S. Rajput y Y. S. Kim (2013) Secondary growth and occurrence of laticifers in the root of papaya (*Carica papaya* L.), *Acta Botanica Gallica*. Taylor & Francis, 160(3–4), 255–260.
- Reape, T. J. y P. F. McCabe (2010) Apoptotic-like regulation of programmed cell death in plants, *Apoptosis*, 15(3), 249–256.

- Reimers, K., C. Y. U. Choi, V. Bucan y P. M. Vogt (2008) The Bax Inhibitor-1 (BI-1) family in apoptosis and tumorigenesis., *Current molecular medicine*, 8(2), 148–156.
- Renny-Byfield, S. y J. F. Wendel (2014) Doubling down on genomes: Polyploidy and crop plants, *American Journal of Botany* , 101(10), 1711–1725.
- Reza, S. H., N. Delhomme, N. R. Street, P. Ramachandran, K. Dalman, O. Nilsson, E. A. Minina y P. V Bozhkov (2018) Transcriptome analysis of embryonic domains in Norway spruce reveals potential regulators of suspensor cell death, *PLOS ONE*. Public Library of Science, 13(3), p. e0192945.
- Rice, P., I. Longden y A. Bleasby (2000) EMBOS: the European molecular biology open software suite, *Trends in genetics*. Elsevier Current Trends, 16(6), 276–277.
- Robinson, K. S., a Clements, a C. Williams, C. N. Berger y G. Frankel (2011) Bax inhibitor 1 in apoptosis and disease., *Oncogene*. Nature Publishing Group, 30(21), 2391–2400.
- Robinson, M. D., D. J. McCarthy y G. K. Smyth (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, 26(1), 139–140.
- Rojas-Rivera, D. y C. Hetz (2014) TMBIM protein family: ancestral regulators of cell death., *Oncogene*. Nature Publishing Group, 34(October 2013), 1–12.
- Ruberti, C., Y. Lai y F. Brandizzi (2018) Recovery from temporary endoplasmic reticulum stress in plants relies on the tissue-specific and largely independent roles of bZIP28 and bZIP60, as well as an antagonizing function of BAX-Inhibitor 1 upon the pro-adaptive signaling mediated by bZIP28, *The Plant Journal*. John Wiley & Sons, Ltd, 93(1), 155–165.
- Sanchez, P., M. De Torres Zabala y M. Grant (2000) AtBI-1, a plant homologue of Bax Inhibitor-1, suppresses Bax-induced cell death in yeast and is rapidly upregulated during wounding and pathogen challenge, *Plant Journal*, 21(4), 393–399.
- Schachtman, D. P. y J. Q. D. Goodger (2008) Chemical root to shoot signaling under

- drought, *Trends in Plant Science*, 13(6), 281–287.
- Schaefer, R. J., J.-M. Michno y C. L. Myers (2017) Unraveling gene function in agricultural species using gene co-expression networks, *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1860(1), 53–63.
- Serin, E. A. R., H. Nijveen, H. W. M. Hilhorst y W. Ligterink (2016) Learning from Co-expression Networks: Possibilities and Challenges, *Frontiers in Plant Science*, 7, p. 444.
- Shabala, S., R. G. White, M. A. Djordjevic, Y.-L. Ruan y U. Mathesius (2016) Root-to-shoot signalling: integration of diverse molecules, pathways and functions, *Functional Plant Biology*, 43(2), 87–104.
- Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski y T. Ideker (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks, *Genome Research*. Cold Spring Harbor Laboratory Press, 13(11), 2498–2504.
- Sierla, M., D. L. Prole, N. Saraiva, G. Carrara, N. Dinischiotu, A. Vaattovaara, M. Wrzaczek, C. W. Taylor, G. L. Smith y B. Feys (2019) Golgi anti-apoptotic proteins are evolutionarily conserved ion channels that regulate cell death in plants, *bioRxiv*, p. 859678.
- Silva, J. A. T. da, Z. Rashid, D. T. Nhut, D. Sivakumar, A. Gera, M. T. Souza y P. Tennant (2007) Papaya (*Carica papaya* L.) biology and biotechnology, *Tree and Forestry Science and Biotechnology*, 1(1), 47–73.
- Singh, D. y A. Laxmi (2015) Transcriptional regulation of drought response: a tortuous network of transcriptional factors, *Frontiers in Plant Science*, 6, p. 895.
- Sojo, V., C. Dessimoz, A. Pomiankowski y N. Lane (2016) Membrane Proteins Are Dramatically Less Conserved than Water-Soluble Proteins across the Tree of Life, *Molecular Biology and Evolution*, 33(11), 2874–2884.
- Soucy, S. M., J. Huang y J. P. Gogarten (2015) Horizontal gene transfer: building the web

- of life, *Nature Reviews Genetics*. Nature Publishing Group, 16(8), 472–482.
- Spiegelman, Z., G. Golan y S. Wolf (2013) Don't kill the messenger: Long-distance trafficking of mRNA molecules, *Plant Science*, 213, 1–8.
- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics*. Oxford Univ Press, 30(9), 1312–1313.
- van Stelten, J., F. Silva, D. Belin y T. J. Silhavy (2009) Effects of Antibiotics and a Proto-Oncogene Homolog on Destruction of Protein Translocator SecY, *Science*, 325(5941), 753 LP – 756.
- Su, G., A. Kuchinsky, J. H. Morris, D. J. States y F. Meng (2010) GLay: community structure analysis of biological networks, *Bioinformatics*, 26(24), 3135–3137.
- Tang, H., J. E. Bowers, X. Wang y A. H. Paterson (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage, *Proceedings of the National Academy of Sciences*, 107(1), 472–477.
- Taylor-Brown, E. y H. Hurd (2013) The first suicides: a legacy inherited by parasitic protozoans from prokaryote ancestors, *Parasites & Vectors*, 6(1), p. 108.
- Thieme, C. J., M. Rojas-Triana, E. Stecyk, C. Schudoma, W. Zhang, L. Yang, M. Miñambres, D. Walther, W. X. Schulze, J. Paz-Ares, W.-R. Scheible y F. Kragler (2015) Endogenous Arabidopsis messenger RNAs transported to distant tissues, *Nature Plants*. Macmillan Publishers Limited, 1, p. 15025.
- Tom, S. y W. G. D. (2006) The effects of stress on plant cuticular waxes, *New Phytologist*, 171(3), 469–499.
- Trapnell, C., L. Pachter y S. L. Salzberg (2009) TopHat: discovering splice junctions with RNA-Seq, *Bioinformatics*, 25(9), 1105–1111.
- Tuteja, N., P. Ahmad, B. B. Panda y R. Tuteja (2009) Genotoxic stress in plants: Shedding light on DNA damage, repair and DNA repair helicases, *Mutation Research/Reviews in Mutation Research*, 681(2), 134–149.

- Unamba, C. I. N., A. Nag y R. K. Sharma (2015) Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants, *Frontiers in Plant Science*, 6, p. 1074.
- Urano, K., Y. Kurihara, M. Seki y K. Shinozaki (2010) “Omics” analyses of regulatory networks in plant abiotic stress responses, *Current Opinion in Plant Biology*, 13(2), 132–138.
- Urasaki, N., K. Tarora, A. Shudo, H. Ueno, M. Tamaki, N. Miyagi, S. Adaniya y H. Matsumura (2012) Digital Transcriptome Analysis of Putative Sex-Determination Genes in Papaya (*Carica papaya*), *PLOS ONE. Public Library of Science*, 7(7), 1–9.
- Van Bel, M., T. Diels, E. Vancaester, L. Kreft, A. Botzki, Y. Van de Peer, F. Coppens y K. Vandepoele (2018) PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics, *Nucleic Acids Research*, 46(D1), D1190–D1196.
- Vishwanath, S. J., C. Delude, F. Domergue y O. Rowland (2015) Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier, *Plant Cell Reports*, 34(4), 573–586.
- Wang, H., Honglei Wang, H. Shao y X. Tang (2016) Recent Advances in Utilizing Transcription Factors to Improve Plant Abiotic Stress Tolerance by Transgenic Technology, *Frontiers in Plant Science*, 7, p. 67.
- Wang, Y., H. Tang, J. D. DeBarry, X. Tan, J. Li, X. Wang, T. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger y A. H. Paterson (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, *Nucleic Acids Research*, 40(7), e49–e49.
- Wang, Y., S. P. Ficklin, X. Wang, F. A. Feltus y A. H. Paterson (2016) Large-Scale Gene Relocations following an Ancient Genome Triplication Associated with the Diversification of Core Eudicots, *PLOS ONE. Public Library of Science*, 11(5), p. e0155637.
- Watanabe, N. y E. Lam (2006) Arabidopsis Bax inhibitor-1 functions as an attenuator of biotic and abiotic types of cell death, *Plant Journal*, 45(6), 884–894.

- Watanabe, N. y E. Lam (2008) BAX inhibitor-1 modulates endoplasmic reticulum stress-mediated programmed cell death in arabidopsis, *Journal of Biological Chemistry*, 283(6), 3200–3210.
- Watanabe, N. y E. Lam (2009) Bax inhibitor-1, a conserved cell death suppressor, is a key molecular switch downstream from a variety of biotic and abiotic stress signals in plants, *International Journal of Molecular Sciences*, 10(7), 3149–3167.
- Weis, C., R. Hückelhoven y R. Eichmann (2013) LIFEGUARD proteins support plant colonization by biotrophic powdery mildew fungi, *Journal of Experimental Botany*, 64(12), 3855–3867.
- Wheeler, T. J., J. Clements y R. D. Finn (2014) Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models, *BMC Bioinformatics*, 15(1), p. 7.
- Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Xu, G., S. Wang, S. Han, K. Xie, Y. Wang, J. Li y Y. Liu (2017) Plant Bax Inhibitor-1 interacts with ATG6 to regulate autophagy and programmed cell death, *Autophagy*. Taylor & Francis, 13(7), 1161–1175.
- Xu, P., S. J. Rogers y M. J. Roossinck (2004) Expression of antiapoptotic genes bcl-xL and ced-9 in tomato enhances tolerance to viral-induced necrosis and abiotic stress, *Proceedings of the National Academy of Sciences of the United States of America*, 101(44), 15805–15810.
- Xu, Q. y J. C. Reed (1998) Bax inhibitor-1, a mammalian apoptosis suppressor identified by functional screening in yeast., *Molecular cell*, 1(3), 337–346.
- Xue, D., X. Zhang, X. Lu, G. Chen y Z.-H. Chen (2017) Molecular and Evolutionary Mechanisms of Cuticular Wax for Plant Drought Tolerance, *Frontiers in Plant Science*, 8, p. 621.
- Xue, G. P., J. Drenth y C. L. McIntyre (2015) TaHsfA6f is a transcriptional activator that regulates a suite of heat stress protection genes in wheat (*Triticum aestivum* L.)

- including previously unknown Hsf targets, *Journal of Experimental Botany*, 66(3), 1025–1039.
- YAMAGAMI, A., M. NAKAZAWA, M. MATSUI, M. TUJIMOTO, M. SAKUTA, T. ASAMI y T. NAKANO (2009) Chemical Genetics Reveal the Novel Transmembrane Protein BIL4, Which Mediates Plant Cell Elongation in Brassinosteroid Signaling, *Bioscience, Biotechnology and Biochemistry*, 73(2), 415–421.
- Yoshida, T., J. Mogami y K. Yamaguchi-Shinozaki (2014) ABA-dependent and ABA-independent signaling in response to osmotic stress in plants, *Current Opinion in Plant Biology*, 21, 133–139.
- Yoshida, T., J. Mogami y K. Yamaguchi-Shinozaki (2015) Omics Approaches Toward Defining the Comprehensive Absciscic Acid Signaling Network in Plants, *Plant and Cell Physiology*, 56(6), 1043–1052.
- Yu, Y., J. Xiao, Y. Yang, C. Bi, L. Qing y W. Tan (2015) Ss-Bi1 encodes a putative BAX inhibitor-1 protein that is required for full virulence of *Sclerotinia sclerotiorum*, *Physiological and Molecular Plant Pathology*. Elsevier Ltd, 90, 115–122.
- Yue, H., S. Nie y D. Xing (2012) Over-expression of Arabidopsis Bax inhibitor-1 delays methyl jasmonate-induced leaf senescence by suppressing the activation of MAP kinase 6, *Journal of Experimental Botany*, 63(12), 4463–4474.
- Zhang, J.-L. y H. Shi (2013) Physiological and molecular mechanisms of plant salt tolerance, *Photosynthesis Research*, 115(1), 1–22.
- Zhang, Jingxia, F. Wang, C. Zhang, Junhao Zhang, Y. Chen, G. Liu, Y. Zhao, F. Hao y Jun Zhang (2018) A novel VIGS method by agroinoculation of cotton seeds and application for elucidating functions of GhBI-1 in salt-stress response, *Plant Cell Reports*, 37(8), 1091–1100.
- Zhao, S., Y. Guo, Q. Sheng y Y. Shyr (2014) Advanced heat map and clustering analysis using heatmap3, *BioMed Research International*. Hindawi, 2014.
- Zhao, T., R. Holmer, S. de Bruijn, G. C. Angenent, H. A. van den Burg y M. E. Schranz

(2017) Phylogenomic Synteny Network Analysis of MADS-Box Transcription Factor Genes Reveals Lineage-Specific Transpositions, Ancient Tandem Duplications, and Deep Positional Conservation, *The Plant Cell*.

Zhao, T. y M. E. Schranz (2017) Network approaches for plant phylogenomic synteny analysis, *Current Opinion in Plant Biology*, 36, 129–134.

Zhao, Y., H. Tang y Y. Ye (2012) RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data, *Bioinformatics*, 28(1), 125–126.

Zhou, Jinhua, T. Zhu, C. Hu, H. Li, G. Chen, G. Xu, S. Wang, Jianfeng Zhou y D. Ma (2008) Comparative genomics and function analysis on BI1 family, *Computational Biology and Chemistry*, 32(3), 159–162.