

**UNIVERSIDAD NACIONAL DEL LITORAL**  
**Facultad de Bioquímica y Ciencias Biológicas**



**Tesis para la obtención del Grado Académico  
de Doctor en Ciencias Biológicas**

**“Conservación y divergencia  
funcional entre miembros de la familia de  
factores de transcripción HD-Zip.  
Análisis molecular, evolutivo y  
de las redes de regulación en las que  
participan”**

**Agustín Lucas Arce**

**Directora de Tesis: Dra. Raquel Lía Chan**

**Instituto de Agrobiotecnología del Litoral**

**- 2012 -**

# Índice general

<b>Índice de Figuras</b>	<b>v</b>
<b>Índice de Tablas</b>	<b>vii</b>
<b>Abreviaturas</b>	<b>viii</b>
<b>Resumen</b>	<b>x</b>
<b>Abstract</b>	<b>xiii</b>
<b>Agradecimientos</b>	<b>xvi</b>
<b>Antecedentes</b>	<b>xix</b>
<b>I Introducción general</b>	<b>1</b>
I.1 La vida: complejidad y ordenamiento . . . . .	2
I.2 Evolución de los niveles de organización . . . . .	2
I.3 Los factores de transcripción . . . . .	3
I.3.1 La transcripción en procariotas y eucariotas . . . . .	4
I.3.2 Clasificación de los factores de transcripción . . . . .	5
I.4 La familia de factores de transcripción HD-Zip . . . . .	7
I.4.1 Primeros miembros descubiertos . . . . .	7
I.4.2 Las cuatro subfamilias de factores de transcripción HD-Zip . . . . .	8
I.5 Los HD-Zip I y II a nivel molecular . . . . .	10
I.5.1 El homeodominio y su interacción con el ADN . . . . .	10
I.5.2 El cierre de leucinas, dominio encargado de la dimerización . . . . .	11
I.6 Función de los factores de transcripción HD-Zip . . . . .	13
I.6.1 La subfamilia HD-Zip I . . . . .	14
I.6.1.1 Relación con las vías de señales de ABA y otras fitohormonas	14
I.6.1.2 Respuestas a estímulos externos . . . . .	17
I.6.1.3 Regulación del desarrollo . . . . .	17
I.6.2 La subfamilia HD-Zip II . . . . .	18
I.7 <i>Arabidopsis thaliana</i> como modelo de estudio en la era post-genómica . . . . .	20
<b>Objetivos generales</b>	<b>22</b>
<b>II Materiales y Métodos</b>	<b>23</b>
II.1 Materiales y Métodos correspondientes al Capítulo III . . . . .	24

II.1.1	Búsqueda y obtención del conjunto de secuencias de FTs HD-Zip I . . .	24
II.1.2	Alineamiento de secuencias y análisis filogenético . . . . .	24
II.1.3	Reconocimiento de motivos y modificaciones post-traduccionales . . .	25
II.1.4	Cepas bacterianas . . . . .	25
II.1.5	Transformación de bacterias . . . . .	26
II.1.6	Construcciones genéticas . . . . .	26
II.1.7	Obtención y purificación de ácidos nucleicos . . . . .	28
	II.1.7.1 Minipreparación de ADN plasmídico . . . . .	28
	II.1.7.2 Minipreparación de ADN genómico de <i>Arabidopsis thaliana</i>	30
II.1.8	Reacciones enzimáticas . . . . .	30
	II.1.8.1 Digestión de ADN con endonucleasas de restricción . . . . .	30
	II.1.8.2 Ligación de moléculas de ADN . . . . .	30
	II.1.8.3 Reacción en cadena de la polimerasa (PCR) . . . . .	31
II.1.9	Análisis de ácidos nucleicos . . . . .	31
	II.1.9.1 Electroforesis de ADN en geles de agarosa . . . . .	31
	II.1.9.2 Purificación de fragmentos de ADN . . . . .	32
II.1.10	Cultivo, transformación y simple híbrido en levaduras . . . . .	32
II.1.11	Tratamiento de plantas de <i>Arabidopsis thaliana</i> . . . . .	33
	II.1.11.1 Cultivo de plantas en tierra . . . . .	33
	II.1.11.2 Desinfección de semillas y cultivo <i>in vitro</i> . . . . .	33
II.1.12	Transformación de plantas de <i>Arabidopsis thaliana</i> . . . . .	33
	II.1.12.1 Selección de transformantes y obtención de líneas homocigotas	34
II.1.13	Cuantificación de transcritos por RT-PCR en tiempo real . . . . .	35
II.1.14	Tratamiento con etileno . . . . .	35
II.1.15	Cuantificación del aserrado de las hojas . . . . .	36
II.2	Materiales y Métodos correspondientes al Capítulo IV . . . . .	36
	II.2.1 Características y procesamiento de la información transcriptómica . . .	36
	II.2.1.1 Filtrado de los datos . . . . .	37
	II.2.2 Reconstrucción de la red con ARACNE . . . . .	38
	II.2.3 Análisis de promotores . . . . .	39
	II.2.4 Estudio de regulación conjunta . . . . .	39
	II.2.5 Estudio funcional con términos GO . . . . .	40

<b>III Fuentes de diversidad funcional en factores de transcripción HD-Zip de la subfamilia I</b>	<b>41</b>	
III.1 Introducción . . . . .	42	
	III.1.1 Familias génicas de FTs vegetales . . . . .	42
	III.1.2 Análisis filogenéticos de los FTs HD-Zip I . . . . .	42
	III.1.3 Diversidad funcional en la subfamilia HD-Zip I . . . . .	43
	III.1.3.1 Patrones de expresión de los genes HD-Zip I de <i>A. thaliana</i> .	45
	III.1.3.2 Expresión ectópica y constitutiva de los genes HD-Zip I . . .	47
	III.1.4 Evidencia de funcionalidad de regiones fuera del dominio HD-Zip . . .	48
III.2 Objetivos . . . . .	50	
III.3 Resultados . . . . .	51	
	III.3.1 Base de datos utilizada para el análisis . . . . .	51
	III.3.2 Clados hallados mediante el análisis filogenético . . . . .	52
	III.3.2.1 Manipulación y alineamiento de las secuencias . . . . .	53

III.3.2.2	Conservación de residuos clave en el reconocimiento del ADN	53
III.3.2.3	Árboles filogenéticos y grupos con región carboxilo terminal (RCTs) similares	54
III.3.3	Motivos conservados en las RCTs	55
III.3.3.1	Las RCTs serían blancos de modificaciones post-traduccionales	60
III.3.4	Motivos conservados en las RNTs	62
III.3.5	Las RCTs poseen la capacidad de transactivar en levaduras	64
III.3.6	La RCT participa en la determinación de la función de la proteína HD-Zip	65
III.3.6.1	El aserrado se incrementa en las plantas con el FT quimérico que posee la RCT de HaHB1	67
III.3.6.2	Las plantas que expresan el FT quimérico H4-H1 presentan una sensibilidad intermedia al etileno	68
III.4	Discusión de resultados	70
III.4.1	Reconstrucción filogenética y definición de grupos	70
III.4.1.1	Comparación con la reconstrucción filogenética obtenida en <i>A. thaliana</i>	72
III.4.2	El motivo AHA y la activación de la transcripción	73
III.4.3	Modificaciones post-traduccionales	75
III.4.3.1	La fosforilación en la regulación de la unión al ADN	75
III.4.3.2	La sumoilación como un nuevo mecanismo de regulación	76
III.4.4	Las RNTs poseerían una funcionalidad limitada	77
III.4.4.1	Otros posibles mecanismos de regulación	77
III.4.5	Relevancia funcional de la RCT en la planta	78
III.4.6	Un modelo de la relación estructura/función de los FTs HD-Zip I	80
III.5	Conclusiones	81

#### **IV Redes de expresión génica en las que participan los factores de transcripción HD-Zip de *Arabidopsis thaliana***

IV.1	Introducción	83
IV.1.1	Las redes de regulación génica	84
IV.1.1.1	Propiedades emergentes	84
IV.1.2	Ingeniería reversa de redes de regulación	85
IV.1.2.1	Niveles de expresión génica	85
IV.1.2.2	Los microarreglos de ADN	86
IV.1.2.3	Algoritmos para la reconstrucción de redes de regulación génica	86
IV.1.3	La información mutua	87
IV.1.4	Algoritmo para la reconstrucción de redes celulares precisas	88
IV.1.4.1	Cálculo de la información mutua	88
IV.1.4.2	Remoción de interacciones indirectas	90
IV.1.5	Redes de regulación en las que participan los FTs HD-Zip I y II	91
IV.2	Objetivos	93
IV.3	Resultados	94
IV.3.1	Datos analizados	94
IV.3.2	Filtrado de los datos	94
IV.3.3	Obtención de los módulos de actividad transcripcional	98
IV.3.4	Análisis de los módulos de actividad transcripcional	98

IV.3.5	Análisis de los promotores en los genes de los módulo de actividad transcripcionals (MATs) . . . . .	99
IV.3.5.1	Búsqueda del motivo unido con mayor afinidad . . . . .	99
IV.3.5.2	Búsqueda de motivos <i>de novo</i> . . . . .	103
IV.3.6	Solapamiento entre los MATs de diferentes FTs HD-Zip . . . . .	103
IV.3.7	Estudios de expresión coordinada dentro de los MATs . . . . .	105
IV.3.8	Análisis de la participación de los genes HD-Zip en diferentes estreses .	109
IV.3.8.1	Subgrupos de <i>probesets</i> corregulados . . . . .	110
IV.3.9	Análisis del enriquecimiento de términos GO . . . . .	113
IV.4	Discusión de resultados . . . . .	115
IV.4.1	Red de regulación transcripcional reconstruida con ARACNE . . . . .	115
IV.4.2	Elementos en <i>cis</i> en los promotores de los genes de los MATs . . . . .	116
IV.4.3	Redundancia funcional entre miembros de las subfamilias I y II . . . . .	118
IV.4.3.1	Los MATs HD-Zip frente a las hipótesis previas . . . . .	119
IV.4.4	Coordinación de la expresión en los MATs . . . . .	120
IV.4.4.1	Los genes HD-Zip participarían en las respuestas al calor . . . . .	121
IV.4.4.2	Los genes de los MATs son regulados en forma negativa o positiva en un mismo estrés . . . . .	122
IV.4.5	Los MAT desde una perspectiva funcional . . . . .	123
IV.4.5.1	Algunas consideraciones finales sobre el análisis funcional . . . . .	126
IV.5	Conclusiones . . . . .	127
	<b>Conclusiones generales</b>	<b>128</b>
	<b>Apéndice</b>	<b>129</b>
V.1	Secuencias utilizadas en los análisis filogenéticos . . . . .	130
V.2	Alineamiento de dominios HD-Zip de las 178 proteínas . . . . .	134
V.3	Árbol filogenético HZ+OG . . . . .	138
V.4	Tablas del análisis de enriquecimiento de términos GO . . . . .	139
	<b>Bibliografía</b>	<b>163</b>

# Índice de Figuras

I.1	Regulación del inicio de la transcripción . . . . .	6
I.2	Dominios y motivos en las proteínas HD-Zip . . . . .	9
I.3	Interacción entre el homeodominio y el ADN . . . . .	10
I.4	Homodímero HD-Zip interaccionando con el ADN . . . . .	12
III.1	Filogenia de los genes HD-Zip I. . . . .	44
III.2	Alineamiento del dominio HD-Zip . . . . .	54
III.3	Regiones carboxilo terminales de cada grupo . . . . .	56
III.4	Árboles HZT y CST . . . . .	57
III.5	Abundancia de Trp en las RCT. Histograma que muestra la abundancia de Trp dividiendo las RCTs en 10 porciones. . . . .	58
III.6	Motivos hallados en las RCTs . . . . .	59
III.7	Secuencias de los motivos de las RCT . . . . .	60
III.8	Frecuencia de los diferentes residuos en los sitios de sumoilación presentes en las RCT. . . . .	61
III.9	Motivos hallados en las RNTs . . . . .	63
III.10	Secuencias de los motivos de las RNT . . . . .	64
III.11	Funcionalidad del motivo AHA en ensayos de simple híbrido en levaduras . . . . .	65
III.12	Construcciones expresadas por las plantas transgénicas. . . . .	66
III.13	Ensayo de aserrado de hojas . . . . .	67
III.14	Ensayo de triple respuesta . . . . .	69
III.15	Modelo funcional de un FT HD-Zip I . . . . .	80
IV.1	Diagrama de la información . . . . .	88
IV.2	Estimador de densidad de núcleo Gaussiano . . . . .	89
IV.3	Análisis de DPI . . . . .	91
IV.4	Distribución de estadísticos aplicados a la expresión de los <i>probesets</i> . . . . .	96
IV.5	Remoción de los <i>probesets</i> según los criterios estadísticos . . . . .	97
IV.6	Distribución de tamaños de los MATs de todos los FT . . . . .	100
IV.7	Distribución de tamaños de los MATs de los HD-Zip . . . . .	100
IV.8	Tamaño de los MATs . . . . .	101
IV.9	Búsqueda de motivos <i>de novo</i> . . . . .	104
IV.10	Solapamiento entre los MATs de diferentes FTs HD-Zip . . . . .	105
IV.11	Expresión global del MAT de AtHB12 . . . . .	106
IV.12	Expresión global del MAT de AtHB1 . . . . .	107
IV.13	Expresión global del MAT de HAT2 . . . . .	108
IV.14	Ejemplo de la determinación de <i>probesets</i> con expresión coordinada . . . . .	109

IV.15	Coordinación de la expresión del MAT de AtHB12 diferentes tratamientos . . .	110
IV.16	Regulación conjunta de genes de los MATs frente a diferentes estreses . . . . .	111
IV.17	Perfiles de expresión de los <i>probesets</i> corregulados . . . . .	112
IV.18	Relaciones de los términos GO en MAT de AtHB12 . . . . .	113
V.1	Alineamiento de dominios HD-Zip . . . . .	134
V.2	Árbol filogenético HZ+OG . . . . .	138

# Índice de Tablas

I.1	Principales transiciones en la evolución . . . . .	3
I.2	La familia HD-Zip en distintas especies . . . . .	9
I.3	Función de los HD-Zip I . . . . .	15
I.4	Función de los HD-Zip II . . . . .	19
II.1	Oligonucleótidos utilizados para los clonados . . . . .	29
III.1	Distribución de proteínas por especie . . . . .	52
III.2	Proteínas por taxones principales . . . . .	53
III.3	Valores de <i>Bootstrap</i> de los diferentes grupos en los árboles HZT y CST . . . . .	55
III.4	Recuento de aminoácidos en los sitios SUMO . . . . .	61
III.5	Análisis de Kruskal-Wallis de los resultados de aserrado. . . . .	68
III.6	Comparación de las reconstrucciones filogenéticas . . . . .	73
IV.1	Resumen de los resultados del filtrado. . . . .	96
IV.2	Parámetros de diferentes conjuntos de MATs . . . . .	99
IV.3	Motivos en los promotores del conjunto de genes completo . . . . .	102
IV.4	Búsqueda de motivos en lo promotores de los genes de los MATs . . . . .	102
IV.5	Términos GO hallados en el MAT de AtHB12 . . . . .	114



## Abreviaturas

<b>35SCaMV</b>	35S del virus del mosaico de la coliflor
<b>aa</b>	aminoácidos
<b>ABA</b>	ácido abscísico
<b>ACC</b>	ácido 1-aminociclopropano-1-carboxílico
<b>Aux</b>	auxinas
<b>ADN</b>	ácido desoxirribonucleico
<b>ADNc</b>	ADN copia
<b>AHA</b>	<i>Aromatic, large Hydrophobic, Acidic context</i>
<b>ARN</b>	ácido ribonucleico
<b>ARNm</b>	ARN mensajero
<b>ARNmi</b>	ARN micro
<b>ARN Pol</b>	ARN polimerasa
<b>ARNr</b>	ARN ribosomal
<b>ARNt</b>	ARN de transferencia
<b>bZip</b>	región básica-cierre de leucinas
<b>bHLH</b>	región básica-hélice-rulo-hélice
<b>CK</b>	citoquininas
<b>DHL</b>	doble híbrido en levaduras
<b>EMSA</b>	ensayos de retardo en geles
<b>ET</b>	etileno
<b>FT</b>	factores de transcripción
<b>GA</b>	giberelina
<b>BD</b>	dominio de unión a ADN del factor de transcripción de levaduras GAL4
<b>GUS</b>	$\beta$ -glucuronidasa

<b>HALZ</b>	cierre de leucinas asociado a un homeodominio
<b>HB</b>	<i>homeobox</i>
<b>HD</b>	homeodominio
<b>HD-SAD</b>	dominio asociado a START
<b>HD-Zip</b>	homeodominio-cierre de leucinas
<b>IM</b>	información mutua
<b>JA</b>	ácido jasmónico
<b>LZ</b>	cierre de leucinas
<b>pb</b>	pares de bases
<b>MAT</b>	módulo de actividad transcripcional
<b>ML</b>	máxima verosimilitud o <i>maximum likelihood</i>
<b>min</b>	minutos
<b>PCR</b>	reacción en cadena de la polimerasa
<b>PCRC</b>	PCR cuantitativa
<b>RCT</b>	región carboxilo terminal
<b>RNT</b>	región amino terminal
<b>RT</b>	transcripción reversa
<b>SA</b>	ácido salicílico
<b>SLN</b>	señal de localización nuclear
<b>SUMO</b>	<i>small ubiquitin-related modifier</i>
<b>START</b>	<i>steroidogenic acute regulatory protein-related lipid transfer</i>
<b>TBP</b>	proteína de unión a la caja TATA
<b>WT</b>	salvaje
<b>Zip</b>	cierre de leucinas

## Resumen

Las plantas, como el resto de los seres vivos, poseen complejos mecanismos de regulación encargados del control preciso de la expresión de sus genes. Los factores de transcripción (FTs) son proteínas que desempeñan una función esencial en estas redes de regulación. Son capaces de unir el ADN en las regiones promotoras de sus genes blanco y alterar su transcripción.

Existen numerosas familias de FTs definidas principalmente por el dominio de unión a ADN que poseen sus integrantes. La familia de HD-Zip está compuesta por proteínas que unen el ADN a través del homeodominio, y además dimerizan a través de un cierre de leucinas, condición requerida para la unión al ADN. De las cuatro subfamilias en las que es subdividida, la subfamilia I consiste en proteínas sin otros dominios o motivos descritos, mientras que las proteínas de la subfamilia II poseen un motivo de regulación redox y una región conservada de función desconocida.

Los factores de transcripción de la subfamilia I están involucrados en variados procesos, muchos de los cuales tienen relación con la fitohormona ácido abscísico (ABA). Cuando han sido expresados bajo la dirección del promotor fuerte y constitutivo 35S del virus del mosaico de la coliflor (*35SCaMV*), generaron plantas transgénicas con fenotipos diferentes. El hecho de que estas proteínas unan *in vitro* con máxima afinidad la misma secuencia, CAATNATTG, sugiere que habría regiones fuera del dominio HD-Zip responsables de la diversidad funcional.

Esto motivó la búsqueda de posibles regiones funcionales de la proteína por fuera del dominio HD-Zip. Se analizaron las secuencias de 178 proteínas de la subfamilia I de diferentes especies, algunas de ellas con su genoma completamente secuenciado. Una reconstrucción filogenética y una inspección de sus regiones carboxilo terminales (RCTs) permitió definir seis grupos monofiléticos con gran similitud en estas regiones.

La búsqueda de motivos conservados en las RCTs indicó la presencia de secuencias específicas de cada grupo y de una secuencia en la porción final de la mayoría de las proteínas con las características del motivo AHA, activador de la transcripción. Mediante ensayos de simple híbrido en levaduras se demostró que el motivo AHA del FT AtHB13 posee la capacidad de transactivación.

Otros motivos de las RCTs presentaron sitios putativos para dos mecanismos de modificación post-traducciona, fosforilación y sumoilación. Aunque escasa, existe evidencia de estas modificaciones en FTs HD-Zip I, con lo que los resultados indicarían que estos mecanismos de regulación actuarían sobre una gran parte de las proteínas de esta subfamilia, aunque habría diferencias entre grupos.

Para investigar la funcionalidad *in planta* de las RCTs, se expresó en *A. thaliana* una proteína quimérica con el dominio HD-Zip de la proteína HaHB4 fusionada a la RCT de HaHB1, empleando el promotor *35SCaMV*. Ambas proteínas de girasol han sido extensamente estudiadas y se contaba con las plantas que las expresaban en forma salvaje bajo la dirección del

promotor *35SCaMV*. Se evaluaron fenotipos característicos generados por cada proteína salvaje: la insensibilidad al etileno observada cuando se expresa HaHB4 y el aserrado aumentado en hojas generado por la expresión de HaHB1. Las plantas que expresaban la proteína quimérica presentaron un aumento significativo en el aserrado de las hojas en una de las líneas y su insensibilidad al etileno fue leve, señalando que la identidad de la RCT tuvo un efecto importante en el fenotipo generado por la proteína.

El conjunto de los resultados permitió concluir que los FTs HD-Zip I serían capaces de activar la transcripción a través del dominio AHA y que sus RCTs serían regulados por fosforilación y sumoilación. De esta forma, las regiones fuera del dominio HD-Zip tendrían una importante participación en la determinación de las vías de señales en las que cada FT de esta subfamilia es capaz de actuar. Por una parte, la diversidad de motivos AHA podría ser responsable de una interacción diferencial con la maquinaria transcripcional; y por otra, los mecanismos de regulación podrían ser distintos sobre diferentes proteínas de la misma subfamilia.

Para estudiar las vías de señales en las que cada proteína HD-Zip participa se emplearon herramientas de ingeniería reversa de redes de regulación transcripcional. Utilizando un algoritmo basado en teoría de la información, se analizó la expresión de 9617 genes en 269 ensayos de microarreglos de *A. thaliana* que involucraban diferentes tratamientos de estrés abiótico, obteniéndose los potenciales genes blanco de 25 proteínas HD-Zip I y II.

Los promotores de estos conjuntos de genes blanco, llamados módulos de actividad transcripcional (MATs), fueron analizados en búsqueda del motivo CAATNATTG. Éste no se encontró enriquecido en relación a la proporción existente en los promotores del total de los genes incluidos en el análisis, sugiriendo que algunas secuencias similares también podrían ser reconocidas con suficiente afinidad para que se establezca una unión funcional. La búsqueda *de novo* de motivos permitió hallar el elemento ABRE en varios promotores de genes de los MAT. Este motivo es reconocido por FTs involucrados en la respuesta a ABA, lo que indicaría los FTs HD-Zip y estas proteínas podrían regular la expresión en forma conjunta.

El posible entrecruzamiento de las vías reguladas por los diferentes FTs fue evaluado analizando el grado de solapamiento entre MATs. Esto mostró que sólo es significativo entre algunos genes filogenéticamente relacionados, pero no sería relevante en otros casos.

A nivel global, los perfiles de expresión de los genes de los MATs presentaron, para algunos FTs, una marcada diferencia en los niveles de expresión media entre vástago y raíz. Esto sugirió una potencial diferencia en la forma que actúan subgrupos de genes del MAT en cada tejido, lo que podría ser responsabilidad de la acción del FT HD-Zip.

Se investigó también qué proporción de los genes de cada MAT poseía expresión coordinada en las diferentes condiciones de estrés. Considerando que esta medición es indicador del grado de participación del FT en el tratamiento, este análisis reveló que la mayoría de los FTs HD-Zip I y II tiene una actividad importante en estrés por calor, una función no asociada a estas proteínas previamente.

Al explorar los perfiles de los genes con expresión coordinada se observó que en algunas condiciones estresantes había subgrupos de genes inducidos y subgrupos de genes reprimidos. A pesar de no ser evidencia directa, esto podría estar asociado a la capacidad demostrada de los FTs de la subfamilia I de activar y reprimir la transcripción.

Finalmente, se estudiaron las categorías funcionales de los genes de los MATs que estadísticamente se encontraban enriquecidas. Varios de los FTs se asociaron a funciones ya encontradas por otros autores, brindando cierto sustento a los genes blanco hallados. Por otro lado, se encontraron varias categorías funcionales no descritas que podrían ser investigadas en futuros trabajos.

En síntesis, este trabajo permitió proponer un mecanismo de acción y regulación que presentarían los FTs HD-Zip I, resaltando la importancia de regiones proteicas no descritas previamente y proponiendo potenciales nuevas fuentes de diversidad funcional entre miembros. La reconstrucción de una red de regulación transcripcional posibilitó la definición de genes blancos putativos de las proteínas de las subfamilias I y II. Diferentes análisis de estos conjuntos brindaron un panorama global de las vías reguladas por los FTs de ambas subfamilias, permitiendo explorarlas desde una perspectiva comparativa y funcional, lo que indicó la existencia de similitudes y particularidades en la actividad de estas proteínas.

## Abstract

Plants, like the rest of the living beings, have complex regulatory mechanisms responsible for the precise control of their gene expression. Transcription factors (TFs) are proteins which play an essential role in these regulation networks. They are capable of binding DNA in the promoter regions of their target genes, altering their transcription.

There are multiple TF families mainly defined by the DNA binding domain of their members. The HD-Zip family is formed by proteins which bind DNA through the homodomain, and also dimerize through a leucine zipper, condition required for DNA binding. Of the four subfamilies in which it is divided, subfamily I comprises proteins lacking other known domains or motifs, whereas subfamily II proteins have a motif implicated in redox regulation and a conserved region of unknown function.

Subfamily I TFs are involved in various processes, many are related to the phytohormone ABA. When they have been expressed with the strong and constitutive promoter 35S cauliflower mosaic virus (*35SCaMV*), they resulted in transgenic plants with different phenotypes. The fact that these proteins bind *in vitro* with maximum affinity the same sequence, CAATNATTG, suggests that there should be regions outside the HD-Zip domain responsible for the functional diversity.

This motivated the search for potentially functional regions of the protein outside the HD-Zip domain. A total of 178 HD-Zip I proteins from different species were analyzed, including some species with complete genome sequences. A phylogenetic reconstruction and the scrutiny of their carboxy-terminal regions (CTRs) allowed the definition of six monophyletic groups with high similarity in these regions.

The search for conserved motifs in the CTRs indicated the presence of group-specific sequences and of a sequence in the end of most of the proteins with the characteristics of the AHA transcriptional-activation motif. Yeast simple hybrid assays demonstrated that the AHA motif of the AtHB13 TF has transactivation capability.

Other motifs in the CTRs exhibited putative target sites for two post-translational modifications, phosphorylation and sumoylation. Although limited, there is evidence that these modifications take place in HD-Zip I TFs. Therefore, the results suggest that these mechanisms regulate most of the proteins from this subfamily, although there would be differences among groups.

To test for the functionality of the CTR *in planta*, a chimeric protein was expressed in *A. thaliana* which consisted of the HD-Zip domain of HaHB4 fused to the CTR of HaHB1, using the *35SCaMV* promoter. Both sunflower FTs have been extensively studied and plants expressing wild-type versions under the control of the *35SCaMV* promoter were available. Characteristic phenotypes conferred by each wild-type protein were evaluated: the insensitivity to ethylene generated by the expression of HaHB4 and the increase in leaf serration produced by HaHB1

expression. Plants expressing the chimeric protein displayed a significant rise in leaf serration in one of the lines and they were slightly insensitive to ethylene, indicating that the identity of the CTR had an important effect on the phenotype generated.

As a whole, these results indicate that FTs HD-Zip I would be capable of transcription activation through the AHA motif and their CTRs would be regulated by phosphorylation and sumoylation. Consequently, the regions outside the HD-Zip domain would have an important role in determining the signaling pathways in which the TFs from this family act. On the one hand, the diversity of AHA motifs could be responsible for a differential interaction with the basal transcriptional machinery; on the other hand, the regulatory mechanisms could be different among proteins of the subfamily.

To study the signaling pathways in which the HD-Zip proteins are involved, techniques for the reverse engineering of transcriptional regulatory networks were applied. Using an algorithm based on information theory, the expression of 9617 genes from 269 microarray experiments was analyzed. They involved different abiotic stress treatments of *A. thaliana* plants and allowed obtaining the potential target genes for 25 HD-Zip I and II proteins.

The promoters of these target gene sets, called modules of transcriptional activity (MTAs), were analyzed searching for the motif CAATNATTG. This sequence was not enriched in relation to the proportion existing in the promoters of the complete set of genes analyzed, suggesting that similar motifs could also be recognized with enough affinity to establish functional binding. *De novo* search of motifs revealed the presence of the ABRE element in several promoters of genes belonging to the MTAs. This motif is recognized by TFs involved in ABA responses which could indicate that HD-Zip TFs and these proteins potentially act in coordination to regulate gene expression.

The possible crosstalk between HD-Zip signaling pathways was tested evaluating the degree of overlap between MTAs. The results showed that the overlap would be significant only among some phylogenetically close genes, but would not be relevant in most cases.

From a global perspective, the gene expression profiles of the MTAs displayed, for some TFs, a distinctive difference in average expression levels between shoot and root. This suggests a potential difference in the way subgroups of genes from the MTAs act in each tissue, which could be a consequence of the action of the HD-Zip TF.

The proportion of genes from each MTA with coordinated expression in each stress condition was also investigated. Considering this parameter as an indicator of the degree of participation of the TF in the treatment, this analysis revealed that most of the HD-Zip I and II TFs would have an important activity in heat stress, a function not previously described for these proteins.

The inspection of the expression profiles of the coordinated genes showed that in some stress conditions there were subgroups of induced genes and subgroups of repressed genes. In spite of not being direct evidence, this could be associated with the proven capability of subfamily I TFs of activating and repressing transcription.

Finally, the enriched functional categories of the genes in the MTAs were investigated. Several TFs were related to functions previously discovered by other authors, supporting to some extent the sets of target genes found. In addition, other functional categories not previously described were associated with many TFs and could be the subject of future works.

In summary, this work allowed the proposal of the mechanisms of action and regulation that HD-Zip I would have, revealing the importance of previously uncharacterized regions of the proteins and suggesting potential new sources of functional diversity among members. The engineering of the transcriptional regulatory network enabled the definition of sets of putative target genes for proteins from subfamilies I and II. Different analyses of these sets provided a global perspective of the signaling pathways regulated by the TFs of both subfamilies; comparative and functional exploratory studies indicated the existence of common and particular features in the activity of these proteins.



## Agradecimientos

El presente manuscrito es la materialización última de un largo proceso. Su realización requirió de mucho esfuerzo y trabajo a cambio de un gran aprendizaje, en el sentido más amplio de la palabra. Pero este logro no fue alcanzado en soledad, por el contrario, son muchas las personas e instituciones que, en mayor o menor medida, han hecho un aporte imprescindible. Por esto deseo expresar mi más sincero agradecimiento.

A las Instituciones:

- A la Facultad de Bioquímica y Ciencias Biológicas de la Universidad Nacional del Litoral porque gracias a las oportunidades que me ha brindado realicé mi carrera de grado, me inicié en la docencia y he realizado el presente trabajo de Tesis.
- Al Consejo Nacional de Investigaciones Científicas y Tecnológicas por haberme posibilitado el desarrollo de la Tesis a través de las becas de posgrado que me fueron otorgadas.
- A la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) por los subsidios otorgados a los proyectos que incluyen el presente trabajo de Tesis.
- Al Instituto de Agrobiotecnología del Litoral (UNL/CONICET) porque desde su creación ha fortalecido la interacción entre los diferentes grupos de investigación que lo forman.
- Al Centro Internacional de Métodos Computacionales en Ingeniería (UNL/CONICET) y al Laboratorio de Bioinformática del Instituto Leloir por permitirme emplear sus recursos computacionales.

A mi directora y compañeros del grupo de trabajo:

- A la Dra. Raquel Chan, directora de esta Tesis, por brindarme la invaluable oportunidad de trabajar bajo su dirección. Por estar siempre presente, por su inagotable energía y entusiasmo, por su confianza y paciencia; y por todas sus enseñanzas sobre ciencia y sobre la carrera científica. Por su vocación, sus valores y su amistad.
- A Ariel Chernomoretz por abrirme las puertas de su laboratorio, por su disponibilidad y por tener siempre la mejor predisposición para responder mis inquietudes. Su participación, orientación y discusión de resultados fue indispensable en la realización del Capítulo IV del presente trabajo de Tesis. También deseo agradecer al resto de los integrantes del grupo por recibirme cálidamente en su laboratorio.
- A Jesi por su calidez, su empatía y su imperturbable buen humor. Por ayudarme en mi primera experiencia y participación en la formación de RRHH. Además, por su trabajo y dedicación en la obtención de los resultados de la Sección III.3.6 del presente trabajo de Tesis.

- A Juli por su alegría y entusiasmo, por sus consejos y por tantas charlas, mate de por medio. También por hacerme participe de su incansable labor científica.
- A Jorge (Toto) por los muchos buenos momentos compartidos discutiendo sobre ciencia y sobre cualquier otro tema que surgiera. Por su ilimitada curiosidad, conocimiento e intereses.
- A Carlos por las muchas enseñanzas, charlas y consejos, por todas las vivencias dentro y fuera del laboratorio.
- A Fede por todas las charlas e intereses mutuos. Por su admirable energía y la alegría con la que emprende todos sus proyectos, que sabe contagiar.
- A Pablo por su empuje, su inigualable humor y predisposición para pasar buenos momentos. También por sus valiosos consejos.
- A Delfi por su buen humor y predisposición que hacen más alegre y fácil el trabajo en grupo.
- A Mati por su compañerismo y por la gran dedicación y compromiso con los que aborda cada actividad.
- A Kari por estar siempre dispuesta a dar una mano, por las muchas charlas compartidas sobre ciencia y sobre la vida.
- A Flavia por su buena predisposición y lo ameno que hace el trabajo diario.
- A Ariel por su entusiasmo, interés y buena onda.
- A Vero por su valioso trabajo diario y por la ayuda que nos brinda. A Seba por su disponibilidad para resolver problemas y su cordialidad. A Pablo D.D. por su colaboración y su simpatía.

A los compañeros de la Facultad y de otros ámbitos académicos:

- Al Dr. Martín Crespi, a Philippe Laporte y al grupo de Gif-Sur-Yvette por recibirme en su laboratorio y ayudarme en todo lo que necesité durante mi estadía.
- A los docentes de la cátedra por el trabajo compartido.
- Al Dr. Daniel González por sus opiniones y consejos.
- A Eli, Ivi, Caro, Raulo, Edu y Grace por los momentos compartidos, los muy buenos recuerdos y la ayuda que siempre me han brindado. A Fran, Luli, Lea y Nata por su entusiasmo, compromiso y buena predisposición.

- A Daniel, Carlos F., Matías A., Lucas, César, Cristian, Carlitos R. y Diego por los muchos asados, partidos y reuniones que hemos compartido y disfrutado enormemente.
- A Vicky, Jesi, Ludmi, Aye, Nori, Álvaro y Flor por la amistad que nos une y por las innumerables experiencias vividas.
- Al Dr. Mario Storti por su ayuda en el uso de las computadoras de alto rendimiento del CIMEC.

A mis amigos:

- A Sergio, JuanMa y Raúl Hermosi por las charlas, discusiones, consejos y experiencias que son parte de quien soy.
- A Raúl Hernández, Rodrigo, Iván, Fede N., Nico, Pablo O., Seba, Pablo S., Virgi y Diego por su amistad y todos los buenos e inolvidables momentos vividos.

A mi familia:

- A la familia “política” por parte de Palu y por parte de mis hermanos (Nadia y Ale) que me ha recibido con los brazos abiertos y que ya considero mía.
- A mis abuelos, tíos y primos por todo su amor y su apoyo.
- A mis padres Claudia y Roberto, y a mis hermanos Sofi y Mauri. Porque les debo *todo*, a pesar de que no pidan nunca nada a cambio.
- A mi amor, Palu. Porque ya dejé de buscar, te encontré. . .

## Antecedentes

El trabajo que se describe en esta Tesis para optar al título de Doctor en Ciencias Biológicas ha sido realizado en el Instituto de Agrobiotecnología del Litoral (CONICET-UNL) ubicado en la Facultad de Bioquímica y Ciencias Biológicas de la Universidad Nacional del Litoral. Los resultados que se presentan han sido dados a conocer en forma parcial en las siguientes publicaciones:

Publicaciones relacionadas directamente con el trabajo de tesis:

- Arce, A.L.; Raineri, J.; Capella, M.; Cabello, J.V. y Chan, R.L. (2011) *Uncharacterized conserved motifs outside the HD-Zip domain in HD-Zip subfamily I transcription factors; a potential source of functional diversity*. BMC Plant Biology, 11(1), 42.
- Arce, A.L. y col. (2012) *Analysis of transcriptional regulatory networks of Arabidopsis HD-Zip I and II transcription factors*. Manuscrito en preparación.

Publicaciones relacionadas indirectamente con el trabajo de Tesis:

Las siguientes publicaciones no contienen resultados volcados en este manuscrito de Tesis. Sin embargo, las herramientas y conocimientos adquiridos por la participación en estos trabajos fueron aplicados a la obtención de esos resultados.

- Manavella, P.A.; Arce, A.L.; Dezar, C.A.; Bitton, F.; Renou, J.P.; Crespi, M. y Chan, R.L. (2006) *Cross-talk between ethylene and drought signalling pathways is mediated by the sunflower Hahb-4 transcription factor*. The Plant Journal, 48(1), 125–137.
- Arce, A.L.; Cabello, J.V. y Chan, R.L. (2008) *Patents on plant transcription factors*. Recent Patents on Biotechnology, 2(3), 209-217.
- Cabello, J.V.; Arce, A.L. y Chan, R.L. (2010). *Methods and compositions for stress tolerance in plants*. Plant Bioscience Limited-CONICET-UNL. WO/2010/139993.
- Cabello, J.V.; Arce, A.L. y Chan, R.L. (2012). *The homologous HD-Zip I transcription factors HaHB1 and AtHB13 confer cold tolerance via the induction of pathogenesis-related and glucanase proteins*. The Plant Journal, 69(1), 141–153.
- Arce, A.L.; Giacomelli, J.I.; Ribichich, K.F. y Chan, R.L. (2012) *Sunflower atypical transcription factors and miRNAs playing a key role in responses to abiotic stresses*. 18<sup>th</sup> International Sunflower Conference. Mar del Plata & Balcarce, Buenos Aires, Argentina.

# Capítulo I

## Introducción general

*“La vida parece ser el comportamiento ordenado y reglamentado de la materia, que no está asentado exclusivamente en su tendencia de pasar del orden al desorden, sino basado en parte en un orden existente que es mantenido”*

*Erwin Schrödinger*

### I.1 La vida: complejidad y ordenamiento

¿Cuál es el origen de la vida? Éste es, tal vez, uno de los interrogantes fundamentales de la biología que más debates y controversias ha generado. La historia de las teorías que se han propuesto está naturalmente asociada a las definiciones mismas de vida o ser vivo imperantes en el momento. A su vez, estas definiciones han sido determinadas no sólo por los paradigmas reinantes en la biología en general, sino en la física y en la química. Pero además, por sus obvias implicaciones más allá del ámbito de las ciencias naturales, estas definiciones han sido igualmente determinadas por la filosofía y los sistemas de creencias prevalecientes.

En el ya clásico libro “¿Qué es la vida?” (1944), Erwin Schrödinger propone que los seres vivos escapan a la tendencia de la materia a pasar a un estado de mayor desorden. Esta aparente contradicción con la segunda ley de la termodinámica está fundamentalmente asociada a que los seres vivos son sistemas abiertos que intercambian energía y materia (e información) con el medio, generando “orden del desorden” que produce su metabolismo.

Trasladando entonces esta idea al origen de la vida y su evolución, podemos reconocer que los organismos más complejos o “sofisticados”, cuya aparición en la tierra es más reciente, poseen un mayor “grado de ordenamiento”<sup>1</sup>. Desafortunadamente, no existiendo un consenso acerca de una medida o parámetro que permita cuantificar el grado de complejidad de los seres vivos (complejidad entendida como complejidad funcional o estructural, no algorítmica; Adami, 2002), se torna difícil dar rigurosidad a esta afirmación para cualquier caso. Independientemente de esto, si reducimos el análisis a casos extremos, como las transiciones que se describen a continuación, este aumento en la complejidad se pone en clara evidencia.

### I.2 Evolución de los niveles de organización

Maynard Smith y Szathmáry (1995) caracterizaron lo que definen como las principales transiciones en el curso de la evolución (Tabla I.1). Estas transiciones fueron seleccionadas esencialmente en función de los cambios en la forma de transmisión de la información de una generación a la siguiente. A modo de ejemplo, si las moléculas se replicaban originalmente en forma independiente, su asociación en cromosomas generó una replicación coordinada; a otro nivel, la aparición de los organismos pluricelulares dio lugar a diferentes tipos celulares y a la división celular con funciones especializadas, crecimiento o reproducción.

La aparición de estos niveles de organización a lo largo de la evolución sustenta, sin lugar a dudas, que ha habido un incremento en la complejidad y el grado de ordenamiento de los seres vivos y las entidades que los precedieron durante la etapa de evolución prebiótica.

Una particularidad que resulta de estas transiciones, más allá de algunas excepciones, es la

---

<sup>1</sup>Sin que esto implique una mejor adaptación al medio en términos darwinianos.

## I. Introducción general

---

**Tabla I.1:** Principales transiciones en la evolución (adaptado de Maynard Smith y Szathmáry (1995)).

Moléculas con capacidad de replicación	→	Poblaciones de moléculas en compartimentos
Moléculas con replicación independiente	→	Cromosomas
ARN como gen y enzima	→	ADN + proteínas (código genético)
Procariotas	→	Eucariotas
Clones asexuales	→	Poblaciones sexuales
Protistas	→	Animales, plantas y hongos (diferenciación celular)
Individuos solitarios	→	Colonias (castas sin capacidad reproductiva)
Sociedades de primates	→	Sociedades humanas (lenguaje)

imposibilidad de las entidades de menor nivel de replicarse independientemente una vez que forman parte de la entidad de nivel superior. Maynard Smith y Szathmáry (1995) consideran que el principal factor que otorga una ventaja evolutiva en las etapas tempranas de la transición es el elevado grado de relación o similitud genética entre las entidades. Por otro lado, el establecimiento del nuevo nivel de organización podría ser favorecido por otros dos procesos:

- Irreversibilidad producto de la contingencia: es la pérdida de la capacidad de replicación independiente consecuencia de mecanismos secundarios no necesariamente relacionados a las presiones selectivas que hacen que la transición sea evolutivamente favorable.
- Control central: son diferentes mecanismos que favorecen la coexistencia de las entidades de menor nivel otorgando una ventaja evolutiva.

Adicionalmente, las transiciones son seguidas generalmente de una división de tareas o funciones entre las entidades de menor nivel. La especialización permite mejoras en el desempeño y por lo tanto en la aptitud para la supervivencia. Por ejemplo, en los organismos procariotas todas los procesos ocurren en un mismo compartimento, mientras que en los eucariotas hay compartimentos especializados en los que se desarrollan diferentes funciones celulares. La asociación en colonias de los insectos eusociales es otro ejemplo que ilustra claramente este fenómeno ya que la división de tareas entre individuos involucra inclusive una de las funciones esenciales de los organismos, la reproducción.

### I.3 Los factores de transcripción

Durante la etapa de evolución prebiótica, una de estas transiciones implicó el pasaje de un “mundo de ARN”, en el que el ARN cumplía simultáneamente la función de contener la

“información genética” y de poseer actividad enzimática, al sistema presente en los organismos actuales, en el que el ADN contiene la información genética y las proteínas ejercen la actividad enzimática (Maynard Smith y Szathmáry, 1995). Esta división de tareas requirió de la aparición de mecanismos moleculares que permitieran el flujo de la información. Es así como nació lo que hoy es conocido como el “Dogma central de la Biología” (Crick, 1970), que establece que la síntesis de una proteína requiere una primera etapa de *transcripción* del ADN a ARN mensajero (ARNm) y una segunda de *traducción* de éste a proteína.

Más precisamente, la transcripción es el proceso mediante el cual una enzima ARN polimerasa (ARN Pol) sintetiza ARN copiando una hebra molde de ADN, lo que da origen a los cuatro tipos principales de ARN: el ARNm, el ARN ribosomal (ARNr), el ARN de transferencia (ARNt) y los pequeños ARNs (entre ellos los ARNmi).

Los factores de transcripción (FTs) son proteínas que regulan principalmente el inicio de este proceso y su origen podría hallarse en etapas tempranas posteriores a la transición de este “mundo de ARN” al de ADN y proteínas. Representan un ejemplo de control central, posibilitando la inducción o represión coordinada de diferentes genes. Además, su aparición constituye un caso de división de tareas, permitiendo una mayor versatilidad en el control sobre la expresión a lo largo del desarrollo y en respuesta a estímulos internos o externos.

### I.3.1 La transcripción en procariotas y eucariotas

Los FTs fueron descubiertos por Jacob y Monod (1961) en organismos procariotas. En *Escherichia coli* la ARN Pol es multimérica, siendo la subunidad  $\sigma$  variable y fundamental en su función, ya que es la responsable de la especificidad de unión al ADN (Ptashne y Gann, 2002). En esta bacteria se han reconocido tres mecanismos principales de regulación de la transcripción:

- El reclutamiento regulado: FTs activadores se encargan de reclutar la ARN Pol con la subunidad  $^{70}\sigma$  a los genes que serán transcritos. A la acción de estos activadores se puede oponer la actividad de proteínas represoras que generalmente evitan la unión de la ARN Pol al promotor.
- La activación de la polimerasa: la ARN Pol con la subunidad  $^{54}\sigma$  es capaz de reconocer los promotores en forma independiente, pero en un estado inactivo. La enzima pasa a la forma activa mediante cambios alostéricos inducidos por una proteína activadora.
- La activación del promotor: la ARN Pol con la subunidad  $^{70}\sigma$  también se une en forma independiente e inactiva a un pequeño grupo de promotores, pero en este caso la activación depende de que una proteína específica induzca cambios en el ADN del promotor, lo que permite entonces el inicio de la transcripción.

En organismos eucariotas también hay varias ARN polimerasas, siendo la ARN Pol II la encargada de sintetizar los ARNm, es decir, transcribir genes que codifican proteínas (aunque



también transcribe genes de ARNmi; Jones-Rhoades y col., 2006). Sin embargo, en eucariotas la transcripción requiere una cantidad de proteínas muy superior a la de procariotas. Adicionalmente, el ADN eucariota se asocia a histonas por lo que puede ser *a priori* inaccesible a la maquinaria transcripcional (Venters y Pugh, 2009).

El paso inicial en la transcripción de cualquier gen es la formación del complejo de preiniciación (CPI); éste es un gran complejo proteico que incluye a la ARN Pol II, a los factores de transcripción generales y al complejo mediador (Figura I.1; Venters y Pugh, 2009). Los factores de transcripción generales incluyen a TFIIA, -B, -D, -E, -F, y -H; y participan en la maquinaria basal de transcripción de la gran mayoría de los genes, reclutando la ARN Pol II y permitiendo que comience la polimerización a partir del sitio de inicio de la transcripción. De ellos, TFIID es a su vez un complejo importante ya que contiene a la proteína de unión a la caja TATA (TBP), que funciona como nucleador en el ensamble del CPI, y a los factores asociados a TBP (Venters y Pugh, 2009).

Respecto al complejo mediador todavía no se conoce con certeza si es indispensable para la formación del CPI. Sin embargo, se ha demostrado que participa en los CPIs de una gran cantidad de genes y posee al menos dos funciones importantes: reclutar proteínas modificadoras de la cromatina y así facilitar el acceso de la maquinaria transcripcional, e interactuar directamente con algunos FTs sirviendo de este modo de “puente” entre éstos y la maquinaria basal de transcripción (Kim y Chen, 2011). Se han hallado otras proteínas con la capacidad de actuar como cofactores posibilitando entonces la interacción indirecta entre el FT y la maquinaria basal de transcripción (Liu y Karmarkar, 2008; Wray y col., 2003).

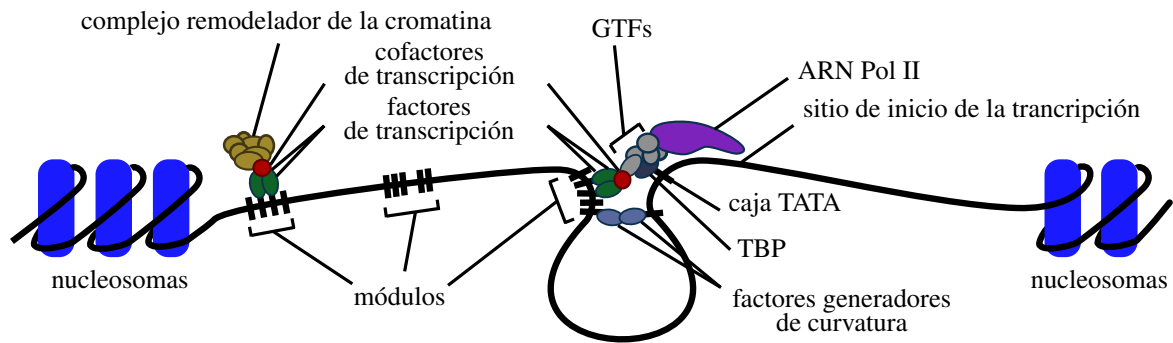
Una vez que el CPI se ha formado, la fosforilación de residuos específicos en el dominio carboxilo terminal de la ARN Pol II sirve de señal para que la transcripción comience (Venters y Pugh, 2009). Las etapas que siguen, de elongación y terminación, poseen mecanismos particulares de regulación, sin embargo, el principal mecanismo que controla la identidad del gen que será transcrito es el reclutamiento de la maquinaria transcripcional a cargo de los FTs (Venters y Pugh, 2009).

### I.3.2 Clasificación de los factores de transcripción

Los FTs son proteínas capaces de reconocer y unir secuencias específicas de ADN, denominadas elementos reguladores en *cis*, presentes en las regiones promotoras de los genes cuya expresión activan o reprimen. La ubicación de los genes que codifican FTs no posee relación con la ubicación de los elementos que éstos unen, por lo que los FTs son elementos de acción en *trans*.

Estas proteínas suelen poseer una estructura modular que exhibe, en general, al menos dos tipos de dominios: un dominio capaz de reconocer y unir el ADN y un dominio de interacción proteína-proteína que media, directa o indirectamente, la activación o represión de la transcripción

## I. Introducción general



**Figura I.1:** Esquema de la maquinaria basal de transcripción (adaptado de Wray y col., 2003). No se ilustra el complejo mediador pero su actividad está representada en el complejo remodelador de la cromatina y en los cofactores de transcripción que participan de las interacciones entre los FTs y la maquinaria basal de transcripción. Los factores generadores de curvatura están encargados de alterar la topología del ADN y los módulos representan conjuntos de elementos *in cis*.

(Brivanlou y Darnell, 2002; Frankel y Kim, 1991; Riechmann, 2000).

La modularidad de los FTs puede ilustrarse mediante el caso de la proteína de levaduras GAL4. Si su dominio de unión a ADN y su dominio de activación son expresados en forma independiente dentro de una célula como fusiones a otras proteínas, basta que éstas interactúen para que se forme un dímero funcionalmente análogo al FT GAL4 salvaje. En otras palabras, se forma un dímero con capacidad de reconocer la misma secuencia de ADN y de activar la transcripción. Éste es el fundamento de la técnica de doble híbrido en levaduras, y el mecanismo puede extenderse a pares de dominios de unión a ADN y de activación provenientes de proteínas diferentes (Ptashne y Gann, 2002).

Existe un repertorio limitado de dominios de unión a ADN, lo cual permite clasificar los FTs en diferentes superfamilias (Riechmann, 2000). A su vez son subdivididos en familias y subfamilias en base a características estructurales y funcionales, principalmente asociadas a particularidades de estos dominios o a la presencia de dominios adicionales u otros motivos conservados en el FT. Estas clasificaciones suelen estar respaldadas por análisis filogenéticos (Ariel y col., 2007a; Gong y col., 2004; Heim y col., 2003; Hosoda y col., 2002; Toledo-Ortiz y col., 2003; Zhang y Wang, 2005).

En plantas se han identificado numerosas familias de FTs, pero sólo una fracción relativamente pequeña ha sido caracterizada funcionalmente (Gong y col., 2004; Riechmann, 2000). Esta identificación ha sido realizada esencialmente en plantas cuyo genoma está secuenciado, como *Arabidopsis thaliana* u *Oryza sativa*, para las cuales una comparación con FTs animales indica la existencia de entre 1500 y 2000 proteínas con esta función, de acuerdo a los criterios empleados (Mitsuda y Ohme-Takagi, 2009; Riechmann, 2000; Xiong y col., 2005). Este número representa, por ejemplo, alrededor de un 6 % del total de genes del genoma de *A. thaliana*, proporción sensiblemente superior a la observada en otros organismos como *Drosophila melano-*

*gaster*, *Caenorhabditis elegans* y *Saccharomyces cerevisiae* (4.5, 3.5, y 3.5 %, respectivamente; Riechmann, 2000). Esto podría sugerir que la regulación transcripcional en *A. thaliana* y otras plantas es más compleja o tiene una participación más importante en la regulación génica.

El gran número de genes que codifican FTs en plantas está asociado, en parte, a que muchas de las familias también presentes en organismos de otros reinos poseen un número muy superior de miembros. Por ejemplo, la familia MADS posee 2, 2 y 4 miembros en *D. melanogaster*, *Caenorhabditis elegans* y *Saccharomyces cerevisiae*, respectivamente; mientras que tiene más de 100 miembros en *A. thaliana*. Además, hay varias familias de FTs exclusivas de plantas, ya sea porque poseen un dominio de unión a ADN que ha sido hallado solamente en proteínas de este reino (por ejemplo, las familias AP2/EREBP, NAC, TCP y WRKY), o porque dominios no específicos de plantas se presentan en combinaciones exclusivas de proteínas de este reino (Riechmann, 2000).

## I.4 La familia de factores de transcripción HD-Zip

### I.4.1 Primeros miembros descubiertos

Los factores de transcripción con el dominio de unión a ADN homeodominio (HD) fueron descubiertos en plantas cuando se logró identificar y secuenciar el gen *Knotted-1* de maíz (Vollbrecht y col., 1991). Sus alelos mutantes se caracterizaban por generar en la planta un desarrollo alterado de las hojas. El mismo año se publicó el descubrimiento en *A. thaliana* de dos factores de transcripción con HD, AtHB1 (*Arabidopsis thaliana* homeobox 1) y AtHB2. Éstos poseían además un dominio de cierre de leucinas (Zip o LZ) inmediatamente adyacente hacia el extremo carboxilo terminal del HD (Ruberti y col., 1991). Esta asociación entre un HD y un LZ (también llamado HALZ por LZ asociado a un HD) es exclusiva de proteínas del reino vegetal y permitió definir la nueva familia HD-Zip de FTs (Ariel y col., 2007a; Ruberti y col., 1991). Cabe destacar que en 1992 no estaba secuenciado el genoma de *A. thaliana* ni de ninguna otra planta, por lo cual el reconocimiento de estos dominios y su aislamiento posterior se sustentó en la homología con los HD animales.

Ya en el trabajo de Ruberti y col. (1991) se reconoció que la presencia del HD indicaba que AtHB1 y AtHB2 eran potenciales FTs. Esta hipótesis se basaba en evidencia preexistente del estudio de proteínas que presentaban el mismo dominio halladas en una gran variedad de especies animales y de hongos. Además, el alineamiento del LZ con el dominio de cierre de leucinas de los FTs de la familia bZip (región básica-cierre de leucinas) les permitió proponer que este dominio permitía que los HD-Zip dimerizaran, como lo hacían los FTs bZip. De hecho, la posición relativa entre el HD y el LZ es análoga a la del cierre de leucinas y la región básica de unión a ADN de los FTs bZip (Ruberti y col., 1991).

Posteriormente se hallaron los genes *HAT22* (Schena y Davis, 1992), *AtHB3* (Mattsson y col.,

1992) y *AtHB4* (Carabelli y col., 1993), también de *A. thaliana*. El primer FT HD-Zip hallado en una especie diferente a *Arabidopsis thaliana* fue *HAHB-1* de girasol (Chan y Gonzalez, 1994).

### I.4.2 Las cuatro subfamilias de factores de transcripción HD-Zip

Basándose únicamente en el grado de similitud de a pares entre los dominios HD-Zip de los pocos miembros hallados al momento, Carabelli y col. (1993) reconocieron la existencia de dos clados claramente identificables de proteínas HD-Zip y los llamaron HD-Zip I y II. Posteriormente se descubrieron *Glabra2* (Rerie y col., 1994) y *AtHB8* (Baima y col., 1995), los primeros miembros identificados de las subfamilias HD-Zip IV y III, respectivamente.

Una vez secuenciado el genoma de *A. thaliana* pudo reconocerse que la familia HD-Zip contiene 48 miembros en esta especie. El total de sus miembros puede subdividirse en las cuatro subfamilias propuestas en función de las siguientes características (Ariel y col., 2007a):

1. conservación en el dominio HD-Zip (característica suficiente para clasificar las proteínas mediante un análisis filogenético).
2. estructura génica.
3. presencia de motivos o dominios conservados adicionales.
4. función.

Basándose en la presencia de otros dominios o motivos, las proteínas de cada subfamilia quedan caracterizadas de la de la siguiente forma (Figura I.2, Ariel y col., 2007a; Mukherjee y col., 2009):

- HD-Zip I: poseen el dominio HD-Zip únicamente. No se han reconocido otros dominios o motivos específicos.
- HD-Zip II: además del dominio HD-Zip, presentan en la región carboxilo terminal (RCT) el motivo CPSCE involucrado en la regulación redox de estas proteínas (Tron y col., 2002), y en la región amino terminal (RNT) el motivo ZIBEL de 10 aminoácidos hallado también en los FTs con HD de la familia BEL (Mukherjee y col., 2009).
- HD-Zip III: poseen un motivo CPSCE (Mukherjee y col., 2009; Tron y col., 2002), el dominio START (*steroidogenic acute regulatory protein-related lipid transfer*) y el dominio asociado a START (HD-SAD). En el extremo carboxilo terminal presentan el dominio llamado MEKHLA (Mukherjee y Bürglin, 2006). Tienen además una inserción de cuatro aminoácidos entre las hélices II y III del HD que no está presente en proteínas de las otras tres subfamilias HD-Zip (Mukherjee y col., 2009).

## I. Introducción general

- HD-Zip VI: poseen, al igual que los FTs HD-Zip III, los dominios START y HD-SAD; pero carecen del dominio MEKHLA. Además presentan un motivo conservado CESVV potencialmente relacionado al motivo CPSCE presente en proteínas HD-Zip II y III.

### HD-Zip I



### HD-Zip II



### HD-Zip III



### HD-Zip IV



100 aa

**Figura I.2:** Dominios y motivos característicos presentes en los FTs de las cuatro subfamilias HD-Zip (adaptado de Ariel y col., 2007a; Mukherjee y col., 2009).

La composición resultante de las subfamilias en *A. thaliana* es la siguiente: 17 miembros en la subfamilia I, 10 miembros en la subfamilia II, 5 miembros en la subfamilia III y 16 miembros en la subfamilia IV. Como se ilustra en la Tabla I.2, la composición relativa de las subfamilias es similar entre especies filogenéticamente distantes con genoma secuenciado, más allá de ciertas variaciones, habiendo una notable diferencia con la composición de la licopodiofita *Selaginella moellendorffii*.

**Tabla I.2:** Composición de la familia HD-Zip en distintas especies (Adaptado de Mukherjee y col., 2009).

Subfamilia	Eudicotiledóneas		Moncotiledóneas		Licopodiofitas	Briofitas
	<i>At</i>	<i>Pt</i>	<i>Os</i>	<i>Zm</i>	<i>Sm</i>	<i>Pp</i>
I	17	22	14	24	4	17
II	10	16	14 (2)	17	2	7
III	5	8	7 (3)	8	3	5
IV	16	15	12	21	4	4

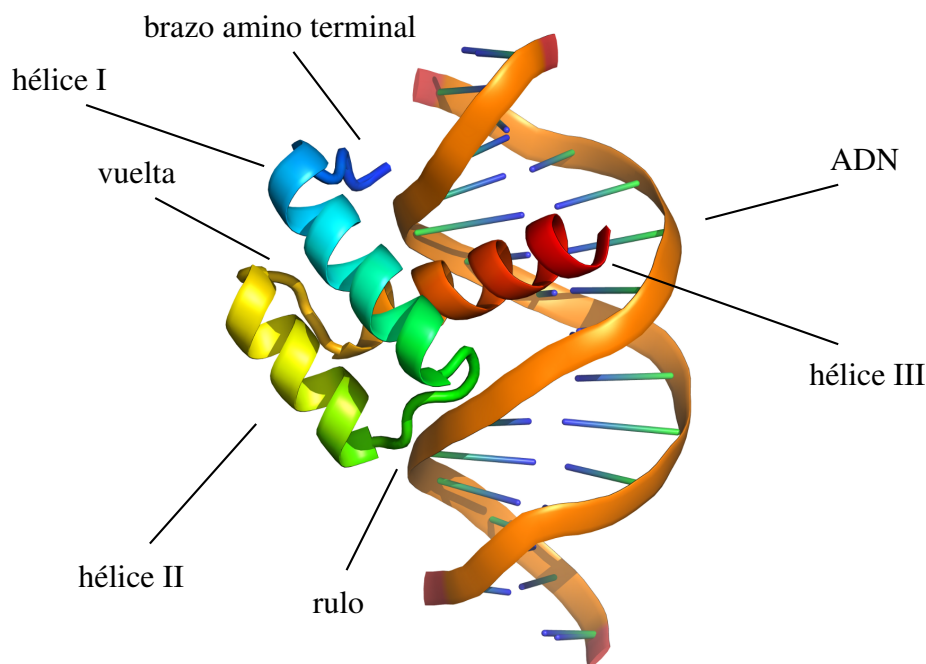
*At*: *Arabidopsis thaliana*, *Pt*: *Populus trichocarpa*, *Os*: *Oryza sativa* (pseudogenes putativos entre paréntesis), *Zm*: *Zea mays*, *Sm*: *Selaginella moellendorffii* y *Pp*: *Physcomitrella patens*.

Debido a que el presente trabajo se focaliza en las subfamilias I y II, el desarrollo de las siguientes secciones está dirigido especialmente a lo que se conoce respecto a los miembros de estas dos subfamilias.

### I.5 Los HD-Zip I y II a nivel molecular

#### I.5.1 El homeodominio y su interacción con el ADN

El dominio HD contiene alrededor de 60 aminoácidos (aa) y es codificado por una secuencia nucleotídica denominada caja homeótica u *homeobox* (*HB*, Ariel y col., 2007a). La estructura consiste en tres  $\alpha$ -hélices (ver Figura I.3); las hélices II y III están unidas por una vuelta formando un motivo similar al de hélice-vuelta-hélice encontrado en varios FTs de organismos procariotas (Palena y col., 1999). La hélice III, o hélice de reconocimiento, se inserta en el surco mayor del ADN e interacciona específicamente con diferentes bases y con el esqueleto azúcar-fosfato. Las hélices I y II, unidas por un rulo, se ubican en forma antiparalela y a su vez perpendicular a la hélice III. El extremo amino terminal del dominio (aproximadamente nueve aa) forma un brazo desordenado que contacta el ADN ubicándose en el surco menor; esta interacción es esencial para lograr una unión de elevada afinidad *in vitro* (Palena y col., 2001).



**Figura I.3:** Estructura de un homeodominio interaccionando con el ADN (imagen generada con fines ilustrativos a partir la estructura de un HD de *Drosophila melanogaster* uniendo el ADN, <http://www.rcsb.org/pdb/explore/explore.do?structureId=1FJL>).

A pesar de que la mayor parte de las proteínas con HD son capaces de unir ADN como monómeros, los FTs HD-Zip sólo pueden hacerlo *in vitro* luego de dimerizar, lo que logran a través del LZ (Gonzalez y col., 1997; Palena y col., 1997, 1999; Sessa y col., 1993, 1997). Esta diferencia está asociada fundamentalmente a la estructura del rulo entre las hélices I y II. El HD de HaHB4 de girasol fue capaz de unir ADN *in vitro* en forma monomérica tan sólo sustituyendo

este rulo por el de ENGRAILED de *D. melanogaster*, un FT con esta capacidad (Tron y col., 2004).

Diferentes ensayos de SELEX (evolución sistemática de ligandos por enriquecimiento exponencial) y de unión a oligonucleótidos en geles de retardo, con o sin competencia, permitieron determinar las secuencias reconocidas *in vitro* por los FTs HD-Zip. Sessa y col. (1993) identificaron por SELEX que el HD-Zip I AtHB1 reconocía con máxima afinidad la secuencia pseudopalindrómica de nueve pares de bases (pb) CAAT(A/T)ATTG, mientras que el HD-Zip II AtHB2 reconocía prácticamente la misma secuencia, pero con una preferencia por una G/C en lugar de una A/T en la posición cinco o central. De igual manera se identificaron los motivos unidos por HaHB4 (Palena y col., 1999), AtHB5 (Johannesson y col., 2001) de la subfamilia I, resultando en la misma secuencia CAAT(A/T)ATTG. El aislamiento y la expresión de seis proteínas de arroz, tres de cada subfamilia, permitió observar en ensayos de retardo en geles (EMSA) una preferencia por la posición central A/T para las proteínas de la subfamilia I, y G/C para las de la subfamilia II (Meijer y col., 2000). Estos resultados concordaron con lo observado previamente para la proteína OsHOX1, otro HD-Zip II de arroz (Meijer y col., 1997).

Johannesson y col. (2001) propusieron que en *A. thaliana* esta preferencia por la posición central no se extiende a todos miembros de la subfamilia I. Por medio de competencias, hallaron que AtHB5, AtHB6 y AtHB16 reconocen el mismo pseudopalíndromo con elevada afinidad pero son insensibles a la identidad de la base en esta posición, mientras que AtHB3 y AtHB13 sí prefieren A/T en esta posición. Aunque también evaluaron la interacción de AtHB7 y AtHB12, no lograron unión de AtHB7 a ningún pseudopalíndromo y sólo una débil unión de AtHB12 sin preferencias por la posición central. Esto demostró, sin embargo, la capacidad de casi todos estos miembros de unir una secuencia extremadamente similar.

La capacidad de unión de diferentes FTs a estos motivos fue comprobada también *in vivo*. Varios estudios utilizaron promotores artificiales con repeticiones de los motivos dirigiendo la expresión de genes reporteros. De esta forma, la coexpresión del FT y la regulación del gen reportero permitieron verificar la unión tanto en ensayos en levaduras (Meijer y col., 2000), como en ensayos en células vegetales (Aoyama y col., 1995; Henriksson y col., 2005; Himmelbach y col., 2002; Meijer y col., 1997). Recientemente la capacidad de unión de HB1 (MtHB1), una proteína HD-Zip I de *Medicago truncatula*, al motivo identificado por SELEX fue verificada *in planta* por medio de un ensayo de coinmunoprecipitación de la cromatina (ChIP, Ariel y col., 2010).

### I.5.2 El cierre de leucinas, dominio encargado de la dimerización

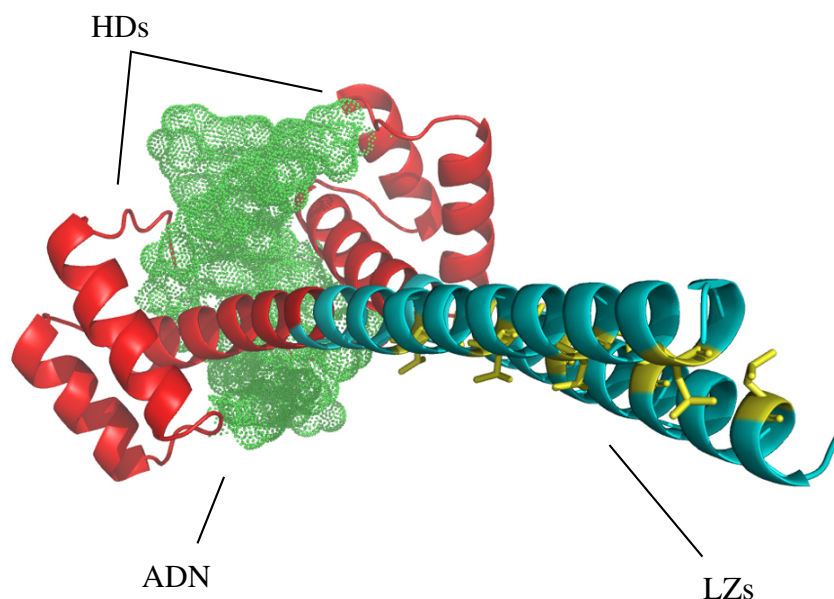
La dimerización es un requisito funcional para numerosos FTs de diversas familias del reino eucariota (Amoutzias y col., 2008). El LZ es uno de los dominios que median esta interacción y en plantas se presenta en los FTs de las familias bZip, bHLH (región básica-hélice-rulo-hélice),

## I. Introducción general

WRKY y HD-Zip (Amoutzias y col., 2008; Rushton y col., 2010).

A nivel estructural, el LZ es una  $\alpha$ -hélice anfipática que consiste en repeticiones no idénticas de siete aa (heptadas) en las que las posiciones de los residuos se nombran de la *a* a la *g* (Landschulz y col., 1988; para una revisión, ver Vinson y col., 2006). Las posiciones *a* y *d* (ocupadas mayoritariamente por leucinas, de ahí el nombre del dominio) son hidrofóbicas y generan un patrón de “prominencias y huecos” (del inglés, “*knobs and holes*”) a lo largo de la interfase de interacción, lo cual crea un núcleo hidrofóbico que estabiliza el dímero (Vinson y col., 2006). La estructura resultante, como ilustra el dímero de proteínas HD-Zip de la Figura I.4, es de hélices paralelas entrelazadas y es del tipo denominado *coiled-coil*.

En las proteínas HD-Zip el LZ posee alrededor de 42 aa (seis heptadas) y su importancia radica, por un lado, en que el HD no posee la afinidad suficiente para unir ADN en forma monomérica, entonces la dimerización aumenta la afinidad al aumentar el número de contactos con el ADN (Tron y col., 2004). Además, se ha demostrado que la sustitución de la primera leucina del LZ por una prolina o la inserción de dos aa entre el HD y el LZ, a pesar de no reducir sensiblemente la eficiencia de dimerización, eliminan la capacidad de unión del homodímero al motivo de nueve pb (Sessa y col., 1993). Esto indica que el LZ posiciona los HDs de modo que interactúen eficientemente con el elemento de ADN que reconocen.



**Figura I.4:** Estructura de un homodímero de proteínas HD-Zip interactuando con el ADN (adaptado de Ariel y col., 2007b). La imagen proviene de un modelado por homología obtenido empleando las estructuras conocidas por cristalografía de un HD unido al ADN y los LZs de un dímero de proteínas bZip.

Varios trabajos han demostrado que los FTs HD-Zip I son incapaces de heterodimerizar con los de la subfamilia II. Ensayos *in vitro* con AtHB1 y AtHB2 probaron que cuando se inmoviliza



uno o el otro, sólo son capaces de capturar una versión marcada de la misma proteína (Sessa y col., 1993). También se han realizado experimentos en *Escherichia coli*, en los que los LZ de las proteínas de girasol HaHB1 y HaHB10 fueron fusionados a una porción del represor  $\lambda$  incapaz de dimerizar. Esta quimeras poseen la actividad del represor salvaje gracias a que dimerizan a través del LZ, con lo cual la coexpresión de un LZ no fusionado al dominio del represor tiene un efecto dominante negativo sobre su actividad si hay interacción. Con esta estrategia se mostró *in vivo* que los LZ de estas proteínas sólo homodimerizan (Gonzalez y col., 1997).

Con la proteína de arroz OsHOX1 se probó que era capaz de homodimerizar y heterodimerizar con AtHB2 de *A. thaliana* en ensayos de doble híbrido en levaduras (DHL) (Meijer y col., 1997). Posteriormente se evaluó la capacidad de heterodimerización de varios miembros de las subfamilias I y II de arroz. Utilizando DHL no se produjo heterodimerización entre miembros de la subfamilia I con miembros de la II. Sin embargo, las tres proteínas HD-Zip II fueron capaces de heterodimerizar entre sí en ensayos de DHL, y por EMSAs se demostró la capacidad de heterodimerización para el único par de proteínas HD-Zip I evaluado (Meijer y col., 2000). También utilizando DHL pero con proteínas de *C. plantagineum*, se observó homo- y heterodimerización en tres de los cuatro FTs de la subfamilia I estudiados, y en los tres FTs de la II, pero no se registró heterodimerización entre proteínas de diferentes subfamilias (Deng y col., 2002; Frank y col., 1998).

La capacidad de heterodimerización dentro de la subfamilia I fue también estudiada en *A. thaliana* (Johannesson y col., 2001). Se realizaron ensayos en los que se inmovilizó AtHB5 y se evaluó la capacidad de capturar diferentes FTs HD-Zip I marcados radioactivamente. Se observó que poseía la capacidad de homodimerizar y de interactuar con AtHB6, AtHB7 y AtHB12. La interacción con AtHB16 fue muy débil y con AtHB1 prácticamente nula. Esto demostró que aunque habría heterodimerización entre AtHB5 y otros miembros de la subfamilia I, ésta no se observa con cualquiera de los demás miembros y en dos de los tres casos en que se produce, la afinidad sería claramente inferior a la del homodímero.

Es importante destacar que más allá de los resultados que sustentan la posibilidad de heterodimerización entre algunos miembros dentro de las subfamilias I y II *ex planta*, la relevancia funcional *in vivo* depende *a priori* de la coexpresión de los genes en la misma célula.

## I.6 Función de los factores de transcripción HD-Zip

Los FTs HD-Zip han sido ampliamente estudiados a nivel funcional más allá de los análisis moleculares y bioquímicos. Los primeros resultados acerca de la función de los HD-Zip provinieron del estudio de AtHB2 y AtHB4 (AtHB4), pertenecientes a la subfamilia II de *A. thaliana* (Carabelli y col., 1993). Los estudios de expresión génica mostraron que ambos genes se encontraban inducidos en oscuridad y particularmente en condiciones enriquecidas en luz de la región espectral denominada luz rojo lejano (RL, aproximadamente una longitud de onda entre

725 y 735 nm, Franklin, 2008). Estas observaciones llevaron a concluir que ambos genes podían estar involucrados en la respuesta de evasión de la sombra (Carabelli y col., 1993; Franklin, 2008).

Posteriormente se fueron encontrando funciones para otros miembros de las diferentes subfamilias (Ariel y col., 2007a) y, aunque se ha avanzado mucho en el entendimiento de los procesos en los que están involucrados estos FTs, todavía resta mucho por comprender acerca de su función y sus mecanismos de acción en las vías de señales en las que participan.

A pasar de que, como fue mencionado, esta tesis no profundizará sobre los miembros de las subfamilias HD-Zip III y IV, es relevante mencionar a grandes rasgos lo conocido acerca de su función. Los genes de la subfamilia III son principalmente reguladores del desarrollo y están involucrados en la regulación del meristema apical del tallo, de los haces vasculares y de los dominios adaxiales de los órganos laterales, mientras que algunos miembros han sido específicamente asociados al desarrollo vascular (Ariel y col., 2007a). Cabe destacar que en *A. thaliana* todos los HD-Zip III son regulados post-transcripcionalmente por ARNmi. Con respecto a los genes de la subfamilia IV, la mayoría ha sido relacionada a eventos del desarrollo que ocurren en la epidermis, entre ellos: el desarrollo de los tricomas, la acumulación de antocianina en la capa subepidermal de las hojas, el desarrollo de los pelos radiculares y la acumulación de aceites en las semillas (Ariel y col., 2007a; Shi y col., 2012).

### I.6.1 La subfamilia HD-Zip I

Las funciones de varios FTs de la subfamilia I de diferentes especies han sido caracterizadas y se ha reconocido la participación de estos FTs en una gran diversidad de procesos (ver Tabla I.3). Estos estudios han permitido definir algunas características generales en la función de los miembros de esta subfamilia.

#### I.6.1.1 Relación con las vías de señales de ABA y otras fitohormonas

Muchos de los miembros caracterizados están involucrados en las vía de señales de ABA o de respuesta a condiciones de estrés en las que participa esta fitohormona, como la deshidratación, el estrés osmótico y el estrés salino.

Dos de los primeros genes HD-Zip I que mostraron una función en estas vías fueron *AtHB7* y *AtHB12* (Lee y Chun, 1998; Söderman y col., 1996). La expresión de ambos es fuertemente inducida por ABA en la mayor parte de la planta como lo muestran los ensayos de *Northern blot* y los ensayos con sus promotores (Henriksson y col., 2005; Olsson y col., 2004). La inducción por ABA de *AtHB7* y *AtHB12* no se observa en las mutantes *abi1* y *abi2*. El estrés hídrico induce sus expresiones a través de esta hormona pues en una mutante en su síntesis, *aba-3*, no se verifica inducción. En ensayos de inhibición del crecimiento de la raíz con ABA, las plantas que expresaban *AtHB12* bajo la dirección del promotor 35S del virus del mosaico de la

**Tabla I.3:** Algunos de los principales genes HD-Zip I caracterizados funcionalmente con una síntesis de la función descripta o sugerida basándose en la evidencia experimental disponible.

Especie	HD-Zip	Función	Referencias
<i>A. thaliana</i>	<i>AtHB1</i>	Desarrollo del parénquima en empalizada. Desarrollo fotomorfogénico.	Aoyama y col., 1995
	<i>AtHB5</i>	Regulador positivo de la respuesta a ABA.	Johannesson y col., 2003; Söderman y col., 1994
	<i>AtHB6</i>	Regulador negativo de la respuesta a ABA. Interacciona con ABI1.	Söderman y col., 1999, 1994 Himmelbach y col., 2002; Lechner y col., 2011
	<i>AtHB7</i>	Fuerte inducción por ABA y déficit hídrico.	Olsson y col., 2004; Söderman y col., 1994
	<i>AtHB12</i>	Es dependiente de ABA a través de ABI1 y ABI2.	Hjellström y col., 2003; Söderman y col., 1996
		Fuerte inducción por ABA y déficit hídrico.	Lee y Chun, 1998; Olsson y col., 2004
		Es dependiente de ABA a través de ABI1 y ABI2.	Lee y col., 2001; Son y col., 2010
	<i>AtHB13</i>	Respuesta a sacarosa. Tolerancia a bajas temperaturas.	Cabello y col., 2012; Hanson y col., 2001, 2002
	<i>AtHB16</i>	Inicio de la floración y respuesta a luz azul.	Wang y col., 2003
	<i>AtHB20</i>	Sensibilidad a ABA en la germinación. Desarrollo vascular en hojas y cotiledones.	Barrero y col., 2010; Mattsson y col., 2003
		Desarrollo de hojas.	Kim y col., 2007
		Inductor de CAL en floración.	Saddic y col., 2006
<i>C. plantagineum</i>	<i>CpHB7</i>	Deshidratación. Regulador negativo de ABA.	Deng y col., 2006, 2002
<i>H. annuus</i> (girasol)	<i>HaHB1</i>	Respuesta a sacarosa. Tolerancia a bajas temperaturas.	Cabello y col., 2012; Chan y Gonzalez, 1994
	<i>HaHB4</i>	Respuesta a deshidratación, ABA y hervíboros. Integración de las vías de ET y JA.	Dezar y col., 2005a,b; Gago y col., 2002 Manavella y col., 2006, 2008a,b
<i>H. vulgare</i> (cebada)	<i>HvHOX1</i>	Regulación del desarrollo de las espiguillas.	Gottwald y col., 2009; Sakuma y col., 2010
<i>N. attenuata</i> (tabaco)	<i>NaHD20</i>	Integración de la respuesta a la deshidratación mediada por ABA con las transiciones florales.	Ré y col., 2011
<i>M. truncatula</i>	<i>MtHB1</i>	Regulación medida por ABA de la emergencia de raíces laterales en condiciones de estrés.	Ariel y col., 2010

Hormonas: ABA, ácido abscísico; Aux, auxinas; CK, citoquininas; ET, etileno; JA, ácido jasmónico.

## I. Introducción general

---

coliflor (*35SCaMV*) mostraron mayor sensibilidad. No se vio lo mismo en ensayos de inhibición de la germinación, respuesta también controlada por ABA (Olsson y col., 2004).

*AtHB6* también participa en la respuesta a ABA. Los resultados son similares a los obtenidos con *AtHB7* y *AtHB12*. Su expresión es inducida por esta hormona, por estrés hídrico y osmótico. La inducción por estrés es dependiente de ABA puesto que en la mutante *aba-3* no se observó. Los niveles de expresión son también dependientes de ABI1 y ABI2 (Söderman y col., 1999). Por otra parte, *AtHB6* fue recuperado en una búsqueda por DHL de proteínas que interactuaran con ABI1, lo que indica un nuevo nivel de regulación por ABA además al transcripcional (Himmelbach y col., 2002). Por otro lado, plantas *35SCaMV::AtHB6* mostraron insensibilidad a ABA en la inhibición de la germinación y en la regulación del cierre estomático, pero no en la inhibición del crecimiento de la raíz (Himmelbach y col., 2002). Esto sugiere que sería un regulador negativo de la respuesta a esta hormona.

Otro gen de *A. thaliana* que participa en las respuestas a ABA es *AtHB5*. Ensayos realizados con su promotor fusionado a un gen reportero indican que la exposición al ABA resulta en una redistribución de la expresión en diferentes tejidos, mientras que los ensayos de *Northern blot* muestran una reducción de los niveles de ARNm en plántulas (Johannesson y col., 2003). Las mutantes *abi1*, *abi3* y *abi5* presentaron una reducción en los niveles de transcripto, no así las mutantes *abi2* y *abi4*. Las mutantes *athb5* no mostraron fenotipo, mientras que las plantas que lo expresaron en forma ectópica y constitutiva presentaron hipersensibilidad a ABA, lo que indicaría que es un regulador positivo de la respuesta a esta hormona (Johannesson y col., 2003).

*AtHB20* de *A. thaliana* participaría en el desarrollo temprano del tejido vascular de hojas, el cual es mediado por auxinas (Aux Mattsson y col., 2003). Más recientemente se halló que también está involucrado en la regulación de la germinación en respuesta a ABA (Barrero y col., 2010).

En otras especies de plantas también se encontraron genes que participan en la respuesta a ABA. *CpHB7* de *C. plantagineum* al ser expresado en forma ectópica y constitutiva afectó las respuestas a esta hormona (Deng y col., 2006). *HaHB4* de *H. annuus* fue inducido por ABA y se halló que está involucrado en la respuesta al déficit hídrico (Dezar y col., 2005a). *MtHB1* de *M. truncatula* regula la emergencia de las raíces laterales en la respuesta al estrés mediada por ABA (Ariel y col., 2010). *NaHD20* de *N. attenuata* regula los niveles de ABA afectando la apertura de las inflorescencias, las flores y el tiempo de floración (Ré y col., 2011).

Algunos FTs HD-Zip I están involucrados en las vías de señales de otras hormonas. *HaHB4* tendría una función compleja ya que las plantas de *A. thaliana* transformadas con la construcción *35SCaMV::HaHB4* mostraron sensibilidad reducida al ET (Manavella y col., 2006), lo cual les permitió sobrevivir el déficit hídrico; mientras que presentaron niveles de JA y ET aumentados y menor acumulación de ácido salicílico (SA) en respuesta al daño o al ataque de herbívoros (Manavella y col., 2008b). *AtHB12* participaría en la vía de señales de giberelina (GA) (Son y col., 2010). *AtHB53* (*AtHB53*) es inducido por Aux y reprimido por citoquininas (CK) en

## I. Introducción general

---

raíces de *A. thaliana* (Son y col., 2005). LeHB1 de *Lycopersicon esculentum* (tomate) participa en la vía de ET (Lin y col., 2008).

### I.6.1.2 Respuestas a estímulos externos

Numerosos estímulos externos, generalmente estresantes, son capaces de inducir la expresión de FTs HD-Zip I y estos participan luego en las respuestas de adaptación de la planta. Naturalmente, siendo el ABA una hormona central en la respuesta al estrés hídrico, salino y osmótico (Cutler y col., 2010), no resulta sorprendente que muchos de los genes caracterizados estén involucrados en estas respuestas.

La expresión de *AtHB7* y *AtHB12* es fuertemente inducida por déficit hídrico y cuando se los expresó en *A. thaliana* bajo la dirección del promotor *35SCaMV* se observó un fenotipo con similitud al de las plantas sometidas a este estrés (Olsson y col., 2004). La expresión en *A. thaliana* de otro FT, *HaHB4*, controlado por el mismo promotor *35SCaMV* permitió obtener una tolerancia muy significativa al déficit hídrico (Dezar y col., 2005a). *CpHB7* de *C. plantagineum* es otro gen de la subfamilia I que ha sido relacionado al estrés hídrico (Deng y col., 2006).

*AtHB13* de *A. thaliana* y su ortólogo en girasol, *HaHB1*, están involucrados en la adaptación de sendas especies a condiciones de *freezing*, temperaturas inferiores a los 0°C (Cabello y col., 2012). En su mecanismo de acción se observó que participan proteínas estabilizadoras de membranas de las relacionadas a patogénesis y glucanasas. También participarían de la vía de señales de los azúcares (Cabello y col., 2010; Hanson y col., 2001).

Por otra parte, tanto *AtHB1* como *AtHB16* participarían en respuestas a la luz. El primero regulando de alguna forma el proceso de deetiación, ya que plantas de tabaco que lo expresan poseen un fenotipo de deetiación constitutiva (Aoyama y col., 1995); y el segundo en la respuesta a la luz azul y en la regulación del inicio de la floración en respuesta a cambios en el fotoperiodo (Wang y col., 2003).

### I.6.1.3 Regulación del desarrollo

Como se mencionó, ya sea en respuesta a estímulos externos o internos, varios genes HD-Zip I regulan diferentes procesos del desarrollo.

En muchos casos en los que estos genes fueron expresados usando el promotor *35SCaMV* se observaron alteraciones en la expansión celular. Las plantas transgénicas de *AtHB7* (Hjellström y col., 2003), *AtHB12* (Olsson y col., 2004), *HaHB1* (Cabello y col., 2012), *HaHB4* (Dezar y col., 2005b) desarrollaron diferente severidad en esta alteración, visible por el acortamiento del tallo. Las plantas transgénicas con la construcción *35SCaMV::AtHB16* mostraron menor expansión en células de las hojas, mientras que las silenciadas presentaron el fenotipo opuesto (Wang y col., 2003).

Las plantas transgénicas *35SCaMV::AtHB13* mostraron una expansión celular asimétrica

especialmente en cotiledones cuando se las cultivó con sacarosa (Hanson y col., 2001). Este fenotipo también se observó con plantas transgénicas *35SCaMV::HaHB1* (Cabello y col., 2010).

El gen *AtHB23* participaría en el desarrollo de la hoja, posiblemente en relación a procesos específicos de la cara adaxial (Kim y col., 2007). *AtHB1* también actuaría en hojas en el desarrollo del parénquima en empalizada, ya que su expresión con un promotor *35SCaMV* en tabaco altera específicamente el desarrollo de este tejido (Aoyama y col., 1995).

En otros estudios se hallaron genes asociados al desarrollo floral: *AtHB16* involucrado en la regulación del inicio de la floración en relación al fotoperiodo (Wang y col., 2003), *AtHB51* como un regulador junto a *LEAFY (LFY)* de *CAULIFLOWER (CAL)* en la transición del meristema vegetativo al floral (Saddic y col., 2006), y *NaHD20* en la coordinación de las transiciones florales con las respuestas a las condiciones de estrés (Ré y col., 2011).

Recientemente se descubrió que *MtHB1* de *M. truncatula* controla la emergencia de raíces laterales, particularmente coordinando este proceso con las condiciones estresantes a las que esté sometida la raíz, por ejemplo, el estrés salino (Ariel y col., 2010).

### I.6.2 La subfamilia HD-Zip II

Los FTs de esta subfamilia no han sido caracterizados tan ampliamente como los de la subfamilia I (ver Tabla I.4). No obstante, uno de ellos, *AtHB2*, ha sido estudiado extensamente y ha demostrado jugar un papel esencial en una importante respuesta de la planta denominada síndrome de evasión de la sombra. Ésta consiste en una alteración del programa de desarrollo de la planta en respuesta a una disminución en la relación de la luz roja sobre la luz rojo lejano percibida (R:RL), indicador de que hay otras plantas haciéndole sombra. Entre las manifestaciones fenotípicas principales y más visibles están: la elongación del hipocotilo o tallo, hiponastía de las hojas y menor expansión de cotiledones y hojas (Franklin, 2008).

*AtHB2* se expresa en oscuridad, y en luz se induce rápida y reversiblemente ante una baja en la relación R:RL (Carabelli y col., 1996, 1993; Roig-Villanova y col., 2006). Esta inducción es controlada por HFR1, que comienza a reprimirlo a partir de los 30 min (Sessa y col., 2005). La expresión ectópica y constitutiva de *AtHB2* en *A. thaliana* empleando el promotor *35SCaMV* genera el fenotipo característico del síndrome de evasión de la sombra, mientras que se observa una respuesta contraria cuando se expresa una construcción antisentido que lo silencia (Schena y col., 1993; Steindler y col., 1997). Estos fenotipos son fundamentalmente el producto del control de la expansión celular por la alteración del transporte de Aux (Steindler y col., 1999).

Además de *AtHB2*, varios de los genes de la subfamilia II de *A. thaliana* participarían también en el síndrome de evasión de la sombra. Entre ellos *AtHB4* (Sorin y col., 2009), y posiblemente *HAT2* y *HAT3* (Ciarbelli y col., 2008).

Se comprobó que las proteínas HD-Zip II funcionan principalmente como represores de la transcripción (Sawa y col., 2002; Steindler y col., 1999). *AtHB2* es capaz de inhibir su

## I. Introducción general

**Tabla I.4:** Lista de algunos de los principales genes HD-Zip II caracterizados funcionalmente. Se provee una síntesis de la función descrita o sugerida a partir de las evidencias experimentales.

HD-Zip	Función	Referencias
<i>Arabidopsis thaliana</i>		
<i>AtHB2</i>	Síndrome de evasión de la sombra. Elongación del hipocotilo y del tallo.	Carabelli y col., 1996, 1993 Kunihiro y col., 2011; Ohgishi y col., 2001 Schena y col., 1993; Sessa y col., 2005 Steindler y col., 1997, 1999
<i>AtHB4</i>	Síndrome de evasión de la sombra.	Sorin y col., 2009
<i>HAT2</i>	Morfogénesis mediada por Aux.	Sawa y col., 2002
<i>HAT22</i>	Respuesta a CK. Síntesis de clorofila y senescencia.	Köllmer y col., 2011
<i>Oryza sativa</i>		
<i>OsHOX1</i>	Transporte de Aux. Diferenciación provascular.	Meijer y col., 2000, 1997 Scarpella y col., 2002, 2000, 2005
<i>Helianthus annuus</i>		
<i>HaHB10</i>	Regulación del inicio de la floración. Regulación de JA y SA ante patógenos.	Dezar y col., 2011 Rueda y col., 2005

propia transcripción (Ohgishi y col., 2001). Por otro lado, en ensayos con fusiones del tipo *35SCaMV::HD-ZipII:VP16:GR* en las que se convierte al FT en un activador inducible gracias al motivo activador viral VP16 y al dominio del receptor de glucocorticoides; se observó que las proteínas quiméricas que incluían a HAT2 o a AtHB2 eran capaces de activar la transcripción del resto de los genes de la subfamilia (Ciarbelli y col., 2008; Sawa y col., 2002). Esto indicaría que estos FTs podrían conformar una compleja red de regulación.

*OsHOX1* de arroz también ha sido caracterizado funcionalmente. Se verificó su expresión en regiones en las que se origina tejido vascular y se comprobó que regula el comienzo de la ontogénesis provascular (Scarpella y col., 2000, 2005). Logra esto regulando el transporte de Aux (Scarpella y col., 2002, 2000). Además, se vio que es inducido por Aux y sacarosa (Scarpella y col., 2005).

En girasol, *HaHB10* fue asociado a la regulación del inicio de la floración y a la respuesta al estrés biótico mediante la regulación de los niveles de las hormonas JA y SA (Dezar y col., 2011). Dos genes de *C. plantagineum* también fueron relacionados a respuestas al estrés, más particularmente a la deshidratación a través de ABA (Deng y col., 2002; Frank y col., 1998). Por otro lado, se halló en *A. thaliana* que *HAT2* y *HAT22* responden a sequía y a diversas hormonas (Huang y col., 2008) e independientemente se observó que este último es regulado por CK y su expresión con el promotor *35SCaMV* redujo los niveles de clorofila y aceleró la senescencia (Köllmer y col., 2011).

Esta evidencia pone de manifiesto que los genes HD-Zip II participan en varios procesos

del desarrollo, principalmente regulando la expansión celular en el síndrome de evasión de la sombra. La mayoría responden a cambios en las condiciones de iluminación pero otros a diversos estímulos externos y participan en las vías de señales de diferentes hormonas, en especial de Aux.

### I.7 *Arabidopsis thaliana* como modelo de estudio en la era post-genómica

A partir de la secuenciación de los genomas de varios organismos, hemos entrado en la denominada “era post-genómica”. Cuando antes generalmente la máxima aspiración en el estudio de un gen o una proteína implicaba su caracterización en forma puntual, ahora se puede ambicionar ubicarlo dentro de una red de interacciones, y así conocer cuál es su función inmerso en la complejidad del organismo. Pero quizás lo más relevante es la nueva posibilidad de realizar estudios globales. Las herramientas disponibles han posibilitado llevar la escala de la investigación de lo individual o particular, a lo general u “ómico”, al nivel del organismo.

Sin embargo, como toda evolución, el progreso (o “proceso”) es continuo más allá de los hitos que lo marcan. Es muy probable que las tecnologías de secuenciación de nueva generación junto a otros numerosos desarrollos en las áreas de la proteómica, la metabolómica y la bioinformática propulsen, o estén propulsando, el advenimiento de una nueva etapa, en la que se observen nuevamente grandes cambios en la forma en que se plantean las investigaciones. Estas herramientas permiten que los estudios en genómica, genómica funcional y biología molecular comiencen a alcanzar la escala de las poblaciones y los ecosistemas. Quizás el avance reciente de las ramas de la metagenómica (Medini y col., 2008) y la ecología genómica (Weigel, 2011) sean un buen ejemplo de esto.

La biología vegetal no es ajena a estos cambios y posiblemente para quien se desempeñe en este campo no sea necesario describir a *Arabidopsis thaliana*, y explicar las ventajas y la importancia de trabajar con este modelo vegetal. Más allá de esto, considerando que el presente trabajo se concentra en parte en este organismo (especialmente en lo desarrollado en el Capítulo IV) y dado que una gran proporción del conocimiento del que se dispone en biología molecular y genómica funcional de plantas proviene de estudios en *A. thaliana*; creo fundamental dedicar algunos párrafos a su introducción.

*Arabidopsis thaliana* es una planta dicotiledónea anual cuyo genoma haploide de  $1.25 \times 10^9$  pb, repartido en cinco cromosomas. Su genoma fue el primero de vegetal en ser secuenciado hace más de una década (Initiative, 2000). Además, su reducido tamaño (alrededor de 35 cm de altura) y su ciclo de vida corto (alrededor de 8 semanas) la hacen un organismo fácilmente manipulable y multiplicable en un laboratorio. Su adopción como especie modelo ha promovido que se desarrollen una gran cantidad de métodos y protocolos para ensayos de laboratorio, entre



## I. Introducción general

---

los cuales puede resaltarse el sencillo método de transformación genética por *floral dip* (Clough y Bent, 1998). También se encuentran disponibles grandes bancos de mutantes que cubren una elevada proporción de sus genes (Alonso y col., 2003).

Por otro lado, particularmente a partir de la secuenciación de su genoma se ha desarrollado una vasta diversidad de herramientas informáticas para el estudio de *A. thaliana*, entre las cuales merecen destacarse especialmente aquellas mantenidas por TAIR (*The Arabidopsis Information Resource*, Lamesch y col., 2011).

## Objetivos generales

Desde el descubrimiento de los factores de transcripción HD-Zip, hace alrededor de 20 años, se han hecho grandes avances en su investigación y es vasto el conocimiento que de ellos se dispone. Como se ha desarrollado en el Capítulo I, en la actualidad se posee información sobre la dotación completa de los genes de esta familia de varias especies vegetales y las relaciones filogenéticas entre los miembros de algunas de ellas. Se conoce además, en mayor o menor medida, la función de una considerable cantidad de genes HD-Zip y se han hecho progresos significativos en el entendimiento de cómo estos FTs dimerizan y reconocen el ADN.

La presente Tesis fue desarrollada en un grupo de investigación que comenzó el trabajo con los FTs HD-Zip prácticamente desde que fueron descubiertos. El mismo ha realizado numerosos aportes al conocimiento sobre los FTs de las subfamilias I y II, que involucraron desde estudios bioquímicos hasta la caracterización funcional de FTs de diferentes especies vegetales.

Con todo, resta mucho por dilucidar en los diferentes aspectos de la función de estas proteínas, y mayor es el desconocimiento cuando los interrogantes son llevados a una escala genómica. Escasa es la información acerca de numerosas características generales de las proteínas de la familia HD-Zip, notoriamente en relación a los genes que regulan y a los procesos que explican la diversidad funcional de sus miembros.

La disponibilidad de grandes bases de datos públicas, gracias a proyectos internacionales de genómica y transcriptómica, han abierto la posibilidad de abordar estas cuestiones. El procesamiento de esta información requiere de la computación, y por lo tanto de las herramientas bioinformáticas. Los resultados obtenidos pueden ser luego puestos a prueba a nivel experimental, y la retroalimentación entre ambos acercamientos potencia su capacidad de resolver los interrogantes y brinda una visión más global acerca de estas subfamilias de FTs.

Con esta noción se planteó abordar el estudio de las subfamilias de FTs HD-Zip I y II; intentando, desde una óptica globalizadora, capturar generalidades y particularidades de sus miembros que nos permitan hacer aportes a la comprensión de la diversidad funcional de los FTs de este grupo, y de su evolución.

# **Capítulo II**

## **Materiales y Métodos**

### II.1 Materiales y Métodos correspondientes al Capítulo III

#### II.1.1 Búsqueda y obtención del conjunto de secuencias de FTs HD-Zip I

Las secuencias aminoacídicas de los FTs miembros de la subfamilia HD-Zip I de diferentes especies fueron obtenidas a través de la herramienta *Conserved Domain Architecture Retrieval Tool* (CDART) del NCBI (Geer y col., 2002). Las secuencias de aquellos miembros pertenecientes a especies cuyo genoma está secuenciado, como *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa*, *Zea mays*, *Physcomitrella patens* y *Selaginella moellendorffii* fueron reemplazadas por las secuencias utilizadas por Mukherjee y col. (2009). La inspección de dichas secuencias reveló que el tripéptido “YES” (Tyr-Glu-Ser) había sido accidentalmente removido de las secuencias publicadas. Esto fue confirmado por los autores (Bürklin T.R., comunicación personal) y debió ser corregido.

La redundancia de secuencias fue verificada usando el programa *skipredundant* del paquete EMBOSS (Rice, 2000) y los resultados fueron inspeccionados y manualmente curados. Las secuencias con la región carboxilo terminal (RCT) trunca no pertenecientes a las especies con genoma secuenciado fueron removidas. Finalmente se añadieron manualmente tres secuencias, MthB1 de *M. truncatula* (también conocido como HB1) y HaHB1 y HaHB11 de girasol.

Los conjuntos de secuencias de las regiones amino terminales (RNT), HD-Zip y RCTs fueron generadas reconociendo los dominios HD y LZ con el programa HMMer v2.3.2 (Eddy, 1998). Para esto se usaron los correspondientes modelos HMM obtenidos de la base de datos PFAM, PF00046.21 y PF02183.10 (<http://pfam.sanger.ac.uk>; Sonnhammer y col., 1997). Los informes fueron analizados y las secuencias truncadas mediante la escritura de rutinas en el lenguaje Perl, haciendo uso del paquete de herramientas Bioperl (Stajich y col., 2002). De esta forma se construyeron los tres conjuntos de secuencias con las porciones proteicas de las regiones amino terminales (RNT), HD-Zip y RCTs.

#### II.1.2 Alineamiento de secuencias y análisis filogenético

Los alineamientos de las secuencias de los HD-Zip y las proteínas completas fueron realizados con el programa MAFFT v6 mediante el algoritmo L-INS-i (Kato y Toh, 2008). El resultado fue inspeccionado manualmente y curado usando el editor de alineamientos múltiples Seaview4 (Gouy y col., 2010). También fueron alineadas las secuencias de las RCTs, separándolas por grupos, lo cual permitió generar la Figura III.3 que presenta la conservación en estas regiones. Los *sequence logos* que se muestran en la Figura III.2 y la Figura III.3 fueron generados con el programa Weblogo (Crooks y col., 2004).

Previo al análisis filogenético, se determinó el modelo más adecuado de evolución proteica empleando el programa ProtTest (Abascal y col., 2005), obteniéndose en los tres casos un modelo con una matriz de sustitución JTT (Jones y col., 1992), considerando sitios invariantes (Reeves,

## II. Materiales y Métodos

---

1992) y una tasa de sustitución entre sitios con una distribución  $\Gamma$  (modelo JTT+I+G Yang, 1993).

Las reconstrucciones filogenéticas fueron realizadas empleando este modelo en el programa PhyML 3.0 que utiliza un algoritmo basado en máxima verosimilitud o *maximum likelihood* (ML) (Guindon y Gascuel, 2003). Se utilizó la opción NNI de búsqueda de topología del árbol. Se usaron 100 réplicas de *bootstrap* para los árboles HZT y HZT+OG, y 144 réplicas para el árbol CST. El árbol HZT+OG contó adicionalmente con tres proteínas HD-Zip II como grupo externo (*outgroup*), HAT1, HAT22 and AtHB17. Su posición relativa fue utilizada para enraizar este árbol y posteriormente los árboles HZT y CST.

El árbol CST fue obtenido empleando el conjunto o *cluster* de computadoras Aquiles perteneciente al “Centro Internacional de Métodos Computacionales en Ingeniería” (CIMEC, <http://www.cimec.org.ar>, CONICET and UNL). Éste consiste en 80 microprocesadores Intel Pentium® 4 cuyo trabajo en paralelo permitió completar la reconstrucción en aproximadamente 11 horas de cómputo.

Los árboles definitivos fueron obtenidos usando el programa *consense* del paquete PHYLIP (Felsenstein, 1989). Las imágenes que se muestran en la Figura III.4 fueron generadas con el programa FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

### II.1.3 Reconocimiento de motivos y modificaciones post-traduccionales

El programa MEME (Bailey y col., 2006) fue usado para la identificar motivos en las RNTs y RCTs. Los parámetros seleccionados fueron: límite de motivo 20, ancho mínimo de motivo 6, mínimo número de sitios por motivo 6 y valor e umbral de 0,1. Los gráficos de distribución de motivos (Figura III.6 y Figura III.9) fueron generados usando el paquete genoPlotR (Guy y col., 2010) para el lenguaje estadístico R (R Development Core Team, 2011).

La predicción de sitios putativos de fosforilación fue realizada en las RNTs y RCTs de las 178 proteínas. Para esto se usó la versión *stand-alone* del programa NetPhos 3.1 (Blom y col., 1999), con un umbral de 0,9.

La predicción de las señal de localización nuclear (SLN) fue realizada con el programa NLStradamus (Nguyen Ba y col., 2009) en las secuencias completas de los FTs, empleando un umbral de 0,6.

### II.1.4 Cepas bacterianas

Se emplearon cepas bacterianas de las especies *Escherichia coli* y *Agrobacterium tumefaciens* cuyas características se detallan a continuación:

- *Escherichia coli* (*E. coli*) cepa DH5 $\alpha$  (Hanahan, 1983): Utilizada en diversos procedimientos rutinarios de clonado en el laboratorio de Biología Molecular. Posee una porción del

gen *lacZ* (*lacZ* $\Delta$ *M15*) por lo cual presenta el sistema de selección de colonias blancas y azules, disponible en ciertos plásmidos. Permite obtener ácido desoxirribonucleico (ADN) plasmídico de buena calidad y alto rendimiento (*endA1*). Provee mayor estabilidad al inserto presente en el plásmido con que se la transforme (*recA1*).

- *Agrobacterium tumefaciens* (*A. tumefaciens*) cepa LBA440 (Ooms y col., 1982): Posee el plásmido Ti desarmado *pAL4404*, con resistencia a estreptomycin, que se obtiene de reemplazar la región del ADN-T del plásmido *pTiAch5* de la cepa salvaje Ach5 por el plásmido *pBR322*. Presenta resistencia cromosomal a rifampicina.

### II.1.5 Transformación de bacterias

Tanto las células de *Escherichia coli* como las de *Agrobacterium tumefaciens* fueron transformadas por el método de electroporación en un equipo Gene pulser (Bio-Rad, Hercules, CA, USA).

Primero se prepararon células competentes de dichas bacterias según el protocolo recomendado por el fabricante para cada especie. El pulso eléctrico fue realizado en cubetas de 0,1 cm para *E. coli* y 0,2 cm para *A. tumefaciens* con un potencial de 1,8 y 2,5 kV, respectivamente; ajustando el resto de las variables del equipo según las recomendaciones del fabricante. A continuación se agregó 1 ml de medio LB estéril y se incubó la suspensión 1 hora a 37 °C en el caso de *E. coli*, y 3 horas a 28 °C en el caso de *A. tumefaciens*. Finalmente, el cultivo de células fue centrifugado a 4.000 g durante 5 minutos (min), el sedimento celular resuspendido en 100  $\mu$ l de medio LB estéril, y sembrado en placas de Petri que contenían medio LB sólido suplementado con el antibiótico correspondiente al vector de transformación empleado. En los casos en que los vectores permitían la discriminación de colonias blancas y azules, el medio fue suplementado con IPTG (120  $\mu$ g/ml) y X-gal (67  $\mu$ g/ml). Las placas de Petri fueron incubadas durante 16 horas a 37 °C en el caso de *E. coli* y 30 horas a 28 °C en el caso de *A. tumefaciens*.

### II.1.6 Construcciones genéticas

Las construcciones genéticas empleadas fueron las siguientes:

- *pGBKT7+BD-A13RCT*: la RCT de AtHB13 fue obtenida amplificando por reacción en cadena de la polimerasa (PCR) este fragmento con los oligonucleótidos A13Y2HCtF y A13Y2HCtR a partir de un clon Riken, código RAFL09-47-L06. Luego de la purificación del fragmento se clonó en los sitios *Eco RI/Bam HI* del vector *pGBKT7*.
- *pGBKT7+BD-A13WxA*: se usaron los oligonucleótidos A13Y2HCtF y A13WxAR para amplificar la RCT con los triptofanos mutados a partir del mismo clon Riken mencionado en el punto anterior.

## II. Materiales y Métodos

---

- *pGBKT7+BD-A13dAHA*: se usaron los oligonucleótidos A13Y2HCtF y A13dAHAR para amplificar la RCT sin el motivo AHA a partir del mismo clon Riken mencionado previamente.
- *pGBKT7+BD-A13AHA*: se usaron los oligonucleótidos A13AHAF y A13Y2HCtR para amplificar el motivo AHA a partir del mismo clon Riken mencionado previamente.
- *pGBKT7+BD-AtHB1*: a partir de una transcripción reversa de ARNm de *A. thaliana* utilizada como molde se amplificó *AtHB1* con los oligonucleótidos ATHB1F y ATHB1R, y se clonó en los sitios *Eco RI/Bam HI* del vector *pGBKT7*. Esta construcción fue preparada y gentilmente provista por el Lic. Matías Capella.
- *pBI121+HaHB4*: el ADN copia (ADNc) del gen *HaHB4* fue clonado en trabajos previos (Dezar y col., 2005b) en el vector *pBI121* que permite su expresión en plantas bajo el control del promotor ubicuo, fuerte y constitutivo 35S del virus del mosaico de la coliflor (*35SCaMV*).
- *pBI121+HaHB1*: el ADNc de *HaHB1* fue previamente aislado por Chan y Gonzalez (1994) de una biblioteca construida en el vector lambda gt10. Este fragmento fue clonado luego por Rueda (2008) en los sitios *Eco RI* del vector pMTL22 y obtenido mediante digestión *Bam HI/Sac I* para ser clonado en los mismos sitios del vector *pBI121*. De esta forma, se puede transformar *A. thaliana* y la expresión de *HaHB1* será controlada por el promotor *35SCaMV*.
- *pBI121+H4SCT*: el ADNc de *HaHB4* sin su RCT fue obtenido por amplificación por PCR con los oligonucleótidos Transf1 y H4-WCT-R a partir del clon *35SCaMV:HaHB4*, y clonado directamente en el vector *pBI121* en los sitios *Bam HI/Sac I*.
- *pBI121+H1SCT*: el cDNA sin la RCT fue obtenido por amplificación por PCR con los oligonucleótidos H1-WCT-R y H1atF a partir de la construcción *pBI121+HaHB1*. Luego fue clonado directamente con los sitios *Bam HI/Sac I* del vector *pBI121*.
- *pBI121+H4-H1*: para obtener esta construcción quimérica se realizaron en primer lugar dos amplificaciones independientes que permitieran generar fragmentos de ADN parcialmente solapados, como describe Higuchi y col. (1988). La región 5' de *HaHB4* comprendiendo el dominio HD-Zip fue amplificada por PCR con los oligonucleótidos Transf1 y H1\*H4R a partir de la construcción *pBI121+HaHB4*; mientras que la región 3' de *HaHB1* comprendiendo la RCT fue amplificada por PCR con los oligonucleótidos H1\*H4F y H1CDS-R usando ADN de la construcción *pBI121+HaHB1* como molde. Ambos productos fueron purificados entonces, mezclados en una proporción mol a mol, desnaturalizados y luego hibridados. El producto de esta hibridación fue extendido con la

## II. Materiales y Métodos

---

enzima Klenow y amplificado por PCR con los oligonucleótidos Transf1 y H1CDS-R. Este producto fue purificado y clonado en el vector *pCR2.1-TOPO* (Invitrogen®). Finalmente, el cDNA mutado fue purificado y digerido con las enzimas *Bam HI/Sac I* para su clonado en el vector *pBII21* en los mismos sitios.

Los oligonucleótidos utilizadas para estas construcciones se muestran en la Tabla II.1.

### II.1.7 Obtención y purificación de ácidos nucleicos

#### II.1.7.1 Minipreparación de ADN plasmídico

La obtención de ADN plásmidico a partir de células de *E. coli* y *A. tumefaciens* se realizó empleando el método de lisis alcalina descrito por Birnboim y Doly (1979). Primero se realizó un cultivo de células bacterianas en medio LB adicionado con ampicilina o kanamicina según el vector con el que fueron transformadas. En el caso de *A. tumefaciens* se agregó además rifampicina (gen de resistencia presente en el genoma) y estreptomycin (gen de resistencia presente en el vector Ti desarmado). Las bacterias se cultivaron a la temperatura correspondiente a cada cepa hasta saturación (alrededor de 16 horas en agitador orbital a 180 rpm, Forma Scientific®, Marietta, OH, USA), se tomaron 1,5 ml y se centrifugaron a 5.000 g durante 5 min. El sedimento celular fue resuspendido en 100 µl de solución I e incubado durante 5 min en hielo. Luego se le adicionaron 200 µl de solución II, se agitó por inversión e incubó nuevamente 5 min en hielo. Posteriormente se le agregó 150 µl de la solución III, se agitó y volvió a incubar 5 min en hielo. A continuación se centrifugó a 12.000 g y 4 °C durante 10 min colectándose el sobrenadante al que se le realizó una extracción con fenol:cloroformo (fase acuosa:fenol:cloroformo en proporción 2:1:1). Luego de una agitación vigorosa se centrifugó a 10.000 g durante 10 min a temperatura ambiente y se tomó la fase acuosa. El ADN plasmídico fue precipitado mediante el agregado de 1 volumen de isopropanol, seguido de incubación a -20 °C durante 30 min y centrifugación a 13.000 g y 4 °C durante 20 min. El sedimento fue lavado mediante el agregado de 500 µl de etanol al 70 % (v/v) y centrifugado durante 15 min a 13.000 g y 4 °C. Finalmente el sedimento seco fue resuspendido en 20 µl de agua destilada previamente esterilizada.

Soluciones empleadas:

- Solución I: Tris-HCl 25 mM (pH 8,0); glucosa 50 mM; EDTA 10mM.
- Solución II: NaOH 0,2 N; SDS 1 % (p/v).
- Solución III: KAc 5M (pH 5,2).



**Tabla II.1:** Oligonucleótidos utilizados para los clonados

Nombre	Secuencia	Sitio de Restricción	Uso
A13Y2HCtF	5'-ggCgAATTCTCAATAAATCTAAACAAGAAACTg-3'	<i>Eco RI</i>	Simple híbrido
A13Y2HCtR	5'-gCggATCCTCAATTgTACTgTTgCTgATCAA-3'	<i>Bam HI</i>	Simple híbrido
A13AHAF	5'-ggCgAATTcGAAgAATgTATCgTAAC-3'	<i>Eco RI</i>	Simple híbrido
A13dAHAR	5'-gCggATCCTTACATgTTACTgATACTATTCTC-3'	<i>Bam HI</i>	Simple híbrido
A13WxAR	5'-gCggATCCTCAATTgTACTgTTgCTgATCAAAGCgCTggCgCAAAACCAgA-3'	<i>Bam HI</i>	Simple híbrido
ATHB1F	5'-gCggAATTCATggAATCCAATTCgTTTTTC-3'	<i>Eco RI</i>	Simple híbrido
ATHB1R	5'-gCggATCCTAAggCCATCCCCAgAAA-3'	<i>Bam HI</i>	Simple híbrido
Transf1	5'-gCggATCCACCATgTCTTTCAACAAGTA-3'	<i>Bam HI</i>	H4-H1 clonado
H1CDS-R	5'-ggggAgCTCTCAATTgAATTgTggTTgTTCC-3'	<i>Sac I</i>	H4-H1 clonado
H1*H4F	5'-gTggAggTgTAAATAAATAggAgCCAgC-3'		H4-H1 clonado
H1*H4R	5'-TATTTTTTAgCACCTCCAAGTgATgAgTAgg-3'		H4-H1 clonado
H4-WCT-R	5'-CCCgAgTCTCTAATCTTCCACCgCTgCCAC-3'		H4WCT clonado
H1-WCT-R	5'-ggCgAgCTCTCATCCTTCTgTTTCTTTTATgTTgAgg-3'	<i>Sac I</i>	H1WCT clonado
H1atF	5'-ggggATCCgCTgATgACTTgCACTggAATggC-3'	<i>Bam HI</i>	H1WCT clonado
H1qR	5'-CCAACCATggCCAAAACCCTg-3'		H1 qRT-PCR
H1qF	5'-ggCCggCAGATCATCAACTTC-3'		H1 qRT-PCR
H4qR	5'-gCCgAgTCTTAGAACTCCAACCACCTTTTg-3'		H4 qRT-PCR
H4qF	5'-CgCgATCAAAATCgAggCAGATTg-3'		H4 qRT-PCR
H4quimqF	5'-CgCgATCAAAATCgAggCAGATTg-3'		H4-H1 qRT-PCR
H1AsTH4R	5'-TCCTTCTgTTCTTTTATgTTgAgg-3'		H4-H1 qRT-PCR
UBCQPCR-R	5'-CAGTggACTCgTACTTgTTCTTgT-3'		qRT-PCR
UBC9GENOM-F	5'-gTTTTTggAAATgTTgACAaggAC-3'		qRT-PCR

## II. Materiales y Métodos

---

### II.1.7.2 Minipreparación de ADN genómico de *Arabidopsis thaliana*

Para analizar mediante la técnica de PCR la población de plantas de *A. thaliana* transformadas, se realizaron preparaciones de ADN total empleando un método rápido descrito por Li y Chory (1998). Se colocó una hoja joven (de aproximadamente 29 días) en un tubo y se la maceró durante 15 s. Después de agregar 700 µl de solución amortiguadora de extracción, se agitó vigorosamente durante 10 s y se centrifugó a 18.000 g durante 1 min. Se recuperó el sobrenadante y se realizó una extracción con 1 volumen de la mezcla fenol-cloroformo 1:1 (v/v). Luego de centrifugar 10 min a 10.000 g, se recuperó nuevamente el sobrenadante y se realizó una segunda extracción sólo con cloroformo (1 vol). Posteriormente, se centrifugó durante 10 min a 10.000 g, se tomaron 600 µl del sobrenadante y se agregó 1 volumen de isopropanol, para luego incubar 1 hora a -20 °C. La mezcla alcohólica se centrifugó a 10.000 g durante 10 min; el precipitado se lavó con etanol 70 % y se secó a temperatura ambiente. Finalmente se resuspendió en 100 µl de agua destilada estéril.

Soluciones empleadas:

- Solución amortiguadora de extracción: Tris-HCl 200 mM (pH 8,0), NaCl 250 mM, EDTA 25 mM, SDS 0,5 % (p/v).

### II.1.8 Reacciones enzimáticas

#### II.1.8.1 Digestión de ADN con endonucleasas de restricción

La digestión de los diferentes fragmentos de ADN fue realizada con las enzimas mencionadas en las correspondientes descripciones de los clonados realizados, tanto para la preparación del vector y el inserto, como para la identificación de colonias positivas. Se empleó en cada caso la solución amortiguadora adecuada a la enzima o par de enzimas empleadas según las recomendaciones del fabricante. Se utilizaron entre 1 y 5 U de la enzima por cada µg de ADN a digerir, siendo el tiempo de digestión 2 horas y la temperatura de 37 °C.

#### II.1.8.2 Ligación de moléculas de ADN

Los fragmentos de ADN obtenidos por corte fueron ligados al vector correspondiente utilizando 1 U de la enzima T4 ADN ligasa (Promega<sup>®</sup>, Madison, WI, USA) en un volumen de reacción de 10 µl amortiguado mediante el uso de la solución provista por el fabricante. Los volúmenes de vector e inserto fueron tales que la relación molar entre los mismos fuera 5:1. La reacción se incubó durante 16 horas a 4 °C.

Por otra parte los productos de amplificación por PCR fueron ligados al vector *pCR 2.1-TOPO* mediante la acción enzimática de una topoisomerasa. El vector se encuentra en forma lineal y posee un residuo de desoxitimidina 3' protuberante en cada uno de sus extremos, al cual se haya enlazada una topoisomerasa. La reacción de ligación se realiza a temperatura ambiente

## II. Materiales y Métodos

durante 20 min en una solución amortiguadora provista por el fabricante con una relación molar de vector a inserto de 5:1.

### II.1.8.3 Reacción en cadena de la polimerasa (PCR)

La amplificación de fragmentos de ADN mediante PCR fue realizada en un volumen de 50 µl para las reacciones cuyo fin era obtener el inserto para su clonado, y de 20 µl para reacciones de identificación de clones positivos de *E. coli*, de *A. tumefaciens* e identificación de plantas transformadas de *A. thaliana*. En todos los casos la reacción fue amortiguada empleando la solución provista por el fabricante, adicionándose diferentes soluciones y agua estéril a los fines de obtener la composición que se presenta en la Tabla II.2.

**Tabla II.2:** Composición de la solución en que se desarrollo la PCR.

Componente	Concentración final
Solución de amortiguación	1x
MgCl <sub>2</sub>	2 mM
dATP, dCTP, dGTP, dTTP	0,2 mM (c/u)
Oligonucleótido directo	0,01 mM
Oligonucleótido reverso	0,01 mM
Molde	1-5 ng
Taq ADN polimerasa (Invitrogen®, Carlsbad, CA, USA)	0,03 U/µl

La reacción se desarrolló en un termociclador PTC-100 (MJ Research®, Waltham, MA, USA) empleando el siguiente programa: 30 ciclos, cada uno compuesto por una primera etapa de 1 min a 94 °C, una segunda etapa de 1 min y 30 s a la temperatura de deshibridación (Td) correspondiente a los oligonucleótidos utilizados, y finalmente una etapa de 2 min a 72 °C. Luego de estos 30 ciclos se realizó una etapa de 10 min a 94 °C. La Td adoptada correspondió a la menor de las temperaturas calculadas para cada uno de los oligonucleótidos específicos según la fórmula empírica:

$$Td = 2 \times (A + T) + 4 \times (C + G) - 5$$

en la que A, T, C y G representan el número de nucleótidos en el cebador que hibridan sobre el molde y cuya base nitrogenada es adenina, timina, citosina y guanina; respectivamente.

## II.1.9 Análisis de ácidos nucleicos

### II.1.9.1 Electroforesis de ADN en geles de agarosa

Se emplearon geles de agarosa para el aislamiento de fragmentos resultantes de cortes con enzimas de restricción y productos de PCR; así como el análisis de clones obtenidos. El sistema

## II. Materiales y Métodos

---

fue del tipo submarino (Ausubel y col., 1987). La concentración de agarosa utilizada varió entre 0,7 y 2,0 % (p/v) dependiendo del tamaño de los fragmentos a separar. Los geles se prepararon en una solución de TAE 1x con una concentración de 0,3 µg/ml de bromuro de etidio. La siembra se realizó diluyendo previamente la muestra diez veces en solución de siembra. La separación electroforética se llevó a cabo en una solución TAE 1x a un voltaje constante entre 1 y 5 V/cm de gel. Una vez terminada la separación electroforética, la visualización de los fragmentos de ADN se realizó en un transluminador de luz UV (UVP®, Upland, CA, USA).

En los casos en los que fue necesario estimar la longitud de los fragmentos de ADN separados electroforéticamente se sembró en el mismo gel el marcador de tamaños moleculares que se obtiene de la digestión del ADN genómico del bacteriófago λ con la endonucleasa de restricción *Hind III*. El producto de esta digestión es una mezcla equimolar de fragmentos de ADN de 23.130, 9.416, 6.557, 4.361, 2.322, 2.027, 564 y 125 pares de bases (pb).

Soluciones empleadas:

- TAE 1x: Tris-Ac 20 mM (pH 8,5), EDTA 1 mM.
- Solución de siembra: azul de bromofenol 0,25 % (p/v), xilencianol FF 0,25 % (p/v), glicerol 30 % (v/v).

### II.1.9.2 Purificación de fragmentos de ADN

Los fragmentos de ADN obtenidos por digestión enzimática con endonucleasas de restricción o por PCR fueron purificados a partir de geles de agarosa. La separación electroforética fue realizada en las mismas condiciones descritas en la Sección II.1.9.1 y una vez alcanzado un grado adecuado de resolución de la banda de ADN, se escindió esa porción de gel. El fragmento de ADN fue purificado a partir de este segmento de gel empleando el kit comercial *GFX PCR DNA and Gel Band Purification Kit* (Amersham®, Piscataway, NJ, USA). El protocolo utilizado fue el recomendado por el fabricante.

### II.1.10 Cultivo, transformación y simple híbrido en levaduras

Los cultivos y la transformación de las células de la levadura *Saccharomyces cerevisiae* AH109 (Clontech®) fueron realizados en medio YPDA o medio sintético mínimo (SD) suplementado con un *dropout* de aminoácidos deficiente en triptófano o histidina según el caso (Sherman, 1991).

Las células de levadura fueron transformadas con las construcciones adecuadas siguiendo el protocolo de acetato de litio (Gietz y col., 1992). Las células fueron seleccionadas por prototrofia para triptófano en un medio SD mínimo sin ese aminoácido.

La capacidad de transactivar fue ensayada mediante el cultivo de las levaduras en un medio SD mínimo sin histidina.

### II.1.11 Tratamiento de plantas de *Arabidopsis thaliana*

#### II.1.11.1 Cultivo de plantas en tierra

Se utilizó *Arabidopsis thaliana* Heyhn. ecotipo Columbia (Col-0). Las semillas fueron compradas a Lehle Seeds® (Tucson, AZ). Las plantas fueron cultivadas directamente en tierra en una cámara de cultivo a 22-24°C en un fotoperiodo de día largo (16 horas de iluminación con una mezcla de tubos fluorescentes tipo blanco frío y GroLux) a una intensidad aproximada de  $150 \mu Em^{-2}s^{-1}$  en macetas de 8 cm de diámetro y 7 cm de altura.

#### II.1.11.2 Desinfección de semillas y cultivo *in vitro*

Las semillas fueron desinfectadas previamente a su siembra en placas de Petri para evitar el crecimiento de hongos. Para esto se hizo un tratamiento con 2 soluciones, en primer lugar un lavado con etanol 70 % (v/v) durante 5 min, y luego con lavandina 10 % y SDS 1 % durante 15 min. A partir de este momento la manipulación de las semillas se realizó en condiciones de esterilidad empleándose una cabina de flujo de aire horizontal (Forma Scientific, Marietta, OH, USA). A continuación se retiró esta solución y se hicieron 3 lavados con agua estéril para eliminar los restos de la última solución que pudieran afectar la germinación.

Para los casos en que las semillas pertenecían a una transformación (filial 1 o F1), éstas se resuspendieron en una cantidad suficiente de agar 0,1 % estéril (aproximadamente 8 ml) para poder distribuir las uniformemente en placas de Petri de 150 mm conteniendo medio MS suplementado con 50  $\mu g/ml$  de kanamicina. Cuando las semillas provenían de plantas previamente seleccionadas (filial 2 o F2, y posteriores) se esterilizaron aproximadamente 20 de cada línea transformante y se resuspendieron en 100-150  $\mu l$  de agar 0,1 % estéril. Se sembraron 8 líneas diferentes por cada placa de Petri de 150 mm que contenía medio MS suplementado con 50  $\mu g/ml$  de kanamicina. Las placas fueron incubadas a 4 °C durante 2 días para interrumpir la dormancia y llevadas a cámara de cultivo (condiciones de cultivo en el punto siguiente).

### II.1.12 Transformación de plantas de *Arabidopsis thaliana*

Las transformantes de *A. thaliana* fueron obtenidas a partir de plantas salvajes del ecotipo Columbia-0 como se describió más arriba. Para cada transformación se prepararon 8 macetas con 10 semillas en cada una y se dejaron crecer en una bandeja con agua sellada con film autoadherente transparente. Luego de la primera semana de incubación en la cámara de cultivo, se retiró el film.

El método de transformación empleado fue el de inmersión o *floral dip* (Clough y Bent, 1998). Las plantas se cultivaron hasta floración, aproximadamente 4 semanas, y cuando los pedúnculos florales se separaron de su yema axilar próxima se cortaron las inflorescencias, sin dañar las hojas caulinares o las yemas axilares cercanas. Entre los 2 y 3 días siguientes se cortaron las

## II. Materiales y Métodos

---

nuevas inflorescencias surgidas mediante el procedimiento descrito. Esto se repitió una o dos veces más hasta el momento en que todas las inflorescencias presentaron al menos 4 flores no abiertas (sin pétalos visibles). La transformación se realizó dos días después del último corte.

El método de *floral dip* consiste básicamente en sumergir las inflorescencias en una suspensión de bacterias *A. tumefaciens* portadoras del plásmido con el ADN-T que se desea insertar en la planta. La suspensión bacteriana se obtuvo preparando primero un cultivo de 30 ml en medio LB suplementado con los antibióticos kanamicina, estreptomycin y rifampicina (cada uno de los 3 en una concentración de 50 µg/l), que se hizo crecer 16 horas a 28 °C en un agitador orbital a 180 rpm. Con este cultivo se inocularon 600 ml de medio LB suplementado con los mismos antibióticos, y distribuido en Erlenmeyers cuya capacidad permitiera una adecuada superficie de intercambio gaseoso. El cultivo se hizo crecer hasta alcanzar la fase estacionaria, aproximadamente 16 horas a 28 °C en un agitador orbital a 180 rpm. Las células fueron cosechadas por centrifugación a 5.500 g durante 20 min y el sedimento fue resuspendido en 1 l de solución de sacarosa al 5 % conteniendo 300 µl del detergente Silwet L-77 (OSi Specialties®, Danbury, CT, USA). El preparado se colocó en un vaso de precipitado y las células fueron mantenidas en suspensión mediante un agitador magnético. La porción aérea de las plantas fue entonces sumergida casi completamente durante 30 a 60 s en el vaso de precipitado tratando de evitar el contacto del líquido con las hojas de la roseta. Cada maceta fue colocada posteriormente en posición horizontal en una bandeja que fue sellada con film transparente autoadherente para evitar su desecación. Esta bandeja se llevó a la cámara de cultivo. Al día siguiente se le retiró el film y las plantas se pusieron en posición vertical hasta finalizar su ciclo de vida. Aproximadamente luego de 6 semanas, cuando las vainas se secaron, se cosecharon las semillas para lo cual fue necesario retirar manualmente los restos de vainas, hojas, etc. Éstas se almacenaron a 4 °C preservando su viabilidad para estudios posteriores.

### II.1.12.1 Selección de transformantes y obtención de líneas homocigotas

La selección de transformantes se realizó mediante el marcador de selección para plantas presente en el ADN-T, el gen de resistencia a kanamicina *nptII*. Las semillas fueron desinfectadas como se describe en la Sección II.1.11.2 y sembradas en medio MS suplementado con vitaminas y kanamicina 50 µg/ml. Las placas fueron llevadas a una cámara de cultivo con las condiciones ya especificadas y, aunque el medio permitió la germinación de la gran mayoría de las semillas (entre 95 y 99 %) aproximadamente a los 7 días, el efecto de la kanamicina comenzó a visualizarse sobre las plantas no transformadas: los cotiledones presentaron un aspecto clorótico. En un periodo de 3 a 5 días adicionales las plantas transformadas lograron desarrollar hojas verdaderas. En ese momento se procedió a transplantarlas a tierra. Se continuó con su cultivo en cámara hasta que sus vainas maduraron, se cosecharon las semillas de cada planta en forma individual y se almacenaron a 4 °C para su análisis posterior.

## II. Materiales y Métodos

---

Para la selección de líneas homocigotas, las plantas transformadas e identificadas como positivas en el análisis de PCR (descrito en la Sección II.1.8.3) se sometieron al siguiente procedimiento. En primer lugar se cosecharon las semillas de cada planta F1 en forma independiente y se sembraron alrededor de 20 en placas de Petri que contenían medio MS suplementado con kanamicina. Las plantas seleccionadas por su resistencia al antibiótico fueron transplantadas a tierra y se cosecharon nuevamente sus semillas. La siembra en placas de Petri se repitió y aquellas líneas en las que todas las plantas sobrevivieron a la kanamicina fueron consideradas homocigotas. Cuando el grado de supervivencia fue inferior al 100 % se repitió el procedimiento hasta lograr que todas las plántulas de la línea sobrevivieran al tratamiento.

### II.1.13 Cuantificación de transcritos por RT-PCR en tiempo real

La cuantificación de transcritos por RT-PCRc fue realizada en cada una de las líneas transgénicas estudiadas. El ARN de las plantas fue preparado con el reactivo Trizol<sup>®</sup> según las instrucciones del fabricante (Invitrogen<sup>®</sup>).

La reacción de retrotranscripción (RT) fue realizada en dos pasos:

1. En un volumen final de 5  $\mu$ l se colocó 1  $\mu$ g de ARN junto con 0,1  $\mu$ M de oligonucleótidos poli-dTV (consiste en 1 nucleótido variable, ubicado en el extremo 3', seguido de 24 nucleótidos timina). Esta mezcla se incubó durante 5 min a 70 °C para luego ser colocada inmediatamente en hielo (este tratamiento permite desnaturalizar las estructuras secundarias del ARN).
2. Se adicionó a la solución anterior 3  $\mu$ l de la solución amortiguadora provista por el fabricante, 1  $\mu$ l de dNTPs (20 mM), 100 U de transcriptasa reversa M-MLV (Promega<sup>®</sup>) y se completó el volumen hasta 15  $\mu$ l con agua bidestilada estéril. Esta mezcla se incubó 1,5 horas a 42 °C y luego la enzima se inactivó a 80 °C durante 5 min seguidos de 30 s a 94 °C.

La PCR cuantitativa (PCRc) se realizó en un termociclador MJ-Chromo 4<sup>®</sup>, en un volumen final de 20  $\mu$ l conteniendo 8 pmol de cada oligonucleótido, 2 mM de MgCl<sub>2</sub>, 10  $\mu$ l de una dilución 1/25 de la reacción de RT y 0,12  $\mu$ l de Platinum<sup>®</sup>Taq (Invitrogen<sup>®</sup>). La fluorescencia fue medida a 78-80 °C durante 40 ciclos de PCR. Como gen de referencia o *housekeeping* se usó la actina. Los oligonucleótidos utilizados para las reacciones se encuentran en la Tabla II.1.

### II.1.14 Tratamiento con etileno

Para este tratamiento las semillas fueron primero esterilizadas superficialmente y distribuidas en placas de Petri con medio MS (Murashige y Skoog, 1962) con 0,8 % de Agar, con una adición de ácido 1-aminociclopropano-1-carboxílico (ACC), precursor de la fitohormona etileno, para

alcanzar una concentración final 5  $\mu\text{M}$ . Las plantas fueron incubadas dos días en oscuridad a 4 °C y luego transferidas a cámara de cultivo por 3-5 días manteniendo las condiciones de oscuridad. El fenotipo fue observado posteriormente mediante el empleo de un microscopio óptico.

### II.1.15 Cuantificación del aserrado de las hojas

La cuantificación del aserrado de las hojas fue realizada en hojas de la roseta de plantas de 21 días. Éstas fueron cortadas y fotografiadas con un escáner plano, para luego procesar las imágenes con el programa LAMINA (Bylesjö y col., 2008). Éste está diseñado para reconocer diferentes parámetros de las hojas, entre ellos el aserrado.

## II.2 Materiales y Métodos correspondientes al Capítulo IV

### II.2.1 Características y procesamiento de la información transcriptómica

Los datos de expresión fueron obtenidos del proyecto AtGenExpress de análisis transcriptómico de *A. thaliana* var. Col-0 (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>), desarrollado por diferentes grupos de investigación bajo la coordinación del programa AFGN (*Arabidopsis Functional Genomics Network*, <http://www.uni-tuebingen.de/plantphys/AFGN/>). Se seleccionó en particular el conjunto de experimentos correspondientes a los tratamientos de estrés abiótico, que incluye 298 microarreglos. Los niveles de expresión se encontraban normalizados por el método GC-RMA (Wu y col., 2004) y la señal de fondo había sido ajustada, por lo que no fue necesario realizar procesamientos adicionales.

Para que se comprendan en profundidad las características de los tratamientos de las plantas utilizadas para la obtención de este conjunto de datos, se hará una descripción detallada de los procedimientos empleados por sus autores. Los experimentos fueron realizados por diferentes grupos de investigación especializados en los estreses particulares. Las plantas fueron inicialmente cultivadas por 13 días a 24 °C en condiciones de esterilidad y día largo (16 horas luz/8 horas de oscuridad). El sustrato utilizado fue medio MS (Murashige y Skoog, 1962) complementado con 0,5 % de agar y 0,5 % de sacarosa. Luego de esto se cambió el medio de cultivo por uno nuevo de iguales características pero sin sacarosa. Luego de cinco días adicionales en estas condiciones de incubación y a las tres horas posteriores a la transición oscuridad a luz (noche/día), las plantas de 18 días fueron sometidas a los siguientes tratamientos:

- 1) Plantas control: las plantas fueron manipuladas al igual que las plantas tratadas, omitiendo el estímulo estresante.
- 2) Estrés por frío (4 °C): las plantas fueron colocadas en hielo en una cámara a 4 °C. La luz provista fue de 60  $\mu\text{Einstein.cm}^{-2}.\text{s}^{-1}$ .



## II. Materiales y Métodos

---

- 3) Estrés osmótico: se añadió manitol al medio en una concentración de 300 mM, el medio fue agitado y las plantas se regresaron a la cámara de cultivo.
- 4) Estrés salino: se agregó NaCl al medio para alcanzar una concentración de 150 mM.
- 5) Estrés por déficit hídrico: las plantas fueron estresadas con una corriente de aire seco durante 15 min hasta alcanzar un 10 % de pérdida de peso fresco.
- 6) Estrés genotóxico: se utilizó bleomicina (1.5 µg/ml) y mitomicina (22 µg/ml).
- 7) Estrés oxidativo: se adicionó al medio metilviológeno a una concentración final de 10 mM.
- 8) Estrés por luz UV: las plantas se expusieron durante 15 min a 1.18 W/m<sup>2</sup> de luz provista por tubos Phillips™ TL40W/12.
- 9) Estrés por daño mecánico: las plantas fueron dañadas con pequeñas agujas.
- 10) Estrés por calor: las plantas fueron sometidas a una temperatura de 38 °C. Las muestras fueron tomadas luego de 0,25; 0,5; 1 y 3 horas de tratamiento; y se las dejó 1, 3, 9 y 24 horas recuperándose del estrés a 25 °C.
- 11) Estrés por calor en células en suspensión: Las células fueron tratadas de la misma forma que las plantas del inciso anterior.

Los tiempos de cada tratamiento fueron: 0; 30 min; 1; 3; 6; 12 y 24 horas. En los casos en los que se verifican respuestas rápidas, como estrés UV, daño mecánico, calor y déficit hídrico; se tomó una muestra a los 15 min. El tratamiento por calor consistió de un tiempo máximo de tres horas y posterior periodo de recuperación. A estos ensayos se los manipuló en forma independiente en los análisis por estrés y se los llamó “calor con recuperación”.

Luego de los tratamientos, las plantas fueron divididas en vástago y raíz para la toma de muestras. Se tomaron muestras con mezclas o *pools* de nueve plantas para cada ensayo de microarreglos, y a su vez se realizaron dos réplicas biológicas por condición, tiempo y tejido.

### II.2.1.1 Filtrado de los datos

La matriz de datos fue filtrada en función, en primer lugar, de la variabilidad entre réplicas y de la calidad de los microarreglos. La variabilidad fue determinada calculando las distancias euclidianas entre las mediciones de todos los microarreglos y haciendo luego un agrupamiento jerárquico mediante el método de “enlace promedio”. Hecho esto se inspeccionó el dendrograma resultante y aquellas réplicas que no formaron un grupo independiente fueron removidas. Se removieron de esta forma 29 microarreglos de la matriz, cuyos códigos son: X0011, X0012, X0111, X0112, X0211, X0212, X0311, X0312, X0021, X0022, X0121, X0122, X0721, X0722, X0421, X0422 C2\_1, X1321, X1322, X6611, X6612, X3522,

## II. Materiales y Métodos

---

X3521, X8623, X8624, X8423, X8424, X8712, X8715 (información sobre el tratamiento correspondiente al código en <http://www.weigelworld.org/resources/microarray/AtGenExpress/>). Quedaron entonces un total de 269 microarreglos.

Con el objetivo de utilizar sólo aquellos genes o *probesets* con un perfil de expresión informativo se realizó un filtrado de los *probesets*. Para esto se calculó el promedio, la desviación estándar ( $\sigma$ ) y el coeficiente de variación (CV) de cada perfil. El análisis por histogramas de la distribución de los tres parámetros permitió determinar los valores de corte o umbral. Se utilizó un criterio de filtrado en el que se requirió que la expresión del *probeset* tuviera  $K$  valores sobre un umbral  $A$  ( $KsA$ ), con  $K = 4$  y  $A = 5$ . El umbral de la  $\sigma$  fue 0,289; correspondientes al percentil 10 de las  $\sigma$ . Finalmente, el umbral elegido para el CV fue 0,104; correspondiente a la mediana de los CVs.

El filtrado de los *probesets* dejó 9612. Se reincorporaron seis *probesets* de HD-Zip y se removió el *probeset* de *AtHB54* que hibrida con un producto génico diferente; con lo que quedaron 9617 *probesets*.

### II.2.2 Reconstrucción de la red con ARACNE

El programa ARACNE requiere la información sobre los *probesets* que corresponden a factores de transcripción. Se obtuvo entonces una lista de genes con esta función de la base de datos AGRIS (<http://arabidopsis.med.ohio-state.edu/AtTFDB/>; Yilmaz y col., 2011). De los 1851 FTs recuperados, 830 estuvieron incluidos en la matriz de expresión filtrada.

Para seleccionar los diferentes parámetros que debieron proveerse al programa ARACNE se siguieron las recomendaciones de Margolin y col. (2006b). Para la selección del ancho de banda utilizado en el cálculo de las densidades de probabilidad con el estimador de núcleo Gaussiano se probaron los tres algoritmos posibles. Se hicieron reconstrucciones de prueba repetidas y el que más reproducibilidad tuvo fue el de “partición adaptativa” (del inglés, *adaptive partitioning*). El valor  $p$  para evaluar estadísticamente los valores de IM obtenidos fue  $1 \times 10^{-10}$ . La tolerancia para la remoción de relaciones mediante el algoritmo de *Data Processing Inequality* (DPI) fue de 0,005. Estos últimos dos valores se encuentran dentro del rango recomendado por los autores y también fueron variados en reconstrucciones de prueba para evaluar su impacto en la red resultante.

Utilizando estos parámetros se hizo la reconstrucción de la red con ARACNE a partir de 1000 réplicas de muestreo con reemplazo o *bootstrap*. La ejecución del programa se realizó en un grupo o *cluster* de cinco CPU Intel Quad-Core® (total de 20 núcleos) perteneciente al “Instituto Leloir” (Ciudad Autónoma de Buenos Aires) y requirió alrededor de 27 horas para completarse. De esta forma se obtuvieron 1000 reconstrucciones de la red que fueron combinadas en una única utilizando la rutina *getconsensus* provista por los autores del programa ARACNE, para la

## II. Materiales y Métodos

---

cual se seleccionó el valor  $p$  de  $1 \times 10^{-15}$  (Margolin y col., 2006b).

La manipulación de la matriz de datos y el posterior análisis de los módulos de actividad transcripcional (MATs) obtenidos de ARACNE fue realizada mayoritariamente con el lenguaje y entorno R para la estadística computacional (R Development Core Team, 2011). Además, en varios casos se utilizaron paquetes pertenecientes al proyecto “Bioconductor” (Gentleman y col., 2004).

### II.2.3 Análisis de promotores

Las secuencias de los promotores de los genes asociados a los *probesets* fueron obtenidas de la base de datos TAIR (versión 10, [ftp://ftp.arabidopsis.org/home/tair/Sequences/blast\\_datasets/TAIR10\\_blastsets/upstream\\_sequences/](ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR10_blastsets/upstream_sequences/)). Se utilizaron dos conjuntos de secuencias: 1000 y 3000 pb corriente arriba del sitio de inicio de la transcripción. La búsqueda de los motivos CAATNATTG y AATNATT se realizó con el lenguaje R y el paquete “Biostrings” (<http://bioc.ism.ac.jp/2.6/bioc/html/Biostrings.html>).

Para la búsqueda de motivos *de novo* se empleó el paquete “cosmo” (Bembom y col., 2007), también dentro del entorno R. Con el propósito de buscar motivos pseudopalindrómicos se impusieron restricciones al algoritmo. Se estableció una longitud mínima del motivo de 9 nucleótidos y máxima de 15 nucleótidos. Se definió que el motivo a buscar tuviera tres regiones: las dos repeticiones invertidas del palíndromo y una región central de longitud variable. Se utilizó el método TCM que busca la posible aparición de más de un motivo por secuencia.

### II.2.4 Estudio de regulación conjunta

La regulación conjunta de los genes en los diferentes tejidos y tratamientos fue investigada calculando primeramente la correlación entre los perfiles de expresión de todos los *probesets* en cada condición. Este valor de correlación ( $c$ ) fue convertido en una medida de disimilitud o distancia ( $d$ ) restándolo de la unidad y dividiendo el resultado por dos ( $d = (1 - c)/2$ ). Esta matriz fue sometida a un agrupamiento jerárquico mediante el método de “enlace promedio” con lo cual se obtuvo un dendrograma. Los *probesets* regulados en forma conjunta fueron seleccionados intersecando el dendrograma a un nivel de distancia de 0,15 y haciendo el recuento de los *probesets* que forman parte de clados de cinco o más *probesets*. Este valor fue considerado un indicador de la participación de cada gen HD-Zip en el estrés y tejido.

Los gráficos de la Figura IV.17 (p. 112) presentan los perfiles de expresión de los *probesets* regulados en forma coordinada según el criterio descrito más arriba. Para su preparación se generó un perfil de expresión normalizado ( $\mathbf{x}_{norm}$ ), es decir, se obtuvieron la media ( $\bar{x}$ ) y la desviación estándar ( $s$ ) de los valores de expresión de ese *probesets* en ese tejido y estrés ( $\mathbf{x}$ ) y

se hizo el siguiente cálculo:

$$x_{norm} = \frac{x - \bar{x}}{s}$$

### II.2.5 Estudio funcional con términos GO

El enriquecimiento de los términos GO fue analizado utilizando el programa desarrollado por Berenstein y Chernomoretz (2012, manuscrito en preparación). Primero se buscan los términos GO asociados a cada gen por las bases de datos. Luego se hace una prueba hipergeométrica para evaluar si un término está estadísticamente enriquecido. Ésta depende de cuantos genes asociados al término hay en el MAT y del total de genes que posee el MAT, así como de cuantos genes asociados al término GO hay en total y del total de genes analizados. Para esta prueba se elige un valor  $p$  que permitirá evaluar estadísticamente la hipótesis de enriquecimiento.

Luego, el programa evalúa si es posible seleccionar un término GO de mayor jerarquía que sea capaz de resumir la información de los términos GO enriquecidos sin incluir un número excesivo de términos GO que no se encontraron enriquecidos. El criterio de esta elección se basa en teoría de la información y los términos GO elegidos son denominados MICA.

En el presente trabajo se consideraron los términos GO con un número mínimo de cinco genes asociados y un número máximo de 300; en los que quedaran seleccionados un número mínimo de tres genes. Se utilizó un valor  $p$  de 0,1 para evaluar estadísticamente el enriquecimiento.

## **Capítulo III**

# **Fuentes de diversidad funcional en factores de transcripción HD-Zip de la subfamilia I**

### III.1 Introducción

#### III.1.1 Origen de las familias génicas de factores de transcripción vegetales

A lo largo de la evolución de las plantas, los eventos de duplicación génica y genómica han sido relativamente frecuentes, estos últimos particularmente en angiospermas (Blanc y Wolfe, 2004). De hecho, en comparación con los animales, las plantas son organismos mucho más tolerantes a la poliploidización y a la aneuploidización (Heslop-Harrison y Schwarzacher, 2011). Luego de estos eventos, en una gran cantidad de casos se han dado procesos de reorganización genómica que condujeron nuevamente a un organismo diploide, cuyo pasado evolutivo lo convierte en un paleopoliploide (Blanc y Wolfe, 2004).

El resultado de estas duplicaciones es la coexistencia inmediata de más de una copia de un mismo gen, generalmente con funciones prácticamente idénticas y por ende redundantes. Los procesos evolutivos subsiguientes llevan habitualmente a la pérdida de alguna de las copias, o a la conservación de ambas mediante la subfuncionalización (división de la función del gen ancestral entre las copias) o neofuncionalización (adquisición de nuevas funciones) de los genes duplicados (Lynch y Conery, 2000).

En *Arabidopsis thaliana*, un análisis por categorías funcionales mostró que la mayoría de los genes duplicados codifica FTs y otras proteínas con funciones de regulación (Maere y col., 2005). Esto permite explicar la gran multiplicidad de las familias génicas de FTs en ésta y otras especies vegetales.

#### III.1.2 Análisis filogenéticos de los FTs HD-Zip I

La reconstrucción de las relaciones evolutivas entre los genes permite proponer una historia de las duplicaciones ocurridas y así estimar su potencial grado de similitud funcional. Para esto es necesario aplicar herramientas de inferencia filogenética molecular cuyo resultado prototípico es el árbol filogenético

El estudio de las relaciones evolutivas entre los genes que codifican proteínas HD-Zip ha sido abordado en algunos trabajos. El trabajo pionero fue realizado utilizando el HD de las escasas proteínas de las subfamilias I y II conocidas en ese momento, en conjunto con proteínas con HD de otras familias (Chan y col., 1998). En un estudio posterior se empleó el dominio HD-Zip de 49 proteínas de las 4 subfamilias, de las cuales 11 correspondían al helecho *Craterostigma plantagineum* y el resto a otras especies, mayoritariamente a *A. thaliana* (Aso y col., 1999). Sakakibara y col. (2001) aislaron 10 genes que codifican HD-Zip del musgo *Physcomitrella patens* y realizaron un árbol filogenético con el método de máxima verosimilitud o *maximum likelihood* (ML) empleando una región de 84 aa de 96 proteínas de diferentes especies, entre las que había representantes de las cuatro subfamilias.

En un trabajo específico sobre la subfamilia I de *A. thaliana*, la reconstrucción filogenética fue realizada con los dominios HD-Zip mediante el método de máxima parsimonia (Henriksson y col., 2005). Los resultados fueron respaldados parcialmente por información de la estructura intrón-exón de cada gen y por la correspondencia entre las regiones genómicas duplicadas y los genes parálogos hallados (Henriksson y col., 2005). En función de este análisis se clasificaron los genes en seis clados, como se muestra en la Figura III.1:  $\alpha$  (*AtHB3*, *AtHB13*, *AtHB20* y *AtHB23*),  $\beta$  (*AtHB1*, *AtHB5*, *AtHB6* y *AtHB16*),  $\gamma$  (*AtHB7* y *AtHB12*),  $\delta$  (*AtHB21*, *AtHB40* y *AtHB53*),  $\varepsilon$  (*AtHB22* y *AtHB51*) y  $\phi$  (*AtHB52* y *AtHB54*).

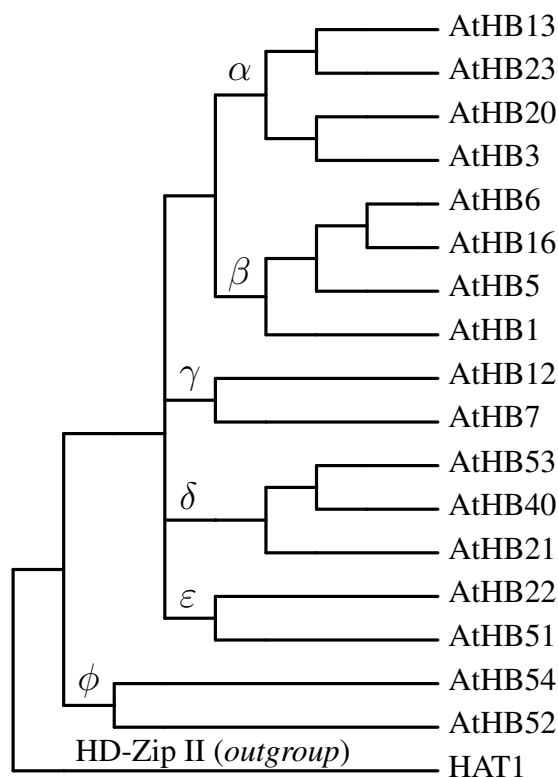
En otro análisis filogenético principalmente orientado a estudiar proteínas de arroz (*Oryza sativa*) se utilizaron para la reconstrucción las secuencias aminoacídicas de los dominios HD-Zip de miembros de las subfamilias I, II y III de diferentes especies (Agalou y col., 2008). Dentro de los FTs HD-Zip I, las relaciones filogenéticas de las proteínas de *A. thaliana* incluidas permitieron recuperar sólo cuatro ( $\alpha$ ,  $\gamma$ ,  $\delta$  y  $\varepsilon$ ) de los seis grupos monofiléticos hallados por Henriksson y col. (2005).

Recientemente se realizó una reconstrucción más amplia que incluyó todas las proteínas con HD de especies vegetales (más correctamente, del grupo Archaeplastida) con genoma secuenciado: seis embriofitas o plantas terrestres (*Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa*, *Zea mays*, *Selaginella moellendorffii* y *Physcomitrella patens*), tres algas verdes (*Ceratopteris richardii*, *Ostreococcus lucimarinus* y *Ostreococcus tauri*) y un alga roja (*Cyanodioschizon merolae*; Mukherjee y col., 2009). Para esto se utilizó únicamente el dominio HD y el método elegido fue ML. No hallaron proteínas HD-Zip en las algas verdes o el alga roja, por lo que los autores concluyeron que sólo se encontrarían en embriofitas.

#### III.1.3 Diversidad funcional en la subfamilia HD-Zip I

El gran número de miembros que posee la familia HD-Zip en diferentes especies, y particularmente la subfamilia I (ver Tabla I.2), permite anticipar que estos FTs deberían tener funciones diferentes, en mayor o menor medida, para haber sido conservados a lo largo de la evolución. Lo expuesto en la Sección I.6 pone de manifiesto que así es. Aunque es posible encontrar algunas características comunes a gran parte de los miembros de este grupo, los procesos particulares en los que participan son muy variados.

Si consideramos estos FTs desde una perspectiva molecular es posible reconocer cierto grado de homogeneidad dentro del grupo. Las proteínas cuya capacidad de unión a ADN fue evaluada *in vitro*, *AtHB1* (Sessa y col., 1993), *HaHB4* (Palena y col., 1999), *AtHB3*, *AtHB5*, *AtHB6*, *AtHB13*, *AtHB16* (Johannesson y col., 2001) y tres proteínas de arroz (*OsHOX4*, *OsHOX5* y *OsHOX6*; Meijer y col., 2000); unen con máxima afinidad la secuencia CAAT(A/T)ATTG. Esto también fue comprobado *in vivo* empleando promotores artificiales con repeticiones de este motivo en ensayos de transformación transitoria de células vegetales (Aoyama y col.,



**Figura III.1:** Filogenia de los genes HD-Zip I de *A. thaliana* adaptada de Henriksson y col. (2005). La raíz del árbol fue elegida gracias a la inclusión de una proteína HD-Zip II en el análisis.

1995; Henriksson y col., 2005; Meijer y col., 2000), o con el propio promotor en el caso de AtHB6 (Himmelbach y col., 2002). La unión específica a esta secuencia pseudopalindrómica fue verificada también por coimmunoprecipitación de la cromatina (ChIP) para MtHB1 de *Medicago truncatula* (conocido también como HB1; Ariel y col., 2010).

Aunque se ha propuesto que para AtHB5, AtHB6, AtHB16 la identidad de la posición central podría no tener efecto en la afinidad de unión (Johannesson y col., 2001), el conjunto de las evidencias experimentales obtenidas indica que la unión a diferentes elementos en *cis* no sería un factor distintivo relevante para explicar las diferencias funcionales entre las proteínas HD-Zip I.

Si evaluamos ahora los resultados obtenidos por diferentes grupos de investigación respecto a las capacidades de dimerización de estas proteínas nos encontramos con que, aunque no serían capaces de heterodimerizar con proteínas de la subfamilia II (Gonzalez y col., 1997; Meijer y col., 2000; Sessa y col., 1993), la heterodimerización entre miembros de una misma subfamilia sí es un evento posible (Deng y col., 2002; Johannesson y col., 2001; Meijer y col., 2000). Lógicamente esta heterodimerización requiere de la coexpresión de dos de estos genes en el mismo tejido y estadio de desarrollo.



### III.1.3.1 Patrones de expresión de los genes HD-Zip I de *A. thaliana*

Quizás el patrón de expresión de cada gen HD-Zip I sea la característica distintiva que permita explicar más sencillamente, al menos en forma parcial, las diferencias funcionales. Se puede ilustrar este razonamiento revisando estudios previos sobre los promotores de los genes de esta subfamilia de *A. thaliana* en fusiones al gen que codifica la  $\beta$ -glucuronidasa (*GUS*).

El análisis histoquímico de plantas transformadas con 2000 pares de bases (pb) del promotor de *AtHB5* fusionadas a *GUS* mostró que este promotor dirige la expresión en la plántula entera apenas germina y a medida que ésta se desarrolla, la expresión va restringiéndose al hipocotilo, siendo máxima en la región de transición entre éste y la raíz. El tratamiento de las plántulas con ABA produjo una disminución de la expresión en el hipocotilo y un incremento en la región de transición. A medida que la planta crece comenzó a detectarse expresión en los peciolos de los cotiledones y en las hojas jóvenes. En plantas adultas se observó expresión en el tejido vascular de las zonas de diferenciación de la raíz principal y de las laterales, y en el tejido vascular de hojas caulinares, tallos y pedicelos de flores y silicuas, pero no en el resto de la flor (Johannesson y col., 2003).

En ensayos similares, unos 1400 pb del promotor de *AtHB6* (perteneciente al clado  $\beta$ , al igual que *AtHB5*, Henriksson y col., 2005) dirigieron la expresión de *GUS* a la epidermis de los cotiledones y a la región de división de la raíz en plántulas. Con el desarrollo comenzó a detectarse en los primordios de las hojas, en las raíces laterales y fuertemente en los estomas; no se registró expresión en hipocotilo. En plantas adultas la expresión en hojas se restringió al tejido vascular, a las células en división de raíces y a los pistilos de las flores. El ABA y el déficit hídrico indujeron la expresión pero no cambiaron el patrón tisular (Söderman y col., 1999).

*AtHB7* y *AtHB12*, genes parálogos del clado  $\gamma$  (Henriksson y col., 2005), poseen patrones de expresión que presentan cierta similitud (Olsson y col., 2004). En plántulas cultivadas en condiciones normales, el promotor de *AtHB12* dirigió la expresión de *GUS*, pero en niveles bajos, en el primordio de hojas, en los peciolos en desarrollo y en la parte basal de las hojas jóvenes de la roseta y los primordios de las raíces laterales. En estas condiciones la expresión de *GUS* controlada por el promotor de *AtHB7* fue detectable sólo en el ápice de las raíces. El tratamiento con ABA produjo un incremento de la expresión dirigida por el promotor de *AtHB12* en todos los tejidos mencionados y además en los cotiledones y las hojas en forma completa, en la vasculatura de las raíces y en sus ápices. El tratamiento con ABA en las plantas transformadas con el promotor de *AtHB7* también produjo un incremento muy marcado en los niveles de expresión de *GUS* que se expandió a casi todos los tejidos. En las plantas adultas ambos promotores dirigieron la expresión a las yemas axilares, a las anteras y a las semillas en desarrollo. Con el promotor de *AtHB12* también se detectó tinción en la vasculatura de las hojas y del tallo de la inflorescencia. El tratamiento de estrés hídrico incrementó fuertemente la expresión dirigida por ambos promotores, se observó actividad *GUS* en el tallo de la inflorescencia y en las

### III.1. Introducción

---

hojas de la roseta, especialmente en el tejido vascular, y en los órganos de la flor (Hjellström y col., 2003; Olsson y col., 2004).

El estudio de 2000 pb del promotor de *AtHB20* del clado  $\alpha$  mostró que el mismo dirige la expresión al tejido vascular de las hojas durante todo su desarrollo (Mattsson y col., 2003). El análisis del promotor de *AtHB13* del mismo clado permitió observar tinción en la región basal del hipocotilo y de los cotiledones en plántulas, específicamente en la vasculatura. En plantas maduras la actividad *GUS* se observó en tejido vascular de la base de las hojas y particularmente en la unión de éstas con el tallo; en flores se dio un patrón análogo en los sépalos y los pétalos (Hanson y col., 2002). El promotor de *AtHB23*, también del clado  $\alpha$ , fue activo específicamente en el meristema apical del tallo, en los primordios de las hojas y en la cara adaxial de las hojas jóvenes; en flores sólo se observó tinción en la punta de las silicuas (Kim y col., 2007).

En el caso de *AtHB51* los estudios del promotor mostraron expresión en hojas jóvenes, principalmente en el ápice, y luego del pasaje al estadio reproductivo se observó en el primordio de la flor y en diferentes órganos y estadios de su desarrollo (Saddic y col., 2006). La expresión en hojas se restringió a los márgenes y con la maduración se conservó en los márgenes proximales abarcando la zona de aserrados. No se observó expresión en hipocotilo o raíces.

Un análisis conjunto por *Northern blot* de los transcritos de todos los genes de la subfamilia permitió reconocer, por un lado, que en plántulas enteras de 10 días los tratamientos con ABA y NaCl indujeron fuertemente los transcritos de *AtHB7* y *AtHB12* y en menor medida los de *AtHB6*, *AtHB21*, *AtHB40* y *AtHB53*; pero redujeron los niveles de ARNm de *AtHB3*, *AtHB5*, *AtHB23* y *AtHB52*. *AtHB1* y *AtHB16* fueron reprimidos por el tratamiento con NaCl pero no con ABA. El tratamiento con bajas temperaturas (entre 2 y 4°C) indujo los niveles de los transcritos de *AtHB6*, *AtHB7*, *AtHB12*, *AtHB40* y *AtHB53* mientras que reprimió los de *AtHB1*, *AtHB16* y *AtHB52*. El resto de los genes no fueron afectados en su expresión por estos tratamientos o sus transcritos no fueron detectados en estas condiciones (Henriksson y col., 2005).

En el mismo estudio se investigó el efecto en plántulas de 4 días de crecimiento en luz blanca, azul u oscuridad (30 min). *AtHB52* fue inducido por luz azul y más fuertemente en oscuridad. *AtHB53* fue inducido por oscuridad pero no por luz azul. *AtHB3*, *AtHB23*, *AtHB1* y *AtHB16* mostraron sólo una leve inducción en oscuridad. Por el contrario, *AtHB5*, *AtHB6*, *AtHB7*, *AtHB12*, *AtHB13* y *AtHB20* presentaron mayores niveles de transcritos en luz blanca que en las otras condiciones de iluminación. La expresión de *AtHB21*, *AtHB22*, *AtHB40*, *AtHB51* y *AtHB54* fue baja o no pudo ser detectada. Los niveles de los transcritos de *AtHB1*, *AtHB7*, *AtHB12*, *AtHB13* y *AtHB20* en luz azul fueron aproximadamente la mitad de los medidos en luz blanca (Henriksson y col., 2005).

Por último, en ese trabajo también se analizó la expresión en las raíces, las hojas, el tallo, las flores y las silicuas de plantas adultas, y en plántulas enteras a los 5 y a los 12 días (Henriksson y col., 2005). En este caso la técnica utilizada fue RT-PCR, la cual en principio no permite reconocer más que la ausencia o presencia de transcritos. Los genes de los clados  $\beta$  (*AtHB1*,

### III.1. Introducción

---

*AtHB5*, *AtHB6* y *AtHB16*),  $\gamma$  (*AtHB7* y *AtHB12*) y los genes *AtHB20*, *AtHB21*, *AtHB51* se expresaron prácticamente en todos los tejidos evaluados, pero el resto mostró mayor o menor especificidad de tejido. La detección de todos los transcritos indicó que ninguno de los 17 HD-Zip identificados es un pseudogén (Henriksson y col., 2005).

#### III.1.3.2 Expresión ectópica y constitutiva de los genes HD-Zip I

Los estudios de expresión realizados en *A. thaliana* involucraron una elevada proporción de los genes de la subfamilia y la diversidad de patrones de expresión encontrados permite por sí misma explicar parcialmente las diferencias funcionales entre éstos. Sin embargo, muchos de los trabajos que investigaron su función utilizaron como herramienta la expresión ectópica y constitutiva en *A. thaliana* empleando el promotor *35SCaMV*, lo cual hace que el fenotipo de las plantas transgénicas resultantes sea directamente comparable. Esto se debe a que, al margen de las variaciones que surjan producto del lugar de inserción de la construcción genética, el patrón de expresión del transgén tendrá una gran similitud. Haciendo este análisis es posible notar que hay grandes diferencias en los fenotipos generados, particularmente en la expresión de genes que no están estrechamente relacionados a nivel filogenético.

Las plantas transgénicas de *A. thaliana* con la construcción *35SCaMV::AtHB7* o *::AtHB12*, del clado  $\gamma$ , presentan un fenotipo similar. Se observó un retraso en el desarrollo, un acortamiento de los peciolo, una reducción del área foliar y otras características morfológicas características de plantas salvajes sometidas a sequía (Hjellström y col., 2003; Olsson y col., 2004). Las plantas de *A. thaliana* con la construcción *35SCaMV::HaHB4*, que codifica una proteína con elevada similitud a los anteriores, desarrollaron un fenotipo en principio similar, pero además presentaron insensibilidad al etileno y un severo retraso en la senescencia, una elevada tolerancia a la sequía y al ataque de herbívoros (Dezar y col., 2005b; Manavella y col., 2006, 2008b). Ninguno de estos fenotipos fue informado en los trabajos en los que se investigaron *AtHB7* y *AtHB12* (Hjellström y col., 2003; Olsson y col., 2004).

La construcción *35SCaMV::AtHB13* y la de su ortólogo de girasol *35SCaMV::HaHB1* (*HaHB1*), generaron plantas de *A. thaliana* con un retraso en el crecimiento y un pronunciado aserrado en las hojas (Cabello y col., 2012; Hanson y col., 2001). El tratamiento de plántulas con sacarosa produjo un angostamiento de los cotiledones (Cabello y col., 2010; Hanson y col., 2001). La expresión ectópica y constitutiva de *AtHB23*, del mismo clado que *AtHB13*, no generó un fenotipo distintivo, tampoco cuando las plantas transgénicas fueron sometidas al mismo tratamiento con azúcares que las plantas *35SCaMV::AtHB13* (Kim y col., 2007).

Las plantas de *A. thaliana* con la construcción *35SCaMV::AtHB16* mostraron un acortamiento del tallo y una reducción en el tamaño de las hojas. Pero además de estas características presentadas por plantas transgénicas de otros genes HD-Zip I, se observó que tenían un retraso en el pasaje del estadio vegetativo al reproductivo que fue menos sensible a cambios en el

fotoperiodo, y mayor sensibilidad a la inhibición de la luz azul sobre la elongación del hipocotilo (Wang y col., 2003).

Éstos son algunos de los ejemplos que ponen en evidencia que hay diferencias funcionales que no pueden explicarse solamente por los patrones de expresión diferentes de los genes HD-Zip. Si a esto sumamos la homogeneidad en el grupo respecto al elemento en *cis* reconocido, y teniendo en cuenta que el dominio HD-Zip representa aproximadamente un tercio de la proteína completa en esta subfamilia, resulta interesante preguntarse si la diversidad funcional entre sus miembros reside parcialmente en las diferencias estructurales que presentan las proteínas fuera del dominio conservado HD-Zip.

#### III.1.4 Evidencia de funcionalidad de regiones fuera del dominio HD-Zip

Existe evidencia respecto a la funcionalidad de regiones que se encuentran fuera del dominio HD-Zip. Lee y col. (2001) hallaron que AtHB12 tiene la capacidad de transactivar en experimentos de simple híbrido en levaduras, y que esta capacidad estaría asociada a su región carboxilo terminal (RCT). Fusiones de diferentes partes de la proteína al dominio de unión a ADN de GAL4 demuestran que la proteína completa o su RCT son capaces de transactivar, mientras que la fusión únicamente al dominio HD-Zip no. Estas construcciones permitieron reconocer que la actividad se restringiría a los aa 180 a 235, la porción final del extremo carboxilo.

Posteriormente se mostró que *AtHB12* es capaz de complementar funcionalmente levaduras mutantes sensibles a NaCl por deficiencia en una calcineurina, pero sólo cuando la proteína posee su RCT completo. A pesar de que esta evidencia fue obtenida en un organismo de otro reino, podría indicar que la RCT de *AtHB12* tiene importancia para la funcionalidad del FT en el contexto de una red de regulación (Shin y col., 2004).

El estudio del gen *VRS1* de cebada (*Hordeum vulgare*) también permitió proponer la funcionalidad de la RCT. La cebada posee el potencial de desarrollar tres espiguillas sobre cada nodo del raquis. En la variedad salvaje se aborta el desarrollo de las espiguillas laterales dando un fenotipo llamado “de dos hileras”, pero la variedad domesticada desarrolla las tres espiguillas generando el fenotipo de “seis hileras”. Se descubrió que el gen responsable de esta alteración en el desarrollo era el alelo recesivo del HD-Zip *vrs1* (Komatsuda y col., 2007). Luego se descubrió que la cebada posee otro gen muy similar, *HvHOX2*, posiblemente proveniente de una duplicación reciente. Resulta interesante que el alineamiento de las proteínas *HvHOX2*, *vrs1* y proteínas muy similares de otras especies mostró que la RCT de *vrs1*, que es 14 aa menor a la de *HvHOX2*, carece de un motivo conservado localizado en la porción terminal (Sakuma y col., 2010). Los autores propusieron que esta región podría ser responsable de la interacción de *HvHOX2* con algún coactivador, y que la falta de esta región en *VRS1* impide dicha interacción. Esto generaría una proteína que interfiere con la actividad de *HvHOX2* en los primordios de las espiguillas laterales, interrumpiendo así su desarrollo.

La hipótesis de la interacción con un cofactor estaría parcialmente respaldada por un trabajo que demostró que HaHB4 interactúa *in vitro* con la proteína StMBF1 de papa (*Solanum tuberosum*), un conocido coactivador transcripcional cuyo ortólogo de *Drosophila melanogaster* es capaz de interactuar con TBP. Sin embargo, la interacción se observó utilizando únicamente el dominio HD-Zip de HaHB4, sin su RCT (Zanetti y col., 2004).

Otro caso es el de la proteína TL (Tendril-less) de arveja, codificada por un gen cuyo alelo mutante *tl* es responsable de un fenotipo de conversión de zarcillos en folíolos. Esto sucede porque se libera la inhibición para completar el desarrollo laminar (Hofer y col., 2009). Notoriamente, esta alteración causada por la supresión de la expresión de *TL* también es generada por un alelo mutante que expresa una proteína que carece únicamente de los últimos 12 aa.

Basándose entonces en las diferentes observaciones realizadas por distintos grupos de investigación, resulta interesarse preguntarse cuál es la importancia y la función que desarrollan las regiones y residuos que se encuentran fuera del dominio HD-Zip. Los variados fenotipos observados en las plantas transgénicas sugieren que estas regiones podrían, en algún grado, ser factores determinantes de las vías de señales en las cuales cada proteína puede participar.

### III.2 Objetivos

En función de lo expuesto, se formuló la hipótesis de que las regiones aminoacídicas que se encuentran por fuera del dominio conservado HD-Zip son potenciales factores involucrados en la determinación de la función de los diferentes FTs HD-Zip I. A partir de esto se propuso el objetivo de investigar las características de estas regiones y evaluar su importancia a nivel funcional. Para lograrlo, se plantearon los siguientes objetivos particulares:

1. Obtener secuencias de proteínas HD-Zip I de varias especies y hacer con éstas una reconstrucción filogenética.
2. Reconocer clados o grupos de secuencias relacionadas que permitan identificar proteínas con potenciales propiedades comunes.
3. Analizar las regiones amino y carboxilo terminales (RNT y RCT) en búsqueda de posibles características funcionales.
4. Evaluar la importancia de la RCT del FT de girasol HaHB1 mediante la expresión en *Arabidopsis thaliana* de una versión quimérica que contenga una fusión de la RNT y el dominio HD-Zip de HaHB4 con esta RCT. Junto a diferentes controles, estudiar los fenotipos de las plantas frente a estímulos característicos de los genotipos que expresan las proteínas HaHB1 y HaHB4 salvajes.

## III.3 Resultados

### III.3.1 La base de datos utilizada para el estudio contiene 178 proteínas HD-Zip I de diferentes especies

Para poder encontrar características relevantes fuera de los dominios conocidos en las proteínas HD-Zip I se decidió comenzar con un análisis de secuencias *in silico*. En primer lugar, esto requirió de la elección de un conjunto de secuencias o *dataset*. Con este fin se combinaron las secuencias aminoacídicas de FTs HD-Zip I halladas en especies con genoma secuenciado (Mukherjee y col., 2009), con las secuencias obtenidas a través de la herramienta de búsqueda CDART (*Conserved Domain Architecture Retrieval Tool*; Geer y col., 2002). Esta herramienta permite recuperar secuencias de la base de datos de proteínas no redundantes del NCBI (*National Center for Biotechnology Information*) con una arquitectura de dominios similar a una proteína dada. Sin ser uno de los métodos más exhaustivos para este tipo de búsquedas, permitió obtener un número satisfactorio de proteínas para el análisis que se pretendía abordar.

En primer lugar, se observó que las secuencias utilizadas por Mukherjee y col. (2009) no presentaban el tripéptido YES (Tyr-Glu-Ser) en aquellos casos en que la secuencia depositada en bases de datos públicas sí lo tenían. Este error fue confirmado por los autores (Bürglin, T.R., comunicación personal) y corregido para este estudio.

Las secuencias obtenidas de CDART fueron buscadas ingresando el dominio HD-Zip de la proteína AtHB13 con la que se obtuvieron tanto secuencias de proteínas HD-Zip I como HD-Zip II, ambas familias con una arquitectura de dominios muy similar. Las proteínas de la subfamilia II fueron removidas mediante un alineamiento con MAFFT (Kato y Toh, 2008). El reordenamiento de las secuencias en función de la similitud permitió diferenciar fácilmente las proteínas de una y otra subfamilia, no sólo por la presencia del motivo CPSCE (sólo presente en la subfamilia II), sino por la similitud en el dominio HD-Zip.

Estos conjuntos de secuencias fueron combinados y las secuencias redundantes fueron reconocidas y removidas utilizando la herramienta *skipredundant* del paquete de programas EMBOSS (Rice, 2000). Por último, se agregaron manualmente las secuencias de HaHB1, HaHB11 y MtHB1, estudiadas en el grupo donde se realizó el trabajo de Tesis; porque no habían sido recuperadas mediante la búsqueda realizada. El conjunto final contuvo 178 secuencias aminoacídicas. En la Tabla III.1 se muestra la distribución de secuencias por especie y en la Tabla III.2 se muestra la distribución de secuencias por grupos taxonómicos.

### III.3. Resultados

**Tabla III.1:** Se resumen las especies representadas en la base de datos y el número de proteínas presentes de cada una.

Abreviatura	Especie	Taxón <sup>a</sup>	Número de miembros
At	<i>Arabidopsis thaliana</i>	Dicot	17
Bn	<i>Brassica napus</i>	Dicot	1
Brs	<i>Brassica rapa</i>	Dicot	1
Cp	<i>Craterostigma plantagineum</i>	Dicot	3
Cr	<i>Catharanthus roseus</i>	Dicot	2
Cr	<i>Ceratopteris richardii</i>	Helecho	3
Dc	<i>Daucus carota</i>	Dicot	4
Gh	<i>Gossypium hirsutum</i>	Dicot	1
Gm	<i>Glycine max</i>	Dicot	11
Ha	<i>Helianthus annuus</i>	Dicot	3
Lj	<i>Lotus japonicus</i>	Dicot	1
Mt	<i>Medicago truncatula</i>	Dicot	4
Nt	<i>Nicotiana tabacum</i>	Dicot	1
Os	<i>Oryza sativa</i>	Monocot	14
Pg	<i>Picea glauca</i>	Conífera	1
Pp	<i>Physcomitrella patens</i>	Musgo	17
Ps	<i>Picea sitchensis</i>	Conífera	5
Ps	<i>Pisum sativum</i>	Dicot	1
Pt	<i>Populus trichocarpa</i>	Dicot	21
Pv	<i>Phaseolus vulgaris</i>	Dicot	3
Rc	<i>Ricinus communis</i>	Dicot	7
Sb	<i>Sorghum bicolor</i>	Monocot	7
Sd	<i>Solanum demissum</i>	Dicot	1
Sl	<i>Solanum lycopersicum</i>	Dicot	2
Sl	<i>Silene latifolia</i>	Dicot	1
Sm	<i>Selaginella moellendorffii</i>	Helecho	4
Ta	<i>Triticum aestivum</i>	Monocot	2
Vv	<i>Vitis vinifera</i>	Dicot	15
Ze	<i>Zinnia elegans</i>	Dicot	1
Zm	<i>Zea mays</i>	Monocot	24

<sup>a</sup>Dicot: dicotiledóneas, Monocot: monocotiledóneas

#### III.3.2 El análisis filogenético de los FTs HD-Zip I permitió reconocer seis clados

La reconstrucción filogenética involucró la generación de tres árboles filogenéticos denominados HZT, HZT+OG y CST. En el árbol HZT se utilizó para el alineamiento múltiple únicamente el dominio HD-Zip, y fue considerada la reconstrucción filogenética de referencia por incluir sólo sitios homólogos a todas las proteínas (aunque se consideraron *indels* o inserciones-delecciones en los pocos casos que fue necesario). Para el árbol HZT+OG se utilizaron las mismas secuencias que para el árbol HZT, pero se añadieron los dominios HD-Zip de tres proteínas de *A. thaliana*



### III.3. Resultados

**Tabla III.2:** Se presentan clasificadas según el taxón las proteínas de la base de datos y el número de especies representadas.

Grupo taxonómico	Número de especies	Número de miembros
Musgos (Briofitas)	1	17
Helechos (Pteridofitas)	2	7
Coníferas	2	6
Monocotiledóneas	5	47
Dicotiledóneas	22	101

de la subfamilia HD-Zip II (*AtHB17*, *HAT1* y *HAT22*). El objetivo fue utilizarlas como grupo externo o *outgroup*, y así poder ubicar la raíz del árbol HZT y CST.

Finalmente, el árbol CST fue construido alineando las secuencias completas de los 178 FTs, con lo cual, se consideraron sitios no homólogos a todas las proteínas. En este caso el objetivo fue que las regiones N y C terminales (RNT and RCT, respectivamente) homólogas a subgrupos de proteínas sirvieran para guiar la definición de clados de proteínas con estrecha relación filogenética.

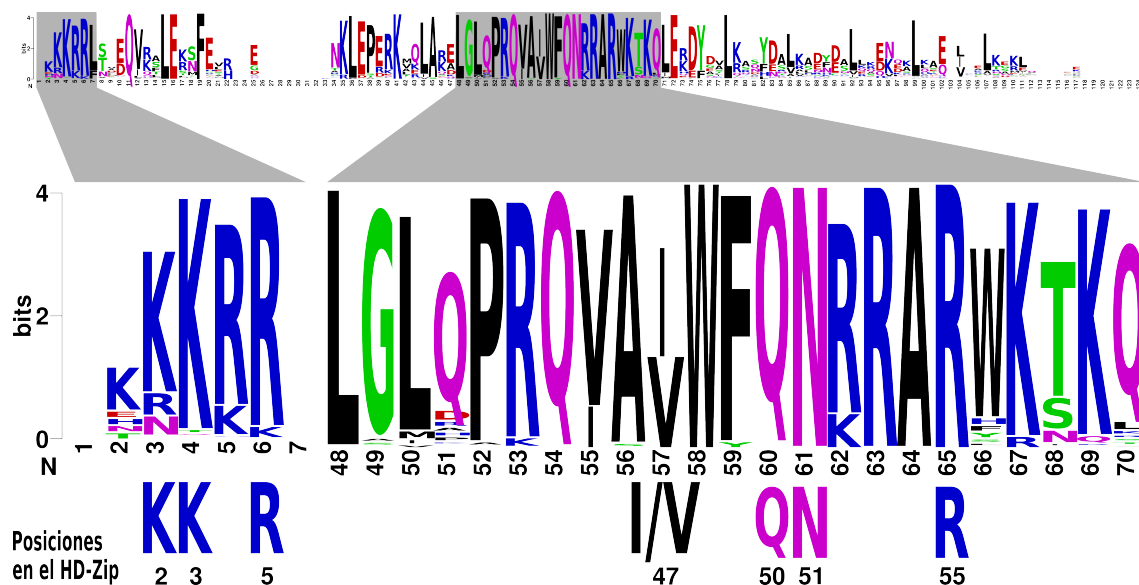
#### III.3.2.1 Manipulación y alineamiento de las secuencias

Los conjuntos de secuencias procesados para realizar los alineamientos múltiples en la construcción de los árboles HZT y HZT+OG se obtuvieron removiendo las RNT y RCT de las proteínas, dejando así únicamente el dominio HD-Zip. Esto fue realizado reconociendo el HD y el cierre de leucinas (LZ) mediante el programa HMMER (Eddy, 1998) utilizando los modelos ocultos de Markov o *hidden Markov models* (HMM) de los dominios obtenidos de la base de datos PFAM (Finn y col., 2010). Esta manipulación de secuencias fue realizada utilizando comandos (o *scripts*) escritos en el lenguaje de programación Perl haciendo uso además del conjunto de módulos Bioperl (Stajich y col., 2002).

Cada uno de los conjuntos de secuencias resultante fue entonces alineado mediante el programa MAFFT (Katoh y Toh, 2008) y los resultados fueron inspeccionados y curados manualmente.

#### III.3.2.2 La conservación de residuos clave en el HD-Zip sugiere una escasa variación en las secuencias de ADN reconocidas

El alineamiento de las 178 proteínas permitió determinar el grado de conservación de los residuos previamente caracterizados como fundamentales en la determinación de la especificidad de unión al ADN de estos FTs (Palena y col., 2001; Sessa y col., 1997; Tron y col., 2001, 2005). La Figura III.2 permite observar un diagrama del tipo *sequence logo* que simplifica el análisis del alineamiento completo (ver Apéndice V.2). Puede verificarse que las posiciones del alineamiento correspondientes a los principales residuos responsables del reconocimiento y de la especificidad



**Figura III.2:** *Sequence logo* resumiendo el alineamiento del dominio HD-Zip. Puede verse claramente la elevada conservación de los residuos fundamentales para la determinación de la especificidad de unión a ADN. Los números que figuran debajo del *sequence logo* representan la posición del residuo en el alineamiento, la cual no corresponde con la posición en el HD por la presencia de inserciones en algunas proteínas del alineamiento.

de unión, es decir: Lys 2, Lys 3, Arg 5, Ile/Val 47, Gln 50, Asn 51 y Arg 55; muestran un muy elevado grado de conservación entre las 178 proteínas analizadas.

Esto sugiere, en términos generales, que no debería haber una variación significativa en la especificidad de la secuencia de ADN de los elementos en *cis* reconocidos por las proteínas HD-Zip I analizadas.

#### III.3.2.3 Árboles filogenéticos y grupos con RCTs similares

Cada alineamiento fue ingresado en el programa PhyML (Guindon y Gascuel, 2003) para la reconstrucción de los árboles filogenéticos basados en el principio de ML. El modelo de evolución en cada caso fue determinado previamente con el programa ProtTest (Abascal y col., 2005). En los tres casos se usó el modelo JTT+I+G (Jones y col., 1992; Reeves, 1992; Yang, 1993). Además se hicieron 100 réplicas de *bootstrap* para los árboles HZT y HZT+OG, y 144 réplicas para el árbol CST.

La raíz del árbol HZT+OG (ver Apéndice V.2) se ubicó en la rama que une al clado que se generó entre las proteínas HD-Zip II (*outgroup*) con el clado de las proteínas HD-Zip I. El sitio de unión de esta rama en el clado HD-Zip I fue utilizado para enraizar los árboles HZT y CST, que se muestran enfrentados en la Figura III.4. Las ramas con un soporte estadístico de *bootstrap* menor a 50 % fueron colapsadas, excepto en el caso de los grupos del árbol HZT que tuvieron un soporte mayor al 50 % en el árbol CST (ver la Tabla III.3 y la explicación a continuación).

### III.3. Resultados

La observación de las RCTs en el alineamiento de las proteínas completas mostró que hay regiones de conservación significativa (ver Figura III.3). La correlación entre los clados hallados en los árboles filogenéticos y la conservación en las RCTs permitió definir 6 grupos (nombrados con números romanos, del I al VI) de los cuales el grupo I fue subdividido en Ia, Ib y Ic debido a las secuencias divergentes de sus RCTs. Este criterio de agrupamiento involucró sólo clados con un elevado soporte estadístico en el árbol HZT y/o en el árbol CST. En la Tabla III.3 se exponen las diferencias en los valores de *bootstrap* obtenidos en cada árbol.

**Tabla III.3:** Valores de *bootstrap* de los diferentes grupos en los árboles HZT y CST. Los valores son porcentuales y se resalta en negrita el mayor.

Grupo	I	Ia	Ib	Ic	II	III	IV	V	VI
HZT	<b>100</b>	99	<b>83</b>	40	53	37	21	95	99
CST	90	<b>100</b>	69	<b>93</b>	<b>69</b>	<b>82</b>	<b>63</b>	pf <sup>a</sup>	<b>100</b>

<sup>a</sup>parafilético

Puede verse que a pesar de que los grupos Ic, II, III y IV no tienen un soporte muy elevado en el árbol HZT, esos valores se incrementan sensiblemente en el árbol CST (ver Tabla III.3) debido a que para la reconstrucción se usan regiones homólogas a subgrupos de proteínas.

Los grupos I, II, III, V y VI están compuestos por proteínas de plantas monocotiledóneas y dicotiledóneas que forman subclados en concordancia con su relación evolutiva. Quedaron excluidos de los seis grupos algunos FTs de plantas dicotiledóneas y las 27 proteínas de musgos, helechos y coníferas. Las 17 proteínas HD-Zip de *Physcomitrella patens* formaron un clado independiente que fue llamado *Pp*. Las proteínas de especies de angiospermas con genoma secuenciado (*Arabidopsis thaliana*, *Oryza sativa*, *Zea mays* y *Populus trichocarpa*) presentaron al menos un miembro en cada uno de los grupos, a excepción de *P. trichocarpa* que no tiene miembros en el grupo III y *A. thaliana* en el IV. Este último grupo está formado exclusivamente por proteínas de plantas dicotiledóneas.

#### III.3.3 Las proteínas HD-Zip de cada grupo presentan motivos conservados en sus RCTs

Como se mencionó anteriormente, la importancia de la RCT había sido puesta en evidencia en trabajos anteriores (Hofer y col., 2009; Lee y col., 2001; Sakuma y col., 2010; Shin y col., 2004), por lo cual resultó de interés explorar en mayor detalle estas regiones. La inspección de los alineamientos permitió observar el evidente enriquecimiento en Trp presente en las regiones finales de las RCTs. En la Figura III.5 se presenta esta observación en forma cuantitativa.

Para profundizar este análisis se empleó un método de descubrimiento de motivos mediante el programa MEME (Bailey y col., 2006). Usando únicamente la secuencia de las RCTs, un

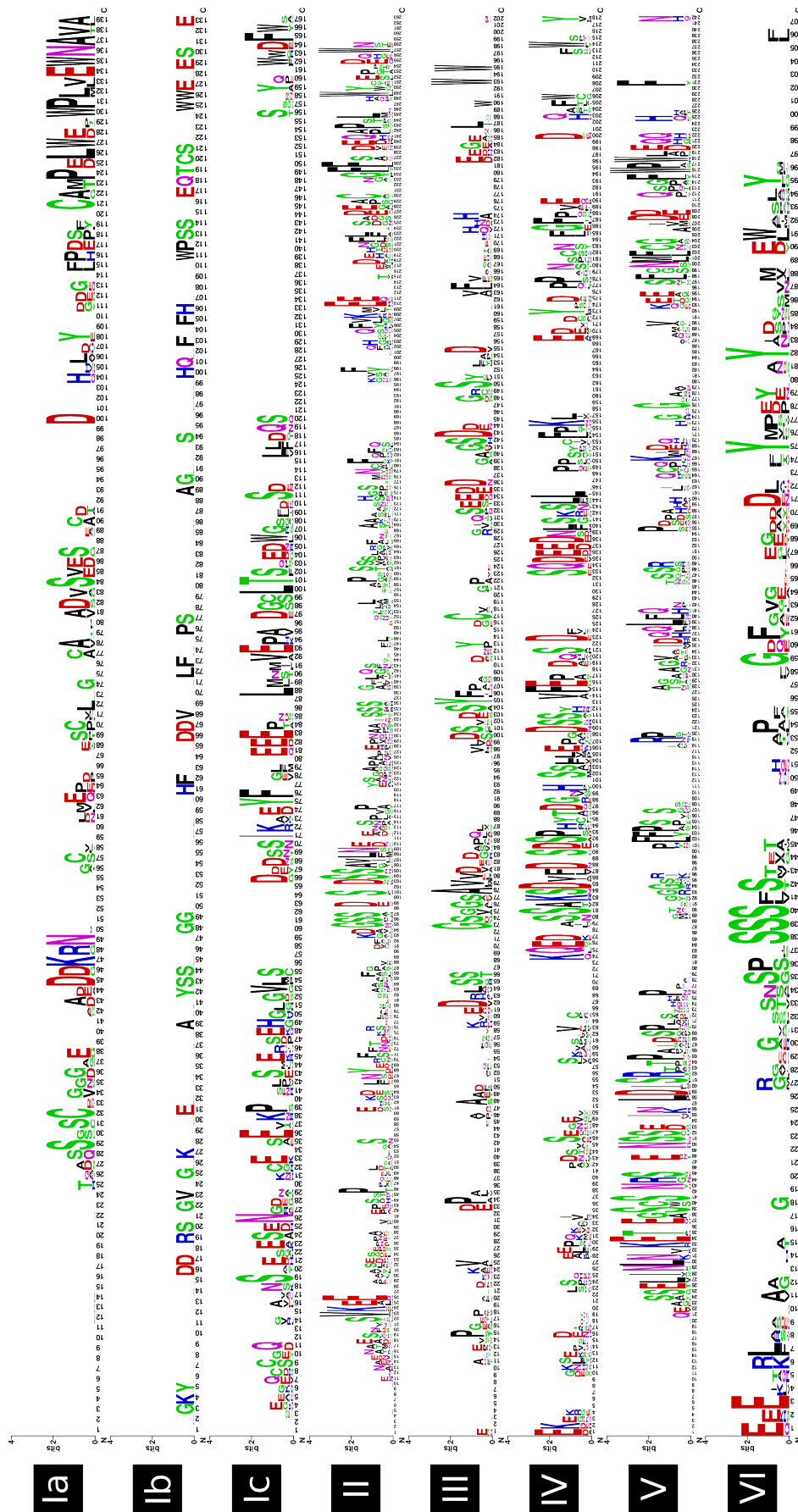


Figura III.3: *Sequence logos* mostrando el grado de conservación de las RCTs en las proteínas de los diferentes grupos. La altura de cada letra representa el grado de conservación del aminoácido en esa posición.

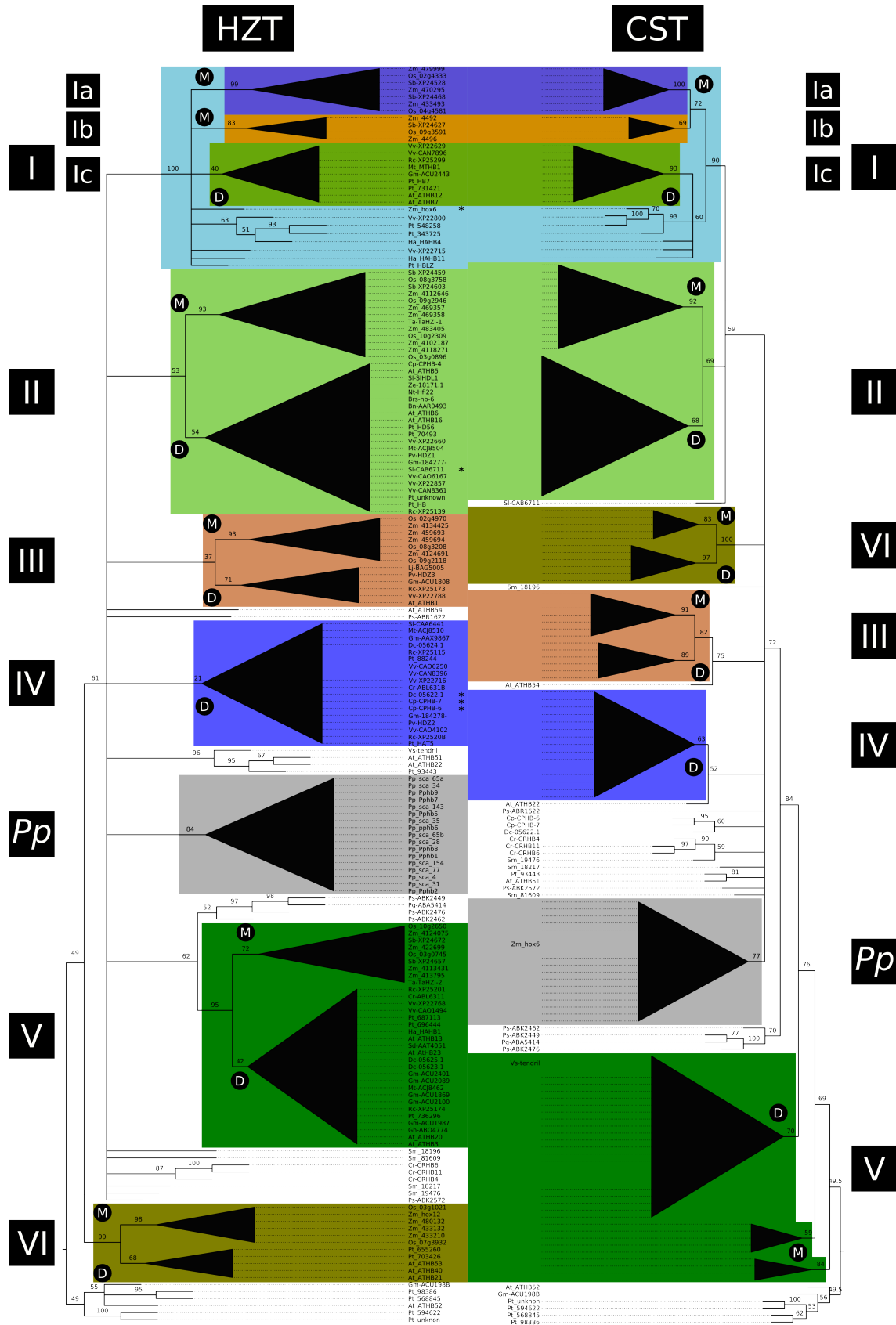
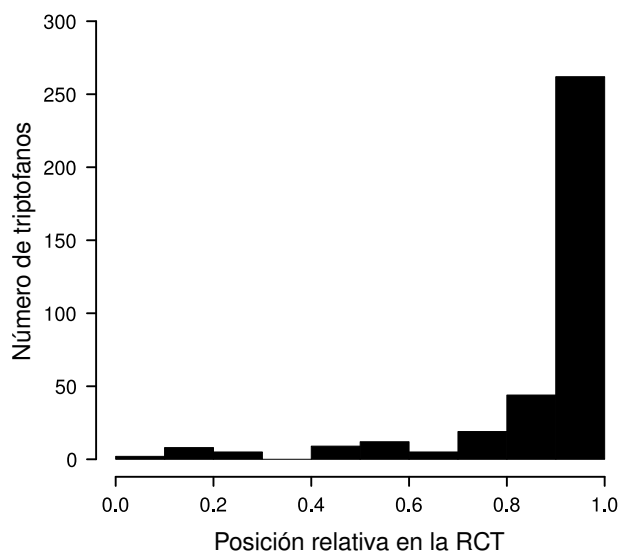


Figura III.4: Árboles HZT (dominios HD-Zip) y CST (proteínas completas). En el árbol CST se dejó solamente el nombre de aquellas proteínas ausentes en el grupo correspondiente del árbol HZT, y se agregó un asterisco a las proteínas del árbol HZT que no aparecieron en el grupo del árbol CST. Clados de proteínas de especies: monocotiledóneas **M** y dicotiledóneas **D**.



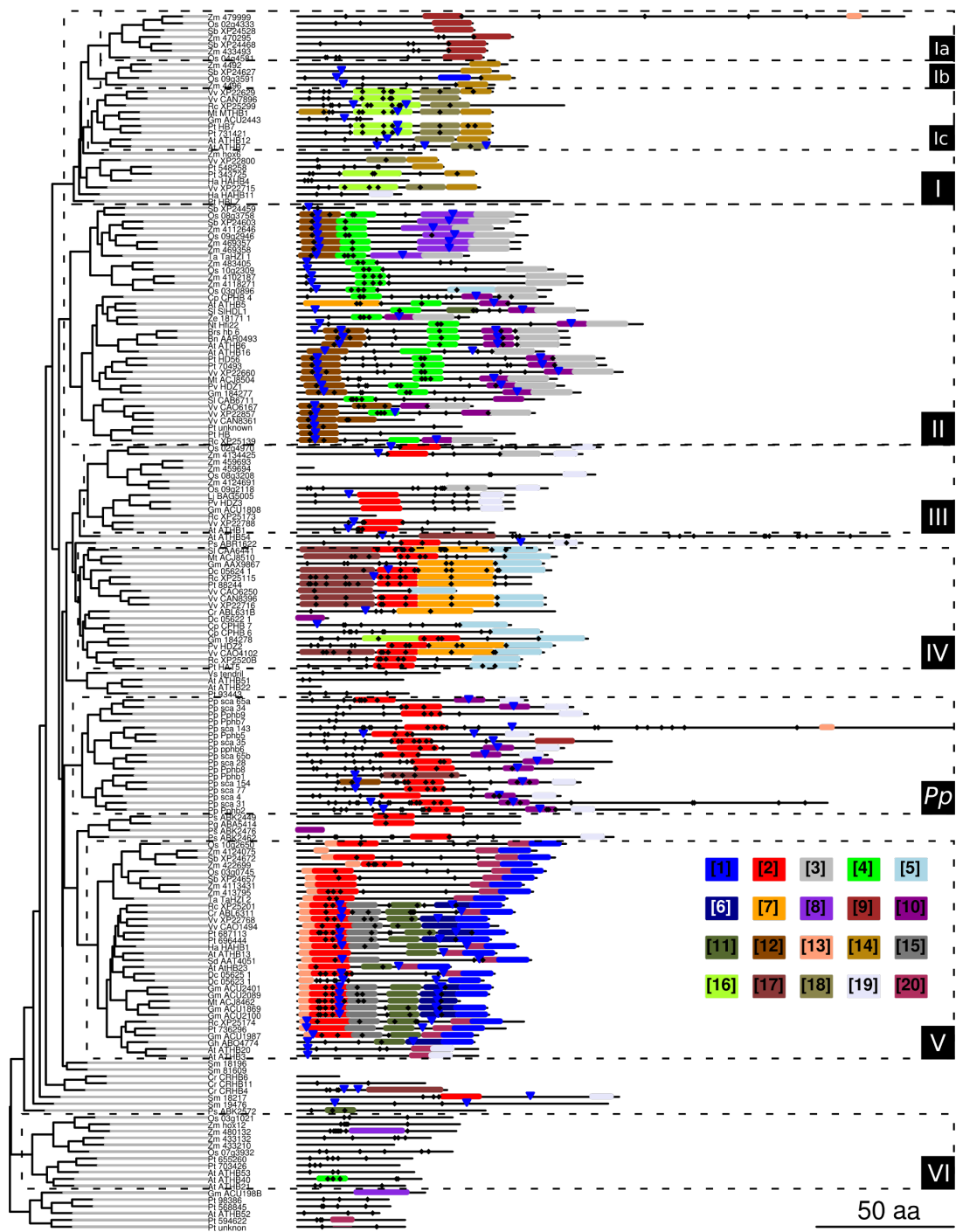
**Figura III.5:** Abundancia de Trp en las RCT. Histograma que muestra la abundancia de Trp dividiendo las RCTs en 10 porciones.

número máximo de motivos a buscar de 20, con una longitud mínima de 6 sitios; se obtuvieron motivos con un valor  $e$  entre  $4,3 \times 10^{-279}$  y  $3,9 \times 10^{-27}$ . La distribución de estos motivos en las diferentes RCTs se muestra en la Figura III.6 y su composición en la Figura III.7.

La mayor parte de los motivos encontrados fueron total o altamente específicos de cada grupo, a excepción del grupo VI que no presentó motivos distintivos. El motivo 2 constituyó un caso particular ya que está presente en la mayoría de los FTs de los grupos III, IV y V, y en muchas proteínas de *P. patens*. Este motivo se caracteriza por poseer un enriquecimiento en Ser, dos de ellas ocupando posiciones conservadas separadas por seis residuos, y varios aminoácidos aromáticos.

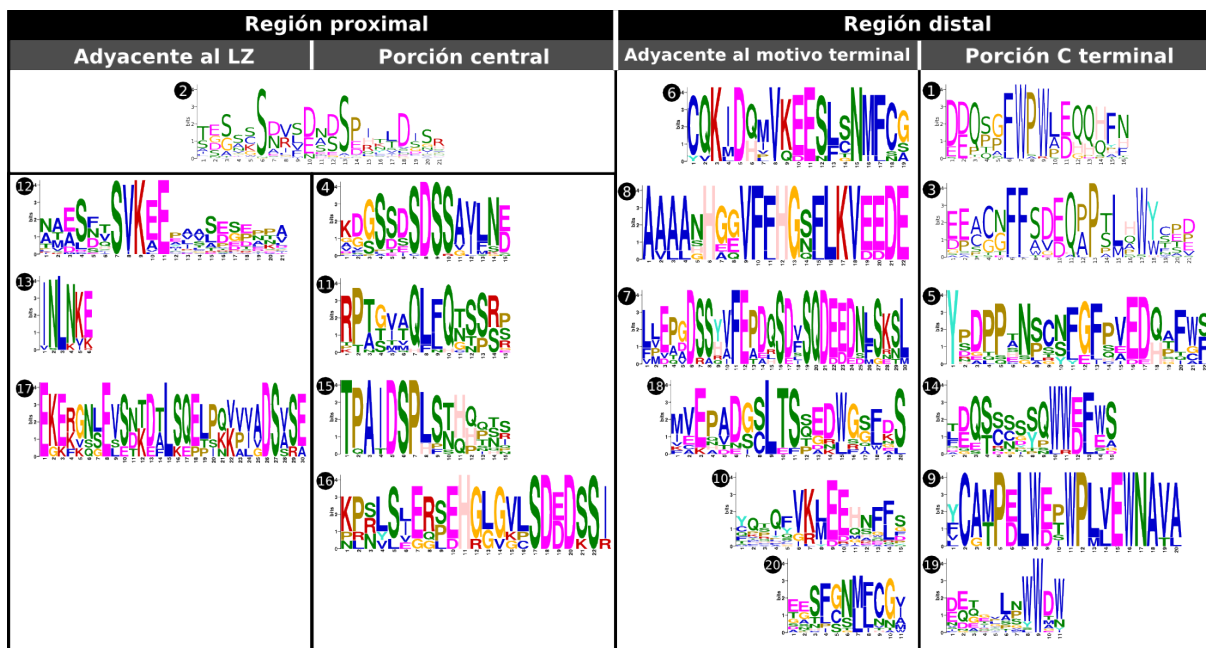
Un análisis de la distribución de estos motivos permitió definir, a grandes rasgos, dos porciones en las RCTs, una fue llamada porción proximal, adyacente al LZ; y la otra porción distal, comprendiendo la región final de la proteína (Figura III.6). En la primera se identificaron hasta tres motivos concatenados (grupos II, V y IV) y/o un motivo ubicado aproximadamente en el centro de la RCT, mientras que la segunda presentó un motivo que comprendió los últimos residuos de la proteína, que en los grupos Ic, II, IV, V y *Pp* estuvo acompañado por un motivo adyacente hacia el extremo N terminal.

La observación de la composición de los motivos reveló que todos aquellos situados en el extremo terminal de la proteína presentaban uno o más Trp en posiciones relativas muy conservadas en cada motivo (motivos 1, 3, 5, 9, 14 y 19; Figura III.7). También se observó con elevada frecuencia la Phe, otro aminoácido aromático, presente en la mayoría de los motivos de la porción distal (motivos 1,3,5,7, 10, 14 y 20). Adicionalmente, varias posiciones estuvieron ocupadas por residuos ácidos y algunas por Pro (motivos 1,3,5, 7 y 9). Estas características se



**Figura III.6:** Distribución de los motivos hallados mediante el programa MEME en las RCTs. El rombo negro (◆) simboliza residuos potencialmente fosforilables. El triángulo azul invertido (▼) simboliza sitios putativos de sumoilación.

### III.3. Resultados



**Figura III.7:** Secuencias de los motivos hallados en las RCTs. Los resultados se presentan como *sequence logos* divididos según la posición en la RCT.

asemejan a las de los motivos activadores de la transcripción *Aromatic, large Hydrophobic, Acidic context* (AHA) hallados en los FTs del tipo HSF (*Heat Stress Transcription Factor*; Döring y col., 2000).

#### III.3.3.1 Las RCTs serían blancos de modificaciones post-traduccionales

En los motivos de la porción proximal de las RCTs, los residuos presentes con mayor frecuencia fueron Ser y aminoácidos ácidos (Figura III.7). Las Ser son residuos potencialmente fosforilables y este tipo de modificación post-traducciona es un mecanismo de regulación de importancia en los FTs (Heazlewood y col., 2007), implicado en la regulación de AtHB6 según ensayos *in vitro* (Himmelbach y col., 2002). En consecuencia, se realizó una predicción *in silico* de los residuos (Ser, Thr o Tyr) con elevadas posibilidades de ser fosforilados. Esto fue realizado con el programa NetPhos (Blom y col., 1999) utilizando un umbral de 0,9. Las predicciones pusieron en evidencia que una gran parte de las Ser conservadas serían blancos preferenciales de esta modificación, particularmente aquellas presentes en los motivos 2, 4, 6, 7, 10, 12, 16, 17 y 18 (ver Figura III.6 y Figura III.7).

Otra posible modificación postraducciona estudiada fue la sumoilación (Miura y col., 2007). Esta modificación se produce por la unión covalente de la proteína pequeña modificadora relacionada a ubiquitina o *small ubiquitin-related modifier* (SUMO) a un residuo de Lys. Ocurre preferentemente en sitios  $\Psi$ KXE/D, siendo  $\Psi$  un residuo hidrofóbico grande, X cualquier aminoácido y E/D, Glu o Asp (Miura y col., 2007). Este péptido corto aparece con elevada frecuencia en los motivos 6, 8 y 10 de la porción distal adyacente al motivo terminal, y en el

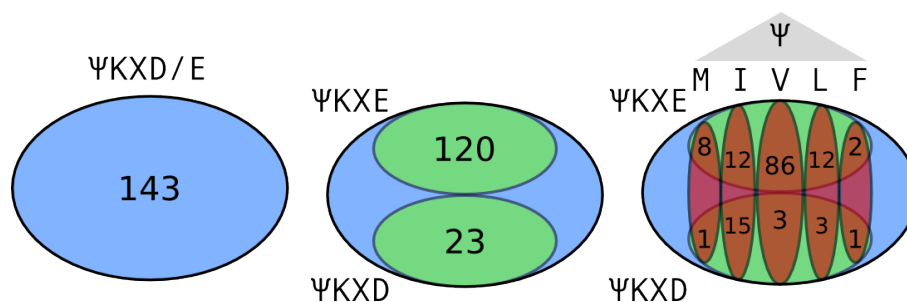


### III.3. Resultados

motivo 12 de la porción proximal. Se buscó entonces este tetrapéptido degenerado en las RCTs, considerando  $\Psi$  como Phe, Val, Ile, Met o Leu. Los resultados se resumen en la Tabla III.4 y la Figura III.8.

**Tabla III.4:** Recuento de aminoácidos en los sitios SUMO

Posición				Número de motivos	Proporción [ %]
1	2	3	4		
$\Psi$	E/D			143 (95 proteínas)	100
	E			120 (92 proteínas)	84
	D			23 (22 proteínas)	16
V	K	X		89	62
I				27	19
L			E/D	15	11
M				9	6
F				3	2



**Figura III.8:** Frecuencia de los diferentes residuos en los sitios de sumoilación presentes en las RCT.

El 62 % de los sitios posee una Val como residuo hidrofóbico en la primera posición, mientras que una considerable mayoría, 84 %, posee un Glu en su cuarta posición. Los motivos de sumoilación, marcados en la Figura III.6, aparecen principalmente en los grupos I (Ib y Ic), II, V y en el grupo *Pp*, en las RCTs de la mayoría de los FTs de los grupos II y V se los halla dos veces por proteína.

Por otro lado y con el fin de evaluar de alguna forma la relevancia de estos resultados, se buscó en las RCTs un motivo similar, pero en el que la cuarta posición fue cualquiera de los 20 aminoácidos excepto Glu y Asp ( $\Psi KX-[E/D]$ ). Se encontraron 82 motivos en 63 proteínas. Esto da una relación de número de motivos de sumoilación a número de motivos con la cuarta posición no canónica ( $\#(\Psi KXE/D)/\#(\Psi KX-[E/D])$ ) de 1.74; lo que indica un importante enriquecimiento en los motivos blanco de esta modificación.

#### III.3.4 Las RNTs también presentan algunos motivos conservados

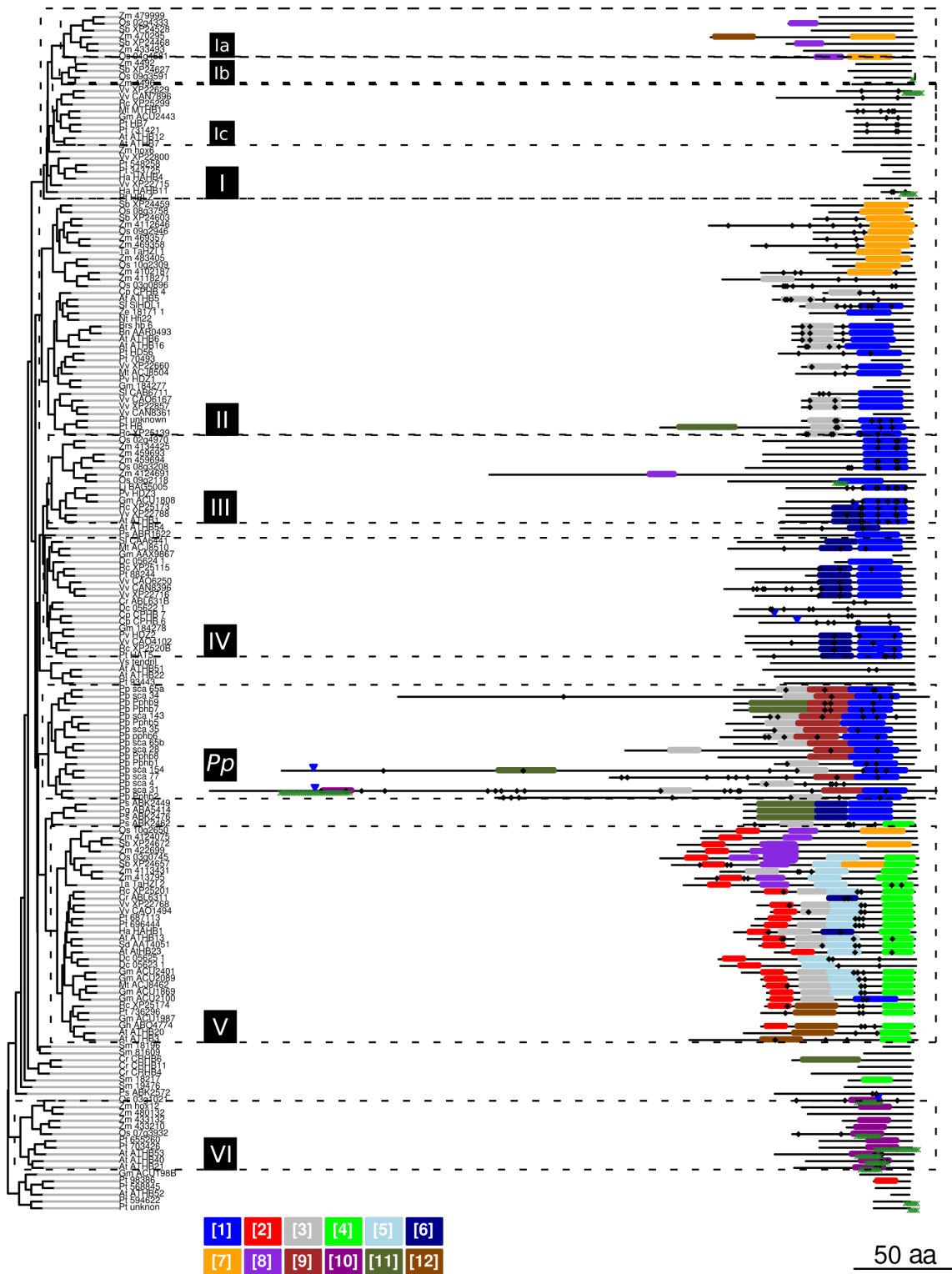
Las RNTs se analizaron usando una estrategia similar de descubrimiento de motivos a la usada para las RCTs. El programa MEME permitió hallar 12 motivos con valores  $e$  de  $2,6 \times 10^{-231}$  hasta  $3,7 \times 10^{-4}$ . La distribución de estos motivos se ilustra en la Figura III.9, mientras que su secuencia se muestra en la Figura III.10.

En cierto grado, la división en grupos se vio respaldada por la distribución de motivos en las RNTs, sin embargo, se presentaron varias excepciones. El motivo 1 estuvo ampliamente distribuido en los grupos II, III, IV y *Pp*, los subgrupos Ia, Ib y Ic carecieron de motivos distintivos y los FTs de plantas monocotiledóneas y de dicotiledóneas fueron divididos por motivos no compartidos en el grupo II. Por otro lado, las proteínas del grupo VI, que no tuvieron motivos característicos en sus RCTs, se distinguieron por presentar el motivo 10 en sus RNTs.

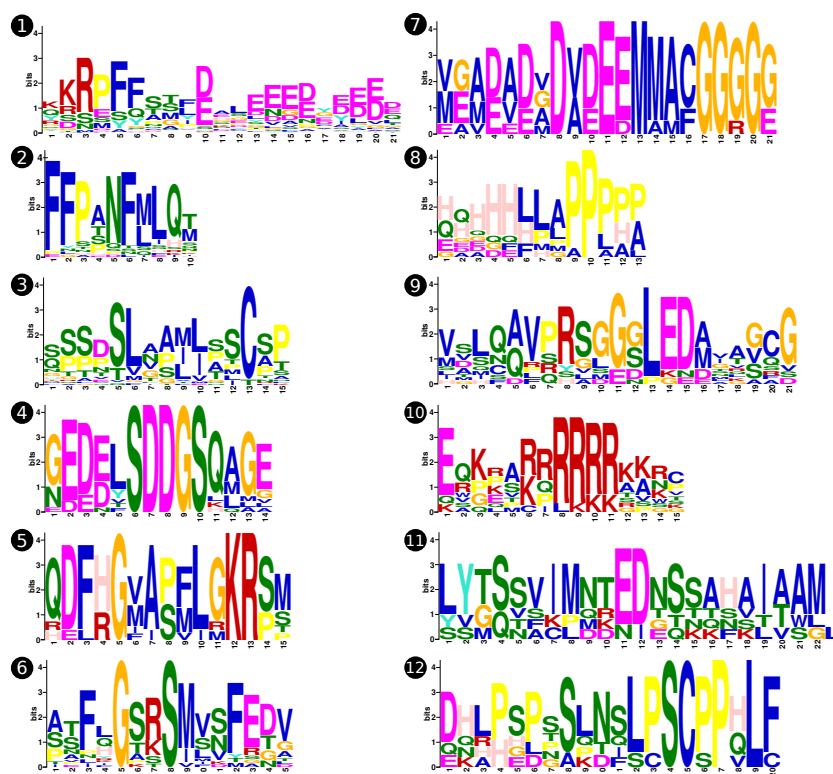
Análogamente a lo realizado con las RCTs, se exploró la posible funcionalidad de las RNTs. Usando el programa NetPhos se evaluaron los potenciales sitios de fosforilación con un valor umbral de 0.9; esto fue indicado sobre la Figura III.9. Los residuos de Ser en los motivos 1 (principalmente del grupo I), 3 y 6 (en las posiciones 10 con elevada frecuencia) fueron los mejores candidatos para esta modificación postraduccional, ya que además de la predicción informática, se observó una elevada conservación entre proteínas.

También se empleó el programa NLStradamus (Nguyen Ba y col., 2009) para predecir la existencia de señales de localización nuclear (SLN) que les permitan a los FTs ser importados al núcleo y así regular la transcripción. Para esto se utilizaron las secuencias proteicas completas, lo que permitió hallar los motivos en 16 de las 178 proteínas. De éstos, tres se encontraron en la RCT (proteínas ATHB54, Pp\_sca\_35 y Pp\_sca\_143), y de las 13 restantes, seis se hallaron en el grupo VI que estuvo compuesto por 11 miembros. Estas SLNs quedaron incluidas en los motivos 10 que fueron compartidos por 10 de estas proteínas. En consecuencia, a pesar de que el algoritmo predijo sólo seis SLNs, es posible que el motivo 10 cumpla esta función en las proteínas de este grupo.

Finalmente y a modo de comparación con las RCTs se evaluó la presencia de motivos de sumoilación. La búsqueda del péptido  $\Psi$ KXE/D arrojó ocho apariciones, siete de las cuales tienen una Glu en su última posición y cuatro una Val en la primera. Por otro lado, el motivo  $\Psi$ KX-[E/D] fue hallado 59 veces en 53 proteínas, con lo cual la relación  $\#(\Psi$ KXE/D)/ $\#(\Psi$ KX-[E/D]) es de 0,14, lo cual indica un enriquecimiento mucho menor al observado en las RCTs (1,74, ver p. 61). Estos resultados muestran que, a pesar de que pueda resultar interesante que los escasos motivos degenerados hallados posean una frecuencia de aminoácidos similar a la observada en las RCTs, la sumoilación no sería una modificación relevante en las RNTs según estas predicciones.



**Figura III.9:** Distribución de los motivos hallados mediante el programa MEME en las RNTs. El rombo negro (◆) simboliza residuos potencialmente fosforilables. El triángulo azul invertido (▼) simboliza sitios putativos de sumoilación. La cruz verde (×) indica residuos que formarían parte de una señal de localización nuclear (SLN)



**Figura III.10:** Secuencias de los motivos hallados en las RNTs.

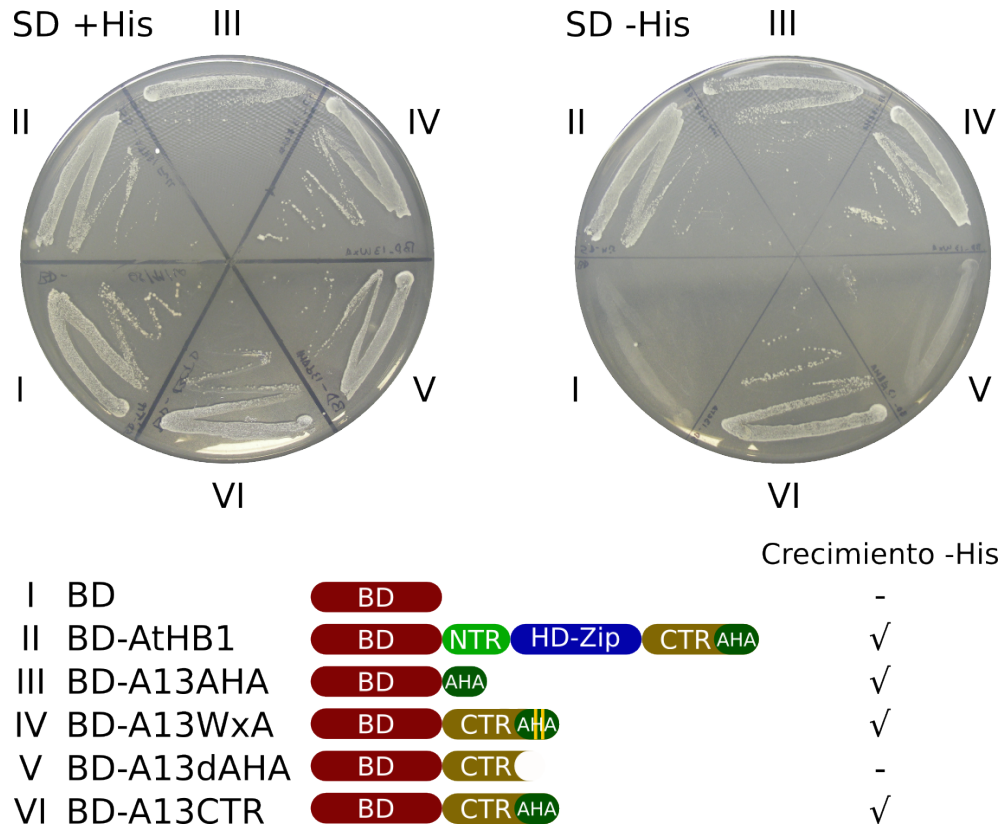
#### III.3.5 Las RCTs poseen la capacidad de transactivar en levaduras

Basándonos en la predicción de que en el extremo terminal de las RCTs existe un motivo tipo AHA con capacidad de transactivación, decidimos emplear un sistema de simple híbrido en levaduras para poner a prueba esta hipótesis. La inspección de la RCT de *AtHB13*, el alineamiento de las secuencias del grupo V al que pertenece (ver Figura III.4) y de los motivos predichos por el programa MEME (ver Figura III.6), permitió definir que el motivo AHA más probable estaba comprendido en una región conservada de 28 aa en la porción final del extremo carboxilo de la proteína.

Se prepararon entonces cuatro construcciones con fusiones al dominio de unión a ADN del factor de transcripción de levaduras GAL4 (BD), empleando los fragmentos de ADN de *AtHB13* que codifican la RCT completa (BD-A13RCT), la RCT en la que los Trp del motivo AHA fueron mutados a Ala (BD-A13WxA), la RCT sin el motivo AHA (BD-A13dAHA) o el motivo AHA solo (BD-A13AHA, Figura III.11). Como control negativo de la reacción se utilizó el dominio BD sin fusionar, y como control positivo se utilizó una construcción en la que el dominio BD estaba fusionado a la proteína completa *AtHB1* (BD-*AtHB1*), obtenida y analizada previamente en el laboratorio. Los resultados se presentan en la Figura III.11.

Puede observarse que las levaduras transformadas con las seis construcciones crecieron normalmente en un medio mínimo SD con His. En el medio SD sin His, en el que el crecimiento es indicador de transactivación y por ende, de que el gen revierte la auxotrofía para la His,

### III.3. Resultados



**Figura III.11:** Funcionalidad del motivo AHA en ensayos de simple híbrido en levaduras. En los paneles superiores se ve el crecimiento en el medio con His (SD+His) y sin His (SD-His). Abajo se ilustran las diferentes proteínas expresadas.

crecieron sólo las levaduras transformadas con las construcciones en las que el AHA está presente, aún en la versión con los Trp mutados por Ala. Esto indica que el AHA posee capacidad de transactivación, que en principio los Trp no serían esenciales para esta actividad, y que no hay en el resto de la RCT una región capaz de transactivar por sí sola. Cabe recordar que estos resultados, al provenir de un ensayo en levaduras, más allá de ser *in vivo*, no reflejan necesariamente lo que ocurre en plantas. Es posible, por ejemplo, que la proteína con la que el dominio AHA interactúa *in planta* requiera de la presencia de los Trp.

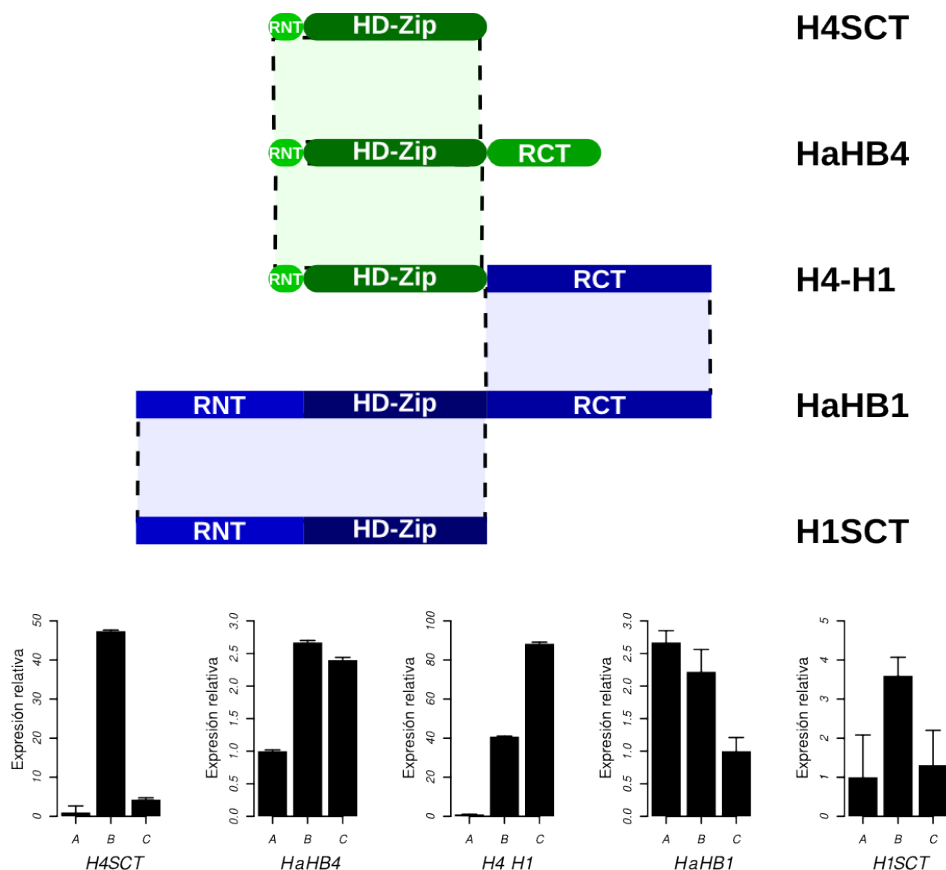
#### III.3.6 La RCT participa en la determinación de la función de la proteína HD-Zip<sup>1</sup>

Para evaluar la importancia y la relación estructura/función de la RCT en una proteína HD-Zip I, se escogieron, para hacer un análisis funcional *in planta*, las proteínas de girasol HaHB1 y HaHB4 ampliamente estudiadas y caracterizadas en el laboratorio. La estrategia adoptada consistió en la comparación fenotípica de plantas transgénicas expresando sendos FTs

<sup>1</sup>Estos ensayos fueron realizados junto a la Lic. Jéssica Raineri.

### III.3. Resultados

controlados por el promotor *35SCaMV* (obtenidas en trabajos previos; Cabello y col., 2012; Dezar y col., 2005b), con plantas en las que el mismo promotor dirigía la expresión de una proteína quimérica compuesta por la RNT y el dominio HD-Zip de HaHB4 fusionados a la RCT de HaHB1 (H4-H1). Sumado a esto, se prepararon plantas transgénicas con dos construcciones, a modo de controles, en las cuales se habían removido las RCTs de estos FTs, *35SCaMV::H1SCT* y *35SCaMV::H4SCT*. Un esquema de las construcciones y de los niveles de transcritos en estas plantas se presenta en la Figura III.12.



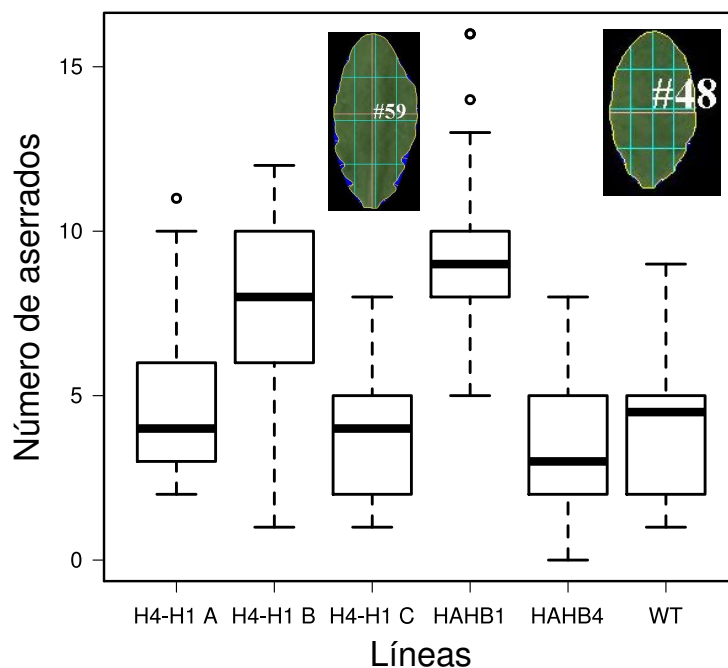
**Figura III.12:** Construcciones expresadas por las plantas transgénicas. Se presenta un esquema de las diferentes proteínas expresadas en *A. thaliana* bajo el control de promotor *35SCaMV* y los niveles de expresión de cada una de las tres líneas utilizadas para los ensayos.

Los fenotipos elegidos para realizar la comparación consistieron en fenotipos distintivos cuantificables. Por un lado, el aserrado de las hojas de la roseta, ya que las plantas transgénicas con la construcción *35SCaMV::HaHB1* poseen un número de aserrados superior al de las plantas salvajes (WT; Cabello y col., 2012). Y por otro, siendo que las plantas transformadas con la construcción *35SCaMV::HaHB4* presentan una insensibilidad al etileno, se estudió esta característica a través de un ensayo de triple respuesta (Manavella y col., 2006).

#### III.3.6.1 El aserrado se incrementa en las plantas con el FT quimérico que posee la RCT de HaHB1

Para analizar el fenotipo de aserrado se cortaron hojas de la roseta de plantas transgénicas de 21 días de los genotipos *35SCaMV::HaHB1* (línea B), *35SCaMV::HaHB4* (línea B), *35SCaMV::H4-H1* (líneas A, B y C) y plantas WT. Estas hojas fueron escaneadas y las imágenes procesadas con el programa LAMINA (Bylesjö y col., 2008). Este programa reconoce en las imágenes el contorno de las hojas y a partir de esto puede determinar un gran número de parámetros foliares, entre ellos el aserrado.

Los resultados, que se muestran en la Figura III.13, fueron sometidos a un análisis de varianza por rangos de Kruskal-Wallis y las construcciones fueron clasificadas en diferentes grupos según una comparación de a pares con un valor p de 0,05 (Tabla III.5). Se observa que las hojas de *35SCaMV::HaHB1* poseen un aserrado sensiblemente mayor al de las plantas WT, mientras que las plantas *35SCaMV::HaHB4* no presentan diferencias con las salvajes. De las tres líneas de plantas *35SCaMV::H4-H1*, solamente la línea B mostró un aserrado incrementado, como el de las plantas *35SCaMV::HaHB1*.



**Figura III.13:** Aserrado de las hojas de la roseta de plantas transgénicas transformadas con diferentes construcciones. Las plantas de la línea B que expresan la quimera H4-H1 presentaron un aserrado similar al de las plantas que expresan HaHB1. Se agregaron dos hojas analizadas informáticamente para ilustrar el proceso.

**Tabla III.5:** Análisis de Kruskal-Wallis de los resultados de aserrado.

Línea	Rango	Grupos
HAHB4 B	50,36	A
H4-H1 C	58,53	A B
WT	66,88	A B
H4-H1 A	78,53	B
H4-H1 B	120,03	C
HAHB1 B	143,17	C

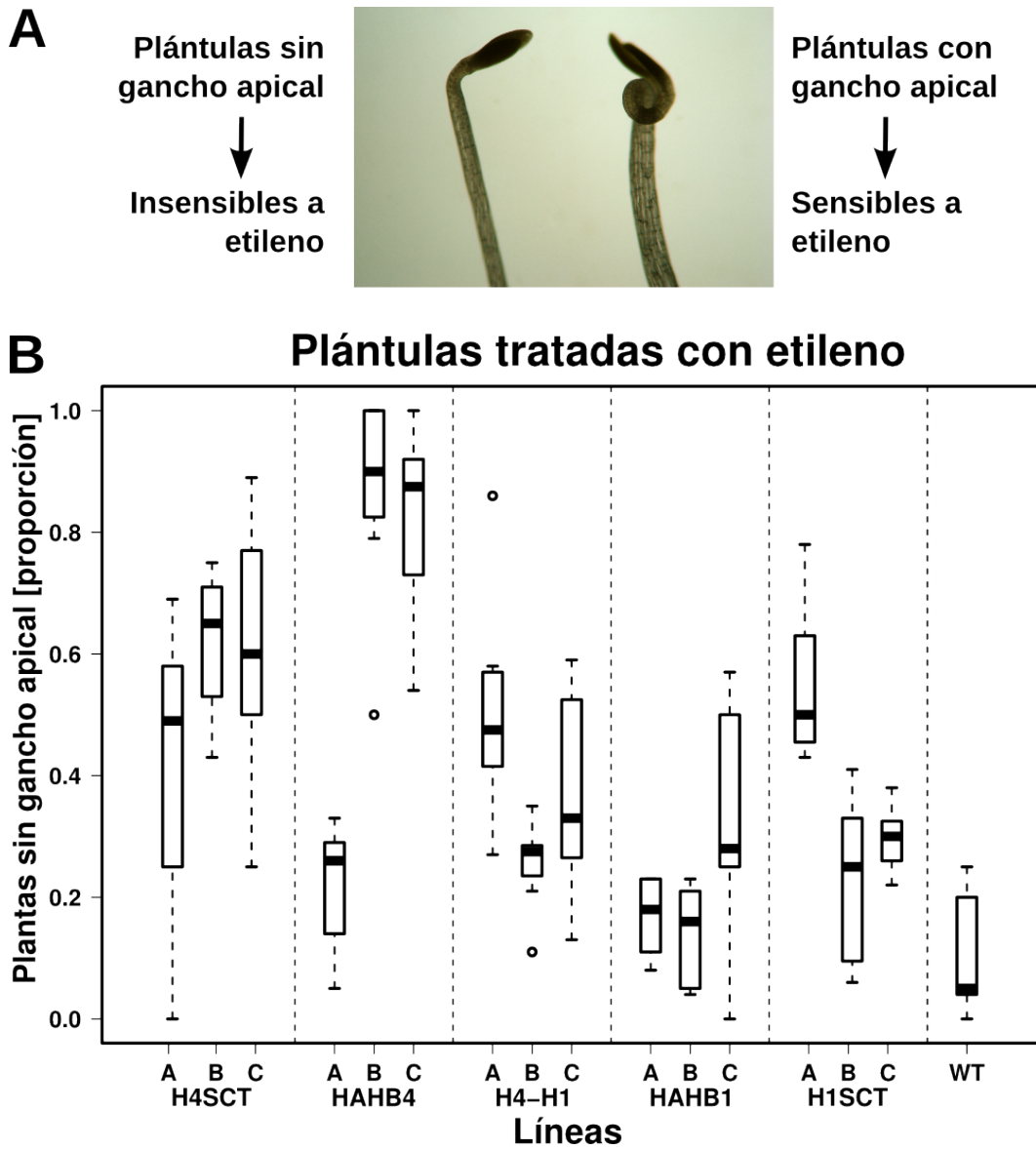
#### III.3.6.2 Las plantas que expresan el FT quimérico H4-H1 presentan una sensibilidad intermedia al etileno

El estudio de la sensibilidad al etileno (ET) fue realizado evaluando la triple respuesta. El ensayo consistió en hacer germinar y cultivar las plantas en oscuridad y en presencia de ácido 1-aminociclopropano-1-carboxílico (ACC), un precursor de la hormona ET. En estas condiciones las plántulas desarrollan el fenotipo denominado de triple respuesta porque presenta tres características distintivas: el torcimiento del hipocótilo en la región cercana a los cotiledones, llamado "gancho apical" (ilustrado en la Figura III.14); el engrosamiento del hipocótilo y su acortamiento. De estas tres respuestas, la más sencilla de cuantificar es la curvatura del gancho apical, que fue considerada una variable binaria: "presenta" vs. "no presenta". Los resultados de este ensayo se muestran en la Figura III.14.

En primer lugar, se observó que las plantas transformadas con la construcción *35SCaMV::HaHB4* mostraban un alto grado de insensibilidad, como se esperaba. Sólo la línea A, de menor expresión (ver Figura III.12), no presentó el fenotipo anticipado. Las plantas WT y las plantas con la construcción *35SCaMV::HaHB1* tuvieron una elevada sensibilidad al tratamiento, lo cual se esperaba para que fuera un ensayo comparativo diferencial.

Las líneas de las plantas de la construcción *35SCaMV::H4SCT* mostraron una sensibilidad intermedia, y un poco mayor a ésta fue la sensibilidad de las plantas transformadas con las construcciones *35SCaMV::H4-H1* y *35SCaMV::H1SCT*. El hecho de que la remoción de la RCT de HaHB4 (plantas *35SCaMV::H4SCT*) haya tenido un efecto menor sobre el fenotipo que la sustitución de éste por la RCT de HaHB1 (plantas *35SCaMV::H4-H1*) podría indicar que el efecto de HaHB4 en esta respuesta no estaría asociado estrechamente a la funcionalidad de su RCT.





**Figura III.14:** Ensayo de triple respuesta. A. Se muestra una imagen de dos plántulas sometidas al tratamiento con ACC, en las que se observa el fenotipo según la sensibilidad. B. Resultados del ensayo en un diagrama de cajas obtenidos con las proporciones de ocho grupos de 20 plantas por línea.

## III.4 Discusión de resultados

La motivación para caracterizar las regiones no pertenecientes al dominio HD-Zip en las proteínas de la subfamilia I surgió del intento de profundizar en la comprensión de la diversidad funcional de sus miembros. Existen numerosas variables determinantes de la función de un FT que lo pueden diferenciar de otros miembros de su familia.

A nivel molecular, la capacidad de reconocer una u otra secuencia de ADN representa una diferencia potencial entre FTs ya que es determinante de las regiones promotoras a las que podrá unirse y, por consiguiente, determinará qué genes será capaz de regular. La evidencia experimental disponible indica que las proteínas HD-Zip I poseen la misma especificidad de unión a ADN, reconociendo con máxima afinidad el elemento de secuencia CAAT(A/T)ATTG (Johannesson y col., 2001; Meijer y col., 2000; Palena y col., 1999; Sessa y col., 1993). En consecuencia, la capacidad de reconocer diferentes promotores no sería un factor determinante generador de diferencias funcionales. Por supuesto, la accesibilidad a estos promotores depende del estado de compactación de la cromatina en el tejido dado (Venters y Pugh, 2009), con lo cual los promotores reconocidos pueden depender del tejido en el que se esté expresando el FT, y de las condiciones a las que esté sometido.

Además, es importante tener en cuenta que la capacidad de heterodimerización podría ser un factor relevante ya que entonces una proteína HD-Zip podría afectar la actividad de otra o podrían formarse dímeros con una función diferente. Los trabajos previos muestran que la heterodimerización entre miembros de la misma subfamilia sería posible (Deng y col., 2002; Johannesson y col., 2001; Meijer y col., 2000).

Como se explicó en la Sección III.1.3.1, los patrones de expresión de los distintos genes HD-Zip I son significativamente diferentes entre miembros. Esto determina en qué tejidos son, en principio, capaces de actuar estas proteínas, y por lo tanto es un factor que constituye una fuente de diversidad funcional muy importante.

Podría considerarse que el uso del promotor *35SCaMV* como herramienta para la expresión ectópica y constitutiva de diferentes genes HD-Zip I tienen un efecto igualador en los patrones de expresión. Habiéndose minimizado entonces el efecto de la expresión diferencial natural y considerando que no habría una preferencia significativa por diferentes promotores, la variedad de fenotipos observados (ver Sección III.1.3.2, p. 47) sugiere fuertemente la existencia de factores adicionales no caracterizados, responsables de las diferentes funciones de cada gen/proteína.

### III.4.1 Reconstrucción filogenética y definición de grupos

Se buscó un conjunto de secuencias de proteínas HD-Zip I de diferentes especies, desde musgos hasta plantas monocotiledóneas, que incluyó el total de los FTs de esta subfamilia de seis especies con genoma secuenciado. Con estas secuencias se realizó primariamente un

### III.4. Discusión de resultados

---

alineamiento de los dominios HD-Zip que fue la base de la reconstrucción filogenética de referencia (árbol HZT).

El análisis del alineamiento de estos dominios permitió observar una elevada conservación de los residuos responsables de determinar la especificidad de la secuencia de ADN reconocida (Figura III.2, p. 54; Palena y col., 2001; Sessa y col., 1997; Tron y col., 2001, 2005). Esta observación confirma los datos experimentales informados previamente (Johannesson y col., 2001; Meijer y col., 2000; Palena y col., 1999; Sessa y col., 1993) indicando que no habría diferencias sustanciales en las secuencias que estos FTs son capaces de unir. En consecuencia, los genes regulados potencialmente por estos FTs no estarían determinados por estas diferencias sino por otros factores que podrían afectar dicha unión (por ejemplo, la compactación de la cromatina, la expresión del gen en el tejido o la presencia de proteínas reguladoras del FT).

En la reconstrucción filogenética se reconocieron y definieron grupos de proteínas relacionadas. El agrupamiento se basó en el sustento estadístico que apoyaba su origen monofilético en este árbol y/o en un árbol realizado con las proteínas completas (árbol CST). Además, la observación de una elevada conservación en las RCTs de cada grupo no sólo brindó mayor respaldo a estos agrupamientos, sino que motivó la subdivisión del grupo I en Ia, Ib y Ic. El objetivo final de este agrupamiento fue investigar la posible similitud funcional de proteínas filogenéticamente relacionadas (ortólogos putativos) y la potencial participación de las secuencias fuera del dominio HD-Zip.

Es esperable que genes ortólogos posean una similitud de secuencia por fuera de los dominios conservados comparativamente mayor a la de genes más distantes evolutivamente, sin que ésta implique necesariamente la funcionalidad de estas regiones. Podría argumentarse que el tiempo de divergencia luego de la duplicación ha sido relativamente menor. Sin embargo, es importante destacar las notorias diferencias en el grado de similitud de las secuencias a lo largo de las RCTs (ver Figura III.3, p. 56). Esta heterogeneidad es un indicador de que ciertas regiones podrían haberse conservado más por ser funcionales.

El reconocimiento de motivos en las RCTs (y en las RNTs en menor medida) dio respaldo a esta hipótesis. En las RCTs se hallaron motivos con una distribución particular, identificándose una región proximal, adyacente al LZ, y una distal en la región final de las proteínas (ver Figura III.6, p. 59 y Figura III.9, p. 63). Las secuencias que separaron los motivos (que podrían llamarse secuencias “espaciadoras”) fueron de longitud variable y una inspección preliminar sugiere que poseerían una composición con escasa diversidad de aminoácidos. Además, la distribución por grupos de los motivos en las RCTs fue altamente específica, lo que da mayor sustento a los agrupamientos.

Finalmente se definieron seis grupos (I al VI) compuestos por proteínas de plantas monocotiledóneas y dicotiledóneas, a excepción del grupo IV conformado exclusivamente por proteínas de especies dicotiledóneas. Todos los grupos incluyeron al menos una proteína de las plantas con genoma secuenciado, a excepción del grupo III que no contuvo proteínas de *P. trichocarpa*, y el

### III.4. Discusión de resultados

---

grupo IV que no incluyó proteínas de *A. thaliana* y tampoco de especies monocotiledóneas. La ausencia de proteínas de coníferas, helechos y musgos en cualquiera de los siete grupos es un indicador de su surgimiento luego de la división entre gimnospermas y angiospermas. El grupo Pp quedó conformado exclusivamente por las 17 proteínas del musgo *Physcomitrella patens* lo que podría señalar que existe una escasa diversificación a nivel molecular de la subfamilia en esta planta no vascular, o que estas proteínas son muy distantes del resto de las proteínas evaluadas. Una reconstrucción filogenética que involucrara más proteínas de los diferentes taxones principales ayudaría a resolver este interrogante. También es importante resaltar que el grupo Pp no quedó ubicado en una posición basal como podría esperarse por la relación filogenética que esta especie tienen con las demás especies incluidas en el estudio.

La conservación en las RCTs constituye un fuerte indicador de homología entre las proteínas de los grupos. Algunos de ellos con bajo soporte estadístico en el árbol HZT, como los grupos III y el IV (ver Tabla III.3, p. 55), tuvieron valores de *bootstrap* significativos en el árbol CST. En consecuencia, la búsqueda de grupos de genes filogenéticamente relacionados mediante un árbol construido con las secuencias completas resultó satisfactoriamente complementaria a la reconstrucción con el dominio HD-Zip común a todas las proteínas.

#### III.4.1.1 Comparación con la reconstrucción filogenética obtenida en *A. thaliana*

Una comparación de este agrupamiento con el realizado exclusivamente con la subfamilia I de genes HD-Zip de *A. thaliana* (Henriksson y col., 2005) mostró similitudes y diferencias (resumidas en la Tabla III.6). Hubo correspondencia directa entre los grupos I y  $\gamma$  (*AtHB7* y *AtHB12*), V y  $\alpha$  (*AtHB3*, *AtHB13*, *AtHB20* y *AtHB23*), y VI y  $\delta$  (*AtHB21*, *AtHB40* y *AtHB53*). Los genes del grupo  $\beta$  en el árbol HZT quedaron en el grupo II (*AtHB5*, *AtHB6* y *AtHB16*) y III (*AtHB1*).

Los genes del grupo  $\varepsilon$  (*AtHB22* y *AtHB51*), aunque formaron un grupo monofilético en el árbol HZT, no fueron considerados en ningún agrupamiento por conformar un grupo pequeño de cuatro integrantes sin representantes de las otras especies con genoma secuenciado, excepto por una proteína de *P. trichocarpa*. Los genes del grupo  $\phi$  (*AtHB52* y *AtHB54*) no fueron agrupados en el árbol HZT ni formaron un grupo monofilético. Adicionalmente, los grupos  $\varepsilon$  y  $\phi$  conformados por proteínas codificadas por genes supuestamente parálogos, fueron los únicos grupos que según Henriksson y col. (2005) no contaban con la evidencia de hallarse en regiones genómicas duplicadas durante la evolución.

Es interesante destacar que el análisis de la distribución de intrones y exones realizado por Henriksson y col. (2005) mostraba un patrón diferente entre *AtHB1* y el resto de los genes del grupo  $\beta$  que aparecen separados en el árbol HZT. Es más, la misma ruptura del grupo  $\beta$  fue propuesta por una reconstrucción que involucró 33 proteínas HD-Zip I de diferentes especies (Agalou y col., 2008).

### III.4. Discusión de resultados

**Tabla III.6:** Comparación del agrupamiento de los FTs en el árbol HZT vs. la reconstrucción de Henriksson y col. (2005)

HZT	Henriksson y col. (2005)	Observaciones
I	$\gamma$	iguales
II	$\beta$	menos <i>AtHB1</i>
III	$\beta$	sólo <i>AtHB1</i>
IV	–	sin miembros en <i>A. thaliana</i>
V	$\alpha$	iguales
VI	$\delta$	iguales
–	$\epsilon$	genes no agrupados en HZT
–	$\phi$	genes no agrupados en HZT

#### III.4.2 El motivo AHA y la activación de la transcripción

La porción final de las RCTs mostraron un enriquecimiento significativo en Trp (ver Figura III.5, p. 58). La búsqueda de motivos en estas regiones permitió reconocer que las porciones finales poseen un motivo con características similares a los motivos AHA descritos para FTs HSF (ver Figura III.7, p. 60; Döring y col., 2000), por poseer:

- aminoácidos aromáticos: principalmente Trp, Pro, Phe y Tyr.
- aminoácidos hidrofóbicos grandes: Leu y Val (la Phe y el Trp pueden considerarse en este grupo también).
- aminoácidos ácidos: Asp y Glu.

En las proteínas HSF, los motivos AHA pueden estar en repeticiones y tienen una función de activación, lo cual fue demostrado tanto en ensayos de simple híbrido en levaduras como de transactivación en protoplastos de tabaco (Döring y col., 2000; Kotak y col., 2004). En estas proteínas se ha demostrado la importancia del Trp en el motivo, ya que en diferentes casos su substitución por un aminoácido como la Ala anuló su actividad.

Varios antecedentes sustentan la noción de que las proteínas HD-Zip I tienen capacidad de transactivación. La expresión transitoria en ensayos de bombardeo de hojas de tabaco demostró que *AtHB1* es capaz de activar la expresión del gen de la luciferasa fusionado a un promotor que contenía seis copias del elemento unido *in vitro*, CAATTATTG (Aoyama y col., 1995). También en ensayos transitorios, pero en células de arroz en suspensión, y en ensayos de simple híbrido en levaduras se demostró que dos proteínas de arroz poseían la capacidad de activación (Meijer y col., 2000).

El estudio de *AtHB12* mediante simple híbrido en levaduras había mostrado también su capacidad de transactivar, pero además había restringido esta actividad a los últimos 55 aa de la proteína, la porción final de su RCT (Lee y col., 2001). La inspección de su secuencia y la

### III.4. Discusión de resultados

---

presencia de motivos en esta porción de la proteína permite concluir que contiene un motivo AHA (ver Figura III.6, p. 59).

Posteriormente se demostró que las proteínas de *A. thaliana*, AtHB5, AtHB6, AtHB7 y AtHB12 también son capaces de transactivar en ensayos transitorios en hojas de *A. thaliana* (Henriksson y col., 2005), un ensayo similar al realizado previamente con AtHB1 en hojas de tabaco (Aoyama y col., 1995).

En el presente trabajo de tesis se demostró mediante estudios de simple híbrido en levaduras la funcionalidad del motivo AHA predicho para el FT AtHB13 (ver Sección III.3.5, p. 64 y Figura III.11, p. 65). La RCT completa de AtHB13 y una región de 28 aa conteniendo el AHA fueron capaces de transactivar. La remoción de esta porción final de la RCT fue suficiente para que se perdiera esta actividad. Por último, la sustitución de los dos Trp que contiene el motivo AHA por Ala no tuvo un efecto sensible sobre su capacidad de transactivación. Este último resultado sugiere que la importancia de estos aminoácidos en el motivo de AtHB13, podría no ser igual a la que tienen estos aminoácidos en los motivos de los FTs HSF. No obstante, es importante resaltar que esta observación podría no ser del todo extrapolable a lo que ocurre en *A. thaliana*.

A esta evidencia directa de transactivación se le puede sumar la evidencia funcional de la importancia de esta porción de la RCT de proteínas HD-Zip I. Sakuma y col. (2010) encontraron que el gen *VRS1*, responsable del fenotipo de cebada de “dos hileras”, codifica una proteína cuya única diferencia importante con la codificada por el gen parálogo HvHOX2, y otras muy similares de diferentes especies, es la ausencia de un motivo conservado de la porción C terminal. La observación de la secuencia presentada en este trabajo permite concluir que posee las características de un motivo AHA.

En un estudio realizado en arveja se halló que el gen HD-Zip I *Tendrill-less* (*TL*) era responsable de la inhibición del desarrollo laminar necesaria para la formación de los zarcillos (Hofer y col., 2009). Los alelos mutantes de este gen generaron la “transformación” de zarcillos en hojas, en algunos de ellos porque la mutación evita la expresión del gen. Resulta interesante que uno de los alelos que produce este fenotipo codifica una proteína a la que sólo le faltan los últimos 12 aa del extremo carboxilo terminal (Hofer y col., 2009). La inspección de esta secuencia reveló que en esos 12 aa habría un motivo de tipo AHA.

Un trabajo reciente demostró que *MtHB1* de *M. truncatula* es un regulador negativo directo del gen *LBD1*, lo cual indica que una proteína HD-Zip I puede actuar como represor (Ariel y col., 2010). *MtHB1* es capaz de unir el promotor de *LBD1*, como lo demostraron un ensayo de ChIP y estudios de expresión de *GUS* en plantas mutantes de *MtHB1* y WT en las que este promotor poseía una mutación del motivo reconocido. Además, la comparación de los promotores de *MtHB1* y *LBD1*, y el estudio de los niveles de expresión de *LBD1* en plantas mutantes de *MtHB1* y WT demostraron que esta regulación es negativa.

### III.4. Discusión de resultados

---

En principio, la capacidad de MtHB1 de reprimir la transcripción de *LBD1* puede resultar contradictoria con la evidencia que propone que las proteínas de esta subfamilia son activadores transcripcionales. Sin embargo, existen antecedentes de FTs con HD de *D. melanogaster* que tienen ambas actividades (Choi y col., 1999; Yu y col., 2005). Posiblemente, lo que este trabajo pone en evidencia es que los mecanismos de acción de estas proteínas en la planta pueden ser más complejos de lo anticipado.

En resumen, en función de la evidencia existente se logró identificar *in silico* la presencia de motivos activadores AHA en la porción final de las RCTs de la mayoría de las 178 proteínas estudiadas. La posterior confirmación de esta predicción a través de la caracterización de un caso, la RCT de AtHB13, permite proponer que los FTs de esta subfamilia serían, en términos generales, potenciales activadores de la transcripción, y que esta actividad puede atribuirse al motivo AHA presente en sus RCTs.

### III.4.3 Modificaciones post-traduccionales

#### III.4.3.1 La fosforilación en la regulación de la unión al ADN

La elevada conservación en las RCTs y la presencia de otros motivos además de los AHA motivaron la búsqueda de posibles funciones adicionales de las secuencias en estas regiones, las que podrían justificar su conservación.

La fosforilación es un tipo de modificación post-traducciona que ha sido descrita para AtHB6 (Himmelbach y col., 2002). Esta proteína fue recuperada en un rastreo por doble híbrido en levaduras (DHL) de proteínas que interactuaban con la fosfatasa ABI1, un regulador esencial de la vía de señales de ABA. La interacción entre AtHB6 y ABI1 disminuyó sensiblemente al remover la RNT y desapareció al eliminar las regiones RNT y RCT. La capacidad de AtHB6 de unirse a la secuencia CAAT(T/A)ATTG *in vitro* fue inhibida por la fosforilación con proteína quinasa A, indicando que es necesario que la proteína esté desfosforilada para poder unir el ADN.

En consecuencia, se realizó una predicción de los aminoácidos (Ser, Thr y Tyr) con un alta posibilidad de ser fosforilables mediante el programa NetPhos (Blom y col., 1999), que se basa en las características de la secuencia alrededor del residuo. Se halló un gran número de sitios putativos de fosforilación en las RCTs y algunos en las RNTs (ver Figura III.6, p. 59 y Figura III.9, p. 63). Es importante destacar que muchos de los aminoácidos con elevada conservación en los motivos pertenecientes mayormente a la porción proximal de la RCT fueron considerados fosforilables por el programa, particularmente en los motivos 2, 4, 6, 7, 10, 12, 16, 17 and 18 (ver Figura III.7, p. 60).

## III.4. Discusión de resultados

### III.4.3.2 La sumoilación como un nuevo mecanismo de regulación

La sumoilación es una modificación post-traducciona que tiene una función muy importante en la regulación de numerosos FTs (Miura y Hasegawa, 2010; Miura y col., 2007). Consiste en la unión covalente de SUMO a un residuo de Lys. El SUMO es una pequeña proteína similar a la ubiquitina, cuya unión tiene principalmente tres potenciales efectos:

- inhibir la interacción preexistente con una proteína dada.
- favorecer o posibilitar una nueva interacción con otra proteína. Lo que ha demostrado conducir a cambios en la localización subcelular o al reclutamiento de proteínas modificadoras de la cromatina u otras proteínas represoras de la transcripción.
- alterar conformacionalmente el FT produciendo a un cambio funcional.

La inspección de los motivos 6, 8, 10 y 12 reveló la presencia de la secuencia putativa de sumoilación,  $\Psi$ KXE/D. Una búsqueda más exhaustiva permitió identificar esta secuencia en las RCTs de poco más de la mitad de las proteínas analizadas, siendo especialmente abundante en ciertos grupos (Ib, Ic, II, proteínas de plantas dicotiledóneas del V y Pp; ver Figura III.6, p. 59).

La comparación de la abundancia de esta secuencia en las RCTs en relación a las RNTs, en las que se hallaron pocas; y el enriquecimiento en contraste con una secuencia similar ( $\Psi$ KX-[E/D]) fueron indicadores de que este hallazgo no sería azaroso. Además, se observó una frecuencia muy elevada de Val en la primera posición y Glu en la cuarta (ver Tabla III.4, p. 61 y Figura III.8, p. 61), lo que indica una mayor especificidad de los motivos encontrados.

En un trabajo reciente se identificaron proteínas recuperadas como conjugadas a SUMO en *A. thaliana* (Miller y col., 2010). De las 357 proteínas halladas, 81 (23 %) correspondían a FTs, entre los cuales se encontró AtHB6. Éste pertenece al grupo II y contiene un sitio SUMO en su RCT (ver Figura III.6, p. 59). Este hallazgo constituye una evidencia sumamente significativa que respalda la hipótesis de que la sumoilación es una modificación de importancia en la regulación de los FTs HD-Zip I. El hecho de que AtHB6 haya sido el único miembro encontrado podría deberse, entre otros factores, a que en las condiciones en las que fue obtenido el extracto proteico es la única proteína de esta subfamilia sumoilada o en suficiente abundancia para su detección.

En otro trabajo de publicación reciente se halló en *A. thaliana* que diferentes miembros de la familia MATH-BTB, proteínas encargadas del reconocimiento específico de blancos para su ubiquitinación y degradación, interactúan y regulan la estabilidad de AtHB6 (Lechner y col., 2011). Una de estas proteínas, BPM3 (BPM3) fue capaz de interactuar en ensayos de DHL con AtHB5 y AtHB16, del mismo grupo II que AtHB6, pero no con AtHB1, lo que sugiere que la degradación proteolítica no regularía todos los miembros de la subfamilia, o al menos no a través de los mismos actores.

En su conjunto, la evidencia experimental señala que habría múltiples niveles de regulación post-traducciona actuando sobre AtHB6 (Himmelbach y col., 2002; Lechner y col., 2011;



### III.4. Discusión de resultados

---

Miller y col., 2010). Los resultados obtenidos en el presente trabajo indican que al menos la fosforilación y la sumoilación podrían estar involucradas en la regulación de una gran parte de los FTs HD-Zip I.

Es importante resaltar que es conocido que varias proteínas son modificadas por ambos mecanismos, ubiquitinación y sumoilación. En algunos casos ambas modificaciones ocurren en la misma Lys de la proteína blanco (Wilkinson y Henley, 2010).

Un ejemplo que merece mención es el de la regulación de un FT extensamente caracterizado, ABI5 de la familia bZip. Al igual que varios genes HD-Zip I, este FT participa de la vía de señales de ABA, en la cual posee una función importante (Cutler y col., 2010). ABI5 es fosforilado en respuesta a ABA (Lopez-Molina y col., 2001) y, en ausencia de la fitohormona, es degradado por ubiquitinación y posterior proteólisis en el proteosoma 26S. Esta degradación puede ser inhibida por sumoilación, la cual lo estabiliza pero lo inactiva (Cutler y col., 2010). Este caso demuestra que todos estos mecanismos son capaces de actuar en la regulación de una única proteína, que además participa en la vía de señales de ABA, así como muchos de los HD-Zip I estudiados (ver Sección I.6.1.1, p. 14).

#### III.4.4 Las RNTs poseerían una funcionalidad limitada

Las RNTs presentaron una menor diversidad de motivos y una distribución con una especificidad por grupos también inferior (Figura III.9, p. 63). Los grupos que presentaron mayor número de motivos y cobertura de sus RNTs fueron el V y el Pp.

La predicción de sitios fosforilados involucró un menor número de proteínas y no coincidió tan extensamente con residuos conservados en los motivos. Sin embargo, el ensayo con AtHB6 en el cual se demuestra que la remoción de la RCT causa una reducción sensible en la capacidad de interacción con la proteína ABI1 (Himmelbach y col., 2002), indica que los sitios de fosforilación podrían ser importantes.

De todas formas, es importante destacar que el motivo 10, exclusivo del grupo VI, grupo que carece de motivos específicos en las RCTs, posee una señal de localización nuclear putativa en seis de los 10 motivos presentes. Esto podría indicar que dicho motivo tiene la función de dirigir el FT al núcleo.

Se hallaron muy pocos sitios potenciales de sumoilación por lo cual, aún si los hallados fueran funcionales, esta modificación no sería importante o extendida en las RNTs.

##### III.4.4.1 Otros posibles mecanismos de regulación

Con el objetivo de brindar un panorama más completo acerca de los diferentes niveles de regulación post-traducciona que podrían actuar sobre las proteínas HD-Zip I, se mencionarán dos mecanismos más que participan de la regulación de los FTs HD-Zip de la subfamilia III.

Por un lado, se ha hallado que proteínas pequeñas denominadas LITTLE ZIPPERs (ZPRs)

### III.4. Discusión de resultados

---

tienen la capacidad de dimerizar con los FTs impidiendo así que unan el ADN (Kim y col., 2008; Wenkel y col., 2007). Las ZPRs tienen entre 70 y 100 aa y consisten esencialmente en un LZ. Se ha demostrado que no son capaces de dimerizar con AtHB2 de la subfamilia II, lo que podría indicar que regulan únicamente miembros de la subfamilia III, aunque harían falta experimentos específicos para arribar a esta conclusión (Kim y col., 2008). De todas formas, aún si estas proteínas, u otras similares, no reconocen FTs HD-Zip I, se abre la posibilidad de investigar si FTs de la subfamilia I con HDs no funcionales son capaces de inhibir la actividad de otros por heterodimerización, como se propuso para HvHOX2 y VRS1 de cabada (Sakuma y col., 2010).

Por otro lado, se ha descubierto que el dominio MEKHLA de la proteína REV, un FT HD-Zip III, reconoce su propio LZ evitando la dimerización (Magnani y Barton, 2011). Esta función inhibitoria no estaría presente en al menos dos de las cuatro proteínas restantes de la subfamilia. Aunque no existan indicios de que algo similar pueda ocurrir con los FTs HD-Zip I, esta hipótesis tampoco ha sido puesta a prueba específicamente, por ejemplo con las RCTs.

#### III.4.5 Relevancia funcional de la RCT en la planta

Para evaluar la relevancia funcional de las diferencias en las RCTs de proteínas HD-Zip I se adoptó una estrategia que aprovechó los fenotipos caracterizados en *A. thaliana* generados por la expresión de los FTs de girasol, HaHB1 y HaHB4. Además de las plantas transgénicas de *A. thaliana* con las construcciones *35SCaMV::HaHB1* (Cabello y col., 2010) y *35SCaMV::HaHB4* (Dezar y col., 2005b) preparadas previamente, se transformaron plantas con tres construcciones más: *35SCaMV::H4-H1* con la RNT y el dominio HD-Zip fusionado a la RCT de HaHB1; *35SCaMV::H1SCT* y *35SCaMV::H4SCT* en las que se removió la RCT de HaHB1 y HaHB4, respectivamente.

En estas plantas se analizaron dos fenotipos fácilmente cuantificables y característicos de las construcciones de referencia. Estos parámetros fueron: el número de aserrados de las hojas de la roseta, el cual está aumentado en las plantas *35SCaMV::HaHB1* (Cabello y col., 2010); y la sensibilidad al ET mediante el recuento de plantas con el gancho apical desarrollado en la triple respuesta, ausente en las plantas *35SCaMV::HaHB4* (Manavella y col., 2006).

La cuantificación del número de aserrados mostró un comportamiento de las plantas *35SCaMV::HaHB4* y *35SCaMV::H4-H1*, líneas A y C, similar al de las plantas WT (Figura III.13, p. 67 y Tabla III.5, p. 68). Por otra parte, las plantas *35SCaMV::HaHB1* y *35SCaMV::H4-H1* línea B presentaron un aserrado significativamente mayor al resto. Este resultado muestra que la RCT de HaHB1 fue capaz de conferirle a la proteína quimérica H4-H1 la capacidad de alterar el fenotipo de aserrado, claramente en al menos una línea, cuando se la expresó en plantas transgénicas bajo el control del promotor *35SCaMV*.

El ensayo de triple respuesta produjo la insensibilidad descrita en las plantas

### III.4. Discusión de resultados

---

*35SCaMV::HaHB4*, y la sensibilidad esperable en las plantas *35SCaMV::HaHB1* y WT. Las plantas *35SCaMV::H4SCT* mostraron una insensibilidad intermedia ligeramente inferior a la generada por la expresión de *HaHB4*; y luego con una sensibilidad menor se ubicaron las plantas *35SCaMV::H4-H1* y *35SCaMV::H1SCT*. Esto demuestra, por un lado, que el dominio HD-Zip de *HaHB4* fue suficiente para provocar un fenotipo de insensibilidad significativo en las plantas *35SCaMV::H4SCT*. Sin embargo, la fusión a la RCT de *HaHB1* en las plantas *35SCaMV::H4-H1* disminuyó esta respuesta. Posiblemente si se hubiera contado en su momento con plantas que expresaran un FT quimérico inverso, es decir, H1-H4; se habría obtenido información valiosa para facilitar la interpretación de las observaciones.

El análisis de estos resultados no es tan claro, pero parece indicar que el dominio HD-Zip de *HaHB4* tendría un rol importante en la capacidad de la proteína de participar en la respuesta al ET. La adición de la RCT de *HaHB1* afecta esta actividad, pero produce un fenotipo intermedio, no totalmente equiparable al de las plantas *35SCaMV::HaHB1*. Sin embargo, se ve en el estudio del número de aserrados, que el dominio HD-Zip de *HaHB4* fusionado a la RCT de *HaHB1* es capaz incrementar esta característica, como *HaHB1* entero lo logra al ser expresado ectópicamente y constitutivamente.

Si se analizan las RCTs de ambos FTs se notará que *HaHB4* posee una RCT atípicamente corta para un miembro de la subfamilia I (64 aa), en la que no se encontraron motivos conservados ni sitios de sumoilación. La observación de su secuencia muestra que tiene un posible motivo AHA, pero con la particularidad de poseer un residuo básico. Pertenece al grupo I, pero al no tener una RCT conservada quedó fuera de los subgrupos Ia, Ib o Ic (Figura III.6, p. 59). *HaHB1* pertenece al grupo V y posee los motivos 2, 6, 11, 13, 15 y 20. Tiene además un sitio de sumoilación y un motivo AHA típico.

Las marcadas diferencias de las RCTs podrían determinar que *HaHB1* sea sumoilada y *HaHB4* no, y que sus motivos AHA distintos les permitan interactuar con diferentes proteínas de la maquinaria basal de transcripción o diferentes cofactores. Es más, la capacidad del dominio HD-Zip de *HaHB4* solo (plantas *35SCaMV::H4SCT*), de producir un elevado grado de insensibilidad al ET podría indicar que *HaHB4* actúa en esta vía mediante un mecanismo que prescinde de su RCT. Por ejemplo, heterodimerizando y alterando la función de otros FTs HD-Zip I, o bloqueando por competencia la unión a elementos en *cis* de diferentes promotores.

En contraste, la RCT de *HaHB1* por si sola sería capaz de “alterar la función de *HaHB4*”, confiriéndole a su HD-Zip la capacidad de participar en las vías de señales que conducen al fenotipo de aserrado aumentado.

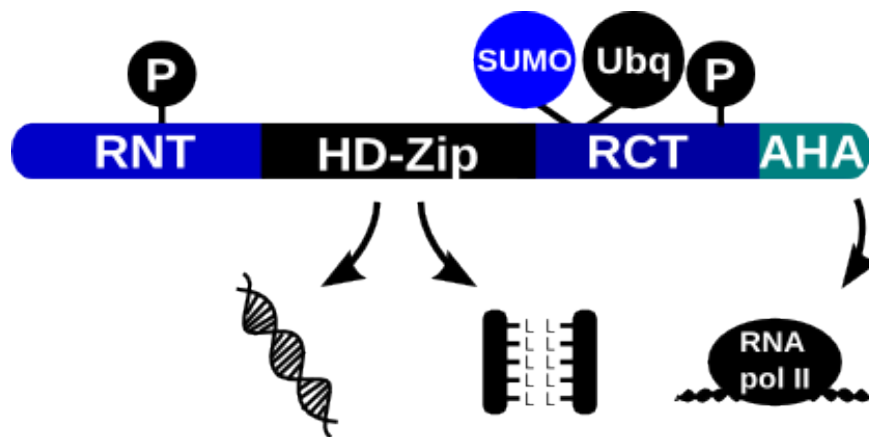
Estos resultados destacan la participación de la RCT en la determinación de la función de un FT HD-Zip I. La divergencia entre las RCT de los diferentes grupos, pero conservando una elevada similitud entre sus integrantes, podrían indicar que éstos poseen mecanismos de acción y regulación con cierta especificidad de grupo. La capacidad de interactuar a través del motivo AHA o de otros motivos con diferentes cofactores o proteínas reguladoras podría ser una fuente

de divergencia funcional entre estos FT en un mismo organismo.

#### III.4.6 Un modelo de la relación estructura/función de los FTs HD-Zip I

A modo de resumen de los resultados obtenidos más lo conocido previamente sobre los FTs HD-Zip I, se puede postular un modelo más completo respecto a la función de los diferentes dominios o motivos presentes en estas proteínas (Figura III.15).

El LZ le provee la capacidad a la proteína de dimerizar, con lo cual el HD adquiere la función de reconocimiento del ADN, posicionando la proteína en el promotor del gen que activará o reprimirá. Regulando esta actividad se suma la fosforilación/desfosforilación de residuos específicos en la RNT y/o RCT, y la sumoilación en la RCT. Además, la estabilidad de la proteína sería potencialmente controlada por ubiquitinación. Una vez unido al ADN, el FT sería capaz de contactar directa o indirectamente la maquinaria de transcripción a través del dominio AHA, cuya naturaleza le daría especificidad a esta interacción.



**Figura III.15:** Modelo de la función de los diferentes dominios y motivos presentes en los FTs HD-Zip I.

## III.5 Conclusiones

En función de los resultados obtenidos se arribó a las siguientes conclusiones

1. Las reconstrucciones filogenéticas basadas en el dominio HD-Zip y las secuencias completas de las 178 proteínas HD-Zip I permitieron definir seis grupos de proteínas estrechamente relacionadas. El árbol con las secuencias completas junto a la inspección de las RCTs demostraron ser un recurso valioso para dicha definición.
2. Los grupos estuvieron formados únicamente por proteínas de plantas monocotiledóneas y eudicotiledóneas, lo que indicó que la aparición de estas proteínas es anterior a la separación de estos taxones, pero posiblemente posterior a la separación entre angiospermas y gimnospermas.
3. Las RCTs presentaron motivos conservados, muchos de los cuales fueron específicos de los diferentes grupos.
4. El motivo presente en las porciones finales de las RCT presentó las características del motivo activador AHA. La capacidad de transactivación de este motivo en la proteína AtHB13 fue demostrada mediante ensayos de simple híbrido en levaduras. Sumando esta evidencia a los trabajos que destacaban la importancia de esta porción de la proteína es posible proponer que los FTs HD-Zip I poseen la capacidad de activar la transcripción. Además, las variaciones en esta secuencia podrían ser responsables de determinar las proteínas con las que interactúa cada FT para ejercer esta actividad, difiriendo especialmente entre miembros de distintos grupos.
5. Las RCTs también presentaron varios sitios putativos de fosforilación y sumoilación en motivos con residuos significativamente conservados. La evidencia previa respecto de que algunos FTs HD-Zip I son regulados por estos mecanismos y por ubiquitinación indicarían que la RCT tendría una función importante en la regulación de la actividad de la proteína.
6. La RNT cumpliría una función menor, poseyendo algunos residuos putativos de fosforilación y un señal de localización nuclear en las proteínas de uno de los grupos.
7. La importancia funcional de la RCT *in planta* fue evaluada en los ensayos con una proteína quimérica (H4-H1) y versiones sin RCT de las proteínas HaHB1 y HaHB4. Los resultados destacan la participación de esta región en la determinación de las vías de señales en las cuales es capaz de actuar el FT.

## Capítulo IV

# Redes de expresión génica en las que participan los factores de transcripción HD-Zip de *Arabidopsis thaliana*

*“Sistema no lineal disipativo alejado del equilibrio que evoluciona irreversiblemente”.*

*Definición de ser vivo - Ilya Prigogine*

*“Cualquier sistema autónomo con capacidad evolutiva abierta”.*

*Definición de ser vivo - Kepa Ruiz-Mirazo, Juli Peretó y Álvaro Moreno*

### IV.1 Introducción

La definición de ser vivo dada por Prigogine está marcada por su condición de químico especializado en el área de la termodinámica (más allá de los alcances y aplicaciones a otras disciplinas que él y otros le dieron a sus trabajos). Ésta pone especial énfasis, a través de la definición de sistema disipativo, en el hecho de que los seres vivos son sistemas abiertos auto-organizados que mantienen un estado de no equilibrio intercambiando materia y energía (e información) con su medio (Prigogine, 1988). Se resalta entonces que un ser vivo necesita, para no morir, mantenerse en un continuo estado de no equilibrio, lo cual está asociado, pero no se reduce, al mantenimiento de la homeostasis interna, estado interno que implica una estabilidad sólo alcanzada gracias a mecanismos de autorregulación complejos y activos.

Lo que esta definición no considera está implícito en la provista por Ruiz-Mirazo y col. (2004) y en otras definiciones formuladas por diversos autores (Moreira y López-García, 2009; Ruiz-Mirazo y col., 2004). La capacidad de evolucionar está asociada directamente a la de herencia y por ende a la de información. Como lo mostraron Avery y col. (1944) e innumerables estudios sucesivos, esta información está esencialmente contenida en el ADN, molécula presente en todos los seres vivos.

Ambas propiedades de los seres vivos, la de sistema complejo autorregulado y la de contenedor de información, están lejos de ser independientes; por el contrario, son virtualmente indisociables. El ADN contiene información gracias a que ésta puede ser “decodificada” por la maquinaria que conforma ese ser vivo, y en consecuencia esta maquinaria condiciona el “significado” de esa información. Inversamente, el sistema autorregulado debe su funcionamiento en gran medida a las enzimas y demás proteínas, las cuales están codificadas en el ADN. Retomando la idea de las principales transiciones a lo largo de la evolución, no resulta sorprendente que en un origen contener la información genética y catalizar reacciones químicas fueran tareas esencialmente desempeñadas por una misma molécula, el ARN.

En última instancia, la información genética es transmitida dando lugar a proteínas y diferentes tipos de moléculas de ARN no codificantes (siendo las principales, ARNt, ARNr y ARNm). El flujo de esta información está dado por el “Dogma central de la Biología” (Crick, 1970); de ADN a ARN mediante la transcripción, y de ARN (ARNm) a proteína a través de la traducción. En conjunto, estos procesos son denominados “expresión génica”.

La conformación y composición del organismo dependerá entonces de la regulación precisa, en forma espacial y temporal, de la expresión génica y de sus productos, por lo que existen múltiples mecanismos capaces de controlarla que actúan en varios niveles. En organismos eucariotas, algunos de ellos son: la condensación de la cromatina, la metilación del ADN, la regulación del inicio de la transcripción, el empalme o *splicing* alternativo, la regulación de la estabilidad del ARNm, la regulación de la traducción, las modificaciones post-traduccionales de la proteína, la regulación de su localización subcelular y de su degradación (Wray y col., 2003).

## IV.1. Introducción

---

Es a través de estos procesos que el organismo coordina la expresión de diferentes genes, logra llevar adelante su programa de desarrollo y responder a los estímulos externos del medio en el que se encuentra.

### IV.1.1 Las redes de regulación génica

Uno de los mecanismos centrales en la regulación de la expresión génica es la regulación del inicio de la transcripción, siendo los factores de transcripción (FT) las principales proteínas encargadas de esta función. El número de genes que es capaz de controlar un FT es muy variable, y generalmente varios FTs actúan en la regulación de la expresión de cada gen (Wray y col., 2003).

De esta forma, si consideramos el conjunto de los genes y los FTs que los regulan, ya sea a nivel de una vía de señales, de una célula o de un organismo; es fácil reconocer que se forma una red. Ésta se denomina red de regulación génica, o más precisamente en este caso, red de regulación transcripcional.

Estas redes suelen representarse por lo que en matemática se denomina “grafos” (Barabási y Oltvai, 2004). Los genes son simbolizados por elementos puntuales denominados nodos o vértices, y entre ellos se trazan líneas o aristas que indican que existe una interacción. En el caso de las redes de regulación transcripcional la línea se traza entre un FT y el gen que regula, que puede ser a su vez un FT o no.

#### IV.1.1.1 Propiedades emergentes

La red puede reconstruirse a partir de la evidencia que surge de estudios analíticos, en los que se considera un número limitado de genes. Pero en la práctica, considerando las decenas de miles de genes que contiene un organismo eucariota complejo y al gran número de interacciones posibles, este enfoque no permite conocer más que una porción muy pequeña de los procesos que ocurren. Sin cuestionar la indudable importancia y confiabilidad de los resultados obtenidos de este modo, es una metodología que sin duda nos brinda un panorama restringido de la complejidad real de la red.

Existen otros enfoques para el estudio de las redes de regulación que han sido posibilitados por el desarrollo de tecnologías de medición simultánea de un gran número de componentes o parámetros de la red. Estos enfoques pertenecen a un área de la biología denominada “Biología de Sistemas”, y su utilización no implica únicamente un cambio de escala respecto al volumen de información analizada, sino la noción de que los estudios globales permiten reconocer nuevas propiedades o características (“propiedades emergentes”) que no son reconocibles a partir de estudios de componentes individuales (Long y col., 2008).

Por ejemplo, una de las características generales de las redes de regulación es la de ser redes “libres de escala” (Barabási, 1999). Esto significa que la probabilidad  $P(k)$  de que un componente



## IV.1. Introducción

---

posea un número  $k$  de interacciones o vértices sigue una ley de potencia ( $P(k) \sim k^{-\gamma}$ ; siendo generalmente,  $2 \leq \gamma \leq 3$ ). En consecuencia, la mayoría de los genes tendrá pocas interacciones y unos pocos interaccionan con un gran número de genes. Este conocimiento ha permitido formular modelos sobre la evolución de las redes (crecimiento con conexión preferencial a la red) o sobre el flujo de información en la red (fenómeno de “mundo pequeño”; Barabási, 1999; Barabási y Oltvai, 2004).

Otra propiedad importante de las redes de regulación génica es que son jerárquicas y modulares (Ravasz y Barabási, 2003). Hay conjuntos (o *clusters*) de genes altamente relacionados entre sí, con escasa interacción con genes por fuera del conjunto. Por sobre ellos hay genes que coordinan o regulan diferentes conjuntos, y así sucesivamente, generando capas o jerarquías. Esta propiedad se corresponde con el concepto de cascadas de señales, en el que un FT de jerarquía alta desencadena la expresión de otros FTs que a su vez controlan otros genes y así se desencadena una respuesta a la señal o estímulo (Brivanlou y Darnell, 2002).

### IV.1.2 Ingeniería reversa de redes de regulación

#### IV.1.2.1 Niveles de expresión génica

La obtención de los datos constituye el primer paso para la reconstrucción de una red de regulación génica. Los avances tecnológicos en el área del análisis de ácidos nucleicos han provisto, a lo largo del tiempo, las herramientas más poderosas para los estudios moleculares a nivel de sistemas. De hecho, el desarrollo relativamente reciente de las tecnologías de secuenciación de alto rendimiento (también conocidas como tecnologías de secuenciación de nueva generación; Metzker, 2009) parece indicar que se mantendrá este “desfasaje” en el potencial de generación de información de las tecnologías empleadas en genómica y transcriptómica, respecto a las utilizadas en proteómica y metabolómica.

En consecuencia, a pesar de que las proteínas son el producto final de la expresión génica, y los principales efectores y coordinadores de las funciones celulares, la mayor facilidad para cuantificar ARNm ha hecho que la mayor parte de los estudios se base en su medición. Si además se tiene en cuenta la complejidad de la regulación génica, la manipulación de los datos y la interpretación de los resultados requiere de la asunción de algunos supuestos (Margolin y Califano, 2007). En el caso de las redes de regulación transcripcional, se asume que la medición del transcripto de un FT es una medida directa de su actividad y que el nivel de ARNm de cualquier gen es producto exclusivo de la acción de los FTs que lo controlan; aun conociendo la existencia de mecanismos de regulación de estas especies, como la modificación post-transcripcionales de las proteínas y la acción de los ARNm sobre el ARNm.

### IV.1.2.2 Los microarreglos de ADN

Existen diferentes metodologías para la cuantificación de los niveles existentes de un determinado transcripto en una célula o tejido en un momento dado. La primera desarrollada fue la del *Northern blot* (Alwine y col., 1977), en el que una muestra de ARNm separada electroforéticamente es transferida e inmovilizada sobre una membrana. Luego se la enfrenta a una sonda de ADN o ARN “marcada”, generalmente en forma radioactiva, que es complementaria al ARNm que se quiere cuantificar. Esta técnica no presentó metodologías alternativas hasta la invención de la reacción en cadena de la polimerasa (PCR; Saiki y col., 1985) y de su aplicación para cuantificar ARNm (Heid y col., 1996; Higuchi y col., 1993). La técnica de PCR cuantitativa (PCRc) mejora su potencial continuamente, siendo actualmente la de mayor uso para la cuantificación de mensajeros por su posibilidad de medir varios simultáneamente y su alto nivel de sensibilidad (Pfaffl, 2010).

Sin embargo, estas metodologías sólo permiten la medición simultánea de los niveles de expresión de un número relativamente limitado de genes. Las mediciones a escala genómica se convirtieron en una posibilidad recién luego de que Schena y col. (1995) desarrollaran, casualmente para *Arabidopsis thaliana* (*A. thaliana*), la técnica de los microarreglos de ADN (en inglés, *DNA microarrays*). Ésta consiste, a grandes rasgos, en la disposición sobre una superficie o matriz de un arreglo conocido de sondas que hibridan con derivados de los ARNm marcados fluorescentemente. La medición posterior de los niveles de fluorescencia permite la cuantificación de los niveles de los transcriptos (Kohane y col., 2002). La miniaturización de este arreglo posibilita disponer en una pequeña superficie de todas las sondas necesarias para cuantificar virtualmente el total de los ARNm de un organismo, o inclusive un número suficiente de sondas para cubrir el genoma completo en los denominados *tiling arrays* (Mockler y col., 2005).

### IV.1.2.3 Algoritmos para la reconstrucción de redes de regulación génica

Los datos que se obtienen de un microarreglo de ADN pueden analizarse mediante métodos de agrupamiento. Éstos se basan en la suposición de que aquellos genes que poseen un patrón de expresión similar, habiendo variadas formas de medir esa similitud, estarán funcionalmente relacionados (Margolin y Califano, 2007). Pero de esta forma no se están reconociendo los mecanismos de regulación ni cuáles son las relaciones directas existentes. Margolin y col. (2006a) demostraron que, utilizando una red de regulación génica sintética, las interacciones directas e indirectas no pueden ser distinguidas utilizando simplemente umbrales de similitud.

El objetivo de la ingeniería reversa (o deconvolución) de las redes de regulación es reconstruir, a partir de múltiples observaciones o mediciones de diferentes estados del sistema, las relaciones o interacciones existentes entre sus componentes. Los principales métodos que han sido empleados para esta reconstrucción son: los modelos lineales, las redes Bayesianas y los métodos basados

## IV.1. Introducción

---

en la teoría de la información (Alvarez-Buylla y col., 2007; Margolin y Califano, 2007).

La ingeniería reversa de redes de expresión génica ha sido aplicada en *A. thaliana* en varios trabajos (Alvarez-Buylla y col., 2007; Long y col., 2008). Por ejemplo, Needham y col. (2009) aplicaron un algoritmo basado en redes bayesianas con la particularidad de iniciar la red a partir de un grupo de genes de interés e introducir iterativamente el resto de los genes, logrando reducir el tiempo de cómputo. Esta estrategia fue aplicada a la base de datos de NASC (*Nottingham Arabidopsis Stock Centre*) con 2.904 microarreglos y puesta a prueba con una red caracterizada extensamente, la red de regulación génica del reloj circadiano. Los resultados obtenidos fueron satisfactorios ya que mostraron ser consistentes con los modelos existentes sobre esta red (Needham y col., 2009).

### IV.1.3 La información mutua

La Teoría de la Información es una área de la matemática y de las ciencias de la computación iniciada por el trabajo de Shannon (1948), que sentó sus bases fundamentales. Fue desarrollada originalmente para disponer de herramientas matemáticas que permitieran abordar el problema del “ruido” en la correcta transmisión de mensajes, y así comprender las limitaciones teóricas de la comunicación. Sin embargo, sus formulaciones tuvieron un alcance mucho más amplio (Yeung, 2008).

En la ingeniería reversa de redes de regulación génica, una de las medidas utilizadas de asociación entre genes es la información mutua (IM), que pertenece a la Teoría de la Información. Para explicarla, en primer lugar deben introducirse las siguientes definiciones (Yeung, 2008):

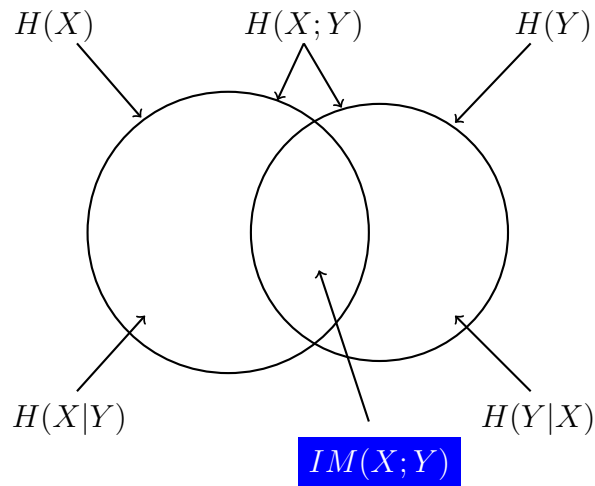
**Independencia estadística** Las variables aleatorias  $X$  e  $Y$  son independientes si  $p(x, y) = p(x)p(y)$ , para todo  $x$  e  $y$ .

**Información mutua** Para las variables aleatorias  $X$  e  $Y$ , la información mutua entre  $X$  e  $Y$  se define como

$$IM(X, Y) = \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (\text{IV.1})$$

Puede decirse que la IM es una medida de la reducción en la incertidumbre acerca del valor de  $X$  que se produce al conocer el valor de  $Y$ . Como la IM es simétrica, lo mismo puede decirse del valor de  $Y$ , si lo que se conoce es  $X$ . Esta interpretación puede representarse mediante el diagrama de la información (Figura IV.1).

Si analizamos la medida de IM en términos de la definición de independencia estadística, es posible reconocer que ante el caso extremo de que las variables  $X$  e  $Y$  sean completamente independientes, el valor de  $\frac{p(x,y)}{p(x)p(y)}$  será unitario, su logaritmo cero y en consecuencia la IM será nula.



**Figura IV.1:** Diagrama de la información. La incertidumbre respecto a las variables  $X$  o  $Y$  es medida mediante la entropía  $H(X)$  e  $H(Y)$  de las variables. Puede visualizarse que a mayor  $IM(X, Y)$ , menor es la incertidumbre de una variable dado que se conoce la otra ( $H(X|Y)$  u  $H(Y|X)$ ).

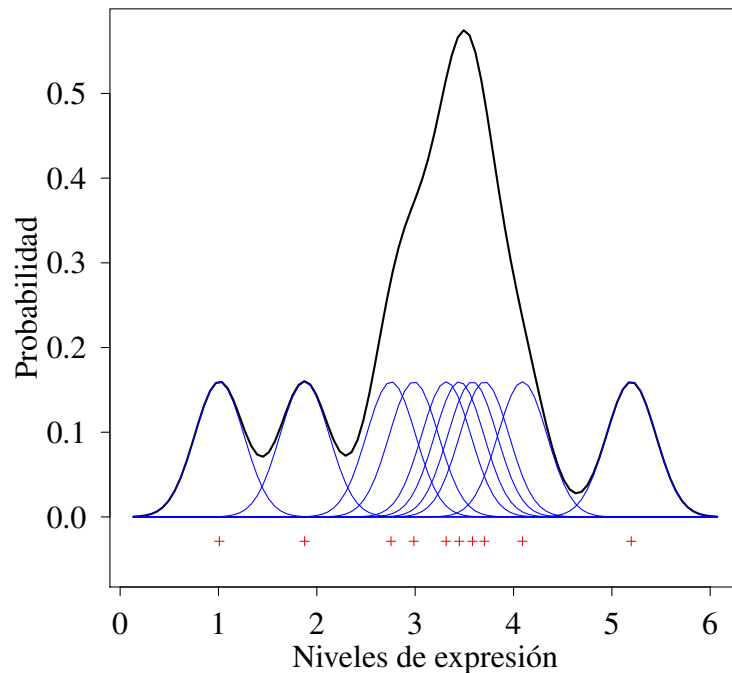
### IV.1.4 Algoritmo para la reconstrucción de redes celulares precisas

El algoritmo para la reconstrucción de redes celulares precisas, ARACNE (del inglés, *Algorithm for the Reconstruction of Accurate Cellular Networks*) es un programa de reconstrucción de redes basado en un algoritmo que cuenta, entre otras cualidades, con la capacidad de distinguir interacciones directas de interacciones indirectas (Margolin y col., 2006a). Los principios en los que se basa y su implementación son, *a priori*, lo suficientemente generales para que pueda ser empleado para la reconstrucción de redes de diversa naturaleza (biológicas, sociales, ingenieriles). No obstante, fue desarrollado para la ingeniería reversa de redes de regulación transcripcional y validado en redes de regulación de linfocitos B humanos (Basso y col., 2005).

ARACNE define un vértice como una dependencia estadística irreductible entre los perfiles de expresión de dos genes, es decir, una relación que no puede ser explicada como un artefacto producto de otra dependencia estadística de la red (Margolin y col., 2006a). Basándose en esta definición, el algoritmo se divide, a grandes rasgos, en dos etapas: el cálculo de las dependencias estadísticas, para lo cual utiliza la medida de IM; y el reconocimiento de aquellas que son estadísticamente irreductibles.

#### IV.1.4.1 Cálculo de la información mutua

El cálculo de las IMs requiere previamente del cálculo de las probabilidades de que cada gen se exprese en diferentes niveles (probabilidades marginales,  $p_X(x)$  y  $p_Y(y)$ ) y la probabilidad conjunta de expresión en diferentes niveles ( $p_{X,Y}(x, y)$ , ver Ecuación IV.1). En otras palabras,



**Figura IV.2:** Estimación de densidad utilizando un estimador no paramétrico de núcleo Gaussiano.

esto implica obtener una distribución de probabilidades para la expresión de cada gen y una distribución conjunta de probabilidades para el par.

La forma en que ARACNE calcula estas distribuciones de probabilidad es a través de un estimador de densidades de núcleo (del inglés, *kernel*) Gaussiano (Margolin y col., 2006a). Es una metodología no paramétrica de estimación de densidades que tiene alguna similitud con el histograma. En un histograma se discretiza el rango de valores que toma la variable dividiéndolo en intervalos, y se cuenta el número de valores comprendidos en cada uno. Finalmente se obtiene la función de distribución de probabilidades normalizando el área de las barras. En el caso del estimador de núcleo Gaussiano, se considera una función normal centrada en cada valor y la función de densidad de probabilidad es la curva que resulta de la sumatoria normalizada de estas funciones (ejemplo de la metodología en la Figura IV.2). Para el caso de la distribución conjunta de densidades se hace el cálculo en dos dimensiones utilizando la distribución normal bivariada como función estimadora.

Finalmente, el algoritmo filtra los valores de IM considerados estadísticamente no distintos de cero, usando un umbral  $IM_0$ . Calcula este umbral a partir de un valor  $p$  específico ( $p_0$ , definido por el usuario), para la hipótesis nula de independencia en la expresión del par de genes considerado.

Cabe destacar que el cálculo de la IM no involucra asunciones o modelos acerca de la relación de dependencia existente entre el par de genes interaccionantes, por lo cual admite tanto relaciones lineales como no lineales. Aunque las redes Bayesianas poseen la misma ventaja,

requieren muchos más recursos computacionales (Margolin y col., 2006a).

También es esencial resaltar que la simetría inherente al cálculo de la IM hace que a partir de la aplicación del algoritmo no se pueda conocer la direccionalidad de la interacción. ARACNE no brinda información acerca de cuál de los dos genes que interactúan es el regulador y cuál el regulado, por lo cual las interacciones deben considerarse, *a priori*, bidireccionales o adireccionales (Margolin y col., 2006a). No obstante, la información que se pueda aportar sobre qué genes son FTs y cuáles no, permite inferir la dirección de la interacción; excepto en el caso en el que los genes relacionados sean ambos FTs.

Una asunción importante que realiza ARACNE es que no hay interacciones de orden mayor a dos. Es decir, no considera estadísticamente las interacciones que surgirían del análisis de las probabilidades conjuntas de los niveles de expresión de tres o más genes y de su IM. La razón es que en problemas de escala genómica, para estimar la densidad probabilidad conjunta de un par de genes es necesario un muestreo de 100 o más tratamientos (en este caso microarreglos); pero esta estimación para tres genes requiere un número de tratamientos de alrededor de un orden de magnitud mayor (Margolin y col., 2006a). De por sí, la realización de 100 o más microarreglos ya representa una cantidad elevada desde una perspectiva de costos. Podría considerarse que estadísticamente el gran número de tratamientos necesarios para aplicar esta técnica es el “costo” de no formular modelos respecto a las relaciones existentes entre los pares de genes; y el beneficio es la posibilidad de recuperar tanto interacciones lineales como no lineales.

### IV.1.4.2 Remoción de interacciones indirectas

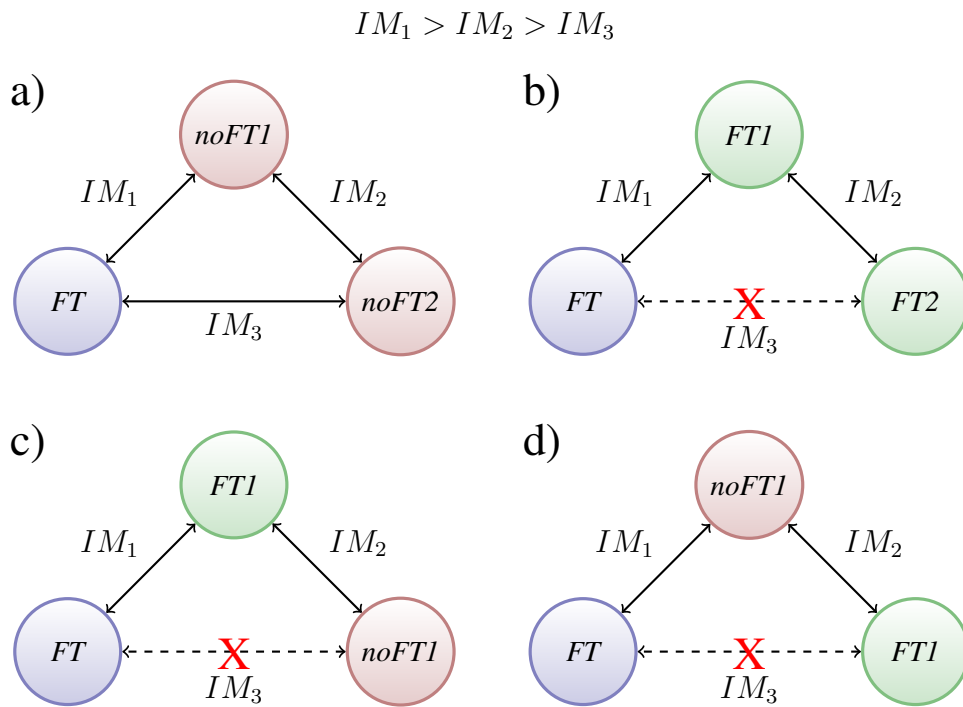
La segunda etapa en la reconstrucción de la red transcripcional con ARACNE es reconocer las relaciones estadísticas irreducibles. Para esto, ARACNE utiliza un algoritmo al que denomina *Data Processing Inequality* (DPI; Margolin y col., 2006a).

Considerando los tres genes  $g_1$ ,  $g_2$  y  $g_3$ , y las tres posibles IMs entre los pares, el fundamento del DPI es que si la interacción entre  $g_1$  y  $g_3$  es a través de  $g_2$ , entonces se cumplirá que:

$$IM(g_1, g_3) \leq \min [IM(g_1, g_2); IM(g_2, g_3)]$$

Lo que significa que la IM entre los genes relacionados indirectamente ( $IM(g_1, g_3)$ ) será igual o menor a las IMs de los pares de genes directamente relacionados ( $IM(g_1, g_2)$  y  $IM(g_2, g_3)$ ). Basándose en esta desigualdad, ARACNE compara las IMs de todos los triplete de genes de la red y remueve aquellas aristas o interacciones en las que se cumple la desigualdad, es decir, en las que la interacción de los genes del par con un intermediario es mayor a la interacción que tienen entre sí (Margolin y col., 2006a).

Sin embargo, al estar reconstruyendo una red transcripcional surgen algunas excepciones. ARACNE utiliza información provista por el usuario sobre qué genes son FTs para condicionar la remoción de aristas. De esta forma, centrando el análisis en un FT ( $g_1$ ), pueden darse cuatro



**Figura IV.3:** Tripletes posibles durante el análisis de DPI. En los incisos a) al d) se plantean las posibles funciones de los genes que forman un triplete durante el análisis de DPI. En cada caso el FT sobre el cual se está centrando el análisis figura en la esquina inferior izquierda del triplete.

situaciones, ilustradas en la Figura IV.3:

- a) ninguno de los genes  $g_2$  y  $g_3$  son FTs: no se puede fundamentar a través de interacciones transcripcionales indirectas que la  $IM(g_2; g_3)$  no sea la menor de las tres IM. Entonces, no se remueve la interacción entre el FT y  $g_3$  (entre los que la IM era menor).
- b) los tres genes son FTs: se remueve la interacción porque la relación indirecta puede darse a través de uno de los FT.
- c)  $g_2$  es un FT y  $g_3$  no: como en el caso anterior, la relación puede explicarse transcripcionalmente. También puede ocurrir que  $g_2$  controle ambos genes.
- d)  $g_2$  no es un FT y  $g_3$  sí: se remueve la interacción porque la relación puede surgir indirectamente ya que ambos FTs,  $g_1$  y  $g_3$ , son reguladores de  $g_2$ .

### IV.1.5 Redes de regulación en las que participan los FTs HD-Zip I y II

Se han realizado numerosos estudios funcionales de los genes HD-Zip I y II, como se describió en la Sección I.6 (p. 13). Sin embargo, el enfoque utilizado ha sido generalmente analítico, es decir, se han estudiado uno o unos pocos FTs en respuestas puntuales con estrategias

no sistémicas. Aun en el caso de los trabajos en *A. thaliana* de Henriksson y col. (2005) y de Ciarbelli y col. (2008), en los que se estudiaron todos los miembros de las subfamilias I y II, respectivamente; los resultados fueron generados con un enfoque y con herramientas no sistémicas.

Podría considerarse una excepción, por ejemplo, el uso de microarreglos de ADN para el estudio de tres FTs de girasol. En el caso de *HaHB4*, los microarreglos permitieron reconocer las vías de señales que se encontraban modificadas en plantas de *A. thaliana* que lo expresaban en forma ectópica y constitutiva. Esto permitió descubrir su participación principalmente en las vías de señales de las fitohormonas ET y JA (Manavella y col., 2006, 2008b). El estudio de *HaHB1* mostró que regulaba las vías de señales de respuesta a temperaturas de congelamiento, a través de la inducción de proteínas estabilizadoras de membranas con actividad anticongelamiento (*antifreeze*; Cabello y col., 2012). Por último, se utilizaron microarreglos para demostrar que el gen HD-Zip II *HaHB10* participa en la transición del estadio vegetativo al reproductivo, y en las vías de señales de SA y JA en respuesta a la infección por patógenos (Dezar y col., 2011).

El estudio de *MtHB1* de *M. truncatula* constituye un caso similar. El análisis de ensayos de microarreglos realizados previamente con el objetivo de detectar genes con expresión diferencial frente a estrés salino, permitió identificar la participación de *MtHB1* y *LBD1* en esta respuesta. En el promotor de *LBD1* hay un elemento CAATNATTG, lo que condujo finalmente a demostrar que era un blanco directo de este FT HD-Zip (Ariel y col., 2010).

Los trabajos mencionados, más allá de utilizar herramientas de análisis global, tuvieron como principal objetivo emplearlas con fines exploratorios que permitieran dilucidar la participación de los diferentes FTs en distintas vías de señales. No se ha realizado al momento un estudio sistémico de las redes transcripcionales en las que están involucrados los genes HD-Zip I y II que permita proponer cuáles son los potenciales genes que regulan y posibilite un análisis comparativo entre miembros.



### IV.2 Objetivos

Se planteó el empleo de una metodología de ingeniería reversa de redes de regulación transcripcional para reconocer conjuntos de potenciales genes blanco de los FT HD-Zip I y II. Los resultados permitirán hacer un análisis comparativo entre los miembros. Particularmente, los objetivos propuestos fueron los siguientes:

- 1) Obtener mediciones transcriptómicas de plantas tratadas con diferentes estreses abióticos de una base de datos pública. Remover microarreglos con réplicas muy variables y genes con perfiles de expresión poco informativos.
- 2) Obtener una red de regulación transcripcional con el programa ARACNE.
- 3) Analizar la presencia del elemento en *cis* reconocido por los FTs HD-Zip en los promotores de cada conjunto de genes blanco (o MAT). Utilizar también una estrategia de búsqueda de elementos *de novo*.
- 4) Comparar el grado de redundancia funcional entre los FTs estudiando la intersección o solapamiento de los MATs de cada uno.
- 5) Reconocer la participación de los FTs en los diferentes estreses abióticos a través de la expresión coordinada de los genes de cada MAT.
- 6) Estudiar funcionalmente los MAT para conocer en qué vías de señales está involucrado cada FT.

### IV.3 Resultados

#### IV.3.1 Datos analizados

En primer lugar, se seleccionó el conjunto de datos a analizar. La necesidad de contar con un número muy elevado de ensayos de microarreglos para poder aplicar el algoritmo ARACNE impulsó la búsqueda de la información en las grandes bases de datos públicas.

La matriz de expresión utilizada para el análisis provino del proyecto a gran escala AtGenExpress de análisis transcritoómico de *A. thaliana* var. Col-0, coordinado por el AFGN (*Arabidopsis Functional Genomics Network*). Los experimentos del proyecto AtGenExpress fueron realizados con el microarreglo ATH1 de genoma completo de Affymetrix® (Redman y col., 2004). Éste brinda la posibilidad de cuantificar simultáneamente la expresión de alrededor de 24.000 genes que codifican proteínas. En rigor, en el microarreglo hay dispuestos conjuntos de oligonucleótidos o *probeset* para la medición de cada ARNm. Sin embargo, en algunos casos estos *probesets* son capaces de hibridar con los productos de más de un gen, y algunos ARNm son reconocidos por más de un *probeset*. Por esto se utilizará la palabra *probeset* para referirse a cada medición que se obtiene en el microarreglo y a la especie que se cuantifica, pues no siempre una medición corresponde al ARNm de un único gen.

En particular, el conjunto de datos de expresión empleado correspondió a los obtenidos en tratamientos de estrés del tipo abiótico: déficit hídrico, daño mecánico (o *wounding*), calor, frío, alta salinidad, condiciones oxidativas, estrés genotóxico y luz UV. Las características de los tratamientos se describen en la Sección II.2.1 (p. 36). Este conjunto de microarreglos fue estudiado por Kilian y col. (2007) con la intención de hallar conjuntos de genes regulados en forma conjunta y contrastar estos conjuntos entre los diferentes estreses. No se utilizaron los datos para hacer una reconstrucción de las redes de regulación que dan origen a estas observaciones.

Cada una de las condiciones mencionada fue evaluada en diferentes tiempos y además cada plántula fue dividida en vástago y raíz para su procesamiento. Finalmente, cada una de estas condiciones de tiempo y tejido fue analizada por duplicado, completando un total de 298 microarreglos.

La elección de ensayos realizados en diferentes condiciones de estrés abiótico respondió al conocimiento de que los genes HD-Zip, en especial los de la subfamilia I, desempeñan una función importante en las respuestas a este tipo de estrés (ver Sección I.6, p. 13).

#### IV.3.2 Filtrado de los datos

Los 298 microarreglos disponibles en la base de datos se encontraban normalizados por lo cual no requirieron mayor procesamiento. Con el objetivo de emplear ensayos confiables, en primer lugar se analizó la calidad de los microarreglos y el nivel de variación entre duplicados. Basándose en este criterio se descartaron 29 microarreglos, quedando 269 para continuar con el

análisis.

Luego, con el objetivo de emplear los *probesets* cuya expresión tuviera un contenido informativo significativo y confiable para el análisis mediante ARACNE (Margolin y col., 2006b), se empleó un criterio de filtrado de los *probesets* basado en tres parámetros estadísticos:

1)  $K$  muestras sobre  $A$  ( $KsA$ )

Consistió en eliminar aquellos *probesets* en los cuales no hay al menos  $K$  tratamientos que superan un valor umbral de expresión  $A$ . El objetivo fue remover *probesets* cuya expresión no superara un cierto número de veces el propio “ruido de fondo”.

2) Desviación estándar ( $\sigma$ )

Se utilizó el estadístico desviación estándar para eliminar los *probesets* inferiores a cierto umbral. El propósito fue eliminar los *probesets* con bajo contenido informativo debido a variaciones de expresión poco significativas asociadas a la variabilidad propia de la cuantificación en el microarreglo, o a que su variabilidad real fuera escasa.

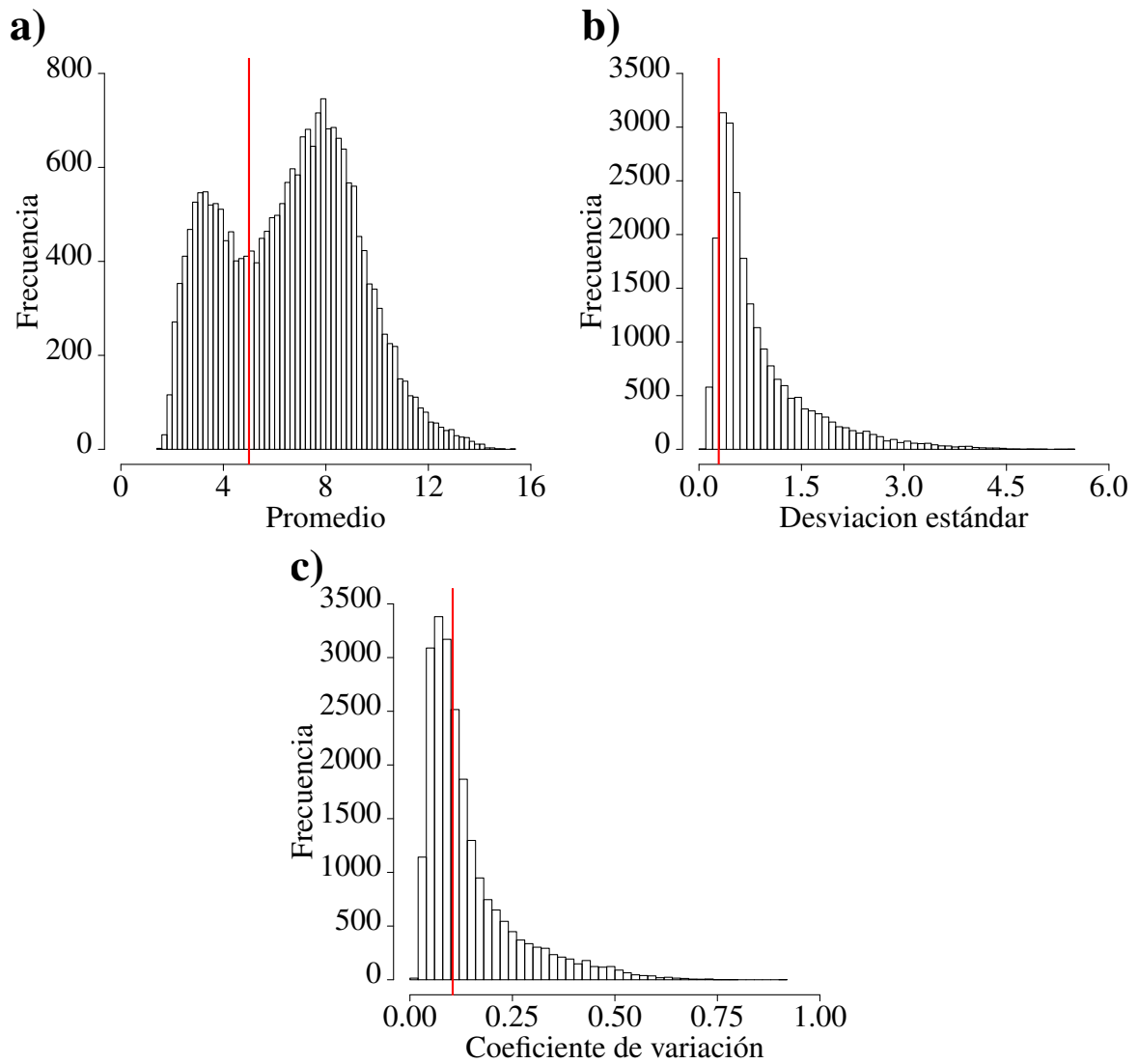
3) Coeficiente de variación (CV)

El estadístico empleado fue el coeficiente de variación y se eliminaron los *probesets* inferiores a cierto valor. En este caso el objetivo es, al igual que con la  $\sigma$ , remover genes con un perfil de expresión poco informativo. A diferencia de la  $\sigma$ , el CV, al calcularse normalizando con el promedio, permite descartar genes cuya escasa variación relativa en la expresión causa que su patrón de expresión sea poco informativo.

Previo al empleo de estos criterios, se hizo un análisis de las distribuciones de los estadísticos mencionados para los 22810 *probesets* analizados. El propósito fue guiar la elección de los valores umbrales a adoptar. Se analizó la distribución de los promedios de los niveles de expresión de los *probesets* para así reconocer el nivel que podría considerarse “ruido de fondo”. El histograma (Figura IV.4, inciso a) mostró una distribución con dos picos fácilmente distinguibles. Se eligió entonces un valor umbral de expresión de  $A = 5$  que se encontraba alrededor del mínimo entre los picos, y un número de muestras a superar bajo,  $K = 4$ , para definir así el criterio  $KsA$ .

En relación a la desviación estándar, el histograma mostró una distribución asimétrica con un único pico (Figura IV.4, inciso b). Se seleccionó el valor umbral que eliminara el 10 % de las mediciones con menor desviación estándar.

El coeficiente de variación presentó una distribución muy similar en forma a la de la desviación estándar (Figura IV.4, inciso c). Considerando que es el estadístico que da una mejor estimación respecto a la variabilidad de la expresión de los *probesets* en términos relativos, se decidió elegir un valor umbral que eliminara la mitad de los *probesets* con menor coeficiente de variación. Esta remoción, que puede parecer excesiva, responde en parte a la gran demanda computacional que involucra la manipulación posterior de los datos.



**Figura IV.4:** Histogramas ilustrando la distribución de los *probesets* según diferentes parámetros estadísticos respecto a su expresión en la matriz analizada. En las tres gráficas se indica con una línea vertical el umbral empleado para la remoción de los *probesets*.

**Tabla IV.1:** Resumen de los resultados del filtrado.

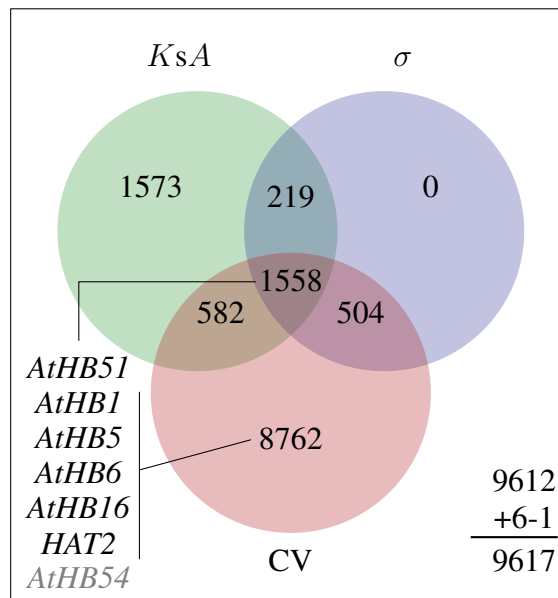
Dataset	Totales	Removidos	Finales
Microarreglos	298	29	269
Genes o <i>probeset</i>	22810	13193	9617

### IV.3. Resultados

Una inspección de los *probeset* correspondientes a los HD-Zip I y II permitió detectar que el *probeset* 264974\_at asociado al FT *AtHB54* de la subfamilia I (*locus* AT1G27045) reconocía el producto del *locus* contiguo (AT1G27050). A causa de esto, dicho *probeset* fue removido del análisis y en consecuencia no fue posible analizar la expresión de este FT. Otro gen de esta subfamilia, *AtHB22*, no pudo ser estudiado por no tener ningún *probeset* asociado en el microarreglo ATH1.

De los *probesets* de los 25 genes HD-Zip I y II restantes, el filtrado removió seis: *AtHB1*, *AtHB5*, *AtHB6*, *AtHB16*, *AtHB51* y *HAT2*. Estos *probeset* fueron manualmente agregados luego del filtrado a la matriz de expresión de modo de obtener sus genes blanco.

Como resultado del filtrado y de la reincorporación de los *probesets* correspondientes a genes HD-Zip I y II se obtuvo un total de 9617 *probesets* para el análisis. En la Tabla IV.1 se resumen estos valores y en la Figura IV.5 se presenta la distribución, según el criterio, de los *probesets* removidos. Se resalta también que *AtHB51* fue removido por los tres criterios, mientras que los cinco *probesets* de HD-Zip restantes sólo fueron eliminados por el criterio menos conservador de los aplicados, el coeficiente de variación.



**Figura IV.5:** Diagrama de Venn que contrasta el número de *probesets* removidos según cada criterio. El número de genes no removido fue 9612, pero finalmente se analizaron 9617 por la reincorporación de los seis *probesets* HD-Zip removidos y la eliminación del *probeset* de *AtHB54*.

### IV.3.3 Obtención de los módulos de actividad transcripcional

Una vez obtenida la matriz de expresión filtrada, se eligieron los parámetros para hacer el cálculo de la red con el programa ARACNE. Teniendo en cuenta que el objetivo era reconocer interacciones transcripcionales directas, se hizo el cálculo sólo de las relaciones de los FTs con el resto de los genes y no se calcularon interacciones entre pares de genes que no fueran FTs. En la Figura IV.3 inciso a) puede verse que esto no afecta la reconstrucción de la red transcripcional.

Como se explicó previamente, ARACNE calcula primero la IM de estos genes contra el resto de los genes (o *probesets*) de la matriz. En un paso siguiente, se eliminan las potenciales interacciones indirectas aplicando el algoritmo de DPI. Como este algoritmo no eliminará las interacciones en tripletes que sólo contengan un FT, no fue necesario hacer el cálculo de los IM entre genes que no eran FTs.

La lista de genes con actividad transcripcional fue obtenida de la base de datos AGRIS (Yilmaz y col., 2011). De los 1851 FTs recuperados, 830 quedaron en la matriz de expresión filtrada.

El cálculo con el programa ARACNE se realizó con 1000 réplicas de *bootstrap* (o muestreo con reemplazo). Además, siguiendo las recomendaciones de Margolin y col. (2006b) y haciendo algunas optimizaciones, se seleccionaron los siguientes parámetros:

- método a emplear para el estimador de núcleo Gaussiano y ajuste del mismo a la matriz utilizada.
- valor  $p$  para considerar un valor de IM estadísticamente diferente de cero.
- tolerancia para el algoritmo DPI, parámetro que evita la remoción de aristas cuando las diferencias entre las menores IMs del triplete no superan un umbral dado.

Una vez calculadas las 1000 réplicas de la red, se obtuvo una red consenso aplicando una rutina provista por los autores de ARACNE (Margolin y col., 2006b). También en este caso fue necesario seleccionar un valor  $p$  que permitiera darle significado estadístico a la frecuencia de aparición de las aristas en el conjunto de las réplicas de la red.

El resultado final fue una lista de los genes blanco putativos directos por cada uno de los 830 FTs analizados. A cada uno de estos conjuntos de genes (o *probesets*) regulados putativamente en forma directa por cada uno de los FTs se lo llamó "MAT" (se incluyó al FT en su propio MAT).

### IV.3.4 Análisis de los módulos de actividad transcripcional

Para obtener algunas características generales sobre la red, se analizó la distribución de los tamaños de los 830 MATs. El histograma de la Figura IV.6 muestra que el número de MATs con una cantidad dada de conexiones se corresponde, a grandes rasgos, con la ley de potencia propuesta para las redes de regulación génica (Barabási, 1999). Puede verse cómo la gran mayoría

### IV.3. Resultados

de los FTs posee un número de conexiones relativamente “bajo”, como lo muestra el gráfico de cajas bajo el histograma.

En relación a los FTs HD-Zip I y II, la distribución de tamaños puede visualizarse en el histograma de la Figura IV.7 y la información detallada se presenta en la Figura IV.8, en la cual se ha agregado la información sobre las relaciones filogenéticas existentes entre los miembros. Para esto se extrajeron las relaciones entre los genes HD-Zip I de *A. thaliana* del árbol HZT (Figura III.4, p. 57, ver Sección III.3.2.3, p. 54), y se combinaron con la reconstrucción realizada por Ciarbelli y col. (2008) con los miembros de la subfamilia II.

En la Tabla IV.2 se muestra un resumen de diferentes parámetros que permiten la comparación del tamaño de los MATs hallados para todos los FTs estudiados y para los subconjuntos de los HD-Zip I y II.

**Tabla IV.2:** Comparación de parámetros de diferentes conjuntos de MATs.

Tamaño del MAT	Conjuntos de MATs			
	HD-Zip I	HD-Zip II	HD-Zip I y II	Total de FTs
Mínimo	70	33	33	8
Mediana	348	179	284	258
Media	465	256	381	392
Máximo	1398	1006	1398	2709

### IV.3.5 Análisis de los promotores en los genes de los MATs

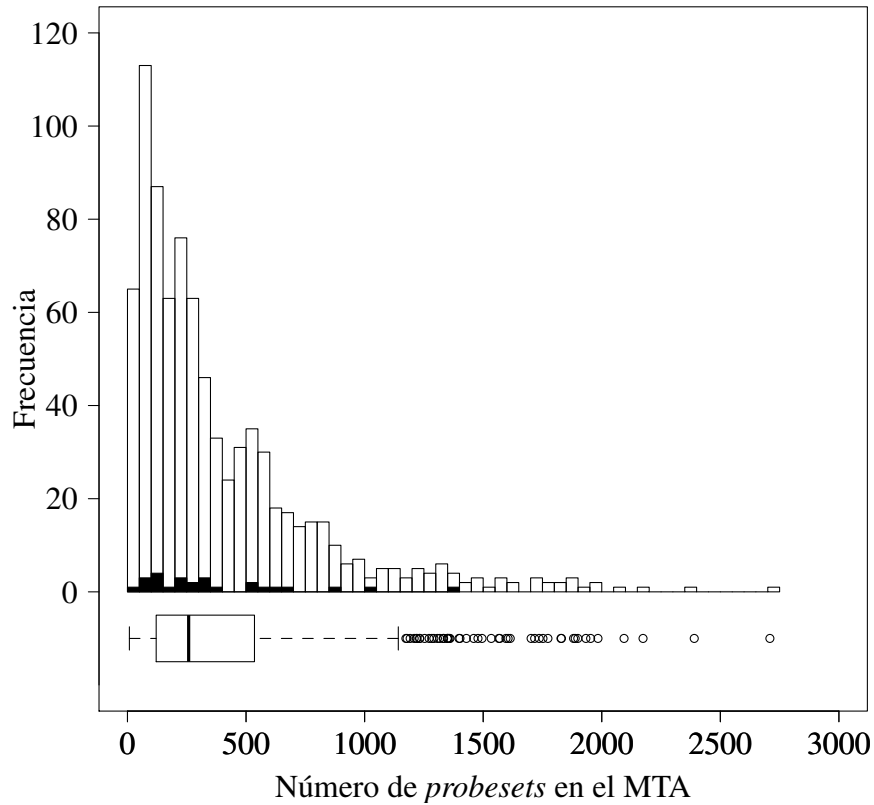
#### IV.3.5.1 Búsqueda del motivo unido con mayor afinidad

Varios trabajos han determinado que las proteínas HD-Zip I y II unen con máxima afinidad una secuencia de ADN muy similar, CAATNATTG (no considerando la preferencia por la base central; Johannesson y col., 2001; Meijer y col., 2000; Palena y col., 1999; Sessa y col., 1993). Considerando que los genes de cada MAT son en principio blancos directos del FTs HD-Zip, se decidió buscar este motivo en los promotores de los genes correspondientes a los *probesets* de cada MAT.

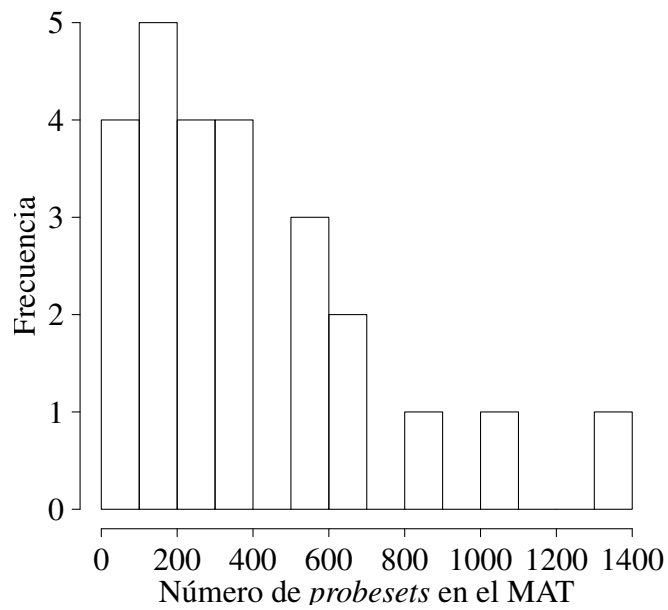
Es importante destacar, sin embargo, que al desconocer la direccionalidad de la interacción entre el gen HD-Zip y el resto de los genes del MAT, existe la posibilidad que aquellos que a su vez son FTs, sean reguladores del HD-Zip y no regulados por éste.

Se definieron entonces dos longitudes de las regiones promotoras a analizar: 1000 y 3000 pb corriente arriba respecto al sitio de inicio de la transcripción. Los promotores fueron obtenidos a partir de la base de datos de secuencias de TAIR (versión 10).

En primer lugar, se hizo una búsqueda global reconociendo este motivo en los promotores de todos los *probesets* incluidos en el análisis. Además, se buscó una secuencia más corta, AATNATT. Los resultados se presentan en la Tabla IV.3. No todos los *probesets* pudieron ser

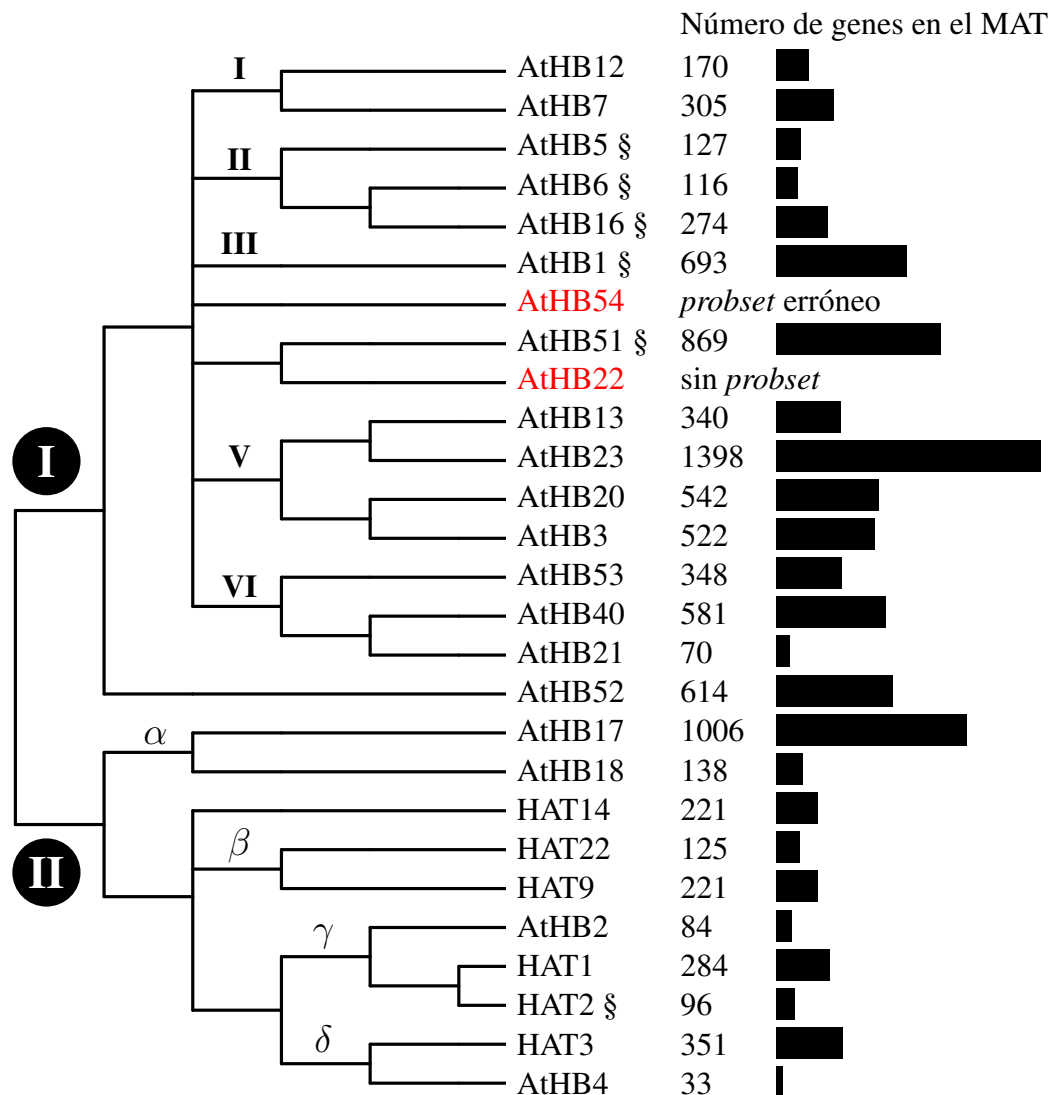


**Figura IV.6:** Histograma que presenta la distribución de tamaños de los MATs de todos los FTs. La porción en negro de las barras corresponde a los MATs de los FTs HD-Zip



**Figura IV.7:** Histograma que presenta la distribución de tamaños de los MATs de los HD-Zip I y II





**Figura IV.8:** Tamaño de los MATs obtenidos. Se agregaron barras proporcionales al tamaño de los MATs para facilitar la comparación entre genes. Se presenta un árbol filogenético con las relaciones entre los FTs según lo presentado en la Sección III.3.2.3 (p. 54, HD-Zip I) y el trabajo de Ciarbelli y col. (2008, HD-Zip II).

asociados a un gen y por tanto a un promotor, y en algunos casos dos *probesets* se encontraban asociados al mismo gen. Por esto, del total de los *probesets* del análisis se realizó la búsqueda sobre 9050 promotores. Los porcentajes de presencia del motivo CAATNATTG en los promotores de 1000 y 3000 pb (o 1 y 3 kpb) no fueron muy elevados (4,21 y 9,96 %, respectivamente). Una búsqueda con un motivo acotado a las siete bases centrales del pseudopalíndromo, AATNATT, indicó que el mismo es prácticamente ubicuo, especialmente si se considera una región promotora de 3000 pb. A pesar de esto, se realizó la búsqueda con ambos motivos.

La búsqueda realizada sobre los promotores de los MATs produjo los resultados presentados en la Tabla IV.4. Se aprecia que, en términos generales, los porcentajes no son muy diferentes a los que se encuentran en el conjunto total de datos, según el caso. Con lo cual, puede decirse

### IV.3. Resultados

que no hay un enriquecimiento significativo de los motivos unidos con máxima afinidad en los promotores de los genes de los MATs.

**Tabla IV.3:** Resumen de la búsqueda de motivos en los promotores del conjunto de genes completo. Se los conjuntos de secuencias de 1000 y 3000 pb corriente arriba obtenidos de TAIR (versión 10).

<i>Probesets</i>	Número		Porcentaje	
	1 kbp	3 kbp	1 kbp	3 kbp
Total	9617			
Con anotación TAIR	9050		100,00	
Con CAATNATTG	381	901	4,21	9,96
Con AATNATT	6398	8421	70,70	93,05

**Tabla IV.4:** Búsqueda del motivo unido con mayor afinidad por los HD-Zip I y II y de un motivo relacionado en los promotores de los genes de los MATs.

HD-Zip	Genes en el MAT			Motivo CAATNATTG				Motivo AATNATT			
	Nro Total	Con promotor Nro	Porcent.	Número 1kbp	3kbp	Porcentaje 1kbp	3kbp	Número 1kbp	3kbp	Porcentaje 1kbp	3kbp
Subfamilia I											
AtHB12	170	162	95	7	16	4,3	9,9	122	152	75	94
AtHB7	305	296	97	12	33	4,1	11,1	220	279	74	94
AtHB5	127	121	95	5	13	4,1	10,7	86	116	71	96
AtHB6	116	113	97	7	12	6,2	10,6	88	109	78	96
AtHB16	274	259	95	13	20	5,0	7,7	187	243	72	94
AtHB1	693	676	98	35	75	5,2	11,1	477	628	71	93
AtHB51	869	789	91	30	70	3,8	8,9	536	723	68	92
AtHB13	340	324	95	14	41	4,3	12,7	241	307	74	95
AtHB23	1398	1316	94	52	135	4,0	10,3	875	1207	66	92
AtHB20	542	496	92	26	53	5,2	10,7	366	469	74	95
AtHB3	522	470	90	20	46	4,3	9,8	365	441	78	94
AtHB53	348	328	94	17	32	5,2	9,8	238	308	73	94
AtHB40	581	532	92	27	52	5,1	9,8	396	496	74	93
AtHB21	70	65	93	2	5	3,1	7,7	52	62	80	95
AtHB52	614	542	88	18	47	3,3	8,7	376	506	69	93
Subfamilia II											
AtHB17	1006	927	92	48	114	5,2	12,3	657	846	71	91
AtHB18	138	128	93	9	19	7,0	14,8	95	119	74	93
HAT14	221	208	94	14	29	6,7	13,9	158	197	76	95
HAT22	125	123	98	5	14	4,1	11,4	92	117	75	95
HAT9	221	213	96	16	29	7,5	13,6	157	200	74	94
AtHB2	84	81	96	4	8	4,9	9,9	57	75	70	93
HAT1	284	270	95	10	27	3,7	10,0	203	256	75	95
HAT2	96	89	93	2	6	2,2	6,7	64	82	72	92
HAT3	351	339	97	10	34	2,9	10,0	241	315	71	93
AtHB4	33	28	85	0	0	0,0	0,0	20	26	71	93

### IV.3.5.2 Búsqueda de motivos *de novo*

Como alternativa a la búsqueda del motivo unido por los FTs HD-Zip, se hizo una búsqueda de motivos *de novo*. Esto implica reconocer motivos estadísticamente enriquecidos en los promotores sin aportar información sobre la secuencia particular que se desea buscar. Sin embargo, se impusieron ciertas condiciones al programa empleado, COSMO (Bembom y col., 2007), con el objetivo de evaluar si era posible encontrar un motivo pseudopalindrómico degenerado similar a CAATNATTG. La restricción forzó la búsqueda de pseudopalíndromos con una longitud mínima de nueve nucleótidos y con una región central no específica de uno o más nucleótidos.

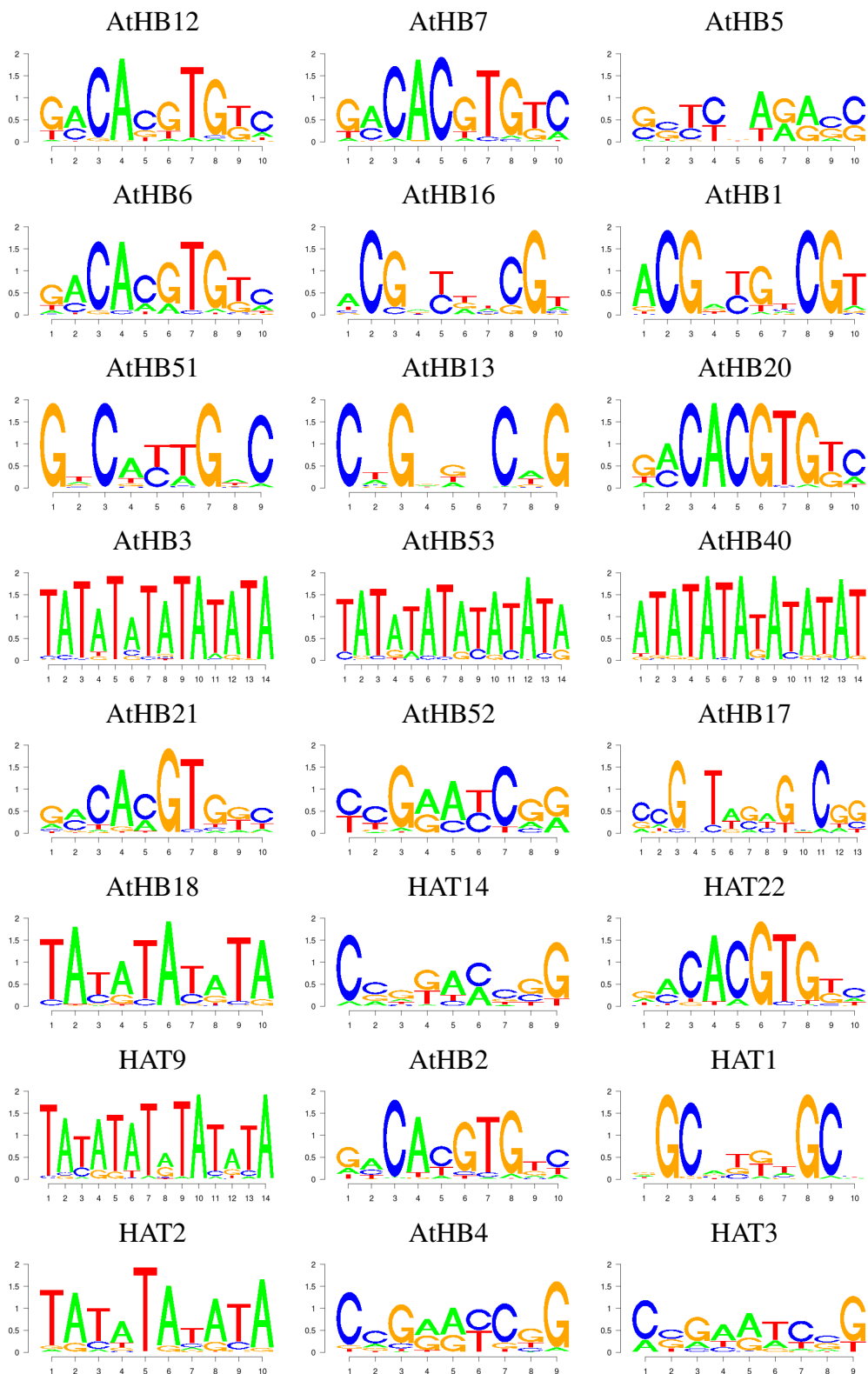
El algoritmo se aplicó únicamente en el conjunto de promotores de 1000 pb empleado en el análisis del punto anterior, ya que su aplicación a los promotores de 3000 pb requiere una capacidad de cómputo mayor a la que se contaba. De hecho, aun con los promotores de 1000 pb, la limitación interna del programa ocasionó que utilizara sólo 50 secuencias de cada conjunto de datos.

Los resultados se presentan en forma de *sequence logos* en la Figura IV.9. En los MATs de *AtHB3*, *AtHB53*, *AtHB40*, *HAT9* y *HAT2* puede verse que se obtuvo un motivo compuesto por una repetición de As y Ts, bases muy frecuentes en los promotores y que por el alto grado de repetición posiblemente carezca de especificidad (Vinces y col., 2009). Del resto de los casos debe resaltarse la presencia del motivo CACGTG en los promotores de los MATs de *AtHB12*, *AtHB7*, *AtHB6*, *AtHB20*, *AtHB21*, *HAT22* y *AtHB2*. Esta secuencia es muy similar al motivo ABRE (ACGTGC) presente en promotores de genes de respuesta a ácido abscísico (ABA Choi y col., 2000).

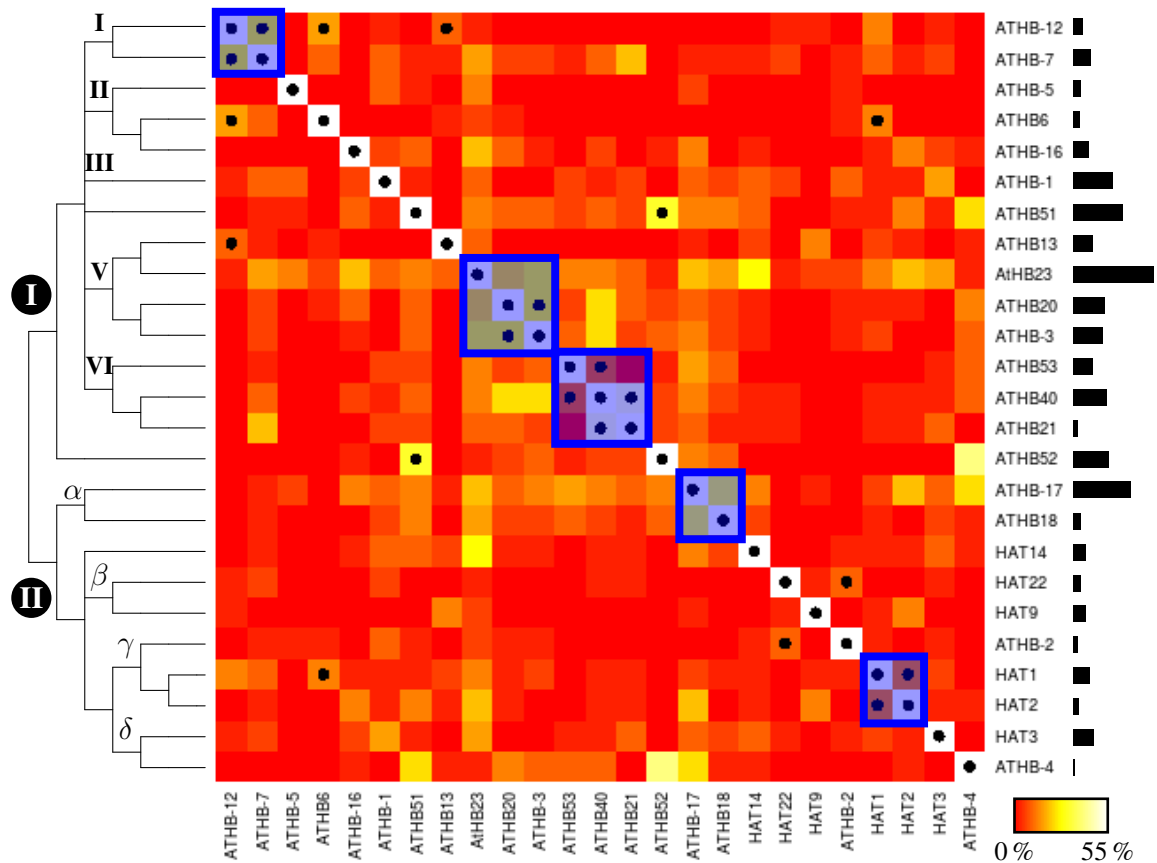
### IV.3.6 Solapamiento entre los MATs de diferentes FTs HD-Zip

Con el propósito de conocer el potencial nivel de redundancia funcional entre genes, se investigó también el grado de solapamiento entre los MATs obtenidos para los genes HD-Zip, evaluándolo especialmente desde una perspectiva filogenética.

Para esto se hicieron todas las comparaciones de a pares posibles, y se consideró el valor porcentual de genes solapados sobre el menor de los MATs del par. El resultado se presenta en la Figura IV.10. Además, sobre la figura se resaltan algunas características importantes. Por un lado, se señalan los casos en los que los MATs del par contienen mutuamente a los mismos HD-Zip comparados, lo que sería un indicador de que uno regula al otro. Y por el otro, se destacan cuatro grupos de genes con estrecha relación filogenética, que presentan un elevado grado de solapamiento entre sí, posible indicador de redundancia funcional.



**Figura IV.9:** Búsqueda de motivos *de novo*. Los motivos fueron hallados empleando el programa COSMO en los conjuntos de promotores de cada MAT. El resultado se muestra en forma de *sequence logos*.



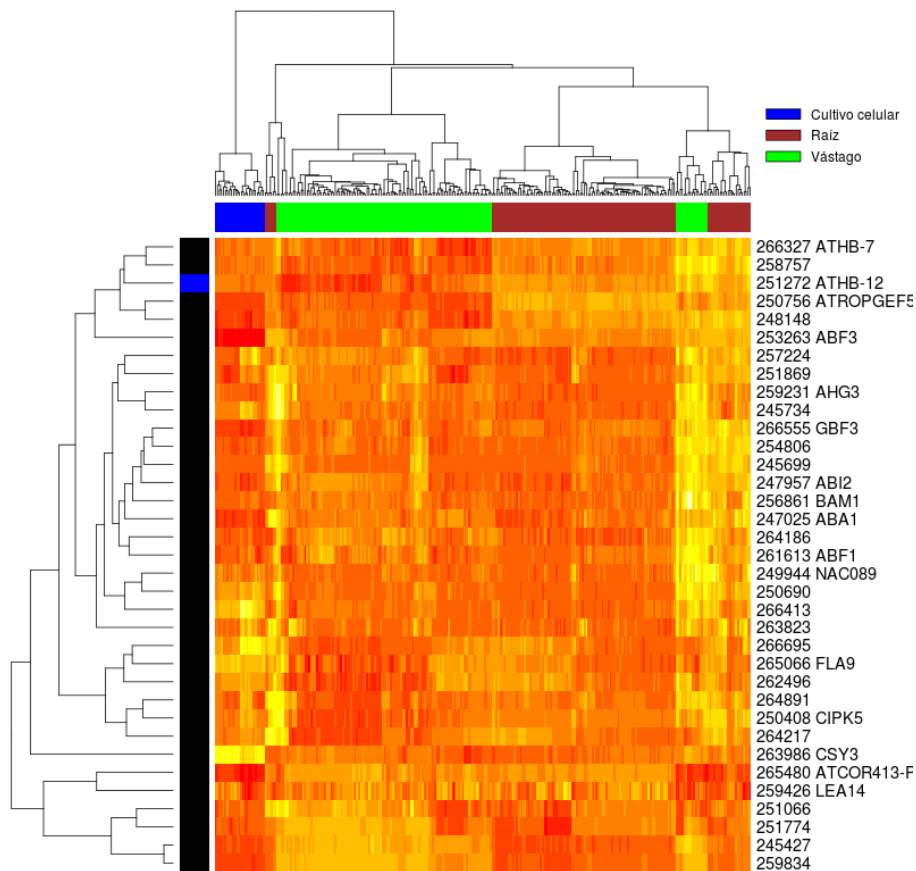
**Figura IV.10:** Solapamiento entre los MATs de diferentes FTs HD-Zip. Se representa mediante un gráfico tipo mapa de calor, “saturado” a un valor de 55 %. Los puntos dentro de los cuadrados indican que los MAT comparados contienen a su vez a ambos genes HD-Zip. Los rectángulos azules destacan el elevado solapamiento existente entre algunos MAT de genes filogenéticamente relacionados. Las barras negras a la derecha del gráfico representan el tamaño relativo de los MATs. La discrepancia en la nomenclatura de algunos genes HD-Zip con las empleada en el resto del manuscrito se debe a que dentro del entorno R se utilizaron los nombres que figuran en las bases de datos de TAIR.

### IV.3.7 Estudios de expresión coordinada dentro de los MATs

Los MATs fueron obtenidos empleando el conjunto completo de los datos, sin distinción por tratamientos ni tejidos. Por esto, su conformación no provee ninguna información en cuanto a la respuesta de los *probesets* del MAT, y por lo tanto del HD-Zip, en las diferentes condiciones.

Para abordar este punto, se comenzó analizando en forma global los perfiles de expresiones correspondientes a cada MAT. Para esto se generaron mapas de calor en los que se visualizaran los perfiles de expresión normalizados de todos los *probesets* del MAT en cada tratamiento y tejido. A modo de ejemplo, se presentan los resultados del MAT de AtHB12 (Figura IV.11), del MAT de AtHB1 (Figura IV.12) y del MAT de HAT2 (Figura IV.13)

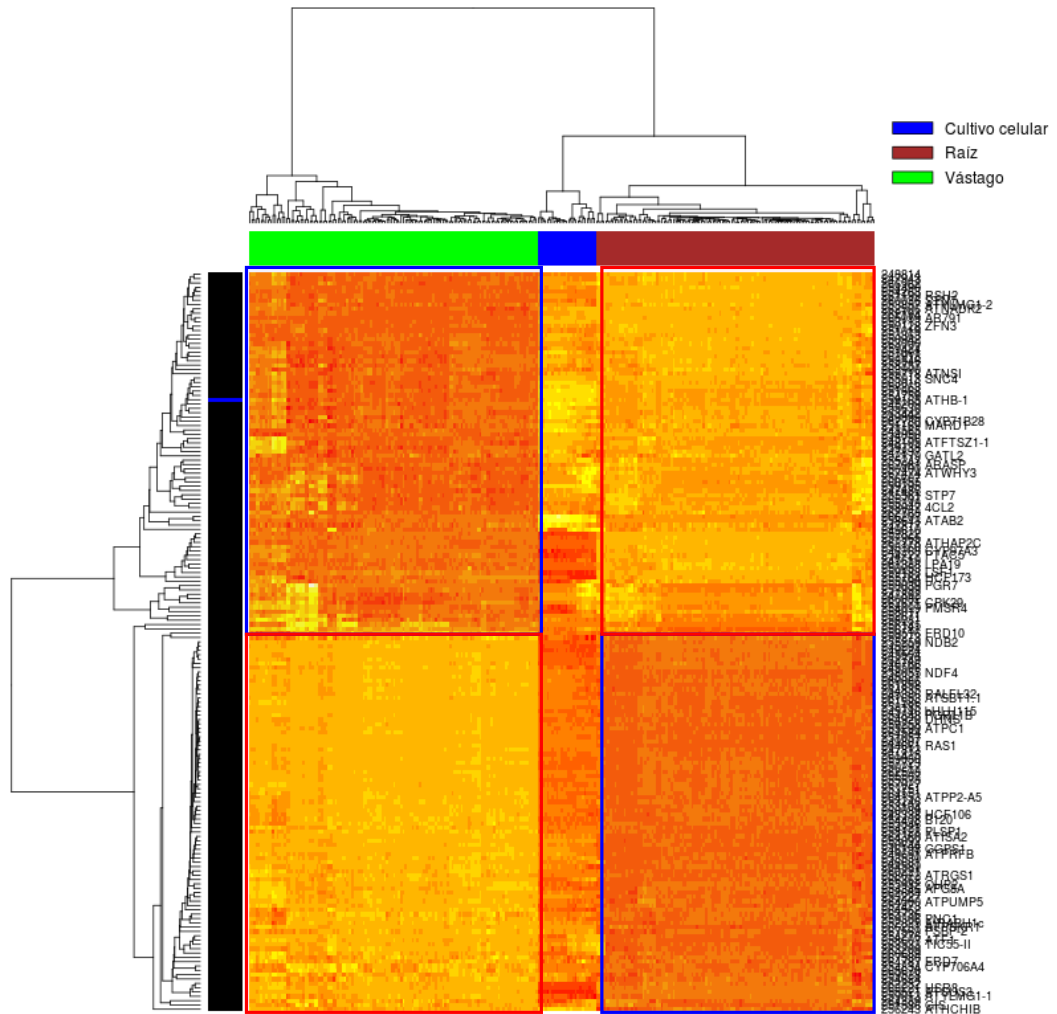
Una de las principales observaciones que este análisis permitió realizar fue que, mientras algunos MATs no presentaron patrones evidentes en los perfiles de expresión (por ejemplo,



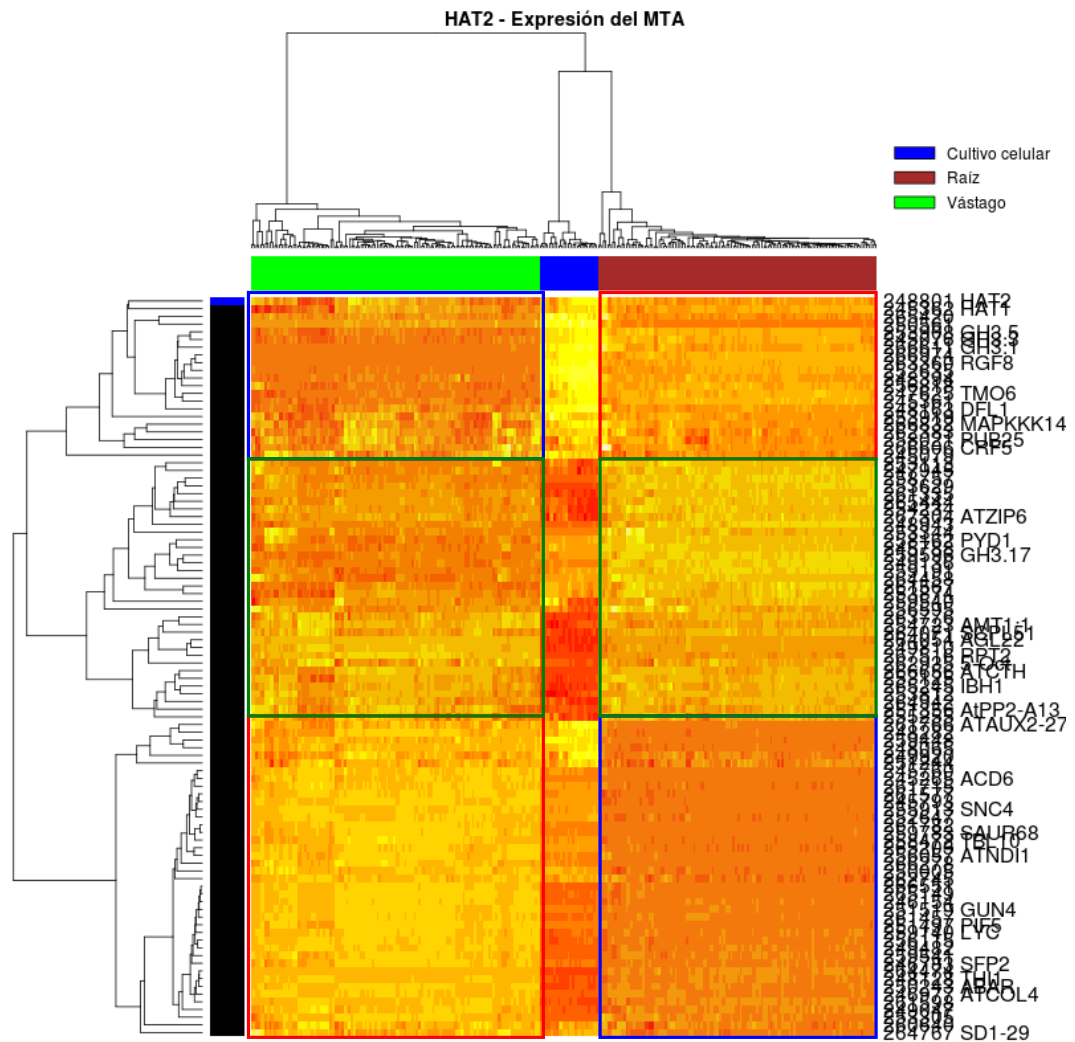
**Figura IV.11:** Gráfico tipo mapa de calor representando la expresión de los *probesets* del MAT de *AtHB12*. Los valores de expresión mayores se presentan en tonos de amarillo y los menores en tonos de rojo. La barra negra a la izquierda es interrumpida por una línea azul que indica el perfil del HD-Zip al que pertenece el MAT. La barra superior indica el tejido utilizado para el ensayo de microarreglo de esa columna. Por razones de presentación, la partícula “\_at” al final del identificador de Affimetrix fue omitida.

*AtHB12*), otros mostraron *probeset* con perfiles de expresión notoriamente diferentes entre vástago y raíz. Un ejemplo claro lo constituyen la expresión de los *probesets* en el MAT de *AtHB1*. De este modo se puso en evidencia, por un lado, que los niveles de expresión entre tejidos son significativamente diferentes en algunos MATs. Concretamente, se observa que los niveles medios de expresión son diferentes entre raíz y vástago para muchos *probesets*.

Otra observación relevante en los casos en que se identificaron patrones en los perfiles de expresión, es que además de haber diferencias notorias entre tejidos, hubo una partición del MAT en subgrupos de *probesets*. Un subgrupo presentó mayor expresión en un tejido respecto al otro, y ocurrió lo inverso para otro subgrupo de *probesets*. Cabe aclarar que no en todos los MAT en los que se visualizaron patrones diferenciales en los perfiles de expresión por tejido esta división en dos grupos fue tan clara. Esto puede verse en el caso del MAT de HAT2 (Figura IV.13).



**Figura IV.12:** Gráfico tipo mapa de calor representando la expresión de los *probesets* del MAT de AtHB1. Los valores de expresión mayores se presentan en tonos de amarillo y los menores en tonos de rojo. Los rectángulos superpuestos en el mapa indican regiones de expresión media mayor o menor. La barra negra a la izquierda es interrumpida por una línea azul que señala el perfil del HD-Zip al que pertenece el MAT. La barra superior indica el tejido utilizado para el ensayo de microarreglo de esa columna. Por razones de presentación, la partícula “\_at” al final del identificador de Affimetrix fue omitida.



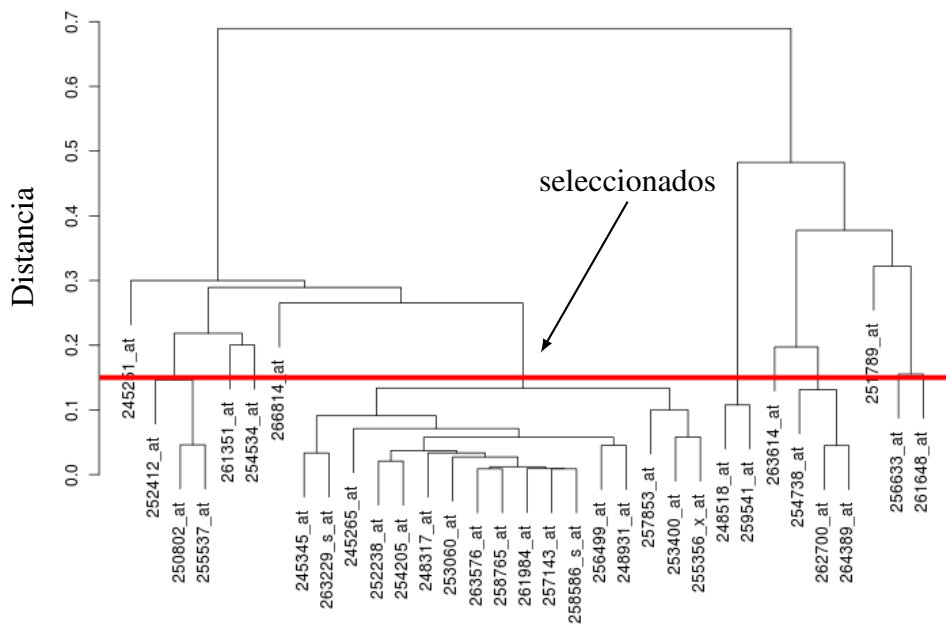
**Figura IV.13:** Gráfico tipo mapa de calor representando la expresión de los *probesets* del MAT de HAT2. Los valores de expresión mayores se presentan en tonos de amarillo y los menores en tonos de rojo. Los rectángulos superpuestos en el mapa indican regiones de expresión media mayor o menor. La barra negra a la izquierda es interrumpida por una línea azul que indica el perfil del HD-Zip al que pertenece el MAT. La barra superior indica el tejido utilizado para el ensayo de microarreglo de esa columna. Por razones de presentación, la partícula “\_at” al final del identificador de Affimetrix fue omitida.



### IV.3.8 Análisis de la participación de los genes HD-Zip en diferentes estreses

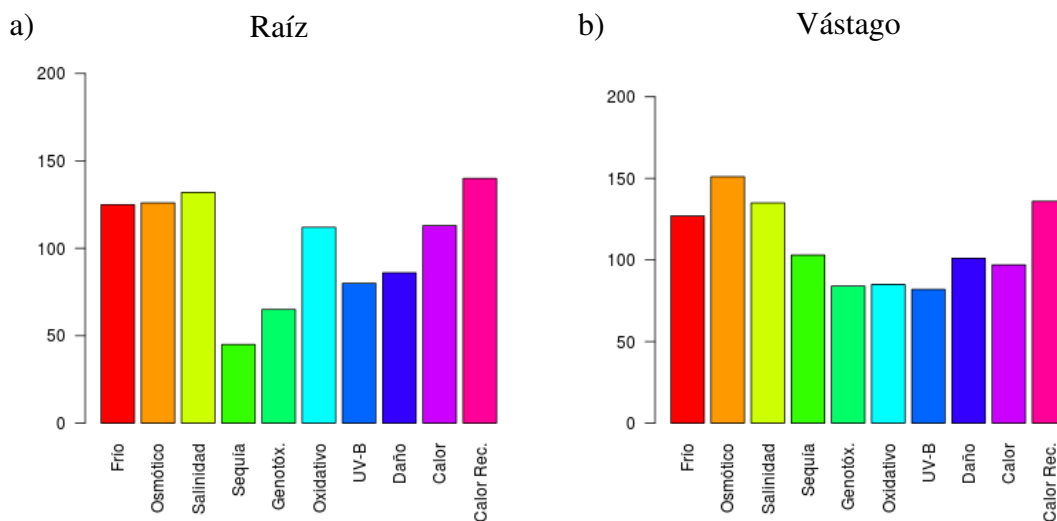
Para poder evaluar el grado de participación que tiene cada uno de los HD-Zip estudiados en las diversas condiciones de estrés abiótico, se adoptó como indicador la medición del número de *probesets* del MAT con expresión coordinada en cada tratamiento.

Para la determinación de los *probesets* con expresión coordinada se utilizó el siguiente criterio. Se cuantificaron las similitudes entre los perfiles de expresión de todos los *probeset* en función del grado de correlación entre dichos perfiles. Esto se convirtió a una medida de distancia y se empleó un algoritmo de agrupamiento de los *probesets* para construir un dendrograma. Se escogió entonces un umbral de distancia bajo el cual los clados de *probesets* con un número mayor a cinco integrantes fueron considerados grupos de genes regulados coordinadamente. Por último, esto se realizó en cada condición y tejido, y el recuento total de los genes coordinados fue tomado como indicador de la participación de un FT dado en la respuesta. Este procedimiento se ilustra en la Figura IV.14, y el resultado en un MAT está ejemplificado en la Figura IV.15.



**Figura IV.14:** Ejemplo de la determinación de *probesets* con expresión coordinada. Corresponde al análisis de los *probesets* coexpresados en el MAT de *AtHB4* en el vástago de plántulas sometidas a estrés por calor con recuperación. La línea roja muestra el umbral de 0,15. El requisito de que al menos cinco *probesets* conformen un grupo hace que el único grupo seleccionado sea el que aparece en la posición central.

Una vez analizados todos los MATs por tratamiento y tejido, se agrupó la información para evaluar características globales de la respuesta de los FTs de ambas subfamilias HD-Zip. La Figura IV.16 es el resultado de trasladar cada gráfico de barras a un gráfico tipo mapa de calor,



**Figura IV.15:** Gráfico de barras que ejemplifica el análisis en un MAT, en este caso el de AtHB12. La altura de las barras corresponde al número de genes con expresión coordinada en cada estrés.

en el que el color representa el número de *probesets* con expresión coordinada. Para que la comparación entre MATs de tamaños diferentes fuera posible, se normalizaron los valores. Se reordenaron tanto las condiciones como los HD-Zip para así encontrar patrones de similitud.

Puede visualizarse que en el vástago (tejido aéreo) los tratamientos por calor, con y sin tiempo de recuperación, son los que generaron el mayor grado de coordinación en la expresión de *probesets* de los diferentes MATs. Respecto al resto de las condiciones puede destacarse la salinidad, el frío y posiblemente el estrés osmótico como otros tratamientos relevantes. Los que menor coordinación generaron fueron los tratamientos de déficit hídrico y estrés genotóxico.

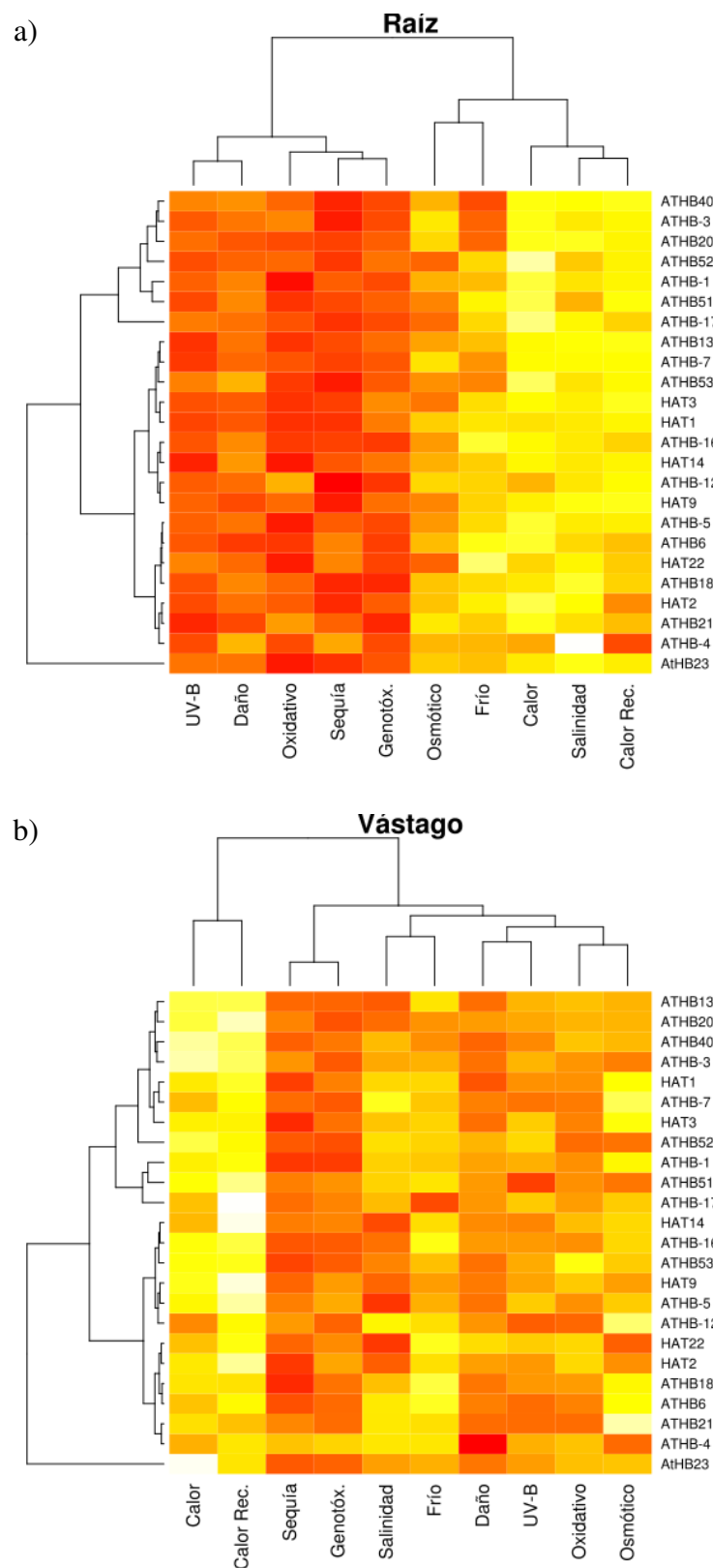
En relación al tejido de raíz, se vio en forma muy notoria que los tratamientos que más coordinación en la expresión generaron fueron: estrés osmótico, frío, calor, calor con recuperación y salinidad.

Si se inspecciona el dendrograma que relaciona las filas podrá notarse que no se verificó en ninguno de los tejidos, mediante este tipo de análisis, mayor similitud entre los perfiles de coexpresión de los MATs de los HD-Zip de una misma subfamilia, respecto a los de la otra.

#### IV.3.8.1 Subgrupos de *probesets* corregulados

El análisis de los *probesets* corregulados por estrés y tejido también fue empleado para investigar la evolución temporal, a lo largo de los tratamientos, de la expresión de los genes controlados por cada HD-Zip. De esta forma se obtuvieron perfiles de expresión de *probesets* que previamente habían sido seleccionados como corregulados por el criterio presentado en la sección anterior.

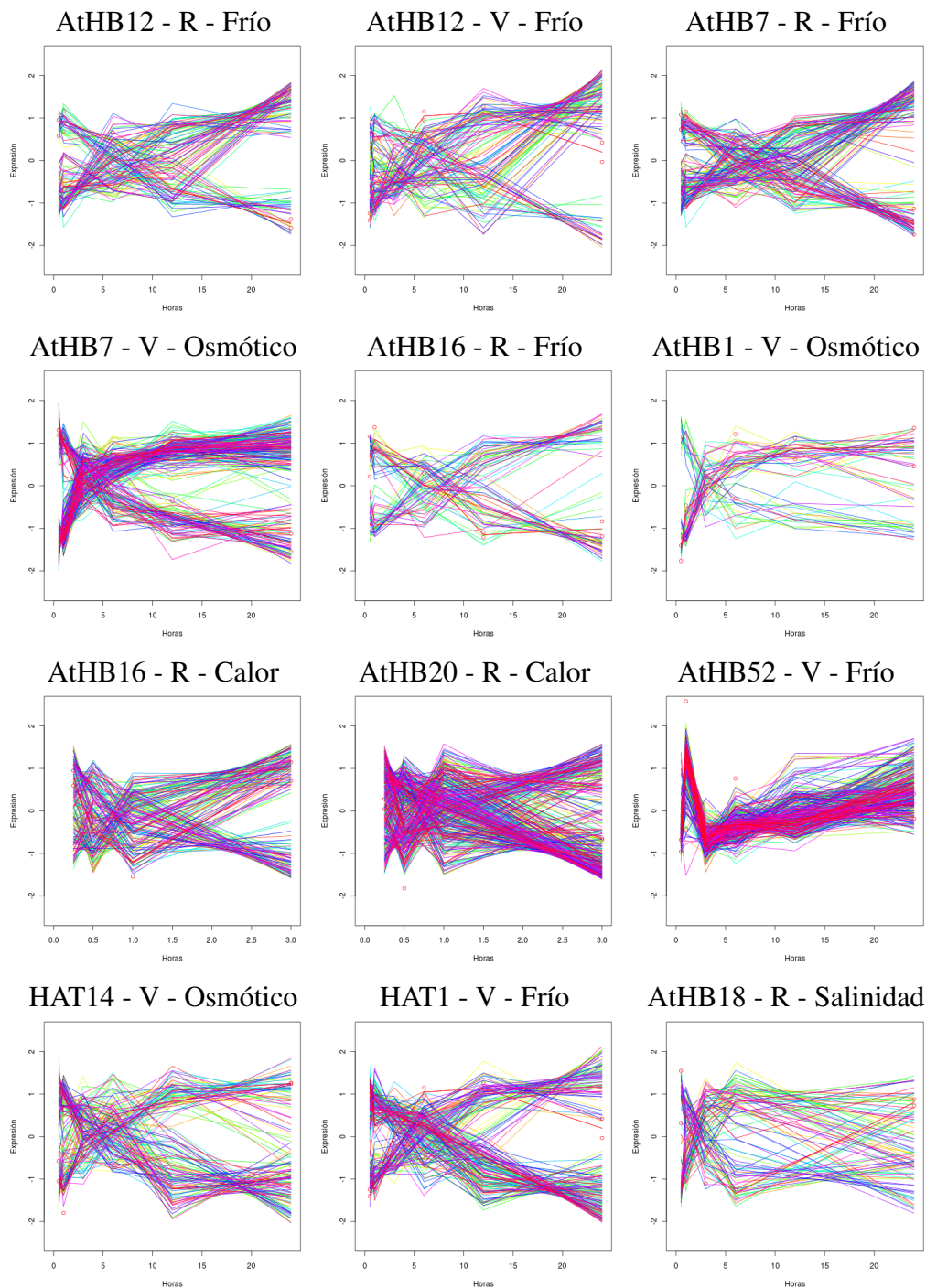
Es así como pudo observarse que en muchos casos los genes de un MAT corregulados en



**Figura IV.16:** Gráfico tipo mapa de calor resumiendo los resultados del análisis de genes coregulados por tejido según la condición. Las tonalidades de amarillo indican mayor número de genes coexpresados. Las columnas y filas fueron reordenadas en función de su similitud.

### IV.3. Resultados

un tratamiento tienen una tendencia, con el transcurso del estrés, a ser inducidos o reprimidos. Dentro de cada grupo pudieron reconocerse conjuntos de *probesets* con perfiles diferentes, sin embargo, es importante resaltar el efecto diferencial neto sobre la expresión.



**Figura IV.17:** Perfiles de expresión de los *probesets* corregulados separados por MAT y tratamiento. Sobre cada perfil se indica el MAT, el tratamiento y el tejido (V: vástago R: raíces) considerados.

### IV.3.9 Análisis del enriquecimiento de términos GO

Por último, se realizó un análisis funcional de los MATs con el propósito de caracterizar las vías de señales en las que participarían los FTs HD-Zip. Esto fue posible gracias al uso de términos GO (ontología de genes o, del inglés, *Gene Ontology*). Éstos consisten en rótulos o etiquetas que se asocian a cada gen y que representan tres posibles características funcionales del propio gen o de su producto (Ontology, 2010):

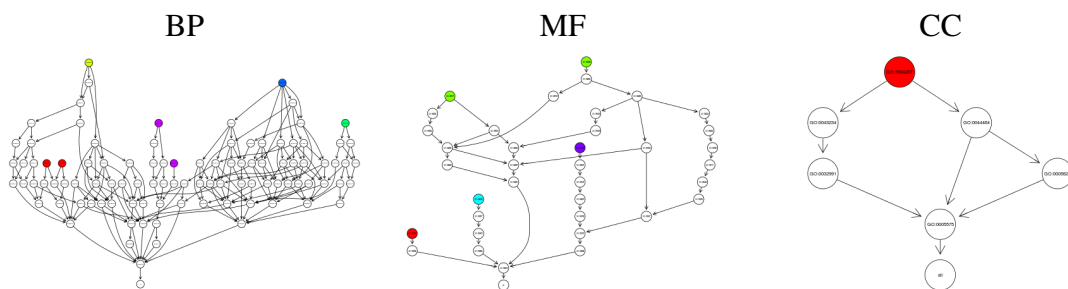
**MF** Función molecular en la que participa.

**BP** Proceso biológico en el que está involucrado.

**CC** Componente celular del que forma parte o donde se localiza.

Este análisis implicó entonces hallar el conjunto de los términos GO asociados a cada gen y evaluar posteriormente qué términos se hallaban estadísticamente enriquecidos en cada MAT. En forma adicional se agruparon los términos relacionados según un algoritmo basado en la teoría de la información para resumir los resultados y facilitar su interpretación (Berenstein y Chernomoretz, 2012, manuscrito en preparación).

En la Tabla IV.5 y en la Figura IV.18 se muestra como ejemplo el resultado obtenido con el MAT de AtHB12. Puede verse que se encontraron enriquecidas las categorías de respuesta al ABA y al estrés abiótico, más particularmente al déficit hídrico y al frío. También resultaron seleccionadas categorías relacionadas a la regulación de la transcripción y a la regulación de la fosforilación de proteínas.



**Figura IV.18:** Gráficos de las relaciones de los términos GO halladas en el MAT de AtHB12. Los nodos coloreados simbolizan términos en los que se hallaron genes. Los nodos de un mismo color fueron agrupados bajo un rótulo de mayor jerarquía que resumiera su función. Los nodos no coloreados muestran los rótulos de mayor jerarquía hasta alcanzar la raíz de la estructura de términos GO. Se han empleado los nombres originales de los términos GO en idioma inglés.

### IV.3. Resultados

**Tabla IV.5:** Términos GO hallados en el MAT de AtHB12. Están separados por filas que indican a la categoría que pertenecen (BP, MF o CC). Sobre los nombres de los términos (GO:XXXXXXX) se muestra un rótulo que resume su función.

ATHB-12 MICA/GOID	Término GO	Genes
<b>BP</b>		
	response to abiotic stimulus	
GO:0009414	response to water deprivation	LEA14 NF-YA5 ATCOR413-PM1 ATHB6 ATHB-7 AT3G02480 AHG3 ATNCED3 BAM1 ATHB-12 ANAC072 ABF3 COR78 ABI2 HAI1 ATMYB96 AAP4 ABA1
GO:0009409	response to cold	ATCOR413-PM1 AHG3 ATICE1 ABI1 LEA4-5 COR27 COR78
	regulation of abscisic acid mediated signaling pathway	
GO:0009788	negative regulation of abscisic acid mediated signaling pathway	ATHB6 AHG3 ABI1 ABI2
	vitamin E metabolic process	
GO:0010189	vitamin E biosynthetic process	HPD G-TMT AT5G53970
	positive regulation of gene expression	
GO:0045941	positive regulation of transcription	ATHAP5B ATHB6 ATICE1 ATHB-12
	transport	
GO:0006857	oligopeptide transport	YSL8 AT2G02020 AT2G37900 AT3G21670
GO:0006865	amino acid transport	AAP1 AT2G41190 AAP4 AT5G65990
<b>MF</b>		
	transcription regulator activity	
GO:0016563	transcription activator activity	ABF1 ATHAP5B ATHB6 ATHB-7 ATICE1 ATHB-12 ANAC072 ABF3
	active transmembrane transporter activity	
GO:0015171	amino acid transmembrane transporter activity	AAP1 AT2G41190 AAP4 AT5G65990
GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	AT1G72700 AT3G27870 ACA8
	DNA binding	
GO:0043565	sequence-specific DNA binding	ATHB13 ATHB6 GBF3 ATMBD11
	phosphoprotein phosphatase activity	
GO:0004722	protein serine/threonine phosphatase activity	HAI2 AHG3 ABI1 ABI2 HAI1
<b>CC</b>		
	protein complex	
GO:0008287	protein serine/threonine phosphatase complex	HAI2 ABI1 ABI2

### IV.4 Discusión de resultados

Las redes de regulación génica son sistemas extremadamente complejos que recién en las últimas dos décadas han comenzado a ser comprendidos desde una perspectiva global. Su estudio se hizo posible gracias a los avances tecnológicos, pero también gracias a los desarrollos en las áreas de la bioinformática y la bioestadística.

Los FTs son actores esenciales en estas redes de regulación génica. El gran número y diversidad de interacciones que son capaces de establecer, ya sea con los promotores de sus genes blanco, con proteínas de la maquinaria transcripcional o con las proteínas que los regulan; hacen que no sea posible capturar la complejidad de su función si no se suman a las metodologías convencionales las herramientas de análisis a nivel de sistemas. Con esta concepción fue que se abordó la caracterización de las vías de señales en las que participan los FTs HD-Zip I y II.

#### IV.4.1 Red de regulación transcripcional reconstruida con ARACNE

A pesar de que el estudio estuvo centrado en los genes HD-Zip I y II, en una primera etapa se hizo la reconstrucción de la red completa. Para esto se emplearon ensayos de microarreglos de bases de datos públicas obtenidos con plantas sometidas a diferentes condiciones de estrés abiótico. De estos datos se removieron los microarreglos cuyos duplicados fueron muy variables y los genes con bajo contenido de información, es decir, genes cuyo rango de expresión era demasiado acotado para reconocer niveles frecuentes y niveles poco frecuentes de expresión. A partir de la matriz resultante, el programa ARACNE generó los MATs de los 830 FTs evaluados.

Inicialmente se evaluó el tamaño de los MATs obtenidos, es decir, el número de genes regulado por cada FT de la red. Su distribución siguió aproximadamente una ley de potencia, esperable para una red de regulación génica (Barabási, 1999). El mayor número de genes blanco fue de alrededor de 2709, un valor superior a los hallados por otros autores en diferentes reconstrucciones de redes de regulación. Por ejemplo, utilizando correlación de Pearson, Mutwil y col. (2010) encontraron que los FTs que controlaban más genes regulaban alrededor de 900 blancos. En el caso del trabajo de Mao y col. (2009) también basado en mediciones de correlación de Person, los FTs con mayor número de genes regulados rondaron los 1200. Needham y col. (2009) hallaron un valor similar, 1254 genes regulados por un único FT, empleando redes Bayesianas sobre datos discretizados. Debido a que estos resultados fueron obtenidos utilizando métodos basados en correlaciones lineales (Pearson) o datos discretizados, no deberían lograr recuperar todas las interacciones, especialmente las no lineales. Por esto se prosiguió con el análisis a pesar de las diferencias observadas.

La distribución de los genes regulados por los FTs HD-Zip, así como el total de los FTs, se comportó aproximadamente como una curva del tipo de ley de potencia. En términos generales, los MATs de la subfamilia I fueron mayores a los de la subfamilia II. Considerando que los datos

fueron obtenidos de microarreglos de estrés abiótico, esto podría estar indicando que los genes HD-Zip I tienen mayor participación en estos estreses que los de la subfamilia II.

##### IV.4.2 Elementos en *cis* en los promotores de los genes de los MATs

Teniendo en cuenta que los genes de los MATs son blancos directos putativos de los FTs HD-Zip, sería esperable que posean el elemento en *cis* unido por estas proteínas en sus promotores. La búsqueda del motivo CAATNATTG en estas secuencias, considerando 1000 o 3000 pb corriente arriba del sitio de inicio de la transcripción, no mostró un enriquecimiento respecto a la abundancia de este motivo en todo el conjunto de secuencias.

Se consideró también el motivo central de este pseudopalíndromo, AATNATT, ya que se ha demostrado que son las bases cuya mutación tiene un mayor efecto en la capacidad del FT de reconocer el ADN, tanto *in vitro* (Palena y col., 1999) como *in vivo* (Aoyama y col., 1995; Henriksson y col., 2005). Este motivo fue hallado en alrededor de un 70 % de los promotores de 1000 pb de los genes del conjunto de datos completo, y este porcentaje no cambió en los promotores de los genes de los MATs. Con los promotores de 3000 pb el porcentaje fue del 93 % para todos los genes analizados, y tampoco varió al buscarlo en los promotores de los genes de los MATs.

Al respecto, en el trabajo presentado por Overnäs (2010) se demostró, mediante ensayos de coimmunoprecipitación de la cromatina (ChIP), la capacidad de AtHB7 y AtHB12 de reconocer diferentes promotores de genes de la vía de señales de la fitohormona ABA. De los nueve promotores unidos por ambas proteínas, sólo dos poseían el motivo CAATNATTG en los 1500 pb evaluados. El resto de los promotores contenía uno o más motivos con el núcleo AATNATT, pero con diferencias en una o en ambas posiciones extremas, las primeras de la secuencia unida por cada HD del dímero.

En relación a esta observación es importante resaltar que Himmelbach y col. (2002) encontraron el motivo CAATTATTA en el promotor del propio gen HD-Zip I *AtHB6*. En ensayos de retardo en geles y de expresión transitoria en protoplastos de *A. thaliana* demostraron la capacidad del FT de unir este motivo, pero no el motivo CAATTGTTA. En un trabajo sobre la proteína CpHB7 de *Craterostigma plantagineum*, se halló que reconocía el motivo GAATTATTA en el promotor del gen LEA CDeT6-19. Esto fue demostrado por ensayos de retardo en geles, experimentos de simple híbrido en levaduras y ensayos transitorios en protoplastos de tabaco. Esta secuencia comparte solo las 6 bases centrales (y la posición central degenerada) con respecto al motivo CAATNATTG (Deng y col., 2006, 2002).

La capacidad de los FTs HD-Zip de reconocer secuencias ligeramente diferentes al motivo unido con máxima afinidad *in vitro* indican que su presencia no sería requisito excluyente para el reconocimiento de un promotor dado; sí posiblemente la presencia del motivo AATNATT. La evidencia presentada respecto a la falta de enriquecimiento del motivo CAATNATTG en los



#### IV.4. Discusión de resultados

---

MATs y la gran abundancia del núcleo AATNATT en todos los promotores, más si se consideran 3000 pb corriente arriba, sugiere la posibilidad de que estos FTs no tengan una muy elevada selectividad respecto a los genes cuyos promotores son capaces de unir. Debe considerarse la posibilidad de que haya otros factores desconocidos afectando la especificidad de la secuencia reconocida.

Diferentes trabajos mostraron que hay FTs que unen un elevado número de promotores. *HY5* es un gen de *A. thaliana* fundamental en el desarrollo fotomorfogénico. Se ha demostrado por ensayos de ChIP seguidos de microarreglos tipo *tiling* (ChIP-chip) que este FT reconoce los promotores de 3894 genes. Las plantas mutantes en este gen tuvieron alterada la expresión de 1144 genes, de los cuales sólo 219 habían sido identificados en el ensayo de ChIP-chip (Lee y col., 2007). En humanos y ratones se han realizado estudios similares obteniéndose en varios casos del orden de las decenas de miles de promotores reconocidos por un único FT (para una revisión, ver MacQuarrie y col., 2011).

En este escenario cobrarían todavía más importancia los mecanismos que regulan o condicionan la actividad de las proteínas HD-Zip. En primera instancia se podría considerar la accesibilidad a los promotores determinada por el estado de compactación de la cromatina. Además, podrían jugar un papel fundamental las proteínas reguladoras, ya sean quinasas, fosfatasas, proteínas involucradas en la sumoilación o en la ubiquitinación de estos FTs. De hecho, la unión de AtHB12 a los promotores de los genes involucrados en las vías de señalización de ABA, demostrada por Overnäs (2010), requirió de la presencia de esta fitohormona, mientras que AtHB7 se unió también en ausencia de este estímulo, destacando la importancia de la regulación.

En el caso de los FTs de la subfamilia II también es importante la capacidad de regulación del estado redox celular. La formación de puentes disulfuro intermoleculares entre las cisteínas del motivo CPSCE sería causal de inactivación (Tron y col., 2002).

Como se mencionó en el capítulo anterior, un factor adicional que podría afectar la funcionalidad de los FTs de la subfamilia I es la posible interacción diferencial con la maquinaria transcripcional a través de los diferentes motivos AHA.

La búsqueda *de novo* de motivos permitió reconocer el enriquecimiento de la secuencia CACGTG en los promotores de varios MATs, tanto pertenecientes a genes de la subfamilia I como de la II. Este motivo es muy similar al ABRE unido por los FTs de la familia GBF, que incluye a la familia ABF involucrada en la respuesta a ABA (Choi y col., 2000). En función de este resultado es posible plantear una alternativa más en cuanto a los mecanismos de acción de los FTs HD-Zip. Sería factible que los FTs de ambas familias actúen en forma conjunta en la regulación de sus genes blanco, lo cual añadiría un condicionante más a la actividad de las proteínas HD-Zip una vez unidas a los promotores.

### IV.4.3 Redundancia funcional entre miembros de las subfamilias I y II

Como se mencionó en la Sección III.1.1 (p. 42), las familias génicas surgieron de eventos de duplicación que generalmente condujeron a la coexistencia de genes con un grado de redundancia funcional muy elevado. Los procesos posteriores de subfuncionalización y neofuncionalización aseguraron que algunos de estos duplicados no se perdieran con el curso de la evolución, resultando en el crecimiento sostenido de estas familias (Lynch y Conery, 2000).

Disponer de un MAT para cada FT HD-Zip permitió la exploración de la potencial redundancia a través del cálculo del solapamiento existente entre los MATs de todos los posibles pares de genes (ver Figura IV.10, p. 105). La interpretación en términos de las relaciones filogenéticas existentes entre los genes indicó que algunos de los grupos más estrechamente relacionados presentarían cierta redundancia funcional, mientras que otros no. En la subfamilia I mostraron un solapamiento elevado el grupo I (AtHB12 y AtHB7); tres miembros del grupo V (AtHB23, AtHB20 y AtHB3, pero no AtHB13) y los tres miembros del grupo VI (AtHB53, AtHB40 y AtHB21). Entre los genes de la subfamilia II presentaron mayor solapamiento los del grupo  $\alpha$  (AtHB18 y AtHB17) y dos miembros del grupo  $\gamma$  (HAT1 y HAT2, excluyendo a AtHB2).

Además, se evaluó si cada MAT del par comparado contenía a los dos genes HD-Zip de estos MATs (vale recordar que cada MAT incluye al HD-Zip al que pertenece). Pudo verse que los dos MATs de AtHB12 y AtHB7 del grupo I contuvieron a ambos genes. Esto concuerda con los resultados de Overnäs (2010) que mostraron que ambos genes son capaces de unir los promotores de su parólogo, y además su propio promotor.

Lo mismo se observó para AtHB20 y AtHB3, excluyendo a AtHB23, el tercer gen del grupo con el que había elevado solapamiento entre los MATs. Es importante resaltar que a nivel filogenético, AtHB20 y AtHB3 están relacionados más estrechamente entre sí, que con AtHB23. Dentro del grupo VI, el MAT de AtHB40 incluyó tanto a AtHB53 como a AtHB21, pero los MATs de estos últimos dos genes no contuvieron a otros genes HD-Zip más allá del propio AtHB40. En la subfamilia II se observó que los MATs de AtHB18 y AtHB17 no incluyeron a ambos genes, pero esto sí ocurrió con los MATs de HAT1 y HAT2.

Más allá de estos grupos también se verificó cierto solapamiento entre los MATs. Particularmente el MAT de AtHB23, el mayor en número de *probesets*, tuvo un nivel de solapamiento medio considerablemente mayor al resto. Exceptuando este caso, no hubo un muy alto grado de solapamiento entre proteínas que no pertenecieran a un mismo grupo, con la posible excepción de AtHB40 con AtHB20 y AtHB3, AtHB51 con AtHB52 y AtHB4 con AtHB52 y AtHB17.

Sin considerar los grupos mencionados, se observaron pocos casos de genes HD-Zip en los MATs de otros genes HD-Zip. Notablemente, el MAT de AtHB12 contuvo a AtHB7, como se mencionó, pero también a AtHB6 y a AtHB13. Esto ocurrió también entre los MATs de AtHB6 y HAT1, AtHB51 y AtHB52, y AtHB2 y HAT2.

## IV.4. Discusión de resultados

---

En conclusión, estos resultados mostraron que existiría cierto grado de redundancia funcional entre algunos FTs HD-Zip, pero no sería tan extendida. Esta redundancia sería particularmente elevada entre algunos de los genes que pertenecen a un mismo grupo o clado, alcanzando niveles máximos de solapamiento que rondan el 50 %.

La presencia de genes HD-Zip en algunos de los MATs de otros HD-Zip también ocurrió notoriamente entre genes relacionados. También hubo algunos casos entre genes con una relación más distante; uno sólo de ellos entre el MAT de un gen HD-Zip I con el de un gen HD-Zip II (AtHB6 y HAT1).

### IV.4.3.1 Los MATs HD-Zip frente a las hipótesis previas

En una revisión reciente se postuló que a través de los FTs HD-Zip I y II se integran las señales de distintos estímulos externos, alterando así los procesos de diferenciación, división y expansión celular en diferentes tejidos (Harris y col., 2011). Esta hipótesis se fundamenta en la capacidad de los miembros de ambas subfamilias de reconocer prácticamente el mismo elemento en *cis*, y de regular procesos del desarrollo celular con resultados similares (generalmente alteraciones en los patrones de expansión celular).

Esta propuesta es interesante y su motivación no carece de fundamentos, sin embargo, los resultados obtenidos respecto al solapamiento de los MATs indicaron que de existir interrelación entre las vías, no sería tan extendida; al menos desde una perspectiva global. Obviando el caso de AtHB23 y AtHB17 con MATs numerosos, no hubo mayor solapamiento entre los genes blanco controlados por los FTs de las diferentes subfamilias. Pude mencionarse la excepción de AtHB4 con AtHB51 y AtHB52, aunque debe tenerse en cuenta que el MAT de AtHB4 es el más pequeño y los MATs de los HD-Zip I tienen más de 600 *probesets* cada uno. Sin embargo, el solapamiento tampoco es nulo o despreciable en todos los casos, y debe considerarse que los microarreglos a partir de los que se realizó el análisis correspondieron exclusivamente a condiciones de estrés abiótico. Serán necesarios más estudios de tipo global para continuar poniendo a prueba esta hipótesis.

Respecto a los FTs HD-Zip II, se ha propuesto que sus miembros conformarían una compleja red interconectada de regulaciones negativas. En las plantas de *A. thaliana* transformadas con la construcción *35SCaMV::AtHB2* el gen endógeno se encontraba reprimido, mientras que las plantas que expresaban una proteína con el dominio HD-Zip de AtHB2 fusionado al motivo activador VP16 permitieron observar que este FT reconoce su propio promotor (Ohgishi y col., 2001). Con estas últimas plantas transgénicas se demostró también que AtHB2 es capaz de reconocer los promotores del resto de los genes HD-Zip II (Ciarbelli y col., 2008). En otro estudio en el que se utilizaron plantas con la construcción *35SCaMV::HAT2* se verificó una disminución en la expresión de los seis genes de esta subfamilia evaluados (Sawa y col., 2002). También en arroz se demostró la actividad transcripcional represora de un FT HD-Zip II (Meijer

## IV.4. Discusión de resultados

---

y col., 1997).

La capacidad de estos FTs de reprimir la expresión podría ser explicada por la presencia de un motivo represor del tipo EAR (Ohta y col., 2001) en las regiones amino terminales (RNT) de muchos de sus miembros, según se halló en un trabajo de búsqueda de estos motivos en *A. thaliana* (Kagale y col., 2010). Una comparación de secuencias permitió observar que este motivo coincidiría con el motivo llamado ZIEBEL en las RNTs de las proteínas HD-Zip II según Mukherjee y col. (2009). En un trabajo reciente se demostró que proteínas del grupo TOPLESS de correpresores interactúan, muy probablemente a través del motivo EAR, con HAT3, HAT14 y HAT22; revelando así un mecanismo molecular para la acción negativa de los FTs de la subfamilia II (Causier y col., 2012).

El análisis de los MATs de los genes HD-Zip II no dio un fuerte sustento a la propuesta de que se formaría una red interconectada. Según estos resultados sólo AtHB2 y HAT22 estarían relacionados transcripcionalmente, y HAT1 con HAT2. El nivel de solapamiento de MATs es significativo entre los parálogos AtHB18 y AtHB17, entre AtHB17 y AtHB4, y en menor medida entre los dos pares de FT que interaccionarían (AtHB2 y HAT22, y HAT1 con HAT2). No obstante, es nuevamente importante remarcar el hecho de que los datos provienen de ensayos realizados en condiciones de estrés abiótico; no puede descartarse, por ejemplo, que si se hubieran utilizado microarreglos de plantas tratadas en distintas condiciones de iluminación la conclusión podría haber sido diferente.

### IV.4.4 Coordinación de la expresión en los MATs

La visualización global de los perfiles de expresión de los genes de cada MAT mostró que en varios de ellos, como los presentados en las Figuras IV.12 (p. 107) y IV.13 (p. 108), los niveles medios de subgrupos de *probeset* fueron más elevados en un tejido (vástago o raíz) respecto al otro; y con otros subgrupos pasó lo opuesto. En varias de estas gráficas puede apreciarse que el gen HD-Zip del MAT queda asociado a alguno de estos subgrupos. Esto sugirió que existe un patrón de expresión claramente diferente entre tejidos para una gran parte de los genes de los MAT.

Esto podría estar sugiriendo que la actividad del gen HD-Zip presenta diferencias importantes entre el vástago y la raíz, aunque haría falta mayor evidencia para sustentar esta hipótesis. Sin embargo, existe al menos un caso en el que se ha descrito este fenómeno. HAT2 participa en la respuesta a la auxinas (Aux). En la porción aérea de la planta (o vástago) su expresión mediante el promotor *35SCaMV* generó fenotipos típicos de las respuestas a esta fitohormona: mayor elongación del hipocotilo, cotiledones epinásticos, y peciolo largos. Por el contrario, estas mismas plantas presentaron en la raíz fenotipos opuestos a los esperados en respuesta a la Aux, como una reducción en la elongación de las raíces laterales (Sawa y col., 2002).

Por otro lado, el efecto de las diferencias en los niveles medios globales por tejido fue tan

## IV.4. Discusión de resultados

---

marcado que podría causar inconvenientes en la detección de patrones de expresión. Los perfiles de algunos genes pertenecientes a diferentes subgrupos podrían tener una buena correlación dentro de los microarreglos de un mismo tejido, pero la diferencia en los niveles “basales” entre tejidos estaría impidiendo reconocerlo. Por esto se procedió a hacer un análisis discriminando las muestras por tejidos, y para capturar mejor las diferencias en los perfiles, también por tratamiento.

Es importante aclarar que para estos análisis no fueron utilizadas las muestras provenientes de cultivos celulares pues éstas habían sido empleadas únicamente en ensayos de estrés por calor. Además, estas muestras representan un sistema artificial y, por lo tanto, en él podrían ocurrir procesos que naturalmente no se verifiquen.

### IV.4.4.1 Los genes HD-Zip participarían en las respuestas al calor

El estudio de la coordinación en la expresión fue empleado para conocer el papel de los genes HD-Zip en la respuesta a los diferentes estreses abióticos. Para esto se seleccionaron genes regulados en forma conjunta en cada uno de los tratamientos diferenciados por tejidos. El criterio de selección involucró una correlación que superara cierto umbral y que correspondiera a grupos de cinco o más genes (ver Figura IV.14, p. 109). Luego se hizo un recuento del número de genes que superaran este criterio para utilizarlo como un indicador del grado de participación del FT al que pertenecía el MAT en ese tratamiento.

Una comparación global de los resultados obtenidos para todos los genes HD-Zip mostró, en primer lugar, diferencias notorias entre raíces y vástago. En raíces hay claramente una mayor participación global en las respuestas de calor, salinidad, frío y, en menor medida, estrés osmótico. En contraste, en el vástago fue evidente la coordinación de genes en la gran mayoría de los MATs, sólo frente a estrés por calor. Podría considerarse que frente a frío y a estrés osmótico hubo una respuesta menor pero significativa, en términos generales; y el resto de los estreses presentó respuestas más heterogéneas entre los diferentes miembros, generalmente menores.

Estas observaciones resaltan que los genes HD-Zip tendrían una mayor participación global a nivel de raíces, particularmente en los estreses mencionados. En tanto que en el vástago la respuesta es más heterogénea entre los diferentes miembros, a excepción del estrés por calor que genera la coordinación relativa de la mayor parte de los genes del MAT.

Esto indica que habría un comportamiento diferencial de los genes HD-Zip I y II entre raíz y vástago. Además, revela que el papel de estos FTs en la respuesta al calor sería muy relevante. No se tiene conocimiento de que haya trabajos relacionando a genes HD-Zip con estrés por altas temperaturas, por lo que el resultado plantea una potencial función en una respuesta no explorada previamente.

Sobre este aspecto es oportuno mencionar que la sumoilación tiene una participación importante en las respuestas al calor en *A. thaliana*. Se ha demostrado que la proteína ligasa de SUMO SIZ1 participa en la tolerancia a altas temperaturas ya que las plantas mutantes en este

## IV.4. Discusión de resultados

---

gen tuvieron una elevada termosensibilidad (Yoo y col., 2006). Los autores propusieron que esta termotolerancia sería independiente del regulón de las proteínas *Heat Stress Transcription Factors* (HSFs), el más estudiado en esta respuesta. Sin embargo, otro trabajo sugiere que la interpretación de estos resultados podría no ser sencilla. Hallaron que un FT HSF que participa en la termotolerancia de *A. thaliana* es sumoilado e inactivado sólo luego de que la planta recibió el estrés y está recuperándose. La sumoilación lo inactiva y se observó que tanto las plantas mutantes en el gen como las transgénicas con la construcción *35SCaMV::AtSUMO1* son termosensibles (Cohen-Peer y col., 2010).

### IV.4.4.2 Los genes de los MATs son regulados en forma negativa o positiva en un mismo estrés

La selección de genes regulados en forma conjunta fue utilizada para investigar las características de los perfiles de expresión considerando la progresión a lo largo del tratamiento. En varios estreses y diferentes MATs pudo observarse la presencia de subgrupos de genes que sufrieron un efecto neto de inducción o represión en su expresión, como producto del tratamiento. Dentro de cada subgrupo los perfiles no fueron homogéneos, por lo que se resalta el efecto neto del tratamiento no considerando que haya una coordinación estrecha en la expresión de los genes de cada subgrupo.

Este resultado sugiere que los FTs HD-Zip I y II podrían estar cumpliendo la función de activadores y represores de la transcripción en una misma condición de estrés y en un mismo conjunto de tejidos, raíz o vástago. Sin embargo, es importante aclarar que ésta no es evidencia directa de que estos FTs poseerían ambas actividades. No se analizó junto a los genes coordinados la expresión del gen HD-Zip que los regularía, y la multiplicidad de perfiles promedio observados podría complicar la interpretación de un análisis de ese tipo.

En relación a la posibilidad de que los FTs de estas subfamilias tengan una doble actividad transcripcional sólo hay evidencia para las proteínas HD-Zip I. Como se discutió extensamente en la Sección III.4.2 (p. 73) y se mostró con AtHB13 (Sección III.3.5, p. 64), se han realizado ensayos, tanto en levaduras como en células vegetales, que demuestran la capacidad de activación transcripcional de los miembros de esta subfamilia. Por otro lado, el trabajo sobre MthB1 de *M. truncatula* indica que este miembro actúa como represor (Ariel y col., 2010).

Los FTs HD-Zip II poseen la capacidad de reprimir la expresión. Esto se demostró mediante numerosos experimentos, como se discutió en la Sección IV.4.3.1 (p. 119). La presencia del motivo EAR hallado en varios miembros brindaría un mecanismo molecular para esta actividad a través de la interacción con el correpresor TOPLESS (Kagale y col., 2010). En un trabajo reciente esta interacción fue demostrada para tres proteínas de la subfamilia: HAT3, HAT14 y HAT22 (Causier y col., 2012). Sin embargo, no hay pruebas de que tengan la capacidad de actuar como activadores. Es más, el motivo EAR tienen dominancia sobre el motivo activador VP16 cuando

## IV.4. Discusión de resultados

---

ambos se presentan fusionados en una misma proteína (Ohta y col., 2001). Sin embargo, puede especularse que la interacción con la proteína TOPLESS habría causado el enmascaramiento y la inhibición funcional del motivo VP16 en los ensayos mencionados. De ser así, no debería descartarse la posibilidad de que un cambio conformacional en la región amino terminal de un FT HD-Zip II oculte el motivo EAR permitiendo la acción de algún motivo activador.

### IV.4.5 Los MAT desde una perspectiva funcional

Disponer de una lista de genes blanco para cada gen HD-Zip permitió también explorar desde una perspectiva funcional los procesos o vías de señales en los que estaría involucrado cada FT. Con este fin se utilizó la anotación de términos funcionales de ontología de genes (términos GO) asociada a cada gen del MAT. Una vez hallados los términos asociados se evaluó el enriquecimiento estadístico de estos términos en cada MAT. Esto es afectado básicamente por el número de genes asociados a ese término GO, cuántos de ellos están en el MAT y el total de genes que posee el MAT. Además, se aplicó un programa basado en teoría de la información para resumir los resultados obtenidos en un número menor, pero informativo, de rótulos o términos funcionales (pertenecientes a la misma anotación GO; Berenstein y Chernomoretz, 2012, manuscrito en preparación).

Los resultados obtenidos son extensos y se presentan en forma de tablas en la Sección V.4 del Apéndice (p. 139). Su inspección permitió reconocer información relevante que se resume para cada gen HD-Zip a continuación.

#### AtHB12 y AtHB7

En ambos MAT se hallaron enriquecidas las categorías relacionadas a estímulos abióticos, más particularmente a la respuesta a déficit hídrico y a bajas temperaturas. También se halló enriquecida la categoría de respuesta a la fitohormona ABA. Estos resultados concuerdan ampliamente con los trabajos previos que involucran a ambos genes específicamente en la respuesta a ABA y al estrés hídrico (Hjellström y col., 2003; Olsson y col., 2004; Söderman y col., 1996), aunque no hay evidencia relacionándolos en la respuesta a las bajas temperaturas.

Otra observación importante no relacionada a la inspección de los términos GO, es la presencia en los MATs de AtHB12 y AtHB7 de los siguientes genes: los dos FTs HD-Zip, *ABI1*, *ABI2* y *AHG3*. Todos estos genes habían sido reconocidos por Overnäs (2010) por tener regiones promotoras a las que se unieron ambos FTs; en el caso de AtHB12 necesitando la inducción por ABA para reconocer algunos de ellos. También se encontró a AtHB6 en el MAT de AtHB12, pero los autores del trabajo citado hallaron que sólo AtHB7 era capaz de unir su promotor.

A pesar de que MtHB1 de *M. truncatula* pertenece filogenéticamente al mismo grupo que AtHB7 y AtHB12, ninguno de los MATs de estos genes incluyó FTs tipo LOB, grupo al que pertenece el blanco de MtHB1, LBD1. Es importante resaltar que los microarreglos utilizados

#### IV.4. Discusión de resultados

---

proviene de ensayos de estrés abiótico, mientras que LBD1 está involucrado en el desarrollo de raíces laterales (Ariel y col., 2010).

##### **AtHB6**

Al igual que en los MATs de AtHB12 y AtHB7, se encontraron enriquecidos los términos GO de respuesta a déficit hídrico y regulación de la vía de señales de ABA. Esta función ya había sido reconocida para AtHB6 (Söderman y col., 1999). Además, se halló el gen *ABI1* en su MAT, lo cual es relevante pues se demostró que AtHB6 establece una interacción proteína-proteína con la fosfatasa ABI1 involucrada en la respuesta a ABA (Himmelbach y col., 2002).

Debe mencionarse también que se hallaron enriquecidas las categorías de aclimatación al frío y respuesta al calor, ambas señaladas por los resultados presentados en la Sección IV.3.8 como vías en las que participan varios FTs HD-Zip.

##### **AtHB16**

Este FT había sido relacionado a la determinación del tamaño de las hojas, la inhibición de la elongación del hipocotilo en respuesta a la luz azul y en la regulación de la floración por el fotoperiodo (Wang y col., 2003). Se hallaron enriquecidos términos GO relacionados a la fotosíntesis y además, puede destacarse que su MAT contiene al fotorreceptor de luz azul *CRY1*.

##### **AtHB1**

Aoyama y col. (1995) demostraron que AtHB1 genera deetiología constitutiva al ser expresado con un promotor *35SCaMV* en tabaco. Además, estas plantas tuvieron un defecto en el desarrollo normal del parénquima en empalizada. En este MAT se hallaron enriquecidos numerosos términos GO asociados a genes involucrados en la fotosíntesis y la organización cloroplastídica, lo que presenta una notable coherencia con los resultados experimentales.

##### **AtHB13**

AtHB13 ha sido asociado a la respuesta a azúcares (Hanson y col., 2001) y a la tolerancia a temperaturas de congelamiento (Cabello y col., 2012). Sin embargo se hallaron enriquecidas categorías relacionadas a estas funciones en el MAT de este FT.

##### **AtHB23**

Este gen se expresa principalmente en la cara adaxial de hojas (Kim y col., 2007) y los términos estadísticamente enriquecidos correspondieron fundamentalmente a la fotosíntesis, desde los complejos fotosintéticos hasta la síntesis de clorofila y otros carotenos, lo que podría indicar cierta concordancia por tratarse de un tejido muy activo en este proceso.



## IV.4. Discusión de resultados

---

### HAT14

En el MAT de este FT de la subfamilia II se encontraron enriquecidos algunos términos GO relacionados a la fotosíntesis, la respuesta a la luz azul y a las altas intensidades de luz. Lo cual, a pesar de no haber sido investigado, es una característica común a varios genes de esta subfamilia (Carabelli y col., 1993; Ciarbelli y col., 2008; Sorin y col., 2009).

### HAT22

En el MAT de HAT22 estuvieron enriquecidos varios términos relacionados a respuestas al estrés: de tipo oxidativo, déficit hídrico, daño mecánico, frío, calor y ataque por hongos. Puede mencionarse que este FT fue reconocido por ser inducido por sequía y múltiples estímulos hormonales (Huang y col., 2008).

### AtHB2

AtHB2 es sin duda el miembro más caracterizado de la subfamilia II de los FTs HD-Zip. Su participación en el síndrome de evasión de la sombra, desencadenado por una alta proporción de luz rojo lejano, es bien conocida y ha sido considerablemente explorada. Sin embargo, aunque apareció enriquecido un término de respuesta a roja o rojo lejano, éste involucra apenas cuatro genes. No se encontró en su MAT a *HFR1* (Sessa y col., 2005), ni a *PIF4* o *5* (Kunihiro y col., 2011); FTs que lo controlarían, aunque se desconoce si lo harían en forma directa. Es importante recordar que esto podría ser atribuido a que los microarreglos fueron realizados en condiciones que podrían evitar que estas relaciones se manifiesten.

Se hallaron además términos asociados a estrés por déficit hídrico y a respuestas a las hormonas citoquininas (CK).

### HAT1

En este MAT se hallaron enriquecidos términos GO relacionados principalmente las vías de señales de ABA, Aux, frío y estrés hídrico. Fue interesante encontrar que entre los genes de respuesta a ABA estuvieron *ABI1*, *ABI2*, *AHG3* y *AtHB6*; todos genes presentes en el MAT de AtHB12 y de AtHB7 (en este último no se encuentra *AtHB6*). De verificarse esto y teniendo en cuenta la importancia de estos genes en la vía de señales de esta hormona, sería evidencia de que podría tener cierto impacto el escaso solapamiento entre los genes regulados por miembros de las diferentes subfamilias, como propusieron Harris y col. (2011).

### HAT2

(Sawa y col., 2002) describieron que HAT2 es inducido rápidamente por Aux y afecta respuestas relacionadas a esta hormona. En su MAT se enriquecieron términos GO asociados a

## IV.4. Discusión de resultados

---

Aux, acertando así a reconocer la función sobre la que se posee evidencia.

### HAT3

Se encontraron en su MAT varios términos GO relacionados a las respuestas a luz, tanto luz azul y roja, como genes de las vías fotosintéticas. Entre los genes de luz azul se encuentra *CRY1*, un fotorreceptor importante que apareció en el MAT de AtHB16. Debe resaltarse que se había sugerido que HAT3 participa las respuestas a cambios de las condiciones de luz, particularmente en condiciones de baja proporción de luz roja a luz rojo lejano (Ciarbelli y col., 2008).

#### IV.4.5.1 Algunas consideraciones finales sobre el análisis funcional

Mediante el análisis del enriquecimiento de términos GO fue posible reconocer categorías funcionales asociadas a las vías de señales en las que participaría cada FT HD-Zip. En varios casos los resultados coinciden con la función hallada en trabajos previos, lo cual respalda de alguna forma la confiabilidad de los MATs obtenidos. Debe tenerse en cuenta que a pesar de la potencialidad que pueda poseer el algoritmo de reconstrucción de redes, hay numerosas fuentes de “ruido” que dificultan la tarea. Por ejemplo, los complejos patrones espaciales de expresión que presentan los genes HD-Zip y un sinnúmero de otros genes son “diluidos” al separar la plántula en dos partes para la toma de muestras.

También es importante resaltar que los datos provienen exclusivamente de ensayos realizados en condiciones de estrés abiótico, por lo que el potencial del análisis estará focalizado en estos tratamientos. Por esto, es limitada la capacidad de reconocer la función de los genes HD-Zip principalmente relacionados a otros procesos.

A pesar de estos condicionantes, se obtuvieron buenos resultados en varios de los casos en los que existía información previa disponible, por lo que puede considerarse que los MATs brindan una lista de muy buenos candidatos a genes blanco. Basándose en esta información se podrá guiar la realización de futuros trabajos experimentales y estudios bioinformáticos.

### IV.5 Conclusiones

Los resultados presentados permitieron arribar a las siguientes conclusiones:

- 1) Los MATs de los FTs HD-Zip I y II obtenidos a partir de la reconstrucción de una red transcripcional basada en ensayos con diferentes estreses abióticos son de tamaños muy variables.
- 2) Los promotores de los genes de los MATs no mostraron enriquecimiento en el elemento en *cis* CAATNATTG unido *in vitro* con máxima afinidad por los FTs HD-Zip I y II. Esto permite proponer que estos FTs tendrían la capacidad de reconocer motivos ligeramente diferentes a éste con afinidad suficiente para poseer actividad transcripcional.
- 3) El reconocimiento de elementos ABRE en los promotores de muchos de los genes de los MATs señalaría que hay una fuerte interrelación con las vías de señales de ABA reguladas por otros FTs, por ejemplo las proteínas ABF.
- 4) El estudio del solapamiento entre los MAT de diferentes FTs indicó un grado intermedio a bajo de potencial redundancia funcional entre miembros, siendo mayor entre algunos FTs con estrecha relación filogenética.
- 5) La inspección de los perfiles de expresión globales de los genes de cada MAT indicó notorias diferencias en los niveles medios de subgrupos de genes entre vástago y raíz. Esto sugeriría un comportamiento distinto de los genes del MAT entre tejidos.
- 6) La coordinación de los genes del MAT en los diferentes tejidos y condiciones de estrés permitió observar una elevada proporción de genes regulados conjuntamente por los FTs HD-Zip I y II en estrés por calor, tanto en vástago como en raíz. Esto plantea una novedosa función para estas proteínas que a la fecha no había sido investigada.
- 7) El análisis de los perfiles de expresión de los genes regulados en forma conjunta a lo largo de los tratamientos mostró subgrupos de genes en varios MAT regulados positivamente y otros regulados negativamente. Deberían realizarse más estudios para establecer si esto es producto de la acción directa de los FTs como activadores y represores de diferentes genes en un mismo tejido y condición de estrés.
- 8) La caracterización funcional de los genes de los MATs permitió proponer las vías de señales en las que participarían los FTs HD-Zip. La coherencia hallada con la función descrita previamente para muchos genes otorga sustento a los resultados, mientras que otras funciones podrían ser tomadas como hipótesis para futuros trabajos.

## Conclusiones generales

El presente trabajo de Tesis tuvo como objetivo investigar generalidades y particularidades funcionales de los miembros de las subfamilias I y II de FTs HD-Zip. Aplicando herramientas bioinformáticas a datos genómicos y transcriptómicos fue posible reconocer mecanismos de acción y regulación de las proteínas HD-Zip I así como identificar las potenciales vías de señales en las que participan las proteínas de ambas subfamilias. Estos resultados fueron corroborados en forma experimental sólo parcialmente y podrán ser utilizados en el futuro para guiar la experimentación sobre los diferentes FTs.

Por otra parte, este enfoque permitió reconocer características comunes y divergentes de estas proteínas que serían difíciles de poner en evidencia a nivel global experimentalmente. De esta forma, se pudieron abordar interrogantes a nivel de un organismo, *Arabidopsis thaliana*, y de diferentes organismos del reino vegetal. Por ejemplo, la presencia del motivo AHA y de los motivos de sumoilación en las proteínas de la subfamilia I fue reconocida gracias a la manipulación de un número considerable de secuencias. En el caso del estudio de las vías de señales, la potencial redundancia funcional entre miembros pudo ser investigada mediante la generación de conjuntos de potenciales genes blanco.

Los resultados alcanzados revelaron un mayor nivel de complejidad en los procesos que gobiernan la acción de las proteínas HD-Zip y brindaron una propuesta concreta acerca de los genes que regulan. Sólo el empleo conjunto de los diferentes enfoques existentes permitirá seguir enriqueciendo la imagen actual sobre la función de estas proteínas.

# Apéndice

## V.1 Secuencias utilizadas en los análisis filogenéticos

Proteína	Especie	Código identificador
At_ATHB1	<i>Arabidopsis thaliana</i>	NP_186796
At_ATHB12	<i>Arabidopsis thaliana</i>	NP_191748
At_ATHB13	<i>Arabidopsis thaliana</i>	NP_177136
At_ATHB16	<i>Arabidopsis thaliana</i>	NP_195716
At_ATHB20	<i>Arabidopsis thaliana</i>	NP_186771
At_ATHB21	<i>Arabidopsis thaliana</i>	NP_179445
At_ATHB22	<i>Arabidopsis thaliana</i>	NP_850266
At_AtHB23	<i>Arabidopsis thaliana</i>	NP_564268
At_ATHB3	<i>Arabidopsis thaliana</i>	NP_568309
At_ATHB40	<i>Arabidopsis thaliana</i>	NP_195392
At_ATHB5	<i>Arabidopsis thaliana</i>	NP_201334
At_ATHB51	<i>Arabidopsis thaliana</i>	NP_195999
At_ATHB52	<i>Arabidopsis thaliana</i>	NP_200209
At_ATHB53	<i>Arabidopsis thaliana</i>	NP_201471
At_ATHB54	<i>Arabidopsis thaliana</i>	NP_174025
At_ATHB6	<i>Arabidopsis thaliana</i>	NP_565536
At_ATHB7	<i>Arabidopsis thaliana</i>	NP_182191
Bn-AAR0493	<i>Brassica napus</i>	AAR04932
Brs-hb-6	<i>Brassica rapa subsp. pekinensis</i>	AAF73482
Cp-CPHB-4	<i>Craterostigma plantagineum</i>	AAL57494
Cp-CPHB-6	<i>Craterostigma plantagineum</i>	AAL57496
Cp-CPHB-7	<i>Craterostigma plantagineum</i>	AAL57497
Cr-ABL6311	<i>Catharanthus roseus</i>	ABL63116
Cr-ABL631B	<i>Catharanthus roseus</i>	ABL63115
Cr-CRHB11	<i>Ceratopteris richardii</i>	BAA34245
Cr-CRHB4	<i>Ceratopteris richardii</i>	BAA34238
Cr-CRHB6	<i>Ceratopteris richardii</i>	BAA34240
Dc-05622.1	<i>Daucus carota</i>	BAA05622
Dc-05623.1	<i>Daucus carota</i>	BAA05623
Dc-05624.1	<i>Daucus carota</i>	BAA05624
Dc-05625.1	<i>Daucus carota</i>	BAA05625
Gh-ABO4774	<i>Gossypium hirsutum</i>	ABO47743
Gm-184277-	<i>Glycine max</i>	AAF01764
Gm-184278-	<i>Glycine max</i>	AAF01765
Gm-AAX9867	<i>Glycine max</i>	AAX98670
Gm-ACU1808	<i>Glycine max</i>	ACU18084
Gm-ACU1869	<i>Glycine max</i>	ACU18697
Gm-ACU1987	<i>Glycine max</i>	ACU19873
Gm-ACU198B	<i>Glycine max</i>	ACU19872
Gm-ACU2089	<i>Glycine max</i>	ACU20897
Gm-ACU2100	<i>Glycine max</i>	ACU21007
Gm-ACU2401	<i>Glycine max</i>	ACU24015
Gm-ACU2443	<i>Glycine max</i>	ACU24436
Ha_HAHB1	<i>Helianthus annuus</i>	AAA63765
Ha_HAHB11	<i>Helianthus annuus</i>	
Ha_HAHB4	<i>Helianthus annuus</i>	AAA63768
Lj-BAG5005	<i>Lotus japonicus</i>	BAG50054
Mt_MTHB1	<i>Medicago truncatula</i>	HD1374
Mt-ACJ8462	<i>Medicago truncatula</i>	ACJ84623
Mt-ACJ8504	<i>Medicago truncatula</i>	ACJ85047
Mt-ACJ8510	<i>Medicago truncatula</i>	ACJ85107
Nt-Hfi22	<i>Nicotiana tabacum</i>	AAM48290
Os_02g4333	<i>Oryza sativa Japonica Group</i>	NP_001047582
Os_02g4970	<i>Oryza sativa Japonica Group</i>	NP_001048008
Os_03g0745	<i>Oryza sativa Japonica Group</i>	A2XD08
Os_03g0896	<i>Oryza sativa Japonica Group</i>	NP_001049219
Os_03g1021	<i>Oryza sativa Japonica Group</i>	NP_001049280

## Apéndice

Proteína	Especie	Código identificador
Os_04g4581	<i>Oryza sativa Japonica Group</i>	Q01IK0
Os_07g3932	<i>Oryza sativa Japonica Group</i>	NP_001060106
Os_08g3208	<i>Oryza sativa Japonica Group</i>	NP_001061807
Os_08g3758	<i>Oryza sativa Japonica Group</i>	NP_001062064
Os_09g2118	<i>Oryza sativa Japonica Group</i>	A3BYC1
Os_09g2946	<i>Oryza sativa Japonica Group</i>	EEE69885
Os_09g3591	<i>Oryza sativa Japonica Group</i>	Q651Z5
Os_10g2309	<i>Oryza sativa Japonica Group</i>	Q6V9S9
Os_10g2650	<i>Oryza sativa Japonica Group</i>	EAZ16012
Pg-ABA5414	<i>Picea glauca</i>	ABA54144
Pp_Pphb1	<i>Physcomitrella patens</i>	AB028072
Pp_Pphb2	<i>Physcomitrella patens</i>	AB028073
Pp_Pphb5	<i>Physcomitrella patens</i>	AB028076
Pp_pphb6	<i>Physcomitrella patens</i>	AB028077
Pp_Pphb7	<i>Physcomitrella patens</i>	AB028078
Pp_Pphb8	<i>Physcomitrella patens</i>	AB028079
Pp_Pphb9	<i>Physcomitrella patens</i>	
Pp_sca_143	<i>Physcomitrella patens</i>	
Pp_sca_154	<i>Physcomitrella patens</i>	
Pp_sca_28	<i>Physcomitrella patens</i>	
Pp_sca_31	<i>Physcomitrella patens</i>	
Pp_sca_34	<i>Physcomitrella patens</i>	
Pp_sca_35	<i>Physcomitrella patens</i>	
Pp_sca_4	<i>Physcomitrella patens</i>	
Pp_sca_65a	<i>Physcomitrella patens</i>	
Pp_sca_65b	<i>Physcomitrella patens</i>	
Pp_sca_77	<i>Physcomitrella patens</i>	
Ps-ABK2449	<i>Picea sitchensis</i>	ABK24498
Ps-ABK2462	<i>Picea sitchensis</i>	ABK24625
Ps-ABK2476	<i>Picea sitchensis</i>	ABK24769
Ps-ABK2572	<i>Picea sitchensis</i>	ABK25728
Ps-ABR1622	<i>Picea sitchensis</i>	ABR16224
Pt_343725	<i>Populus trichocarpa</i>	
Pt_548258	<i>Populus trichocarpa</i>	XP_002297947
Pt_568845	<i>Populus trichocarpa</i>	
Pt_594622	<i>Populus trichocarpa</i>	XP_002328030
Pt_655260	<i>Populus trichocarpa</i>	XP_002310795
Pt_687113	<i>Populus trichocarpa</i>	XP_002315797
Pt_696444	<i>Populus trichocarpa</i>	XP_002311597
Pt_703426	<i>Populus trichocarpa</i>	XP_002327822
Pt_70493	<i>Populus trichocarpa</i>	ABK95942
Pt_731421	<i>Populus trichocarpa</i>	XP_002320889
Pt_736296	<i>Populus trichocarpa</i>	XP_002323851
Pt_88244	<i>Populus trichocarpa</i>	XP_002317992
Pt_93443	<i>Populus trichocarpa</i>	
Pt_98386	<i>Populus trichocarpa</i>	XP_002330748
Pt_HAT5	<i>Populus trichocarpa</i>	XP_002312706
Pt_HB	<i>Populus trichocarpa</i>	
Pt_HB7	<i>Populus trichocarpa</i>	XP_002302659
Pt_HBLZ	<i>Populus trichocarpa</i>	
Pt_HD56	<i>Populus trichocarpa</i>	ABK95942
Pt_unknon	<i>Populus trichocarpa</i>	XP_002309768
Pt_unknown	<i>Populus trichocarpa</i>	
Pv-HDZ1	<i>Phaseolus vulgaris</i>	AAK84885
Pv-HDZ2	<i>Phaseolus vulgaris</i>	AAK84886
Pv-HDZ3	<i>Phaseolus vulgaris</i>	AAK84887
Rc-XP25115	<i>Ricinus communis</i>	XP_002511518
Rc-XP25139	<i>Ricinus communis</i>	XP_002513970
Rc-XP25173	<i>Ricinus communis</i>	XP_002517348
Rc-XP25174	<i>Ricinus communis</i>	XP_002517497
Rc-XP25201	<i>Ricinus communis</i>	XP_002520111

## Apéndice

Proteína	Especie	Código identificador
Rc-XP2520B	<i>Ricinus communis</i>	XP_002520146
Rc-XP25299	<i>Ricinus communis</i>	XP_002529960
Sb-XP24459	<i>Sorghum bicolor</i>	XP_002445988
Sb-XP24468	<i>Sorghum bicolor</i>	XP_002446870
Sb-XP24528	<i>Sorghum bicolor</i>	XP_002452838
Sb-XP24603	<i>Sorghum bicolor</i>	XP_002460380
Sb-XP24627	<i>Sorghum bicolor</i>	XP_002462710
Sb-XP24657	<i>Sorghum bicolor</i>	XP_002465782
Sb-XP24672	<i>Sorghum bicolor</i>	XP_002467274
Sd-AAT4051	<i>Solanum demissum</i>	AAT40518
Sl-CAA6441	<i>Solanum lycopersicum</i>	CAA64417
Sl-CAB6711	<i>Solanum lycopersicum</i>	CAB67118
Sl-SIHDL1	<i>Silene latifolia</i>	BAD27254
Sm_18196	<i>Selaginella moellendorffii</i>	XP_002978617
Sm_18217	<i>Selaginella moellendorffii</i>	XP_002968958
Sm_19476	<i>Selaginella moellendorffii</i>	
Sm_81609	<i>Selaginella moellendorffii</i>	XP_002964248
Ta-TaHZI-1	<i>Triticum aestivum</i>	ABC86566
Ta-TaHZI-2	<i>Triticum aestivum</i>	ABC86567
Vs-tendril	<i>Vicia sativa</i>	ACI42916
Vv-CAN7896	<i>Vitis vinifera</i>	CAN78962
Vv-CAN8361	<i>Vitis vinifera</i>	CAN83612
Vv-CAN8396	<i>Vitis vinifera</i>	CAN83969
Vv-CAO1494	<i>Vitis vinifera</i>	CBI22185
Vv-CAO4102	<i>Vitis vinifera</i>	CBI32024
Vv-CAO6167	<i>Vitis vinifera</i>	CBI19202
Vv-CAO6250	<i>Vitis vinifera</i>	CBI15277
Vv-XP22629	<i>Vitis vinifera</i>	XP_002262950
Vv-XP22660	<i>Vitis vinifera</i>	XP_002266089
Vv-XP22715	<i>Vitis vinifera</i>	XP_002271523
Vv-XP22716	<i>Vitis vinifera</i>	XP_002271692
Vv-XP22768	<i>Vitis vinifera</i>	XP_002276889
Vv-XP22788	<i>Vitis vinifera</i>	XP_002278872
Vv-XP22800	<i>Vitis vinifera</i>	CBI34588
Vv-XP22857	<i>Vitis vinifera</i>	XP_002285743
Ze-18171.1	<i>Zinnia elegans</i>	BAB18171
Zm_4102187	<i>Zea mays</i>	
Zm_4112646	<i>Zea mays</i>	
Zm_4113431	<i>Zea mays</i>	
Zm_4118271	<i>Zea mays</i>	
Zm_4124075	<i>Zea mays</i>	
Zm_4124691	<i>Zea mays</i>	
Zm_4134425	<i>Zea mays</i>	
Zm_413795	<i>Zea mays</i>	
Zm_422699	<i>Zea mays</i>	
Zm_433132	<i>Zea mays</i>	
Zm_433210	<i>Zea mays</i>	
Zm_433493	<i>Zea mays</i>	
Zm_4492	<i>Zea mays</i>	ACF86119
Zm_4496	<i>Zea mays</i>	ACR37611
Zm_459693	<i>Zea mays</i>	
Zm_459694	<i>Zea mays</i>	
Zm_469357	<i>Zea mays</i>	
Zm_469358	<i>Zea mays</i>	
Zm_470295	<i>Zea mays</i>	
Zm_479999	<i>Zea mays</i>	
Zm_480132	<i>Zea mays</i>	
Zm_483405	<i>Zea mays</i>	
Zm_hox12	<i>Zea mays</i>	
Zm_hox6	<i>Zea mays</i>	
ATHB17	<i>Arabidopsis thaliana</i>	NP_178252.2



## Apéndice

---

Proteína	Especie	Código identificador
HAT1	<i>Arabidopsis thaliana</i>	NP_193476.1
HAT22	<i>Arabidopsis thaliana</i>	NP_195493.1



# Apéndice

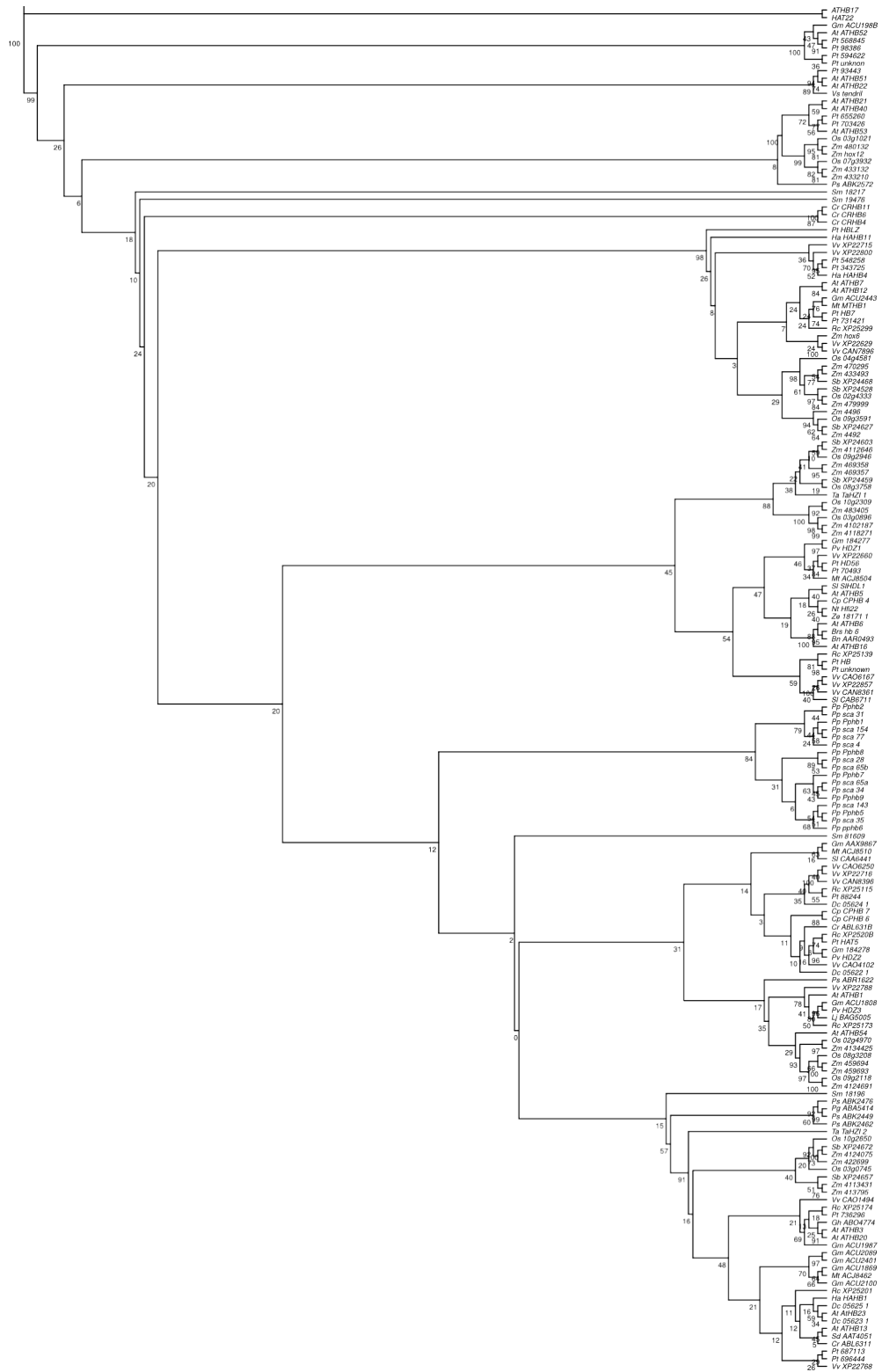
At_AtHB23	T	K	O	L	E	K	D	Y	D	M	L	K	R	O	F	E	S	L	F	D	E	N	V	L	O	T	O	N	O	K	I	O	A	Q	.	V	M	A	L	K	S	R	E	P	.	.	.	I	E	100		
At_ATHB13	T	K	O	L	E	K	D	Y	D	T	L	K	R	O	F	E	S	L	F	D	E	N	V	L	O	T	H	N	O	K	I	O	A	Q	.	I	M	G	L	K	S	R	E	Q	.	.	.	T	E	100		
Gm-ACU2089	T	K	O	L	E	K	D	Y	D	V	L	K	R	O	F	E	A	V	K	S	D	N	D	A	L	O	A	O	N	O	K	I	O	A	Q	.	I	L	A	L	K	S	R	E	P	.	.	.	T	E	100	
Gm-ACU2401	T	K	O	L	E	K	D	Y	D	V	L	K	R	O	F	E	A	V	K	S	D	N	D	A	L	O	A	O	N	O	K	I	O	A	Q	.	I	L	A	L	K	S	R	E	P	.	.	.	T	E	100	
Gm-ACU1869	T	K	O	L	E	K	D	Y	D	L	L	K	R	O	F	E	A	T	K	A	D	N	D	A	L	O	A	O	N	O	K	I	O	T	E	.	I	L	A	L	K	S	R	E	P	.	.	.	T	E	100	
Mt-ACU8462	T	K	O	L	E	K	D	Y	D	V	L	K	R	O	F	E	A	T	K	A	D	N	D	A	L	O	A	O	N	O	K	I	O	T	E	.	I	L	A	L	K	S	R	E	P	.	.	.	T	E	100	
Gm-ACU2100	T	K	O	L	E	K	D	Y	D	L	L	K	R	O	F	E	A	T	K	A	D	N	D	A	L	O	F	O	N	O	K	I	O	T	E	.	I	L	A	L	K	S	R	E	P	.	.	.	T	E	100	
Rc-XP25201	T	K	O	L	E	K	D	Y	D	L	L	K	R	O	F	E	A	T	K	A	D	N	D	A	L	O	A	O	N	O	K	I	H	A	E	.	I	M	A	L	K	S	R	E	P	.	.	.	T	E	100	
Vv-XP22768	T	K	O	L	E	K	D	Y	D	L	L	K	R	O	F	E	A	V	K	A	E	N	D	A	L	O	A	O	N	O	K	I	H	A	E	.	M	L	A	L	K	S	R	E	P	.	.	.	T	E	100	
Cr-ABL6311	T	K	O	L	E	K	D	Y	D	I	L	K	R	O	F	E	A	T	K	A	E	N	D	A	L	O	A	O	N	O	K	I	H	A	E	.	I	M	A	L	K	S	R	E	P	.	.	.	T	E	100	
Pt_687113	T	K	O	L	E	K	D	Y	D	L	L	K	R	O	F	E	A	T	K	A	E	N	D	A	L	O	A	O	N	O	K	I	H	A	E	.	I	L	T	L	K	S	R	E	P	.	.	.	T	E	100	
Ha_HAHB1	T	K	O	L	E	K	D	Y	D	A	L	K	R	O	F	E	A	V	K	A	E	N	D	S	L	O	S	N	H	K	I	H	A	E	.	I	M	A	L	K	S	R	E	P	.	.	.	A	E	100		
Pt_696444	T	K	O	L	E	I	D	Y	D	L	L	K	R	O	F	E	A	V	K	A	E	N	N	A	L	O	T	O	N	O	R	L	H	A	E	.	I	L	A	L	K	S	R	E	P	.	.	.	T	E	100	
Vv-CA01494	T	K	O	L	E	K	D	Y	D	L	L	K	R	O	F	E	A	V	K	A	D	N	D	A	L	O	A	O	N	K	I	H	A	E	.	L	L	A	L	K	S	R	E	P	.	.	.	K	G	100		
Sd-AAT4051	T	K	O	L	E	K	D	Y	E	V	L	K	R	O	F	E	D	A	I	K	A	E	N	D	A	L	O	T	O	N	O	K	I	H	A	E	.	I	M	S	L	K	S	R	E	Q	.	.	.	P	T	100
Rc-XP25174	T	K	O	L	E	K	D	Y	D	V	L	K	R	O	F	E	D	A	L	K	A	D	N	E	V	L	O	T	O	N	K	I	H	A	E	.	L	M	D	L	K	S	R	E	S	.	.	.	N	D	100	
Pt_736296	T	K	O	L	E	K	D	Y	E	V	L	K	R	O	F	E	D	A	L	K	A	D	N	D	A	L	O	A	O	N	K	I	H	A	E	.	L	L	S	L	K	S	R	E	S	.	.	.	N	E	100	
Gh-ABO4774	T	K	O	L	E	K	D	Y	D	A	L	K	R	O	F	E	A	L	K	A	D	N	D	A	L	O	A	O	N	K	I	N	A	E	.	L	L	A	L	K	I	K	S	.	.	.	N	E	100			
Dc-05625.1	T	K	O	L	E	K	D	Y	D	L	L	K	S	O	F	E	A	V	K	A	E	N	D	S	L	O	S	H	N	O	K	I	H	A	Q	.	I	M	A	L	K	S	R	E	P	.	.	.	T	E	100	
Dc-05623.1	T	K	O	L	E	K	D	Y	D	L	L	K	S	O	F	E	S	L	K	A	E	N	D	S	L	O	S	R	N	K	A	I	H	A	Q	.	I	L	A	L	K	S	R	E	P	.	.	.	T	D	100	
At_ATHB3	T	K	O	L	E	K	D	Y	D	S	L	K	R	O	F	E	V	L	K	S	D	N	D	S	L	L	A	H	N	K	I	H	A	E	.	L	V	A	L	K	S	R	E	P	.	.	.	K	E	100		
Sb-XP24672	T	K	O	L	E	K	D	Y	D	A	L	R	R	O	L	D	A	R	A	E	N	D	T	L	S	H	N	K	I	O	A	E	.	I	M	A	L	K	G	G	G	.	.	.	G	G	100					
Zm_4124075	T	K	O	L	E	K	D	Y	D	A	L	R	R	O	L	D	A	R	A	E	N	D	A	L	S	H	N	K	I	O	A	E	.	I	M	A	L	K	G	G	G	.	.	.	G	G	100					
Os_10g2650	T	K	O	L	E	K	D	F	D	A	L	R	R	O	L	D	A	R	A	E	N	D	A	L	S	L	N	S	K	I	H	A	E	.	I	V	A	L	K	G	G	A	.	.	.	A	A	100				
Sb-XP24657	T	K	O	L	E	K	D	Y	D	A	L	R	R	O	L	D	A	V	K	A	D	N	D	A	L	S	H	N	K	I	O	A	E	.	I	L	A	L	K	G	G	R	E	.	.	.	A	G	100			
Zm_4113431	T	K	O	L	E	K	D	Y	D	A	L	R	R	O	L	D	A	V	K	A	D	N	D	A	L	S	H	N	K	I	O	A	E	.	I	L	S	L	K	G	R	E	A	.	.	.	G	G	100			
Zm_413795	T	K	O	L	E	K	D	Y	D	A	L	R	R	O	L	D	A	V	K	A	D	N	D	A	L	S	H	N	K	I	O	A	E	.	I	L	X	L	K	G	R	E	A	.	.	.	A	G	100			
Os_03g0745	T	K	O	L	E	K	D	Y	D	A	L	R	R	O	L	D	A	V	K	A	E	N	D	A	L	N	H	N	K	I	O	A	E	.	I	V	A	L	K	G	R	E	A	.	.	.	A	S	100			
Ta-TaHZI-2	T	K	O	L	E	K	D	Y	D	V	L	K	R	O	F	E	A	V	K	A	E	N	D	A	L	S	H	N	K	I	O	S	E	.	I	L	G	L	K	E	R	E	.	.	.	A	A	100				
Zm_422699	T	K	O	L	E	K	D	Y	D	A	L	R	R	O	L	D	A	R	A	E	N	D	A	L	S	H	N	K	I	O	T	E	.	D	L	R	L	R	E	R	P	O	.	.	.	L	K	100				
At_ATHB20	T	R	O	L	E	K	D	Y	D	S	L	K	R	O	F	E	S	L	K	S	D	N	A	S	L	L	A	N	K	I	L	A	E	.	V	M	A	L	K	N	R	E	C	.	.	.	N	E	100			
Gm-ACU1987	T	K	O	L	E	K	E	Y	E	V	L	K	R	O	F	E	A	V	K	A	D	N	D	S	L	K	A	O	N	O	K	I	H	A	E	.	L	O	T	L	K	S	R	D	.	.	.	S	E	90		
At_ATHB6	T	K	O	L	E	K	D	Y	G	V	L	K	T	O	F	E	S	L	R	H	N	F	D	S	L	R	R	D	N	E	S	L	O	E	.	I	S	K	L	K	A	K	I	N	.	.	.	G	G	100		
Brs-hb-6	T	K	O	L	E	K	D	Y	G	V	L	K	T	O	F	E	S	L	R	H	N	F	D	S	L	R	R	D	N	E	S	L	O	E	.	I	G	K	L	K	A	K	I	N	.	.	.	G	E	100		
Bn-AAR0493	T	K	O	L	E	K	D	Y	G	V	L	K	T	O	F	E	S	L	R	H	N	F	D	S	L	R	R	D	N	E	S	L	O	E	.	I	G	K	L	K	A	K	I	N	.	.	.	G	E	100		
At_ATHB16	T	K	O	L	E	K	D	Y	G	V	L	K	T	O	F	E	S	L	R	H	N	F	D	S	L	R	R	D	N	E	S	L	O	E	.	I	S	K	I	K	A	K	V	N	.	.	.	G	E	100		
Vv-XP22660	T	K	O	L	E	R	D	Y	G	I	L	K	A	N	Y	E	T	L	K	L	N	Y	D	A	I	O	H	D	N	E	A	L	K	E	.	I	R	E	L	K	S	R	L	N	.	.	.	E	E	100		
Pt_HD56	T	K	O	L	E	R	D	Y	G	V	L	K	A	N	Y	E	S	L	K	H	N	F	D	A	I	O	O	D	N	E	A	L	K	E	.	I	R	E	L	K	A	K	I	N	.	.	.	E	E	100		
Pt_70493	T	K	O	L	E	R	D	Y	G	V	L	K	A	N	Y	E	S	L	K	H	N	F	D	A	I	O	O	D	N	E	A	L	K	E	.	I	R	E	L	K	A	K	I	N	.	.	.	E	E	100		
Gm-184277-	T	K	O	L	E	R	D	Y	G	V	L	K	A	N	Y	E	S	L	K	L	N	Y	D	T	L	O	O	D	H	E	A	L	K	E	.	I	K	E	L	K	S	R	L	V	.	.	.	Q	E	100		
Pv-HDZ1	T	K	O	L	E	R	D	Y	G	V	L	K	A	N	Y	E	S	L	K	L	N	Y	D	T	L	O	O	D	H	E	A	L	K	E	.	I	K	E	L	K	S	R	L	.	.	.	L	Q	100			
Nt-Hfi22	T	K	O	L	E	R	D	Y	G	V	L	K	S	N	F	D	A	L	K	H	N	Y	E	S	L	K	H	D	N	E	A	L	K	E	.	I	L	E	L	K	S	R	V	.	.	.	T	E	100			
Ze-18171.1	T	K	O	L	E	R	D	Y	G	V	L	K	S	N	F	D	A	L	K	H	K	Y	E	S	L	K	O	D	N	E	S	M	V	K	.	I	K	E	L	K	S	K	L	.	.	.	E	E	100			
Sl-S1HDL1	T	K	O	L	E	R	D	Y	G	V	L	K	S	N	F	D	A	L	K	L	N	Y	D	S	L	H	D	K	D	S	L	N	O	.	I	K	M	L	K	S	R	L	N	.	.	.	E	K	100			
At_ATHB5	T	K	O	L	E	R	D	Y	G	V	L	K	S	N	F	D	A	L	K	R	N	R	D	S	L	O	R	D	S	L	G	O	.	I	K	E	L	K	A	K	I	N	.	.	.	V	E	100				
Cp-CPHB-4	T	K	N	L	E	R	D	Y	G	V	L	K	S	N	F	D	A	L	K	H	D	F	E	T	L	K	R	D	N	E	S	L	K	E	.	I	H	E	L	K	S	R	L	N	.	.	.	E	D	100		
Mt-ACJ8504	T	K	O	L	E																																															

Vv-XP22716	KKRRLLADDOVQFLERNFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Vv-CAN8396	KKRRLLADDOVQFLERNFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Rc-XP25115	KKRRLLADDOVQFLERNFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Pt_82244	KKRRLLADDOVQFLERNFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Dc-05624.1	KKRRLLKADDOIQFLERNFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Cp-CPHB-7	KKRRLLSDDOVRFLDKSFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARYK	54
Gm-AAAX9867	KKRRLLSDEOVKFLDKSFEV...E	NKLEPERKMIWLAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Mt-ACJ8510	KKRRLLSDDOVRFLDKSFEV...E	NKLEPERKTKLAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Cp-CPHB-6	KKRRLLSADDOVRYLERTFEV...E	NKLEPERKAKLAKDLGLGLOP	ROVAIWF	ONRRARYK	55
Sl-CAA6441	KKRRLLTDNOVOFLDKSFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Sm_18196	KKRRLLSDEOVKFLDKSFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARWK	54
Dc-05622.1	KKRRLLSDEOVKFLDKSFEV...E	NKLEPERKVO LAKDLGLGLOP	ROVAIWF	ONRRARYK	54
Ps-ABK2476	KKRRLLTLEOVRALPKNFEM...A	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pg-ABA5414	KKRRLLTLEOVRALPKNFEM...A	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Ps-ABK2449	KKRRLLTLEOVRALPKNFEM...A	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_Pphb2	KKRRLLTFDOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_Pphb1	KKRRLLTFDOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_31	KKRRLLTFDOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_154	KKRRLLTFDOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_77	KKRRLLTFDOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_4	KKRRLLTFDOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_65a	KKRRLLSFDVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_34	KKRRLLSFDVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_pphb6	KKRRLLTFDOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_Pphb5	KKRRLLSFEVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_Pphb8	KKRRLLSFEVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_28	KKRRLLSFDVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_65b	KKRRLLSFDVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_143	KKRRLLTLDVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_sca_35	KKRRLLTLDVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_Pphb7	KKRRLLSFDVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_Pphb9	KKRRLLSFDVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pp_Pphb5	KKRRLLTLEOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Sm_81609	KKRRLLSDDOVRSLERHFEV...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Os_09g2118	KKRRLLTAEVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Zm_4124691	TKRRLLTAEVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	57
Ps-ABK2462	KKRRLLTFQOVKRLERKSEV...A	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Ps-ABK2572	KKRRLLSVOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Sm_19476	KKRRLLSVOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Sm_18217	KKRRLLSVOVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
At_ATHB22	KKKKMSSEOLKFLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	62
At_ATHB51	KKRRLLSSEOLKFLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_93443	KKRRLLSSEOLKFLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Vs-tendrill	KKNLKTSNOVDALERSEHE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Cr-CRHB11	KKRRLLTAEVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Cr-CRHB6	KKRRLLTAEVRSLELNFEM...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Cr-CRHB4	KKRCGLTVDVRYLERNFEV...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
At_ATHB7	NOKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	56
At_ATHB12	NOKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Vv-XP22629	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	56
Vv-CAN7896	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	56
Rc-XP25299	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Pt_HB7	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Pt_731421	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Gm-ACU2443	IKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Mt_MTHB1	NTKRRFTEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Zm_hox6	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Vv-XP22715	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Sb-XP24627	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Zm_4496	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Os_09g3591	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Zm_4492	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_HBLZ	NTRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Ha_HAHB11	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	56
Sb-XP24468	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Zm_433493	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Os_04g4581	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Sb-XP24528	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Os_02g4333	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Zm_479999	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	55
Zm_470295	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	57
Vv-XP22800	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_548258	YKRRFTEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_343725	NKRRFSDEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Ha_HAHB4	GKRRFTEOIKSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
At_ATHB21	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
At_ATHB40	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_655260	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_703426	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
At_ATHB53	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Os_03g1021	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Zm_480132	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Zm_hox12	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Os_07g3932	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Zm_433132	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Zm_433210	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
At_ATHB54	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
At_ATHB52	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Gm-ACU198B	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_568845	NKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_98386	NKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_594622	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54
Pt_unknon	KKRRFTEOIRSLERSEFE...E	NKLEPERKMO LAKALGLGLOP	ROVAIWF	ONRRARWK	54

Apéndice

Vv-XP22716	KOLEKDFGALQASVNSIKAEYENLLEKDEFKTE.VILITDKLL...VK	100
Vv-CAN8396	KOLEKDFGALQASVNSIKAEYENLLEKDEFKTE.VILITDKLL...VK	100
Rc-XP25115	KOLEKDYDVLQTSVNSIKADYDALLOEDRDKAE.VNLLITDKLL...LK	100
Pt_88244	KOLEKDYEVLLQSSVNGTKADYDNLFKKEKTKAE.VNLLITNELL...LK	100
Dc-05624.1	KOLEKDYDVLQTSVNSIKADYDNLFLAEDRDKAE.VLDLITDKLL...LK	100
Cp-CPHB-7	KOLEKDYDALKSNYDRKEDFDALYSENEKTKTE.VNLTAEKLL...GK	100
Gm-AAx9867	KOLEKDYDVLQTSVNSIKADYDNLFLAEDRDKAE.VARLTKVLL...GR	100
Mt-ACJ8510	KOLEKDYDVLQTSVNSIKADYDNLFLAEDRDKAE.VASLTKVLL...ER	100
Cp-CPHB-6	KOLEKDYDVLQTSVNSIKADYDNLFLAEDRDKAE.VASLTKVLL...ER	100
Sl-CAA6441	KOLEKDYDELNRNRYDTLKSNNLLEKDEFDRTE.VFRITGKLF...IK	100
Sm_18196	KOLEKDYDVLQTSVNSIKADYDNLFLAEDRDKAE.V...FC	90
Dc-05622.1	KOLEKDYDVLQTSVNSIKADYDNLFLAEDRDKAE.V...FC	100
Ps-ABK2476	KOLEKDFNVLKQDYDALKQDYDNLMEENNLQAM.VGNYTLFLF...IL	100
Pg-ABA5414	KOLEKDFNVLKQDYDALKQDYDNLMEENNLQAM.IERMSSKSQ...SC	100
Ps-ABK2449	KOLEKDFNVLKQDYDALKQDYDNLMEENNLQAM.IERMSSKSQ...SC	100
Pp_Pphb2	KOLEKDYEVLLTLDVNRKSEFEAVLOEQELOGE.HIVNFGOME...98	98
Pp_Pphb1	KOLEKDYEVLLTLDVNRKSEFEAVLOEQELOGE.MECTTEKIQ...TA	100
Pp_sca_31	KOLEKDYEVLLTSDVNRKSEFEAVLOEQELOGE.IECTTGKIQ...IS	100
Pp_sca_154	KOLEKDYEVLLSLDYNOKNKFFDDVVOEQELOGE.MDCLRGKLF...TP	100
Pp_sca_77	KOLEKDYEVLLNLDVNRKSEFEAVLOEQELOGE.VKTLKEKSP...MP	100
Pp_sca_4	KOLEKDYESLTSVNGTKADYDNLFLAEDRDKAE.VERTKQLQ...TV	100
Pp_sca_65a	KOLEKDYETLQDYNRKSDFFAVLKDKNLKEDE.PASPAQSEK...SD	100
Pp_sca_32	KOLEKCDYDALQTSVNSIKADYDNLFLAEDRDKAE.VNRLKGIAP...EV	100
Pp_pphb6	KOLEKDYEALAAADKSTKHDYDLVLAENNLKAE.VORLSGKAP...TS	100
Pp_Pphb8	KOLEKDYEFELNSGYSKLKRDFEKVLEAVLKAE.LVRLSAKIL...PK	100
Pp_sca_28	KOLEKDYEFELNSGYLKLKVEFETALRDEFKAE.VORLSGKTS...SQ	100
Pp_sca_65b	KOLEKDYEMLNSGYLKLKADYFETALRDEFKAE.VORLSGKTS...PO	100
Pp_sca_143	KOLEKDYETLAADYKTLMADYEHVVEVNCIRAE.VVRLTGGETP...PS	100
Pp_sca_35	KOLEKDYETLEAGFKRKADYELVLDVNYLKAEO.LQRLSGDT...98	98
Pp_Pphb7	KOLEKDYETLKAAYDRKADFEAVLDTSAKAE.VSRKGISN...DD	100
Pp_Pphb9	KOLEKDYGALAKDYNRKSEFEAVSRDNGYKAE.VNRLKGIAP...DD	100
Pp_Pphb5	KOLEKDYETLESYDRKADYFEOVLSBNHKAENNLKAE.FTCVDSNMF...AL	100
Sm_81609	KOLEKDYDALKENLDRGDKYKSLKKEOELEAE.VCOTFFLFFV...LS	100
Os_09g2118	KOLEKDFDRLRAAHDDELLAGRTALAAENLRSQ.VILITTEKLO...AN	100
Zm_4124691	KOLEKDFDRLRAAHDDELLAGRTALAAENLRSQ...92	92
Ps-ABK2462	KOLEKDYDLDLQKQYDDEKKNYDILLQENKHFKAERLNRESGND...DQ	100
Ps-ABK2572	KOLEKDYDLDLQKQYDDEKKNYDILLQENKHFKAERLNRESGND...DQ	100
Sm_19476	KOLEKDYESLTSVNGTKADYDNLFLAEDRDKAE.V...AA	100
Sm_18217	KOLEKDYDVLKAAVESHA...ENKRLKAO.LSDKLLVDGDHEKK...96	96
At_ATHB22	NKOLEHLYESLROEFDITVSRKELLOEELIQKSM.IREQ...101	101
At_ATHB51	KOLEHLYDLSLRQEVVSRKQMLHDEVKLRAL.LRDQGLIKK...QI	100
Pt_93443	KOLEHLYDNLKQEFDSVSKKQKLOEVMKIKAV.VREQATRQK...VS	100
Vs-tendril	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...86	86
Cr-CRHB11	NKOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Cr-CRHB6	NKOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Cr-CRHB4	NKOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
At_ATHB7	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
At_ATHB12	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Vv-XP22629	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Vv-CAN7896	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Rc-XP25299	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Pt_HB7	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Pt_731421	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Gm-ACU2443	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Mt_MTHB1	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Zm_hox6	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Vv-XP22715	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Sb-XP24627	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Zm_4496	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Os_09g3591	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Zm_4492	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Pt_HBLZ	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Ha_HAHB11	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Sb-XP24468	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Zm_433493	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Os_04g4581	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Sb-XP24528	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Os_02g4333	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Zm_479999	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Zm_470295	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Vv-XP22800	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Pt_548258	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Pt_343725	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
Ha_HAHB4	KOLEHLYDNLKQEF...NOKIQD.VMTLKEK...96	96
At_ATHB21	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	100
At_ATHB40	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	100
Pt_655260	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
Pt_703426	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	100
At_ATHB53	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
Os_03g1021	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
Zm_480132	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
Zm_hox12	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
Os_07g3932	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
Zm_433132	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
Zm_433210	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
At_ATHB54	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	93
At_ATHB52	NKRLHEEYTKLKNAYETTVVEKCR.LDSVIHKEQ.LYEAEREIQ...RL	103
Gm-ACU198B	TQSLHEVQHTLQSKHAAALSDKAKLEHVOVFLQDE.LKRRARNQLA...LF	100
Pt_568845	TQSLHEVQHTLQSKHAAALSDKAKLEHVOVFLQDE.LKRRARNQLA...LF	100
Pt_98386	TQSLHEVQHTLQSKHAAALSDKAKLEHVOVFLQDE.LKRRARNQLA...LF	100
Pt_594622	TQSLHEVQHTLQSKHAAALSDKAKLEHVOVFLQDE.LKRRARNQLA...LF	100
Pt_unknon	TQSLHEVQHTLQSKHAAALSDKAKLEHVOVFLQDE.LKRRARNQLA...LF	100

### V.3 Árbol filogenético HZ+OG



**Figura V.2:** Árbol filogenético HZ+OG. Fue realizado con los dominios HD-Zip de 178 proteínas de la subfamilia I y tres proteínas de la subfamilia II gracias a las cuales se posicionó la raíz del árbol.

## V.4 Tablas del análisis de enriquecimiento de términos GO

Las tablas que se encuentran a continuación presentan los resultados del análisis de términos GO de cada MAT. La primer columna contiene el código de dos letras (BP, MF o CC) que indica el tipo de terminología GO que figura bajo él, luego aparecen los términos MICA que resumen la función de los términos GO que anteceden y finalmente el identificador o rótulo del término GO (GOID). La segunda columna posee una descripción del término GO y la tercera los genes pertenecientes al término que forman parte del MAT.

ATHB-12		
MICA/GOID	Término GO	Genes
<b>BP</b>		
	response to abiotic stimulus	
GO:0009414	response to water deprivation	LEA14 NF-YA5 ATCOR413-PM1 ATHB6 ATHB-7 AT3G02480 AHG3 ATNCED3 BAM1 ATHB-12 ANAC072 ABF3 COR78 ABI2 HAI1 ATMYB96 AAP4 ABA1
GO:0009409	response to cold	ATCOR413-PM1 AHG3 ATICE1 ABI1 LEA4-5 COR27 COR78
	regulation of abscisic acid mediated signaling pathway	
GO:0009788	negative regulation of abscisic acid mediated signaling pathway	ATHB6 AHG3 ABI1 ABI2
	vitamin E metabolic process	
GO:0010189	vitamin E biosynthetic process	HPD G-TMT AT5G53970
	positive regulation of gene expression	
GO:0045941	positive regulation of transcription	ATHAP5B ATHB6 ATICE1 ATHB-12
	transport	
GO:0006857	oligopeptide transport	YSL8 AT2G02020 AT2G37900 AT3G21670
GO:0006865	amino acid transport	AAP1 AT2G41190 AAP4 AT5G65990
<b>MF</b>		
	transcription regulator activity	
GO:0016563	transcription activator activity	ABF1 ATHAP5B ATHB6 ATHB-7 ATICE1 ATHB-12 ANAC072 ABF3
	active transmembrane transporter activity	
GO:0015171	amino acid transmembrane transporter activity	AAP1 AT2G41190 AAP4 AT5G65990
GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	AT1G72700 AT3G27870 ACA8
	DNA binding	
GO:0043565	sequence-specific DNA binding	ATHB13 ATHB6 GBF3 ATMBD11
	phosphoprotein phosphatase activity	
GO:0004722	protein serine/threonine phosphatase activity	HAI2 AHG3 ABI1 ABI2 HAI1
<b>CC</b>		
	protein complex	

## Apéndice

ATHB-12			
MICA/GOID	Término GO	Genes	
GO:0008287	protein serine/threonine phosphatase complex	HAI2	ABI1 ABI2

ATHB-7			
MICA/GOID	Término GO	Genes	
BP			
regulation of abscisic acid mediated signaling pathway			
GO:0009788	negative regulation of abscisic acid mediated signaling pathway	ANAC002	AHG3 ABI1 ABI2
organic substance transport			
lipid transport			
GO:0006869	lipid transport	AT1G48750	AT1G72700 AT2G37870 ATLTP1 AT2G45180 AT3G27870 AT4G33550 LTP4
GO:0006865	amino acid transport	AT1G31820	AAP1 LAX3 AT2G41190 AT5G65990
response to abiotic stimulus			
GO:0009269	response to desiccation	ALDH7B4	CLO-3 ATRD22
GO:0009631	cold acclimation	ERD10	ATCAX1 ATDI8
MF			
dioxygenase activity			
GO:0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	ATLOX1	ALPHA CLO-3 ARD4
transcription regulator activity			
GO:0016563	transcription activator activity	ANAC002	ATHB-7 ATICE1 ATHB-12 ANAC072 ABF3 KNAT4
amine transmembrane transporter activity			
GO:0015171	amino acid transmembrane transporter activity	AT1G31820	AAP1 LAX3 AT2G41190 AT5G65990
CC			
organellar ribosome			
GO:0009547	plastid ribosome	RPS9	CS17 RPL9

ATHB-5			
MICA/GOID	Término GO	Genes	
BP			
response to abiotic stimulus			
GO:0080167	response to karrikin	AT2G16660	AT2G41040 AT3G61220 ATGPX7 ATOPT4



## Apéndice

ATHB6 MICA/GOID	Término GO	Genes
<b>BP</b>		
response to abiotic stimulus		
GO:0009414	response to water deprivation	AtGolS2 ERD14 ATCOR413-PM1 ATHB6 AT3G05640 AHG3 ATNCED3 BAM1 ATHB- 12 ATCBF3 ATCBF1 COR78 ABI2
GO:0009631	cold acclimation	ATCOR413-PM1 ATCBF2 ATCBF3 ATCBF1
GO:0009408	response to heat	AT4G17250 ABI1 APG6 ABI2
response to organic substance		
GO:0009788	negative regulation of abscisic acid mediated signaling pathway	ATHB6 AHG3 ABI1 ABI2
GO:0010200	response to chitin	ATWRKY6 AT1G64380 AT-HSFB2B AT4G28140
carbohydrate metabolic process		
GO:0016051	carbohydrate biosynthetic process	ATGSL06 AtGolS2 APL4 AT2G22190 GATL2 F8H
GO:0006073	cellular glucan metabolic process	ATGSL06 APL4 BAM1 XTH15
dephosphorylation		
GO:0006470	protein dephosphorylation	HAI2 ABI1 ABI2
<b>MF</b>		
phosphoprotein phosphatase activity		
GO:0004722	protein serine/threonine phosphatase activity	HAI2 AT3G05640 AHG3 ABI1 AT5G27930 ABI2
transcription regulator activity		
GO:0016563	transcription activator activity	ATHB6 ATHB-12 ATCBF2 ATCBF3 ATCBF1
<b>CC</b>		
protein complex		
GO:0008287	protein serine/threonine phosphatase complex	HAI2 ABI1 ABI2
intrinsic to plasma membrane		
GO:0046658	anchored to plasma membrane	COBL2 AtENODL2 AT4G29360

ATHB-16 MICA/GOID	Término GO	Genes
<b>BP</b>		
photosynthesis, light reaction		
GO:0009773	photosynthetic electron transport in photosys- tem I	HCEF1 ATPD PSAN
GO:0042549	photosystem II stabilization	OEC33 STN8
chlorophyll metabolic process		
GO:0015995	chlorophyll biosynthetic process	AT1G74470 CLA CHLM ABAR
<b>MF</b>		

## Apéndice

ATHB-16			
MICA/GOID	Término GO	Genes	
transaminase activity			
GO:0008453	alanine-glyoxylate transaminase activity	AOAT1 AGT	
transition metal ion binding			
GO:0030145	manganese ion binding	PSBY AT3G04200 GLP9	
acyltransferase activity			
GO:0016407	acetyltransferase activity	AT1G72030	AT4G28030 AT5G47720 AT5G67430
CC			
chloroplast thylakoid			
GO:0030095	chloroplast photosystem II	PSB27 AT1G51400 PSBY OEC33	
GO:0009543	chloroplast thylakoid lumen	PSB27 AT1G12250 AT1G51400 OEC33 PSAN	
chloroplast stroma			
GO:0010287	plastoglobule	AT1G32220 AT1G78140 OEC33 ATPD	
plastid envelope			
GO:0010319	stromule	HCEF1 ATPD EMB2728	

ATHB-1			
MICA/GOID	Término GO	Genes	
BP			
alkaloid metabolic process			
GO:0009821	alkaloid biosynthetic process	AT1G35190 ATSSL4 SSL5 AT3G51440 AT3G51450 AT3G57010	
macromolecule localization			
GO:0045036	protein targeting to chloroplast	ATTOC33 ABA3 PSB29 TIC55-II CPHSC70- 2EAT HCF106 CIA2	
GO:0065002	intracellular protein transmembrane transport	ABA3 AT1G36390 PSB29 HCF106	
GO:0015914	phospholipid transport	AT1G17500 AT1G72700 LP2	
cellular carbohydrate metabolic process			
GO:0005986	sucrose biosynthetic process	AT1G43670 FLN2 FLN1 ATSUS3 ASUS1	
GO:0019575	sucrose catabolic process, using beta-fructofuranosidase	AT1G62660 FLN2 FLN1 AT4G34860	
GO:0005982	starch metabolic process	ATISA2 ATSS3 AT1G43670 LSF1 ATSUS3 BAM3	
plastid organization			
GO:0010027	thylakoid membrane organization	PTAC4 PSB29 FTSH2 PLSP1 ATOBGC	
GO:0009658	chloroplast organization	ARASP ATYLMG1-1 ATAB2 CHUP1 AtCNFU1 KAC1 AtCEST ATFTSZ1-1	
photosynthesis, light reaction			
GO:0009773	photosynthetic electron transport in photosystem I	NDF4 HCEF1 PGRL1B PGR5-LIKE	

## Apéndice

ATHB-1 MICA/GOID	Término GO	Genes
GO:0010206	photosystem II repair	TLP18.3 DEGP2 DEGP1
GO:0010207	photosystem II assembly	LPA19 HCF173 PSB29
	response to bacterium	
GO:0042742	defense response to bacterium	ABA3 ESP ATLOX1 AT1G77920 ATWRKY40 ATWRKY17 ACA4 AT3G04790 ESM1 ATERF-4 ATERF011 AT4G32260 ATNPR3 RPS6 TGA1
MF		
	amine-lyase activity	
GO:0016844	strictosidine synthase activity	ATSSL4 SSL5 AT3G51440 AT3G51450 AT3G57010
	oxidoreductase activity	
GO:0008794	arsenate reductase (glutaredoxin) activity	AT4G15660 AT4G15680 AT4G15690 AT4G15700
GO:0004033	aldo-keto reductase activity	AT1G06690 AT2G27680 AT5G53580
GO:0016701	oxidoreductase activity, acting on single do- nors with incorporation of molecular oxygen	ATLOX1 ALPHA TIC55-II AT4G15093 AT5G64250
GO:0015035	protein disulfide oxidoreductase activity	AT1G06830 AT4G15660 AT4G15680 AT4G15690 AT4G15700
CC		
	chloroplast thylakoid	
GO:0009535	chloroplast thylakoid membrane	AT1G06690 PQL1 AT1G14345 AT1G50020 TLP18.3 PTAC4 PSAE-2 PSB29 AT2G24020 AT2G26340 AT2G27290 FTSH2 EDA3 AT2G35490 DEGP2 AT3G04790 NDF4 AT3G17930 PLSP1 AT3G25480 DEGP1 AT3G47070 AT3G51510 AT3G53470 PSRP5 AT4G01150 AT4G02530 ATPC1 PGRL1B PTAC5 PGR5-LIKE AT4G23890 PSB28 AT4G32260 AT5G03880 AT5G08050 AT5G13410 AT5G14910 AT5G37360 APE1 NDH18 AtCEST HCF106 AT5G58260
GO:0009543	chloroplast thylakoid lumen	LPA19 PQL1 TLP18.3 AT2G26340 AT2G40400 AT2G43560 DEGP1 AT4G02530 AT5G13410
	plastid nucleoid	
GO:0009508	plastid chromosome	PTAC4 ATWHY3 HMR PTAC13 PTAC5
GO:0042644	chloroplast nucleoid	ATNTH2 FLN2 FLN1
	plastid envelope	
GO:0009527	plastid outer membrane	ATTOC33 PSB29 ATSFR2 CHUP1 CRL
GO:0010319	stromule	ATHM1 AT1G67280 PSB29 PGK1 HCEF1
	extracellular region	

## Apéndice

---

ATHB-1 MICA/GOID	Término GO	Genes
GO:0048046	apoplast	ATSBT1.1 ATHM1 AT1G09310 AT1G09750 AT1G29660 AGT PGK1 ESM1 AT3G16370 HCEF1 CSP41A RALFL32 AT4G20830 AT5G03350 CPHSC70-2EAT AT5G58330 XTH6

---

---

ATHB51 MICA/GOID	Término GO	Genes
BP		
response to hormone stimulus		
GO:0009733	response to auxin stimulus	AT1G19830 AT1G29440 AT1G29450 AT1G29460 AT1G29490 SAUR68 AT1G48690 AT1G56150 AT1G75590 GH3.1 AT3G03830 AT3G03840 AT3G25880 AT3G53250 AT4G13790 AT4G34760 ATSAUR15 AtMYB76
MF		
tetrapyrrole binding		
GO:0020037	heme binding	CYP87A2 CYP705A25 CYP705A9 AT2G35380 AT3G03670 CYP72A7 CYP72A11 CYP705A20 CYP705A33 CYP71B17 CYP71B35 AT4G26010 AT5G19880 AT5G19890 CYP86B1 CYP71B14 AT5G58390
CC		
plastid thylakoid membrane		
GO:0009535	chloroplast thylakoid membrane	AT1G52220 AT1G54500 TLP18.3 PSAG AT1G68590 AT2G01870 AT2G26340 AT2G26500 AT3G15110 NDF4 AT3G17930 AT3G63160 AT4G01150 PSAD-1 PETC LHCB5 PSAL AT4G28025 AT4G32260 LHCB4.1 AT5G03880 PSBH

---

---

ATHB13 MICA/GOID	Término GO	Genes
BP		
response to abiotic stimulus		

---

## Apéndice

ATHB13		
MICA/GOID	Término GO	Genes
GO:0009414	response to water deprivation	ERD6 AT-GTL1 ATCOR413-PM1 AHK1 ATHH2 AHG3 ATHB-12 ATPLC ATMYB96 AAP4
GO:0080167	response to karrikin	RSW10 AGP26 ATGSTU27 ATMYB94 AMT1;1 AT4G35110 AT5G57785
	cellular response to stimulus	
GO:0043562	cellular response to nitrogen levels	PLDALPHA4 PDLZ2 PYD1
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	AT1G10850 RKF1 RKL1 ATCLV1 AT3G28040 BAM2 BAM3
	small molecule metabolic process	
GO:0009112	nucleobase metabolic process	HGPT APT2 PYD1
GO:0009098	leucine biosynthetic process	ATLEUD1 IPMI1
	transmembrane transport	
GO:0080170	hydrogen peroxide transmembrane transport	GAMMA-TIP GAMMA-TIP2
	MF	
	hydrolase activity, acting on ester bonds	
GO:0004620	phospholipase activity	AT1G49740 PLDALPHA4 AT2G30550 PDLZ2 ATPLC
GO:0008081	phosphoric diester hydrolase activity	AT1G49740 PLDALPHA4 PDLZ2 ATPLC
	lyase activity	
GO:0016835	carbon-oxygen lyase activity	AT1G04680 ATLEUD1 ATCYSD1 AT3G53190 IPMI1 AT4G02610
	water transmembrane transporter activity	
GO:0015250	water channel activity	GAMMA-TIP ATHH2 GAMMA-TIP2
	CC	
	intrinsic to membrane	
GO:0031225	anchored to membrane	AtENODL8 AGP8 ATHH2 AGP26 AT3G06035 AT3G53190 AT4G16140 AT5G19250 AT5G26330 AT5G42720 FLA12 AT5G64080

AtHB23		
MICA/GOID	Término GO	Genes
	BP	
	catabolic process	
GO:0010304	PSII associated light-harvesting complex II catabolic process	FTSH8 FTSH1 FTSH2 NYC1 NOL FTSH5
GO:0019464	glycine decarboxylation via glycine cleavage system	AT1G32470 AT1G60990 AtGLDP2 SHM1
GO:0006096	glycolysis	GAPA-2 GAPB ATHXK4 GAPA AT5G56350 PFK7 AT5G56720 AT5G58330 AT5G63680

## Apéndice

AtHB23 MICA/GOID	Término GO	Genes
	photosynthesis, light reaction	
GO:0009765	photosynthesis, light harvesting	PSB27 PSAO FTSH1 TLP18.3 LHCA3 LHCA1 LUT2 PSBH
GO:0010207	photosystem II assembly	LPA1 HCF173 PSBR PSB29 MSP-1
GO:0009767	photosynthetic electron transport chain	PSAG ATFD2 PGR5 FTSH2 PETC PGRL1B FTSH5 PSAN MSP-1
	response to light stimulus	
GO:0010196	nonphotochemical quenching	CP24 NPQ4 PETC LHCB5
GO:0009644	response to high light intensity	ATMYBL2 PGR5 FTSH2 CAO AT-HSFA7A ATSDX1 SEP1 OHP FTSH5 MSP-1
	enzyme linked receptor protein signaling pathway	
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	AT1G14390 AT1G25320 RKF1 AT1G34110 RKL1 AT1G75640 BRL2 AT2G25790 AT3G03770 AT3G24240 AT3G28040 BAM2 AT4G28650 AT4G36180 AT4G39270 AT5G24100 AT5G41180 AT5G45780 AT5G48940 AT5G63930
	cellular metabolic process	
GO:0016120	carotene biosynthetic process	Z-ISO PDE181 PDE226 LUT2
GO:0016123	xanthophyll biosynthetic process	CYP97A3 B1 LUT2
GO:0015996	chlorophyll catabolic process	ACD1 NYC1 NOL ATCLH2
GO:0015995	chlorophyll biosynthetic process	ALB-1V AT1G74470 ATG4 ACSF CH-42 PORB ABAR
GO:0010189	vitamin E biosynthetic process	ATHPT ATSDX1 AT5G53970
	response to metal ion	
GO:0046688	response to copper ion	DRT112 HMA5 CSD2 ATLAC15
	DNA conformation change	
GO:0006265	DNA topological change	ATSP011-1 ATTOPII BIN5 TOP3A
	cellular homeostasis	
GO:0045454	cell redox homeostasis	AT1G07700 AT1G56500 AT2G30540 ATTRX AT3G53220 AT3G62950 AT3G62960 AT4G15700 AT4G31240 CAT2 GASA4 AT5G42150
	stem cell development	
GO:0019827	stem cell maintenance	PLT2 SAV3 AGL22 AGL20 PLT1 AGL24
	MF	
	tetrapyrrole binding	
GO:0016168	chlorophyll binding	CP24 NPQ4 LHCA3 AT1G76570 CAB4 LHCA1 LHCB5 SEP1 LHCB4.1 PSBA
	oxidoreductase activity	
GO:0008943	glyceraldehyde-3-phosphate dehydrogenase activity	GAPA-2 GAPB ALDH11A3 GAPA

## Apéndice

AtHB23 MICA/GOID	Término GO	Genes
GO:0004601	peroxidase activity	AT1G44970 AT2G18980 AT2G43480 AT3G01190 AT4G11230 AT4G11290 AT4G25090 AT4G26010 AT4G33420 CAT2 APX5 AT5G15180 AT5G51890 AT5G66390 RHS19
	carbon-carbon lyase activity	
GO:0016832	aldehyde-lyase activity	THA1 FBA1 FOLB1 AT3G56060 FBA2
	transition metal ion binding	
GO:0005506	iron ion binding	CYP77B1 CYP87A2 CYP71B28 AT1G35190 AT1G44970 AT2G18980 AT2G37700 AT2G43480 AT3G01190 CYP72A11 AT3G19000 CYP705A22 CYP71B24 CYP71B4 CYP71B36 AT4G11230 AT4G11290 CYP702A2 CYP705A3 AT4G16770 AT4G25090 AT4G26010 CYP709B3 AT4G33420 APX5 AT5G15180 CYP86B1 CYP81D1 AT5G58660 AT5G66390 RHS19
	endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 3'-phosphomonoesters	
GO:0016892	endoribonuclease activity, producing 3'-phosphomonoesters	AT1G14210 AT1G14220 AT5G59980
	receptor activity	
GO:0004888	transmembrane receptor activity	AT1G17600 AT1G17610 AT1G51270 AT1G56520 BRL2 AT3G04210 AT3G44630 AT4G19510 AT4G23510 AT5G40100 AT5G44870 AT5G45060 AT5G45200 AT5G45210 AT5G45220 AT5G46450 AT5G58120
	protein kinase regulator activity	
GO:0016538	cyclin-dependent protein kinase regulator activity	CYCA2;4 CYCB2;1 CYCB1;4 CKS2 CYC2 ICK3
	CC	
	chloroplast stroma	
GO:0010287	plastoglobule	CP24 PSAF AT1G52590 LHCA3 AT1G79600 PSAE-2 FBA1 AT2G35490 RCA CAB4 LHCA1 PSAD-1 LHCB5 PSAL AT4G31390 ATSDX1 FBA2 LHCB4.1 MSP-1 PSBA
	protein complex	
GO:0030076	light-harvesting complex	NPQ4 LHCA3 ABA4 AT1G76570 CAB4 LHCA1 LHCB5 LHCB4.1 LHCB3
GO:0009538	photosystem I reaction center	PSAD-2 PSAF PSAE-2 PSAH-1 PSAD-1 PSAL

## Apéndice

AtHB23		
MICA/GOID	Término GO	Genes
GO:0030095	chloroplast photosystem II	PSB27 OE23 AVDE1 AT1G51400 PSBY PSBQ MSP-1
GO:0009654	oxygen evolving complex extracellular region	OE23 AT3G05410 PSBQ MSP-1 PSBH
GO:0048046	apoplast	ATCPO-I OE23 AT1G11860 GAPA- 2 BGLU40 AT1G29660 AT1G29670 PRK GAPB ATHPR1 ATGSTU20 FBA1 CSD2 XTH32 AT2G37660 RCA ATPPC2 ANP3 AT3G14067 AT3G16530 ATPREP1 AT3G20820 GAPA RALFL27 ATXTR8 BGAL2 ROC4 PSBQ RALFL32 AtENODL2 SHM1 FBA2 EXGT-A4 ATGSL1 ATLAC15 ATBXL1 AT5G58330 BGAL6 MSP-1
	plastid thylakoid lumen	
GO:0009543	chloroplast thylakoid lumen	PSB27 OE23 AVDE1 DRT112 AT1G51400 TLP18.3 AT1G76450 AT3G12340 AT3G56140 AT3G57680 PSBQ AT4G17740 AT4G24930 AT5G11450 AT5G13410 PSAN MSP-1
	plastid envelope	
GO:0010319	stromule	AT1G16880 PRK GAPB PSB29 RCA GAPA CAT2 SHM1
GO:0009706	chloroplast inner membrane	COR314-TM2 ACD1 AT3G51140 ACSF MGD1 ATSDX1 ABAR ATOBGC

ATHB20		
MICA/GOID	Término GO	Genes
	BP	
	response to chemical stimulus	
GO:0006979	response to oxidative stress	AT1G30870 AT1G44970 ATH5 AT1G71695 ATKTI1 AT2G43480 AT3G01190 ALPHA- DOX1 AT4G11290 AT4G30170 AT5G06730 AT5G15180 AER AT5G16990 AT5G17820 AT5G42180 AT5G64100 AT5G66390
GO:0009636	response to toxin	ATH5 HMA5 ATMTP3
	MF	
	oxidoreductase activity, acting on peroxide as acceptor	



## Apéndice

ATHB20		
MICA/GOID	Término GO	Genes
GO:0004601	peroxidase activity	RCI3 AT1G30870 AT1G44970 AT1G71695 AT2G37130 AT2G43480 AT3G01190 AT4G11290 AT4G25090 AT4G30170 AT5G06730 AT5G15180 AT5G17820 AT5G42180 AT5G64100 AT5G66390
	metal ion binding	
GO:0020037	heme binding	AT1G30870 AT1G44970 AT1G71695 CYP705A8 AT2G37130 AT2G43480 AT3G01190 CYP705A22 AT4G11290 CYP705A3 AT4G30170 AT5G06730 AT5G15180 AT5G17820 AT5G42180 AT5G64100 AT5G66390
GO:0000287	magnesium ion binding	AT2G35390 AT2G35840 AT3G14490 AT3G58140 AT4G20210 AT4G20230
	enzyme regulator activity	
GO:0004857	enzyme inhibitor activity	ATKTI1 AT2G45220 AT2G47550 RHS12 AT3G47400 PME61 UNE11 SIM AT5G43580 AT5G62340
	binding	
GO:0030246	carbohydrate binding	AT1G61480 ATPP2-A3 AT3G01260 AtGH9C3 AT5G03700 AT5G06740 AT5G35370 AtPP2-A8 AtPP2-A6
	hydrolase activity, acting on ester bonds	
GO:0004091	carboxylesterase activity	AT1G29660 AT1G74460 AT1G75880 AT2G45220 AT2G47550 RHS12 AT3G27950 AT3G47400 PME61 AT3G62280 UNE11 AT4G19420 AT5G24230 AT5G62340
	CC	
	Golgi apparatus part	
GO:0017119	Golgi transport complex	BP80-3;3 BP80-3;2

ATHB-3		
MICA/GOID	Término GO	Genes
	BP	
	response to chemical stimulus	
GO:0006979	response to oxidative stress	AT1G30870 ATKTI1 AT2G43480 AT3G01190 AT3G49960 AT4G11290 AT4G26010 AT4G30170 AER AT5G16980 AT5G16990 AT5G17820 RHS18 AT5G24070 AT5G42180 ATNUDT2 AT5G64100 AT5G66390 RHS19

## Apéndice

ATHB-3 MICA/GOID	Término GO	Genes
GO:0009636	response to toxin	ATGSTU26 HMA5 AtNIT4
GO:0009269	response to desiccation	RCI3 ALDH7B4 AT2G46140
	cellular catabolic process	
GO:0010304	PSII associated light-harvesting complex II catabolic process	FTSH8 NYC1 NOL
GO:0015996	chlorophyll catabolic process	ACD1 NYC1 NOL
	cell wall organization or biogenesis	
GO:0016998	cell wall macromolecule catabolic process	AT1G02360 AT1G56680 AT2G43600 AT2G43610
GO:0009664	plant-type cell wall organization	AT1G23720 AT3G54580 ATHRGP1 AT5G35190 ATEXP14 BGAL6
	root system development	
GO:0048829	root cap development	ANAC015 LAX3 ANAC070
	MF	
	oxidoreductase activity	
GO:0004601	peroxidase activity	RCI3 AT1G30870 AT2G43480 AT3G01190 AT3G49960 AT4G11290 AT4G25090 AT4G26010 AT4G30170 AT5G17820 RHS18 AT5G24070 AT5G42180 AT5G64100 AT5G66390 RHS19
GO:0050664	oxidoreductase activity, acting on NADH or NADPH, oxygen as acceptor	ATRBOHB AT4G25090 ATRBOHC
GO:0009815	1-aminocyclopropane-1-carboxylate oxidase activity	AT1G12010 AT1G77330 AT2G30840
	metal ion binding	
GO:0020037	heme binding	AT1G30870 AT2G43480 AT3G01190 CYP705A22 AT3G49960 AT4G11290 CYP702A2 CYP705A2 CYP705A3 AT4G26010 AT4G30170 CYP81D3 AT5G17820 RHS18 AT5G24070 AT5G42180 AT5G64100 AT5G66390 RHS19
GO:0000287	magnesium ion binding	AT1G33750 AT2G35840 AT3G29410 ATSP2 AT4G20210 AT5G48110
	pattern binding	
GO:0008061	chitin binding	AT1G56680 AT2G43600 AT2G43610
	hydrolase activity	
GO:0004568	chitinase activity	AT1G02360 AT1G56680 AT2G43600 AT2G43610
GO:0004185	serine-type carboxypeptidase activity	sepl31 sepl6 sepl3 SCPL7 SCPL34
GO:0004091	carboxylesterase activity	ATPME2 GLIP2 AT2G03980 sks16 AT2G42690 RHS12 AT3G27950 PME61 AT3G62280 AT4G13550 AT5G04960 AT5G24230 PGL4 AT5G62340

## Apéndice

ATHB-3		
MICA/GOID	Término GO	Genes
GO:0016892	endoribonuclease activity, producing 3'-phosphomonoesters	AT1G14210 AT1G14220
	structural molecule activity	
GO:0005199	structural constituent of cell wall	AT1G23720 AT3G54580 ATHRGP1 AT5G35190
	coenzyme binding	
GO:0051287	NAD binding	GAPA-2 FDH AT5G39320 ATNUDT2
	phospholipid transporter activity	
GO:0008526	phosphatidylinositol transporter activity	AT1G19650 COW1
	protein binding	
GO:0005516	calmodulin binding	AT2G24300 ATCNGC14 ATLTP1 ATMLO12 ATMLO15 iqd17 AT4G31320 IQD12 ATMLO10
	CC	
	plant-type vacuole	
GO:0042807	central vacuole	TIP4;1 ATTIP2;3

ATHB53		
MICA/GOID	Término GO	Genes
	BP	
	response to chemical stimulus	
GO:0006979	response to oxidative stress	CSD1 AT1G30870 AT1G32220 NTRC AT4G08780 AT4G26010 AT5G05340 AT5G06730 AT5G17820 ANAC092 AT5G44380
	programmed cell death	
GO:0006915	apoptosis	CW9 AT1G59780 AT1G69550 AT1G72850 AT3G04220 AT5G36930 AT5G46450
	cell wall organization or biogenesis	
GO:0071555	cell wall organization	ATEXP18 ADPG2 AT2G45220 RHS12 ATHRGP1 VTC2 AT5G04970 AT5G35190
	MF	
	oxidoreductase activity	
GO:0004601	peroxidase activity	AT1G30870 AT4G08780 AT4G26010 AT4G36430 AT5G05340 AT5G06730 AT5G17820
GO:0016730	oxidoreductase activity, acting on iron-sulfur proteins as donors	AT2G04700 ACD1
	tetrapyrrole binding	

## Apéndice

ATHB53		
MICA/GOID	Término GO	Genes
GO:0020037	heme binding	CYP87A2 AT1G30870 AT1G73340 CYP718 CYP72A7 AT4G08780 AT4G26010 AT4G36430 AT5G06730 AT5G17820 AT5G45040
	hydrolase activity	
GO:0004091	carboxylesterase activity	AT1G29670 AT1G51440 AT1G62770 AT2G03980 sks16 AT2G27360 AT2G45220 RHS12 AT3G48460 AT4G10950 CDEF1 AT5G04970
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	ATPDR4 ATPDR6 AT2G39350 ATABCC6 AT3G55090 AHA7 PGP2
	CC	
	extracellular region part	
GO:0005615	extracellular space	CDEF1 AT5G10130

ATHB40		
MICA/GOID	Término GO	Genes
BP		NA
	response to chemical stimulus	NA
GO:0006979	response to oxidative stress	AT1G30870 AT1G52200 AT1G71695 AT2G18980 AT2G35380 ATFER4 AT2G41480 AT3G01190 AT3G03670 AT3G21770 AT4G11290 PMSR4 AT4G26010 AT4G30170 AT5G17820 AT5G24070 AT5G42180 AT5G50350 AT5G64100
	triterpenoid metabolic process	NA
GO:0016104	triterpenoid biosynthetic process	HMG2 04C11 BARS1
	MF	NA
	oxidoreductase activity	NA
GO:0004601	peroxidase activity	RCI3 AT1G30870 AT1G71695 AT2G18980 AT2G35380 AT2G37130 AT2G41480 AT3G01190 AT3G03670 AT3G21770 AT4G11230 AT4G11290 AT4G25090 AT4G26010 AT4G30170 AT5G17820 AT5G24070 AT5G42180 AT5G64100
GO:0050664	oxidoreductase activity, acting on NADH or NADPH, oxygen as acceptor	ATRBOHB AT4G11230 AT4G25090 ATRBOHC
GO:0008471	laccase activity	LAC3 LAC13 ATLAC15
	tetrapyrrole binding	NA

## Apéndice

ATHB40					
MICA/GOID	Término GO	Genes			
GO:0020037	heme binding	CYP722A1	AT1G30870	AT1G71695	
		AT2G12190	CYP705A13	AT2G18980	
		CYP705A9	AT2G35380	AT2G37130	
		AT2G41480	AT3G01190	AT3G03670	
		CYP72A14	CYP71B37	CYP94B3	
		AT4G11290	CYP702A2	CYP705A2	
		AT4G26010	AT4G30170	AT5G08250	
		AT5G17820	AT5G24070	AT5G42180	
		AT5G64100			
ATHB21					
MICA/GOID	Término GO	Genes			
BP		NA			
response to water				NA	
GO:0009414	response to water deprivation	AT3G44380	ABF3	GORK	LTI65
MF		NA			
telomeric DNA binding				NA	
GO:0003691	double-stranded telomeric DNA binding	ATBP-1	ATTRP1		
ATHB52					
MICA/GOID	Término GO	Genes			
BP		NA			
lipid localization				NA	
GO:0006869	lipid transport	AT1G54280	AT1G66850	AT1G73780	
		AT3G07450	AT3G43720	AT2S2	AT5G07230
		AT5G62080			
cellular component organization or biogenesis at cellular level				NA	
GO:0080110	sporopollenin biosynthetic process	ACOS5	CYP704B1		
GO:0010411	xyloglucan metabolic process	AtXTH12	AtXTH13		
regulation of timing of meristematic phase transition				NA	
GO:0048510	regulation of timing of transition from vegetative to reproductive phase	DDF1	AGL19	ATHB51	
MF		NA			
oxidoreductase activity				NA	
GO:0004601	peroxidase activity	AT1G30870	AT1G34510	AT2G39040	
		AT3G03670	AT3G49960	AT4G25980	
		AT4G26010	AT5G19880	RHS18	

## Apéndice

ATHB52		
MICA/GOID	Término GO	Genes
GO:0016713	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced iron-sulfur protein as one donor, and incorporation of one atom of oxygen	CYP96A15 CYP704B1 CYP86
	tetrapyrrole binding	NA
GO:0020037	heme binding	AT1G30870 AT1G34510 AT2G39040 CYP718 AT3G03670 AT3G49960 CYP706A7 AT4G26010 CYP82C3 AT5G19880 RHS18 CYP71B14 CYP705A12
	molecular function	NA
GO:0045735	nutrient reservoir activity	AT3G04200 AT3G05950 AT2S2 AtPLAIVC AT5G39130
	hydrolase activity, hydrolyzing O-glycosyl compounds	NA
GO:0008810	cellulase activity	BG3 BG2
	glucosyltransferase activity	NA
GO:0080039	xyloglucan endotransglucosylase activity	AtXTH12 AtXTH13

ATHB-17		
MICA/GOID	Término GO	Genes
BP		
tetrapyrrole metabolic process		
GO:0006778	porphyrin metabolic process	HO3 ATFC-II HEME2 FLU ATG4 CHLM ACD2 NOL HEMC GSA1
GO:0033014	tetrapyrrole biosynthetic process	ATCPO-I ATFC-II HEME2 FLU ATG4 CHLM HEMC GSA1
cellular catabolic process		
GO:0019464	glycine decarboxylation via glycine cleavage system	AtGLDP2 GDCH SHM1
GO:0010304	PSII associated light-harvesting complex II catabolic process	FTSH1 NOL FTSH5
response to chemical stimulus		
GO:0009751	response to salicylic acid stimulus	PRO25 WAK2 RGL AT1G74840 ATMYB63 ATGSTU7 ATMYB38 CRK4 AtMYB109 ACD6 AtMYB18 ATMYB64 AtMYB49 ATMYB59
GO:0015893	drug transport	AT1G15170 ATPDR7 ATPDR11 AT2G16980 ATPDR5 AT2G38330 ABCG37
secondary metabolic process		
GO:0006722	triterpenoid metabolic process	04C11 BARS1 CYP705A5 THAS

## Apéndice

ATHB-17 MICA/GOID	Término GO	Genes		
GO:0009807	lignan biosynthetic process	AT2G21100	AT3G13650	AT4G11190 AT4G23690
	membrane lipid metabolic process			
GO:0006664	glycolipid metabolic process	DGD1	AT4G10060	SQD2 ATMGD2
	generation of precursor metabolites and energy			
GO:0043467	regulation of generation of precursor metabolites and energy	DGD1	OEC33	ASML1 FTSH5
GO:0019684	photosynthesis, light reaction	HCF173	FTSH1	DGD1 DEGP1 OEC33 HCEF1 AT3G60370
		PGRL1B	HMA6	FTSH5
GO:0022900	electron transport chain	AT1G32550	SDH1-2	ATFD3 OEC33 HCEF1 PGRL1B HMA6 FTSH5
	signal transduction			
GO:0035556	intracellular signal transduction	AT1G44050	AT1G61840	AT2G02680 AT3G07000
		AT3G27473	AT3G45530	
		AT4G01920	AT4G01930	AT4G11550
		AT4G14980	AT5G43040	AT5G46390
		AT5G54020	AT5G54040	AT5G59940
	establishment of plastid localization			
GO:0009902	chloroplast relocation	AT1G66840	CHUP1	NPL1
	MF			
	macrolide binding			
GO:0005528	FK506 binding	AT1G20810	ATFKBP62	AT3G60370 AT4G19830
		AT4G26555	AT5G13410	
	isomerase activity			
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	AT1G01940	AT1G20810	CYP40 ATFKBP62 AT3G60370
		ROC4	AT4G19830	AT4G26555 AT5G13410
GO:0016866	intramolecular transferase activity	AT1G22170	AT1G34150	04C11 BARS1 DRT101 THAS
		GSA1		
	tetrapyrrole binding			
GO:0020037	heme binding	CYP87A2	CYP705A27	CYP705A25 AT1G68850
		HO3	AT2G18980	AT2G43480
		AT3G03670	CYP72A11	CYP705A20
		CYP71B34	CYP94D2	CYP74B2
		AT4G33420	APX5	AT5G06730
		AT5G08250	CYP86B1	AT5G24070
		CYP71B12	CYP96A4	AT5G58390
	oxidoreductase activity			
GO:0004601	peroxidase activity	AT1G68850	AT2G18980	AT2G43480 AT3G03670
		AT3G21770	AT4G11230	AT4G33420
		APX5	AT5G06730	AT5G24070 AT5G58390

## Apéndice

ATHB-17		
MICA/GOID	Término GO	Genes
GO:0016706	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	AT1G06650 AT1G14130 AT1G35190 AT2G17720 AT3G19000 AT4G16770 AT5G12270 AT5G20550 AT5G58660
binding		
GO:0030246	carbohydrate binding	RTM1 AT1G11300 PDCB3 AT1G56680 AT1G61610 AtPP2-B2 ARD ATPP2-A3 BGAL8 AT4G11900 ATPP2-A1 AT5G06740 AtPP2-A8
receptor activity		
GO:0004888	transmembrane receptor activity	AT1G69550 AHK4 AT2G17060 AT3G04210 ARK3 AT4G23510 AT5G40100 AT5G44920 AT5G45060 AT5G45210 AT5G45220 AT5G46270 AT5G51630
ATPase activity, coupled to movement of substances		
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	ATPDR7 AT1G17500 AT1G54280 ATPDR11 ATTAP1 ATPDR5 ABCB4 AT3G13900 AT3G28345 ABCG37 AT3G53510 AHA2 HMA6
CC		
chloroplast thylakoid		
GO:0009543	chloroplast thylakoid lumen	AT1G20810 AT2G26340 PSBP-2 PSBTN DEGP1 OEC33 AT3G60370 AT4G19830 PSBQ AT5G13410 AT5G46390 AT5G52970
GO:0009535	chloroplast thylakoid membrane	RPL4 CP24 AT1G16880 FTSH1 TSA1 ZKT AT2G26340 PSBP-2 AT2G45180 FLU PSBTN DEGP1 OEC33 ATG4 AT3G52230 AT3G53470 ROC4 PGRL1B PSBQ CHLM AT4G27700 AT4G32260 SEP1 AT5G03880 AT5G13410 AT5G37360 FTSH5 AT5G55710 PSBL
plastid outer membrane		
GO:0009707	chloroplast outer membrane	ATSFR2 DGD1 CHUP1 AT3G52230 AFT ATMGD2

ATHB18		
MICA/GOID	Término GO	Genes
BP		
regulation of biological quality		
GO:0048767	root hair elongation	PIP5K3 ABCB4 ATCSLD2



## Apéndice

ATHB18		
MICA/GOID	Término GO	Genes
GO:0055082	cellular chemical homeostasis	ATGLR3.4 WAK2 ATGLR1.2
	response to abiotic stimulus	
GO:0080167	response to karrikin	AT1G50900 VTC4 AT3G55720 ZIFL1
	MF	
	ATPase activity, coupled to movement of substances	
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	AT2G01320 AT2G39350 ABCB4 AT3G25610 AT3G55090 AT3G63380
	cis-trans isomerase activity	
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	AT1G20810 AT3G60370 ROC4 AT5G35100
	CC	
	thylakoid part	
GO:0031977	thylakoid lumen	AT1G20810 AT3G60370 ROC4 AT5G35100

HAT14		
MICA/GOID	Término GO	Genes
	BP	
	photosynthesis, light reaction	
GO:0009765	photosynthesis, light harvesting	PSB27 PSAO TLP18.3 LHCA3 LHCA1
GO:0009767	photosynthetic electron transport chain	PGR5 PSAN MSP-1
	DNA replication	
GO:0006261	DNA-dependent DNA replication	ATMCM2 LGO ABP MCM3
	response to light stimulus	
GO:0009644	response to high light intensity	ATMYBL2 PGR5 OHP MSP-1
GO:0009637	response to blue light	IAA19 LHCA1 LHCB4.1
	meristem growth	
GO:0010075	regulation of meristem growth	ATMCM2 BAM2 BAM3
	chlorophyll metabolic process	
GO:0015995	chlorophyll biosynthetic process	AT1G74470 CHLM CIP7
	toxin metabolic process	
GO:0009407	toxin catabolic process	ATGSTU24 ATGSTZ1 ATGSTU27
	response to metal ion	
GO:0046686	response to cadmium ion	ATGLX1 AT1G11860 AT1G56190 ATMYBL2 AT1G74910 AT2G37760 ATGSKB6 AT5G64250
	enzyme linked receptor protein signaling pathway	
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	AT1G79620 AT2G25790 AT3G28040 BAM2 BAM3
	MF	
	tetrapyrrole binding	
GO:0016168	chlorophyll binding	LHCA3 LHCA1 LHCB4.1
	aspartic-type peptidase activity	

## Apéndice

HAT14			
MICA/GOID	Término GO	Genes	
GO:0004190	aspartic-type endopeptidase activity	AT1G09750 AT3G52500	AT1G79720 AT2G36670
	oxidoreductase activity, acting on CH-OH group of donors		
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	AT1G60730 AT5G19440	ATCAD1 FLDH NOL
	CC		
	protein complex		
GO:0030095	chloroplast photosystem II	PSB27	AT1G51400 PSBQ MSP-1
GO:0030076	light-harvesting complex	LHCA3 LHCA1 LHCB4.1	
	extracellular region		
GO:0048046	apoplast	AT1G09750 AT1G79720 BGAL2	AT1G11860 XTH32 AT3G10080 ATXTR8 RALFL32 MSP-1
	plastid thylakoid lumen		
GO:0009543	chloroplast thylakoid lumen	PSB27	AT1G51400 TLP18.3 PSBQ PSAN MSP-1
	chloroplast stroma		
GO:0010287	plastoglobule	LHCA3 LHCA1 LHCB4.1	MSP-1

HAT22			
MICA/GOID	Término GO	Genes	
	BP		NA
	response to chemical stimulus		NA
GO:0010200	response to chitin	JAZ7 ATWRKY33 ATSZF2 ATL6 ATERF-4 AZF2 AT5G46910 AT-HSFB2A	
GO:0006979	response to oxidative stress	AT1G13340 AtGolS1	AT1G19020 ATHSPRO2 PMSR4 HSP18.2
	response to stress		NA
GO:0009414	response to water deprivation	ATWRKY33 ATMRP4 DREB2 ATHVA22D ANAC072 SIP1	AZF2
GO:0009611	response to wounding	CIPK9 AT1G66760 ATCK1 JAZ2 ATMES5	ATMRP4
GO:0009409	response to cold	CIPK9 ATCCR2 ATWRKY33 ATHVA22D LEA4-5	ATSZF2
GO:0050832	defense response to fungus	ATWRKY33 ATSZF2 AT5G39580	ATNPR3
GO:0009408	response to heat	ATWRKY33 AtGolS1 DREB2	HSP18.2
	indole-containing compound metabolic process		NA
GO:0042435	indole-containing compound biosynthetic process	ASA2 ATWRKY33 AT3G57880	
	negative regulation of biological process		NA
GO:0048523	negative regulation of cellular process	AFP4 ATERF-4 ATHB-2 ATHVA22D	SHA1

## Apéndice

HAT22			
MICA/GOID	Término GO	Genes	
	glycoside metabolic process		NA
GO:0016138	glycoside biosynthetic process	FMO SIP1 AT5G65140	
MF		NA	
	transcription regulator activity		NA
GO:0016564	transcription repressor activity	ATERF11 ATERF-4 AZF2	
	transmembrane transporter activity		NA
GO:0015238	drug transmembrane transporter activity	ZF14 AT1G61890 AT1G66760	

HAT9			
MICA/GOID	Término GO	Genes	
BP		NA	
	response to abiotic stimulus		NA
GO:0080167	response to karrikin	SNRK2-8 HYH AT3G52740 AMT1;1 AT4G35110 AT5G05270 AT5G48540 AT5G62210 AT5G62680	
GO:0010224	response to UV-B	HYH AT3G21890 AT5G52250	
	peptide transport		NA
GO:0006857	oligopeptide transport	AT1G18880 NRT1.5 AT3G01350 AT5G62680	

ATHB-2			
MICA/GOID	Término GO	Genes	
BP		NA	
	response to stimulus		NA
GO:0010200	response to chitin	AT1G20823 ATCMPG1 ATERF-4 AT3G46620 ATERF6 AT5G59450	
GO:0009735	response to cytokinin stimulus	ACR4 ARR5 ATHB-2 HAT22	
GO:0071495	cellular response to endogenous stimulus	RPK1 ATERF-4 ARR5 ATSYP122 ATERF6	
	response to abiotic stimulus		NA
GO:0009631	cold acclimation	ATCBF2 ATCBF3 ATCBF1	
GO:0009414	response to water deprivation	ATCAMP25 DREB2 ATCBF3 ATCBF1 CYP707A3 ABA1	
GO:0009639	response to red or far red light	ARR5 ATHB-2 CYP707A3 ABA1	
	regulation of response to stimulus		NA
GO:0048585	negative regulation of response to stimulus	ATERF-4 ATSYP122 BAP1	
GO:0031347	regulation of defense response	ATERF-4 ATSYP122 BAP1	
MF		NA	
	transcription regulator activity		NA
GO:0016563	transcription activator activity	DREB2 ATCBF2 ATCBF3 ATCBF1	

## Apéndice

HAT1		
MICA/GOID	Término GO	Genes
BP		
response to hormone stimulus		
GO:0009788	negative regulation of abscisic acid mediated signaling pathway	ATHB6 AHG3 ABI1 ATPK10 ABI2
GO:0009733	response to auxin stimulus	AT1G29440 AT1G29500 SAUR68 BRL2 IAA19 ATMYB30 ARGOS AT3G60690 AFB1 AT4G34770 BT5 IAA28 HAT2
response to abiotic stimulus		
GO:0009409	response to cold	GAPB ATSERAT2;1 ERD14 ATCOR413-PM1 MYBC1 AHG3 AOX1A ABI1 AT4G30650 LTI65 COR78
GO:0009269	response to desiccation	ERD14 RD2 COR78
response to other organism		
GO:0009624	response to nematode	AAP1 ATGPT2 LAX2 ZIFL2 AT5G52450
heterocycle biosynthetic process		
GO:0006168	adenine salvage	APT2 APT5
GO:0006779	porphyrin biosynthetic process	AT1G74470 GSA2 GUN4 CHLM
dephosphorylation		
GO:0006470	protein dephosphorylation	HAI2 ABI1 AT5G06750 ABI2
organic substance transport		
GO:0008643	carbohydrate transport	ATGPT2 ATHPT AT4G17550
MF		
transferase activity, transferring pentosyl groups		
GO:0003999	adenine phosphoribosyltransferase activity	APT2 APT5
phosphoprotein phosphatase activity		
GO:0004722	protein serine/threonine phosphatase activity	HAI2 AHG3 ATACP5 ABI1 AT5G06750 ABI2
CC		
cell wall		
GO:0009505	plant-type cell wall	ATBXL2 AT1G04040 AT1G21670 AT1G21680 EXL7 AT2G26440 AT3G10720 AT3G43270 IMK2 EXO AT4G16563 sks4 AT5G11420
protein complex		
GO:0008287	protein serine/threonine phosphatase complex	HAI2 ABI1 AT5G06750 ABI2

HAT2		
MICA/GOID	Término GO	Genes
BP		
chemical homeostasis		
GO:0010252	auxin homeostasis	GH3.17 GH3.3 GH3.5 DFL1

## Apéndice

HAT2		
MICA/GOID	Término GO	Genes
	response to hormone stimulus	NA
GO:0009733	response to auxin stimulus	ATAUX2-27 GH3.17 SAUR68 GH3.1 GH3.3 GH3.5 HAT2 DFL1
	pigment metabolic process	NA
GO:0046148	pigment biosynthetic process	LYC GUN4 ABAR
MF		NA
	ligase activity, forming carbon-nitrogen bonds	NA
GO:0016881	acid-amino acid ligase activity	GH3.17 GH3.3 PUB25 GH3.5 DFL1
HAT3		
MICA/GOID	Término GO	Genes
	BP	
	photosynthesis, light reaction	
GO:0042549	photosystem II stabilization	ATCYP38 OEC33 MSP-1
GO:0010207	photosystem II assembly	ATCYP38 OEC33 MSP-1
GO:0009773	photosynthetic electron transport in photosystem I	NDF5 CRR23 PSAP
	response to abiotic stimulus	
GO:0009637	response to blue light	LHCB4.2 JK224 ATBZIP54 ATCRY1 LHCB4.1 PMI15
GO:0080167	response to karrikin	AT1G50900 AT1G74670 AGP26 CAB4 AT4G35110 ZIFL1 AT5G18470
GO:0010114	response to red light	ATRL2 ATEXP3 LHCB4.2 LHCB4.1
	phospholipid metabolic process	
GO:0008654	phospholipid biosynthetic process	AT1G15110 NPC2 CLA ATS2 VTE5 ATMIPS3
	chlorophyll metabolic process	
GO:0015995	chlorophyll biosynthetic process	ALB-1V AT1G74470 CLA CHLM
	chloroplast relocation	
GO:0009903	chloroplast avoidance movement	JK224 PMI15
	regulation of phosphate metabolic process	
GO:0035303	regulation of dephosphorylation	OEC33 MSP-1
	cellular protein metabolic process	
GO:0006457	protein folding	AT1G73655 AT2G24860 AT2G30695 AT2G43560 ATCYP38 AT4G26555 AT5G18140 AT5G35100 AT5G55200
	MF	
	cis-trans isomerase activity	
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	AT1G73655 AT2G43560 ATCYP38 AT4G26555 AT5G35100
	photoreceptor activity	

## Apéndice

HAT3		
MICA/GOID	Término GO	Genes
GO:0009882	blue light photoreceptor activity	JK224 ATCRY1
	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	
GO:0004345	glucose-6-phosphate dehydrogenase activity	G6PD2 G6PD1
	macrolide binding	
GO:0005528	FK506 binding	AT1G73655 AT2G43560 AT4G26555
	tetrapyrrole binding	
GO:0016168	chlorophyll binding	LHCB4.2 CAB4 LHCB4.1
	CC	
	thylakoid part	
GO:0009543	chloroplast thylakoid lumen	PSB27 AT1G12250 AT1G51400 AT2G43560 ATCYP38 OEC33 AT4G15510 AT5G53490 MSP-1
GO:0030095	chloroplast photosystem II	PSB27 AT1G51400 OEC33 AT4G15510 MSP-1
GO:0009522	photosystem I	PSAD-2 PSAO PSAF PSAP
GO:0009654	oxygen evolving complex	OEC33 AT4G15510 MSP-1
	plastid stroma	
GO:0010287	plastoglobule	PSAF AT2G42130 CAB4 OEC33 LHCB4.1 MSP-1
GO:0042646	plastid nucleoid	PSAP PTAC3 MFP1

# Bibliografía

- Abascal, F.; Zardoya, R. y Posada, D. (2005). *ProtTest: selection of best-fit models of protein evolution*. *Bioinformatics*, 21(9), 2104–2105.
- Adami, C. (2002). *What is complexity?* *BioEssays*, 24(12), 1085–1094.
- Agalou, A.; Purwantomo, S.; Overnäs, E.; Johannesson, H.; Zhu, X.; Estiati, A.; de Kam, R.J.; Engström, P.; Slamet-Loedin, I.H.; Zhu, Z.; Wang, M.; Xiong, L.; Meijer, A.H. y Ouwerkerk, P.B.F. (2008). *A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members*. *Plant Molecular Biology*, 66(1-2), 87–103.
- Alonso, J.M.; Stepanova, A.N.; Leisse, T.J.; Kim, C.J.; Chen, H.; Shinn, P.; Stevenson, D.K.; Zimmerman, J.; Barajas, P.; Cheuk, R.; Gadrinab, C.; Heller, C.; Jeske, A.; Koesema, E.; Meyers, C.C.; Parker, H.; Prednis, L.; Ansari, Y.; Choy, N.; Deen, H.; Geralt, M.; Hazari, N.; Hom, E.; Karnes, M.; Mulholland, C.; Ndubaku, R.; Schmidt, I.; Guzman, P.; Aguilar-Henonin, L.; Schmid, M.; Weigel, D.; Carter, D.E.; Marchand, T.; Risseeuw, E.; Brogden, D.; Zeko, A.; Crosby, W.L.; Berry, C.C. y Ecker, J.R. (2003). *Genome-wide insertional mutagenesis of Arabidopsis thaliana*. *Science*, 301(5633), 653–657.
- Alvarez-Buylla, E.R.; Benítez, M.; Dávila, E.B.; Chaos, A.; Espinosa-Soto, C. y Padilla-Longoria, P. (2007). *Gene regulatory network models for plant development*. *Current Opinion in Plant Biology*, 10(1), 83–91.
- Alwine, J.C.; Kemp, D.J. y Stark, G.R. (1977). *Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes*. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5350–5354.
- Amoutzias, G.D.; Robertson, D.L.; Van de Peer, Y. y Oliver, S.G. (2008). *Choose your partners: dimerization in eukaryotic transcription factors*. *Trends in Biochemical Sciences*, 33(5), 220–229.
- Aoyama, T.; Dong, C.h.; Wu, Y.; Carabelli, M.; Sessa, G.; Ruberti, I.; Morelli, G. y Chau, N. (1995). *Ectopic expression of the Arabidopsis transcriptional activator Athb-1 alters leaf cell fate in tobacco*. *The Plant Cell*, 7(11), 1773–1785.

- Ariel, F.D.; Diet, A.; Marion, V.; Gruber, V.; Florian, F.; Chan, R.L. y Crespi, M. (2010). *Environmental regulation of lateral root emergence in Medicago truncatula requires the HD-Zip I transcription factor HBI*. *The Plant Cell*, 22(7), 1–14.
- Ariel, F.D.; Manavella, P.A.; Dezar, C.A. y Chan, R.L. (2007a). *The true story of the HD-Zip family*. *Trends in Plant Science*, 12(9), 419–426.
- Ariel, F.D.; Manavella, P.A.; Giacomelli, J.I. y Chan, R.L. (2007b). *HD-Zip I and II , Transcription Factors Involved in the Adaptive Response to Environmental Stress*. *Plant Stress*, 1(2), 189–196.
- Aso, K.; Kato, M.; Banks, J.A. y Hasebe, M. (1999). *Characterization of homeodomain-leucine zipper genes in the fern Ceratopteris richardii and the evolution of the homeodomain-leucine zipper gene family in vascular plants*. *Molecular Biology and Evolution*, 16(4), 544–552.
- Ausubel, F.; Brent, R.; Kingston, R.; Moore, D.; Seidman, J.; Smith, J. y Struhl, K. (1987). *Current protocols in molecular biology*. New York, New York, Estados Unidos: John Wiley & Sons.
- Avery, O.T.; MacLeod, C.M. y McCarty, M. (1944). *Studies on the chemical nature of the substance inducing transformation of pneumococcal types*. *The Journal of Experimental Medicine*, 79(2), 137–159.
- Bailey, T.L.; Williams, N.; Misleh, C. y Li, W.W. (2006). *MEME: discovering and analyzing DNA and protein sequence motifs*. *Nucleic Acids Research*, 34(Web Server issue), W369–W3673.
- Baima, S.; Nobili, F.; Sessa, G.; Lucchetti, S.; Ruberti, I. y Morelli, G. (1995). *The expression of the Athb-8 homeobox gene is restricted to provascular cells in Arabidopsis thaliana*. *Development*, 121(12), 4171–4182.
- Barabási, A.L. (1999). *Emergence of Scaling in Random Networks*. *Science*, 286(5439), 509–512.
- Barabási, A.L. y Oltvai, Z.N. (2004). *Network biology: understanding the cell's functional organization*. *Nature Reviews Genetics*, 5(2), 101–113.
- Barrero, J.M.; Millar, A.A.; Griffiths, J.; Czechowski, T.; Scheible, W.R.; Udvardi, M.; Reid, J.B.; Ross, J.J.; Jacobsen, J.V. y Gubler, F. (2010). *Gene expression profiling identifies two regulatory genes controlling dormancy and ABA sensitivity in Arabidopsis seeds*. *The Plant Journal*, 61(4), 611–622.
- Basso, K.; Margolin, A.A.; Stolovitzky, G.; Klein, U.; Dalla-Favera, R. y Califano, A. (2005). *Reverse engineering of regulatory networks in human B cells*. *Nature Genetics*, 37(4), 382–390.
- Bembom, O.; Keles, S. y van der Laan, M.J. (2007). *Supervised detection of conserved motifs in DNA sequences with cosmo*. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 8.



- Berenstein, A. y Chernomoretz, A. (2012). Manuscrito en preparación.
- Birnboim, H.C. y Doly, J. (1979). *A rapid alkaline extraction procedure for screening recombinant plasmid DNA*. Nucleic Acids Research, 7(6), 1513–1523.
- Blanc, G. y Wolfe, K.H. (2004). *Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes*. The Plant Cell, 16(7), 1667–1678.
- Blom, N.; Gammeltoft, S. y Brunak, S. (1999). *Sequence and structure-based prediction of eukaryotic protein phosphorylation sites*. Journal of Molecular Biology, 294(5), 1351–1362.
- Brivanlou, A.H. y Darnell, J.E. (2002). *Signal transduction and the control of gene expression*. Science, 295(5556), 813–818.
- Bylesjö, M.; Segura, V.; Soolanayakanahally, R.Y.; Rae, A.M.; Trygg, J.; Gustafsson, P.; Jansson, S. y Street, N.R. (2008). *LAMINA: a tool for rapid quantification of leaf size and shape parameters*. BMC Plant Biology, 8, 82.
- Cabello, J.V.; Arce, A.L. y Chan, R.L. (2010). Methods and compositions for stress tolerance in plants. Plant Bioscience Limited-CONICET-UNL. WO/2010/139993.
- Cabello, J.V.; Arce, A.L. y Chan, R.L. (2012). *The homologous HD-Zip I transcription factors HaHB1 and AtHB13 confer cold tolerance via the induction of pathogenesis-related and glucanase proteins*. The Plant Journal, 69(1), 141–153.
- Carabelli, M.; Morelli, G.; Whitelam, G.C. y Ruberti, I. (1996). *Twilight-zone and canopy shade induction of the Athb-2 homeobox gene in green plants*. Proceedings of the National Academy of Sciences of the United States of America, 93(8), 3530–3535.
- Carabelli, M.; Sessa, G.; Baima, S.; Morelli, G. y Ruberti, I. (1993). *The Arabidopsis Athb-2 and -4 genes are strongly induced by far-red-rich light*. The Plant Journal, 4(3), 469–479.
- Causier, B.; Ashworth, M.; Guo, W. y Davies, B. (2012). *The TOPLESS interactome: a framework for gene repression in Arabidopsis*. Plant Physiology, 158(1), 423–438.
- Chan, R.L.; Gago, G.M.; Palena, C.M. y Gonzalez, D.H. (1998). *Homeoboxes in plant development*. Biochimica et Biophysica Acta, 1442(1), 1–19.
- Chan, R.L. y Gonzalez, D.H. (1994). *A cDNA encoding an HD-zip protein from sunflower*. Plant Physiology, 106(4), 1687–1688.
- Choi, C.Y.; Lee, Y.M.; Kim, Y.H.; Park, T.; Jeon, B.H.; Schulz, R.A. y Kim, Y. (1999). *The homeodomain transcription factor NK-4 acts as either a transcriptional activator or repressor and interacts with the p300 coactivator and the Groucho corepressor*. The Journal of Biological Chemistry, 274(44), 31543–31552.
- Choi, H.i.; Hong, J.h.; Ha, J.o.; Kang, J.y. y Kim, S.Y. (2000). *ABFs, a family of ABA-responsive element binding factors*. The Journal of Biological Chemistry, 275(3), 1723–1730.

- Ciarbelli, A.R.; Ciolfi, A.; Salvucci, S.; Ruzza, V.; Possenti, M.; Carabelli, M.; Fruscalzo, A.; Sessa, G.; Morelli, G. y Ruberti, I. (2008). *The Arabidopsis homeodomain-leucine zipper II gene family: diversity and redundancy*. *Plant Molecular Biology*, 68(4-5), 465–478.
- Clough, S.J. y Bent, A.F. (1998). *Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana*. *The Plant Journal*, 16(6), 735–743.
- Cohen-Peer, R.; Schuster, S.; Meiri, D.; Breiman, A. y Avni, A. (2010). *Sumoylation of Arabidopsis heat shock factor A2 (HsfA2) modifies its activity during acquired thermotolerance*. *Plant Molecular Biology*, 74(1-2), 33–45.
- Crick, F. (1970). *Central dogma of molecular biology*. *Nature*, 227(5258), 561–563.
- Crooks, G.; Hon, G.; Chandonia, J. y Brenner, S.E. (2004). *WebLogo: a sequence logo generator*. *Genome Research*, 14(6), 1188–1190.
- Cutler, S.R.; Rodriguez, P.L.; Finkelstein, R.R. y Abrams, S.R. (2010). *Abscisic acid: emergence of a core signaling network*. *Annual Review of Plant Biology*, 61, 651–679.
- Deng, X.; Phillips, J.; Bra, A.; Engstro, P.; Johannesson, H.; Ouwerkerk, P.B.F.; Ruberti, I.; Salinas, J.; Vera, P.; Iannaccone, R.; Meijer, A.H. y Bartels, D. (2006). *A homeodomain leucine zipper gene from Craterostigma plantagineum regulates abscisic acid responsive gene expression and physiological responses*. *Plant Molecular Biology*, (pp. 469–489).
- Deng, X.; Phillips, J.; Meijer, A.H.; Salamini, F. y Bartels, D. (2002). *Characterization of five novel dehydration-responsive homeodomain leucine zipper genes from the resurrection plant Craterostigma plantagineum*. *Plant Molecular Biology*, 49(6), 601–610.
- Dezar, C.A.; Fedrigo, G. y Chan, R.L. (2005a). *The promoter of the sunflower HD-Zip protein gene directs tissue-specific expression and is inducible by water stress, high salt concentrations and ABA*. *Plant Science*, 169(2), 447–456.
- Dezar, C.A.; Gago, G.M.; Gonzalez, D.H. y Chan, R.L. (2005b). *Hahb-4, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to Arabidopsis thaliana plants*. *Transgenic Research*, 14(4), 429–440.
- Dezar, C.A.; Giacomelli, J.I.; Manavella, P.A.; Ré, D.A.; Alves-Ferreira, M.; Baldwin, I.T.; Bonaventure, G. y Chan, R.L. (2011). *HAHB10, a sunflower HD-Zip II transcription factor, participates in the induction of flowering and in the control of phytohormone-mediated responses to biotic stress*. *Journal of Experimental Botany*, 62(3), 1061–1076.
- Döring, P.; Treuter, E.; Kistner, C.; Lyck, R.; Chen, A. y Nover, L. (2000). *The role of AHA motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2*. *The Plant Cell*, 12(2), 265–278.
- Eddy, S. (1998). *Profile hidden Markov models*. *Bioinformatics*, 14(9), 755–763.

- Felsenstein, J. (1989). *PHYLIP - Phylogeny Inference Package (Version 3.2)*. *Cladistics*, 5, 164 – 166.
- Finn, R.D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J.E.; Gavin, O.L.; Gunasekaran, P.; Ceric, G.; Forslund, K.; Holm, L.; Sonnhammer, E.L.L.; Eddy, S.R. y Bateman, A. (2010). *The Pfam protein families database*. *Nucleic Acids Research*, 38(Database issue), D211–D222.
- Frank, W.; Phillips, J.; Salamini, F. y Bartels, D. (1998). *Two dehydration-inducible transcripts from the resurrection plant Craterostigma plantagineum encode interacting homeodomain-leucine zipper proteins*. *The Plant Journal*, 15(3), 413–421.
- Frankel, A. y Kim, P. (1991). *Modular structure of transcription factors: Implications for gene regulation*. *Cell*, 65(5), 717–719.
- Franklin, K.A. (2008). *Shade avoidance*. *The New Phytologist*, 179(4), 930–944.
- Gago, G.M.; Almoguera, C.; Jordano, J.; Gonzalez, D.H. y Chan, R.L. (2002). *Hahb-4, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower*. *Plant, Cell & Environment*, 25(5), 633–640.
- Geer, L.Y.; Domrachev, M.; Lipman, D.J. y Bryant, S.H. (2002). *CDART: protein homology by domain architecture*. *Genome Research*, 12(10), 1619–1623.
- Gentleman, R.C.; Carey, V.J.; Bates, D.M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A.J.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J.Y.H. y Zhang, J. (2004). *Bioconductor: open software development for computational biology and bioinformatics*. *Genome Biology*, 5(10), R80.
- Gietz, D.; St Jean, A.; Woods, R. y Schiestl, R. (1992). *Improved method for high efficiency transformation of intact yeast cells*. *Nucleic Acids Research*, 20(6), 1425.
- Gong, W.; Shen, Y.; Ma, L.; Pan, Y.; Du, Y.; Wang, D.; Yang, J.; Hu, L.; Liu, X.; Dong, C. y Others (2004). *Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes*. *Plant Physiology*, 135(2), 773–782.
- Gonzalez, D.H.; Valle, E.M. y Chan, R.L. (1997). *Interaction between proteins containing homeodomains associated to leucine zippers from sunflower*. *Biochimica et Biophysica Acta*, 1351(1-2), 137–149.
- Gottwald, S.; Bauer, P.; Komatsuda, T.; Lundqvist, U. y Stein, N. (2009). *TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene HvHox1*. *BMC Research Notes*, 2, 258.
- Gouy, M.; Guindon, S. y Gascuel, O. (2010). *SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building*. *Molecular Biology and*

- Evolution, 27(2), 221–224.
- Guindon, S. y Gascuel, O. (2003). *A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood*. Systematic Biology, 52(5), 696–704.
- Guy, L.; Roat Kultima, J. y Andersson, S.G.E. (2010). *genoPlotR: comparative gene and genome visualization in R*. Bioinformatics, 26(18), 2334–2335.
- Hanahan, D. (1983). *Studies on transformation of Escherichia coli with plasmids*. Journal of Molecular Biology, 166(4), 557–580.
- Hanson, J.; Johannesson, H. y Engström, P. (2001). *Sugar-dependent alterations in cotyledon and leaf development in transgenic plants expressing the HDZhdip gene ATHB13*. Plant Molecular Biology, 2, 247–262.
- Hanson, J.; Regan, S. y Engström, P. (2002). *The expression pattern of the homeobox gene ATHB13 reveals a conservation of transcriptional regulatory mechanisms between Arabidopsis and hybrid aspen*. Plant Cell Reports, 21(1), 81–89.
- Harris, J.C.; Hrmova, M.; Lopato, S. y Langridge, P. (2011). *Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli*. The New Phytologist, 190(4), 823–837.
- Heazlewood, J.L.; Durek, P.; Hummel, J.; Selbig, J.; Weckwerth, W.; Walther, D. y Schulze, W.X. (2007). *PhosPhAt: a database of phosphorylation sites in Arabidopsis thaliana and a plant-specific phosphorylation site predictor*. Nucleic Acids Research, 36(Database issue), D1015–D1021.
- Heid, C.a.; Stevens, J.; Livak, K.J. y Williams, P.M. (1996). *Real time quantitative PCR*. Genome Research, 6(10), 986–994.
- Heim, M.A.; Jakoby, M.; Werber, M.; Martin, C.; Weisshaar, B. y Bailey, P.C. (2003). *The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity*. Molecular Biology and Evolution, 20(5), 735–747.
- Henriksson, E.; Olsson, A.S.B.; Johannesson, H.; Johansson, H.; Hanson, J.; Engström, P. y Söderman, E. (2005). *Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns and phylogenetic relationships*. Plant Physiology, 139(1), 509–518.
- Heslop-Harrison, J.S.P. y Schwarzacher, T. (2011). *Organisation of the plant genome in chromosomes*. The Plant Journal, 66(1), 18–33.
- Higuchi, R.; Fockler, C.; Dollinger, G. y Watson, R. (1993). *Kinetic PCR analysis: real-time monitoring of DNA amplification reactions*. Nature Biotechnology, 11(9), 1026–1030.
- Higuchi, R.; Krummel, B. y Saiki, R.K. (1988). *A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions*. Nucleic Acids

- Research, 16(15), 7351–7367.
- Himmelbach, A.; Hoffmann, T.; Leube, M.; Höhener, B. y Grill, E. (2002). *Homeodomain protein ATHB6 is a target of the protein phosphatase ABI1 and regulates hormone responses in Arabidopsis*. The EMBO Journal, 21(12), 3029–3038.
- Hjellström, M.; Olsson, A.S.B.; Engström, P. y Söderman, E.M. (2003). *Constitutive expression of the water deficit-inducible homeobox gene ATHB7 in transgenic Arabidopsis causes a suppression of stem elongation growth*. Plant, Cell & Environment, 26(7), 1127–1136.
- Hofer, J.; Turner, L.; Moreau, C.; Ambrose, M.; Isaac, P.; Butcher, S.; Weller, J.; Dupin, A.; Dalmais, M.; Le Signor, C.; Bendahmane, A. y Ellis, N. (2009). *Tendrill-less regulates tendrill formation in pea leaves*. The Plant Cell, 21(2), 420–428.
- Hosoda, K.; Imamura, A.; Katoh, E.; Hatta, T.; Tachiki, M.; Yamada, H.; Mizuno, T. y Yamazaki, T. (2002). *Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the Arabidopsis response regulators*. The Plant Cell, 14(9), 2015–2019.
- Huang, D.; Wu, W.; Abrams, S.R. y Cutler, A.J. (2008). *The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors*. Journal of Experimental Botany, 59(11), 2991–3007.
- Initiative, T.A.G. (2000). *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana*. Nature, 408(6814), 796–815.
- Jacob, F. y Monod, J. (1961). *Genetic regulatory mechanisms in the synthesis of proteins*. Journal of Molecular Biology, 3(3), 318–356.
- Johannesson, H.; Wang, Y. y Engström, P. (2001). *DNA-binding and dimerization preferences of Arabidopsis homeodomain-leucine zipper transcription factors in vitro*. Plant Molecular Biology, 45(1), 63–73.
- Johannesson, H.; Wang, Y.; Hanson, J. y Engström, P. (2003). *The Arabidopsis thaliana homeobox gene ATHB5 is a potential regulator of abscisic acid responsiveness in developing seedlings*. Plant Molecular Biology, 51(5), 719–729.
- Jones, D.T.; Taylor, W.R. y Thornton, J.M. (1992). *The rapid generation of mutation data matrices from protein sequences*. Bioinformatics, 8(3), 275–282.
- Jones-Rhoades, M.W.; Bartel, D.P. y Bartel, B. (2006). *MicroRNAs and their regulatory roles in plants*. Annual Review of Plant Biology, 57, 19–53.
- Kagale, S.; Links, M.G. y Rozwadowski, K. (2010). *Genome-wide analysis of ethylene-responsive element binding factor-associated amphiphilic repression motif-containing transcriptional regulators in Arabidopsis*. Plant Physiology, 152(3), 1109–1134.
- Katoh, K. y Toh, H. (2008). *Recent developments in the MAFFT multiple sequence alignment*

- program*. Briefings in Bioinformatics, 9(4), 286–298.
- Kilian, J.; Whitehead, D.; Horak, J.; Wanke, D.; Weinl, S.; Batistic, O.; D'Angelo, C.; Bornberg-Bauer, E.; Kudla, J. y Harter, K. (2007). *The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses*. The Plant Journal, 50(2), 347–363.
- Kim, Y.J. y Chen, X. (2011). *The plant Mediator and its role in noncoding RNA production*. Frontiers in Biology, 6(2), 125–132.
- Kim, Y.K.; Son, O.; Kim, M.; Nam, K.; Kim, G.; Lee, M.; Choi, S.Y. y Cheon, C.I. (2007). *ATHB23, an Arabidopsis class I homeodomain-leucine zipper gene, is expressed in the adaxial region of young leaves*. Plant Cell Reports, 26(8), 1179–1185.
- Kim, Y.S.; Kim, S.G.; Lee, M.; Lee, I.; Park, H.Y.; Seo, P.J.; Jung, J.H.; Kwon, E.J.; Suh, S.W.; Paek, K.H. y Park, C.M. (2008). *HD-ZIP III activity is modulated by competitive inhibitors via a feedback loop in Arabidopsis shoot apical meristem development*. The Plant Cell, 20(4), 920–933.
- Kohane, I.S.; Kho, A.T. y Butte, A.J. (2002). *Microarrays for an Integrative Genomics*. Cambridge, Massachusetts, Estados Unidos: The MIT Press.
- Köllmer, I.; Werner, T. y Schmülling, T. (2011). *Ectopic expression of different cytokinin-regulated transcription factor genes of Arabidopsis thaliana alters plant growth and development*. Journal of Plant Physiology, 168(12), 1320–1327.
- Komatsuda, T.; Pourkheirandish, M.; He, C.; Azhaguvel, P.; Kanamori, H.; Perovic, D.; Stein, N.; Graner, A.; Wicker, T.; Tagiri, A.; Lundqvist, U.; Fujimura, T.; Matsuoka, M.; Matsumoto, T. y Yano, M. (2007). *Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene*. Proceedings of the National Academy of Sciences of the United States of America, 104(4), 1424–1429.
- Kotak, S.; Port, M.; Ganguli, A.; Bicker, F. y von Koskull-Döring, P. (2004). *Characterization of C-terminal domains of Arabidopsis heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization*. The Plant Journal, 39(1), 98–112.
- Kunihiro, A.; Yamashino, T.; Nakamichi, N.; Niwa, Y.; Nakanishi, H. y Mizuno, T. (2011). *Phytochrome-interacting factor 4 and 5 (PIF4 and PIF5) activate the homeobox ATHB2 and auxin-inducible IAA29 genes in the coincidence mechanism underlying photoperiodic control of plant growth of Arabidopsis thaliana*. Plant & Cell Physiology, 52(8), 1315–1329.
- Lamesch, P.; Berardini, T.Z.; Li, D.; Swarbreck, D.; Wilks, C.; Sasidharan, R.; Muller, R.; Dreher, K.; Alexander, D.L.; Garcia-Hernandez, M.; Karthikeyan, A.S.; Lee, C.H.; Nelson, W.D.; Pløetz, L.; Singh, S.; Wensel, A. y Huala, E. (2011). *The Arabidopsis Information Resource*

- (TAIR): *improved gene annotation and new tools*. Nucleic Acids Research, 40(Database issue), 1202–1210.
- Landschulz, W.; Johnson, P. y McKnight, S. (1988). *The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins*. Science, 240(4860), 1759–1764.
- Lechner, E.; Leonhardt, N.; Eisler, H.; Parmentier, Y.; Alioua, M.; Jacquet, H.; Leung, J. y Genschik, P. (2011). *MATH/BTB CRL3 Receptors Target the Homeodomain-Leucine Zipper ATHB6 to Modulate Abscisic Acid Signaling*. Developmental Cell, 21(6), 1116–1128.
- Lee, J.; He, K.; Stolc, V.; Lee, H.; Figueroa, P.; Gao, Y.; Tongprasit, W.; Zhao, H.; Lee, I. y Deng, X.W. (2007). *Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development*. The Plant Cell, 19(3), 731–749.
- Lee, Y.H. y Chun, J. (1998). *A new homeodomain-leucine zipper gene from Arabidopsis thaliana induced by water stress and abscisic acid treatment*. Plant Molecular Biology, 37(2), 377–384.
- Lee, Y.H.; Oh, H.S.; Cheon, C.I.; Hwang, I.T.; Kim, Y.J. y Chun, J.y. (2001). *Structure and Expression of the Arabidopsis thaliana Athb-12, Homeobox Gene*. Biochemical and Biophysical Research Communications, 284(1), 133–141.
- Li, J. y Chory, J. (1998). *Preparation of DNA from Arabidopsis*. Methods in Molecular Biology, 82, 55–60.
- Lin, Z.; Hong, Y.; Yin, M.; Li, C.; Zhang, K. y Grierson, D. (2008). *A tomato HD-Zip homeobox protein, LeHB-1, plays an important role in floral organogenesis and ripening*. The Plant Journal, 55(2), 301–310.
- Liu, Z. y Karmarkar, V. (2008). *Groucho/Tup1 family co-repressors in plant development*. Trends in Plant Science, 13(3), 137–144.
- Long, T.A.; Brady, S.M. y Benfey, P.N. (2008). *Systems approaches to identifying gene regulatory networks in plants*. Annual Review of Cell and Developmental Biology, 24, 81–103.
- Lopez-Molina, L.; Mongrand, S. y Chua, N.H. (2001). *A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America, 98(8), 4782–7.
- Lynch, M. y Conery, J.S. (2000). *The Evolutionary Fate and Consequences of Duplicate Genes*. Science, 290(5494), 1151–1155.
- MacQuarrie, K.L.; Fong, A.P.; Morse, R.H. y Tapscott, S.J. (2011). *Genome-wide transcription factor binding: beyond direct target regulation*. Trends in Genetics, 27(4), 141–148.
- Maere, S.; De Bodt, S.; Raes, J.; Casneuf, T.; Van Montagu, M.; Kuiper, M. y Van de Peer, Y. (2005). *Modeling gene and genome duplications in eukaryotes*. Proceedings of the National

- Academy of Sciences of the United States of America, 102(15), 5454–5459.
- Magnani, E. y Barton, M.K. (2011). *A per-ARNT-sim-like sensor domain uniquely regulates the activity of the homeodomain leucine zipper transcription factor REVOLUTA in Arabidopsis*. *The Plant Cell*, 23(2), 567–5682.
- Manavella, P.A.; Arce, A.L.; Dezar, C.A.; Bitton, F.; Renou, J.P.; Crespi, M. y Chan, R.L. (2006). *Cross-talk between ethylene and drought signalling pathways is mediated by the sunflower Hahb-4 transcription factor*. *The Plant Journal*, 48(1), 125–137.
- Manavella, P.A.; Dezar, C.A.; Ariel, F.D. y Chan, R.L. (2008a). *Two ABREs, two redundant root-specific and one W-box cis-acting elements are functional in the sunflower HAHB4 promoter*. *Plant Physiology and Biochemistry*, 46(10), 860–867.
- Manavella, P.A.; Dezar, C.A.; Bonaventure, G.; Baldwin, I.T. y Chan, R.L. (2008b). *HAHB4, a sunflower HD-Zip protein, integrates signals from the jasmonic acid and ethylene pathways during wounding and biotic stress responses*. *The Plant Journal*, 56(3), 376–388.
- Mao, L.; Van Hemert, J.L.; Dash, S. y Dickerson, J.A. (2009). *Arabidopsis gene co-expression network and its functional modules*. *BMC Bioinformatics*, 10, 346.
- Margolin, A.A. y Califano, A. (2007). *Theory and limitations of genetic network inference from microarray data*. *Annals of the New York Academy of Sciences*, 1115, 51–72.
- Margolin, A.A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Dalla Favera, R. y Califano, A. (2006a). *ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context*. *BMC Bioinformatics*, 7 Suppl 1, S7.
- Margolin, A.A.; Wang, K.; Lim, W.K.; Kustagi, M.; Nemenman, I. y Califano, A. (2006b). *Reverse engineering cellular networks*. *Nature Protocols*, 1(2), 662–671.
- Mattsson, J.; Ckurshumova, W. y Berleth, T. (2003). *Auxin signaling in Arabidopsis leaf vascular development*. *Plant Physiology*, 131(3), 1327–1339.
- Mattsson, J.; Söderman, E.; Svenson, M.; Borkird, C. y Engström, P. (1992). *A new homeobox-leucine zipper gene from Arabidopsis thaliana*. *Plant Molecular Biology*, 18(5), 1019–1022.
- Maynard Smith, J. y Szathmáry, E. (1995). *The major transitions in evolution*. Oxford University Press.
- Medini, D.; Serruto, D.; Parkhill, J.; Relman, D.A.; Donati, C.; Moxon, R.; Falkow, S. y Rappuoli, R. (2008). *Microbiology in the post-genomic era*. *Nature Reviews Microbiology*, 6(6), 419–430.
- Meijer, A.H.; de Kam, R.J.; D’Erfurth, I.; Shen, W. y Hoge, J.H. (2000). *HD-Zip proteins of families I and II from rice: interactions and functional properties*. *Molecular & General Genetics*, 263(1), 12–21.



- Meijer, A.H.; Scarpella, E.; van Dijk, E.L.; Qin, L.; Taal, A.J.; Rueb, S.; Harrington, S.E.; McCouch, S.R.; Schilperoort, R.A. y Hoge, J.H. (1997). *Transcriptional repression by Oshox1, a novel homeodomain leucine zipper protein from rice*. *The Plant Journal*, 11(2), 263–276.
- Metzker, M.L. (2009). *Sequencing technologies — the next generation*. *Nature Reviews Genetics*, 11(1), 31–46.
- Miller, M.J.; Barrett-Wilt, G.a.; Hua, Z. y Vierstra, R.D. (2010). *Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, 107(38), 16512–16517.
- Mitsuda, N. y Ohme-Takagi, M. (2009). *Functional analysis of transcription factors in Arabidopsis*. *Plant & Cell Physiology*, 50(7), 1232–1348.
- Miura, K. y Hasegawa, P.M. (2010). *Sumoylation and other ubiquitin-like post-translational modifications in plants*. *Trends in Cell Biology*, 20(4), 223–232.
- Miura, K.; Jin, J.B. y Hasegawa, P.M. (2007). *Sumoylation, a post-translational regulatory process in plants*. *Current Opinion in Plant Biology*, 10(5), 495–502.
- Mockler, T.C.; Chan, S.; Sundaresan, A.; Chen, H.; Jacobsen, S.E. y Ecker, J.R. (2005). *Applications of DNA tiling arrays for whole-genome analysis*. *Genomics*, 85(1), 1–15.
- Moreira, D. y López-García, P. (2009). *Ten reasons to exclude viruses from the tree of life*. *Nature Reviews Microbiology*, 7(4), 306–311.
- Mukherjee, K.; Brocchieri, L. y Bürglin, T.R. (2009). *A comprehensive classification and evolutionary analysis of plant homeobox genes*. *Molecular Biology and Evolution*, 26(12), 2775–2794.
- Mukherjee, K. y Bürglin, T.R. (2006). *MEKHLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins*. *Plant Physiology*, 140(4), 1142–1150.
- Murashige, T. y Skoog, F. (1962). *A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures*. *Physiologia Plantarum*, 15(3), 473–497.
- Mutwil, M.; Usadel, B.; Schütte, M.; Loraine, A.; Ebenhöf, O. y Persson, S. (2010). *Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm*. *Plant Physiology*, 152(1), 29–43.
- Needham, C.J.; Manfield, I.W.; Bulpitt, A.J.; Gilmartin, P.M. y Westhead, D.R. (2009). *From gene expression to gene regulatory networks in Arabidopsis thaliana*. *BMC Systems Biology*, 3, 85.
- Nguyen Ba, A.N.; Pogoutse, A.; Provart, N. y Moses, A.M. (2009). *NLStradamus: a simple*

- Hidden Markov Model for nuclear localization signal prediction.* BMC Bioinformatics, 10, 202.
- Ohgishi, M.; Oka, A.; Morelli, G.; Ruberti, I. y Aoyama, T. (2001). *Negative autoregulation of the Arabidopsis homeobox gene ATHB-2.* The Plant Journal, 25(4), 389–398.
- Ohta, M.; Matsui, K.; Hiratsu, K.; Shinshi, H. y Ohme-Takagi, M. (2001). *Repression domains of class II ERF transcriptional repressors share an essential motif for active repression.* The Plant Cell, 13(8), 1959–1968.
- Olsson, A.S.B.; Engström, P. y Söderman, E. (2004). *The homeobox genes ATHB12 and ATHB7 encode potential regulators of growth in response to water deficit in Arabidopsis.* Plant Molecular Biology, 55(5), 663–677.
- Ontology, T.G.C. (2010). *The Gene Ontology in 2010: extensions and refinements.* Nucleic Acids Research, 38(Database issue), D331–335.
- Ooms, G.; Hooykaas, P.J.; Van Veen, R.J.; Van Beelen, P.; Regensburg-Tuïnk, T.J. y Schilperoort, R.A. (1982). *Octopine Ti-plasmid deletion mutants of agrobacterium tumefaciens with emphasis on the right side of the T-region.* Plasmid, 7(1), 15–29.
- Overnäs, E. (2010). *Characterization of Members of the HD-Zip I and DREB/ERF Transcription Factor Families and their Functions in Plant Stress Responses.* [uu.diva-portal.org/smash/get/diva2:337900/FULLTEXT01](http://uu.diva-portal.org/smash/get/diva2:337900/FULLTEXT01). Tesis de Doctorado.
- Palena, C.M.; Chan, R.L. y Gonzalez, D.H. (1997). *A novel type of dimerization motif, related to leucine zippers, is present in plant homeodomain proteins.* Biochimica et Biophysica Acta, 1352(2), 203–212.
- Palena, C.M.; Gonzalez, D.H. y Chan, R.L. (1999). *A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein Hahb-4 with DNA.* Biochemical Journal, 341, 81–87.
- Palena, C.M.; Tron, A.E.; Bertoncini, C.W.; Gonzalez, D.H. y Chan, R.L. (2001). *Positively charged residues at the N-terminal arm of the homeodomain are required for efficient DNA binding by homeodomain-leucine zipper proteins.* Journal of Molecular Biology, 308(1), 39–47.
- Pfaffl, M.W. (2010). *The ongoing evolution of qPCR.* Methods, 50(4), 215–216.
- Prigogine, I. (1988). *¿Tan solo una ilusión?* Barcelona, España: Tusquets.
- Ptashne, M. y Gann, A. (2002). *Genes & Signals.* Cold Spring Harbor Laboratory Press.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.
- Ravasz, E. y Barabási, A.L. (2003). *Hierarchical organization in complex networks.* Physical

- Review E, 67(2), 1–7.
- Ré, D.A.; Dezar, C.A.; Chan, R.L.; Baldwin, I.T. y Bonaventure, G. (2011). *Nicotiana attenuata NaHD20 plays a role in leaf ABA accumulation during water stress, benzylacetone emission from flowers, and the timing of bolting and flower transitions*. *Journal of Experimental Botany*, 62(1), 155–166.
- Redman, J.C.; Haas, B.J.; Tanimoto, G. y Town, C.D. (2004). *Development and evaluation of an Arabidopsis whole genome Affymetrix probe array*. *The Plant Journal*, 38(3), 545–561.
- Reeves, J.H. (1992). *Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA*. *Journal of Molecular Evolution*, 35(1), 17–31.
- Rerie, W.G.; Feldmann, K.A. y Marks, M.D. (1994). *The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis*. *Genes & Development*, 8(12), 1388–1399.
- Rice, P. (2000). *The European Molecular Biology Open Software Suite EMBOSS : The European Molecular Biology Open Software Suite*. *Science*, 16(6), 2–3.
- Riechmann, J.L. (2000). *Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes*. *Science*, 290(5499), 2105–2110.
- Roig-Villanova, I.; Bou, J.; Sorin, C.; Devlin, P.F. y Martínez-García, J.F. (2006). *Identification of primary target genes of phytochrome signaling. Early transcriptional control during shade avoidance responses in Arabidopsis*. *Plant Physiology*, 141(1), 85–96.
- Ruberti, I.; Sessa, G.; Lucchetti, S. y Morelli, G. (1991). *A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif*. *The EMBO Journal*, 10(7), 1787–1791.
- Rueda, E.C. (2008). *Estudios funcionales de los factores de transcripción de girasol de tipo HD-Zip en el desarrollo vegetal y la adaptación al medio ambiente*. Tesis de Doctorado.
- Rueda, E.C.; Dezar, C.A.; Gonzalez, D.H. y Chan, R.L. (2005). *Hahb-10, a sunflower homeobox-leucine zipper gene, is regulated by light quality and quantity, and promotes early flowering when expressed in Arabidopsis*. *Plant & Cell Physiology*, 46(12), 1954–1963.
- Ruiz-Mirazo, K.; Peretó, J. y Moreno, A. (2004). *A universal definition of life: autonomy and open-ended evolution*. *Origins of Life and Evolution of the Biosphere*, 34(3), 323–346.
- Rushton, P.J.; Somssich, I.E.; Ringler, P. y Shen, Q.J. (2010). *WRKY transcription factors*. *Trends in Plant Science*, 15(5), 247–258.
- Saddic, L.A.; Huvermann, B.; Bezhani, S.; Su, Y.; Winter, C.M.; Kwon, C.S.; Collum, R.P. y Wagner, D. (2006). *The LEAFY target LMII is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER*. *Development*, 133(9), 1673–1682.

- Saiki, R.; Scharf, S.; Faloona, F.; Mullis, K.; Horn, G.; Erlich, H. y Arnheim, N. (1985). *Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia*. *Science*, 230(4732), 1350–1354.
- Sakakibara, K.; Nishiyama, T.; Kato, M. y Hasebe, M. (2001). *Isolation of homeodomain-leucine zipper genes from the moss *Physcomitrella patens* and the evolution of homeodomain-leucine zipper genes in land plants*. *Molecular Biology and Evolution*, 18(4), 491–502.
- Sakuma, S.; Pourkheirandish, M.; Matsumoto, T.; Koba, T. y Komatsuda, T. (2010). *Duplication of a well-conserved homeodomain-leucine zipper transcription factor gene in barley generates a copy with more specific functions*. *Functional & Integrative Genomics*, 10(1), 123–133.
- Sawa, S.; Ohgishi, M.; Goda, H.; Higuchi, K.; Shimada, Y.; Yoshida, S. y Koshiba, T. (2002). *The HAT2 gene, a member of the HD-Zip gene family, isolated as an auxin inducible gene by DNA microarray screening, affects auxin response in Arabidopsis*. *The Plant Journal*, 32(6), 1011–1022.
- Scarpella, E.; Boot, K.; Rueb, S. y Meijer, A.H. (2002). *The procambium specification gene *Oshox1* promotes polar auxin transport capacity and reduces its sensitivity toward inhibition*. *Plant Physiology*, 130(3), 1349–1360.
- Scarpella, E.; Rueb, S.; Boot, K.J.; Hoge, J.H. y Meijer, A.H. (2000). *A role for the rice homeobox gene *Oshox1* in provascular cell fate commitment*. *Development*, 127(17), 3655–3669.
- Scarpella, E.; Simons, E.J. y Meijer, A.H. (2005). *Multiple regulatory elements contribute to the vascular-specific expression of the rice HD-Zip gene *Oshox1* in Arabidopsis*. *Plant & Cell Physiology*, 46(8), 1400–1410.
- Schena, M. y Davis, R.W. (1992). *HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily*. *Proceedings of the National Academy of Sciences of the United States of America*, 89(9), 3894–3898.
- Schena, M.; Lloyd, A. y Davis, R.W. (1993). *The HAT4 gene of Arabidopsis encodes a developmental regulator*. *Genes & Development*, 7(3), 367–379.
- Schena, M.; Shalon, D.; Davis, R.W. y Brown, P. (1995). *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. *Science*, 270(5235), 467–470.
- Schrödinger, E. (1944). *What is life?* Cambridge, Inglaterra: Cambridge University Press.
- Sessa, G.; Carabelli, M.; Sassi, M.; Ciolfi, A.; Possenti, M.; Mitterpergher, F.; Becker, J.; Morelli, G. y Ruberti, I. (2005). *A dynamic balance between gene activation and repression regulates the shade avoidance response in Arabidopsis*. *Genes & Development*, 19(23), 2811–2815.
- Sessa, G.; Morelli, G. y Ruberti, I. (1993). *The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities*. *The EMBO Journal*, 12(9), 3507–

- 3017.
- Sessa, G.; Morelli, G. y Ruberti, I. (1997). *DNA-binding specificity of the homeodomain-leucine zipper domain*. *Journal of Molecular Biology*, 274(3), 303–309.
- Shannon, C.E. (1948). *The mathematical theory of communication*. *The Bell System Technical Journal*, 27, 379–423 y 623–656.
- Sherman, F. (1991). *Getting started with yeast*. *Methods Enzymology*, 194(3), 3–41.
- Shi, L.; Katavic, V.; Yu, Y.; Kunst, L. y Haughn, G. (2012). *Arabidopsis glabra2 mutant seeds deficient in mucilage biosynthesis produce more oil*. *The Plant Journal*, 69(1), 37–46.
- Shin, D.; Koo, Y.D.; Lee, J.; Lee, H.j.; Baek, D.; Lee, S.; Cheon, C.I.; Kwak, S.S.; Lee, S.Y. y Yun, D.J. (2004). *Athb-12, a homeobox-leucine zipper domain protein from Arabidopsis thaliana, increases salt tolerance in yeast by regulating sodium exclusion*. *Biochemical and Biophysical Research Communications*, 323(2), 534–540.
- Söderman, E.; Hjellström, M.; Fahleson, J. y Engström, P. (1999). *The HD-Zip gene ATHB6 in Arabidopsis is expressed in developing leaves, roots and carpels and up-regulated by water deficit conditions*. *Plant Molecular Biology*, 40(6), 1073–1083.
- Söderman, E.; Mattsson, J. y Engström, P. (1996). *The Arabidopsis homeobox gene ATHB-7 is induced by water deficit and by abscisic acid*. *The Plant Journal*, 10(2), 375–381.
- Söderman, E.; Mattsson, J.; Svenson, M.; Borkird, C. y Engström, P. (1994). *Expression patterns of novel genes encoding homeodomain leucine-zipper proteins in Arabidopsis thaliana*. *Plant Molecular Biology*, 26(1), 145–154.
- Son, O.; Cho, H.t.; Kim, M.; Lee, H.; Lee, M.; Song, E.; Park, J.; Nam, K.; Chun, J.; Kim, H.; Hong, S.K.; Chung, Y.Y.; Hur, C.G.; Cho, H.T. y Cheon, C.I. (2005). *Induction of a homeodomain-leucine zipper gene by auxin is inhibited by cytokinin in Arabidopsis roots*. *Biochemical and Biophysical Research Communications*, 326(1), 203–209.
- Son, O.; Hur, Y.S.; Kim, Y.K.; Lee, H.J.; Kim, S.; Kim, M.R.; Nam, K.H.; Lee, M.S.; Kim, B.Y.; Park, J.; Park, J.; Lee, S.C.; Hanada, A.; Yamaguchi, S.; Lee, I.J.; Kim, S.K.; Yun, D.J.; Söderman, E. y Cheon, C.I. (2010). *ATHB12, an ABA-inducible homeodomain-leucine zipper (HD-Zip) protein of Arabidopsis, negatively regulates the growth of the inflorescence stem by decreasing the expression of a gibberellin 20-oxidase gene*. *Plant & Cell Physiology*, 51(9), 1537–1547.
- Sonnhammer, E.L.; Eddy, S.R. y Durbin, R. (1997). *Pfam: a comprehensive database of protein domain families based on seed alignments*. *Proteins*, 28(3), 405–420.
- Sorin, C.; Salla-Martret, M.; Bou-Torrent, J.; Roig-Villanova, I. y Martínez-García, J.F. (2009). *ATHB4, a regulator of shade avoidance, modulates hormone response in Arabidopsis seedlings*.

- The Plant Journal, 59(2), 266–277.
- Stajich, J.E.; Block, D.; Boulez, K.; Brenner, S.E.; Chervitz, S.a.; Dagdigian, C.; Fuellen, G.; Gilbert, J.G.R.; Korf, I.; Lapp, H.; Lehväsaiho, H.; Matsalla, C.; Mungall, C.J.; Osborne, B.I.; Pocock, M.R.; Schattner, P.; Senger, M.; Stein, L.D.; Stupka, E.; Wilkinson, M.D. y Birney, E. (2002). *The Bioperl toolkit: Perl modules for the life sciences*. Genome Research, 12(10), 1611–1618.
- Steindler, C.; Carabelli, M.; Borello, U.; Morelli, G. y Ruberti, I. (1997). *Phytochrome A, phytochrome B and other phytochrome(s) regulate ATHB-2 gene expression in etiolated and green Arabidopsis plants*. Plant, Cell & Environment, 20(6), 759–763.
- Steindler, C.; Matteucci, A.; Sessa, G.; Weimar, T.; Ohgishi, M.; Aoyama, T.; Morelli, G. y Ruberti, I. (1999). *Shade avoidance responses are mediated by the ATHB-2 HD-zip protein, a negative regulator of gene expression*. Development, 126(19), 4235–4245.
- Toledo-Ortiz, G.; Huq, E. y Quail, P. (2003). *The Arabidopsis basic/helix-loop-helix transcription factor family*. The Plant Cell, 15(8), 1749–1770.
- Tron, A.E.; Bertoncini, C.W.; Chan, R.L. y Gonzalez, D.H. (2002). *Redox regulation of plant homeodomain transcription factors*. The Journal of Biological Chemistry, 277(38), 34800–34807.
- Tron, A.E.; Bertoncini, C.W.; Palena, C.M.; Chan, R.L. y Gonzalez, D.H. (2001). *Combinatorial interactions of two amino acids with a single base pair define target site specificity in plant dimeric homeodomain proteins*. Nucleic Acids Research, 29(23), 4866–4872.
- Tron, A.E.; Comelli, R.N. y Gonzalez, D.H. (2005). *Structure of homeodomain-leucine zipper/DNA complexes studied using hydroxyl radical cleavage of DNA and methylation interference*. Biochemistry, 44(51), 16796–16803.
- Tron, A.E.; Welchen, E. y Gonzalez, D.H. (2004). *Engineering the loop region of a homeodomain-leucine zipper protein promotes efficient binding to a monomeric DNA binding site*. Biochemistry, 43(50), 15845–15851.
- Venters, B.J. y Pugh, B.F. (2009). *How eukaryotic genes are transcribed*. Critical Reviews in Biochemistry and Molecular Biology, 44(2-3), 117–141.
- Vinces, M.D.; Legendre, M.; Caldara, M.; Hagihara, M. y Verstrepen, K.J. (2009). *Unstable tandem repeats in promoters confer transcriptional evolvability*. Science, 324(5931), 1213–1216.
- Vinson, C.; Acharya, A. y Tapparowsky, E.J. (2006). *Deciphering B-ZIP transcription factor interactions in vitro and in vivo*. Biochimica et Biophysica Acta, 1759(1-2), 4–12.
- Vollbrecht, E.; Veit, B.; Sinha, N. y Hake, S. (1991). *The developmental gene Knotted-1 is a*

- member of a maize homeobox gene family*. *Nature*, 350, 241–243.
- Wang, Y.; Henriksson, E.; Söderman, E.; Henriksson, K.; Sundberg, E. y Engström, P. (2003). *The Arabidopsis homeobox gene, ATHB16, regulates leaf development and the sensitivity to photoperiod in Arabidopsis*. *Developmental Biology*, 264(1), 228–239.
- Weigel, D. (2011). *Natural variation in Arabidopsis thaliana: from molecular genetics to ecological genomics*. *Plant Physiology*, 158(January), 2–22.
- Wenkel, S.; Emery, J.; Hou, B.H.; Evans, M.M.S. y Barton, M.K. (2007). *A feedback regulatory module formed by LITTLE ZIPPER and HD-ZIPIII genes*. *The Plant Cell*, 19(11), 3379–3390.
- Wilkinson, K.A. y Henley, J.M. (2010). *Mechanisms, regulation and consequences of protein SUMOylation*. *The Biochemical Journal*, 428(2), 133–145.
- Wray, G.A.; Hahn, M.W.; Abouheif, E.; Balhoff, J.P.; Pizer, M.; Rockman, M.V. y Romano, L.a. (2003). *The evolution of transcriptional regulation in eukaryotes*. *Molecular Biology and Evolution*, 20(9), 1377–1419.
- Wu, Z.; Irizarry, R.A.; Gentleman, R.; Martinez-Murillo, F. y Spencer, F. (2004). *A Model-Based Background Adjustment for Oligonucleotide Expression Arrays*. *Journal of the American Statistical Association*, 99(468), 909–917.
- Xiong, Y.; Liu, T.; Tian, C.; Sun, S.; Li, J. y Chen, M. (2005). *Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots*. *Plant Molecular Biology*, 59(1), 191–203.
- Yang, Z. (1993). *Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites*. *Molecular Biology and Evolution*, 10(6), 1396–1401.
- Yeung, R.W. (2008). *Information Theory and Network Coding*. New York, New York, Estados Unidos: Springer.
- Yilmaz, A.; Mejia-Guerra, M.K.; Kurz, K.; Liang, X.; Welch, L. y Grotewold, E. (2011). *AGRIS: the Arabidopsis Gene Regulatory Information Server, an update*. *Nucleic Acids Research*, 39(Database issue), D1118–D1122.
- Yoo, C.Y.; Miura, K.; Jin, J.B.; Lee, J.; Park, H.C.; Salt, D.E.; Yun, D.J.; Bressan, R.a. y Hasegawa, P.M. (2006). *SIZ1 small ubiquitin-like modifier E3 ligase facilitates basal thermotolerance in Arabidopsis independent of salicylic acid*. *Plant Physiology*, 142(4), 1548–1558.
- Yu, Z.; Syu, L.J. y Mellerick, D.M. (2005). *Contextual interactions determine whether the Drosophila homeodomain protein, Vnd, acts as a repressor or activator*. *Nucleic Acids Research*, 33(1), 1–12.
- Zanetti, M.E.; Chan, R.L.; Godoy, A.; Gonzalez, D.H. y Casalengué, C. (2004). *Homeodomain-leucine zipper proteins interact with a plant homologue of the transcriptional co-activator*

## BIBLIOGRAFÍA

---

*multiprotein bridging factor 1*. *Journal of Biochemistry and Molecular Biology*, 37(3), 320–324.

Zhang, Y. y Wang, L. (2005). *The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants*. *BMC Evolutionary Biology*, 5, 1.