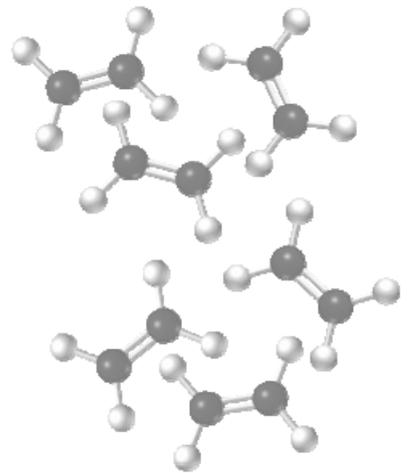


IV - CAPÍTULO 1

Análisis del transcriptoma en plantas transgénicas
que expresan el gen *HAHB4*



IV. - CAPITULO 1

Análisis del transcriptoma en plantas transformadas que expresan el gen *HAHB4*.

IV.1 – Introducción

El factor de transcripción HAHB4 confiere a las plantas de *Arabidopsis* transformadas de forma tal que lo sobreexpresen una serie de características fenotípicas diferenciales como la tolerancia a estrés hídrico. Con el interés de determinar los mecanismos moleculares que intervienen en la generación de estas características fenotípicas se decidió realizar un análisis detallado y comparativo del transcriptoma de las plantas transgénicas con respecto al de las plantas del mismo genotipo sin transformar. Consideramos que obteniendo esta valiosa información, nos acercaríamos a dilucidar la función de HAHB4 en su ambiente genómico natural, la planta de girasol.

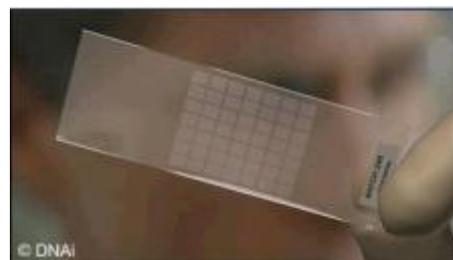
Se han desarrollado numerosas metodologías para realizar el análisis masivo del patrón de expresión génica de una célula o de un organismo. Entre estas estrategias podemos nombrar el análisis serial de expresión genética (SAGE), desplazamiento diferencial (*diferencial display*), los ADNc-AFLPs (del inglés “*amplified fragment length polymorphism*”) y otros. Estas técnicas, que fueron revolucionarias en el momento de su desarrollo, presentan desafortunadamente algunas desventajas como grandes errores experimentales, ser tediosas y sobre todo la imposibilidad de analizar simultáneamente la expresión del genoma completo. Muchos de estos aspectos negativos se atenúan con el uso de microarreglos, una técnica de desarrollo relativamente reciente con respecto a las nombradas. Los microarreglos, además de constituir una técnica extremadamente sensible, permiten el análisis de un número muy grande de genes en forma simultánea. En cierto modo el desarrollo de esta técnica se ha dado gracias a la miniaturización de técnicas moleculares previas que utilizan como principio la hibridización específica de las moléculas de ácidos nucleicos

Intentando recordar un poco la historia, fue a finales de la década de los 80 que una compañía llamada Affymax (actualmente Affymetrix) desarrolló la síntesis de oligonucleótidos sobre una superficie sólida en un sistema miniaturizado. El primer trabajo que utiliza esta técnica fue publicado en el año 1995 en la revista *Science*. Es

notorio que este primer trabajo fuera hecho con plantas de Arabidopsis y por el grupo del Dr. Schena, quien a su vez publicó en el año 1992 el trabajo pionero sobre los genes de tipo HD-Zip, tema central de este trabajo de Tesis (Schena y col., 1995).

En forma resumida, los microarreglos de ADN son soportes sólidos en los que se encuentran inmovilizados de forma ordenada entre cientos y miles de genes o segmentos de genes. Desde el punto de vista metodológico, las secuencias de ADN pueden sintetizarse sobre el soporte o depositarse una vez sintetizadas. Existen principalmente de dos tipos de microarreglos de acuerdo a la forma en la que se obtienen y a la posibilidad de analizar simultáneamente varias condiciones:

a) Microarreglos de puntos de ADN u oligonucleótidos. Se conocen también como microarreglos de dos colorantes y permiten comparar dos muestras distintas en la misma laminilla. En este caso, las moléculas de interés, sean oligonucleótidos o representen ADNc, se depositan directamente sobre la superficie sólida por medio de un robot.



b) Microarreglos de oligonucleótidos de alta densidad que sólo permiten monitorear una condición por arreglo. Estos tienen una precisión y densidad muy alta ya que usan oligonucleotidos cortos con una longitud de hasta 20 a 25 nucleótidos. Se fabrican en forma comercial por compañías como Affymetrix entre otras.



Estas dos tecnologías presentan comparativamente entre sí ciertas ventajas y desventajas. Los de alta densidad son más costosos y no poseen mucha flexibilidad en su diseño. Sin embargo, su gran precisión permite utilizarlos para analizar un número mucho mayor de genes y realizar mayor cantidad de controles internos. Presentan además, una calidad de fabricación muy alta que logra minimizar los errores implícitos de la impresión. Por su lado los de puntos permiten la medición y comparación simultánea de dos muestras de ARN en la misma laminilla y presentan una mayor flexibilidad en su diseño, lo que facilita el diseño experimental.

Para utilizar la técnica de microarreglos correctamente es importante tener en cuenta algunos puntos:

- La información obtenida no es cuantitativa sino que ofrece datos comparativos entre niveles de expresión en distintas situaciones
 - Los microarreglos no ofrecen información sobre la expresión última de los genes, o sea las proteínas, sino sobre la abundancia relativa de los mensajeros
 - La técnica permite detectar diferencias pequeñas en la abundancia de transcritos expresados en distintas condiciones experimentales. Por lo tanto las muestras a comparar deben ser obtenidas siguiendo protocolos estrictos con el fin de evitar artefactos.
- La variabilidad biológica es, en principio, la mayor fuente de variabilidad presente en el sistema, por lo que los datos obtenidos serán más representativos y confiables cuando más replicados biológicos se incluyan en el análisis.
- Existen dos problemas inherentes a la técnica:
 - Los niveles de expresión tiene un alto nivel de variabilidad entre experimentos. Esta variabilidad se debe al gran número de posibles fuentes de errores, tanto de origen aleatorio como sistemático. Es fundamental utilizar técnicas de normalización, pre-procesamiento y filtrado para minimizar los efectos causados por este problema.
 - Debido a los altos costos, se utiliza un número pequeño de muestras comparado con el gran número de variables. Como consecuencia, las técnicas estadísticas tradicionales fracasan y es necesario utilizar técnicas de reducción de dimensionalidad.

Se han escrito revisiones completas sobre estas metodologías y los puntos a tener en cuenta para utilizarlas correctamente. A modo de ejemplo, pueden revisarse los suplementos de la revista *Nature Genomics* “*The Chipping Forecast I* y *II* (enero de 1999, vol 21 numero 1 y diciembre de 2002, vol 32, numero 4 respectivamente).

Actualmente se cuenta con una variante de la tecnología de puntos disponible para los estudios en *Arabidopsis* conocida como microarreglos de CATMA (del inglés “*Complete Arabidopsis Transcriptome MicroArray*”). Estos arreglos presentan una alta calidad de impresión y permiten la detección de casi todos los transcritos de *Arabidopsis* así como muchas de sus variables causadas por el proceso de “corte y empalme” (*splicing*, Hilson *et al.*, 2004).

El proyecto CATMA nació con el objetivo de diseñar y producir microarreglos de alta calidad que contuvieran todos los genes de *Arabidopsis* representados en GSTs (del inglés *Gene Specific Sequence Tags*). CATMA es un consorcio de grupos de investigación de instituciones públicas pertenecientes a ocho países europeos. En la actualidad el proyecto CATMA posee 30,886 GSTs en su base de datos (Crowe y col., 2003). Los GSTs usados por CATMA son fragmentos genómicos amplificados por PCR cuyos tamaños oscilan entre 150 a 500 pb diseñados de forma tal que en ninguno de ellos más del 50% representa secuencias intrónicas. Para el diseño de los GSTs específicos de cada gen de *Arabidopsis* y de los oligonucleótidos necesarios para su amplificación se desarrolló un programa específico llamado SPADS (del inglés “*Specific Primer & Amplicon Design Software*”, Thareau y col., 2003). Este programa permite, diseña y genera GSTs que representen todos los genes de *Arabidopsis* en forma rápida y sistemática. En los microarreglos de CATMA, además de tener representados genes repetidos que sirven como controles del experimento en sí, también están impresos diferentes GSTs que representan un mismo gen. Estos controles son de extrema utilidad para eliminar los artefactos generados por hibridaciones inespecíficas.

En el presente trabajo de Tesis se usaron microarreglos generados por el consorcio CATMA con el objeto de determinar los efectos del transgén *HAHB4* sobre el transcriptoma de *Arabidopsis*. Esta estrategia, como se describe en los capítulos siguientes, nos fue de suma utilidad para comenzar a comprender la función de este gen de girasol. .

IV.2 – Resultados y discusión

IV.2.1 – La expresión de HAHB4 genera cambios significativos en el transcriptoma de Arabidopsis thaliana.

Con el fin de dilucidar cuáles serían los mecanismos moleculares responsables de los cambios fenotípicos generados por la presencia de *HAHB4* en plantas de *Arabidopsis*, particularmente la tolerancia a estrés hídrico, se llevó a cabo un análisis transcriptómico comparando el conjunto de ARNm de plantas transgénicas (35S:*HAHB4*) y salvajes crecidas en condiciones normales o sometidas a estrés hídrico. El microarreglo de CATMA utilizado contiene 25328 GSTs que representan 24576 genes de *Arabidopsis*, más algunos controles y replicados internos. Los resultados indicaron que del total de genes analizados, 3322 presentaron alguna alteración en su nivel de expresión, al menos entre dos de las cuatro muestras.

Para poder analizar estos complejos resultados los genes se dividieron según su patrón de comportamiento, comparando las distintas muestras. En la figura IV.1 se ilustra un análisis de *clusters* realizado con los grupos de genes que variaron su expresión en las distintas muestras ensayadas (salvajes con estrés, transgénicas en condiciones controladas, transgénicas en condiciones de estrés) con respecto a los niveles de esos mismos genes en la muestra de plantas salvajes crecidas en condiciones controladas. En la figura IV.1A se muestra la distribución general de genes ya sean estos reprimidos o inducidos con respecto a la muestra control, mientras que en las IV.1B y IV.1C se graficaron solamente los genes activados o reprimidos respectivamente. En las intersecciones de las tres figuras se indica el número de genes que varía en dos o tres de las muestras. Como puede observarse, el número de genes totales incluidos en cada intersección (IV.1A) no coincide con la sumatoria de los indicados en los subconjuntos conformados por genes reprimidos o inducidos (IV.1B y IV.1C). Estas diferencias se deben a que muchos de los genes de las intersecciones en la figura A corresponden a genes que son reprimidos en una condición y activados en la otra, o sea un comportamiento mixto. Al graficar sólo los activados o sólo los reprimidos, se excluyen aquellos que tienen comportamiento opuesto en las distintas muestras.

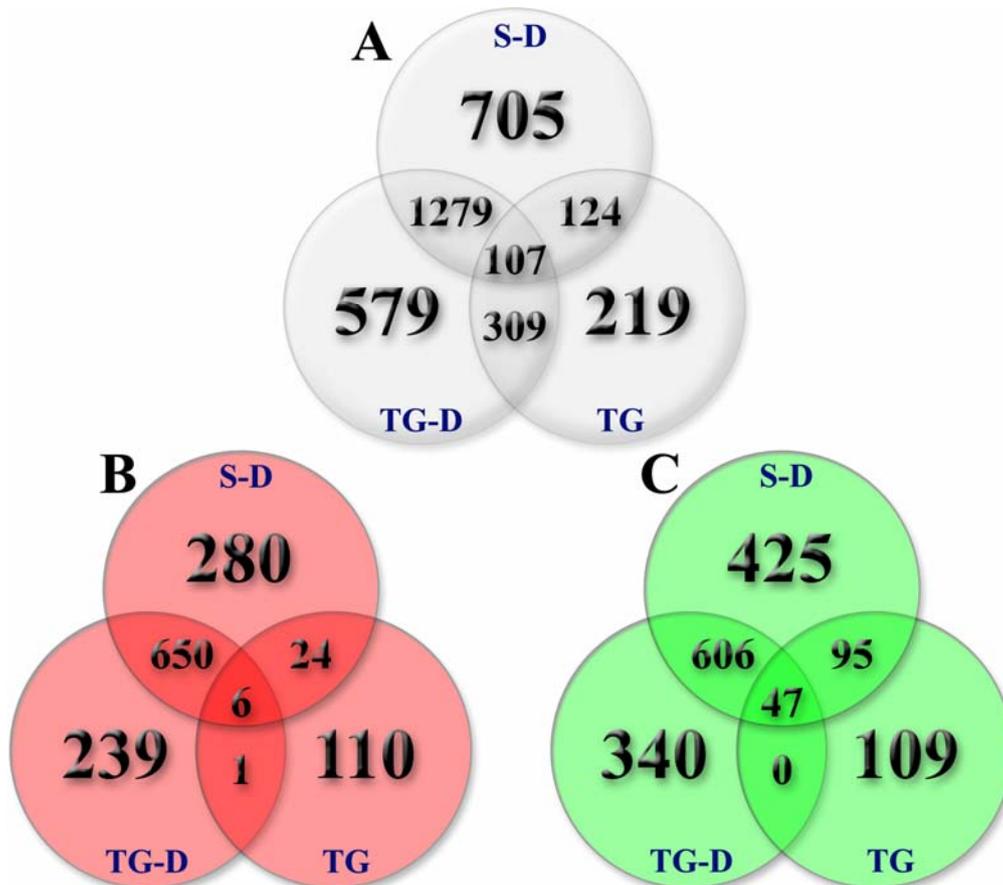


Figura IV.1: Distribución de genes según su expresión en las distintas muestras usadas en el ensayo de microarreglos. A: distribución general; B: sólo genes activados; C: sólo genes reprimidos. S: salvajes; TG: transgénicas de genotipo *35S:HAHB4*; D: estrés hídrico.

En una segunda instancia, los genes se dividieron de acuerdo a su comportamiento general. Esto nos permitió identificar con claridad aquellos grupos de genes cuya expresión es regulada en forma independiente de la presencia de *HAHB4*. En la figura IV.2 se ilustra el comportamiento de los 3322 genes cuya expresión se ve alterada al menos en una de las tres muestras tomando como patrón la muestra control. De esta forma, se identificaron 12 grupos que pueden a su vez subdividirse en dos de comportamiento casi especular. El gráfico muestra cuatro líneas verticales que representan las cuatro muestras (salvajes en condiciones control, salvajes sometidas a estrés hídrico, transgénicas en condiciones control y transgénicas sometidas a estrés hídrico). Para cada gen se considera si es reprimido o inducido en las distintas muestras independientemente de cuánto lo haga. De esta manera, hay un subgrupo de genes que aumenta su expresión en estrés en plantas salvajes y en plantas transgénicas en condiciones control, manteniéndose elevado cuando las plantas transgénicas son sometidas a condiciones de estrés hídrico (subgrupo A.1). Si bien el cuadro es bastante complejo, algo que se explica por la cantidad de genes que cambian su expresión en presencia del transgén, la figura ilustra claramente cuántos y cuáles genes pueden agruparse de acuerdo a cada uno de los comportamientos.

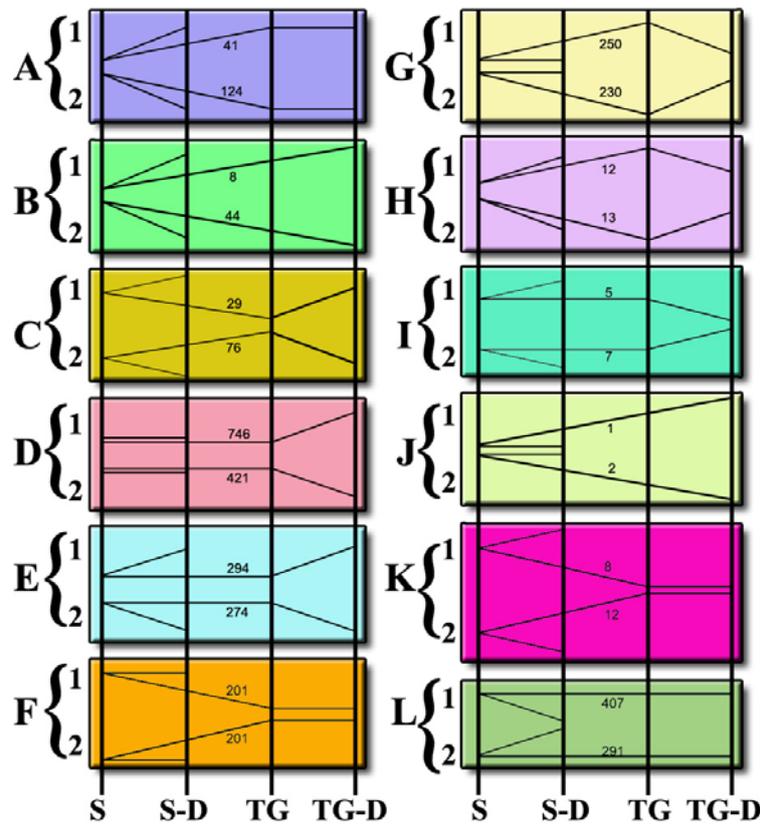


Figura IV.2: Distribución de los genes de acuerdo al comportamiento presentado en las distintas muestras con respecto a las muestras de plantas salvajes en condiciones control. Cada grupo se indica con una letra mayúscula y se subdivide en dos subgrupos de comportamiento opuesto. Los números pequeños dentro de cada cuadro indican la cantidad de genes que componen cada subgrupo. Las líneas internas indican el comportamiento de los genes de cada grupo comparando las muestras provenientes de plantas: salvajes en condiciones control (S); salvajes sometidas a estrés hídrico (S-D); transgénicas (TG) y transgénicas sometidas a estrés hídrico (TG-D).

El análisis representado en esta figura permite, en pos de dilucidar el efecto del transgén, eliminar los grupos de genes que varían su comportamiento debido al estrés, independientemente de la presencia del transgén, así como los que presentan un comportamiento difícil de explicar con el estado del conocimiento actual. Resumidamente, de los más de 3000 genes que varían su comportamiento entre las muestras, sólo 759 lo hacen como consecuencia de la presencia del transgén. Estos genes están representados principalmente en los subgrupos A, B, C, F y G.

IV.2.2 – CAAT(A/T)ATTG, la secuencia unida in vitro por HAHB4 está más representada en las regiones promotoras de los genes que cambian su

expresión como consecuencia de la presencia del transgén que en el conjunto del genoma.

Una vez seleccionados los 759 genes cuya expresión es afectada por la presencia de *HAHB4*, nos pareció interesante analizar cuáles de estos genes podrían actuar como blancos directos de este factor de transcripción. En trabajos previos realizados en nuestro laboratorio se determinó que la proteína codificada por *HAHB4* une *in vitro* en forma preferencial y con mucha afinidad la secuencia pseudopalindrómica CAAT(A/T)ATTG (Palena y col., 1999). Si bien esta misma secuencia es unida con mucha afinidad por otros miembros de la subfamilia I de proteínas HD-Zip y con algo menos de afinidad por los miembros de la subfamilia II, era de esperar que en el conjunto de genes cuya expresión se afecta por la presencia de *HAHB4* estuviera representada con mayor frecuencia que en el genoma total. A fin de corroborar o refutar esta hipótesis, se buscó esta secuencia en las 1000 pb corriente arriba del sitio de inicio de la transcripción de los 759 genes seleccionados, así como en la misma región de todos los otros genes de Arabidopsis (TAIR6 *genome release*), gracias al diseño de un programa bioinformático casero. El resultado de este análisis indicó que un 2,3% de los genes de Arabidopsis presenta en su región promotora cercana al menos una vez el pseudopalíndromo, mientras que un 8,5% de los 759 genes seleccionados también la presenta, sugiriendo un fuerte enriquecimiento en este conjunto de putativos blancos de *HAHB4*. Esto fortalece la hipótesis de que estos genes podrían ser los responsables directos de los efectos de *HAHB4*.

IV.2.3 – Los resultados obtenidos en el ensayo de microarreglos se confirman en estudios por RT-PCR cuantitativa

Los estudios de microarreglos, tal como se describió más arriba, suelen arrastrar varias fuentes de errores. De ahí es que es necesario plantearse determinar la validez de los resultados por métodos alternativos. En general el método de elección consiste en medir los niveles de varios transcritos detectados en el microarreglo como de expresión diferencial entre muestras por la técnica de RT-PCR cuantitativa. Con este fin, en principio se eligieron al azar veintidós genes pertenecientes al conjunto de los 759 modificados en su expresión por la presencia de *HAHB4*. Las medidas se realizaron sobre muestras biológicas nuevas y obtenidas en forma independiente de la originalmente utilizada para hibridar el microarreglo. En la figura IV.3 se ilustran los resultados obtenidos.

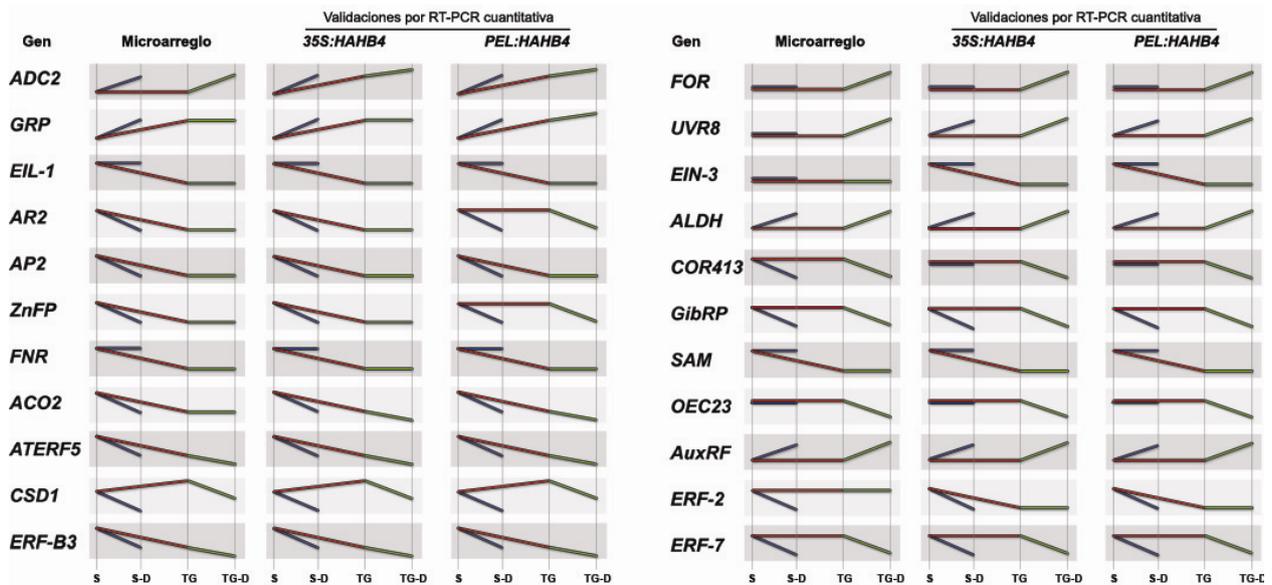


Figura IV.3: Validaciones de los datos obtenidos en el experimento de microarreglos. Representación gráfica de los niveles de expresión de 22 genes medidos por RT-PCR cuantitativa. En la primera columna se ilustra el comportamiento de los genes en el ensayo de microarreglos. En la segunda columna las medidas hechas en muestras obtenidas en forma independiente de plantas transgénicas del genotipo *35S:HAHB4* y en la tercera columna las del genotipo *PEL:HAHB4*. Salvajes (S); salvajes sometidas a estrés hídrico (S-D); transgénicas (TG) y transgénicas sometidas a estrés hídrico (TG-D). La variación entre: S y S-D (líneas azules); S y TG (líneas rojas); TG y TG-D (líneas verdes). Los genes cuya expresión fue medida, codifican las siguientes proteínas: ADC-2 (arginina-decarboxilasa), GRP (proteína rica en glicina), EIL-1 (similar a la proteína insensible a etileno 1), AR-2 (NADPH cyt-reductasa), AP2 (factor de respuesta a etileno con dominio AP2), ZnFP (proteína con dedos de zinc), FNR (ferredoxina-NADP-oxidoreductasa), ACO (ACC oxidasa), ERF-5 (factor 5 de respuesta a etileno), CSD-1 (Cu/Zn-superoxido-dismutasa), ERF-B3 (factor de respuesta a etileno de la subfamilia B-3), FOR (oxidoreductasa dependiente de FAD), UVR-8 (proteína de resistencia a UVB), EIN3 (insensible a etileno 3), ALDH (betaina-aldehido-deshidrogenasa), COR413 (proteína de aclimatación a frío), GibRP (proteína regulada por giberelinas), SAM (S-adenosilmetionina-sintetasa), OEC23 (complejo dependiente de oxígeno del fotosistema II), AuxRF (factor de respuesta a auxinas), ERF-2 (factor de respuesta a etileno 2), ERF-7 (factor de respuesta a etileno 7)

Como se observa en esta figura, existe una muy buena correlación entre los datos obtenidos en el ensayo de microarreglos (primera columna) y las medidas realizadas por RT-PCR en tiempo real (segunda columna). Sólo tres de los genes analizados (EIN-3, ERF-2 y ADC-2) muestran un patrón de expresión ligeramente diferente al obtenido en el microarreglo. Aun así, los niveles de expresión de estos tres genes cuantificados por RT-PCR coinciden con los obtenidos en dos de los tres

replicados biológicos usados en el microarreglo. La tabla IV.1 muestra los valores obtenidos en las validaciones expresados en forma relativa al control de plantas salvajes, tomado este último arbitrariamente como uno.

GEN	Salvaje	Salvaje-D	TG	TG-D
ADC2 (AT4g34710)	1±0.021	2,35±0.161	2,03±0.075	2,84±0.131
GRP (AT2g05520)	1±0.098	3,16±0.034	8,86±0.169	7,92±0.191
EIL1 (AT2g27050)	1±0.138	0,89±0.099	0,21±0.042	0,25±0.076
AR2 (AT4g30210)	1±0.101	0,60±0.013	0,72±0.072	0,65±0.041
AP2 (AT2g27050)	1±0.099	0,38±0.041	0,42±0.067	0,50±0.101
ZnFP (AT1g51700)	1±0.102	0,63±0.011	0,64±0.096	0,58±0.031
FNR (AT1g20020)	1±0.132	1,21±0.114	0,30±0.064	0,34±0.011
ACO (AT1g62380)	1±0.083	0,50±0.021	0,62±0.019	0,46±0.041
ERF5 (AT5g47230)	1±0.043	0,26±0.029	0,50±0.088	0,13±0.039
CSD1 (AT1g08830)	1±0.061	0,23±0.012	3,28±0.061	0,59±0.071
ERF B3 (AT5g61600)	1±0.098	0,45±0.035	0,36±0.012	0,13±0.068
FOR (AT5g67290)	1±0.090	1,04±0.021	1,16±0.081	9,79±0.291
UVR8 (AT5g63860)	1±0.033	1,71±0.030	1,12±0.036	1,74±0.041
EIN 3 (AT3g20770)	1±0.045	0,96±0.106	0,17±0.087	0,20±0.031
ALDH (AT1g74920)	1±0.081	2,14±0.136	1,06±0.084	3,31±0.270
COR413 (AT3g50830)	1±0.044	0,93±0.020	0,99±0.012	0,52±0.012
GibRP (AT2g14900)	1±0.096	0,63±0.069	1,27±0.176	0,44±0.120
SAM (AT2g36880)	1±0.114	1,12±0.139	0,32±0.082	0,30±0.121
OEC23 (AT1g06680)	1±0.111	0,87±0.171	1,02±0.051	0,52±0.093
AuxRF (AT5g62000)	1±0.129	10,44±0.434	0,97±0.238	7,12±0.073
ERF 2 (AT5g47220)	1±0.123	0,32±0.120	0,41±0.076	0,48±0.083
ERF 7 (AT3g20310)	1±0.067	0,24±0.074	1,17±0.126	0,58±0.071

Tabla IV.1: Valores numéricos relativos correspondientes a los niveles de expresión de los genes representados en la figura IV.3. TG: plantas transgénicas 35S:HAHB4; D: sequia. Las desviaciones estándar fueron calculadas para cada muestra usando triplicados biológicos. Las diferencias entre muestras fueron consideradas significativas si su Valor-p fue inferior a 0,01.

Con el fin de analizar el comportamiento de estos genes en plantas que expresan HAHB4 de forma inducible y no constitutiva, se midieron los niveles de transcritos de los mismos en las plantas transformadas con la construcción en la que la expresión del

gen es dirigida por su propio promotor. La figura IV.3 (tercera columna) y la tabla IV.2 muestran los datos obtenidos.

Los resultados obtenidos indican claramente que al menos, este conjunto de genes presenta un comportamiento muy similar en los dos genotipos transgénicos, constitutivo e inducible. Si bien las plantas de estos genotipos presentan características fenotípicas diferenciales, en algunos aspectos conservan la similitud. A modo de ejemplo, ambos son más tolerantes al estrés hídrico que el genotipo salvaje aunque la diferencia de comportamiento es mayor en las plantas que llevan el promotor de expresión constitutiva. El comportamiento de estos genes que podrían ser blancos directos e indirectos de *HAHB4* en los dos genotipos transgénicos podría explicarse también por el hecho de que todos ellos se regulan en condiciones de estrés hídrico (ver figuras y tablas anteriores). Como el promotor de *HAHB4* se regula positivamente por esta condición, dispararía de forma similar la respuesta.

GEN	Salvajes	Salvajes-D	TG	TG-S
ADC2 (AT4g34710)	1±0.029	1,89±0.013	3,65±0.129	7,86±0.349
GRP (AT2g05520)	1±0.072	2,52±0.129	4,90±0.193	5,47±0.123
EIL1 (AT2g27050)	1±0.079	0,91±0.098	0,23±0.019	0,19±0.041
AR2 (AT4g30210)	1±0.134	0,09±0.012	0,96±0.028	0,12±0.051
AP2 (AT2g27050)	1±0.152	0,10±0.061	0,08±0.042	0,02±0.060
ZnFP (AT1g51700)	1±0.114	0,21±0.051	1,28±0.127	0,66±0.095
FNR (AT1g20020)	1±0.182	0,85±0.096	0,32±0.119	0,42±0.042
ACO (AT1g62380)	1±0.098	0,24±0.070	0,52±0.026	0,16±0.034
ERF5 (AT5g47230)	1±0.072	0,43±0.021	0,49±0.031	0,19±0.024
CSD1 (AT1g08830)	1±0.096	0,16±0.072	2,93±0.168	0,29±0.032
ERF B3 (AT5g61600)	1±0.129	0,16±0.072	0,39±0.108	0,12±0.022
FOR (AT5g67290)	1±0.141	0,92±0.031	1,02±0.147	3,35±0.173
UVR8 (AT5g63860)	1±0.144	1,87±0.047	1,03±0.095	1,94±0.118
EIN 3 (AT3g20770)	1±0.125	1,28±0.149	0,31±0.041	0,29±0.129
ALDH (AT1g74920)	1±0.171	2,87±0.151	1,06±0.211	2,95±0.335
COR413 (AT3g50830)	1±0.131	0,94±0.049	0,91±0.045	0,17±0.044
GibRP (AT2g14900)	1±0.043	0,42±0.021	1,02±0.112	0,54±0.059
SAM (AT2g36880)	1±0.123	1,24±0.238	0,24±0.031	0,27±0.035
OE23 (AT1g06680)	1±0.036	0,95±0.038	0,92±0.041	0,60±0.071
AuxRF (AT5g62000)	1±0.129	4,39±0.375	1,12±0.151	4,89±0.298
ERF 2 (AT5g47220)	1±0.133	0,33±0.103	0,44±0.122	0,37±0.084
ERF 7 (AT3g20310)	1±0.141	0,41±0.078	1,09±0.134	0,28±0.105

Tabla IV.2: Valores numéricos correspondientes a los niveles de expresión de genes representados en la figura IV.3 relativos a los medidos en plantas utilizadas como controles, tomados arbitrariamente como uno. TG: plantas transgénicas *PEL:HAHB4*; D: sequia. Las desviaciones estándar fueron calculadas para cada muestra usando triplicados biológicos. Las diferencias entre muestras fueron consideradas significativas si su Valor p- fue inferior a 0,01.

IV.2.4 – El análisis funcional de los genes regulados por HAHB4 indica la participación de este gen en ciertas vías metabólicas.

El conocimiento de cuáles genes, dentro del genoma, podrían actuar como blancos de la acción de HAHB4 nos permitió en un siguiente paso investigar los procesos fisiológicos en los cuales participan y consecuentemente inferir la funcionalidad de HAHB4. Con el objeto de visualizar claramente las vías en las que participan estos genes blancos putativos, se recurrió al programa en línea fatiGO+ (Al-Shahrour y col., 2007; <http://babelomics.bioinfo.cipf.es/fatigoplus/cgi-bin/fatigoplus.cgi>) el cual está incluido en el paquete de programas Babelomics de acceso gratuito. Este programa ha sido diseñado para el análisis funcional de experimentos a escala genómica (Al-Shahrour y col., 2006) y utiliza las anotaciones GO (del inglés *Gene Ontology*) como principio básico para sus algoritmos. Estas anotaciones incluyen para cada gen sus funciones biológicas conocidas. Cada función biológica tiene una anotación GO específica. Hay genes que no poseen anotaciones GO ya que se desconoce su función y viceversa hay genes con varias funciones conocidas que poseen más de una anotación. Teniendo en cuenta estas consideraciones, las anotaciones GO se organizan según distintos niveles de jerarquía que van desde el 1 al 9, siendo el nivel 1 el de mayor jerarquía e incluyendo dentro de sí un grupo de anotaciones de nivel 2 así como el grupo de nivel 2 integra genes de nivel 3 y así sucesivamente. Las anotaciones van desde funciones generales (nivel 1) a funciones específicas (nivel 9). Por ejemplo, un gen puede tener una anotación de nivel 1 = procesos biológicos, nivel 3: síntesis de hormonas, nivel 5: síntesis etileno; nivel 9 = actividad ACC oxidasa.

Teniendo en cuenta estas anotaciones, el programa compara dos listas de genes, en nuestro caso los 759 genes regulados por *HAHB4* versus el genoma completo de *Arabidopsis*, y nos informa cuáles funciones GO están representadas significativamente en el grupo de genes seleccionados. Por ejemplo, si en nuestro grupo de genes el 39% pertenece al grupo “asimilación de nitrato” cuando los genes anotados con esta función

en el total del genoma sólo representan el 1% el programa nos indicará que el grupo de genes seleccionado tiene una probabilidad estadísticamente importante de cumplir funciones en ese proceso. El programa proporciona asimismo un valor p que le da un significado estadístico a la observación. .

La tabla IV.3 muestra los datos más relevantes obtenidos en el análisis con este programa. Estos datos nos orientan hacia algunos procesos biológicos en los que podría estar involucrado el factor de transcripción HAHB4.

Proceso biológico	Nivel GO	Porcentaje GP/ GC	p-Valor
Fotosíntesis	3	6,41% / 0,5%	3,4219e ⁻¹⁰
Procesos de biosíntesis	3	27,76% / 10,12%	2,8122e ⁻⁰⁷
Respuesta mediada por hormonas	6	14,45% / 4,01%	1,1661e ⁻⁰⁶
Modificación de proteínas	6	6,05% / 15,42%	9,5104e ⁻⁰⁵
Respuesta a estrés abiótico	3	10,41% / 4,02	4,4527e ⁻⁰⁴
Utilización de carbono	3	1,78% / 0,08%	2,5015e ⁻⁰³

Tabla IV.3: Resultados del análisis de funciones biológicas en fatiGO+. GP: porcentaje de genes en el grupo problema que cumplen una función biológica anotada en GO; GC: porcentaje de genes que cumplen la misma función tomando en cuenta el genoma completo de Arabidopsis.

El principal problema que presenta este tipo de análisis es que muchos de los genes de Arabidopsis no poseen ninguna anotación GO y además de los que sí tienen, la mayoría sólo incluye los niveles GO más altos (generales). El motivo es que hasta la fecha se desconoce la función de muchos genes de los cuales sí se conoce la secuencia. En muchos casos se conoce la participación del producto de un gen en un dado metabolismo pero no con exactitud qué función cumple. Por este motivo, este tipo de análisis, presentado en la tabla IV.3 sirve a modo de aproximación y no debe tomarse como definitivo.

Con el objeto de profundizar en la función de HAHB4, los resultados obtenidos en fatiGO+ se analizaron con el programa MapMan (Thimm y col., 2004; <http://gabi.rzpd.de/projects/MapMan/>) focalizándonos en los procesos biológicos que el análisis previo había detectado como posibles candidatos. Este nuevo estudio bioinformático nos permitió identificar algunas vías metabólicas en las que HAHB4 estaría involucrado.

En concordancia con los resultados obtenidos previamente utilizando el programa *fatiGO+*, el análisis realizado con el programa *MapMan* indicó que la expresión de genes involucrados en la fotosíntesis, así como también en la respuesta a etileno y ácido jasmónico se encontraban alterados en su expresión en las plantas transgénicas. Un dato interesante que surge de este análisis es el hecho que del subgrupo de genes (231) cuya expresión se modifica tanto en estrés hídrico como en las plantas transgénicas en condiciones tomadas como control (figura *IV.1* intersección S-S y TG), el 35% participa de la fotosíntesis o de la respuesta a etileno. Esta observación nos orienta hacia la hipótesis de que esos procesos podrían ser los que de alguna forma estén modulando la tolerancia estrés hídrico. Esta hipótesis se corroboró experimentalmente y la descripción detallada de cómo *HAHB4* participa de los procesos biológicos mencionados se presenta en los capítulos subsiguientes.

Basándonos en los resultados presentados en este capítulo podemos concluir sintéticamente que:

- el análisis del transcriptoma de plantas salvajes y transgénicas sometidas a estrés hídrico indica que existen 3322 genes con expresión diferencial entre las muestras.
- el ensayo de microarreglos fue validado exitosamente mostrando una correlación entre los efectos regulatorios del transgén expresado en forma constitutiva o inducible.
- Del conjunto de genes que modifican su expresión entre las muestras, 759 lo hacen debido a la presencia del transgén.
- los análisis bioinformáticas de los datos obtenidos en los ensayos de microarreglos y sus validaciones, indican que *HAHB4* estaría involucrado en procesos ligados a fotosíntesis, respuesta a etileno y a ácido jasmónico.