

Tesis Doctoral

Análisis genómico de girasol: desarrollo de colecciones de ESTs y de una plataforma bioinformática para estudios de expresión de genes candidato en respuestas a estreses abióticos

Fernández, Paula del Carmen

2006

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Fernández, Paula del Carmen. (2006). Análisis genómico de girasol: desarrollo de colecciones de ESTs y de una plataforma bioinformática para estudios de expresión de genes candidato en respuestas a estreses abióticos. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Fernández, Paula del Carmen. "Análisis genómico de girasol: desarrollo de colecciones de ESTs y de una plataforma bioinformática para estudios de expresión de genes candidato en respuestas a estreses abióticos". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2006.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES



***ANÁLISIS GENÓMICO DE GIRASOL:
Desarrollo de colecciones de ESTs y de una
plataforma bioinformática para estudios de
expresión de genes candidato en respuestas a
estreses abióticos***

**ING. AGR. PAULA DEL CARMEN FERNANDEZ
DIRECTOR: DRA. RUTH A. HEINZ**

**TESIS PARA OPTAR POR EL TÍTULO DE
DOCTOR EN CIENCIAS BIOLÓGICAS**

**INSTITUTO DE BIOTECNOLOGÍA
CICVyA INTA – CASTELAR
DICIEMBRE 2006**

A mis papás, A Ale y a Salvador:
introducción, nudo y desenlace de esta historia

AGRADECIMIENTOS

Como uno nunca logra nada solo, sino en grupo y con la ayuda de tantas personas necesarias en cuerpo y alma, quiero tratar de mencionarlas aquí brevemente, tratando de no olvidar a ninguna en particular...

- A la Universidad de Buenos Aires, por la beca otorgada en el año 2002 para que pudiera ejecutar este trabajo de tesis, y por la posibilidad de seguir brindando educación pública, gratuita y de calidad. A toda la sociedad que la sostiene.
- Al INTA, por lo que significa como institución, por su renombre y porque gracias a su gente como el Dr. Osvaldo Rossetti y la Dra. Elisa Carrillo tuve el apoyo institucional para desarrollar este trabajo.
- A Ruth, por todo lo que es como persona y profesional. Por su ayuda, su confianza, su tolerancia, su esfuerzo diario en que todo salga bien y una no se sienta “a la deriva”, su buen trato. Y su valiosa experiencia y conocimiento. Una directora presente siempre.
- A Norma, porque sin ella el segundo capítulo de mi tesis me hubiera desmoronado. Por aceptarme y sonreírme con tanta calidez el día de la primera entrevista, por tener más confianza en mí que yo misma, por poder contar con ella para todo. Y porque por cuestiones formales no pudo estar co-dirigiendo esta tesis pero fue co-directora a lo largo de estos años.
- A Esteban por haberme permitido trabajar en su grupo tan valioso y respetado, porque para él el conocimiento tiene valor, y eso, redundantemente, vale, hoy más que nunca.
- A Sergio Lew y todo su grupo de Alselec S.A. y BioAxioma (Gustavo, Darío e Iñaki). Fueron mi mano derecha en el análisis bioinformático de las secuencias y tuvieron mucha paciencia en mis largas horas que compartí la oficina con ustedes. Me sentí siempre como en casa. Gracias de corazón.
- A Julio Di Rienzo por haberme ayudado con todo su *expertise* estadístico, siempre con tantas ganas, logrando que las horas frente a los datos vuelen. Por su pragmatismo académico.
- A Diego Lijavetzky, por haberme recibido junto con Norma y Esteban, y guiarme en los primeros pasos.

- A Luis porque es un AMIGO, un trabajador, una excelente persona y colaborador. Me socorriste muchas veces en todo sin necesidad casi de pedirte. Gracias por las “*real time*” de último momento.
- A Caro, por tus palabras, tu contención, tu ayuda, tus tortas, tus corridas conmigo a la clínica aquel mediodía, y tu compañía. Sos un ejército de salvación.
- A Paola y Dani, los remolinos del laboratorio. Por tantos diálogos “*existenciales*” y “*mundanos*”, por poder hablar de marcadores moleculares y de desigualdad social a la vez, por ayudarme a no diluir el objetivo central de todo esto. Por compartir buenos libros, buena música y las noticias por la tarde. Por ayudarme a no perder el árbol en el bosque.
- A Vero Nishi. No se puede nombrar en un parrafito todo, porque es lo que vos das. Ayudás, escuchás, prestás, decís por favor y gracias todo el tiempo. Sonreís todo el tiempo. Sos un sol.
- A Ana Julia y Andrea Puebla, por ser las profes del lab. Por la experiencia y meticulosidad en el trabajo que saben compartir, por enseñarme cosas, por compartir tantas otras. Por su sensibilidad.
- A Maxi, que estás muy muy lejos y me ayudaste un montón en tu paso por el lab. Te quiero mucho y te agradezco todo lo que me ayudaste, compartiste y lo que seguís compartiendo en la distancia.
- A Betiana por tantas ayudas luego de mi horario de corrida al jardín.
- A Ariel y Flor DV por sacar geles, sacar PCRs, sacar fotos y poner buena cara, por servicio de “*remis incluido*”. Los aprecio un montón.
- A Marisa LB , por sus diálogos francos y sinceros, por ser la madre sustituta del jardín, por el apoyo moral y técnico.
- A Flor Longo, por ofrecerme su compu y tantas otras tantas cosas.
- A Mariana, Ceci VR, Sebas y Fer por todo lo que supieron enseñarme.
- A Julián y Carla por su excelente predisposición siempre. ¡Y por sus mates!
- A Valeria Peralta por su ayuda en el invernáculo.
- A Agustín y Juan por el cuidado con mis plantitas y las ganas en ayudar.

- A todos los integrantes del grupo plantas que ayudan en el trabajo en equipo creando un buen clima de laburo.
- A Isabel, Fabián, Vilma, Majo y Roxana, por estar siempre ayudando en todo y con tantas ganas.
- A Silvio C. por transmitir conocimiento, compartirlo y ayudar en cosas del día a día. A Oscar por sus ganas de colaborar en todo.
- A Rubén y Marisa Farber, espíritus imprescindibles de ánimo, ayuda y paciencia. Es buenísimo trabajar con gente así.
- A Guido y Federico, por su ayuda en la PCR cuantitativa.
- A Paz por su buena onda y siempre con sus ganas de dar una mano.
- A Flor Lucca, por lo que pudimos compartir y la ayuda que me brindaste todo ese tiempo.
- A toda la gente del Instituto de Biotecnología que sonrío al pasar, escucha al hablar y piensa con ganas y entusiasmo siempre.
- A mis amigos Mariela, Liliana, Natalia, Vanesa, Jorge y tantos más. Por ser una extensión de mi pensamiento y entusiasmo. Gracias por bancarme todo este tiempo y no escucharme hablar de otra cosa.
- A mis papás y a mis suegros, por estar ahí siempre, por la ayuda con Salvador. Sin ustedes esto hubiera sido demasiado empinado para los dos.
- A mi tía Marta, mi segunda mamá, que a la distancia sabe todo, pregunta todo y quiere saber de mí. Su curiosidad es un mimo muy grande. Te adoro.
- A mis hermanos Silvia y Roberto por tantos diálogos y momentos compartidos en familia. Y por quererme como soy.
- A mi cuñada Silvina, siempre preocupándote por mí, por como me va, por lo que necesito y por lo que vos tenés y me podés brindar. Por querer saber de todo esto, al menos para hablar un rato juntas. Te quiero mucho.

- A Ale, por el apoyo, el amor, la tolerancia, el respeto y todo lo que tenés para darme desde que te conozco. Por el *know-how* “informático” *unpublished*. Gracias por ayudarme a construir, siempre a mi lado.
- A Salvador, que aprendió a usar la compu a fuerza de verme horas sentada allí, por entender que es para trabajar y por la constancia en su espera. Por su paciencia que me sirvió de motor estos últimos 3 años. Por enseñarme que no todo sale bien, ni cuando uno quiere. Por ser la señal más tangible de que uno aprende todo el tiempo. Y por haber decidido venir con nosotros, a formar parte de nuestro mundo, a desestructurarlo. Te amo.

*“Hay hombres que de su ciencia
tienen la cabeza llena;
hay sabios de todas menas,
mas digo, sin ser muy ducho:
es mejor que aprender mucho
el aprender cosas buenas”*

*Consejos de un padre a su hijo
De “El gaucho Martín Fierro”
José Hernández, 1872*

INDICE

RESUMEN.....	6
SUMMARY	11
I. INTRODUCCIÓN	15
I.1.Descripción del taxón	15
I.2. El girasol.....	15
I.3. El girasol en el mundo.....	16
I.3.1. El girasol en Argentina	18
I.4. El estrés abiótico y las plantas.....	18
I.4.2. Las plantas y la salinidad	19
I.4.3. Las plantas y las bajas temperaturas.....	21
I.4.4. Exigencias climático-ambientales del girasol	23
I.4.5. Efecto de la temperatura en el desarrollo del girasol	24
I.5. Proyectos genómicos.....	24
I.6. Genómica del girasol	26
I.7. Secuencias EST	28
I.8. La Bioinformática.....	30
I.9. La necesidad de la bioinformática en el área de la genómica	31
I.9.1. Predicción de genes	32
I.9.2. Caracterización de secuencias.....	32
I.10. Genómica funcional.....	33
I.10.1. Genómica funcional para la identificación y caracterización de procesos biológicos complejos	33
I.11. Micromatrices en el análisis de expresión génica a gran escala	35
I.11.1. Micromatrices de ADN sintetizados a partir de oligonucleótidos sintéticos.....	36

I.11.2. Micromatrices de ADN sintetizados a partir de ADNc.....	36
I.12. Metodología del uso de las micromatrices.....	37
I.13. Esquema general para el análisis de datos de micromatrices.....	39
I.13.1. Obtención de la imagen digital	40
I.13.2. Análisis de la imagen	41
I.13.3.Grillado o direccionamiento	41
I.13.4. Segmentación.....	43
I.14. Aplicaciones de las micromatrices de ADNc	43
I.14.1. Ventajas y desventajas	44
 II. OBJETIVOS E HIPÓTESIS	 46
II.1. Objetivo general.....	46
II.2. Objetivos específicos	46
II.3. HIPÓTESIS.....	47
 III. MATERIALES Y MÉTODOS.....	 48
III.1 Material biológico.....	48
III.2. Ensayos de estrés térmico y osmótico	48
III.3. Extracción de ARN total de girasol	49
III.4. Preparación de ARNm de girasol a partir de ARN total	49
III.5. Construcción de colecciones de ADNc.....	50
III.5.1. Cuantificación de ácidos nucleicos	51
III.5.2. Ligación de moléculas de ADN	51
III.5.3. Transformación de E. coli	52
III.5.4. Preparación del templado	52
III.5.5. Electroforesis de ADN en geles de agarosa	53
III.6. Reacciones de determinación de la secuencia de ADN.....	53
III.7. Análisis de secuencias.....	54

III.8. Northern blot	54
III.8.1. Amplificación de fragmentos de ADN mediante la reacción en cadena de la polimerasa (PCR) para uso de sondas específicas	55
III.8.2. Purificación de fragmentos de ADN de geles de agarosa.....	55
III.8.3. Pre-hibridación de membranas	55
III.8.4. Marcación de sondas	56
III.8.5. Hibridación, lavados y exposición de membranas	56
III.9. RT-PCR (reacción de PCR a partir de transcripción reversa)	57
III.9.1. PCR cuantitativa (en tiempo real)	58
III.10. Protocolos clásicos de biología molecular.....	59
III.11. Generación de micromatrices de ADN	60
III.11.1. Obtención de ADNc amplificados para impresión.....	60
III.11.2. Impresión de las micromatrices.....	61
III.11.3. Hibridación y escaneo de las micromatrices	62
III.11.4. Preprocesamiento de datos	62
III.11.5. Normalización entre réplicas biológicas	63
III.11.6. Consistencia de las repeticiones	64
III.12. Análisis estadístico. Identificación de grupos de genes con expresión diferencial.....	64
IV. RESULTADOS	65
 CAPÍTULO 1: OBTENCIÓN DE COLECCIÓN DE SECUENCIAS DE TIPO EST	65
IV.1.1. Colecciones de ADNc diferenciales	65
a) Clonoteca de ADNc órgano-específico de flor (estadio R1).....	66
b) Clonoteca órgano-específica de flor (estadio R4).....	67
c) Clonoteca de ADNc órgano-específico de hoja	68
d) Clonoteca de ADNc órgano-específico de tallo.....	69
e) Clonoteca de ADNc órgano-específico de raíz	70
IV.1.i) Análisis bioinformático de secuencias.....	71
IV.1.ii) Detección de unigenes.....	72
IV.1.iii) Búsqueda de similitud en bases de datos públicas.....	73
IV.1.iv) Análisis informático de ESTs	73

IV.1.2. Análisis de ESTs basados en función predicha	75
--	----

IV.1.3. Discusión.....	94
------------------------	----

CAPÍTULO 2: DESARROLLO DE UNA PLATAFORMA BIOINFORMÁTICA Y BASES DE DATOS PARA EL ANÁLISIS, ORDENAMIENTO Y ANOTACIÓN DE SECUENCIAS MOLECULARES 98

IV.2.1. Edición, ensamblaje de secuencias y detección de las redundancias presentes en las colecciones analizadas.....	98
--	----

IV.2.2. Anotación de secuencias.....	99
--------------------------------------	----

IV.2.3. Diseño, creación y mantenimiento de una base de datos.....	100
--	-----

IV.2.4. Diseño, creación y mantenimiento de una plataforma de análisis bioinformático (BioPipeline®)	105
--	-----

IV.2.5. Instalación y puesta en funcionamiento del servidor INTA 01.....	110
--	-----

IV.2.6. Discusión.....	112
------------------------	-----

CAPÍTULO 3: ESTUDIOS DE EXPRESIÓN DE GENES CANDIDATO FRENTE A ESTRESSES ABIÓTICOS.....115

IV.3.1. Tecnología del chip de ADN	115
--	-----

IV.3.2. Micromatrices de ADN	115
------------------------------------	-----

a) Confiabilidad de las repeticiones	115
--	-----

b) Análisis exploratorio de los perfiles de expresión génica	117
--	-----

c) Identificación de genes candidato - Análisis de componentes principales (ACP)	123
--	-----

IV.3.3. Perfiles de expresión por gen candidato	132
---	-----

IV.3.4. Validación por técnicas de PCR cuantitativa (qRT-PCR).....	136
--	-----

IV.3.4.i) Validaciones de clones por PCR cuantitativa (PCR en tiempo real)	136
--	-----

IV.3.5. Envío de datos de micromatrices a una base de datos de reposición pública.....	137
--	-----

IV.3.6. Validación por técnica de Northern Blot.....	138
--	-----

IV.3.7. Discusión.....	139
IV.3.7.1. Diferencias entre frío y salinidad	141
IV.3.7.2. Análisis de expresión por funcionalidad hipotética asociada.....	142
V. CONCLUSIONES.....	151
BIBLIOGRAFÍA.....	154

RESUMEN

El girasol es uno de los principales cultivos oleaginosos del mundo por su volumen de producción y composición en ácidos grasos insaturados. En promedio, durante los últimos cinco años, la producción mundial se ha mantenido estable concentrándose el 72% de la misma entre Argentina y la Unión Europea. A pesar de esto y debido a las condiciones de los precios internacionales y la productividad comparativa con otros cultivos, el área de producción se está desplazando a regiones más áridas a nivel mundial cobrando relevancia la incidencia de estreses abióticos como sequía, salinidad y bajas temperaturas. Pese a la importancia económica del girasol a escala mundial, el grado de avance de los conocimientos públicos sobre el genoma de girasol es limitado cuando se lo compara con cultivos como el arroz, el maíz, la soja, el tomate, el trigo, la papa, la cebada. Sin embargo, en los últimos cinco años, en paralelo al avance de este trabajo, se han iniciado distintos proyectos a nivel nacional e internacional orientados al análisis genómico de girasol con lo cual ha aumentado significativamente la disponibilidad de secuencias genómicas y de ADNc, con el consecuente impacto en el avance de las investigaciones y el conocimiento sobre esta especie.

El objetivo de este trabajo es contribuir al conocimiento de las regiones codificantes del genoma de girasol a partir de una estrategia de genómica funcional que sirva como base para la caracterización transcripcional simultánea de los perfiles de expresión inducidos bajo condiciones de estrés y el desarrollo de marcadores funcionales. Para lo cual se abordó el desarrollo de un banco local de secuencias que se expresan o ESTs ("expressed sequence tags") diferenciales para distintos órganos y/o estadios de desarrollo de girasol y el desarrollo de una plataforma bioinformática para la caracterización de los perfiles transcripcionales de las mismas en respuesta a estrés abióticos.

En una primera etapa se evaluaron distintas estrategias de secuenciación a pequeña y mediana escala para la identificación de genes diferencialmente

expresados en girasol a partir de la generación colecciones de ADNc diferencial. La técnica utilizada se basó en la hibridación substractiva como herramienta para la identificación de genes diferencialmente expresados en un tejido definido, disminuyendo significativamente la redundancia en la secuenciación de clones que representan a genes abundantes y maximizando la detección de los transcritos “raros” o poco abundantes. Esta estrategia permitió un incremento de la eficiencia de secuenciación dentro del marco de un proyecto genómico de pequeña escala que apunta a la identificación de genes de utilidad para propósitos de mejoramiento y la búsqueda de marcadores funcionales. Como resultado, se obtuvieron clonotecas diferenciales a partir de hoja, tallo, raíz y flor en dos estadios de desarrollo (R1 y R4). El uso de diferentes fuentes de ARN como objeto o “tester” y referencia o “driver” fue de utilidad para la selección y detección de transcritos poco abundantes. La especificidad órgano-específica varió dentro de un rango de 75 a 100% de las secuencias no-redundantes (unigenes) dentro de cada clonoteca de ADNc. La clonoteca correspondiente al estadio R4 de flor fue la menos redundante con un 62% de secuencias únicas y la que aportó el número más elevado de secuencias nuevas de girasol. El análisis bioinformático se llevó a cabo mediante la instalación de un servidor local (INTA 01) conteniendo las herramientas de distribución pública para la edición, análisis y ensamblado de secuencias como Phred/Phrap, CAP3 y algoritmos de comparación de secuencias como BLAST y el desarrollo de BioPipeline® una plataforma desarrollada en entorno Windows para el análisis automatizado de secuencias utilizando los mismos programas y algoritmos mencionados anteriormente. La información de secuencias generadas fue depositada en la base dbESTs de NCBI para su distribución pública. Para el manejo local de esta información se desarrolló una base relacional de tipo ACeDB para la integración de la información generada y su anotación funcional. Sobre un total de 919 secuencias que fueron editadas y anotadas, 318 representan secuencias únicas. Las comparaciones contra bases de datos públicos arrojaron un valor del 60% de secuencias no-redundantes con similitud

a secuencias conocidas. El número de genes novedosos predichos varió según la clonoteca analizadas, en un rango que va de 56% en las clonotecas de flor estadio R4 al 16% en las de flor R1. Las comparaciones con los ESTs de girasol depositados y reportados en bases de datos mostraron que 197 secuencias (59,9%) del total) representaban secuencias nuevas, sin similitud significativa con secuencias conocidas. Este enfoque fue exitoso respecto del aislamiento de un número significativo de secuencias reportadas asociadas en su función inferida (probable) a caracteres de importancia agronómica, procesos regulatorios y fisiológicos clave.

En una segunda etapa se abordó la evaluación de la expresión relativa simultánea de los transcritos correspondientes a las secuencias de la base de ESTs desarrollada bajo diferentes condiciones de estrés abiótico utilizando la tecnología de micromatrices de ADN. Se imprimieron los fragmentos representativos de los 318 unigenes anotados en las clonotecas previamente descritas en micromatrices sobre soporte de vidrio. Se evaluaron tres muestras biológicas tanto para tratamiento de frío (estrés térmico), tratamiento de salinidad (estrés salino) y controles representados por plantas crecidas en condiciones regulares. Estos estreses fueron seleccionados considerando al girasol como una planta con grado intermedio-alto de sensibilidad a bajas temperaturas en la zona radicular, así como también particularmente susceptible al desarrollo en condiciones de alta salinidad.

El análisis transcripcional permitió detectar 3 grupos de genes bien diferenciados en sus patrones de expresión. Por una parte el Grupo 2, integrado por 126 genes, no mostró variación en sus niveles medios a través de los tratamientos. En cambio, el Grupo 1 (integrado por 112 genes) evidenció sobre-expresión respecto del control cuando las plantas fueron sometidas a estrés (por frío o por salinidad). De manera análoga, 49 genes conformaron el Grupo 3, pero en este caso se observó una sub-expresión respecto del control en ambos tratamientos. Dentro del grupos 1 y 3, se detectaron perfiles de expresión diferenciales para ambos tipos de estrés, siendo la sobre expresión y/o sub

expresión de los genes evaluados más pronunciada en respuesta al estrés por frío respecto al estrés por salinidad. Finalmente, surgieron 80 genes candidato en la separación de tratamientos de acuerdo a su p-valor en la prueba de comparación de medias por análisis de la varianza y su ubicación en la región periférica del plano de ordenación generado por los dos primeros ejes principales de un Análisis de Componentes Principales (ACP) de la matriz de expresión génica escalada gen a gen por la desviación estándar residual del análisis de la varianza. De estos genes, 7 fueron validados por técnicas de reacción en cadena de la polimerasa cuantitativa (*Real Time PCR*), habiéndose obtenido patrones transcripcionales similares con ambas estrategias de análisis

La diferencia en expresión génica fue analizada en relación al origen de las secuencias en evaluación. Las secuencias provenientes de la colección de hoja mostraron en general patrones de subexpresión génica en respuesta a ambos estreses (la mayoría corresponden a genes asociados al proceso de fotosíntesis) mientras que las secuencias aisladas de colecciones de tallo o flor en estadio R4 mostraron patrones de inducción génica frente a los mismos estreses. Si se considera que los cambios transcripcionales en respuesta a frío y salinidad fueron evaluados en hoja, estos resultados confirman la eficiencia de la técnica de SSH utilizada para la generación de las colecciones de ADNc órgano específicas y explica la alta relación de transcritos que muestra cambios en su perfil en respuesta a estreses en relación a los que no varían su nivel transcripcional en las condiciones evaluadas. Durante la segunda etapa de este trabajo se realizó un análisis de los 80 genes candidatos, detectándose 48 como sobre o sub expresados frente a uno u otro estrés, indicando la implicancia de mecanismos regulatorios comunes a ambos estreses. Asimismo 17 y 12 genes se detectaron inducidos o sub expresados específicamente frente a un estrés y no frente al otro. De acuerdo al análisis funcional comparativo estos genes estarían involucrados en mecanismos de regulación incluyendo procesos de transcripción, traducción, degradación/plegado o interacción de proteínas o

asociados a mecanismos de generación y procesamiento de especies reactivas de oxígeno. De estas secuencias, 12 corresponden a secuencias con funcionalidad desconocida. Sólo 3 de las secuencias analizadas en este trabajo mostraron patrones transcripcionales opuestos incluyendo una con similitud a un transportador de lípidos.

La información generada a partir de la caracterización del banco de ESTs constituye por una parte una fuente de información sobre genes candidatos involucrados en mecanismos de respuesta a estreses abióticos que pueden ser incluidos en ensayos de complementación en plantas transgénicas modelo y asimismo permite identificar secuencias nuevas no descritas anteriormente para el desarrollo de marcadores funcionales a ser utilizados en programas de mejoramiento asistido de este cultivo.

Palabras clave: *Helianthus annuus*; girasol; EST; SSH; ADNc diferencial; genómica funcional; bioinformática, transcriptos órgano- específicos; micromatrices de ADNc; estrés por frío, estrés salino, PCR cuantitativa

.

SUMMARY

Sunflower is one of the most important oil crops worldwide regarding its production volume and composition in unsaturated fatty acids. However due to international price conditions and productivity respect to other crops, sunflower production is progressively expanding to arid regions worldwide. Consequently, abiotic stresses as drought, low temperatures and salinity arise as main constrains nowadays. Sunflower is considered a medium-high tolerant plant crop to low temperatures in the radial zone and sensitive to saline environment. Compared to other crops such as rice, maize, soybean, tomato, wheat, potato and barley genetic and genomic sunflower knowledge was limited until the onset of the present decade. However, in parallel to the advance of this thesis (last five years), different projects aimed to the genome analysis of cultivated sunflower and its related wild species, resulted in a significant increase of the number of available expressed and genome sequences in public databases.

The objective of this work was to develop tools for structural and functional analysis of cultivated sunflower through the development of a model organ-specific ESTs ("expressed sequence tags") database, the creation and actualization of a bioinformatics platform to characterize transcriptional profiles of the represented genes in response to abiotic stress conditions with the aim to identify novel genes as a source of functional molecular markers associated to important traits, under the scope of an alternative, cost-effective strategy to high-throughput sequencing projects.

In the first stage of this work a low scale sequencing strategy based on suppressed subtracted hybridization method was evaluated for the construction of organ-specific cDNA libraries in order to identify differentially expressed genes in sunflower. Differential organ-specific ESTs were generated from leaf, stem, root and flower bud at two developmental stages (R1 and R4). Organ-specificity ranged from 75 to 100% of non-redundant sequences in the different cDNA libraries. Sequence redundancy varied according to the target and driver cDNA used for each library.

The R4 flower cDNA library was the less redundant library with 62% of unique sequences.

Bioinformatics analysis was achieved through the development of BioPipeline®, a bioinformatic tool for sequence quality analysis, contaminant sequence removal and sequence assembly and automatic annotation based on the utilization of free available software as Phred/Phrap, CAP3 and BLAST algorithms using a local server installed at the Institute of Biotechnology, INTA01. Processed sequences were deposited for public distribution in the dbEST database at NCBI. Biological information related to these sequences was locally stored using an ACeDB-based relational database as a laboratory information management system. Out of a total of 919 sequences that were edited and annotated, 318 were non redundant sequences. Comparison against sequences in public databases showed that 60% of non redundant sequences showed significant similarity to known sequences. The number of predicted novel genes varied among the different cDNA libraries, ranging from 56% in the R4 flower to 16 % in the R1 flower bud library. Comparison with sunflower ESTs on public databases showed that 197 of non-redundant sequences (60%) did not exhibit significant similarity to previously reported sunflower ESTs. This approach helped to successfully isolate 33 newly reported unique sequences for sunflower putatively related to responses to biotic and/or abiotic stresses. Application of suppressed subtracted hybridization technology not only enabled the cost effective isolation of differentially expressed sequences but it also allowed the identification of novel sequences in sunflower from a relative small number of analyzed sequences when compared to major sequencing projects.

In a second stage, functional genomic studies were performed using microarray technology. A cDNA glass slide chip containing the 318 organ-specific unigenes was constructed in order to evaluate the expression profile under cold and salinity stress conditions. Three biological leaf samples from treated plants, exposed either to low temperatures or watered with NaCl solution, as well as control plants grew under standard condition, were evaluated by hybridization of fluorescence labeled treated and control RNA to target genes. Transcriptional analysis allowed the

detection of three groups of genes well differentiated in their expression profiles during the exploratory analysis. One group composed by 126 genes did not show variation along treatments; a second group composed by 112 genes showed evidence of up-regulation under abiotic stress (cold or salinity), whereas a SUB-regulation in both treatments was observed in the 49 genes belonging to a third group. During the second phase of this analysis, 80 candidate genes were identified according to the *p-value* obtained by classical ANOVA and their peripheral location in the PCA analysis of the gene matrix. Seven genes were experimentally validated by quantitative real time PCR (qRT-PCR) being the transcriptional patterns derived from the two methods highly similar in all cases. Differences in genes expression were analyzed according to the differential organ-specific cDNA library from which they were isolated. While differential leaf library sequences showed the majority of genes sub-regulated (most of them corresponding to photosynthesis or general metabolism related genes), R4 flower developmental stage and stem libraries showed a larger number of up-regulated genes. Considering that transcriptional changes in response to salt and cold stresses were evaluated in leaf tissue, these results confirmed the efficiency of SSH technique in the generation of the differential organ specific libraries and explained the large *transcription change / transcription unchanged* ratio detected in these assays. Within the 80 candidate a large number of evaluated sequences (48) were either up-regulated or sub-regulated under both abiotic stresses, thus supporting common regulatory mechanisms and general responses to low temperature and salinity. Interestingly 17 sequences were specifically up regulated and 12 were specifically sub-regulated in one stress but not in the other. These genes are potentially involved in different regulatory mechanisms including transcription / translation / protein degradation / protein folding or correspond to genes involved in ROS production or ROS-scavenging and some of them (12) did not show similarity to reported gene products. Only three of the evaluated sequences in this work showed an inverted transcriptional pattern under cold and salinity stress, including one with similarity to a LTP transcript. The information obtained from this EST library characterization not only contributed as a major source of knowledge about candidate genes involved in the

mechanisms of abiotic stress responses in sunflower that can be incorporated in future complementary gene studies in model or sunflower transgenic plants but also allows the detection of new putative functional markers as a key tool for marker assisted selection in sunflower.

Key words: *Helianthus annuus*; sunflower, EST; SSH; differential cDNA; functional genomics; bioinformatics; organ-specific transcriptoma; cDNA microarrays; cold stress, salinity stress, quantitative PCR

I. INTRODUCCIÓN

I.1. Descripción del taxón

La familia Compositae perteneciente a la Clase Magnoliopsida (Dicotiledóneas) es una de las familias más grandes y diversas de todas las plantas con flor, conteniendo un décimo de todas las especies de angiospermas. Se caracteriza por tener una flor compuesta (inflorescencia), de la cual deriva su nombre. Se estima que el número básico cromosómico es nueve. Todas las especies Compositae analizadas hasta la fecha tienen genomas más grandes que 1,1 picogramos de ADN /1C; (Kesseli y Michelmore, 1997). La familia probablemente se originó en América del Norte hace 30 a 100 millones de años atrás (Raven y Axelrod, 1974) y sufrió una rápida y extensiva diversificación produciendo un gran espectro de taxones: contiene más de cuarenta especies económicamente importantes para la vida del hombre, entre ellas: lechuga (*Lactuca sativa* L.; alimentación), girasol (*Helianthus annuus* L.; alimentación y ornamental), topinambur (*Helianthus tuberosus* L., alimentación, forrajero, prebiótico), cártamo (*Carthamus tinctorius* L., alimentación, uso industrial), camomila (medicinal), crisantemo, dalia, margarita y *Zinnia* (ornamentales), entre otras.

I.2. El girasol

El girasol cultivado es una de las pocas plantas que han sido domesticadas en la zona templada de América del Norte y cuyo antecesor silvestre es conocido con precisión. El girasol, *Helianthus annuus* L., comprende tres subespecies principales: *Helianthus annuus ssp lenticularis*, girasol silvestre; *Helianthus annuus ssp annuus*, girasol maleza y *Helianthus annuus ssp macrocarpus*, girasol cultivado por sus frutos comestibles. El primero, que se encuentra difundido en el oeste de América del Norte, es ramificado, con capítulos y aquenios pequeños y pocas flores liguladas. El segundo, se lo encuentra en el Medio Oeste de los Estados Unidos, es también ramificado, pero con capítulos y aquenios de mayor tamaño y número de

flores liguladas mas numerosas que el silvestre. El girasol cultivado es no ramificado, con un solo capítulo con numerosas flores liguladas y aquenios grandes. Está distribuido en casi todo el mundo pero no persiste fuera del cultivo (<http://www.asagir.org.ar>). El girasol cultivado (*Helianthus annuus* L.) es una especie diploide ($2n = 2x = 34$) con un genoma haploide medio de 3000 megabases aproximadamente (Arumuganathan 1991). Pertenece a la subtribu Helianthinae, subfamilia Asteroideae y familia Compositae (Seiler y Rieseberg, 1997). El centro de diversidad de dicha especie es América del Norte (Heiser y Smith, 1955; Heiser, 1976). El género *Helianthus* esta compuesto por 36 especies perennes y 12 especies anuales. Algunas especies salvajes de *Helianthus* han sido importantes fuentes para la identificación de genes de resistencia a enfermedades, tolerancia a la sequía, esterilidad citoplasmática masculina (CMS) y otras características agronómicas importantes (Jan, 1998). El desarrollo del girasol cultivado por cruzamientos interespecíficos (*H. annuus* x *H. petiolaris*) con incorporación del CMS de *H. petiolaris* y los genes restauradores de la fertilidad de *H. annuus* abrieron el camino de la industria de la semilla híbrida de girasol (Leclercq, 1969; Kinman, 1970; Fick y Zimmer, 1976).

El girasol es naturalmente una especie alógama y auto-incompatible, su domesticación ha permitido la auto-compatibilidad y un aumento en el tamaño y en el vigor de la semilla. Esto hace que el girasol cultivado tenga varias características que facilitan su análisis genético, a su vez el desarrollo de germoplasma autopolinizable ha favorecido los programas de mejoramiento como el desarrollo de híbridos. El girasol cultivado puede ser cruzado sin dificultad, tiene una alta producción de semillas (aproximadamente 1000 semillas/planta) y un tiempo generacional corto de 70 a 100 días.

I.3. El girasol en el mundo

El girasol es uno de los principales cultivos oleaginosos del mundo. En promedio, durante los últimos cinco años, la producción mundial se ha mantenido estable concentrándose el 72% de la misma entre Argentina y la Unión Europea. Los

subsidios han sido el motor de la producción girasolera en Europa, mientras que en la Argentina, la expansión de este cultivo se debió fundamentalmente al aumento de productividad y eficiencia en los costos. Según datos de la Cámara de la Industria Aceitera de la República Argentina (Franco, 2005) la producción para la campaña 2005/06 se estima en 25 millones de toneladas (Tabla 1).

Tabla 1. Producción de semilla de girasol (en miles de toneladas)						
Semilla de girasol						
Provisión mundial (en miles de toneladas)						
Item	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06
					Revisado	Estimado
Área cosechada (1.000 HA)	19.540	18.485	19.892	22.918	21.262	22.791
Rendimiento (MT/HEC)	1.18	1.18	1.2	1.17	1.23	1.29
Producción de semilla						
Argentina	2.950	3.720	3.340	2.990	3.650	3.800
Europa del Este	1.657	1.861	1.648	2.295	2.270	1.950
Unión Europea	3.333	3.030	3.718	4.078	4.133	3.765
China	1.954	1.750	1.946	1.820	1.750	1.850
Rusia/Ucrania	7.368	4.936	7.194	9.348	8.001	10.450
EEUU	1.608	1.551	1.112	1.209	930	1.824
India	730	870	1.060	1.160	1.300	1.250
Turquía	630	530	830	560	640	790
Otros	2.880	3.551	3.108	3.467	3.505	3.665
TOTAL	23.110	21.799	23.956	26.927	26.179	29.344
Importación de semilla						
México	23	10	104	38	11	23
Unión Europea	1.999	1.155	1.007	1.473	763	1.000
Otros	704	467	812	1.249	813	801
TOTAL	2.726	1.632	1.923	2.760	1.587	1.824
Semilla partida	21.116	18.514	21.149	23.442	23.115	25.510
Exportación de semilla						
Argentina	94	342	232	44	97	121
EEUU	153	176	122	136	116	225
Rusia/Ucrania	1.768	100	524	1.271	73	560
Otros	711	1.084	1.112	1.277	1.257	957
TOTAL	2.726	1.702	1.990	2.728	1.543	1.863

(Fuente :<http://www.sunflowernsa.com/stats/table.asp?contentID=109&htmlID=74&submit170=View&submit.x=57&submit.y=12>)

1.3.1. El girasol en Argentina

El cultivo de girasol es el segundo en importancia dentro del mercado argentino de producción de semillas oleaginosas. Según estimaciones oficiales (ver Plan Plurianual de CyT) durante el periodo 1998/99 Argentina fue el segundo productor mundial (detrás de la ex URSS) con 6,7 millones de toneladas, lo cual produjo ingresos por 1000 millones de dólares en exportaciones de aceites y granos. Si bien estos niveles de producción no se mantuvieron en los últimos años, fundamentalmente debido a una fuerte baja del precio del aceite de girasol ante un aumento considerable de la disponibilidad del competidor aceite de palma en el mercado mundial, el mismo sigue siendo un producto estratégico para nuestro país. Argentina actúa como formadora de precios para este cultivo. Dada la relevancia del girasol y el desarrollo agroindustrial derivado del mismo para la economía del país, es fundamental que Argentina promueva este cultivo definiendo una política agrícola robusta e independiente y desarrollando un importante paquete tecnológico que nos permita ser competitivos a nivel mundial.

I.4. El estrés abiótico y las plantas

El estrés abiótico es la influencia negativa sobre el desarrollo vegetal generada por el cambio medioambiental. Durante su desarrollo las plantas se encuentran expuestas frecuentemente a factores causantes de estrés, tanto naturalmente como en condiciones agrícolas. Algunas de estas condiciones medioambientales desfavorables varían en pocos minutos u horas (viento, calor), mientras que a otras les toma días o semanas (sequía) e incluso meses (falta de nutrientes) (Taiz y Zeiger, 1998), (Dezar, 2005a). Un hecho importante es que las plantas son organismos sésiles que como tales deben enfrentar cambios ambientales como salinidad, sequía y frío. Si bien la salinidad, el agua y el frío son estreses claramente diferentes entre ellos en su naturaleza física, cada uno dispara diferentes tipos de respuestas en las plantas, activando también reacciones típicas en estos organismos. Asimismo en los últimos años se ha estudiado más profundamente y

en forma concertada la compleja red de transducción de señales en respuesta a estreses abióticos especialmente en plantas modelo como *Arabidopsis* (Zhu JK., 2001). No obstante las respuestas a frío no están relacionadas con las respuestas a salinidad, sequía o regulación por ABA, se pudo detectar una cascada transcripcional que incluyen genes activados por TFs sometidos a bajas temperaturas: CBF y DREB1 (Zhu 2001, Maruyama y col. 2004). Esta respuesta fue también reportada en arroz por Dubouzet y col. 2003).

1.4.2. Las plantas y la salinidad

Si bien el sodio es un micronutriente de fundamental importancia para el desarrollo vegetal, niveles de sodio en solución en exceso son adversos para la producción de granos comerciales, afectando su calidad y cantidad (Gepstein y col. 2006; Blumwald y Grover 2006). Los suelos salinos en zonas áridas y/o semi-áridas de Asia, África y Sudamérica son causa de un vasto número de limitantes agronómicas. En Asia, por ejemplo, 21,5 millones de hectáreas son consideradas no aptas para cultivo por salinidad presunta (de las cuales 12 millones son limitados por salinidad y las restantes 9,5 millones limitantes en su relación alcalino/sódica). Si incluimos a la India en este análisis, 8,6 millones de hectáreas adicionales se suman a esta problemática, de las cuales 3,4 millones son suelos sódicos. Una observación relevante a esta problemática es que el síntoma ambiental de suelos salinos se encuentra acentuado con el aumento de irrigación en suelos agrícolas (Sahi y col. 2006).

Un importante objetivo en el desarrollo agrícola moderno ha sido mejorar la tolerancia a sequía y salinidad en los cultivos de grano. Dado que la mayor parte de los suelos aptos para agricultura se encuentran en procesos de salinización o de irrigación pobre, el impacto regional de la salinidad en la producción de granos se ha tornado preocupante en los últimos años en diversas partes del mundo (Tanji, 1990; Flowers y Yeo, 1995) surgiendo la necesidad de obtener plantas tolerantes a la salinidad. Es por ello que la necesidad de balancear productividad con tolerancia a la salinidad y/o sequía se ha tornado en los últimos años como un punto de

discusión que debe ser resuelto a favor del rendimiento relativo sólo en áreas cultivables. Sin embargo, los rápidos avances que se están llevando a cabo en el conocimiento de mecanismos bioquímicos que podrían participar en las respuestas de las plantas a los estreses abióticos, así como también la elucidación de rutas metabólicas que responden al estrés por salinidad, ofrecen nuevas aproximaciones en la resolución de este problema persistente (Bohnert y Jensen, 1996, Winicov y Bastola, 1997).

Una amplia gama de evidencias sugiere que la respuesta celular de las plantas a condiciones de alta salinidad se encuentra controlada por múltiples genes (Bartels y Sunkar, 2005; Chinnusamy y col. 2005), siendo este proceso un mecanismo complejo que involucra la activación de una cascada de expresión génica actuando de manera concomitante con la represión de genes constitutivos o de mantenimiento (Sahi y col. 2006). En los últimos cinco años el advenimiento de técnicas de macro y micromatrices de ADNc permitió estudiar en forma concertada el perfil de expresión diferencial en plantas sometidas a estrés salino (Seki y col. 2001; Kanesaki y col. 2002; Sahi y col. 2003; Kore-Eda y col. 2004; Marin y col. 2004; Sottosanto y col. 2004; Taji y col. 2004; Shiozaki y col. 2005). A partir de los datos obtenidos de *Arabidopsis thaliana* y *Oryza sativa* surgen las hipótesis de que las plantas toleran la salinidad debido a moléculas efectoras que a través de un mecanismo molecular complejo modulan la etiología del estrés o atenúan sus efectos a través de moléculas reguladoras involucradas en la percepción del estrés (Hasegawa y col. 2000, Sahi y col., 2003).

a) *Bomba Na⁺/K⁺*

Entre las estrategias para mantener un equilibrio K⁺/Na⁺ adecuado en el citosol están la extracción y/o compartimentalización de sodio. Bajo las condiciones fisiológicas típicas, las plantas mantienen una proporción del K⁺/Na⁺ constante. El K⁺ se mantiene relativamente alto, de 100-200 mM y las concentraciones de Na⁺ bajas de 1-10 mM (Wilson y col 1995). Un aumento en la concentración intracelular del Na⁺ establecerá un gradiente de potencial electroquímico que favorecerá el transporte pasivo del sodio al citosol. Los antiportadores de Na⁺/H⁺ median la

compartimentación dentro de la vacuola y la extracción de Na^+ de la célula, estos dos procesos son claves en la acción concertada en la desintoxicación y la regulación osmótica del Na^+ citosólico (Blumwald y col. 2000).

b) Antiportadores Na^+/H^+

Las enzimas estudiadas de plantas halófitas muestran gran sensibilidad a salinidad, lo que sugiere que estas plantas son capaces de mantener constantes los niveles de Na^+ en el citosol. Esto puede lograrse compartimentalizando el Na^+ en vacuolas, o bien, secretando Na^+ . Se reportaron antiportadores Na^+/H^+ de membrana (Blumwald y col. 2000), además del bien caracterizado antiportador Na^+/H^+ de membrana SOS1 de *Arabidopsis thaliana* (Shi y col. 2000). La sobreexpresión de AtNX1, antiportador vacuolar Na^+/H^+ de *Arabidopsis thaliana* (Apse y col. 1999) y tomate (Zhang y col. 2001) confirió mayor tolerancia a salinidad. Asimismo, plantas transgénicas de *Brassica napus* para el mismo gen crecieron normalmente y produjeron semillas en presencia de 200 mM de Na Cl (Zhang y col. 2001), además de mostrar un incremento en los niveles de prolina.

1.4.3. Las plantas y las bajas temperaturas

Las plantas capaces de aclimatarse al frío perciben las bajas temperaturas sobre cero durante el avance del invierno y disparan procesos bioquímicos específicos que resultan en la tolerancia, aún de temperaturas de congelamiento (Thomashow, 1999).

La aclimatación al frío está asociada a la expresión *de novo* de genes, síntesis de proteínas y varios cambios fisiológicos (Guy, 1990). Los mecanismos que contribuyen a la tolerancia incluyen la prevención de la desnaturalización de proteínas inducidas por bajas temperaturas, la prevención de la precipitación de moléculas y la atenuación de las consecuencias de la formación de hielo intercelular. Entre los mecanismos involucrados en la estabilización de las membranas celulares en respuesta al frío, juegan un papel relevante los cambios en la composición de lípidos (Steponkus y col, 1993; Uemura y Steponkus, 1997) y la

acumulación de solutos compatibles (moléculas que pueden acumularse en la célula sin modificar el metabolismo central) en células de plantas sometidas a este estrés. Se han aislado y caracterizado numerosos genes inducibles por bajas temperaturas, en varias especies vegetales incluyendo alfalfa (Wolfrain y col., 1993), *Arabidopsis* (Gilmour col., 1992; Nordin, 1991; Yamaguchi-Shinozaki, 1993), cebada (Dunn y col., 1991) y trigo (Houde y col, 1992, Oullet y col., 1998). Una gran parte de estos codifican proteínas de función conocida que contribuyen a la tolerancia a las bajas temperaturas. Por ejemplo, el gen de *Arabidopsis* FAD8 codifica una desaturasa de ácidos grasos que alteraría la composición lipídica de las membranas (Gibson y col., 1994). También se conocen genes de chaperonas en espinaca y *Brassica napus* que responden a las bajas temperaturas y genes regulados por frío que intervienen en la transducción de señales. (Figura 1).

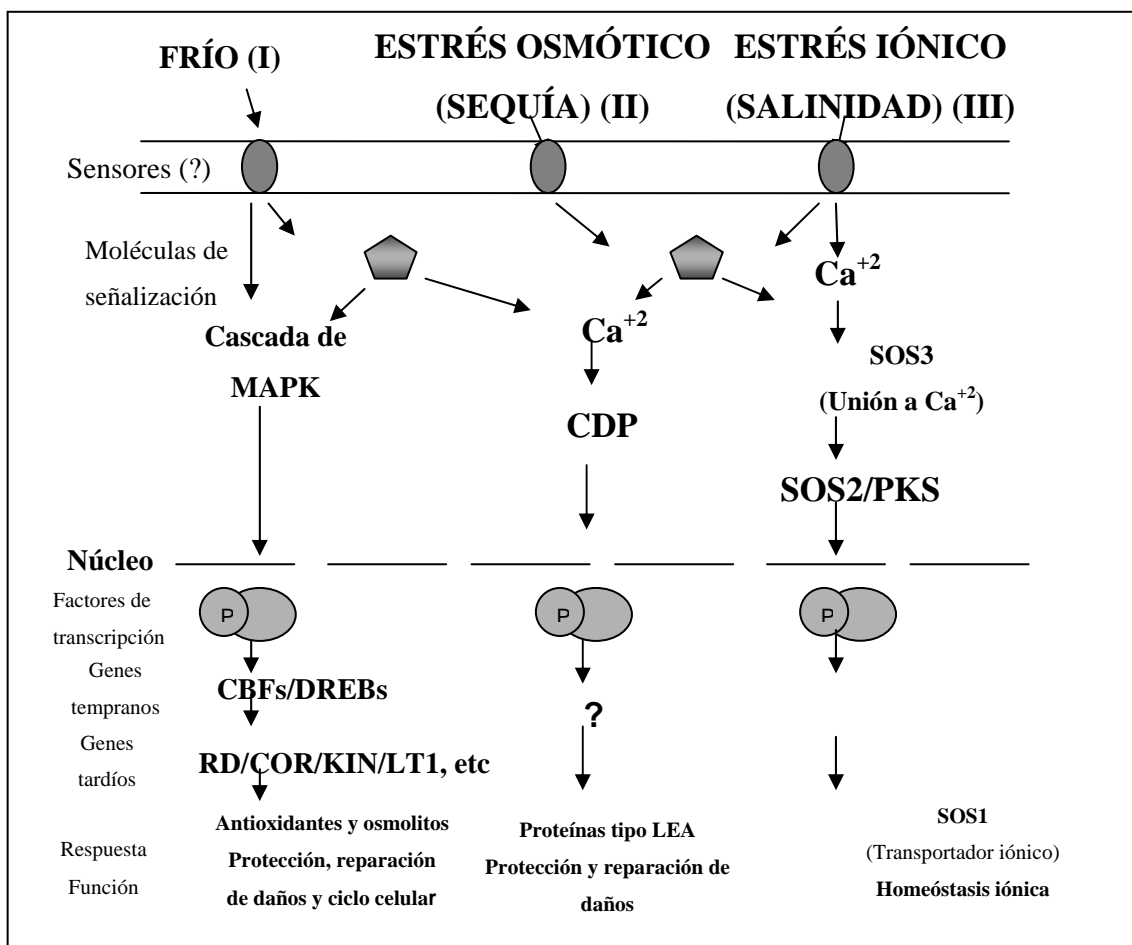


Figura 1 (Puebla AF. y Del Viso F. 2004): Vías principales de señalización en plantas en respuesta a estreses abióticos (frío, sequía y salinidad). La vía I de señalización involucra la producción de antioxidantes y osmolitos involucrando la transducción de señales a través de MAP kinasas. La vía II involucra la producción de diferentes tipos de proteínas (tipo LEA y otras) y está mediada por CDPKs. La vía III involucra a la vía SOS, específica del estrés iónico. No se detallan las moléculas de señalización o segundos mensajeros.

Sin embargo, muchos genes inducidos durante la aclimatación a bajas temperaturas codifican proteínas de función desconocida. Muchas de éstas presentan homología a las proteínas LEA (Late Embryogenesis Abundant). Los polipéptidos LEA son sintetizados en la semilla durante la embriogénesis tardía antes de la desecación y también en plántulas en respuesta a la deshidratación (Ingram y col., 1996). Estas proteínas son altamente hidrofílicas, tienen una composición simple de aminoácidos y en muchas se predicen regiones capaces de formar α -hélices anfipáticas. Otras proteínas, de función desconocida, que no se encuentran agrupadas con las LEA, son las codificadas por genes COR (cold resistant), también designados LT1, KIN, RD y ERD, de *Arabidopsis*. Estas proteínas también presentan regiones con capacidad para formar α -hélices anfipáticas que podrían tener roles en la estabilización de las membranas durante las bajas temperaturas. Algunos de estos genes han sido utilizados para transformar plantas susceptibles a frío bajo promotores constitutivos (35S CaMV) y se han realizado estudios de adquisición de tolerancia a estrés con las plantas transgénicas obtenidas (Holmberg y Bulow, 1998). Los resultados sugieren que la tolerancia a bajas temperaturas depende de varios genes de tolerancia y no de un gen particular.

1.4.4. Exigencias climático-ambientales del girasol

El girasol es un cultivo que se adapta bien a diferentes ambientes (Robinson, 1978), pudiendo tolerar temperaturas bajas y períodos de estrés hídrico. Con relación al fotoperíodo, el girasol se clasifica como especie insensible. Este comportamiento puede ser verificado por la amplia dispersión en el territorio nacional de híbridos y

variedades probadas en las Redes de Ensayo de Girasol. No obstante, algunos cultivares se comportan como plantas de día largo.

El déficit hídrico es el principal factor limitante para el desarrollo de los cultivos en suelos de aptitud agrícola (Boyer, 1982). Además, el girasol es cultivado frecuentemente en condiciones de secano, y a pesar de ser considerado como un cultivo tolerante a la sequía, su producción en granos puede ser afectadas drásticamente (Castro y Bouças Farias, 2005).

1.4.5. Efecto de la temperatura en el desarrollo del girasol

En suelos con aireación y disponibilidad hídrica adecuadas, la temperatura es el factor más limitante en la germinación de la semilla del girasol. Dicho proceso se inhibe con temperaturas de suelo menores a 3-4°C y aumenta exponencialmente de 3°C a 30°C, disminuyendo abruptamente por encima de los 37°C (Macchia y col. 1985; Gay y col. 1991). Las temperaturas bajas durante la germinación retrasan la emergencia, el establecimiento radicular e inducen la formación de plántulas pequeñas. De esta manera, la planta queda expuesta al ataque de patógenos y plagas presentes en el suelo. Varios autores citan como la temperatura basal para el girasol valores entre 4°C y 8°C (Robinson y col., 1967; Sadras y Hall, 1988; Merrien, 1992; Villalobos y Ritchie, 1992). No obstante, las plantas pueden tolerar temperaturas bajas por períodos cortos, principalmente en la fase inicial de desarrollo hasta 4 o 5 hojas (Robinson, 1978), mientras que por debajo de 4°C a 5°C se considera que el girasol no presenta actividad fisiológica.

I.5. Proyectos genómicos

Los importantes avances en genética de avanzada y genómica en los últimos 12 años resultaron en el desarrollo de un mapa genético de girasol saturado basado en el uso de diferentes marcadores moleculares que incluyen RFLP, AFLP y SSR (Berry y col., 1994 y 1995; Gentzbittel y col., 1994, 1995 y 1999; Jan y col., 1998; Gedil y col., 2001; Tang y col., 2002; Paniego y col 2006; Poormajarti y col 2006). El desarrollo de un gran número de microsatélites polimórficos (Paniego y col.,

2001, Tang y col., 2002) demostró su utilidad no solo por su aplicación en la selección asistida sino también en genotipificado (*fingerprinting*) de ADN y estudios de relación genómicos. Además, los proyectos de secuenciación de ADNc han identificado *Expressed Sequence Tags* (ESTs) en diferentes plantas. En la actualidad, se encuentran representadas más de 100 especies de plantas en la división EST de GenBank (www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html) con un total de 39.769.148 secuencias (depositadas al 17 de noviembre de 2006). Sin embargo, los proyectos de *Arabidopsis*, arroz, maíz, tomate y soja aún reúnen más del 50% del total de las entradas. La familia de las asteráceas (*Compositae*) está representada por 466.639 entradas de las cuales 94.110 secuencias corresponden a ESTs de girasol. Dichos proyectos permiten la caracterización de grupos completos de transcritos en organismos blanco y al mismo tiempo proveen una fuente de marcadores genéticos funcionales que pueden asociarse a tratamientos agronómicos importantes, complementando así el uso de marcadores neutros. De esta manera, el uso de marcadores de tipo EST podría facilitar el mapeo genético de aquel gen que afecte directamente al tratamiento o secuencia específica con función predicha por comparación de secuencia (Cato y col., 2001).

Los ESTs generados a partir de colecciones de ADNc podrían representar, idealmente, todos los genes de un órgano o tejido en un estadio de desarrollo específico y/o expuesto a una determinada condición ambiental. Sin embargo, las diferencias encontradas a nivel de expresión entre genes en un determinado tejido tienen rendimientos de ARNm que difieren en abundancia, haciendo dificultosa la captura de ARNm poco abundantes o únicos a partir de dichas clonotecas. Este problema también lleva a la secuenciación redundante de clones que representan los mismos genes expresados, afectando la eficiencia y efectividad de costo de la técnica basada en ESTs (Bonaldo y col. 1996) la cual dificulta la capacidad de laboratorios de bajo presupuesto para llevar a cabo estudios de caracterización a partir de ESTs.

Para evitar este problema se reportaron diferentes estrategias basadas en clonotecas de ADNc normalizadas (Patanjali y col., 1991; Soares y col., 1994, Bonaldo y col., 1996) incluyendo algunas de origen vegetal (Asamizu y col., 2000; Ali y col., 2000,

Reddy y col., 2002). En los estudios que constituyen esta tesis, se reportó por primera vez en girasol el aislamiento y caracterización de ESTs a partir de colecciones de ADNc órgano-específicas construidas por la técnica de hibridación por supresión substractiva (Diatchenko y col., 1996, 1999) como un método alternativo para identificar transcriptos de girasol expresados diferencialmente en un órgano o tejido en diferentes estadios de desarrollo en comparación con otros tejidos vegetales. Además, se analizó la eficiencia de los métodos de substracción y enriquecimiento para cada colección de ADNc generada (Fernández y col 2003).

I.6. Genómica del girasol

Pese a la importancia económica del girasol los estudios genéticos y genómicos en esta especie son limitados, siendo uno de los factores determinantes de ello la complejidad taxonómica del género *Helianthus*. Entre los factores que contribuyen a esta complejidad se encuentran el alto número de especies anuales y perennes, híbridos interespecíficos naturales o artificiales con distintos grados de introgresión por parte de especies silvestres, poliploidía en especies perennes y una amplia distribución geográfica. A pesar de los enormes avances científicos de la biología estructural (incluyendo la secuenciación nucleotídica de los genomas de varias especies) el conocimiento de las regiones cromosómicas en donde tiene lugar la inserción de muchos caracteres de interés teórico y otros importantes para la producción no está suficientemente explorado reduciendo considerablemente su potencial aplicación a la eficiencia de los procesos de mejora genética. Las especies silvestres de *Helianthus* son importantes fuentes de genes de resistencia a enfermedades, tolerancia a la sequía, androesterilidad citoplasmática y otros caracteres de interés agronómico. Los estudios de la citogenética clásica en híbridos han sido importantes herramientas para analizar las relaciones interespecíficas en el género. Los rearrreglos estructurales han sido importantes en la especiación; dado que las distintas especies difieren en una o más translocaciones recíprocas e inversiones paracéntricas y pericéntricas (Chandler y col, 1986). Los rearrreglos

podrían actuar como mecanismos de aislamiento reproductivo y podrían tener algún tipo de valor adaptativo manteniendo supergenes.

La complejidad taxonómica del género se debe a varios factores: ocurrencia de híbridos naturales e introgresión, poliploidía en especies perennes y una amplia distribución geográfica, lo que resulta en una intergradación morfológica en distintas formas con amplia variación fenotípica debida a componentes heredables y no heredables (Seiler y Rieseberg, 1997). Se ha postulado que la hibridación natural y el intercambio interespecífico de genes por retrocruzamientos (introgresión), no solo aumentan la variación genética en ciertas especies, sino que también permiten a estas especies, inicialmente anuales, ampliar su rango de distribución. Particularmente en el caso de la especie más ampliamente distribuida, *H. annuus*, la misma es capaz de expandir su rango por hibridación e introgresión con especies nativas adaptadas localmente (Seiler y Rieseberg, 1997). La variación en el tamaño del genoma es significativa entre muchas especies del género y particularmente entre cultivos del girasol. En distintos genotipos de *H. annuus* se han citado variaciones entre 4,9 y 9,9 pg. (Sims y Price, 1985). Estas significativas diferencias deben ser analizadas para conocer su causa y probable significado adaptativo. El advenimiento de técnicas de citogenética molecular como la hibridación *in situ* (FISH: *fluorescence in situ hybridization*, GISH: *genomic in situ hybridization*) brinda información adicional a la citogenética clásica para el análisis estructural de los genomas (Rocco, 2001). Estas técnicas permiten combinar información citológica acerca de la morfología nuclear o cromosómica con información molecular de la estructura de las secuencias con aplicaciones en diversas áreas como mapeo físico, relaciones genéticas y evolutivas entre especies, organización del genoma y arquitectura nuclear.

Como se mencionara en la sección anterior, el advenimiento de los proyectos genómicos a gran escala ha permitido un sustancial aumento en el número de ESTs descritos para muchas especies vegetales incluyendo al girasol. Actualmente hay depositadas en bases públicas 94.110 ESTs de girasol que representan 33.246 secuencias únicas (http://www.tigr.org/tdb/tgi/hagi/release_notes.html). Asimismo se han caracterizado 2.500 polimorfismos de secuencias únicas y/ o deleciones

pequeñas basados en la comparación de secuencias de ESTs provenientes de dos germoplasmas distintos (<http://cgpdb.ucdavis.edu/>) Por otra parte diferentes proyectos de genómica funcional han posibilitado el estudio de diversas proteínas antimicrobianas de girasol potencialmente implicadas en la resistencia a situaciones de estrés biótico (Regente y col. 1997, Regente y de la Canal 2000, Giudici y col. 2000, Regente y de la Canal 2001) y se ha clonado y analizado la expresión de genes de proteínas antifúngicas de girasol (Urdangarín y col. 2000, Regente y de la Canal 2003) pero aún no se ha caracterizado la diversidad de formas alélicas que pudieran existir para estos u otros genes candidatos de interés. Otro avance importante para esta especie ha sido el aislamiento y caracterización de secuencias de genes correspondientes a factores de transcripción de tipo HD-ZIP involucrados en la respuesta a distintos factores ambientales (Dezar y col. 2005), como estrés hídrico y vías dependientes de oscuridad en plantas (ej. biogénesis de cloroplastos) (Dezar, 2005; Rueda y col. 2005).

I.7. Secuencias EST

Las secuencias ESTs son generadas tomando clones al azar de una colección de ADNc y realizando una sola reacción de secuenciación con una extensión promedio de 300-500 bases. Sin embargo, las secuencias ESTs a menudo se generan a partir de colecciones de ADNc que han sido normalizadas para equalizar la abundancia de clones que corresponden a los diferentes transcritos. Los ESTs secuenciados a partir de estas últimas pueden ser comparados para identificar transcritos que se expresan en una colección y están completamente ausentes en otra, pero si se pretende obtener datos cuantitativos exactos que describan abundancia relativa se debe recurrir a colecciones de ADNc no normalizadas (Pessino y Martelotto, 2004). Si bien existen varias estrategias para la construcción de colecciones de ADNc diferenciales, la mayoría consiste en una técnica de hibridación de una población blanco (*tester*) con otra que contiene ARNm en exceso llamada de referencia (*driver*), y luego una separación de la fracción no hibridada obteniéndose los transcritos de interés. Los métodos tradicionales de hibridaciones sustractivas

(Bonaldo y col. 1996) (Patanjali y col. 1991) (Soares y col. 1991) (Asamizu y col. 2002), (Ali y col. 2002), (Reddy y col. 2002) eran altamente eficientes en la generación de transcritos diferenciales, pero no permitían una representación eficiente de transcritos poco abundantes. Las nuevas estrategias de construcción de colecciones de ADNc diferenciales permiten una eficiencia semejante en la amplificación de transcritos poco abundantes respecto a los más abundantes a través de la incorporación de etapas de supresión de amplificación de éstos últimos. Diatchenko y col. (1996) proponen una estrategia basada en esta metodología que permite ecualizar la abundancia relativa de secuencias específicas a través de la cinética del proceso de hibridación (Figura 2).

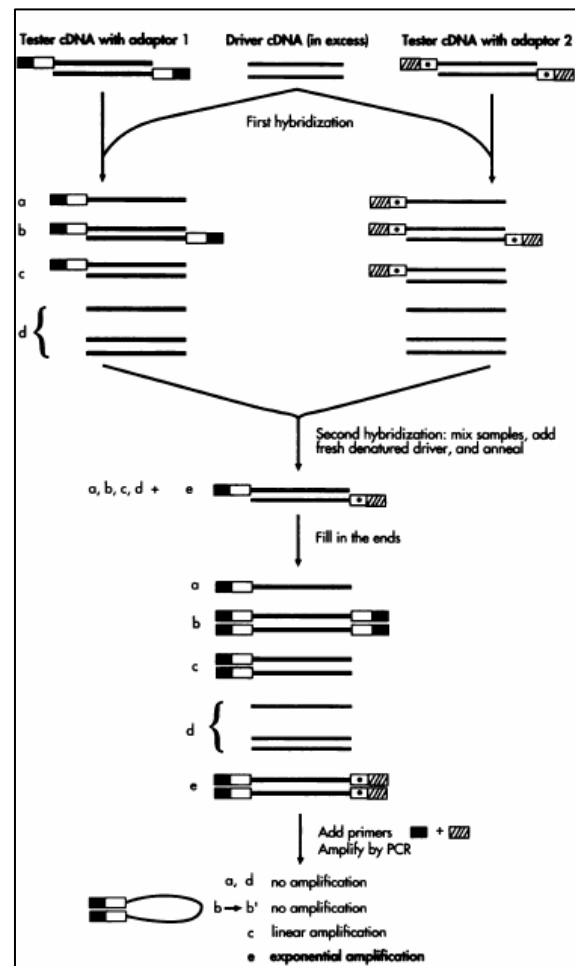


Figura 2: Esquema representativo de la metodología de hibridación sustractiva (SSH). (Diatchenko y col. 1996)

I.8. La Bioinformática

La Bioinformática es el área de las ciencias de la informática que aborda la búsqueda de soluciones a los problemas biológicos con herramientas computacionales. Tanto el análisis genómico como sus disciplinas relacionadas pueden ser abordados desde una perspectiva diferente a partir de la enorme disponibilidad de secuencias moleculares acumuladas en las bases de datos internacionales como GenBank, EMBL, DDBJ y SwissProt. El impacto de los proyectos genómicos durante los últimos 10 años, ha sido por un lado la creciente disponibilidad de secuencias, y por el otro la gran diversidad de datos biológicos acumulados. Esto, junto con el desarrollo de las tecnologías informáticas y la llegada de Internet han transformando la estructura de almacenamiento y acceso de los datos y ha permitido el desarrollo de aspectos fundamentales para el avance de los conocimientos sobre los sistemas biológicos como lo son: la búsqueda de similitud en secuencias moleculares, el análisis comparativo (filogenético, taxonómico y sinténico) de las especies, el análisis de macromoléculas, la ingeniería y diseño de estructuras proteicas, entre otros.

Según la definición del Centro Nacional para la Información Biotecnológica "National Center for Biotechnology Information" (NCBI por sus siglas en Inglés, 2001): "Bioinformática es un campo de la ciencia en el cual confluyen varias disciplinas tales como: biología, computación y tecnología de la información. El fin último de este campo es facilitar el descubrimiento de nuevas ideas biológicas así como crear perspectivas globales a partir de las cuales se puedan discernir principios unificadores en biología. Al comienzo de la "revolución genómica", el concepto de bioinformática se refería sólo a la creación y mantenimiento de base de datos donde se almacena información biológica, tales como secuencias de nucleótidos y aminoácidos. El desarrollo de este tipo de base de datos no solamente significaba el diseño de la misma sino también el desarrollo de interfaces complejas donde los investigadores pudieran acceder a los datos existentes y suministrar o revisar datos. Luego toda esa información debía ser combinada para formar una idea lógica de las actividades celulares normales, de tal manera que los

investigadores pudieran estudiar cómo estas actividades se veían alteradas en estados de una enfermedad, por ejemplo. De esta necesidad deriva el surgimiento del campo de la bioinformática y actualmente el campo más popular es el análisis e interpretación de varios tipos de datos, incluyendo secuencias de nucleótidos y aminoácidos, dominios y estructura de proteínas. El proceso de analizar e interpretar los datos es conocido como biocomputación. Dentro de la bioinformática y la biocomputación existen otras subdisciplinas importantes como son el desarrollo e implementación de herramientas que permitan el acceso, uso y manejo de varios tipos de información. El desarrollo de nuevos algoritmos (fórmulas matemáticas) y estadísticos con los cuales se pueda relacionar partes de un conjunto enorme de datos, como por ejemplo métodos para localizar un gen dentro de una secuencia, predecir estructura o función de proteínas y poder agrupar secuencias de proteínas en familias relacionadas”.

La Medicina Molecular y la Biotecnología constituyen dos áreas prioritarias científico- tecnológicas como desarrollo e innovación tecnológica. El desarrollo en ambas áreas está estrechamente relacionado. En ambas áreas se pretende potenciar la investigación genómica y postgenómica así como de la bioinformática. Debido al extraordinario avance de la genética molecular y la genómica, la Medicina Molecular se constituye como arma estratégica del bienestar social del futuro inmediato. Se pretende potenciar la aplicación de las nuevas tecnologías y de los avances genéticos para el beneficio de la salud. Dentro de las actividades financiadas, existen acciones estratégicas, de infraestructura, centros de competencia y grandes instalaciones científicas. En esta área, la dotación de infraestructura se plasmará en la creación y dotación de unidades de referencia tecnológica y centros de suministro común, como Centros de Bioinformática, que cubran las necesidades de la investigación en Medicina Molecular.

I.9. La necesidad de la bioinformática en el área de la genómica

Las bases de datos relativas a secuencias nucleotídicas y proteicas contienen abundantes y crecientes volúmenes de información. Los proyectos actuales de

secuenciación de genomas en desarrollo y de interacción entre proteínas están suministrando cantidades de información mayores aún. Cuando se dispone de una secuencia nucleotídica o aminoacídica, la búsqueda de secuencias similares en las bases de datos suministra información sobre la probable estructura y función de estas moléculas

1.9.1. Predicción de genes

La predicción de genes y unidades de transcripción sobre la secuencia del ADN genómico puede dividirse en dos etapas. La primera es la partición de la secuenciación en segmentos que podrían corresponder a genes individuales, lo que ha demostrado ser un proceso crítico y el punto débil de los algoritmos diseñados para encontrar genes *de novo*. La segunda fase consistiría en generar un modelo de gen que refleje la estructura más probable del transcripto o transcriptos codificados en esa región. Este proceso se puede realizar de modo exacto cuando se disponen de ADNc completos (proyectos de ESTs) o se conoce la secuencia de una proteína homóloga. Si esto no es así, la predicción *de novo*, aunque menos exacta, es la única herramienta disponible (Venter y col., 2001).

1.9.2. Caracterización de secuencias

La identificación de los posibles genes conduce al siguiente paso, la caracterización de los elementos anotados, lo que se realiza analizando las relaciones de una secuencia con otras de los mismos o diferentes genomas, estudiando motivos estructurales y funcionales conservados y, finalmente, tratando de predecir la función en base a los datos anteriores. La caracterización de secuencias en función de la similitud se ha realizado tradicionalmente usando métodos como BLAST (Altschul y col., 1997) o FASTA (Pearson, 1990), que siguen siendo insustituibles para el análisis de la función probable. Estos programas son capaces de detectar las regiones de similitud entre las secuencias de ADN como de proteínas. De esta manera, se pueden identificar aquellas secuencias previamente descritas que están

relacionadas con la que se está analizando, pudiendo determinar a qué familia pertenecen e inferir el tipo de función que podrían realizar (Terol y col., 2002).

I.10. Genómica funcional

Si la genómica estructural es la rama de la genómica orientada a la caracterización y localización de las secuencias que conforman el ADN de los genes, la genómica funcional consiste en la recolección sistemática de información sobre la función de los genes, mediante la aplicación de aproximaciones experimentales globales que evalúen la función de los genes haciendo uso de la información y elementos de la genómica estructural. Se caracteriza por la combinación de metodologías experimentales a gran escala con estudios computacionales de los resultados.

Para la genómica funcional el objetivo es llenar el vacío existente entre el conocimiento de las secuencias de un gen y su función a nivel molecular, celular y fisiológico, para de esta manera desvelar el comportamiento de los sistemas biológicos. Se trata de expandir el alcance de la investigación biológica desde el estudio de genes individuales al estudio de todos los genes de una célula al mismo tiempo en un momento determinado (Nuez y Carrillo, 2002).

I.10.1. Genómica funcional para la identificación y caracterización de procesos biológicos complejos

La genómica funcional es un área muy amplia de aplicaciones que tiene en común la identificación molecular específica (secuenciación) de las partes codificantes de los genomas y la caracterización de su función, es decir, se concentra en la prospección y caracterización molecular de genes e incluye varias subáreas como son: la caracterización de los ARNs mensajeros que se expresan en determinados tejidos o circunstancias de interés (como puede ser durante el ataque de un hongo fitopatogénico a una planta, durante la expresión de genes de virulencia de una bacteria zoonótica, o en respuesta aun estrés ambiental). Este subárea de la genética funcional se denomina transcriptómica y la caracterización suele realizarse en distintas etapas: en aquellos sistemas en los que no se conoce la secuencia

nucleotídica de su genoma puede involucrar una primera etapa de secuenciación de los ADN copias del ARN mensajero (EST) y caracterización por monitoreo y análisis de expresión concertada de muchos genes simultáneamente utilizando macro- o micromatrices de ADN (*microarrays* o *chips* de ADN).

En el siglo XIX, las observaciones del monje austriaco Gregor Mendel sugirieron que existía una relación entre un rasgo o carácter en un individuo y cierta información (almacenada en las células) que lo determinaba, y esta hipótesis llevó a proponer la existencia de entidades que luego se denominaron como genes. A mediados del siglo XX se postuló un modelo estructural del ADN que permitió comprender las bases moleculares de los eventos que determinan la herencia, la transcripción génica y la replicación conservativa del ADN entre otros, estableciendo los conceptos para el desarrollo de la tecnología que permitiría la manipulación del material genético y la relación de caracteres o enfermedades con un patrón de herencia definido. Posteriormente, en los años '80 se vislumbró la posibilidad de determinar las secuencias completas de los genomas de varios organismos, incluyendo el humano, dando así origen a la "genómica estructural" (Dhand, 2000). Más tarde, surgió el término "genómica funcional" como un concepto futurista, para definir la utilización de acercamientos globales para el entendimiento de las funciones de genes y/o proteínas, que permiten establecer vías genéticas integradas que gobiernan la fisiología de un organismo (Fields y col., 1999).

Actualmente la genómica funcional está impulsando múltiples campos de la biología y la medicina, incorporando nuevas metodologías para el estudio del ADN y las proteínas, con herramientas computacionales para el análisis, archivo y manejo de la información. Ejemplo de una de estas metodologías son las denominadas micromatrices de ADN, que permiten por ejemplo cuantificar la expresión genética del transcriptoma de una célula (Staudt y Brown, 2000). Debido que las micromatrices permiten obtener una visión global y cuantitativa de la expresión génica en el ARN mensajero (ARNm) producido por las células, esta tecnología ofrece grandes ventajas sobre métodos tradicionales que sólo permiten un análisis cualitativo o semicuantitativo de un número muy limitado de genes (Lockhart y

Winzeler, 2000). Las micromatrices constituyen parte de las nuevas metodologías aplicadas a la genómica funcional que están cambiando el modelo reduccionista científico hacia un enfoque más amplio y directo de la interrelación de los componentes que constituyen las células de diferentes organismos, hacia hipótesis dirigidas al entendimiento del funcionamiento global de un sistema celular (Vucmirovic y Tilghman, 2000).

I.11. Micromatrices en el análisis de expresión génica a gran escala

A finales de los años '80, la tecnología que desembocaría en la plataforma *GeneChip* fue desarrollada por cuatro científicos, en Affymax: Stephen Fodor, Michael Pirrung, Leighton Read y Lubert Stryer. El proyecto original estaba destinado a la construcción de péptidos sobre chips, pero desembocó en la capacidad para construir secuencias de ADN sobre micromatrices. La aplicación práctica de esta idea se llevó a cabo por la empresa Affymetrix, que comenzó a actuar como una compañía independiente en el año 1993.

Las micromatrices, por tanto, surgieron de la combinación de las técnicas microelectrónicas y el empleo de materiales biológicos. Se basan en la ultraminiaturización y paralelismo implícito y se concretan en micromatrices de material biológico de alta densidad de integración válidos para realizar distintos tipos de estudios repetitivos con muestras biológicas simples. Si en las micromatrices empleadas en las computadoras personales se consigue una alta densidad de integración de circuitos electrónicos en una oblea de silicio, en las micromatrices biológicas se logra una alta densidad de integración de material genético en una oblea de silicio, cristal o plástico. Los *chips* están divididos en unas pequeñas casillas que actúan cada una a modo de un tubo de ensayo en el que se produce una reacción. El número de estas casillas es muy elevado, llegando incluso a los centenares de miles. (Zepeda Garcia 2003).

I.11.1. Micromatrices de ADN sintetizados a partir de oligonucleótidos sintéticos

Las micromatrices de oligonucleótidos sintéticos se elaboran en una fase sólida similar y contienen fragmentos de oligonucleótidos representativos (generalmente entre 15 y 20 bases) de todos los genes de interés. Una variación de este tipo de matrices son las matrices de polimorfismos que contienen varios oligonucleótidos correspondientes a un sólo gen, con las variaciones en su secuencia que son representativas de polimorfismos de interés. Debido a la extrema miniaturización del sistema se pueden analizar en un único *chip* todas las posibilidades de mutación de un gen simultáneamente. Solo aquellos fragmentos de ADN que hibriden permanecerán unidos tras los lavados y dado que se conocen las secuencias y posiciones de los oligonucleótidos empleados, y mediante un escáner óptico que es capaz de localizar, mediante un proceso similar a la microscopía confocal, las cadenas marcadas con el fluorocromo se detectan los oligos que presentan señal diferencial. Una computadora analiza la información procedente del escáner y ofrece el resultado. La nomenclatura empleada para referirse a estas nuevas tecnologías es diversa y comienza por el término más general que es el de "Biochip" y hace referencia al empleo de materiales biológicos sobre un *chip*. Otros términos más específicos son: "DNA chip", "RNA chip" (según el material empleado) y "Oligonucleotide chip" o "DNA microarray", que hacen referencia al material y a la forma en la que se construye el chip. Existen también términos comerciales con los que referirse a las micromatrices que varían dependiendo de la tecnología empleada.

I.11.2. Micromatrices de ADN sintetizados a partir de ADNc

Las micromatrices de ADN se elaboran en una fase sólida mediante la aplicación automatizada de secuencias cortas correspondientes a los ADNc de un número variable de genes. Los ADN empleados en la construcción de estas micromatrices se obtienen a partir de clonotecas de ADNc que contienen fragmentos representativos de genes en el transcriptoma, los cuales se conocen como "secuencias de expresión cortas" (Expressed Sequence Tags - ESTs) y que son

identificadas con un número de referencia o acceso de fácil identificación en una base de datos. Las micromatrices utilizan soportes de cristal de dimensiones muy pequeñas para la impresión de los ADNc o secuencias EST, y el número de genes aplicados es del orden de decenas de miles. Así, en algunos estudios (Vodkin y col. 2004) se estudian en conjunto más de 22.000 genes, de los cuales aproximadamente un 40% se expresan en el tipo celular utilizado. En cuanto al sistema de detección del nivel de expresión de cada gen, en la tecnología de micromatrices se utilizan colorantes fluorescentes, por lo que se requiere equipos de lectura e instalaciones de elevado costo económico. En contraposición, las macromatrices consisten en membranas de *nylon* donde se depositan fragmentos de ADNc que serán hibridados a sondas radiactivas [P] obtenidas de tejidos o células tratadas en diferentes condiciones. La intensidad de las marcas (*spots*) correspondientes a cada gen es proporcional al nivel de expresión de este gen en la célula en una situación determinada. Las membranas para la realización de estos estudios de matrices pueden obtenerse comercialmente, y la realización de este tipo de análisis puede realizarse de forma bastante sencilla en un laboratorio de biología molecular con instalaciones para utilizar un equipo de captura de imágenes radioactivas o fluorescentes para determinar cuantitativamente la radiactividad o fluorescencia incorporada en cada punto de la membrana. Aunque el número de genes que se analiza es mucho más reducido en comparación con las micromatrices, hay membranas específicamente diseñadas en las que los ADNc de los genes implantados tienen relación con el metabolismo o la situación particular que se quiera analizar.

I.12. Metodología del uso de las micromatrices

Para determinar la expresión global de genes por medio de las micromatrices, se requiere la obtención de ARN mensajero (ARNm) de las células y/o tejidos a estudiar (células/tejidos con una característica especial o que han sido estimuladas) y de células/tejidos control (células normales o que no han sido estimuladas). Seguidamente, a partir del ARNm sintetiza ADN complementario (ADNc) por

medio de una reacción de transcripción reversa en la que simultáneamente se agregan nucleótidos marcados con fluorocromos tales como la Cyanina-3 (Cy-3) y la Cyanina-5 (Cy-5); alternativamente la marcación del ADNc se puede realizar con un isótopo radioactivo. La marcación con estos fluoróforos es en la actualidad la más utilizada, los cuales tienen emisiones en los rangos de 510-550 nm y 630-660 nm, respectivamente. En el proceso tradicional, para la muestra control se utiliza la Cy-3 y para la muestra problema se usa la Cy-5. La Cy-3 emite una fluorescencia verde, mientras que la Cy-5 emite una fluorescencia roja (Staudt y Brown, 2000). Esta marcación diferencial permite no sólo localizar las señales fluorescentes en la micromatriz facilitando la identificación de los genes, sino que también permite la cuantificación de las señales correspondientes a cada uno de ellos como una medida directa del grado de expresión del ARNm correspondiente.

Luego de adicionar el ADNc marcado a la matriz se debe esperar un tiempo prudencial (toda la noche) para favorecer la hibridación. En el caso de las matrices de vidrio o de plástico, los resultados son analizados en un fluorómetro el cual cuantifica la intensidad de la fluorescencia en cada punto de la matriz, correspondiente a un gen (Eyster y Lindhal, 2001). La intensidad para cada fluorocromo reportada por el equipo en cada punto de la matriz es la resultante de la relación $Cy5/Cy3$ (verde/rojo), la cual se expresa como intensidad de fluorescencia en escala logarítmica y en números absolutos: si ésta es 1 (color amarillo) significa que el gen se expresa de manera similar en ambas muestras (control y problema), si es menor que 1 (color verde) significa que el gen se expresó más en las células control y si la relación es mayor de 1 (color rojo) significa que el gen se sobre expresó en la muestra problema. Convencionalmente se emplea como punto de corte de significancia el valor de dos en la escala logarítmica. Finalmente, empleando algoritmos de asociación computacionales y con base en la regulación positiva y/o negativa de los genes bajo estudio, es posible establecer relaciones entre los patrones de expresión de los genes y diversas funciones celulares (Staudt y Brown, 2000). Comúnmente en los estudios de este tipo, estos datos se corroboran realizando análisis de Northern o qRT-PCT (ARNm) o de Western (proteínas) *blots* para aquellos genes que muestran una expresión diferencial (Eyster y Lindhal,

2001), permitiendo al investigador identificar características del sistema por medio de datos que pueden ser entendidos como una "escena" molecular en un estudio biológico

I.13. Esquema general para el análisis de datos de micromatrices

En la Figura 3 se resume en 5 (cinco) pasos el procedimiento necesario para hallar genes diferencialmente expresados utilizando micromatrices de dos canales, utilizando el paquete estadístico R (www.r-project.org), ampliamente difundido por ser libre y gratuito. No obstante, existen otros paquetes comerciales y académicos para el análisis de datos de micromatrices, como se describe en *III.1.1*.

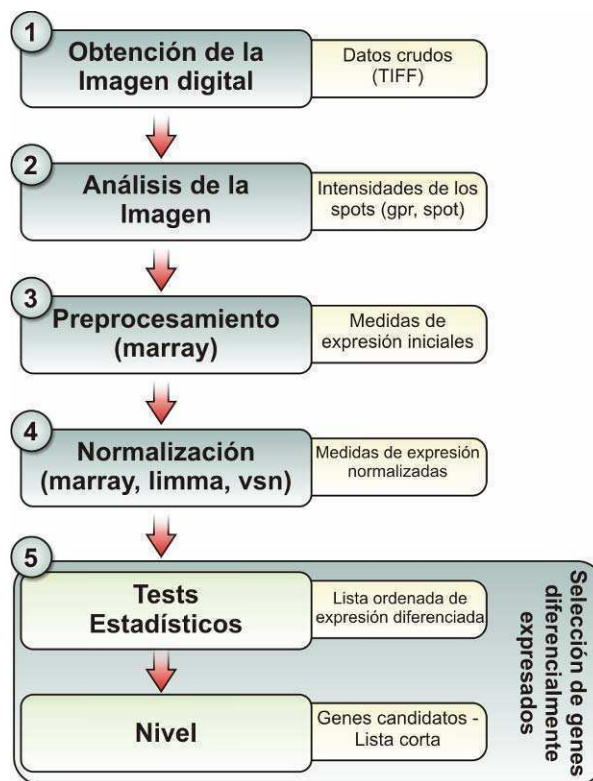


Figura 3 - Esquema general para el análisis de datos de micromatrices (Bello y col. 2006)

I.13.1. Obtención de la imagen digital

La imagen puede ser obtenida por medio de dos instrumentos, *scanners* o *imagers*. Ambos actúan excitando cada fluoróforo de cada molécula blanco mediante una luz monocromática producida por un láser (para el caso de los *scanners*) o luz blanca (*imagers*) y colectando la luz de emisión (fluorescencia) convirtiendo la corriente de fotones en valores digitales que pueden ser almacenados en una computadora como un archivo de imagen (usualmente con formato .TIFF). A cada fluoróforo le corresponde una longitud de onda de excitación y una longitud de onda de emisión diferente.

Para un típico experimento de micromatrices, se producen dos archivos de imagen, una para cada fluoróforo que en general para resumir, se presentan en una sola imagen superponiendo ambas imágenes. Para el caso en el que las intensidades de las señales de un punto o *spot* en ambas imágenes sean similares, este *spot* se mostrará de color amarillo (Figura 4).

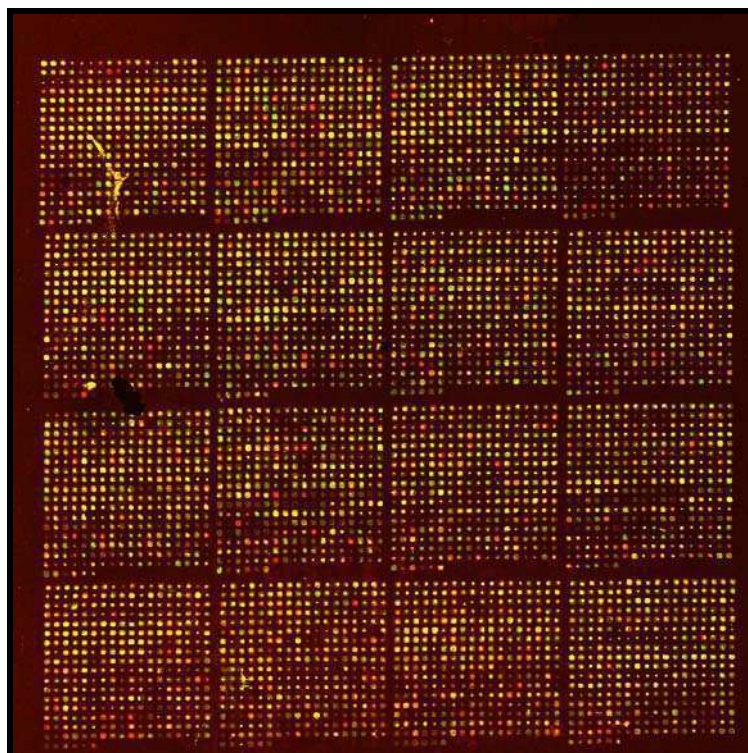


Figura 4: Imagen de una micromatriz de dos canales (Bello y col. 2006)

I.13.2. Análisis de la imagen

El primer propósito del análisis de la imagen es obtener las intensidades del fondo inespecífico o *background* y las señales específicas o *foreground* para los canales rojo y verde de cada *spot* en la micromatriz. Un propósito secundario del análisis de la imagen es obtener medidas cualitativas para cada *spot* que podría ser usado para marcar *spots* o *arrays* defectuosos o para evaluar la reproducibilidad de cada *spot*. Luego del análisis de la imagen obtenemos archivos del tipo .gpr, .spot, etc. dependiendo del software utilizado. Estos archivos contienen, entre otras cosas, las intensidades de cada canal correspondientes a los *spots* (*foreground*), *background*, medidas cualitativas (medidas de variabilidad, tamaño del *spot*, medidas de circularidad, entre otras).

El análisis de la imagen puede separarse en tres etapas, como lo demuestra la Figura 5:

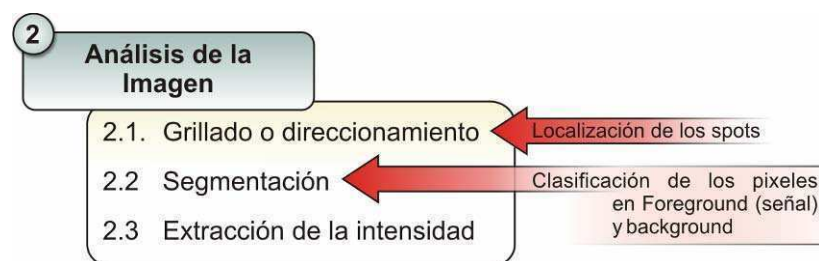


Figura 5: Esquema metodológico del análisis de la imagen (Bello y col. 2006)

I.13.3. Grillado o direccionamiento

Antes de segmentar la imagen primero se debe identificar la localización de cada *spot*. Este proceso es llamado grillado o direccionamiento. Bajo condiciones ideales los *spots* en cada bloque se encuentran equidistantes y en el mismo lugar

donde fueron impresos por las agujas. Sin embargo pequeñas variaciones durante la impresión de los *spots* puede causar irregularidades significativas en la imagen.

Para localizar los *spots* se debe localizar un rectángulo o cuadrado que los contenga. Estos rectángulos pueden obtenerse por medio de métodos de grillado automático o manual. Para el segundo caso es necesario estimar ciertos números de parámetros como la separación entre filas y columnas; pequeñas traslaciones individuales de los *spots*, etc.

I.13.4. Segmentación

La segmentación de la imagen puede ser definida como el proceso de partición en dos diferentes regiones, cada una con propiedades diferentes. La segmentación permite la clasificación de píxeles como *foreground* o *background*, utilizando una *spot mask*.

I.14. Aplicaciones de las micromatrices de ADNc

A pesar de ser una tecnología muy reciente y que, por lo tanto, está aún en vías de experimentación, actualmente las microcromatrices están siendo aplicados en:

- Monitorización de expresión génica: permite determinar cual es el patrón de expresión génica y cuantificar el nivel de expresión de manera simultánea para un elevado número de genes. Esto permite realizar estudios comparativos de activación de determinados genes en tejidos sanos y enfermos y determinar así la función de los mismos (Garzón y Jiménez, 2002).
- Detección de mutaciones y polimorfismos: Permite el estudio de todos los posibles polimorfismos y la detección de mutaciones en genes complejos (Garzón y Jiménez, 2002).
- Secuenciación: Si bien se han diseñado algunas micromatrices para secuenciación de fragmentos cortos de ADN, no existe aún en el mercado ningún chip que permita secuenciar de novo secuencias largas de ADN.
- Diagnóstico clínico y detección de microorganismos: Posibilitan la identificación rápida empleando unos marcadores genéticos de los patógenos (Amaratunga y Cabrera, 2004).
- Relevamiento y toxicología de fármacos: el empleo de las micromatrices permite el analizar los cambios de expresión génica que se dan durante la administración de un fármaco de forma rápida, así como la localización de

nuevas posibles dianas terapéuticas y los efectos toxicológicos asociados (Amaratunga y Cabrera, 2004).

- Seguimiento de terapia: las micromatrices permiten valorar rasgos genéticos que pueden tener incidencia en la respuesta a una terapia. (Zepeda Garcia , 2003)
- Medicina preventiva: El conocimiento y posible diagnóstico de ciertos caracteres genéticos asociados a determinadas patologías permite una prevención de las mismas antes de que aparezcan los síntomas (Zepeda Garcia, 2003)

I.14.1. Ventajas y desventajas

Las micromatrices aunque muy convenientes para el estudio global de la expresión de genes, tienen algunas desventajas con respecto a otras metodologías. A continuación se listan las principales ventajas y desventajas de esta metodología.

Las ventajas son:

- Facilitan el estudio de la expresión de miles de genes en forma simultánea en un sistema celular determinado (Seki y col., 2002).
- Permiten establecer una conexión biológica racional entre la función de un producto génico y su patrón de expresión, así como su relación con los patrones de expresión de muchos otros genes simultáneamente, es decir, facilitan identificar el grupo de genes expresados en una célula en un momento determinado. Esto permite analizar a fondo los sistemas bioquímicos y regulatorios que están operando y que funciones puede o no realizar la célula en un momento dado.
- La viabilidad técnica: es una tecnología de fácil transferencia (de hecho ya existe en nuestro medio tecnología adecuada para construirlas), la medición depende de una hibridación sencilla y hay suficiente información de librerías de secuencias para construirlas (Staudt y Brown, 2000).

Sus desventajas son:

- El costo elevado de su elaboración.
- La necesidad de fluoróforos y de *software* especializados en el análisis de la información que dependen en forma importante del proveedor que los suministra (Zepeda Garcia, 2003)
- La reproducibilidad de los datos no siempre es consistente. Este problema se ha resultado parcialmente haciendo mediciones por duplicado (dos puntos en la matriz que corresponden al mismo gen).
- La repetición de un experimento a otro puede mostrar variaciones importantes en el patrón de expresión lo cual hace necesario en algunos casos obtener múltiples réplicas biológicas (Amaratunga y Cabrera, 2004).
- Se requiere validación de los patrones de expresión para aquellos genes detectados como genes candidatos por técnicas de análisis transcripto específicas (Northern blot/ qRT-PCR) (Chuaqui y col., 2002) (Firestein y col., 2002).

II. OBJETIVOS E HIPÓTESIS

II.1. Objetivo general

Abordar el análisis genómico de la regiones funcionales del genoma de girasol a partir de la caracterización de perfiles de transcritos para diferentes órganos y estadios florales, con especial énfasis en procesos relacionados con respuestas a estreses biótico y abióticos, y con el propósito de establecer una plataforma efectiva para la exploración y caracterización de la diversidad molecular para el cultivo de girasol que facilite el descubrimiento de marcadores moleculares funcionales y genes candidatos para caracteres genéticos complejos de interés para el mejoramiento.

II.2. Objetivos específicos

- Construcción y ordenamiento de clonotecas de cDNA diferenciales de hoja, tallo, raíz y flores en distintos estadios de desarrollo - R1 y R4 - (Schneiter y Miller, 1981) de girasol.
 - Identificación, aislamiento, clonado y secuenciación de secuencias genéticas que se expresan o ESTs ("expressed sequence tags").
 - Desarrollo de una plataforma bioinformática para el análisis y anotación funcional de las secuencias generadas.
 - Diseño, mantenimiento y curado de una base de datos que integre las secuencias obtenidas con su función hipotética y demás características de manera relacional.
 - Caracterización de los patrones de expresión de transcritos durante procesos de estrés abiótico mediante el uso de micromatrices de ADNc.
 - Validación de transcritos expresados diferencialmente a través de la técnica de PCR en tiempo real (qRT-PCR) y/o *northern blot*.
-

II.3. HIPÓTESIS

La construcción de un banco local de secuencias ESTs a partir de colecciones de ADNc sustraídas órgano-específicas permitirá ampliar el número de secuencias de girasol especialmente aquellas funcionalmente relacionadas con respuestas a estreses abióticos dentro del marco de un proyecto de genómica de girasol local. Asimismo el desarrollo de herramientas bioinformáticas de aplicación al análisis de las secuencias generadas en el marco de este proyecto tendrá impacto en el desarrollo de otros proyectos genómicos de interés agronómico. Finalmente el estudio concertado de los perfiles de expresión de los transcritos correspondientes al banco de ESTs generado posibilitará la detección de genes candidatos involucrados en la respuesta a bajas temperaturas y salinidad con posible aplicación en el desarrollo de marcadores moleculares funcionales para asistir al mejoramiento del cultivo de girasol.

III. MATERIALES Y MÉTODOS

III.1 Material biológico

Para la construcción de las bibliotecas de ADN copia (ADNc), se utilizaron plántulas de girasol de la línea pública endocriada HA 89 (Fernández et al. 2003). Las plántulas fueron crecidas en invernáculo bajo condiciones controladas (20-24°C y fotoperíodo de 16 h luz/ 8 h oscuridad), y luego trasplantadas al campo durante la estación de crecimiento de manera de obtener plantas maduras. Se cosecharon diferentes órganos de plantas de dos meses, a saber: hojas, tallos, raíces y primordios florales de 1 a 2 cm de diámetro (estadios R1 y R2) y de 3 a 4 cm de diámetro (estadios R3 y R4).

III.2. Ensayos de estrés térmico y osmótico

Se crecieron plantas en invernáculo en las mismas condiciones previamente descritas, a partir de las cuales se obtuvo ARN total que se utilizó en la evaluación de perfiles transcripcionales. El diseño constó de 9 macetas de 5 litros de capacidad, con un sustrato mezcla de 70% tierra, 30% perlita con 3 plantas cada una: 3 macetas correspondieron a plantas control (C), 3 macetas a plantas sometidas a estrés térmico por frío (F) y 3 macetas a plantas sometidas a estrés por salinidad (S). El estrés por frío consistió en exponer las plantas de 15 días a una temperatura de 12 °C en cámara durante 24 horas previo a su cosecha (Huang y col. 2005), y el estrés por salinidad consistió en el riego de las mismas durante 3 días con 200 ml por maceta de solución 150mM de NaCl (adaptado de Liu y Bard, 2003), siendo esta concentración la máxima tolerable por una planta sin inducción de respuestas patológicas (Kreps y col. 2002; Seki y col. 2002; Taji y col. 2004). Todas ellas fueron regadas de la misma manera (un riego diario por la tarde) antes y después de recibir el estrés, excepto las sometidas a estrés osmótico que recibieron un riego diferencial, según se explicó precedentemente. Las plantas

control fueron crecidas bajo las mismas condiciones que las plantas utilizadas para la construcción de las bibliotecas de ADNc. Se cosecharon y conservaron en nitrógeno líquido los siguientes órganos de la planta aéreos y subterráneos, a saber: hojas, tallos y raíces, y luego fueron almacenadas a -80 °C hasta su procesamiento.

III.3. Extracción de ARN total de girasol

El ARN empleado para los ensayos de Northern Blot fue extraído empleando el reactivo TRIzol® (Invitrogen). Aproximadamente 2 g de tejido congelado a -80 °C fue pulverizado con nitrógeno líquido y homogeneizados en 20 ml de reactivo TRIzol®. Se procedió de acuerdo con las instrucciones sugeridas por el proveedor (Invitrogen, USA), con la modificación de la precipitación con 0,1 vol. de NaAc 3M pH 5,2 y 2 vol de etanol absoluto a -80°C hasta ser analizado. Se analizó la integridad y calidad del ARN por la técnica de electroforesis en gel de agarosa 1,5% utilizando buffer de corrida ME (400 mM MOPS, 100mM acetato de Na, 10 mM EDTA pH 8.0, todo solubilizado en agua DEPC). Se utilizó un buffer para la muestra desnaturante que contenía: formamida deionizada 60%, ME 1,2X, formaldehído 8,5%, azul de bromofenol/xilen cianol 0,01% P/V (Life Technologies, adaptado de Current Protocols in Molecular Biology, Vol. 1). Las muestras se incubaron 5 min a 65 °C e inmediatamente fueron transferidas a hielo, previo a la siembra en el gel.

III.4. Preparación de ARNm de girasol a partir de ARN total

La purificación de ARNm a partir de ARN total de tejido vegetal se realizó a partir de 200-500 ug de RNA total usando el sistema NucleoTrap® mRNA Midi Purification (Clontech, USA), siguiendo las instrucciones del fabricante. El método de purificación se basa en el uso de oligo d(T) biotinilado que hibridiza, en solución, con la cola de poli(A)+ de los ARNm. El ARNm se eluye de la fase sólida añadiendo agua. Luego de aislado, el ARNm fue cuantificado a través de la absorbancia a luz UV a 260 nm (GenQuant pro, Amersham-Pharmacia, UK).

III.5. Construcción de colecciones de ADNc

Las clonotecas de ADNc diferenciales fueron construidas a partir de diferentes tejidos incluyendo hoja, tallo, raíz y flor, éste último a su vez en diferentes estadios fenológicos (R1 y R4 según la escala descriptiva de Schneiter y Miller). Para ello se utilizó PCR-Select cDNA Subtraction Kit® (Clontech, USA). En la etapa inicial, el ADNc fue sintetizado a partir de 0.5-2.0 µg de poly (A+) ARN de dos tipos de tejido. El tejido “blanco” (llamado *tester*) y el tejido de referencia (llamado *driver*). Una vez sintetizado, el ADNc fue digerido con RsaI, que corta dejando extremos romos de 400 pb en promedio aproximadamente. La definición de tejido blanco y tejido de referencia depende de los intereses específicos de cada órgano. La colección de ADNc de hoja fue arrestada contra el ADNc del estadio floral R4, mientras que las colecciones de ADNc de tallo, raíz y estadio floral R4 fueron arrestadas contra la población de ADNc de hoja. Por otro lado, la colección de ADNc de flor temprana (R1) fue arrestada con el ADNc de flor tardía (R4). En todos los casos, las poblaciones de ADNc fueron procesadas siguiendo las instrucciones del fabricante, con algunas modificaciones. El tejido “blanco” fue subdividido en dos mitades, y cada mitad fue ligada a diferentes adaptadores. Se realizaron dos rondas de hibridación con un exceso del ADNc de tejido de referencia. Las condiciones de hibridación fueron realizadas según lo recomendado por el proveedor. Los productos resultantes fueron sometidos a dos ciclos de PCR con oligonucleótidos específicos como adaptadores de modo de amplificar secuencias diferencialmente expresadas. Dichas amplificaciones se realizaron en un termociclador de DNA PT 100 (MJ Research, USA). La mezcla de reacciones de la PCR contenía 10x PCR buffer, 0.2 mM dNTPs, 0.4 µM PCR primer 1 y Advantage cDNA polimerasa (Clontech, USA). La PCR se realizó bajo las siguientes condiciones: desnaturalización a 94°C (30 seg) seguido de 27 ciclos, y cada ciclo a su vez consistió en un paso de desnaturalización a 94°C (30 seg), un paso de hibridación a 66°C (30 seg) y un paso de elongación a 72°C (10 min). La segunda mezcla de reacciones de PCR contenía 10x PCR buffer, 0,2 mM dNTPs, 0.4 µM oligonucleótido anidado 1, 0,4 µM oligonucleótido anidado 2 y Advantage

cDNA polimerasa. El ciclo de PCR constó de 12 ciclos, y cada ciclo a su vez constaba a su vez de un paso de desnaturalización a 94°C (30 seg), un paso de anillado a 66° (30 seg) y un paso de elongación a 72°C (1,5 min).

Las moléculas de ADNc se seleccionaron por tamaño y aquellas mayores a 250 pb fueron clonadas de manera no direccional en el vector pGem-T-Easy Vector® (Promega, USA). La ligación se realizó a 4°C durante 48 h y el producto de ligación resultante se utilizó para transformar *Escherichia coli* (cepa XL1-blue) mediante electroporación (Pulse Controller, BioRad, USA), según se detalla en 3.5.2 y 3.5.3., respectivamente..

III.5.1. Cuantificación de ácidos nucleicos

La concentración del ADN y el ARN en solución se estimó a partir de la absorbancia a 260 nm, considerando que $D.O_{(260\text{ nm})} = 1$ equivale a 50µg/ml de ADN o a 40µg/ml de ARN. En el caso de las preparaciones de ARN, también se midió la absorbancia a 280 nm, pico máximo de absorción de las proteínas, para verificar que la relación de $Abs_{.260}/Abs_{.280}$ fuese cercana a 2, indicando que el ARN analizado correspondía a una preparación de alta calidad, prácticamente libre de proteínas. Para las mediciones se utilizó el espectrofotómetro Gen Quant Pro (Amersham Pharmacia Biotech, Inc.)

III.5.2. Ligación de moléculas de ADN

La ligación de fragmentos de ADN se llevó a cabo utilizando 1U de T4 ADN ligasa (Promega), en un volumen de reacción de 10 ul empleando la solución amortiguadora de reacción provisto por el fabricante de la enzima. Se utilizaron cantidades de inserto y vector tales que la relación molar entre ambos fuera de 5 a 1. La incubación se realizó durante 16 horas a 4 °C y se utilizó como vector de clonado el plásmido pGEM®-T easy (Promega), siguiéndose el protocolo descrito por el fabricante.

III.5.3. Transformación de *E. coli*

Para la preparación de células competentes de *E. coli* y su posterior transformación se ha seguido el método descrito por Ausbel y col. (1997) en “Current Protocols in Molecular Biology” con algunas modificaciones. Las células de *E. coli* de la cepa XL1-Blue se cultivaron toda la noche en medio LB (peptona de carne 10g/l, extracto de levadura 5g/l, NaCl 10g/l; pH 7.0) a 37 °C con agitación en presencia del antibiótico adecuado.. Con este cultivo saturado se inocularon 200 ml de medio LB fresco y se dejó crecer durante 4-5 horas hasta alcanzar una $DO_{(600)} = 0,4$. Las células fueron cosechadas por centrifugación a 3000 g durante 20 minutos a 4 °C. El sedimento obtenido fue lavado con 10 ml de Tris-HCl 10mM (ph 7,5), EDTA 1 mM. Finalmente, el sedimento fue resuspendido en 20 ml de medio LB y se fraccionaron alícuotas de 50 μ l. Una vez obtenidas las células competentes, los tubos se mantuvieron en hielo y se agregaron entre 1 μ l de la mezcla de ligación. El choque eléctrico se realizó en cubetas de 0,1 cm (Bio-Rad) y las condiciones de electroporación empleadas fueron las recomendadas en el manual del fabricante del equipo (*Gene PulserTM*, Bio-Rad Laboratorios Inc., USA). Inmediatamente después del disparo se adicionó 1 ml de medio LB a la suspensión de células y la mezcla se incubó durante 1 hora a 37 °C. Finalmente, alícuotas de 200 μ l se sembraron en placas que contenían LB suplementado con ampicilina y en presencia de X-gal e IPTG para la detección de clones recombinantes. Las placas fueron incubadas a 37 °C durante 18 horas.

III.5.4. Preparación del templado

Las colecciones de ADNc fueron plaqueadas en medio sólido Luria Bertani (LB, según descripción en 3.5.3.) conteniendo ampicilina. Para preparar el medio sólido se añade 15g/l de agar. Los clones recombinantes se seleccionaron por actividad de β - galactosidasa en un medio que contiene X-GAL e IPTG. Se seleccionaron al azar las colonias blancas y se picaron en placas de 384 pocillos conteniendo Freezing Media (36 mM K_2HPO_4 , 13,2 mM KH_2PO_4 , 1,7 mM citrato de sodio, 0,4

mM MgSO₄, 6,8 mM (NH₄)₂ SO₄, en medio LB y 4,4 % glicerol), se crecieron durante toda la noche y posteriormente se almacenaron a -70°C en las mismas placas y en criotubos individuales por duplicado de manera de evitar la posible contaminación cruzada entre clones dentro de los pocillos de la placas. Los plásmidos recombinantes se aislaron utilizando REAL 96 *prep kit* (Qiagen, Germany) según recomendaciones del proveedor. El tamaño de inserto de los clones recombinantes fue examinado por electroforesis de los productos de digestión con la enzima *EcoRI* en gel de agarosa 1,2 % en buffer TAE buffer 1X (Tris-Ac 20 mM pH 8,5, EDTA 1 mM).

III.5.5. Electroforesis de ADN en geles de agarosa

La electroforesis de los fragmentos de ADN que se querían detectar y/o cuantificar se llevó a cabo en geles de agarosa 1% (p/v), empleándose solución TAE 1X (Tris-Acetato 20 mM, pH 8,5, EDTA 1 mM) tanto para la preparación del gel como para la solución de corrida. La concentración de bromuro de etidio utilizada fue de 0,3ug/ml y el volumen de muestra se mezcló con 1/10 vol. de solución de siembra (azul de bromofenol 0,25% p/v y xilencianol FF 0,25% p/v). El gel se corrió a voltaje constante entre 1 y 5 V por cm de gel. Para la visualización del ADN en el gel se empleó un transiluminador de luz UV. En todos los casos en los que fue necesario estimar la longitud de los fragmentos de ADN separados mediante electroforesis en geles se empleó como marcador 1 KB Ladder (Invitrogen, USA) según recomendaciones del fabricante.

III.6. Reacciones de determinación de la secuencia de ADN

La secuenciación de los plásmidos recombinantes se realizó en los siguientes secuenciadores respectivamente: Laboratorio de Alta Complejidad, IMyZA – CICVyA – INTA Castelar, Argentina; Centro de Biología Molecular e Engenharia Genética - CBMEG Universidade Estadual de Campinas, Sao Paulo, Brazil y/o Department of Plant Pathology, Kansas University. El orden

reverso de secuenciación fue realizado utilizando el cebador SP6 sólo cuando las secuencias directas a partir del cebador T7 fallaron o fueron poco informativas debido a una longitud corta.

III.7. Análisis de secuencias

Las secuencias correspondientes al vector y las secuencias no informativas fueron removidas automáticamente usando rutinas bioinformáticas. Las secuencias procesadas fueron convertidas a formato FASTA y posteriormente procesadas mediante una rutina “pile-up” (Biopipeline®), tarea realizada en colaboración con profesionales del área bioinformática (S.L., Bioaxioma S.A.). Para las búsquedas de homología se consultó el GenBank a través de los portales de National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) y DNA Database of Japan (<http://www.ddbj.nig.ac.jp>), utilizando el algoritmo BLAST (Altschul y col., 1997). El desarrollo, uso y mantenimiento de esta plataforma y su servidor, así como también el de la base de datos generada para este trabajo se describen en el capítulo 4: Resultados y discusión.

III.8. Northern blot

Un total de 20 ug de ARN total de cada muestra fueron analizados por electroforesis en geles de agarosa 1,5% en un buffer de corrida MOPS 1X y voltaje constante (80 V). Previamente a ser sembradas las muestras fueron desnaturalizadas por incubación a 65°C durante 5 minutos con 3 volúmenes del buffer de carga denaturalizante. Para visualizar los ARN bajo luz UV se agregó al gel 0,3 ug/ml de bromuro de etidio. El ARN fue transferido, durante toda la noche, a una membrana Hybond N+ (Amersham Pharmacia Biotech, Inc.) según las instrucciones del proveedor. Luego de la inmovilización, se dejó secar las membranas y el ARN fue fijado por exposición a luz UV durante 3-4 min. De esta manera las membranas quedaron preparadas para ser prehibridadas e hibridadas con la sonda radioactiva.

III.8.1. Amplificación de fragmentos de ADN mediante la reacción en cadena de la polimerasa (PCR) para uso de sondas específicas

Los volúmenes de reacción fueron de 20 o 50 ul, empleándose 0,75 a 1,5 U, respectivamente, de enzima Taq ADN polimerasa (Invitrogen, USA) por reacción. Se utilizó la solución amortiguadora de reacción provista por el fabricante llegando a una concentración final de $MgCl_2$ 2 mM; dNTP 0,2mM c/u y 0,4 μ M de cada oligonucleótido específico. A esta mezcla de reacción se le incorporó el ADN molde y luego se le agregó la enzima. Las reacciones se llevaron a cabo en un termociclador Eppendorf y se utilizó el siguiente programa, en donde la temperatura de hibridación se estableció de acuerdo con la composición de bases de los oligonucleótidos utilizados, aplicando la siguiente relación para su cálculo: $T_m = 2x(A+T) + 4x(G+C) - 5$ °C . Programa: 30 ciclos: 1 minuto 30 seg. a T_m , 2 minutos a 72 °C; 1 ciclo: 10 minutos a 72 °C.

III.8.2. Purificación de fragmentos de ADN de geles de agarosa

Una vez corridas las muestras de ADN mediante la técnica de electroforesis en agarosa según lo descrito en 3.5.5. e identificadas mediante visualización sobre transiluminador UV, las bandas correspondientes a los fragmentos esperados se aislaron utilizando el kit comercial GFX™ PCR DNA y Gel Band Purification Kit (Amersham Pharmacia Biotech, Inc.).

III.8.3. Pre-hibridación de membranas

El buffer utilizado para las prehibridaciones e hibridaciones fue el sugerido por Amersham Pharmacia Biotech, Inc. para las membranas de Nylon N+ o solución de hibridación UltraHyb (Ambion, USA). Las prehibridaciones se realizaron durante 3 horas a 42 °C en tubos de hibridación con rotación continua en un horno de hibridación marca Biometra.

III.8.4. Marcación de sondas

Se prepararon sondas marcadas con α -³²P-dCTP (actividad específica 3000 Ci/ug, New England Nuclear, EUA) mediante la técnica de iniciación al azar o “random priming” (Prime a Gene, Promega,EUA) según se muestra en Tabla 2:

Componentes	[final]
Buffer 10X	1X
Hexámeros al azar	0,25 μ M
Fragmento Klenow de la ADN polimerasa	1 UE
dNTPs (AGT)	0,03 mM
α - ³² P-dCTP	0,02 mCi
ADN (desnaturalizado)	30 ng
dH ₂ O	Llevar a 50 ul

Tabla 2 - Componentes de la mezcla maestra de reacción de marcación de sondas

La reacción se incubó durante 1 h a 37°C y luego se desnaturalizó la sonda marcada radioactivamente incubando la reacción durante 10 min con NaOH 0,4N.

III.8.5. Hibridación, lavados y exposición de membranas

La sonda se adicionó a los tubos de hibridación conteniendo las membranas en la solución de hibridación y se mantuvo en contacto con las mismas durante 18 hs a 42°C en un horno de hibridación con rotación continua, marca Biometra. Posteriormente, las membranas se lavaron 2 veces con 2X SSC / 0,1% SDS durante 5 min y 2 veces con 0,1X SSC/0,1% SDS durante 15 min para remover la sonda hibridada inespecíficamente, también a 42°C. Las membranas se expusieron en placas autorradiográficas de alta sensibilidad (Amersham Pharmacia Biotech, Inc.) o Biomax MS (Kodak) a distintos tiempos según la

actividad específica de la sonda y según la intensidad de señal esperada. Esta intensidad se estimó previamente con un contador Geiger, sin la tapa protectora.

III.9. RT-PCR (reacción de PCR a partir de transcripción reversa)

Para la síntesis de la primera cadena de ADNc, templado para la PCR posterior, se utilizaron 0,5 µg de ARN, utilizando 200 unidades de la enzima Superscript III (Invitrogen), en presencia de solución amortiguadora de primera cadena (50mM TrisClH pH 8,3, 75mM KCl, 3mM MgCl₂), 150ng/ml de hexanucleótidos al azar (Invitrogen), 10mM de dNTPs y 0,1 M DTT en 20 ml de reacción. Paralelamente se llevó a cabo la misma reacción sin la enzima RT, para evaluar la presencia de ADN contaminante. Los oligonucleótidos utilizados se detallan en la Tabla 3:

EF127F	5'GCATTGGGCAGATCTTGTTT 3'
EF127R	5'GTCCCCTTTGGAGGCAGTA 3'
F126F	5' CAAAATGCAACGACCCATTA 3'
F126R	5' TCTGTACGCCCTCATGTTCA 3'
F231F	5' CAACAAAAGCAGACGCTGAA 3'
F231R	5' AGCATGTGGTGTGGACAG 3'
F455F	5 GCCGAGGTACAAACTGGAGA 3'
F455R	5' TGAGCATGATCTGAATATCTTGAA3'
F543F	5' ACGGAAGCGTTGTTTGGTAA3'
F543R	5' TCAACATCCCACAGAAACGA3'
F557F	5' CGCAATTGCTATTGATGGAA3'
F557R	5' ACACCGGTATGGTTGATGCT3'
H110F	5' ACGCGAGTCGGTTGTTTTAT3'
H110R	5' TCATTTTCTCCACCCATGGTA3'
H123F	5' GGCAGGTACCAGGGGTTATT3'
H123R	5' GAGGTTTCATTCCGTCGTTGT3'
H136F	5' TTTGCAAGGATGAATGGTGA3'
H136R	5' GTGACCCGAACTCCTTGGTA 3'
H360F	5' GGCAGCCAATCCTCTTGATA 3'
H360R	5' CGACTCCGCCAAATACAGAT3'

H385F 5' TTCAGCCCGGAAAGAATATG 3'
H385R 5' AACTTTGCAGTGGGACCATC 3'
T283F 5' CTCACGAAAGCTTCCTGCTT 3'
T283R 5' GCAGGTACTCGGTTTGTTC 3'
T340F 5' AAGACGGTGGATTTGAGGTG 3'
T340R 5' AACCTTGCCTGCTTTCTCA 3'
F550F 5' CAGAGACGTTCTTGCGTTGA 3'
F550R 5' CGCACACAACAAAGAAATGG 3'
F227F 5' TTGTCACAGACGAGGAGGAA 3'
F227R 5' GGCAGGTCCGTTATCACACT 3'
Actina F 5' AGGGCGGTCTTTCCAAGTAT
Actina R 5' ACATACATGGCGGGAACATT

Tabla 3 – Oligonucleótidos diseñados para PCR cuantitativa

III.9.1. PCR cuantitativa (en tiempo real)

Las reacciones de PCR cuantitativa se realizaron utilizando el kit comercial QuantiTect™ SYBR® Green PCR (QIAGEN), siguiendo las instrucciones del fabricante. Los oligonucleótidos fueron sintetizados utilizando el programa Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Las reacciones se llevaron a cabo en un volumen final de 25 µl, según se describe en la Tabla 4:

Componentes	[final]
2X QuantiTect SYBR Green PCR	1X
Oligonucleótido A	0,3 uM
Oligonucleótido B	0,3 uM
ADNc	60 ng
H ₂ O bidestilada libre de RNAsa	Llevar a 25 ul

Tabla 4 – Componentes de la mezcla maestra de reacción de PCR cuantitativa

Las reacciones se llevaron a cabo en un equipo ABI Prism 7000 HT Sequence Detection System (Applied Biosystems) utilizando el siguiente programa: 2 min a 50°C, 10 min a 95°C y 15 seg a 95°C por 40 ciclos y 1 min a 60°C. Una vez terminada la amplificación se creó la curva de disociación, medida como cambios en los niveles de fluorescencia en función de la temperatura, de modo de detectar los productos no específicos. El programa de disociación utilizado fue el siguiente: 15 seg a 95°C, 15 seg a 60°C seguido de una suave rampa de 20 min desde 60°C a 95°C. Se realizaron 3 réplicas de cada reacción con *actina* como gen control de normalización (expresión constitutiva). La cuantificación de los cambios relativos en la expresión génica se realizó utilizando el método $2^{-\Delta\Delta CT}$ según lo descrito por Livak y Schmittgen (2001).

III.10. Protocolos clásicos de biología molecular

Las técnicas de minipreparación y maxipreparación de plásmidos y digestión de ADN con enzimas de restricción se realizaron según describe Sambrook y col., 1989.

III.11. Generación de micromatrices de ADN

III.11.1. Obtención de ADNc amplificados para impresión

Para la impresión de micromatrices de ADNc sobre soporte de vidrio, se amplificaron por PCR los fragmentos representativos de los unigenes anotados en las bibliotecas previamente caracterizadas. Las reacciones se realizaron en volúmenes de 50 μ l según se detalla a continuación, utilizando el kit Expand Long Expand Template PCR System (Roche Diagnostics, Inc.), siguiendo especificaciones del fabricante (Tabla 5).

Componentes	[final]
Buffer 10X	1X
dNTPs (AGT)	0,2 mM
Oligonucleótidos	0,25 mN
Taq polimerasa	0,75 U
ADN (desnaturalizado)	60 ng
H ₂ O bidestilada	Llevar a 50 μ l

Tabla 5 – Componentes de la mezcla maestra de reacción de PCR para amplificación de clones de ADNc y su uso en la impresión de micromatrices

Los oligonucleótidos utilizados fueron diseñados sobre la zona *lacZ* del vector (Figura 6), siendo la secuencia la siguiente (Invitrogen, 85586B04 y 85586B05, respectivamente):

Secuencia 5' a 3' (LacZ1): GCT TCC GGC TCG TAT GTT GTG TG

Secuencia 5' a 3' (LacZ2): AAA GGG GGA TGT GCT GCA AGG CG

El programa de ciclado y la termocicladora fueron los mismos citados en III.8.1, excepto la temperatura de anillado de los oligonucleótidos que se programó siguiendo una rampa de descenso de 60°C a 55°C, acorde a las T_m de cada uno de ellos.

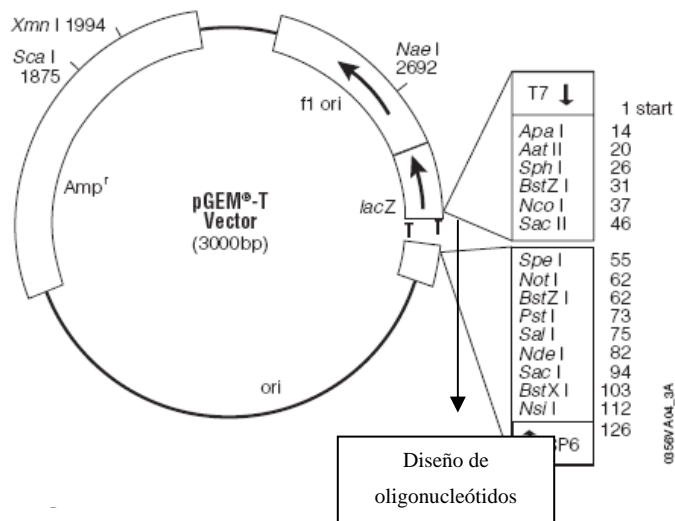


Figura 6. Mapa circular del vector pGEM®-T Easy y sus puntos de referencia.

III.11.2. Impresión de las micromatrices

Se verificó la presencia de los productos obtenidos en geles de agarosa 1%, sembrándose un marcador de masa molecular (Low DNA Mass Ladder, Invitrogen 10068013) de modo de cuantificar las concentraciones entre 100 y 200 ng/ul para lograr una buena calidad de impresión. Para la purificación de los productos se utilizaron placas filtro de 96 wells (Millipore MANU3050) sobre una bomba de vacío tipo manifold de Millipore. La impresión de las micromatrices de ADN se realizó a través de un servicio ofrecido por Gentrion Genomic Services (Gentrion, Suc Argentina), utilizando el equipo VersArray Chip Writer (BioRad). Se imprimieron 6 grillas de 8X8 puntos (supergrilla) por cuadruplicado sobre soportes de vidrio (Ultragap II, Corning Systems). Se incluyeron como control la secuencia codificante para actina como control positivo, y tres clones correspondientes a genes de ratón como control negativo (Números de acceso GenBank: NM009060, NM008690, NM019476).

III.11.3. Hibridación y escaneo de las micromatrices

Las micromatrices obtenidos fueron utilizadas para evaluar la expresión relativa de los transcritos representados en forma simultánea en tejidos control y tratado mediante hibridación de sondas marcadas con distintos fluoróforos Cy3 y Cy5 (Amersham Pharmacia Biotech, Inc.). Se evaluaron los ARNs extraídos de 3 muestras biológicas de control, de 3 muestras tratadas con frío y 3 muestras crecidas bajo condiciones de alta salinidad. Se utilizó el sistema “SuperScript Indirect RNA Amplification System” (Invitrogen, EUA) para amplificar el RNA partiendo de 800 ng RNA extraído de cada muestra basado en el método de síntesis de DNA complementario *in situ* (Eberwine y *col.* 1992). El RNA ya amplificado y habiendo incorporado aminoalil UTP en la transcripción *in vitro*, se marcó fluorescentemente con Cy3 o Cy5 por incubación en medio alcalino con ésteres de Cy3 o Cy5 respectivamente. Los vidrios impresos fueron escaneados (usando canales para ambos fluoróforos) a tres intensidades diferentes usando el lector de fluorescencia VersArray Chip Reader (BioRad). Las imágenes obtenidas fueron analizadas utilizando el programa Spotfinder (www.tm4.org/spotfinder.html), cuantificándose la intensidad de señal para cada spot. Luego, se realizó una integración de los datos de las múltiples imágenes escaneadas.

III.11.4. Preprocesamiento de datos

La normalización y el análisis de expresión global se realizaron a través de programas generados utilizando el programa R (University of Auckland) a través de su versión R 1.9.0 (<http://www.rproject.org>). Los datos de cada vidrio fueron normalizados corrigiendo la dependencia por intensidad (LOWESS). Se realizó una integración de los datos de las réplicas técnicas dentro de cada soporte (cada *spot* fue impreso 4 veces) y las réplicas técnicas entre soportes (intercambio de fluoróforos o “dye-swap”). Se generó una matriz de expresión génica cuyo análisis se focalizó en la identificación de genes con expresiones diferenciales. Todos los

análisis relativos a la exploración de la matriz génica se realizaron con el software Infostat 2006® (Grupo Infostat, FCA, Córdoba, Argentina).

III.11.5. Normalización entre réplicas biológicas

Los métodos utilizados en la normalización entre réplicas biológicas tienen como supuesto que la gran mayoría de los genes no alteran su expresión entre tratamientos. Bajo esta suposición, la ecualización de las escalas o la ecualización de los cuantiles no modifican substancialmente la posibilidad de detectar patrones de expresión diferencial. Sin embargo, en el contexto de la presente aplicación, no pareció razonable aplicar estas normalizaciones (adicionales) ante el peligro de destruir o confundir los patrones de expresión diferencial, aún cuando se observó una importante diferencia de escala entre el caso control y los tratados que podría sugerir lo contrario. Para interpretar esta diferencia de escalas se tuvo en cuenta que cuando se hace referencia a la “escala” de las expresiones génicas por vidrio no se está haciendo referencia a la escala de una variable aleatoria sino a una pseudoescala que refleja variabilidad de una mezcla de variables aleatorias. En esta mezcla, una mayor variabilidad de las expresiones génicas de los vidrios de un tratamiento respecto de los vidrios de otro tratamiento puede reflejar tanto la variación atribuible al error experimental como también a los efectos de tratamiento que producen la sub- o sobreexpresión relativa de algunos genes. Esto es particularmente esperable en un experimento como el que se consideró en este trabajo, donde el valor esperado de la expresión génica en el control debería ser cero (valor esperado del logaritmo de la razón entre los canales rojo y verde), mientras que si los tratamientos de frío y salinidad produjeran un cambio de expresión génica, eso debería reflejarse en un cambio de las pseudoescalas. (*Di Rienzo JA, 2006, comunicación personal*)

III.11.6. Consistencia de las repeticiones

Para evaluar cuán disímil son las repeticiones biológicas de este experimento se obtuvo una ordenación de las mismas en un plano de ordenación generado mediante análisis de componentes principales aplicado a la matriz de expresiones génicas. Esta aproximación tiene en cuenta la información de todos los genes simultáneamente y permite establecer si una repetición es atípica o no.

III.12. Análisis estadístico. Identificación de grupos de genes con expresión diferencial.

El análisis de la matriz de expresión génica se focalizó en la identificación de genes con expresiones diferenciales. Existen diversas estrategias para obtener un listado de genes candidatos. La estrategia más sencilla es aplicar una prueba de hipótesis tradicional o eventualmente no paramétrica gen a gen. Esta aproximación tiene el inconveniente de generar un gran número de falsos positivos. Los métodos correctivos, basados en el control de la tasa de falsos positivos tienen el inconveniente de generar asimismo una alta tasa de falsos negativos. Por consiguiente la metodología utilizada en este trabajo consistió en complementar las técnicas inferenciales clásicas con un criterio de selección (ver resultados) basado en el ordenamiento de genes y tratamientos en el espacio generado por los dos primeros ejes principales obtenidos de un análisis de componentes principales de la matriz de expresión génica. Posteriormente, y dentro del conjunto de genes seleccionados en el paso anterior, se realizó el reconocimiento de grupos de genes con perfiles de expresión similar mediante la aplicación de un algoritmo de clasificación no supervisada.

IV. RESULTADOS

CAPÍTULO 1: OBTENCIÓN DE COLECCIÓN DE SECUENCIAS DE TIPO EST

IV.1.1. Colecciones de ADNc diferenciales

Se construyeron colecciones de clones (clonotecas) de ADNc mediante la técnica de hibridación sustractiva (SSH) de los siguientes órganos: flor (estadios R1 y R4), hoja, tallo y raíz. En primer lugar se evaluó la eficiencia de dicho procedimiento mediante un análisis de transcritos diferenciales obtenidos de las colecciones de ADNc recíprocas: se usaron preparaciones ARN poliadenilado a partir de flor en estadio R4 y de hoja para generar ADNc “tester” y “driver”, respectivamente, para cada una de las colecciones. La comparación de las secuencias de transcritos diferenciales de ambas colecciones arrojó patrones claramente diferenciales entre las mismas. Dichos datos mostraron a nivel de comparación de secuencias que el 94% (209 de 222 secuencias) y 62,5% (80 de 128 secuencias) de las secuencias analizadas eran únicas para el estadio R4 de flor y hoja, respectivamente (Tabla 6). El alto porcentaje de unigenes órgano-específicas obtenidas a partir de hoja y flor en estadio R4 confirmó la potencialidad de esta estrategia para detectar transcritos inducidos diferencialmente en otros órganos o estadios de desarrollo a través de la construcción de clonotecas adicionales. Estas colecciones incluyeron una de flor en estadio temprano de desarrollo (R1) arrestada con flor de un estadio mas avanzado de desarrollo (R4) y dos colecciones conteniendo transcritos diferenciales de tallo y raíz, estas dos últimas arrestadas con ADNc de hoja.

Total de ESTs	Flor R1 (1)	Flor R4 (2)	Hoja (3)	Tallo (4)	Raíz (5)	Total
Aislados	504	384	268	400	115	1671
Secuenciados	261	269	159	312	72	1073
Analizados	244	222	128	280	45	919
Secuencias diferenciales entre (2) y (3) ^{(a), (b)}		209	80			
Secuencias no redundantes	40	140	63	82	4	329
Porcentaje de secuencias no redundantes ^(b)	16,4	63,0	49,2	28,9	8,8	35,8
Secuencias diferenciales ^(c)	31	131	42	81	4	289
ORF promedio ^(c)	269	220	257	265	225	247
Tamaño de inserto promedio (pb) ^(b)	495	365	443	463	370	441

Tabla 6 – Número de secuencias detectadas exclusivamente en la clonoteca bajo análisis.

^(a) Secuencias detectadas exclusivamente en colecciones de ADNc de flor u hoja

^(b) Número final sobre secuencias analizadas. ^(c) Secuencias detectadas exclusivamente en la colección de ADNc indicada sobre el total de secuencias no redundantes.

a) Clonoteca de ADNc órgano-específico de flor (estadio R1)

La colección de ADNc diferencial de flor en estadio R1 fue sustraída con una población de ARNm del mismo órgano pero en estadio R2. Esta estrategia tuvo como objetivo el aislamiento de transcritos relacionados con el desarrollo de botón floral. La misma constó de 244 clones agrupados en 40 *contigs* (unigenes), clasificados en orden alfabético ascendente, siendo EF101 (abreviatura correspondiente a *early flower*) la primer secuencia correspondiente a la colección, y EF701, la última. En la Tabla 6 se observa el número total de secuencias aisladas, analizadas, diferenciales, tamaño promedio de inserto y de ORF para esta colección, mientras que en la Figura 7 se grafica el análisis de clasificación funcional de dichas secuencias que se expresan o “ESTs” específicos para flor en estadio R1 según su similitud con base de datos, utilizando el programa BLASTX perteneciente al algoritmo BLAST (Altschul y col. 1990).

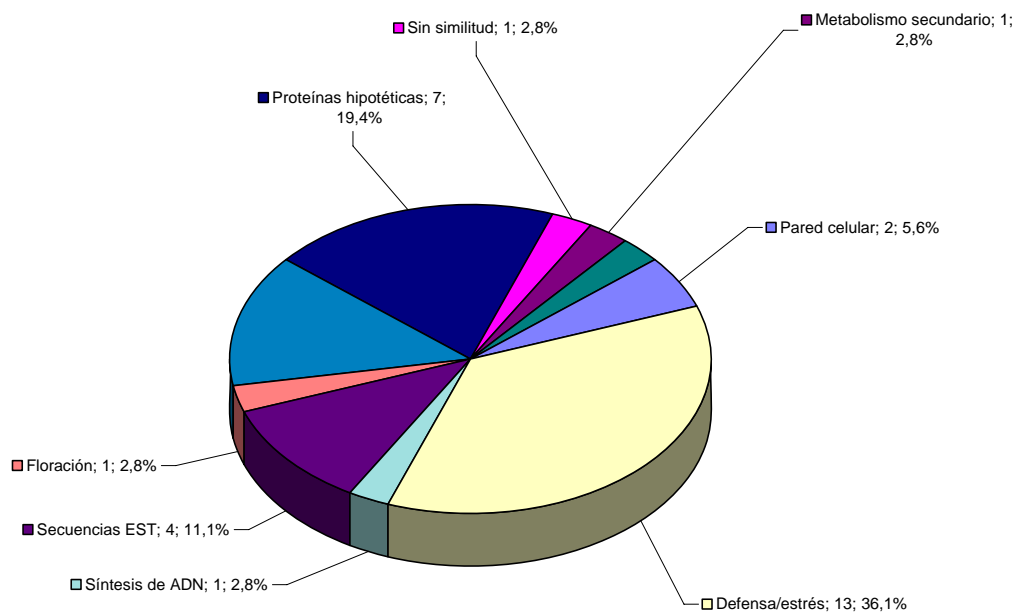


Figura 7 Clasificación funcional de ESTs aislados a partir de la clonoteca diferencial de flor en estadio R1

b) Clonoteca órgano-específica de flor (estadio R4)

Las poblaciones de ARNm de flor en estadio R4 y de hoja fueron utilizadas como fuente de tejido blanco y tejido de referencia, respectivamente, para la construcción de esta clonoteca. La misma constó de 222 clones agrupados en 140 *contigs*, siendo el porcentaje de unigenes de 63%, presentándose como la clonoteca menos redundante obtenida a partir de la presente colección. Las secuencias fueron clasificadas en orden alfabético ascendente, siendo F122 (abreviatura correspondiente a *flor*) la primera secuencia correspondiente a la colección, y F596, la última. En la Tabla 6 pueden observarse los valores paramétricos del total de secuencias analizadas, y en la Figura 8 la clasificación funcional de dichos ESTs.

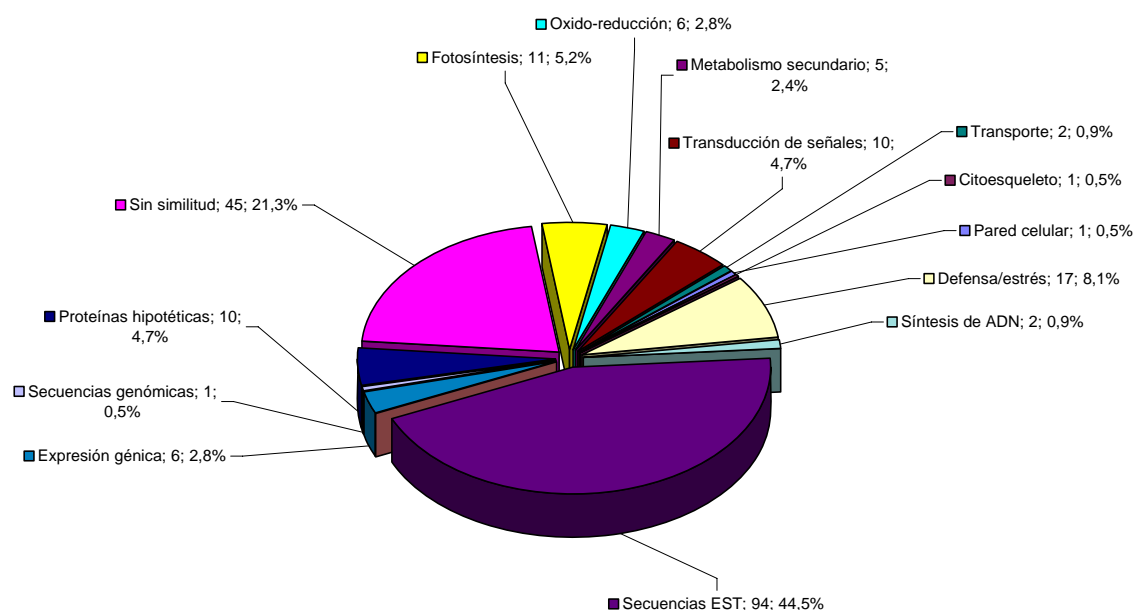


Figura 8 .Clasificación funcional de ESTs aislados a partir de la clonoteca diferencial de flor en estadio R4.

c) Clonoteca de ADNc órgano-específico de hoja

Tal cual se describió en I.1.1.b), el cDNA sintetizado a partir de Poli (A+) ARNm de hoja como fuera descrito anteriormente fue arrestado con mensajeros provenientes de flor en estadio R4 de manera de obtener una colección de transcritos específicos, poco abundantes y poco comunes entre ambos tejidos. El número total de secuencias analizadas fue de 180 agrupadas en 80 *contigs*, siendo el porcentaje de unigenes 49,2%. Las secuencias fueron clasificadas en orden alfabético ascendente, siendo H101 (abreviatura correspondiente a *hoja*) la primera secuencia correspondiente a la colección, y H494, la última. En la Tabla 6 pueden observarse los datos analíticos del total de secuencias analizadas, y en la Figura 9 la clasificación funcional predicha de los ESTs correspondientes.

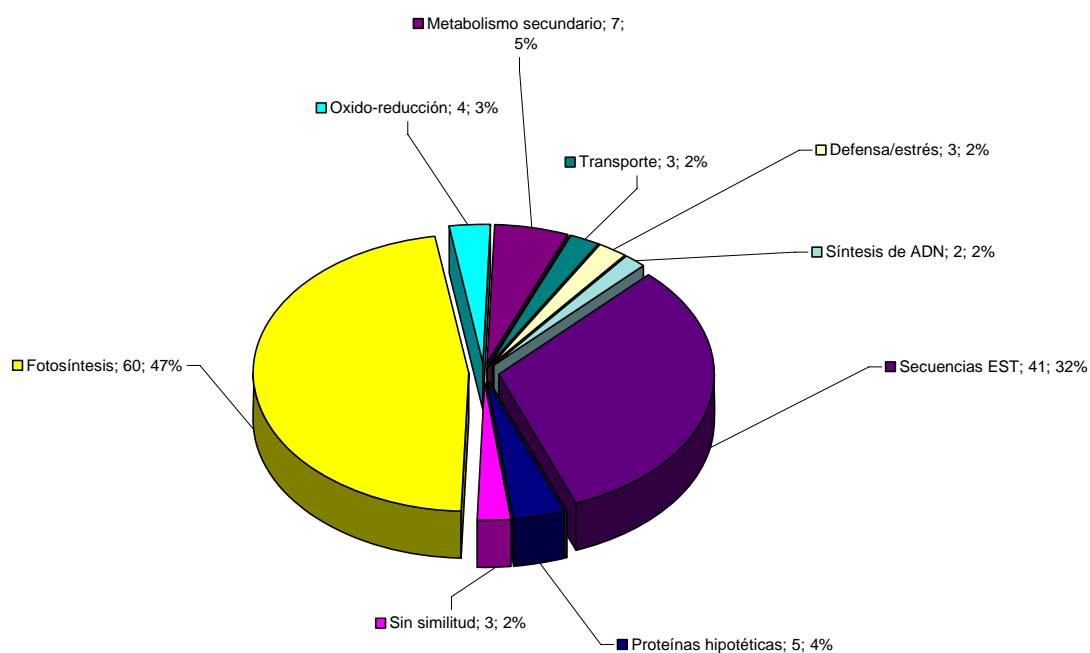


Figura 9: Clasificación funcional de ESTs aislados a partir de la clonoteca diferencial de hoja.

d) Clonoteca de ADNc órgano-específico de tallo

Se construyó una clonoteca diferencial a partir de ADNc sintetizado a partir de ARNm aislado de tallo y arrestado con cDNA provenientes de hoja. El número total de secuencias analizadas fue de 280 agrupadas en 82 *contigs*, siendo el porcentaje de unigenes 28,9%. Las secuencias fueron clasificadas en orden alfabético ascendente, siendo T 103 (abreviatura correspondiente a *tallo*) la primera secuencia correspondiente a la colección, y T 479, la última. En la Tabla 6 pueden observarse los datos analíticos del total de secuencias obtenidas y en la Figura 10 la clasificación funcional predicha de los ESTs correspondientes.

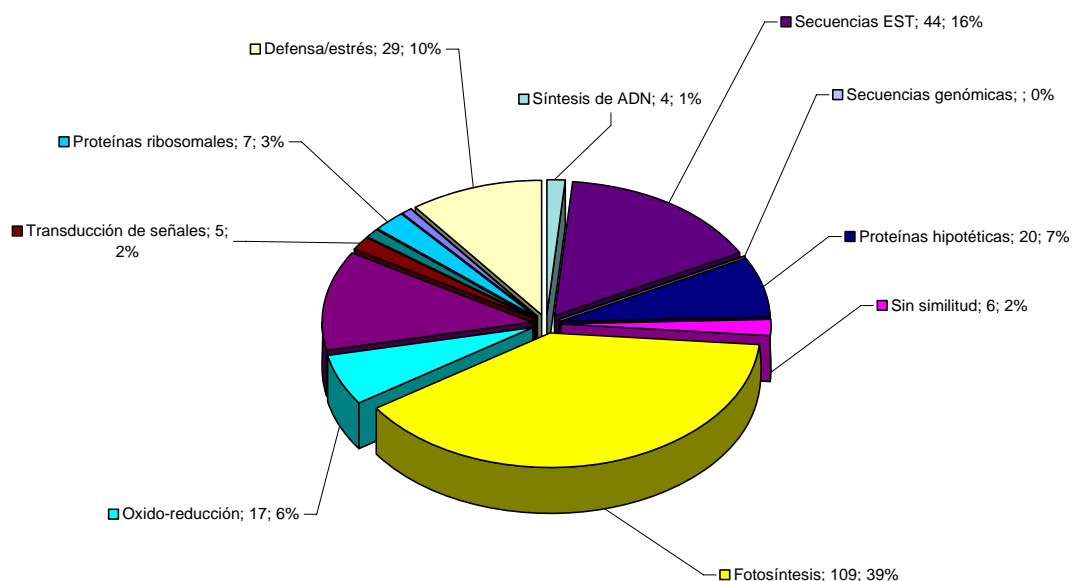


Figura 10: Clasificación funcional de ESTs de la clonoteca diferencial de tallo.

e) Clonoteca de ADNc órgano-específico de raíz

Esta clonoteca no se encuentra altamente representada ya que se obtuvo un número no representativo de clones debido a problemas en la cantidad y calidad de ARNm aislado de raíz que condujo a una clonoteca con alta redundancia de secuencias diferenciales.. El número total de secuencias analizadas fue de 45 agrupadas en 4 *contigs*, siendo el porcentaje de unigenes 8,8%. Las secuencias fueron clasificadas en orden alfabético ascendente, siendo R102 (abreviatura correspondiente a *raíz*) la primera secuencia correspondiente a la colección, y R196, la última. En la Tabla 6 pueden observarse los datos analíticos del total de secuencias obtenidas y en la Figura 11 la clasificación funcional predicha de los ESTs correspondientes



Figura 11: Clasificación funcional de la clonoteca diferencial de raíz.

IV.1.i) Análisis bioinformático de secuencias

La calidad de las secuencias correspondientes a los clones aislados se evaluó teniendo en cuenta la longitud de lectura y el porcentaje de bases indeterminadas. Una vez eliminadas las secuencias flanqueantes del vector los ESTs se ordenaron en una base de datos relacional sobre una plataforma *Windows*[®], según se describe en el ítem *II.1.3*. Las secuencias se compararon entre sí a través de un algoritmo llamado “*Blast local*” para definir las unigenes y a su vez se compararon contra bases de datos internacionales de nucleótidos y proteínas usando los algoritmos BLASTN y BLASTX (Atschul y col. 1990). Los parámetros utilizados para los análisis con BLAST fueron aquellos que los portales de acceso utilizan por defecto, excepto el valor E (*E value*) que fue corregido al valor de ($E=10^{-3}$). En base a la probabilidad de similitud (homología) arrojada por estos programas se asignó una función génica probable a las secuencias analizadas y se definieron grupos que comparten similitud bajo una designación común.

En la Tabla 6 se observa el número total de secuencias aisladas, analizadas, diferenciales, tamaño promedio de inserto y de ORF, por clonoteca de ADNc. La misma expone que un total de 1073 clones no direccionales fueron seleccionados al azar a partir de la secuenciación de las diferentes clonotecas de ADNc, luego de la remoción de las secuencias de calidad pobre y secuencias con similitud a ARN ribosomal. De esta manera, se generaron 919 secuencias legibles, que fueron posteriormente editadas y anotadas. Las secuencias 5' y 3' se encontraron representadas de manera igualitaria en la base de datos de EST.

IV.1.ii) Detección de unigenes

El análisis de redundancia de secuencias fue llevado a cabo por comparación de las mismas utilizando BLASTN a través del programa BioEdit® (Tom Hall, <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Las 919 secuencias generadas en este trabajo fueron archivadas en formato FASTA y cargadas como una base de datos local para ejecutar el programa BLASTN dentro de BioEdit®. Los alineamientos entre ellas fueron analizados y evaluados de manera manual, encontrándose 322 unigenes. Posteriormente, y a modo de automatizar el análisis bioinformático y chequear los unigenes detectados a partir del banco total con mayor precisión, las mismas secuencias fueron comparadas en el servidor local (Inta 01, BioAxioma S.R.L.) desarrollado a partir de este trabajo por Lew y col., 2002. El desarrollo de este programa y la exploración de su interfaz para análisis de secuencias se describen en el Capítulo II de esta sección. Dicha herramienta se aplicó no sólo para estandarizar y automatizar el proceso del análisis y almacenamiento de las secuencias, sino también para detectar secuencias de vector o artefactos remanentes resultantes de la comparación de ellas con bases de datos públicas. El número final obtenido con esta plataforma fue de 318 secuencias que representan unigenes para toda la colección, versus 322 que fue el detectado con el uso de BioEdit®.

IV.1.iii) Búsqueda de similitud en bases de datos públicas

El programa Advanced BLAST (Altschul y col., 1990) fue utilizado para la búsqueda de similitud de secuencias en bases de datos de proteínas. Los 10 primeros hits fueron anotados de acuerdo con su función putativa según BLASTX. Las secuencias que regresaban con similitud no significativa contra la base de datos de proteínas se compararon de manera local utilizando BLASTN contra grupos de secuencias de plantas del tipo STS y EST. Las bases de datos conteniendo aquellas fueron descargadas de bases de datos públicas utilizando “Standalone WWW BLAST Server” del Centro Nacional de Información Biotecnológica (NCBI; ftp://ftp.ncbi.nih.gov/blast/server/current_release). Una vez procesadas, las secuencias fueron exportadas a un archivo en formato FASTA y ordenadas a través de Biopipeline® (Lew y col., 2002). Esta rutina gráfica muestra una matriz procesando la información en una plataforma de Visual Basic la cual alinea los mejores valores (hits) de secuencias en una matriz de puntaje. Las secuencias que mostraban más del 80% de identidad sobre el total del largo de secuencia fueron consideradas idénticas o potencialmente idénticas y fueron asignadas a un grupo específico. Las secuencias que presentaban una identidad entre el 50% y el 80% fueron analizadas exclusivamente de manera manual, alineándolas entre ellas (BLAST 2 sequences: <http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>, Altschul y col., 1992). El alineamiento de secuencias sobre aquellas con alto grado de similitud fue confirmado a través del uso de programas de alineamiento (ClustalW, Thompson y col., 1994). El ensamblaje de ESTs agrupados se realizó utilizando el programa de ensamblado CAP3 (Huang y col., 1996).

IV.1.iv) Análisis informático de ESTs

Biopipeline® arrojó un resultado de 329 secuencias con un total de 318 unigenes (restándole las 11 secuencias correspondientes a posibles contaminaciones con ARN ribosomal), mientras que CAP3 (utilizando un valor de corte de 95% de identidad y un solapamiento mínimo de 25 bases), detectó un total de 29 secuencias

ensambladas (“*contigs*”) y 249 unigenes (secuencias únicas o “*singletons*”), según se muestra en la Figura 12.

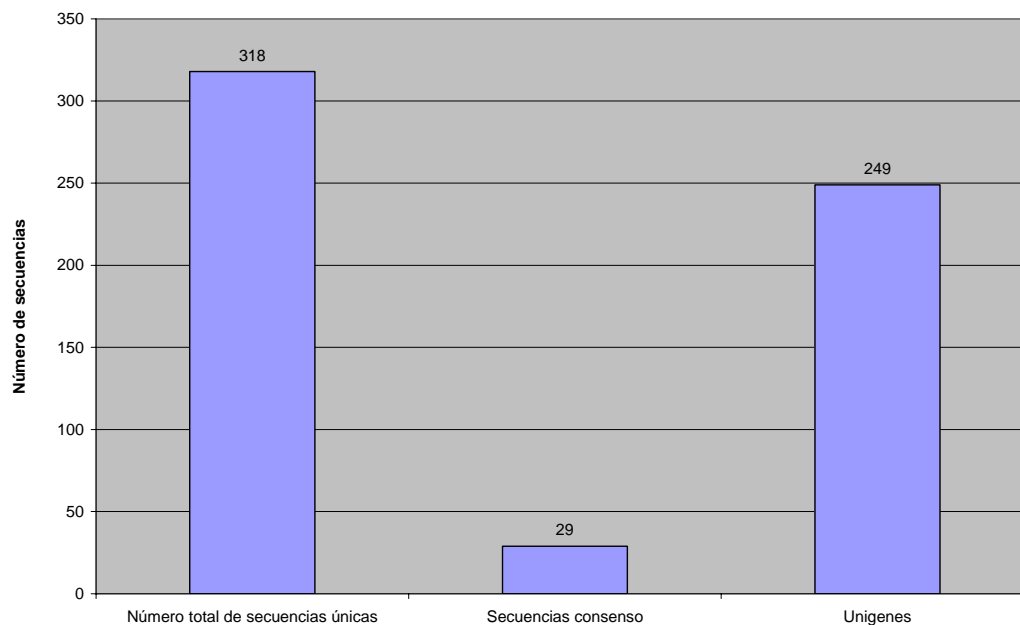


Figura 12: Análisis genómico de ESTs utilizando Cap-3

Por otro lado, la redundancia de secuencias varió a través de las diferentes clonotecas de ADNc (Figura 13).

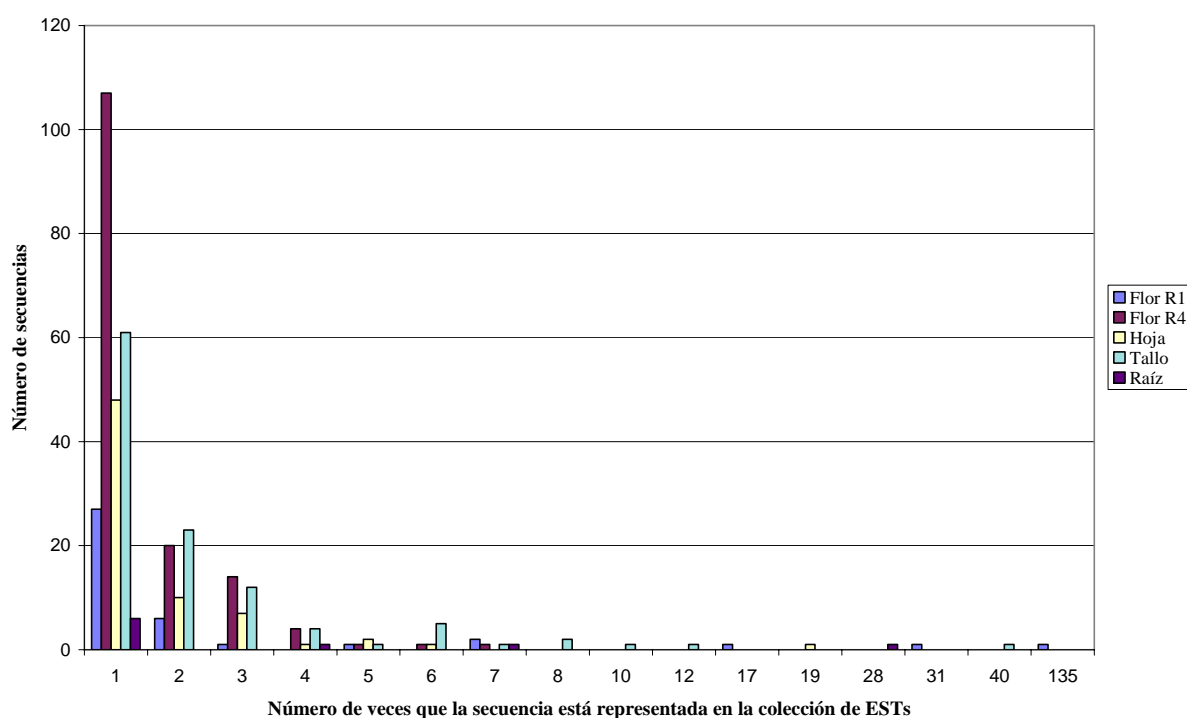


Figura 13 – Frecuencia de clones redundantes entre ESTs de diferentes clonotecas órgano-específicas

La colección menos redundante correspondió al estadio R4 de flor, con un total de 140 unigenes sobre 227 analizadas (62%). En contraste, las clonotecas de ADNc más redundantes fueron la del estadio R1 de flor (13% de unigenes) y la de raíz (7% de unigenes). Las clonotecas de ADNc de hoja y tallo mostraron niveles de redundancia intermedios, con un 49 y 28% de unigenes, respectivamente.

IV.1.2. Análisis de ESTs basados en función predicha

Los ESTs de girasol fueron clasificados en diferentes clases funcionales de acuerdo a su función predicha basada en comparación de secuencias con bases de datos SWISS/Prot/TrEMBL (SWALL). En la Figura 14 se observa la clasificación de los ESTs (Tabla 6) según su homología con bases de datos públicas acorde a la nomenclatura de Gene Ontology y la asistencia del software GOblet (Henning y col., 2003). La nueva anotación incorpora la terminología definida por la iniciativa de Gene Ontology Annotation (GOA; Camon y col., 2003) para la clasificación de

los productos de genes usando una anotación consensuada en las bases de datos que describe de manera consistente que actividad cumple cada proteína. El desarrollo, análisis y caracterización del conjunto de genes aislados en este trabajo así como la anotación de una función definida según GOA, que incluye el número de identificación de GO, la secuencia con mayor similitud evaluada a partir de BLASTX y el número de acceso de la misma en SWALL están descriptos en Fernández y col. (2003), mientras que el listado completo de los ESTs generados en este trabajo, la función putativa otorgada por BLASTX con su puntaje respectivo, el vocabulario asignado por GO y su número de acceso se encuentran como información complementaria del mismo trabajo.

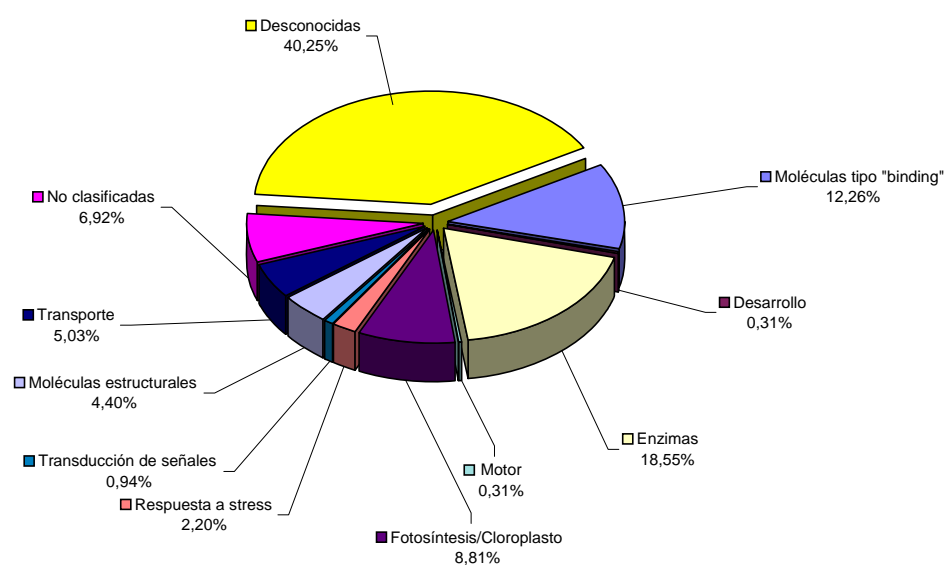


Figura 14: Clasificación según Gene Ontology de secuencias analizadas

Un total de 190 secuencias (60%) sobre 318 no-redundantes no mostró similitud significativa a secuencias conocidas de genes en bases de datos con un nivel de significación basado en el valor E (*E value*) de 10^{-3} y un valor de puntaje mayor a 80. No se encontraron diferencias significativas en el tamaño promedio de inserto entre el grupo de secuencias con similitud a secuencias en bases de datos y el grupo

de secuencias sin similitud a las mismas, lo que indicaría que el tamaño de secuencia aquí reportado no fue limitante para determinar una funcionalidad putativa determinada en GenBank. De las 113.149 entradas correspondientes a la familia de las Compuestas depositadas en bases públicas al momento de realizar las comparaciones de las secuencias generadas en este proyecto, 44.961 correspondían a girasol y el resto a lechuga (Fernández *y col.* 2003), de las cuales sólo 2.061 están anotadas funcionalmente (TIGR Gene Indices, <http://www.tigr.org/tdb/tgi/hagi>). A pesar de esta gran cantidad de información disponible, la comparación de las 318 secuencias no-redundantes generadas en este trabajo contra las 44.961 secuencias de girasol y lechuga (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) mostraron que 197 (59,9%) no tenían similitud significativa con ESTs de girasol previamente reportados, mientras que 228 (69,3%) tampoco lo hacían con ESTs de lechuga. Por otro lado también se realizó un estudio comparativo de los resultados de BLASTX de manera de determinar si los 197 nuevos genes putativos se encontraban representados en GenBank por secuencias previas de girasol de diferentes regiones, ya que los ESTs de este trabajo derivaron de ARN poliadenilado y, por lo tanto, enriquecido en secuencias provenientes del extremo 3', mientras que los ESTs depositados en CGI (Compositae Genome Initiative: <http://compositdb.ucdavis.edu>) se encuentran enriquecidos en la región 5'. Este estudio demostró que la mayoría de los ESTs, si bien comparten la anotación, no comparten la identidad a nivel de nucleótido, lo que significaría que muchas de ellas podrían ser variantes de familia de genes. La clase denominada como "no clasificada" corresponde a secuencias que mostraban similitud significativa con secuencias en SWALL utilizando BLASTX pero para las que no se encontraron un vocabulario GO asociado. Muchas de estas secuencias corresponden a proteínas hipotéticas con función desconocida. La abundancia relativa de categorías de EST varió de acuerdo a la clonoteca analizada (Figura 15). Los ESTs que mostraban similitud no significativa (clase "desconocida") representaron 56% de las secuencias analizadas en la clonoteca de ADNc de flor en estadio R4, mientras que esta categoría fue considerablemente inferior en otras clonotecas, variando del 16 al 47%.

Actualmente, la familia de las Compuestas se encuentra representada en GenBank por 466.639 entradas (24/11/2006), de las cuales 93.793 corresponden a ESTs de girasol cultivado, y el resto a *Helianthus argophyllus*, *Helianthus paradoxus*, *Helianthus exilis*, *Helianthus petiolaris* (girasol silvestre, *Taraxacum officinale*) y lechuga (*Lactuca sativa*) y otras especies del género *Lactuca*, a saber: *Lactuca serriola*, *Lactuca saligna*, *Lactuca virosa* y *Lactuca perennis*: (http://cgpdb.ucdavis.edu/database/php_my_admin/php_my_admin.php). En la Tabla 7 podemos observar que en la actualidad, aún un 14,5% del total del banco local de ESTs son secuencias nuevas para girasol.

Número de ESTs de Compuestas en GenBank						
<i>(http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)</i>						
Especie	Secuencias dbEST	Singletons	Contigs	Total	Puntaje < 80 bits	Porcentaje (sobre 318)
<i>Helianthus annuus</i>	93.793	22.584	10.662	33.246	46	14,47%
<i>Helianthus argophyllus</i>	34.656	5.224	1.018	6.242	125	39,31%
<i>Helianthus exilis</i>	32.827	15.045	5.714	20.759	92	28,93%
<i>Helianthus paradoxus</i>	10.340	4.478	941	5.419	125	39,31%
<i>Helianthus petiolaris</i>	27.484	9.466	4.293	13.759	122	38,36%
<i>Lactuca perennis</i>	29.125	7.420	4.689	12.109	185	58,18%
<i>Lactuca saligna</i>	30.710	7.259	5.161	12.420	180	56,60%
<i>Lactuca sativa</i>	80.849	16.210	11.470	27.680	121	38,05%
<i>Lactuca serriola</i>	55.491	13.823	7.439	21.262	151	47,48%
<i>Lactuca virosa</i>	30.068	8.007	5.098	13.105	178	55,97%
<i>Taraxacum officinale</i>	41.296	10.741	6.283	17.024	176	55,35%

Tabla 7: Número de secuencias de tipo ESTs depositadas en GenBank al 24/11/2006

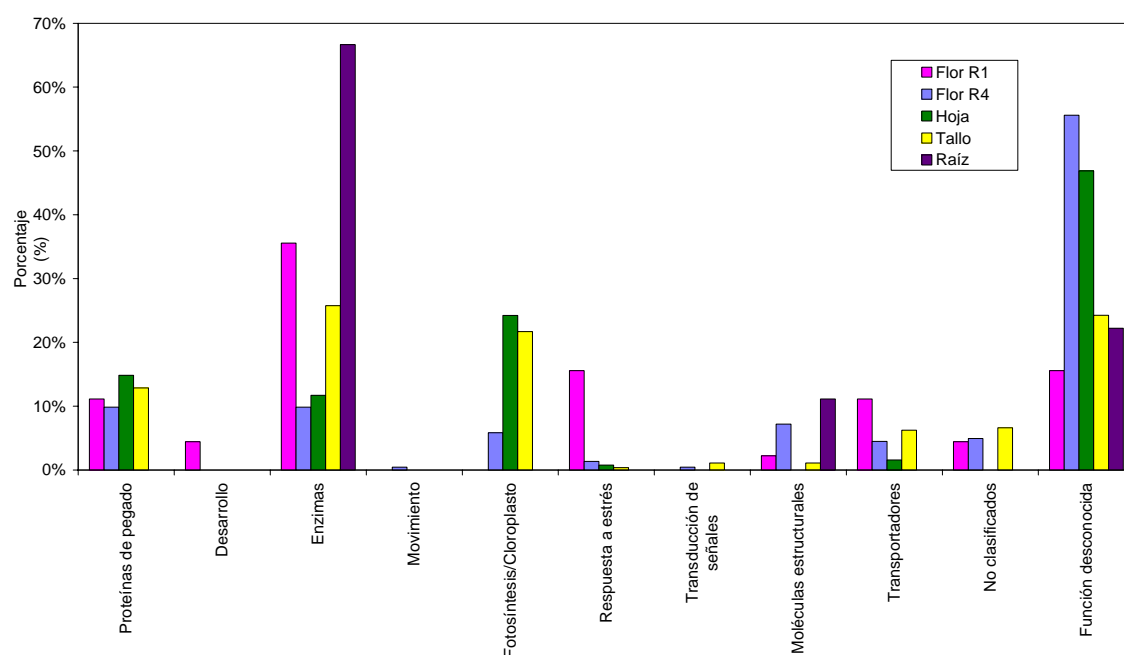


Figura 15: Comparación de ESTs clasificados por función predicha entre las cuatro clonotecas órgano-específicas de ADNc. La clasificación funcional de las secuencias fue realizada según lo descrito en la Figura 14. El porcentaje de ESTs incluidos en cada clase funcional fue comparado entre las cuatro clonotecas de ADNc.

La categoría "proteínas estructurales" y "motoras" así como también secuencias relacionadas con crecimiento celular y metabolismo, incluidas en este trabajo bajo la terminología GO de "enzima", mostraron un bajo nivel de representatividad comparado con las mismas categorías obtenidas de clonotecas de ADNc no-normalizadas (Carson y Botha, 2000), (The Arabidopsis Genome Initiative, AGI: <http://pgec-genome.ars.usda.gov>). Estos resultados sugieren que el paso de normalización que tuvo lugar en la construcción de las clonotecas de ADNc fue eficiente en la reducción del nivel de expresión de transcritos super abundantes representados en los diferentes tejidos analizados. Un estudio similar sobre representatividad de ESTs sobre la categoría funcional de metabolismo celular fue reportado en otras clonotecas de ADNc normalizadas (Ali y Taylor, 2002).

La Tabla 8 incluye la lista completa de la colección anotada según la terminología GO y con el número de acceso otorgado por GenBank., mientras que la Tabla 9 presenta la colección de 33 ESTs no-redundantes relacionados con respuestas a

defensa y *stress*, seleccionados como genes candidato putativos para estudios de expresión.

Tabla 8 -Lista de clones no redundantes con número de acceso de la secuencia generada y su correspondiente definición de identificación GO, similitud BLASTX y número de acceso de la secuencia resultado

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Binding (GO:0005488)					
BU671784	GO:0005524	ATP binding	Q40073	Ribulose biphosphate carboxylase/oxygenaseactivase	6,00E-28
BU671785	GO:0005524	ATP binding	Q9FUI0	Ribulose biphosphate carboxylase/oxygenase activase	7,00E-35
BU671789	GO:0005524	ATP binding	P10896	Ribulose biphosphate carboxylase	4.9e-20
BU671795	GO:0005524	Binding	Q40281	Ribulose biphosphate carboxylase	4,00E-61
BU671806	GO:0005524	ATP binding	Q8LKP8	Heat shock protein 70	2,00E-44
BU671808	GO:0005524	ATP binding	Q41062	Preprotein translocase secA subunit	1,00E-104
BU671809	GO:0005509	Calcium ion binding	Q8LCT9	Hypothetical protein	4.0e-14
BU671811	GO:0003746	Translation elongation factor activity	Q9SZD6	Hypothetical protein (Elongation factor Ts) (EF-Ts)	5.9e-18
BU671817	GO:0003677	DNA binding	Q9LYB9	Hypothetical 21.2 kDa protein.	4,00E-44
BU671832	GO:0003773	Heat shock protein	Q02028	Stromal 70 kDa heat shock-related protein	3,00E-27
BU671833	GO:0003779	Actin binding	Q9FVI2	Actin-depolymeizing factor 1 (ADF 1)	3,00E-27
BU671836	GO:0005524	ATP binding	O81076	Putative AAA-type ATPase	8,00E-21
BU671845	GO:0000287	Magnesium binding	Q93WE2	Magnesium chelatase subunit	4,00E-36
BU671851	GO:0003746	Translation elongation factor	O23755	Elongation factor 2 (EF-2).	1,00E-35
BU671866	GO:0009058	Biosynthesis	Q39049	Magnesium chelatase subunit.	4E-16
BU671867	GO:0030145	Magnesium binding	Q943W1	Oxygen-evolving enhancer protein 1	2,00E-92
BU671870	GO:0005524	ATP binding	Q9FHA9	RuBisCO subunit binding-protein beta subunit precursor, chaperonin	1,30E-09
BU671873	GO:0003746	Translation elongation factor	P17786	Elongation factor 1-alpha (EF-1-alpha).	1,00E-100
BU671885	GO:0003925	Small monomeric	O22413	ADP-ribosylation factor 1.	1,00E-33

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Binding (GO:0005488)					
		GTPase			
BU671890	GO:0003723	RNA binding	Q8LJP5	S-like RNase. Cell division cycle protein 48	3,00E-26
BU671897	GO:0004002	Adenosinetriphosphatase	P54774	homolog. S-adenosylmethionine synthetase	3,00E-71
BU671900	GO:0000287	Magnesium binding	P43282	3 Dicarboxylate/tricarboxylate	2,00E-10
BU671917	GO:0005488	Binding	Q8SF02	carrier.	2,00E-09
BU671923	GO:0003723	Binding	O81363	40S ribosomal protein S4	1,00E-78
BU671930	GO:0005488	Binding	O49875	Adenine nucleotide translocato	3,00E-33
BU671932	GO:0005514	Calcium ion storage	Q40401	Calreticulin precursor.	6,00E-22
BU671940	GO:0003677	DNA-binding	Q9LEW2	Nucleoid DNA-binding protein	5,00E-19
BU671944	GO:0005509	Calcium ion binding	O49301	T26J12.7 protein.	7,00E-13
BU671999	GO:0003743	Heavy metal binding	O82089	Copper homeostasis factor	3,00E-17
		Hydrogen-transporting			
BU672006	GO:0003936	two-sector ATPase	Q34691	F1 ATPase alpha subunit.	4,00E-20
		Translation initiation			
BU672029	GO:0003743	factor B04	QHLK55	Translation initiation factor	5,00E-52
				Peroxisomal Ca-dep solute	
BU672033	GO:0005488	Calcium ion binding	Q9LY28	carrier-like protein	1,00E-33
		Translation initiation			
BU672037	GO:0005509	factor	P41382	Eukariotic initiation factor 4A-10	8,00E-31
BU672042	GO:0005488	Binding	O04610	Putative carrier protein	1,00E-15
		Hydrogen-transporting			
BU672072	GO:0003936	two-sector ATPase	Q9SSS9	Putative H ⁺ -transporting ATP synthase	2,00E-15
		Hydrogen-transporting			
BU672086	GO:0003936	two-sector ATPase	Q9SCC3	ATP synthase beta subunit	1,00E-89
		ATP aynthase gamma			
BU672088	GO:0003936	Caín	P29790	Hydrogen transporting two sector ATPase	1,00E-58
		Hydrogen-transporting			
BU672096	GO:0003936	two-sector ATPase	P29790	ATP synthase gamma chain, chloroplast precursor	2,00E-41
				Mitochondrial half-ABC	
BU672099	GO:0000166	Nucleotide binding	Q9LF78	transporter	1E-41
Transporter (GO:0005215)					
		ATP binding cassette			
BU671850	GO:0004009	(ABC) transporter	Q9ARL6	Putative ABC transporter	4,00E-32
BU671860	GO:0006118	Electron transport	Q9T011	Hypothetical 28.3 kDa protein	1,00E-44
				Triose phosphate/phosphate translocator, chloroplast	
BU671863	GO:0005215	Transporter	P49131	precursor (CTPT).	1,00E-53
BU671880	GO:0008565	Protein transporter	Q93YS8	Hypothetical protein.	3,00E-16
BU671884	GO:0005215	Transporter	Q9M8W5	T6K12.29 protein	3,00E-16

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Transporter (GO:0005215)					
BU671887	GO:0005215	Transporter	Q9ZR68	Aquaporin 1 34 kDa outer mitochondrial	2,00E-84
BU671899	GO:0015288	Porin	P42055	membrane protein porin Polygalacturonase inhibitor	3,00E-38
BU671906	GO:0005489	Electron transport	Q94L58	protein	3,00E-39
BU671975	GO:0005279	Amino acid-polyamine transporter	Q8RZP7	Putative amino acid transport protein. Chloroplast drought-induced	6,00E-08
BU671977	GO:0005489	Electron transporter	O04002	stress protein of 32kDa	6,00E-41
BU671979	GO:0006118	Electron transporter	Q39191	Nitrite reductase Putative thioredoxin-like U5	3,00E-61
BU672024	GO:0005489	Electron transporter	Q8W376	small ribonucleoprotein Chloroplast drought-induced	5,00E-16
BU672102	GO:0005489	Electron transporter	O04002	stress protein of 32kDa Putative component of	2,00E-56
BU671796	GO:0006118	Electron transport	Q9ZR03	cytochrome B6-F complex.	2,00E-69
BU671812	GO:0005489	Electron transporter	Q8W4M1	Putative thioredoxin reductase.	5E-19
BU671950	GO:0006118	Electron transport	Q9ZS34	Geranylgeranyl reductase.	2,00E-08
Enzyme (GO:0003824)					
		Ribulose-bisphosphate			
BU671786	GO:0016984	carboxylase activit activity	P08705	Ribulose-bisphosphate carboxylase activity	2.50E-10
BU671794	GO:0000238	Phosphoethanolamine N methytransferase	Q944HO	Putative phosphoethanolamine N methytrans	2,00E-26
BU671797	GO:0016702	Oxidoreductase, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	Q8LP14	Nine-cis-epoxycarotenoid dioxygenase4.	9,00E-52
BU671801	GO:0004332	Fructose biphosphate aldolase	Q9SXX5	Plastidic aldolase Cyclophilin (EC 5.2.1.8)	2,00E-91
BU671803	GO:0042027	Cyclophilin-type peptidy-prolyl cis-trans isomerase	Q9M530	(Peptidyl-prolyl cis-trans isomerase) (PPIase)	3,00E-18
BU671805	GO:0008246	Electron transfer flavoprotein	Q39640	Glycolate oxidase	2E-94
BU671807	GO:0004553	Hydrolase	Q9M473	Putative -beta 1,3-glucanase	5,00E-38
BU671815	GO:0004497	Monoxygenase	Q9XE94	Geranylgeranyl hydrogenase	5,00E-63

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Enzyme (GO:0003824)					
BU671823	GO:0042132	Fructose-bisphosphatase	Q940F8	Sedoheptulose-bisphosphatase. S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) (AdoMetDC) (SamDC) (Induced stolen tip protein TUB13) [Contains: S-adenosylmethionine decarboxylase alpha chain; S-adenosylmethionine decarboxylase beta chain].	1,00E-46
BU671824	GO:0004014	Adenosylmethionine decarboxylase	Q04694	Unknown protein (EC 3.3.2.3)	3,00E-46
BU671825	GO:0004301	Epoxide hydrolase	Q94K11	(Epoxide hydrolase).	2,00E-71
BU671831	GO:0004365	Glyceraldehyde 3-phosphate dehydrogenase (phosphorylating) activity	Q9ZIC0	Glyceraldehyde phosphate dehydrogenase	0.012
BU671834	GO:0004372	Glycine hydroxymethyltransferase	P34899	Serine hydroxymethyltransferase, mitochondrial precursor (EC 2.1.2.1) (Serine methylase) (Glycine hydroxymethyltransferase) (SHMT).	1,00E-137
BU671835	GO:0004374	Glycine cleavage system	P49361	Glycine dehydrogenase [decarboxylating] A, mitochondrial precursor (EC 1.4.4.2) (Glycine decarboxylase A) (Glycine cleavage system P-protein A).	9,00E-48
BU671838	GO:0004096	Catalase	P45739	Catalase (EC 1.11.1.6).	2,00E-46
BU671841	GO:0004332	Fructose biphosphata aldolase	P93565	homologous to plastidic aldolases	3,00E-21
BU671843	GO:0004356	Glutamate ammonia ligase	O04961	Chloroplastic glutamine synthase Phosphoribulokinase, chloroplast precursor (EC 2.7.1.19) (Phosphopentokinase)	7,00E-56
BU671849	GO:0004849	Uridine kinase	P26302	(PRKASE) (PRK).	5,00E-24
BU671853	GO:0016984	Ribulose-bisphosphate carboxylase activity	Q9AUJ6	Ribulose-1,5-bisphosphate carboxylase small subunit R1 precursor	7.5e-11
BU671857	GO:0008121	Ubiquinol-cytochrome c reductase	Q9FYB6	Rieske FeS protein precursor	1,00E-18

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Enzyme (GO:0003824)					
BU671864	GO:0004332	Fructose-bisphosphate aldolase	Q9SXX5	Plastidic aldolase	1,00E-54
BU671869	GO:0004802	Transketolase	Q42676	Transketolase, chloroplast (EC 2.2.1.1) (TK) (Fragment). Aminomethyltransferase, mitochondrial precursor (EC 2.1.2.10) (Glycine cleavage system T protein) (GCVT).	8,00E-84
BU671872	GO:0004047	Aminomethyltransferase	P49363	system T protein) (GCVT).	2,00E-54
BU671874	GO:0004332	Fructose-bisphosphate aldolase	O22486	Aldolase.	0.00011
BU671875	GO:0016740	Transferase	Q8LAD3	Glu-tRNA(Gln) amidotransferase subunit C. Serine hydroxymethyltransferase 2, mitochondrial precursor (EC 2.1.2.1) (Serine methylase) (Glycine hydroxymethyltransferase)	3.4e-19
BU671879	GO:0004372	Glycine hydroxymethyltransferase	P49358	(SHMT).	7,00E-04
BU671881	GO:0004497	Monoxygenase	Q9SRQ1	Putative cytochrome P450	1,00E-12
BU671886	GO:0004299	Proteasome endopeptidase	O23708	Proteasome subunit alpha type 2	2,00E-17
BU671888	GO:0004553	Hydrolase, hydrolyzing O-glycosyl compounds	Q9M453	Putative beta-1,3-glucanase.	3,00E-58
BU671892	GO:0016301	Kinase	Q9LV28	G binding protein	3,00E-66
BU671894	GO:0016844	Strictosidine synthase	Q9M1B4	Hypothetical 45.6 kDa protein.	2,00E-68
BU671903	GO:0003755	Peptidyl-prolyl cis-trans isomerase	Q9M530	Cyclophilin	3,00E-18
BU671904	GO:0004564	Beta-fructofuranosidase	O81985	1,2-beta-fructan 1F-fructosyltransferase	2,00E-51
BU671907	GO:0008415	Acyltransferase	O24625	Anther-specific protein. S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine adenosyltransferase) (AdoMet synthetase).	6,00E-87
BU671913	GO:0004478	Methionine adenosyltransferase	Q9FUZ1	synthetase).	1,00E-95
BU671924	GO:0004197	Cysteine-type endopeptidase	Q8VWS1	Putative cysteine proteinase	4,00E-42
BU671928	GO:0003755	Peptidyl-prolyl cis-trans isomerase	Q8L5T1	Peptidylprolyl isomerase (Cyclophilin)	3,00E-18
BU671929	GO:0004222	Metalloendopeptidase	O22941	Putative zinc protease	1,00E-61
BU671931	GO:0004672	Protein kinase	P93017	Auxin-repressed protein	0.030

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Enzyme (GO:0003824)					
				ISOLOG (Dormancy-associated protein homolog) (At2g33830/T1B8.13).	
				Putative cytosolic 6-phosphogluconate dehydrogenase (EC 1.1.1.44) (6-phosphogluconate dehydrogenase, decarboxylating).	8,00E-15
BU671943	GO:0004616	Phosphogluconate dehydrogenase (decarboxylating)	O81238		
BU671960	GO:0004766	Spermidine synthase	48658	Spermidine synthase 1 (EC 2.5.1.16)	1,00E-21
BU671968	GO:0042132	Fructosa-bisphosphatase	Q9XF82	Fructose-1,6-bisphosphatase (EC 3.1.3.11).	4,00E-18
BU671972	GO:0008168	Methyltransferase	Q9LW67	Ankyrin-like protein Serine hydroxymethyltransferase (EC 2.1.2.1) (Serine methylase) (Glycine hydroxymethyltransferase)	7,00E-33
BU671976	GO:0004372	Glycine hydroxymethyltransferase	O23254	(SHMT).	2,00E-31
BU671989	GO:0004601	Peroxidase	O64970	Cationic peroxidase 2	1,00E-40
		Phosphogluconate dehydrogenase		6-phosphogluconate dehydrogenase, putative	2,00E-19
BU672005	GO:0004616	Protein-methionine-S-oxide reductase	Q9FWA2		
BU672021	GO:0008113		Q9FF06	Methionine sulfoxide reductase	2,00E-15
		Carbonate dehydratase		Carbonic anhydrase (EC 4.2.1.1) (Carbonate dehydratase).	3,00E-89
BU672057	GO:0004089	1-deoxy-D-xylulose-5-phosphate reductoisomerase	P46511		
BU672065	GO:0030604	Inositol-3-phosphate synthase	Q8L6C5	1-deoxy-D-xylulose 5-phosphate reductoisomerase.	2,00E-18
BU672069	GO:0004512		Q9LW96	Inositol-3-phosphate synthase Putative alanine	1,00E-42
BU672071	GO:0004021	Alanine aminotransferase	Q8LFE9	aminotransferase	7,00E-38
		glyceraldehyde 3-phosphate dehydrogenase		Glyceraldehyde 3-phosphate dehydrogenase B, chloroplast precursor	1,00E-69
BU672076	GO:0004365		P25857		
BU672082	GO:0004185	Serine carboxypeptidase	Q8W2X4	Hypothetical protein	7,00E-13
		Glycine hydroxymethyltransferase		Serine hydroxymethyltransferase 1, mitochondrial precursor	5,00E-71
BU672085	GO:0004372		P49357		
		Fructose-bisphosphatase		Fructose-1,6-bisphosphatase, cytosolic	1,00E-80
BU672090	GO:0042132		P46276		
BU672094	GO:0000234	Phosphoethanolamine N-	Q8LJ10	Putative phosphoethanolamine	1,00E-39

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Enzyme (GO:0003824)					
		methyltransferase		methyltransferase	
BU672105	GO:0008137	NADH dehydrogenase	P58419	NAD(P)H-quinone oxidoreductase chain 4	3,00E-25
BU672106	GO:0004124	Cysteine synthase	Q8Y0X6	Probable cysteine synthase B (CSASE B) protein (EC 4.2.99.8).	9,00E-21
BU672109	GO:0004365	Glyceraldehyde 3-phosphate dehydrogenase	P09044	Glyceraldehyde 3-phosphate dehydrogenase B	3,00E-24
Structural Molecules					
BU671813	GO:0003735	Structural constituent of ribosome	Q8L6Z4	Hypothetical protein.	5,00E-19
BU671848	GO:0003735	Structural constituent of ribosome	Q9P0N7	Mitochondrial 39S ribosomal protein L51 (MRP-L51).	1,00E-73
BU671920	GO:0003735	Structural constituent of ribosome	P17093	40S ribosomal protein S11.	1,00E-27
BU671926	GO:0003735	Structural constituent of ribosome	P29766	60S ribosomal protein L2 (L8) (Ribosomal protein TL2).	1,00E-42
BU671933	GO:0003735	Structural constituent of ribosome	Q9SS17	40S ribosomal protein S24.	1,00E-17
BU671938	GO:0003735	Structural constituent of ribosome	Q8L8W6	Putative ribosomal protein L39	0,004
BU671962	GO:0005198	Structural molecule	Q9M406	Alpha-tubulin 4 (Fragment)	9,00E-13
BU671992	GO:0003735	Structural constituent of ribosome	Q00508	Ribosomal protein L41	2,00E-21
BU672008	GO:0003735	Structural constituent of ribosome	Q9FKC0	60S ribosomal protein L13a-4	5,00E-39
BU672036	GO:0003735	Structural constituent of ribosome	Q9M5L1	40S ribosomal protein S16	9,00E-29
BU672049	GO:0003735	Structural constituent of ribosome	Q9VA87	60S ribosomal protein L21	1,00E-75
BU672052	GO:0003735	Structural constituent of ribosome	Q8LEQ9	50S ribosomal protein L3	7,00E-24
BU672053	GO:0003735	Structural constituent of ribosome	O65082	60S ribosomal protein L15-2.	4,00E-12
BU672107	GO:0003735	Structural constituent of ribosome	Q9ZRW5	Ribosomal protein S28	2,00E-28
Response to Stress (GO:0006950)					
BU671840	GO:0004096	Catalase	P45739	Catalase	1,00E-95
BU671889	GO:0030145	Manganese binding Immediate	O48999	Germin-like protein 3	2,00E-35
BU671909	GO:0016068	hypersensitivity response	Q93YX9	Lipid transfer protein	1,00E-37
BU671910	GO:0016068	Immediate HR response	Q9M6B8	Lipid transfer protein	2,00E-15

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Response to Stress (GO:0006950)					
BU671955	GO:0009607	Response to biotic stimulus	P13046	Pathogenesis-related protein R major form	3,00E-34
BU672016	GO:0004867	Serine protease inhibitor	Q8LNY0	Protease inhibitor 2	2,00E-13
BU672055	GO:0004869	Cysteine protease inhibitor	Q9MB08	Multicystatin.	2,00E-12
Photosynthesis (GO:0015979) / Chloroplast (GO:0009507)					
BU671783	GO:0009765	Photosynthesis light harvesting	Q9M320	Lhca2 protein.	7,00E-22
BU671790	GO:0009507	Chloroplast	P27524	Chlorophyll A-B binding protein CP24 10A, chloroplast precursor (CAB- 10A) (LHCP).	7E-48
BU671791	GO:0009765	Photosynthesis light harvesting	Q00321	Type I (26 kD) CP29 polypeptide.	5E-89
BU671792	GO:0009538	Photosystem I reaction center	Q40430	PSI-H precursor.	3,00E-21
BU671793	GO:0009522	Photosynthesis light harvesting	Q40716	Precursor of rice 22 kDa protein of photosystem II (PSII-S).	8,00E-63
BU671800	GO:0009765	Photosynthesis light harvesting	Q9SDT2	Chlorophyll a/b-binding protein.	1E-57
BU671810	GO:0009507	Chloroplast	Q9SUI5	Photosystem I reaction center subunit psaK, chloroplast precursor (Photosystem I subunit X) (PSI-K).	6,00E-45
BU671818	GO:0009507	Chloroplast	P51277	Hypothetical 41.5 kDa protein in YCF6-CHLB intergenic region (ORF349).	1,00E-33
BU671822	GO:0009765	Photosynthesis light harvesting	Q9XF88	LHCB4.2 protein (Putative chlorophyll a/b-binding protein).	4,00E-36
BU671828	GO:0009765	Photosynthesis light harvesting	O24561	Chlorophyll a/b-binding protein CP29 precursor	4,00E-30
BU671829	GO:0009765	Photosynthesis light harvesting	Q40247	Light-harvesting chlorophyll a/b-binding protein (LHCP) precursor.	2E-97
BU671837	GO:0009765	Photosynthesis light harvesting	Q39939	Homologous early light induced protein.	1E-20
BU671846	GO:0009507	Chloroplast	Q41932	Oxygen-evolving enhancer protein 3-2, chloroplast precursor (OEE3) (16 kDa subunit of oxygen evolving system of photosystem II) (OEC 16 kDa subunit).	4,00E-49
BU671877	GO:0009507	Chloroplast	P06183	Photosystem II 10 kDa polypeptide, chloroplast precursor (Light inducible tissue-specific ST-LS1 protein).	3E-45
BU671949	GO:0009507	Chloroplast	O49347	Photosystem II core complex proteins psbY, Photosystem I	1,00E-14

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Photosynthesis (GO:0015979) / Chloroplast (GO:0009507)					
BU671963	GO:0009507	Chloroplast	Q9SLQ8	Oxygen-evolving enhancer protein 2, chloroplast precursor (OEE2) (23 kDa subunit of oxygen evolving system of photosystem II) (OEC 23 kDa subunit) (23 kDa thylakoid membrane protein) (OEC23).	2,00E-22
BU671991	GO:0015979	Photosynthesis	P46486	Photosystem I reaction centre subunit III	1,00E-15
BU672044	GO:0015979	Photosynthesis	Q41385	Photosystem I reaction center subunit XI, chloroplast precursor	1,00E-16
BU672054	GO:0009765	Photosynthesis harvesting	P12360	Chlorophyll A-B binding protein 6A	7,00E-29
BU672058	GO:0019685	Photosynthesis, reaction	Q8w3D9	Putative dehydrogenase	3,00E-27
BU672060	GO:0009765	Photosynthesis harvesting	Q9M320	Lhca2 protein	3,00E-20
BU672064	GO:0015979	Photosynthesis	Q9S797	Photosystem I ret V, chloroplast precursor (PSI- G). PSI light-harvesting antenna	2,00E-21
BU672068	GO:0009765	Photosynthesis harvesting	Q9SQL2	chlorophyll a/b-binding protein	9,00E-12
BU672070	GO:0009765	Photosynthesis harvesting	Q9SY97	PSI type III chlorophyll a/b-binding protein.	1,00E-108
BU672078	GO:0009507	Chloroplast	P51277	Hypothetical 41.5 kDa protein in YCF6-CHLB intergenic region	1,00E-16
BU672081	GO:0009765	Photosynthesis harvesting	P27524	Chlorophyll A-B binding protein CP24 10A	9,00E-44
BU672091	GO:0009538	Photosystem I reaction centre subunit III	P46486	photosystem I reaction center	2,00E-42
BU672103	GO:0009765	Photosynthesis harvesting	Q9M320	Lhca2 protein	1,00E-84
Signal transduction (GO:0007165)					
BU671820	GO:0004707	MAP kinase signaling	Q9LUC3	MAP kinase protein (Fragment).	4,00E-31
BU671826	GO:0007242	Intracellular cascade	Q9ZVU6	T5A14.12 protein.	4.3e-16
BU671951	GO:0004722	Protein serine/threonine phosphatase	Q9FSV3	Serine/threonine phosphatase	1,00E-25
Development (GO:0007275)					
BU671901	GO:0007275	Development	Q9XGW1	PINHEAD protein (ZWILLE)	3E-13
Motor (GO:0003774)					
BU671959	GO:0005200	Structural constituent of cytoskeleton	Q93XI7	Actin	6,00E-17
Unclassified					
BU671862	No GO terms were found		P04474	Acidic proline-rich protein PRP33 precursor	3,50E-12
BU671939	No GO terms were found		Q9ZRD4	ATGP4.	6,10E-09
BU671891	No GO terms were found		Q8CM04	Conserved hypothetical protein.	8,10E-21

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Unclassified					
				Emb CAB75797.1	
BU671830	No GO terms were found		Q9FN48	(Hypothetical protein).	5.4e-39
BU671934	No GO terms were found		Q9SX29	F24J5.10 protein (Hypothetical protein)	5,40E-56
BU671969	No GO terms were found		O53022	Fused-ccdB protein	2.1e-05
BU671821	No GO terms were found		Q8W4G5	G5bf protein.	2.7e-65
BU671847	No GO terms were found		Q9FLI7	Genomic DNA, chromosome 5, P1 clone:MIO24	2,30E-25
BU671878	No GO terms were found		Q94H99	Hypothetical 15.0 kDa protein.	1,70E-05
BU671918	No GO terms were found		Q9C5D0	Hypothetical 26.0 kDa protein.	3,70E-05
BU671852	No GO terms were found		Q940B4	Hypothetical 49.0 kDa protein.	9.9e-48
BU671871	No GO terms were found		Q9SZJ4	Hypothetical 77.1 kDa protein.	3,40E-19
BU671819	No GO terms were found		Q84P75	Hypothetical protein.	4,60E-43
BU671957	No GO terms were found		Q8LEU2	Hypothetical protein.	3,90E-14
BU671839	No GO terms were found		Q8LF26	Hypothetical protein.	7,10E-16
BU671914	No GO terms were found		Q8LGG2	Hypothetical protein.	6,50E-17
BU671799	No GO terms were found		Q8CE09	Hypothetical protein.	0.0085
BU671814	No GO terms were found		Q8LDL0	Hypothetical protein.	1.0e-13
BU671922	No GO terms were found		Q8GVZ2	OJ1417_E01.23 protein.	6,20E-07
BU671856	No GO terms were found		Q8GTY4	RuBisCO activase (Fragment).	0.034
BU671859	No GO terms were found		Q9ZVL6	T22H22.19 protein	3,90E-46
				T24D18.18 protein	
				(Hypothetical 34.1 kDa protein).	
BU671865	No GO terms were found		Q9S9M7		3,10E-69
BU671782				No similarity	
BU671787				No similarity	
BU671788				No similarity	
BU671802				No similarity	
BU671804				No similarity	
BU671816				No similarity	
BU671827				No similarity	
BU671844				No similarity	
BU671854				No similarity	
BU671855				No similarity	
BU671858				No similarity	
BU671861				No similarity	
BU671868				No similarity	
BU671902				No similarity	
BU671905				No similarity	
BU671908				No similarity	
BU671912				No similarity	
BU671916				No similarity	

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Unclassified					
BU671919				No similarity	
BU671921				No similarity	
BU671925				No similarity	
BU671927				No similarity	
BU671935				No similarity	
BU671936				No similarity	
BU671937				No similarity	
BU671941				No similarity	
BU671942				No similarity	
BU671945				No similarity	
BU671946				No similarity	
BU671947				No similarity	
BU671948				No similarity	
BU671952				No similarity	
BU671953				No similarity	
BU671954				No similarity	
BU671956				No similarity	
BU671958				No similarity	
BU671961				No similarity	
BU671964				No similarity	
BU671965				No similarity	
BU671966				No similarity	
BU671967				No similarity	
BU671970				No similarity	
BU671971				No similarity	
BU671973				No similarity	
BU671974				No similarity	
BU671978				No similarity	
BU671980				No similarity	
BU671981				No similarity	
BU671982				No similarity	
BU671983				No similarity	
BU671984				No similarity	
BU671985				No similarity	
BU671986				No similarity	
BU671987				No similarity	
BU671988				No similarity	
BU671990				No similarity	
BU671993				No similarity	
BU671994				No similarity	

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Unclassified					
BU671995				No similarity	
BU671996				No similarity	
BU671997				No similarity	
BU671998				No similarity	
BU672000				No similarity	
BU672001				No similarity	
BU672002				No similarity	
BU672003				No similarity	
BU672004				No similarity	
BU672007				No similarity	
BU672009				No similarity	
BU672010				No similarity	
BU672011				No similarity	
BU672012				No similarity	
BU672013				No similarity	
BU672014				No similarity	
BU672015				No similarity	
BU672017				No similarity	
BU672018				No similarity	
BU672019				No similarity	
BU672020				No similarity	
BU672022				No similarity	
BU672023				No similarity	
BU672025				No similarity	
BU672026				No similarity	
BU672027				No similarity	
BU672028				No similarity	
BU672030				No similarity	
BU672031				No similarity	
BU672032				No similarity	
BU672034				No similarity	
BU672035				No similarity	
BU672038				No similarity	
BU672039				No similarity	
BU672040				No similarity	
BU672041				No similarity	
BU672043				No similarity	
BU672045				No similarity	
BU672046				No similarity	
BU672047				No similarity	
BU672048				No similarity	
BU672050				No similarity	

IV. RESULTADOS Y DISCUSIÓN

Número de acceso	Identificación GO	Definición función GO	Número de acceso	BLASTX hit	E value
Unclassified					
BU672051				No similarity	
BU672056				No similarity	
BU672059				No similarity	
BU672061				No similarity	
BU672062				No similarity	
BU672063				No similarity	
BU672066				No similarity	
BU672067				No similarity	
BU672073				No similarity	
BU672074				No similarity	
BU672075				No similarity	
BU672077				No similarity	
BU672079				No similarity	
BU672080				No similarity	
BU672083				No similarity	
BU672084				No similarity	
BU672087				No similarity	
BU672089				No similarity	
BU672092				No similarity	
BU672093				No similarity	
BU672095				No similarity	
BU672097				No similarity	
BU672098				No similarity	
BU672100				No similarity	
BU672101				No similarity	
BU672104				No similarity	
BU672108				No similarity	
BU672110				No similarity	

Tabla 9 - ESTs relacionadas con respuesta a estrés biótico y/o abiótico
^a Número de Acceso GenBank otorgados a los ESTs de este trabajo
^b Número de identificación e identificación funcional acorde a la anotación GO
^c Número de acceso GenBank de la secuencia con mayor similitud
^d La búsqueda de similitud fue hecha usando el programa BLASTX
La función putativa de los ESTs fue asignada acorde al valor más alto de similitud (“score”) en GenBank.

AN ^a	GO ID ^b	Funcionalidad GO ^b	AN ^c	BLASTX hit ^d	E value
BU671794	GO:000238	Phosphoethanolamine N methyltransferase	Q944HO	Putative phosphoethanolamine N methyltransferase	2.00E-26
BU671801	GO:0004332	Fructose biphosphate aldolase	Q9SXX5	Plastidic aldolase	2.00E-91
BU671803	GO:0042027	Cyclophilin-type peptidyl-prolyl cis-trans isomerase	Q9M530	Cyclophilin (EC 5.2.1.8) (Peptidyl-prolyl isomerase) (PPIase) cis-trans	3.00E-18
BU671805	GO:0008246	Electron transfer flavoprotein	Q39640	Glycolate oxidase	2E-94
BU671807	GO:0004553	Hydrolase	Q9M473	Putative -beta 1,3-glucanase	5.00E-38
BU671832	GO:0003773	Heat shock protein	Q02028	Stromal 70 kDa heat shock-related protein	3.00E-27
BU671840	GO:0004096	Catalase	P45739	Catalase	1.00E-95
BU671841	GO:0004332	Fructose biphosphata aldolase	P93565	homologous to plastidic aldolases	3.00E-21
BU671845	GO:0000287	Magnesium binding	Q93WE2	Magnesium chelatase subunit	4.00E-36
BU671864	GO:0004332	Fructose-bisphosphate aldolase	Q9SXX5	Plastidic aldolase	1.00E-54
BU671866	GO:0009058	Biosynthesis	Q39049	Magnesium chelatase subunit.	4E-16
BU671867	GO:0030145	Magnesium binding	Q943W1	Oxygen-evolving enhancer protein	2.00E-92
BU671886	GO:0004299	Proteasome endopeptidase	O23708	Proteasome subunit alpha type 2	2.00E-17
BU671887	GO:0005215	Transporter	Q9ZR68	Aquaporin 1	2.00E-84
BU671888	GO:0004553	Hydrolase, hydrolyzing O- glycosyl compounds	Q9M453	Putative beta-1,3-glucanase.	3.00E-58
BU671889	GO:0030145	Manganese binding	O48999	Germin-like protein 3	2.00E-35
BU671904	GO:0004564	Beta-fructofuranosidase	O81985	1,2-beta-fructan fructosyltransferase 1F-	2.00E-51
BU671906	GO:0005489	Electron transport	Q94L58	Polygalacturonase inhibitor protein	3.00E-39
BU671909	GO:0016068	Immediate hypersensitivity response	Q93YX9	Lipid transfer protein	1.00E-37
BU671910	GO:0016068	Immediate hypersensitivity response	Q9M6B8	Lipid transfer protein	2.00E-15
BU671924	GO:0004197	Cysteine-type endopeptidase	Q8VWS1	Putative cysteine proteinase	4.00E-42
BU671928	GO:0003755	Peptidyl-prolyl cis-trans isomerase	Q8L5T1	Peptidylprolyl isomerase (Cyclophilin)	3.00E-18
BU671929	GO:0004222	Metalloendopeptidase	O22941	Putative zinc protease	1.00E-61
BU671944	GO:0005509	Calcium ion binding	O49301	T26J12.7 protein.	7.00E-13
BU671955	GO:0009607	Response to biotic stimulus	P13046	Pathogenesis-related protein R major form	3.00E-34
BU671960	GO:0004766	Spermidine synthase	48658	Spermidine synthase 1 (EC 2.5.1.16)	1.00E-21
BU671972	GO:0008168	Methyltransferase	Q9LW67	Ankyrin-like protein	7.00E-33
BU671989	GO:0004601	Peroxidase	O64970	Cationic peroxidase 2	1.00E-40
BU671977	GO:0005489	Electron transporter	O04002	Chloroplast drought-induced stress protein of 32kDa	6.00E-41
BU672016	GO:0004867	Serine protease inhibitor	Q8LNY0	Protease inhibitor 2	2.00E-13
BU672055	GO:0004869	Cysteine protease inhibitor	Q9MB08	Multicystatin.	2.00E-12
BU672102	GO:0005489	Electron transporter	O04002	Chloroplast drought-induced stress protein of 32kDa	2.00E-56
BU672106	GO:0004124	Cysteine synthase	Q8Y0X6	Probable cysteine synthase B (CSASE B) protein (EC 4.2.99.8).	9.00E-21

IV.1.3. Discusión

En este capítulo se describió la obtención de un banco local de ESTs a partir de clonotecas de ADNc órgano-específicas mediante el uso de la técnica de SSH (Diatchenko y col, 1996, 1999), lo que permite la fácil identificación de un gen o una porción de un transcriptoma a partir de fragmentos representativos de secuencias de ADNc, así como la caracterización del mismo, anotación en base a función predicha por similitud con proteínas asociadas de función conocida o secuencias nucleotídicas (a través de los algoritmos BLASTX y BLASTN, Altschul y col., 1990). La metodología de SSH permite reducir la representación de genes expresados en muy altos niveles y simultáneamente detectar y aislar aquellos genes menos abundantes mediante su enriquecimiento y recuperación, resultando en una normalización de la población de ADNc (Hu y col. 2003). Dicha técnica, en combinación con la de micromatrices de ADNc, permite obtener un perfil a gran escala del transcriptoma en estudio sin la necesidad de clonar los ADNc, de manera preliminar (Shi y col. 2005).

El alto grado de similitud encontrado con otros ESTs de plantas que no pertenecen a la familia taxonómica de las Compuestas indicó que el alto porcentaje de secuencias nuevas para girasol reportado en este trabajo (Fernández y col. 2003) no fue debido a secuencias altamente variables correspondientes a la región no codificante presente en el extremo 3' de ARNm. Estos resultados indican la alta eficiencia de esta técnica en el aislamiento de transcriptos órgano-específicos comparados con otras colecciones de ADNc previamente reportadas. Endo y col. 2000 reportaron que el 58,5% de las secuencias ESTs aisladas a partir del botón floral de *Lotus japonicum* presentaban similitud con secuencias depositadas en bases de datos, mientras que sólo el 11,4% de las mismas eran únicas para dicha fabácea. Otras publicaciones indican que solo el 12% de los ESTs provenientes de una clonoteca ecualizada de ADNc construida a partir de inflorescencias de *Arabidopsis thaliana* fueron unigenes (Takemura y col. 1999). Este número significativo de secuencias órgano específicas únicas aislado de hoja y flor en diferentes estadios de desarrollo

validó la técnica utilizada posibilitando la construcción de otras colecciones sucesivas de tallo y raíz sustraídas con hoja de modo de evitar la alta redundancia de secuencias asociadas a fotosíntesis. Por otro lado, el estadio de flor R1 fue sustraído con R4 con el objetivo de identificar la inducción específica de genes durante estadios tempranos de desarrollo. El alto nivel de redundancia en la clonotecas de flor R1, comparado con R4, puede deberse a diferencias en las estrategias de sustracción. Mientras que la clonoteca de flor R4 fue sustraída con una población de ARNm de otro órgano no relacionado (hoja) permitiendo la detección de transcritos poco abundantes o no representados en hoja, la clonoteca de flor R1 fue sustraída con ARNm del mismo órgano pero en distinto estadio de desarrollo. Los transcritos del mismo órgano/tejido comparten un gran número de ARNm's idénticos y, en consecuencia, un grupo reducido de transcritos diferencialmente expresados permanecieron no sustraídos en un estadio de desarrollo específico. En el caso de la clonoteca de raíz, el análisis de redundancia fue tratado de manera cautelosa debido al pequeño número de moléculas de ADNc que permanecieron no sustraídas luego del paso de hibridación. Las clonotecas de hoja y tallo mostraron niveles más altos de redundancia comparados con la clonoteca de flor en estadio R4. Esto podría deberse a la alta representación que tendrían las secuencias relacionadas con fotosíntesis. El análisis de especificidad por órgano entre las secuencias no redundantes confirmó que una alta proporción de unigenes en cada clonoteca corresponde a secuencias detectadas únicamente en estos tejidos. En las clonotecas de flor estadio R4 y tallo, el 93,5% y 98,7% de las secuencias analizadas fueron únicas para dichas clonotecas, respectivamente (Tabla 6). Un estudio global que incluyó a todas las clonotecas generadas reveló que el 87,8% de las secuencias no-redundantes se expresan diferencialmente.

Otros estudios anteriores reportaron valores similares de genes predichos aislados de diferentes clonotecas normalizadas de ADNc. Asamizu y col., 2002 reportaron que el 45% de los ESTs no-redundantes generados a partir de diferentes tejidos, incluyendo los órganos aéreos, primordios florales, raíces y plántulas crecidas bajo técnicas de cultivo de tejidos fueron reportados como posibles nuevos genes. Reddy y col., 2002 reportaron un 28,2% de nuevas secuencias no redundantes

obtenidas a partir de una clonoteca de ADNc generada a partir de plántulas de arroz sometidas a estrés por sequía. Por otro lado, también se observó en este trabajo una sub-representación de ESTs con función putativa de metabolismo celular comparado con otras clonotecas de ADNc normalizadas (Ali y col. 2002; Tamborindeguy y col. 2004), lo cual evidenciaría aún más la eficiencia en la construcción de estas clonotecas en la disminución del número de transcritos abundantes representados de manera igualitaria en los diferentes órganos analizados.

El número de secuencias depositadas en GenBank desde el año 2003 en que fueron generados y publicados estos resultados hasta el presente ha incrementado significativamente y en particular para la familia de las Compuestas (Tabla 7 del presente capítulo) que incluye actualmente 93.793 secuencias correspondientes al girasol cultivado de las cuales 33.246 corresponden a unigenes y sólo 2.061 están anotadas funcionalmente (TIGR Gene Indices, <http://www.tigr.org/tdb/tgi/hagi>). Al 30 de noviembre de 2006, la comparación de las 318 secuencias no-redundantes generadas en este trabajo contra las 466.639 secuencias de Compuestas (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) mostraron que aún 46 (14,4% del total del banco) no mostraron similitud significativa con ESTs de girasol previamente reportados, mientras que 121 (38%) tampoco lo hacen con ESTs de lechuga. Por otro lado, dentro de las 136 secuencias (42,7%) que no presentaron similitud con secuencias en Gene Ontology, 90 secuencias (70%) mostraron similitud significativa a ESTs de función desconocida, mientras que las remanentes (32% del total del banco) fueron reportadas en este estudio como nuevas para girasol.

En síntesis, la técnica de hibridación sustractiva (SSH) permitió la detección de secuencias nuevas para girasol a partir de un número relativamente pequeño de secuencias analizadas, a pesar del gran volumen de secuencias de esta especie depositadas hasta el momento en bases de datos públicas. De estas secuencias surgen nuevos genes candidato de relevancia (Peluffo y col. 2005) para la realización de un análisis funcional frente a estreses abióticos (Fernández y col.

2005) de compromiso agronómico para el rendimiento del girasol cultivado, según se presentará y discutirá en el Capítulo 3, Sección IV. de este trabajo de tesis.

CAPÍTULO 2: DESARROLLO DE UNA PLATAFORMA BIOINFORMÁTICA Y BASES DE DATOS PARA EL ANÁLISIS, ORDENAMIENTO Y ANOTACIÓN DE SECUENCIAS MOLECULARES

IV.2.1. Edición, ensamblaje de secuencias y detección de las redundancias presentes en las colecciones analizadas

Las secuencias obtenidas fueron procesadas manualmente y usando rutinas informáticas escritas en VB (Visual Basic) mediante las cuales se detectan regiones contaminantes del vector por comparación contra la secuencia completa del mismo y se eliminan de la secuencia cruda. De la misma manera se procesaron las regiones que contenían un alto número de indeterminaciones o de baja calidad de secuencia.

Las secuencias editadas fueron ensambladas a fin de definir las secuencias únicas representadas de las colecciones. El ensamblaje de fragmentos consiste en la unión de fragmentos de secuencia solapantes obtenidos en un proyecto de secuenciación. Esta tarea puede automatizarse en gran medida empleando sistemas informáticos o computadoras personales. La calidad de la lectura requiere que cada posición haya sido confirmada mediante la lectura de varios fragmentos independientes y solapantes con orientaciones en ambos sentidos de lectura, ensamblados y ordenados correctamente. Además es preciso comprobar que los fragmentos leídos realmente corresponden a la secuencia que está siendo estudiada, y no al vector o al sitio de unión múltiple de los oligonucleótidos empleados en el experimento, ni a contaminantes externos.

Todo ello puede facilitarse con la utilización de aplicaciones informáticas que realicen de forma automática la mayoría de las tareas repetitivas en el proceso de lectura, verificación y ensamblado de fragmentos. En este trabajo, la redundancia fue analizada por medio de los sistemas de agrupamiento que funcionaban bajo

versión alfa de Biopipeline®. El análisis posterior de las secuencias consenso se realizó utilizando el programa CAP3 (Huang, 1991) estableciendo como parámetro del análisis un valor de 20 bases y 85% de la secuencia como solapamiento mínimo.

IV.2.2. Anotación de secuencias

Para la anotación de las secuencias se instaló en un servidor local el programa BLAST (BLAST “Stand-Alone”, Altschul y col., 1997) para comparación de secuencias. La disponibilidad de este servidor para realizar alineamientos de secuencias anónimas contra las secuencias depositadas en base de datos públicos evita la tarea de realizar las mismas búsquedas en línea utilizando las plataformas que ofrecen los servidores públicos de NCBI, EBI, DDBJ entre otros centros de reposición de secuencias biológicas. La instalación del servidor BLAST Stand-Alone se realizó en un servidor Linux (IBM) y se habilitaron los servicios de red requeridos para ejecutar estas aplicaciones desde otros terminales conectados a la red interna del Instituto de Biotecnología. Asimismo, se realizaron todos los cambios necesarios para la adecuación de las pantallas acorde a las necesidades de los usuarios como el agregado de enlaces para facilitar los accesos y el manejo dentro de las pantallas, así como agregado de pantallas adicionales.

Para el alineamiento y anotación de secuencias se instalaron en el servidor local las siguientes bases de datos de nucleótidos y proteínas: Índice de Genes de TIGR para *Arabidopsis* (AGI); Girasol (HaGI), Lechuga (LsGI); Arroz (OsGI) y Tomate (LeGI). La base de datos NR de GenBank y la base de datos de vectores Univec también de Genbank. Las bases de proteínas SwissProt; Trembl del EBI; base de proteínas de *Arabidopsis thaliana* y arroz del MIP y las bases de datos Uniprot, Uniprotsw, Uniprot-tr y Uniref100 de la organización “Universal Protein Resource”.

Las secuencias obtenidas se compararon con secuencias nucleotídicas, proteicas y de ESTs disponibles en bases de datos públicas (GenBank, PIR, GenPept, etc.) utilizando programas basados en el algoritmo BLAST (Altschul y col. 1990). En base al grado de similitud entre secuencias se asignó una función génica probable a cada EST analizado. Posteriormente, dichas secuencias fueron anotadas usando el paquete Goblet acorde a las normas del Consorcio de Ontología Génica (GO) (Ashburner y col. 2000). Por otro lado, también se realizaron comparaciones entre los ESTs obtenidos y las secuencias específicas de ESTs públicas depositadas en bases de datos de lechuga y girasol (TIGR Gene Index, <http://www.tigr.org/tdb/tgi/hagi>).

IV.2.3. Diseño, creación y mantenimiento de una base de datos

Las secuencias obtenidas a partir de este trabajo se almacenaron, clasificaron y ordenaron en una base de datos generada con el programa MS Access 2000 (Microsoft Office Profesional, 2000®). Este programa permitió incorporar rutinas de programación en Visual Basic facilitando así la sistematización personalizada de tareas predeterminadas, la generación de macros para búsquedas, la conversión del total de los ESTs a formato FASTA y la importación de los resultados de búsquedas en bases de datos públicas, entre otras.

En la Figura 16 se muestra la estructura de la base de datos generada, sus campos y la interrelación entre ellos.

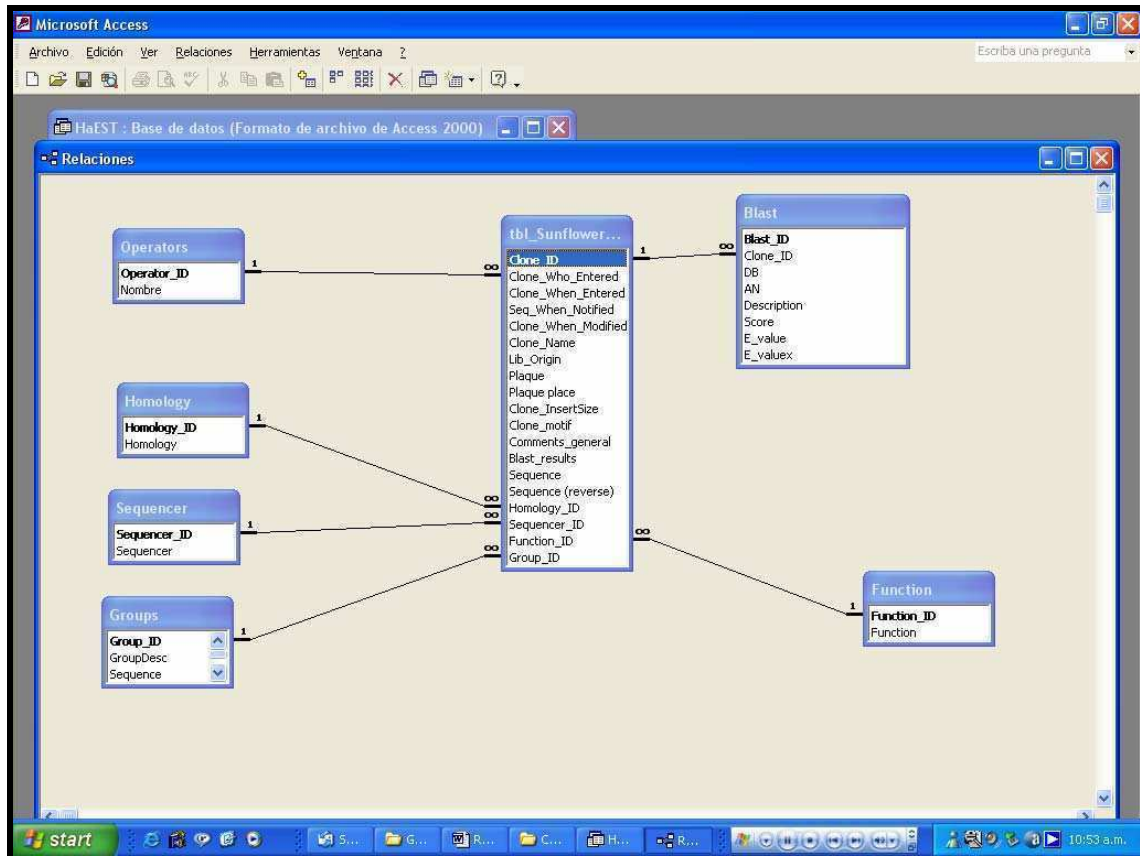


Figura 16: Estructura de la base de datos

La base de datos consta de 7 tablas, según puede verse en la figura previa, a saber: BLAST, Function, Groups, Homology, Operators, Sequencer y tbl_sunflower_EST. Estas tablas se encuentran integradas entre sí de manera que interrelacionan la información cargada en cada una de ellas, posibilitando las consultas múltiples según sea la información que se desea obtener.

La tabla BLAST contiene una serie de rutinas escritas en Visual Basic mediante las cuales se importan los resultados obtenidos en el uso del programa BLAST propiamente dicho. Este programa realiza un alineamiento local para el cual las secuencias deben ser ingresadas en formato FASTA. A su vez, las bases de datos que BLAST consulta para los alineamientos se identifican por medio de unos identificadores FASTA según se detalla en la Tabla 10:

Nombre de la base de datos	Sintaxis del identificador
GenBank	gb accession locus
EMBL Data Library	emb accession locus
DDBJ, DNA Database of Japan	dbj accession locus
NBRF PIR	pir entry
Protein Research Foundation	prf name
SWISS-PROT	sp accession entry name
Brookhaven Protein Data Bank	pdb entry chain
Patents	pat country number
GenInfo Backbone Id	bbs number
General database identifier ^a	gnl database identifier
NCBI Reference Sequence	ref accession locus
Local Sequence identifier	lcl identifier
^a gnl permite a las bases de datos no incluidas en estas listas usar el mismo carácter identificador. Como requisito excluyente, la combinación del segundo y tercer campo deberían ser únicos.	

Tabla 10: Manual de NCBI, <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.table.632>

El uso de estos identificadores fue tenido en cuenta al momento de diseñar las rutinas mencionadas, ya que de otra manera no hubiera sido posible la carga exitosa de los mismos dada la variabilidad de sintaxis entre ellos.

Por otro lado, en la Figura 17 puede observarse la apariencia de un registro cargado a esta base de datos a modo de planilla, y toda la información de interés para el análisis posterior del mismo registrado en campos de la planilla mencionada, utilizando los campos detallados en la figura anterior.

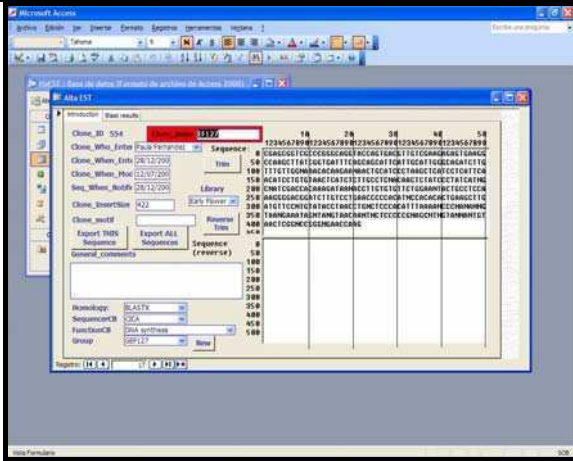
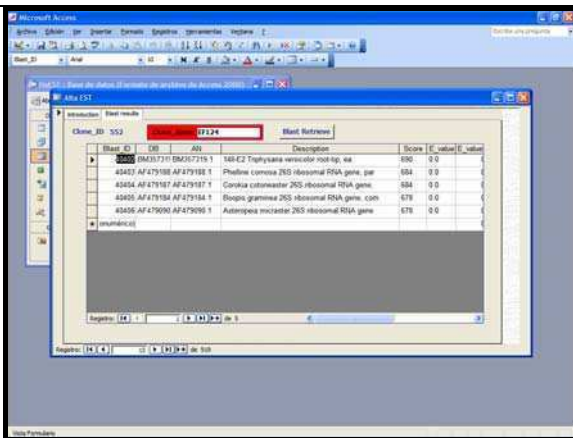
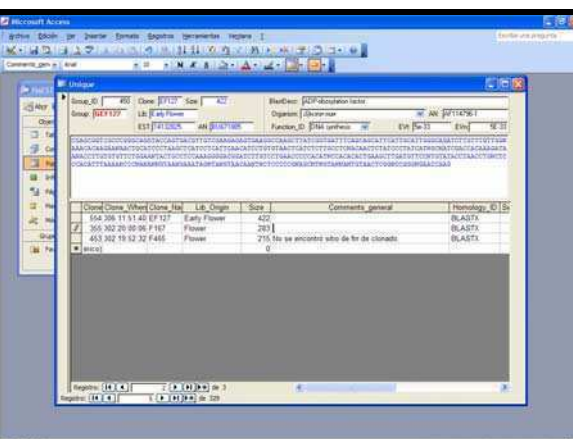
	<p>a) Formulario de alta de EST (Introducción)</p>
	<p>b) Formulario de alta de EST (Blast results)</p>
	<p>c) Formulario de grupos de EST. Dentro de cada grupo se encuentran las secuencias no-redundantes (contigs de la misma secuencia que representan transcritos únicos).</p>

Figura 17: Registro de base de datos generado para el banco local de ESTs

Una vez generada esta base de datos, se planteó la necesidad de migrar a una estructura de tipo relacional, más compatible con formatos estándar de uso general no sólo en el área de bioinformática sino también a nivel de protocolos TCP/IP. Para ello se utilizó el modelo ACeDB, creado para almacenar las secuencias del organismo *C. elegans* (<http://www.acedb.org/>), la cual administra datos biológicos (y de una manera creciente, datos no biológicos) de una manera flexible. Para el almacenamiento de datos en una plataforma ACeDB se comenzó a explorar el esquema de datos de una base equivalente, en este caso la desarrollada para el almacenamiento de datos biológicos de las iniciativas genómicas en gramíneas “GrainGenes” (<http://wheat.pw.usda.gov/GG2/index.shtml>), y sobre esta base se armó un nuevo esquema adecuado para el almacenamiento de datos biológicos en girasol. Para ello se instaló ACeDB (motor de la base) (<http://www.acedb.org/>), se exportaron los datos contenidos en la base de datos ACCESS a archivos con formato del nuevo motor. A partir de los cuales se los adaptó a un nuevo formato compatible para ser importados al nuevo esquema creado en ACeDB. Finalmente se configuró la interfaz ACeDB para abrir, administrar y hacer consultas en la base de datos desde la Web (Figura 18).

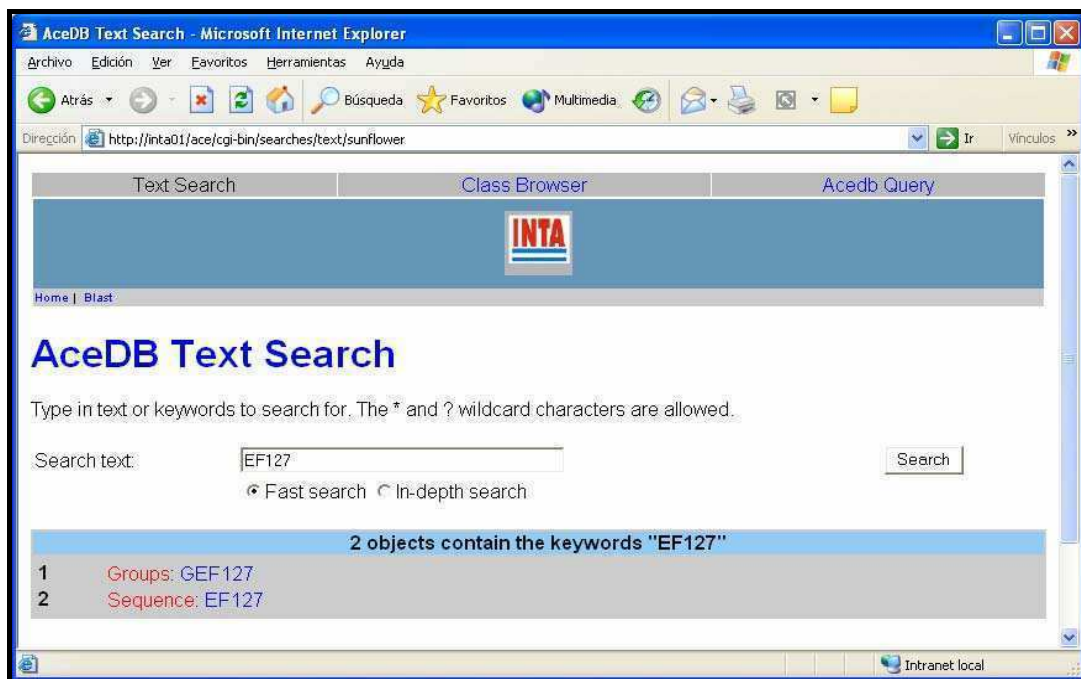


Figura 18: Base de datos de girasol bajo motor ACeDb en el servidor INTA01

Del mismo modo que lo descrito anteriormente, se adaptaron e hicieron los cambios necesarios de las pantallas para mejorar la visualización y facilitar el trabajo en las mismas y se implementaron enlaces para transferir información automáticamente a BLAST y realizar alineamientos de secuencias consultadas en GrainGenes, reduciendo la posibilidad de errores al copiar las secuencias.

IV.2.4. Diseño, creación y mantenimiento de una plataforma de análisis bioinformático (BioPipeline®)

BioPipeline® es un software útil para el procesamiento de datos bioinformáticos, el cual, mediante una interfaz gráfica, permite al usuario definir estructuras de procesamiento (“pipelets”) con gran flexibilidad. Estas estructuras fueron armadas mediante interconexión de objetos o bloques configurables que reciben la información, la procesan y generan un resultado que luego es procesado en el bloque siguiente de la cadena.

BioPipeline® provee la capacidad de conectarse a servidores BLAST (tanto de NCBI como Standalone Servers). En las Figuras 19 y 20 se muestran los componentes principales de BioPipeline® y su esquema de interacción. El software se ejecuta desde una PC Cliente y los datos se intercambian en formato XML.

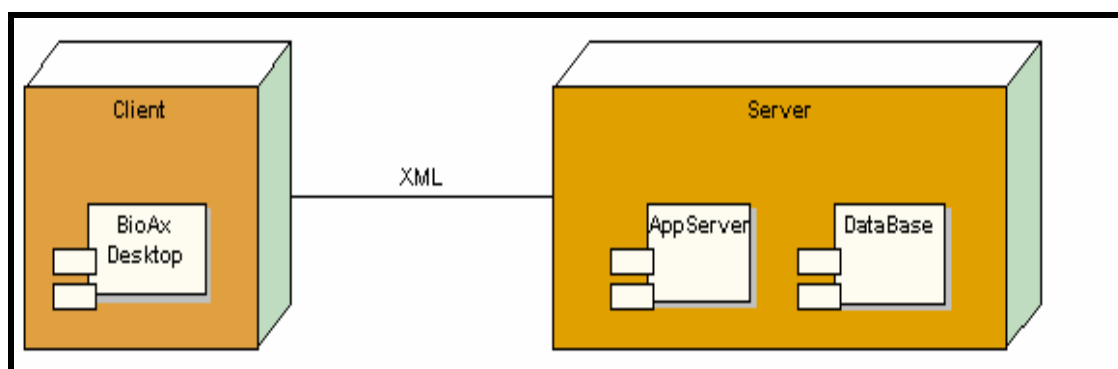


Figura 19: Componentes de Biopipeline®

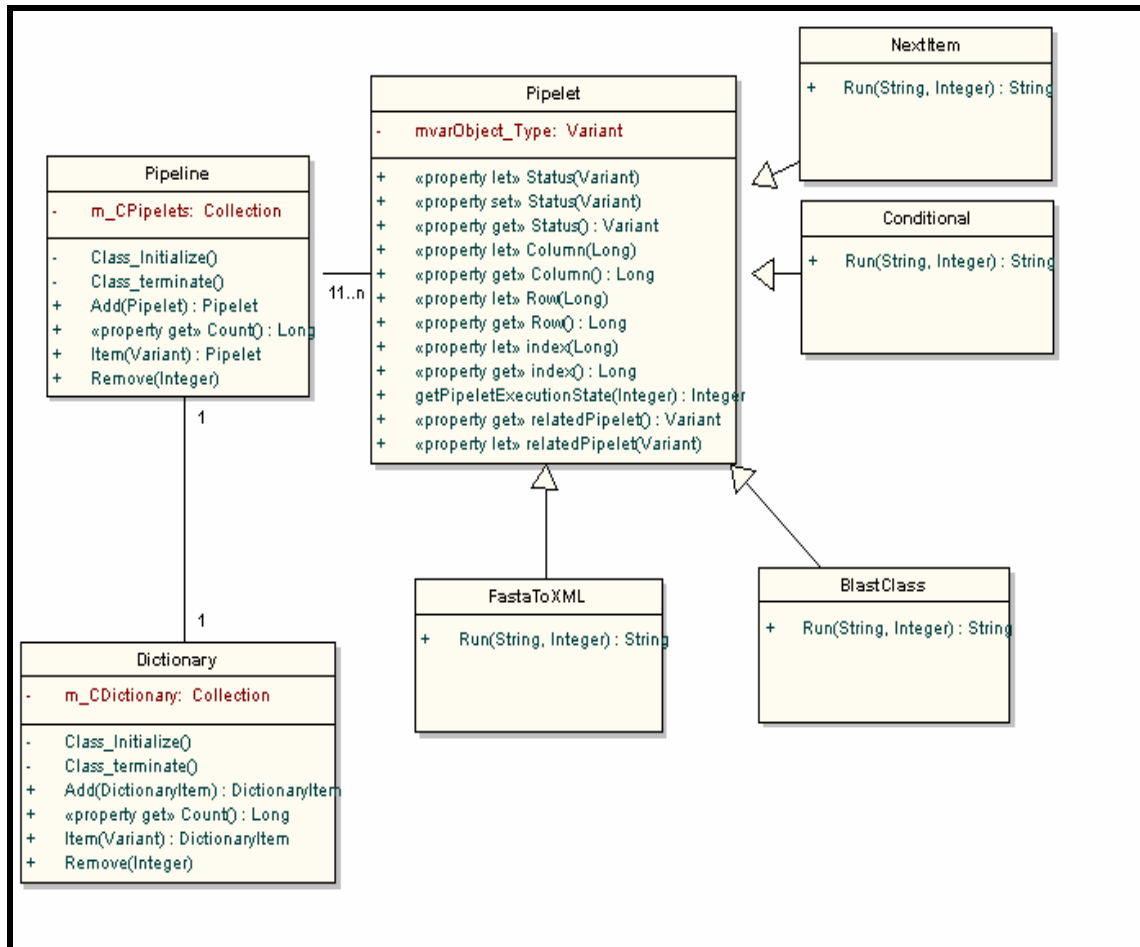


Figura 20: Interacción entre componentes de Biopipeline®

Los resultados finales pueden ser visualizados usando distintas interfaces de presentación de los datos, como puede verse en la Figura 21.

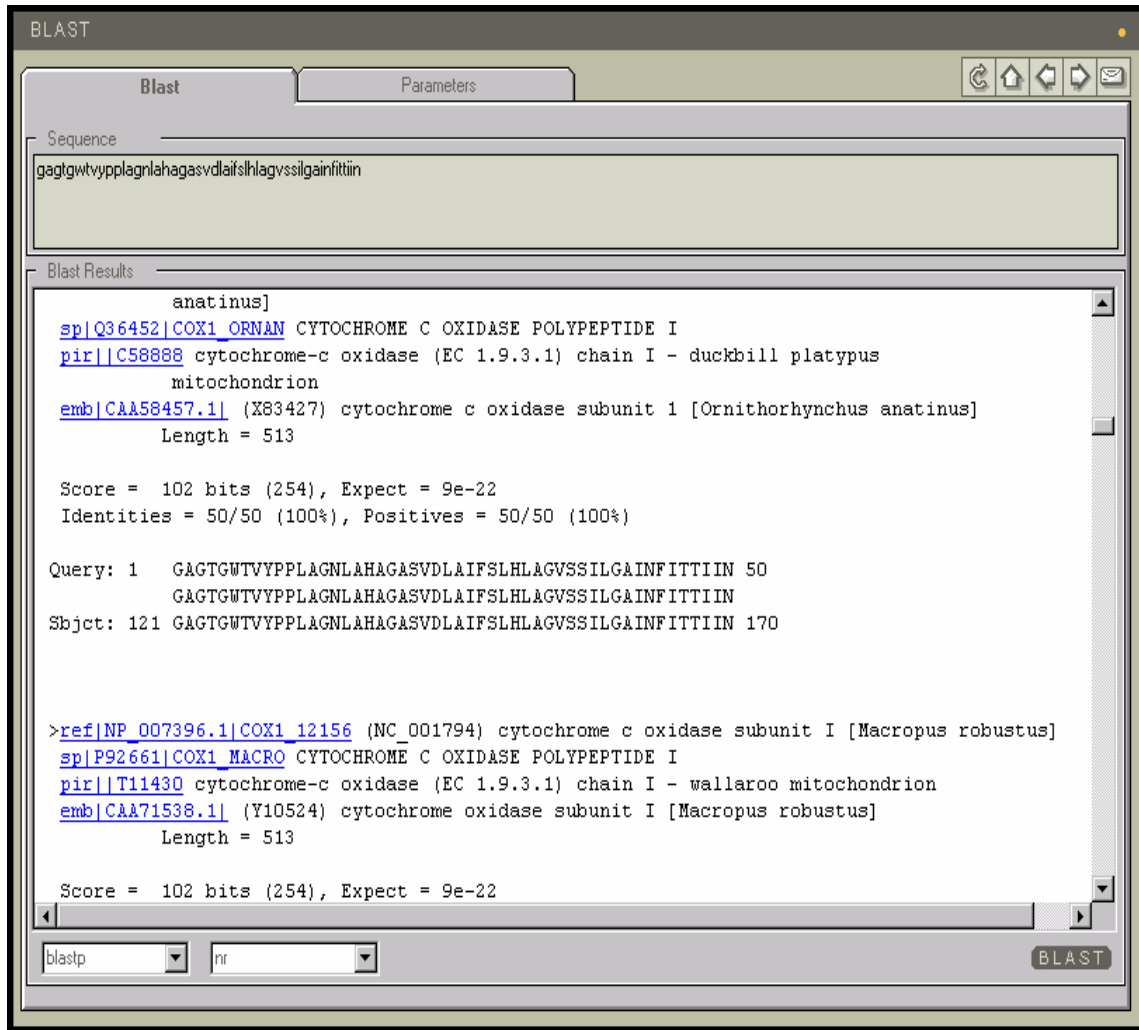


Figura 21: Ejemplo de interfaz de visualización de Biopipeline®

Se procesaron las 3 clonotecas correspondientes a los distintos estadios utilizando el programa BLAST2 para hallar coexpresión de ESTs, mediante una secuencia de rutinas (“pipeline”) conectada a un servidor BLAST2 Standalone. En la Figura 22 observamos dichas rutinas enlazadas, a partir de las cuales se generó el proceso de análisis en Biopipeline®.

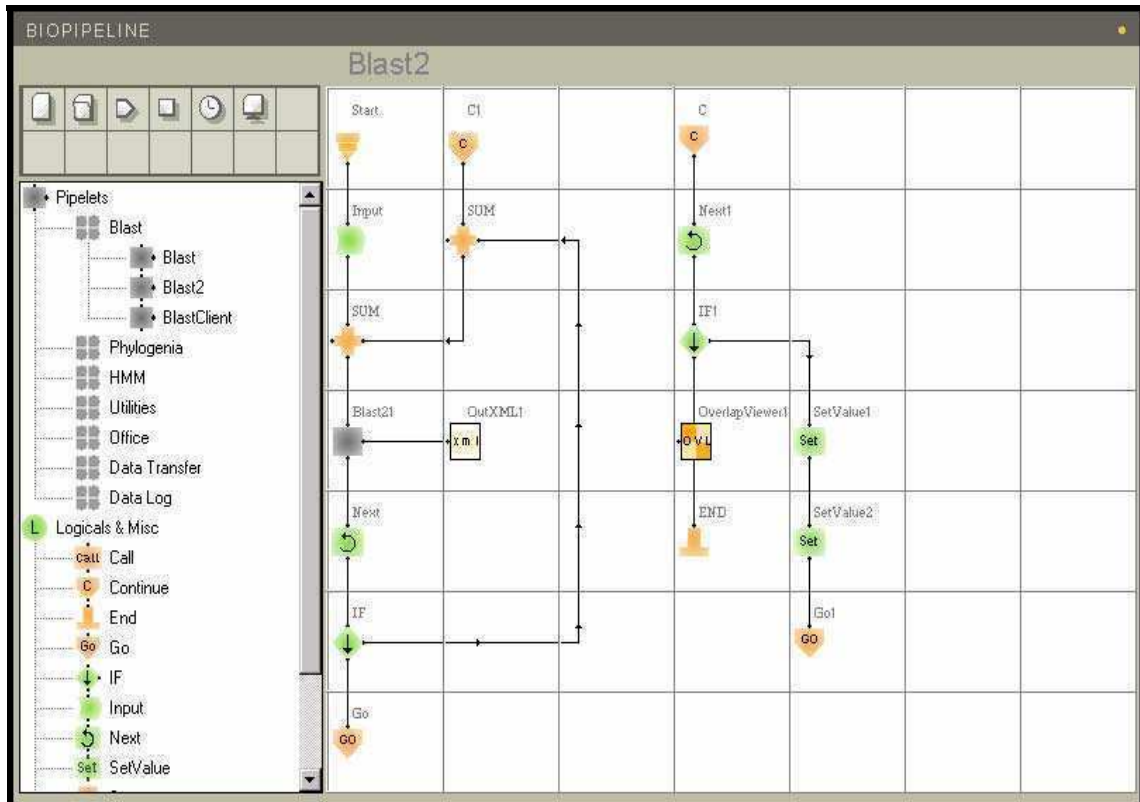


Figura 22: Rutinas enlazadas de Biopipeline®

Utilizando el objeto visor de solapamientos (“overlap viewer”: Figura 23), pudieron detectarse las secuencias redundantes en las clonotecas bajo análisis.

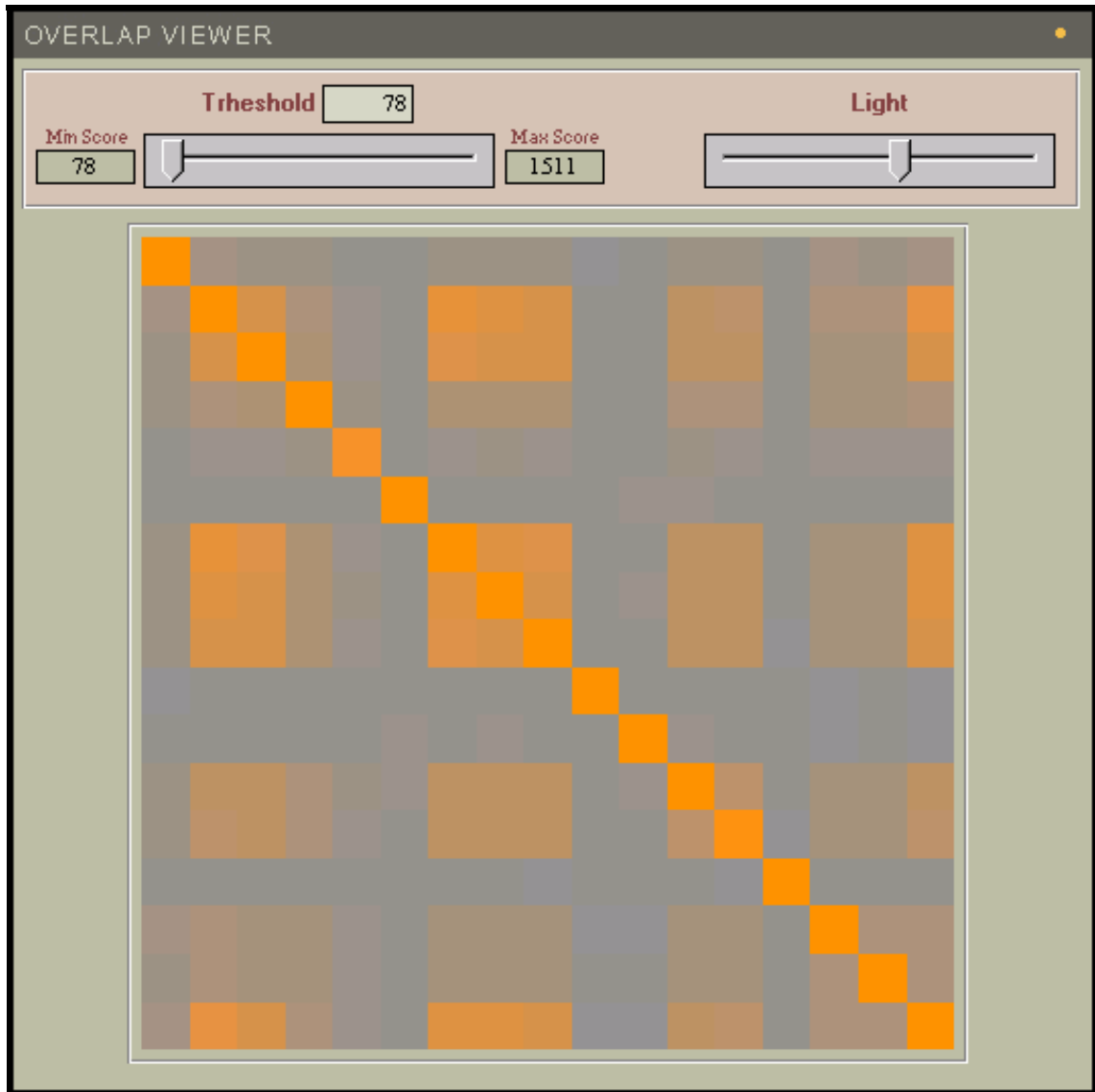


Figura 23: Visor de solapamientos de GenDesk®

IV.2.5. Instalación y puesta en funcionamiento del servidor INTA 01

Una vez desarrollado Biopipeline®, surgió la necesidad de contar con la plataforma en nuestra institución que a su vez contenga las bases de datos de disponibilidad pública vinculadas al análisis genómico para poder ejecutar BLAST de manera local (*standalone*).

Para ello se instaló y se puso en funcionamiento en un servidor local (INTA 01) las herramientas Phred/Phrap para edición de secuencias. Este paquete está siendo usado para editar las secuencias correspondientes a los marcadores de tipo SSR y elaborar los registros en el formato requerido para su depósito en las bases de datos moleculares de acceso público.

Se instaló y se está evaluando el funcionamiento de una plataforma de análisis de EST desarrollada por el Centro Nacional de Investigaciones Genómicas, USA (NCGR) denominada XGI (<http://www.ncgr.org/ourwork>) que enlaza procesos de edición, anotación y almacenamiento de secuencias de manera automática. Este sistema permite el análisis de secuencias a través de búsquedas usando BLAST e InterProScan, la anotación según Gene Ontology (GO), la búsqueda de péptido señal usando PexFinder. La información obtenida a partir de este proceso es almacenada en una base de datos y accesible desde la Web a través de un buscador estándar. Esta plataforma permite la intercomunicación entre bases de datos facilitando el análisis integrado de información disponible para los sistemas en estudio con el objetivo final de identificar nuevos genes para contribuir a un control eficiente de los patógenos o bien al mejoramiento de los cultivos comerciales. Por otra parte esta plataforma permite mantener actualizado el análisis periódico de los datos disponibles, al mismo tiempo que facilita la distribución de la información a través de una interfase Web en construcción. En la Figura 24 se observan la pantalla de bienvenida del servidor y los programas actualmente cargados en INTA01, conjuntamente con las bases de datos desarrolladas para el almacenamiento de las secuencias del presente trabajo.

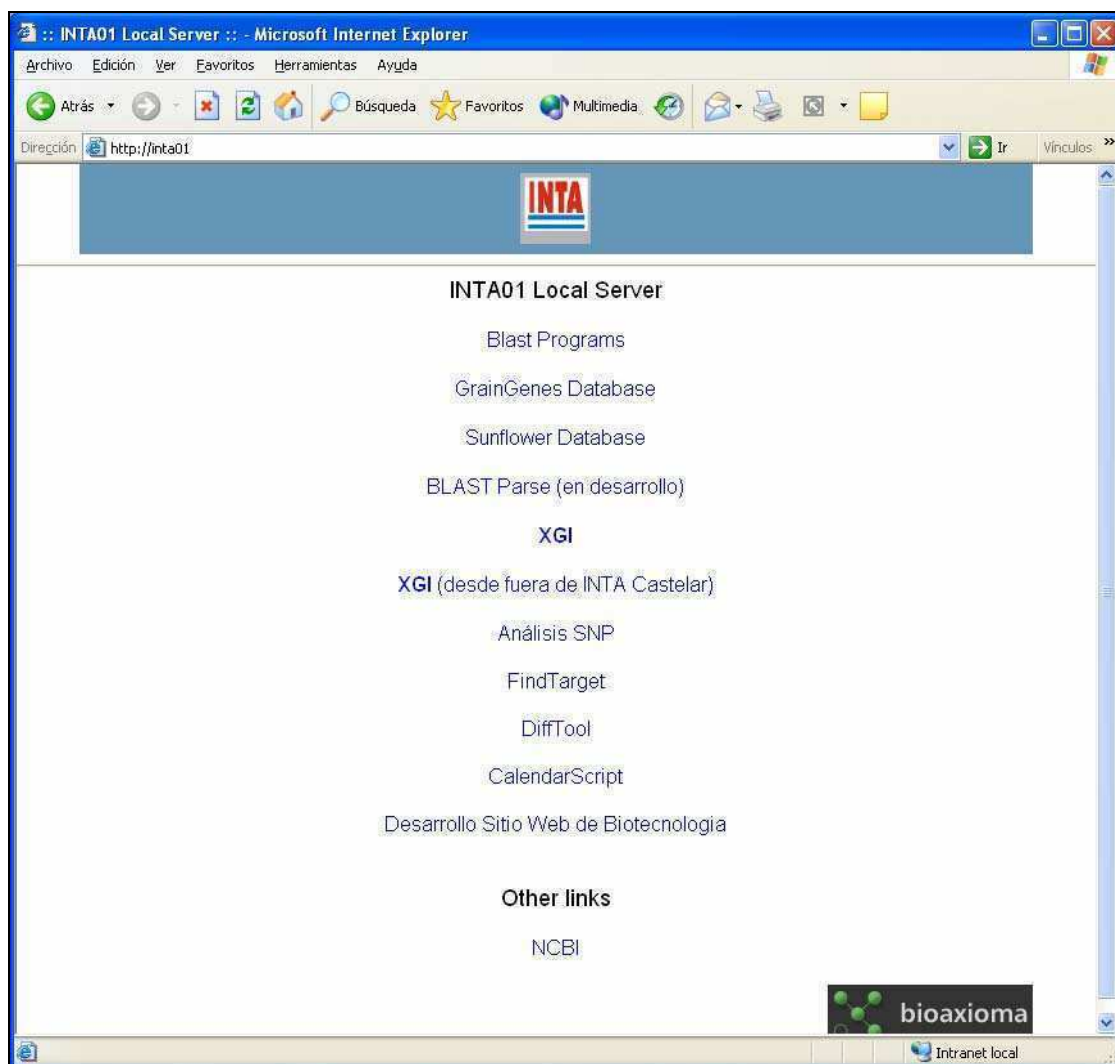


Figura 24: Pantalla de bienvenida de servidor INTA01

IV.2.6. Discusión

Las bases de datos de secuencias constituyen la base de todas las herramientas bioinformáticas usadas en análisis de secuencias. Si bien existe un gran número de iniciativas que ha producido una importante cantidad de bases de datos más o menos especializados con distintos tipos de datos cuya enumeración resultaría larga y compleja, vale la pena aclarar que los centros de referencia en relación con la generación de datos e integración de bases moleculares son el NCBI (<http://www.ncbi.nlm.nih.gov>), el EBI (<http://www.ebi.ac.uk>) y el DDBJ (DNA Database of Japan: <http://www.ddbj.nig.ac.jp>). Este último fue utilizado como un servidor de consulta creciente a lo largo de este trabajo, no solo por su interfaz de integración con otros servidores sino también por la posibilidad de analizar secuencias en bases de datos no presentes en otros espejos del GenBank. Un ejemplo de esto es la base de datos DAD (DDBJ aminoacid sequence database, <http://www.ddbj.nig.ac.jp/ddbjnew/nl17/news-j.html>, que contiene todas las secuencias nucleotídicas traducidas liberadas periódicamente partir de DDBJ más sus actualizaciones, con algunas diferencias según el resto de los portales. El uso de DDBJ fue ampliamente utilizado en este trabajo para aquellas secuencias que no presentaban similitud en las bases de datos tradicionales *a priori*, y luego validadas con los demás portales una vez que las actualizaciones eran llevadas a cabo.

Con respecto a BioPipeline®, su uso demostró aumentar la eficiencia y exactitud del análisis de secuencias a gran escala, llevados a cabo previamente de manera manual, ya sea mediante el uso de algoritmos de alineamiento locales entre secuencias o utilizando el paquete BioEdit® (Tom Hall, <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>), entre otros. El análisis manual de secuencias arrojó un total de 329 secuencias ensambladas (contigs), mientras que el valor final de BioPipeline® fue de 318 secuencias. Sumado a esto, el tiempo de análisis fue lento y exhaustivo, ya que demandó la comparación de cada secuencia en forma individual con las restantes del banco de ESTs y el posterior análisis del

alineamiento, discriminando aquellas con valor de puntaje mayor a 80 (*score* en *bits*, según BLAST), entre 50 y 80 y menor a 80.

Según Dopazo y Valencia (2002) las herramientas bioinformáticas más usadas en el contexto del análisis de secuencias incluyen los métodos de búsqueda de secuencias similares (mediante los algoritmos BLAST y FASTA), el alineamiento múltiple (CLUSTAL), la definición de regiones conservadas con posible significado funcional o el uso de estas regiones para definir familias de genes o proteínas y los métodos filogenéticos para reconstruir las relaciones entre las secuencias. Sin embargo, todas estas herramientas están diseñadas para trabajar con una o pocas secuencias y el procedimiento clásico para saber más sobre una secuencia consisten, básicamente, en buscar información sobre ella en bases de datos. El nuevo desafío del análisis de datos genómicos implica la capacidad de manejar mucha información de carácter difuso. Esto entraña la resolución de dos problemas principales: clasificar en grupos todos los datos y averiguar, *a posteriori*, que información tiene en común los miembros de cada grupo que los distingue entre sí. Acorde con este supuesto, hemos identificado las secuencias que conforman grupos entre sí a través de un ensamblado como el primer acercamiento a su clasificación en grupos, diferenciando aquellas que conforman un contig o secuencia ensamblada de las que no.

Una vez obtenidas las secuencias, clasificadas en grupos y caracterizadas según su funcionalidad hipotética, se las almacenó en una base de datos. Si bien la organización por campos fue de sencilla aplicación y desarrollo y generó un acceso a la información de amplias posibilidades, resultó compleja la incorporación de los datos de similitud según BLAST en la tabla BLAST Search (ver II.1.3.) de la base de datos. Una de las causas de esta dificultad fue dada por la existencia de los identificadores FASTA (Tabla 10) que enumeran a cada secuencia con un formato diferencial, lo cual generaba conflicto en las rutinas de almacenamiento de los archivos con extensión .fas o .txt. Es por ello que Fedorova y col. 2002 mencionan como principal ventaja de los análisis de ESTs *in silico* la manera rápida y efectiva en que pueden ser analizadas un número alto de secuencias.

La metodología aplicada en este trabajo implicará la puesta en común de equipos e información para relevamiento de datos e información biológica disponible actualmente (ESTs, GSS, SNPs, nucleótidos, proteínas, metabolitos, cartografía genética, etc.) lo que permitirá definir el contenido específico de las bases de datos a implementar en el futuro, utilizando como modelo las desarrolladas a lo largo de esta Tesis. En este sentido se diseñarán e implementarán prototipos de bases tomando como referencia estándares de código abierto aceptados internacionalmente y se investigarán recursos y herramientas de visualización de la información (Generis Software Components for Model Organism Databases, Gmod, <http://www.gmod.org>). Asimismo los datos serán incorporados utilizando el vocabulario controlado sobre estándares Gene Ontology (GO) u Open Ontology Consortium (OBO).

No obstante, para que todo este proyecto tenga una pronta puesta en funcionamiento y desarrollo acorde a los supuestos planteados, se torna necesaria la concientización en el reconocimiento de la bioinformática como una disciplina autónoma y necesaria dentro de los proyectos de ciencia y técnica, fundamentalmente en el área de la genómica aplicada al mejoramiento de los cultivos.

CAPÍTULO 3: ESTUDIOS DE EXPRESIÓN DE GENES CANDIDATO FRENTE A ESTRESSES ABIÓTICOS

IV.3.1. Tecnología del chip de ADN

Las micromatrices de ADN constituyen una de las tecnologías disponibles más usadas para el análisis funcional del genoma a gran escala. Existen dos tecnologías básicas para la producción de micromatrices de ADN, una de ellas utiliza como sustrato moléculas de ADNc y la otra oligonucleótidos sintéticos, según fue descrito en la sección I.10. (Introducción). A partir de cualquiera de ellas se produce una matriz de expresiones génicas cuyo tratamiento estadístico es común a cualquier plataforma y de allí el carácter general de los instrumentos de análisis.

Las mayores limitaciones en la utilización de estas tecnologías giran alrededor del análisis de datos. Aún cuando en los últimos años se han hecho muchas contribuciones en este campo, no existen soluciones ampliamente aceptadas y experimentos comparativos dentro y entre plataformas revelan la variabilidad inherente a estas tecnologías tanto a nivel de replicas técnicas como biológicas, haciéndose cada vez mas evidente la dificultad del análisis e interpretación de sus resultados (Michiels y col., 2005).

IV.3.2. Micromatrices de ADN

a) Confiabilidad de las repeticiones

Un análisis preliminar de la diferencia en la expresión génica de algunos genes mostró que la repetición 3 del grupo de vidrios correspondientes al tratamiento “control” se comportaba de manera atípica. Sin embargo concluir sobre el carácter atípico de una repetición en un experimento con tres repeticiones por tratamiento es difícil. Para evaluar cuán disímil son las repeticiones biológicas de este

experimento se consideró el problema desde una perspectiva multivariada ordenando las repeticiones en un plano de ordenación generado mediante análisis de componentes principales aplicado a la matriz de expresiones génicas. Esta aproximación tiene en cuenta la información de todos los genes simultáneamente para ubicar en el plano a las repeticiones. La representación de las repeticiones en los planos definidos por los componentes principales 1,2, 1,3 y 2,3 se muestran en las Figuras 25a,b y c respectivamente. La proporción de variabilidad total retenida por estos tres ejes es del 86%. En estos planos es fácil observar que la repetición número tres del tratamiento “control” se aleja de las otras repeticiones a una distancia considerablemente mayor que la distancia que separa las repeticiones de los otros tratamientos entre si.

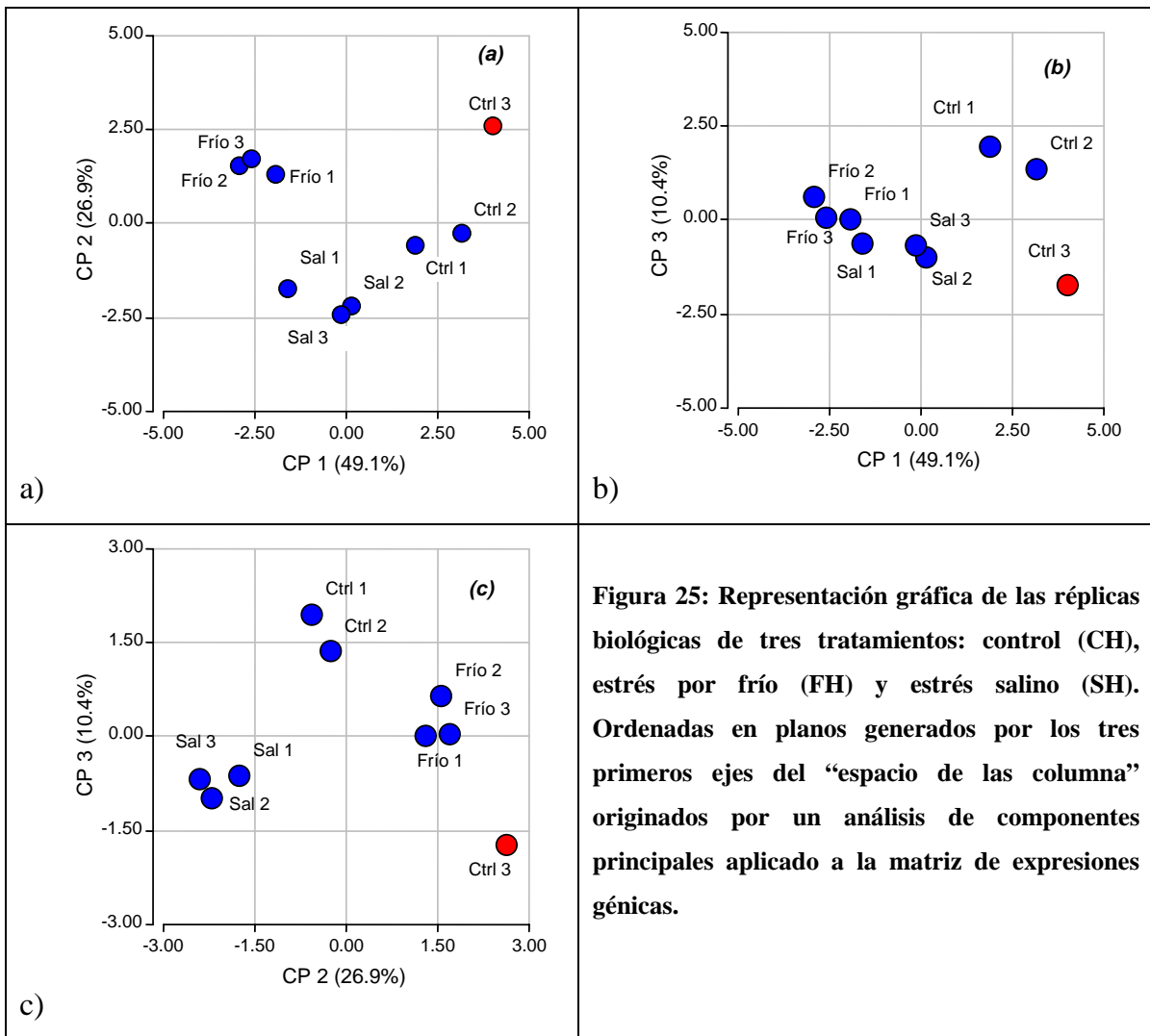


Figura 25: Representación gráfica de las réplicas biológicas de tres tratamientos: control (CH), estrés por frío (FH) y estrés salino (SH). Ordenadas en planos generados por los tres primeros ejes del “espacio de las columna” originados por un análisis de componentes principales aplicado a la matriz de expresiones génicas.

Teniendo en cuenta la observación anterior, la repetición 3 del tratamiento control fue eliminada del análisis subsiguiente.

Para poner a todos los genes en una escala comparable, la expresión de cada gen en cada repetición se dividió por una estimación de la desviación estándar común dentro de tratamientos. Esta estimación corresponde a la raíz cuadrada del cuadrado medio residual del análisis de la varianza univariado de cada gen.

b) Análisis exploratorio de los perfiles de expresión génica

Utilizando el algoritmo *k-centroides* de clasificación no supervisada se agruparon genes según sus perfiles de expresión. El perfil de expresión de cada gen se constituyó por su expresión a través de las distintas repeticiones de los tratamientos comparados. El algoritmo *k-centroides* demandó la especificación del número de grupos que se esperaba detectar. En el contexto de esta aplicación, implicaba el previo conocimiento de cuantos grupos diferentes de expresión génica podía haber presentes. Pero es justamente esto lo que no se sabe y se quiere establecer. Una forma de establecer el número de grupos de expresión génica es corriendo el algoritmo de los *k-centroides* para diferente número de grupos y evaluar la heterogeneidad dentro de los conglomerados que se forman. La relación entre esa medida de heterogeneidad y el número de conglomerados propuestos se grafica para obtener un curva como la que se muestra en la Figura 26. La lógica para establecer el número de conglomerado “óptimo” es la siguiente: la heterogeneidad dentro de conglomerados usualmente disminuye cuando el número de conglomerados, entre los que pueden distribuirse las observaciones, es mayor. Sin embargo, una vez alcanzado el número “natural” de conglomerado, dividir las observaciones en un número mayor de grupos solo produce una declinación suave de la heterogeneidad ya que, desde un punto de vista teórico, lo que se está haciendo es agrupando por similitud de los errores experimentales que se suponen de mucha menor magnitud que las diferencias estructurales que se tratan de identificar. Luego el número óptimo de conglomerados es el número de conglomerados a partir del cual se

observa una reducción apreciable de la tasa de disminución de la heterogeneidad dentro de conglomerados. En el caso que nos ocupa se considera que 3 es el número óptimo de conglomerados, aunque la decadencia de la heterogeneidad sigue una curva bastante suave.

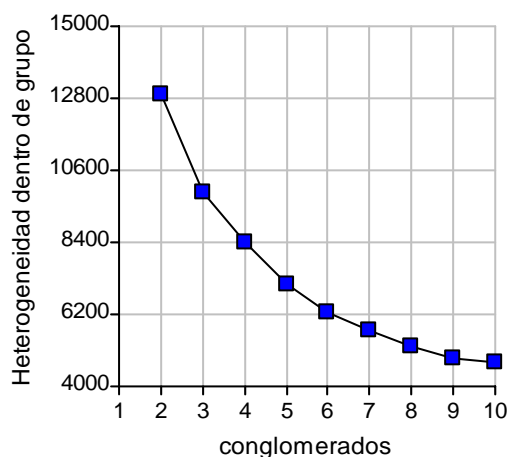


Figura 26: Curva que representa la relación entre el número de conglomerados producidos por el algoritmo de los *k-centroides* y una medida de heterogeneidad común dentro de conglomerados.

Teniendo en cuenta el resultado anterior, se identificaron los genes pertenecientes a cada uno de los tres grupos generados por el algoritmo *k-centroides*. Las bandas de confianza al 95% para perfiles medios de cada grupo se presentan en la Figura 26. La Tabla 11 presenta, para cada uno de los tres grupos de genes identificados por el algoritmo de los *k-centroides*, el número de genes por grupo y la media, la desviación estándar, el mínimo y el máximo de la expresión génica en cada tratamiento.

Como puede observarse, los grupos identificados correspondieron a patrones de expresión génica bien diferenciados. El Grupo 2, integrado por 126 genes (ver Tabla 12 para una enumeración de los genes en este y los otros grupos), no presentó variación en sus niveles medios a través de los tratamientos. Contrariamente al Grupo 2, el Grupo 1 (integrado por 112 genes que se sobre-expresan respecto del control cuando las plantas son sometidas a estrés ya sea por frío o por incremento de la salinidad en el medio de cultivo) presentó sobre-expresión respecto del control para todos los tratamientos evaluados. De manera análoga, los 49 genes que

conformaron el Grupo 3, presentaron una sub-expresión respecto del control en los mismos tratamientos. Tanto en el Grupo 1 como en el Grupo 2, la diferencia de expresión génica con respecto al control fue máxima cuando se consideró el tratamiento por frío.

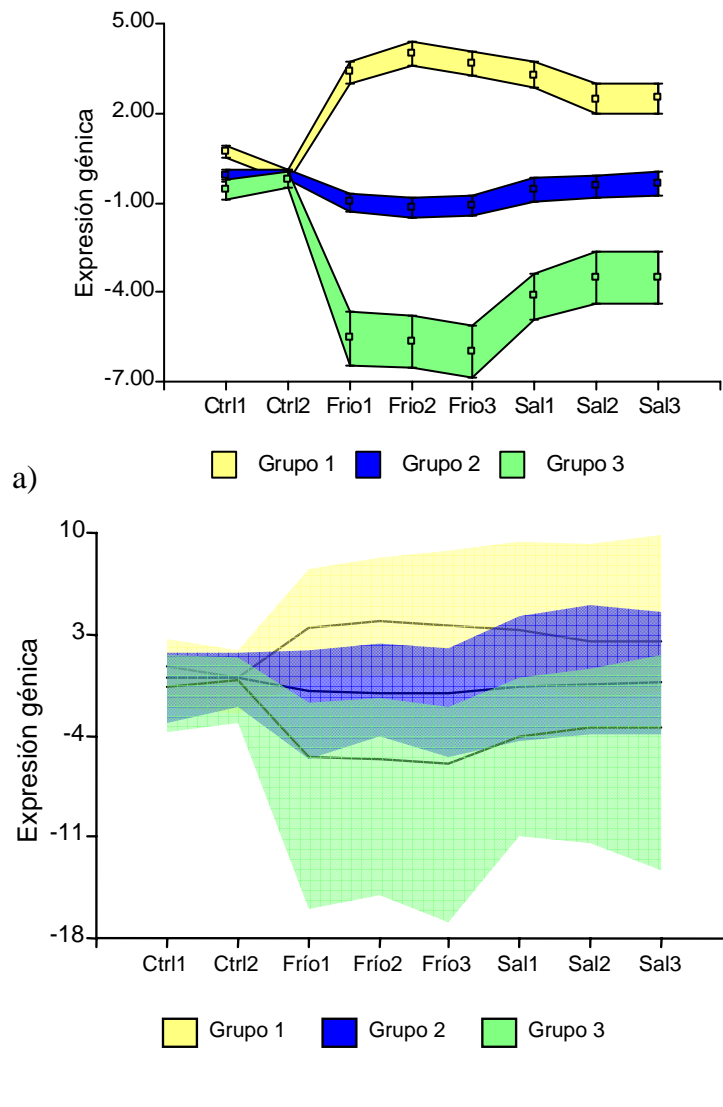


Figura 27: Bandas del confianza al 95% (a) y bandas de predicción al 95%(b) para cada uno de tres grupos de genes agrupados por el algoritmo *k-centroides* de clasificación no supervisada aplicado a la matriz de expresiones génicas (CH1 y CH2 corresponden a las repeticiones biológicas del control, FH1, FH2 y FH3 las repeticiones del tratamiento con frío y SH1, SH2 y SH3 las correspondientes repeticiones del tratamiento salino).

Grupo	Genes	Tratamiento	Media	D.E.	Mín	Máx
1	112	Control	0,0354	0,0811	-0,3010	0,2450
		Estrés por frío	0,3243	0,1934	-0,1197	0,8707
		Estrés salino	0,2260	0,1755	-0,0420	0,8300
2	126	Control	-0,0139	0,0821	-0,3965	0,1350
		Estrés por frío	-0,1015	0,1728	-0,6850	0,3220
		Estrés salino	-0,0564	0,1722	-0,6123	0,4130
3	49	Control	-0,0298	0,0664	-0,3145	0,0610
		Estrés por frío	-0,4297	0,2408	-0,9063	-0,0193
		Estrés salino	-0,2838	0,2349	-0,8700	0,3957

Tabla 11: Estadísticos resumen de la expresión génica bajo distintos tratamientos en tres grupos de genes identificados mediante el algoritmo *k-centroides* aplicado a la matriz de expresiones génicas

Observando la Figura 27 vemos que existen genes con expresión diferencial entre tratamientos. Las listas de genes incluidos en los Grupos 1, 2 y 3 probablemente incluyeron falsos positivos como falsos negativos. Si bien el procedimiento de los *k-centroides* ha producido un agrupamiento que hace que los genes asignados a los Grupo 1 y 3 tengan una probabilidad mayor de ser verdaderos positivos y contrariamente los asignados al Grupo 2 de ser verdaderos negativos, es posible algún grado de confusión, como lo sugiere la variabilidad de los perfiles de respuesta dentro de cada uno de estos grupos (Figura 27, bandas de predicción). Por otra parte la diferencia entre control y tratamientos parece ser de tal magnitud que el algoritmo de clasificación de los *k-centroides* estuviera captando los patrones de aquellos genes que separan principalmente a los tratamientos (frío y salinidad) del control y no distingue los genes que podrían estar expresándose diferencialmente entre los tratamientos estresantes. En consecuencia, los genes que diferencian los tratamientos de estrés podrían quedar ocultos ya sea por que son pocos y/o porque su expresión diferencial no tiene la magnitud necesaria para ser detectada. Una aproximación clásica al problema de la expresión diferencial es

probar, gen a gen, la igualdad de medias entre tratamientos. Estas pruebas se modifican, en el contexto del análisis de datos de micromatrices, para controlar la tasa de falsos positivos (*false discovery rate*). Sin embargo esta corrección se asocia con altas tasas de falsos negativos. Por tal motivo, en esta aplicación, se ha intentado otra aproximación basada en una técnica multivariada de ordenación conjugada con una aproximación clásica de prueba de hipótesis (Fernández y col. 2006).

Grupo 1				Grupo 2					Grupo 3	
EF127	F358	F514	H388	EF121	F226	F552	H448	T340	EF432	H136
EF202	F362	F526	H406	EF136	F228	F568	R141	T348	EF466	H207
EF264	F363	F528	H417	EF161	F232	F569	R163	T349	EF502	H209
EF457	F373	F537	H459	EF167	F234	F588	R184	T362	EF624	H302
F172	F376	F542	R116	EF189	F268	F596	T107	T369	F122	H304
F176	F377	F547	T108	EF201	F284	H101	T111	T370	F126	H311
F180	F379	F554	T110	EF208	F304	H108	T117	T375	F137	H322
F181	F413	F557	T120	EF269	F332	H109	T145	T383	F168	H354
F191	F421	F560	T124	EF274	F333	H112	T147	T387	F171	H378
F202	F426	F561	T128	EF280	F334	H115	T149	T394	F175	H385
F205	F431	F563	T150	EF283	F342	H120	T166	T406	F183	H387
F206	F432	F564	T155	EF316	F343	H121	T182	T415	F209	H393
F211	F440	F572	T174	EF332	F356	H122	T205	T435	F210	H404
F219	F446	F576	T176	EF343	F365	H127	T206	T452	F218	T154
F220	F447	F583	T187	EF421	F369	H133	T213		F221	T221
F237	F449	F587	T189	EF449	F407	H137	T223		F231	T234
F244	F453	F593	T196	EF474	F409	H139	T233		F372	T253
F258	F455	H128	T218	EF606	F412	H210	T240		F401	T277
F259	F458	H132	T243	EF652	F425	H217	T254		F443	T289
F279	F461	H154	T266	EF653	F452	H221	T256		F543	T302
F287	F467	H205	T282	F173	F457	H312	T269		F549	T450
F295	F482	H308	T283	F192	F468	H316	T272		F550	
F305	F484	H315	T307	F201	F469	H317	T275		H110	
F314	F486	H329	T360	F203	F470	H402	T303		H111	
F319	F489	H337	T368	F208	F483	H426	T313		H117	
F336	F491	H360	T423	F212	F493	H432	T318		H123	
F340	F494	H368	T464	F215	F530	H433	T333		H124	
F341	F496	H379	T479	F225	F531	H440	T338		H125	

Tabla 12: Lista de los 287 genes analizados por micromatrices y su clasificación diferencial en grupos

*c) Identificación de genes candidato - Análisis de componentes principales (ACP)**c.2) Generación de gráfico bi-plot por ACP*

Una forma alternativa de abordar la identificación de genes con expresión diferencial es buscar espacios de representación en los que la ubicación de los genes y su relación con los tratamientos permiten inferir el sentido e intensidad de sus asociaciones. Para ello, se utilizó el *biplot* generado por la representación de los genes en el plano generado por los primeros ejes de un análisis de componentes principales de una matriz cuyas filas se correspondieron con genes y sus columnas con las expresiones génicas en los distintos tratamientos (control, frío y salinidad). En esta aplicación del análisis de componentes principales, los objetos que se quieren ordenar son los genes y los atributos son las expresiones de un mismo gen a través de los tratamientos y sus repeticiones. Luego, los puntos, en el *biplot* que aparecieron conectados al centro del mismo, corresponden a los pares (x,y) conformados por los coeficientes con los que las repeticiones de los distintos tratamientos contribuyó a los componentes principales. El gráfico resultante se muestra en la Figura 28. Los segmentos de recta que representan las repeticiones de los distintos tratamientos tienen dirección y longitud. Como puede observarse las repeticiones de un mismo tratamiento se orientan en el mismo sentido y con similar longitud. Este resultado es esperable y confirma la identidad de los tratamientos. Luego podemos identificar una dirección con el tratamiento de frío, otra con el de salinidad y otra al tratamiento control. Aquellos genes ubicados en la dirección “apuntada” por las repeticiones del tratamiento con frío son aquellos que se sobre-expresan en esta condición. De manera similar los genes que se ubican en la dirección de salinidad son los que se sobre-expresan en esa condición. Contrariamente los genes que ubican en las direcciones opuestas son los que se sub-expresan en el tratamiento correspondiente. Aunque la escala en la que se graficaron las repeticiones de los tratamientos fue arbitraria y fue elegida de tal manera que se pudieran visualizar adecuadamente en el espacio de representación de los “genes”,

las longitudes relativas de sus segmentos son indicativas del peso que tienen en la expresión diferencial de los genes. En la representación de la Figura 28, los controles presentaron segmentos comparativamente cortos, indicando que no tenían peso en la ordenación y por lo tanto que no había genes que se sobre- o sub-expresaran en esta condición. Este resultado fue consistente con el hecho de que un *pool* de plantas control (CH) fue utilizado como referencia de todas las repeticiones. El eje principal 1 que retuvo el 76% de la variabilidad total del espacio multivariado original, ordenó los genes según si el tratamiento estresante indujo (derecha) o reprimió (izquierda) su expresión respecto del control. El eje 2 que retuvo el 17% de la variabilidad original estaba asociado con la diferenciación entre el tratamiento salino y frío. Luego en el cuadrante I de la Figura 28 están los genes que se sobre-expresaban en el tratamiento salinidad, los del cuadrante II son los que sobre-expresaban en tratamiento frío, el cuadrante III tiene los genes que se sub-expresaban en tratamiento salinidad y finalmente los del cuadrante IV son los genes sub-expresados en el tratamiento frío.

Con el fin de buscar genes con expresión diferencial se calculó la distancia al origen del plano de ordenación para los puntos que representan a los genes. Esta distancia debió estar asociada al grado de diferenciación de la expresión de los genes. Cuanto más alejados del centro más sobre- o sub-expresados deberían estar los genes en algún tratamiento. De esta manera, aquellos ubicados en la periferia de la nube de puntos deberían ser los genes más prometedores.

Para combinar esta aproximación, basada en la ordenación de genes por ACP, y el criterio clásico de prueba de hipótesis para igualdad de medias se consideran solo aquellos genes que para la prueba clásica de análisis de la varianza tuvieran un *p*-valor menor o igual a 0,05 y que estuvieran a una distancia del origen en la Figura 28, mayor que 9 (correspondiente, aproximadamente al percentil 70 de la distribución de distancias al origen). Este criterio de corte se seleccionó teniendo en cuenta que el gen T110, que había sido previamente validado experimentalmente (ver III.4. del presente capítulo) se hallaba aproximadamente en esa posición (percentil 72). Este procedimiento tuvo por objeto controlar la tasa de falsos positivos y de falsos negativos, eligiendo dentro de un conjunto relativamente

amplio de genes (considerando que más del 50% tienen p -valores menores a 0,05) aquellos que de acuerdo a su ubicación en el plano de ordenación dado por los dos primeros componentes principales se ubicaban en la periferia de la nube de puntos representando la totalidad de los genes evaluados. Los genes seleccionados como candidatos se presentan en la Tabla 14, clasificados según su función obtenida por homología con bases de datos públicas.

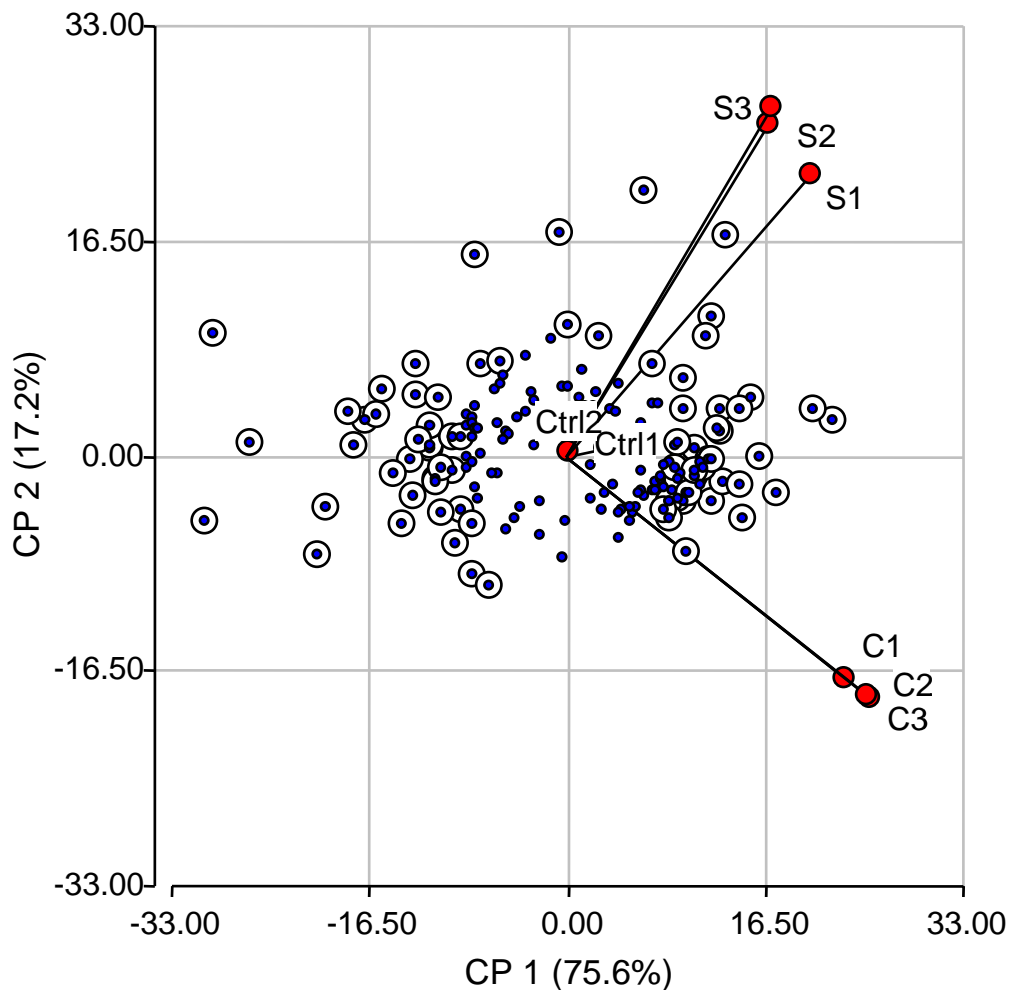


Figura 28: *Bi-plot* que muestra genes con p -valor para un test F menor que 0,05 y con una distancia al origen mayor que el percentilo 70 de la distancia al origen de la distribución (círculos rellenos)

Teniendo en cuenta los grupos de genes identificados por análisis exploratorio y cuya estadística básica se describe en la Tabla 11, los genes listados en la Tabla 14, representan el 38% (42/112) de los genes en el Grupo 1, 4% (5/126) de los genes

del Grupo 2 y el 70% (34/49) de los genes en el Grupo 3. Del total de genes seleccionados 52% corresponden al Grupo 1, 6% al Grupo 2 y 42% al Grupo 3. En cuanto a las frecuencias absolutas por órgano, flor en estadio R1 concentró el 6,3% de los genes más diferencialmente expresados, flor en estadio R4 concentró el 51,8%, hoja el 23,5% y tallo el 24,7% (Tabla 13).

<u>Órgano</u>	<u>Número de genes más diferencialmente expresados</u>	<u>Porcentaje sobre el total (%)</u>
EF	5	6,3
F	37	51,9
H	18	23,5
T	20	24,7
<u>Total</u>	<u>80</u>	100,00

Tabla 13 Frecuencias absolutas de representación por órgano en el número de genes más diferencialmente expresado

c.1) Obtención de conglomerados

Una vez identificados un conjunto de genes candidato es importante reconocer aquellos grupos de genes que presentan perfiles de expresión diferencial. De esta manera, y una vez obtenido los grupos de genes candidato, existen herramientas útiles para agrupar los genes dentro de un número dado de conglomerados. Un punto clave en el análisis de conglomerados es la estimación del número de grupos de genes con patrones diferentes de expresión. Para este propósito, se aplicó nuevamente el algoritmo *k-centroide* de manera secuencial según lo descrito en *IV.3.2.b* (Figura 26), obteniéndose la Figura 29.

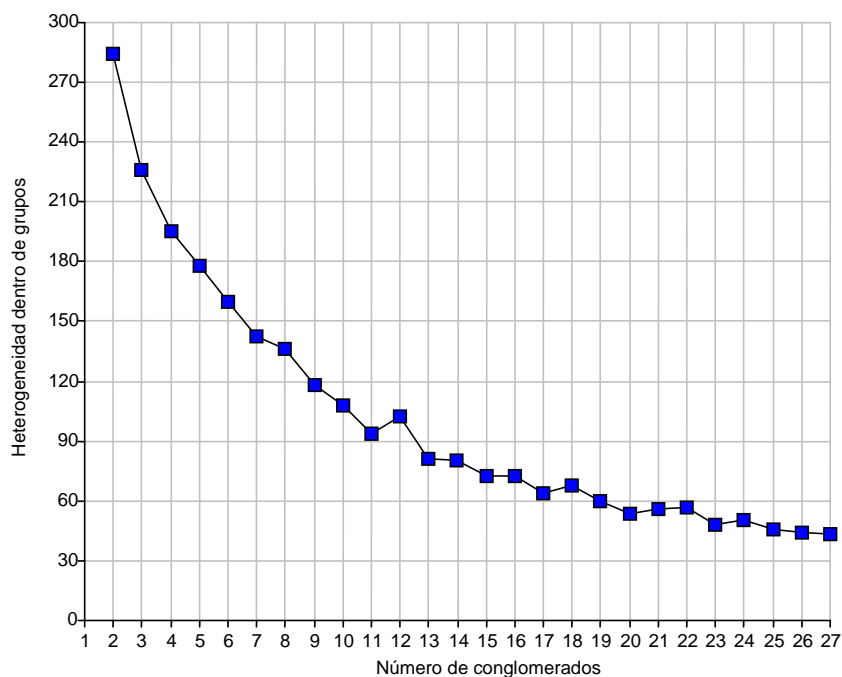


Figura 29: Curva que representa la relación entre el número de conglomerados producidos por el algoritmo de los k-centroides y una medida de heterogeneidad común dentro de conglomerados

Siguiendo idéntico criterio para la selección del número de grupos aplicado en **IV.3.2 b)** (Figura 26) se identificaron 3 conglomerados dentro de los 80 genes candidato. La caracterización de los mismos fue realizada a través de un gráfico de cajas (box-plot) entre las réplicas biológicas (Figura 30).

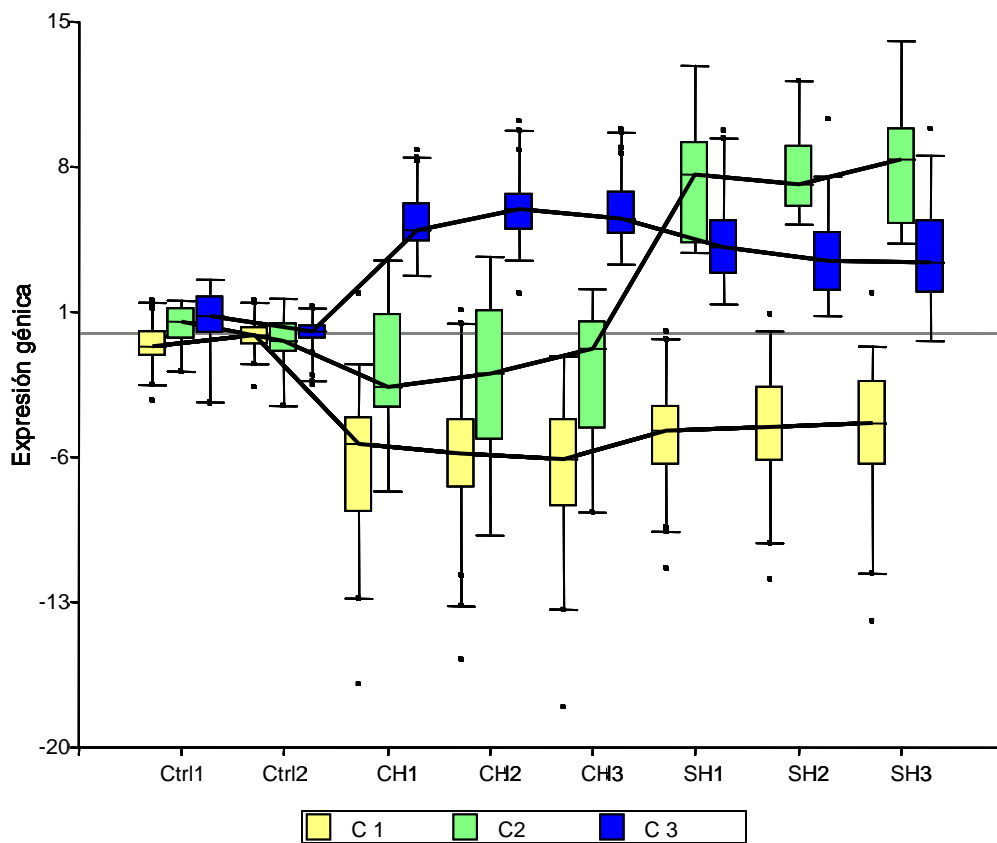


Figura 30: Bandas del confianza al 95% para cada uno de los 3 grupos de genes determinados por el algoritmo *k-centroides* de clasificación no supervisada aplicado a la matriz de expresiones génicas (CH1 y CH2 corresponden a las repeticiones biológicas del control, FH1, FH2 y FH3 las repeticiones del tratamiento con frío y SH1, SH2 y SH3 las correspondientes repeticiones del tratamiento salino. C1: conglomerado 1, C2: conglomerado 2; C3: conglomerado 3)

De este análisis surgieron los siguientes grupos de genes agrupados de la siguiente manera: el conglomerado 1, conformado por 34 genes sub-regulados en frío y salinidad, el conglomerado 2, integrado por 10 genes sobre-regulados en salinidad y el conglomerado 3, con 36 genes que incluye aquellos sobre-regulados tanto en salinidad como en frío. La Tabla 15 presenta dichos genes agrupados por anotación funcional GO mostrando sus niveles de expresión.

Conglomerado 1				Conglomerado 2		Conglomerado 3			
EF432	F401	H136	T234	F171	EF127	F426	F557	T155	
EF502	F443	H209	T253	F192	EF264	F455	F561	T243	
EF624	F543	H302	T289	F202	F176	F467	F572	T283	
F230	F549	H304	T340	H329	F577	F482	F593	T307	
F216	F550	H322		H368	F295	F489	H360	T368	
F137	H110	H354		T107	F305	F491	H406		
F175	H111	H385		T111	F319	F494	T129		
F209	H123	H387		T322	F340	F496	T411		
F210	H124	H411		T187	F373	F514	T120		
F231	H125	T221		T464	F379	F554	T124		

Tabla 14 – Lista de 80 genes seleccionados a través del algoritmo *k-centroide* en tres conglomerados

EST	Clasificación GO	Número de Acceso GenBank	Producto génico	Número de acceso del gen candidato	Frío	Salinidad	Niv el de expresión	
Metabolismo general/Fotosíntesis								
F175	GO:0009522	BU671793	photosystem II 22-K chlorophyll-binding protein – rice	BAA12337.1	SUB	↓	NC	
H136	GO:0005489	BU672102	O04002_SOLTU CDSP32 protein (Chloroplast Drought-induced Stress Protein of 32kDa) general metabolisms	O04002	SUB	↓	NC	
H124	GO:0009765	BU672070	Q9SY97_ARATH PSI type III chlorophyll a/b-binding protein	AAD25555.1	SUB	↓	SUB	↓
H125	GO:0009765	BU672060	(AL132962) Lhca2 protein [Arabidopsis thaliana]	CAB71077.1	SUB	↓	SUB	↓
H302	GO:0009507	BU672078	Magnesium-protoporphyrin monomethyl este	IX Q9M591	SUB	↓	SUB	↓
	GO:0009538	BU672091	SAF_FLATR PHOTOSYSTEM REACTION CENTRE SUBUNIT III PRECURSOR (LIGHT-HARVESTING COMPLEX I 17 KD PROTEIN) (PSI-F) (Flaveria trinerva)	AAA33344.1	SUB	↓	SUB	↓
H385	GO:0042132	BU672090	fructose-1,6-bisphosphatase [Pisum sativum]	5305145	SUB	↓	SUB	↓
T129	GO:0009765	BU671800	Q9SDT2_DAUCA Chlorophyll a/b-binding protein.	AAF20948.1	SOBRE	↑	SOBRE	↑
T340	GO:0004356	BU671843	O04961_HELAN Chloroplastic glutamine synthetase (EC 6.3.1.2) (Fragment) general metabolism	AAN84537.1	NC		SUB	↓
T411	GO:0004332	BU671801	Q9SXX5_NICPA Plastidic aldolase general metabolism	BAA77603.1	SOBRE	↑	SOBRE	↑

IV. RESULTADOS Y DISCUSIÓN

EST	Clasificación GO	Número de Acceso GenBank	Producto génico	Número de acceso del gen candidato	Frío	Salinidad		
Mecanismos de traducción								
F230	GO:0003735	BU671926	Lycopersicon esculentum ribosomal protein L2 protein	CAA45863.1	NC	SUB		↓
F467	GO:0003735	BU672008	Arabidopsis thaliana 60S ribosomal protein L13a protein	BAB09429.1	SOBRE	↑	NC	
F455			Q5I7K2_WHEAT Ribosomal protein S7	Q5I7K2	SOBRE	↑	SOBRE	↑
F550	No GO annotation	BU672017	Translation initiation factor B04 [<i>Helianthus annuus</i>]	21842314	SOBRE	↓	SOBRE	↓
F577	GO:0003723	BU671923	RS4_SOLTU 40S RIBOSOMAL PROTEIN S4 (potato)	AAC24585.1	SOBRE	↑	SOBRE	↑
T111	GO:0003746	BU671873	<i>Solanum lycopersicon</i> elongation factor 1-alpha protein	CAA37212.1	NC		SOBRE	↑
Sin similitud								
H110	No GO annotation	BU672056	Sin similitud en base de datos		SOBRE	↓	SOBRE	↓
Mecanismos de transcripción								
F305	GO:0003677	BU671940	41 kD chloroplast nucleoid DNA binding protein (CND41) [<i>Nicotiana sylvestris</i>]	24430421	SOBRE	↑	SOBRE	↑
F561		BU672034	Zinc finger family protein (Fragment)	Q5ULY2	SOBRE	↑	NC	
T187	GO:0003677	BU671817	MBD4; DNA binding [<i>Arabidopsis thaliana</i>]	15229373	NC	↑	SOBRE	↑
T464	GO:0006118	BU671860	Single-stranded DNA binding [<i>Medicago truncatula</i>]	ABD28355.1	NC		SOBRE	↑
Transducción de señales								
EF127	GO:0003925	BU671885	ADP-ribosylation factor 1.	Q7F270_ORYS A	SOBRE	↑	SOBRE	↑
T234	No GO annotation	BU671830	Extracellular Ca sensing receptor (<i>Arabidopsis thaliana</i>)	Q9FN48	SUB	↓	SUB	↓
Conformación de proteínas/interacciones								
EF264	GO:0004299	BU671886	<i>Arabidopsis thaliana</i> multicatalytic endopeptidase protein	CAA73619.1	SOBRE	↑	SOBRE	↑
F231	GO:0003755	BU671928	Q7XY71_9ROSI Cyclophilin (Fragment)	Q7XY71	SUB		NC	
F443	GO:0003743	BU671999	Copper chaperone [<i>Lycopersicon esculentum</i>]	30039180	SUB	↓	SUB	↓
T124	GO:0005524	BU671806	Q8LKP8_SAUME Heat shock protein	Q8LKP8	SOBRE	↑	SOBRE	↑
T289	GO:0004374	BU671835	Flaveria pringlei P-protein of the glycine cleavage system protein	CAA85353.1	SUB	↓	SUB	↓
Transporte								
	GO:0016068		<i>Gossypium hirsutum</i> lipid transfer protein precursor protein transport	AAF35184.1		↑		↓
EF502		BU671910			SOBRE		SUB	
H360		BU672086	ATP synthase beta subunit [<i>Helianthus annuus</i>]	CAB65035.1	SOBRE	↑	SOBRE	↑
F557	GO:0003936	BU672042	ARATH A_IG002N01.16 protein	AAL07192.1	SOBRE	↑	SOBRE	↑
	GO:0005488		(Putative carrier protein)					
Metabolismo secundario								
H123		BU672069	myo-inositol-1-phosphate synthase (EC 5.5.1.4) - common ice plant	BAA84084.1	SUB	↓	SUB	↓
	GO:0004512							

IV. RESULTADOS Y DISCUSIÓN

EST	Clasificación GO	Número de Acceso GenBank	Producto génico	Número de acceso del gen candidato	Frío	Salinidad
Metabolismo secundario						
H322		BU672083	Q43469_HELAN Delta-8 sphingolipid desaturase	Q9ZRP7	SUB ↓	SUB ↓
T322	GO:0004497	BU671815	Glycine max geranylgeranyl hydrogenase protein.	AAD28640.1	NC	SOBRE ↑
Mecanismos de estrés oxidativo						
H354	GO:0008137	BU672105	NADH-plastoquinone oxidoreductase subunit 4 [<i>Oenothera elata</i> subsp. <i>hookeri</i>]	CAB67220.1	SUB ↓	SUB ↓
		BU671805	Q84LB8_ZANAE Glycolate oxidase (EC 1.1.3.15) (EC 1.1.3.1) ROS machinery	Q84LB8	SOBRE ↑	SOBRE ↑
T120	GO:0008246					
T155	GO:0005489	BU671812	Putative thioredoxin reductase [<i>Hordeum vulgare</i> subsp. <i>vulgare</i>]	38174811	SOBRE ↑	NC
T221	GO:0004301	BU671825	catalytic/ hydrolase [<i>Arabidopsis thaliana</i>] ROS machinery	NP_849507.1	SUB ↓	SUB ↓
T243	GO:0003773	BU671832	HSP7S_PEA Stromal 70 kDa heat shock-related protein, chloroplast precursor	Q02028	SOBRE ↑	NC
T307	GO:0004096	BU671838	<i>Helianthus annuus</i> catalase protein	AAA69866	SOBRE ↑	SOBRE ↑
T368	GO:0009507	BU671846	PSBQ2_ARATH Oxygen-evolving enhancer protein 3-2, chloroplast precursor (OEE3) (16)	Q41932	SOBRE ↑	SOBRE ↑
No clasificadas						
EF432		BU672101			SUB ↓	SUB ↓
EF624		BU671912			SUB ↓	SUB ↓
F137		BU671986			NC	SUB ↓
F171		BU671987			SUB ↓	SOBRE ↑
F176		BU671919			SOBRE ↑	SOBRE ↑
F192		BU671924			NC	SOBRE ↑
F202		BU671945	Q9SL05_ARATH Expressed protein (At2g05620/T20G20.3)	Q9SL05	SUB ↓	SOBRE ↑
F209		BU671949			SUB ↓	NC
F210		BU671950			SUB ↓	NC
F216		BU671927			NC	SUB ↓
F295		BU671931			SOBRE ↑	NC
F319		BU671961			SOBRE ↑	NC
F340		BU671965			SOBRE ↑	SOBRE ↑
F373		BU671980			SOBRE ↑	NC
F379		BU671983			SOBRE ↑	SOBRE
F401		BU671988			SUB ↓	SUB ↓
F406		BU671788			SOBRE ↑	SOBRE ↑
F426		BU672039			SOBRE ↑	SOBRE ↑
F482		BU672013			SOBRE ↑	SOBRE ↑
F489		BU672016			SOBRE ↑	SOBRE ↑
F491		BU672017			SOBRE ↑	SOBRE ↑
F494		BU672019			SOBRE ↑	SOBRE ↑
F496		BU672020			SOBRE ↑	SOBRE ↑
F514		BU672110			SOBRE ↑	SOBRE ↑

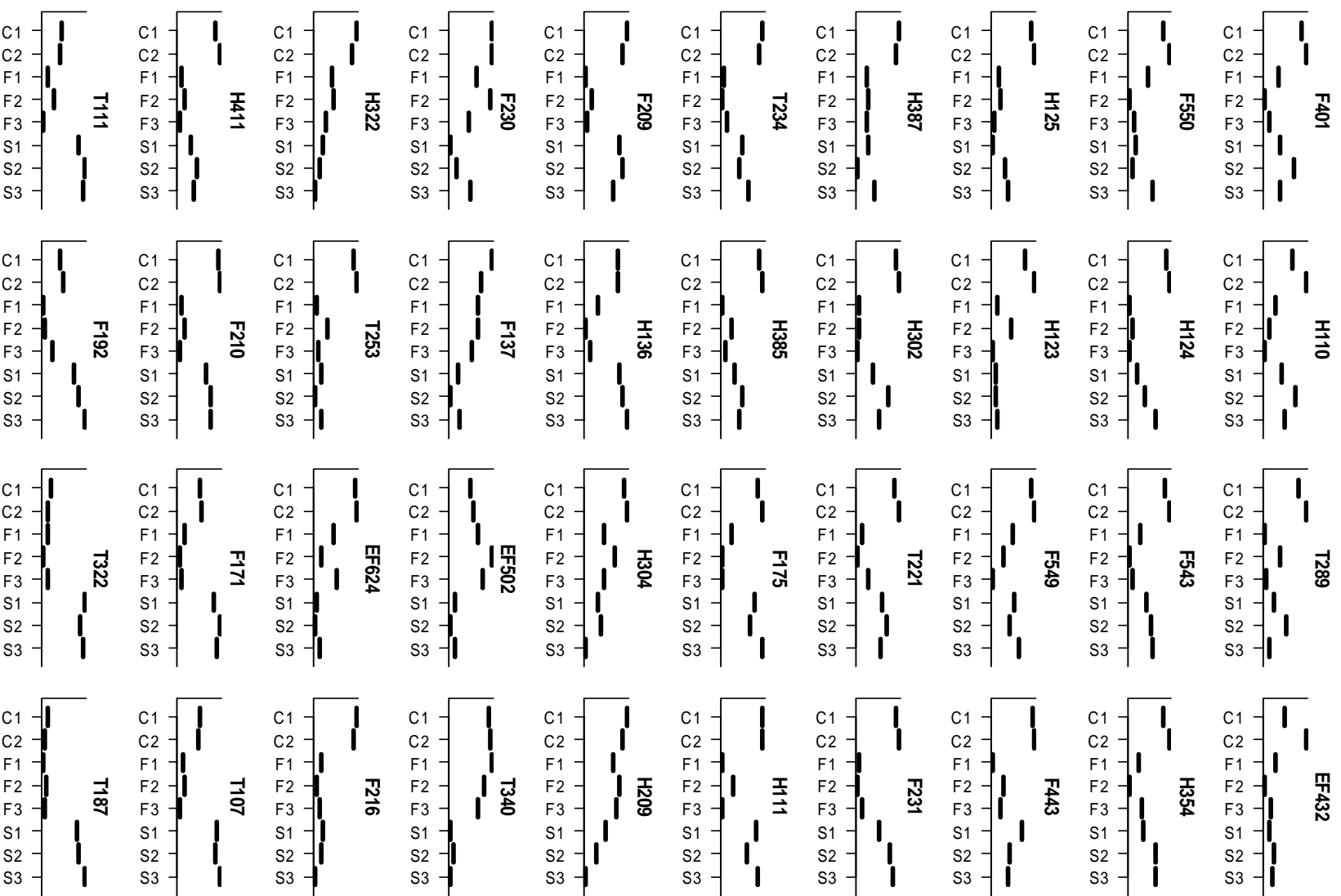
EST	Clasificación GO	Número de Acceso GenBank	Producto génico	Número de acceso del gen candidato	Frío	Salinidad
No clasificadas						
F543		BU672026	similar to unknown protein [Arabidopsis thaliana]	BAB02703.1	SUB ↓	SUB ↓
F549		BU672028			SUB ↓	SUB ↓
F554		BU672032			SOBRE ↑	SOBRE ↑
F572		BU672046			SOBRE ↑	SOBRE ↑
F593		BU672051			SOBRE ↑	NC
H111		BU671789			SUB ↓	NC
H209		BU672074			NC	SUB ↓
H304		BU672079			NC	SUB ↓
H329		BU672084			NC	SOBRE ↑
H368		BU672087			SOBRE ↑	SOBRE ↑
H387		BU672092			SUB ↓	SUB ↓
T107		BU671799			SUB ↓	SOBRE ↑
T253		BU671868			SUB ↓	SUB ↓
T283		BU671875	Os03g0651000 [Oryza sativa (japonica cultivar-group)]	115454377	SOBRE ↑	NC
T479	GO:0016740	BU671862			SOBRE ↑	SOBRE ↑

Tabla 15: Lista de los 80 genes más importantes en la separación de tratamientos, de acuerdo a su p-valor en la prueba de comparación de medias por análisis de la varianza y su ubicación en la región periférica del plano de ordenación generado por los dos primeros ejes principales de un ACP de la matriz de expresión génica escalada gen a gen por la desviación estándar residual del análisis de la varianza. SOBRE: sobre-expresión; SUB: sub-expresión; NC: sin cambio

IV.3.3. Perfiles de expresión por gen candidato

Para cada uno de los 80 genes seleccionados como genes candidato se graficó el perfil de expresión según su comportamiento en cada tratamiento (Di Rienzo J., Comunicación Personal). (Figura 29). También se construyó el gráfico de colores (“*heatmap*”) asociado a ese patrón diferencial de modo de analizar cualitativamente el gradiente diferencial por tratamiento y por gen (Figura 30). De estos resultados surgieron genes candidato de interés para su análisis de expresión diferencial, de los cuales diez fueron seleccionados para validar mediante la técnica de PCR cuantitativa (qRT-PCR) y/o *northern blot*, según se detalla en III.3. Entre ellos figuran los siguientes unigenes de la colección: EF127, EF264, EF502, H360, F171, F455, F550, F446, T110 y T340. Estos genes candidato presentan un perfil de expresión diferencial y variable según los tratamientos, como puede verse en las Figuras 31 y 32. En estas figuras podemos observar otra manera de presentar la

información obtenida a través de gráficos que muestran perfiles de expresión de cada gen candidato por funcionalidad hipotética asignada *in silico*



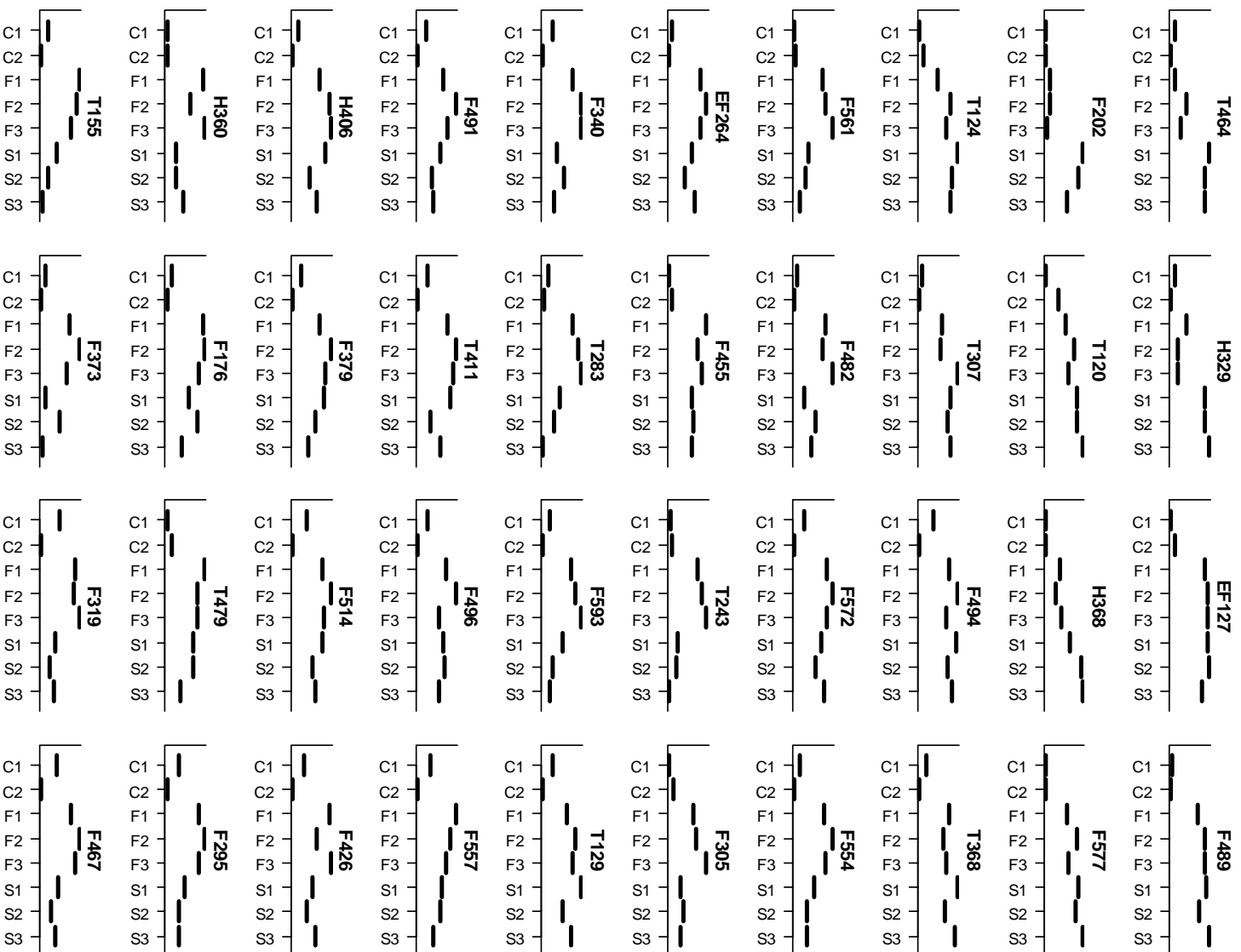


Figura 31 : Perfil de expresión de los 80 genes candidato
C1 = control hoja 1, C2 = control hoja 2, F1 = frío hoja 1, F2 = frío hoja 2, F3 = frío hoja 3, S1:
salinidad hoja 1, S2 = salinidad hoja 2, S3 = salinidad hoja 3

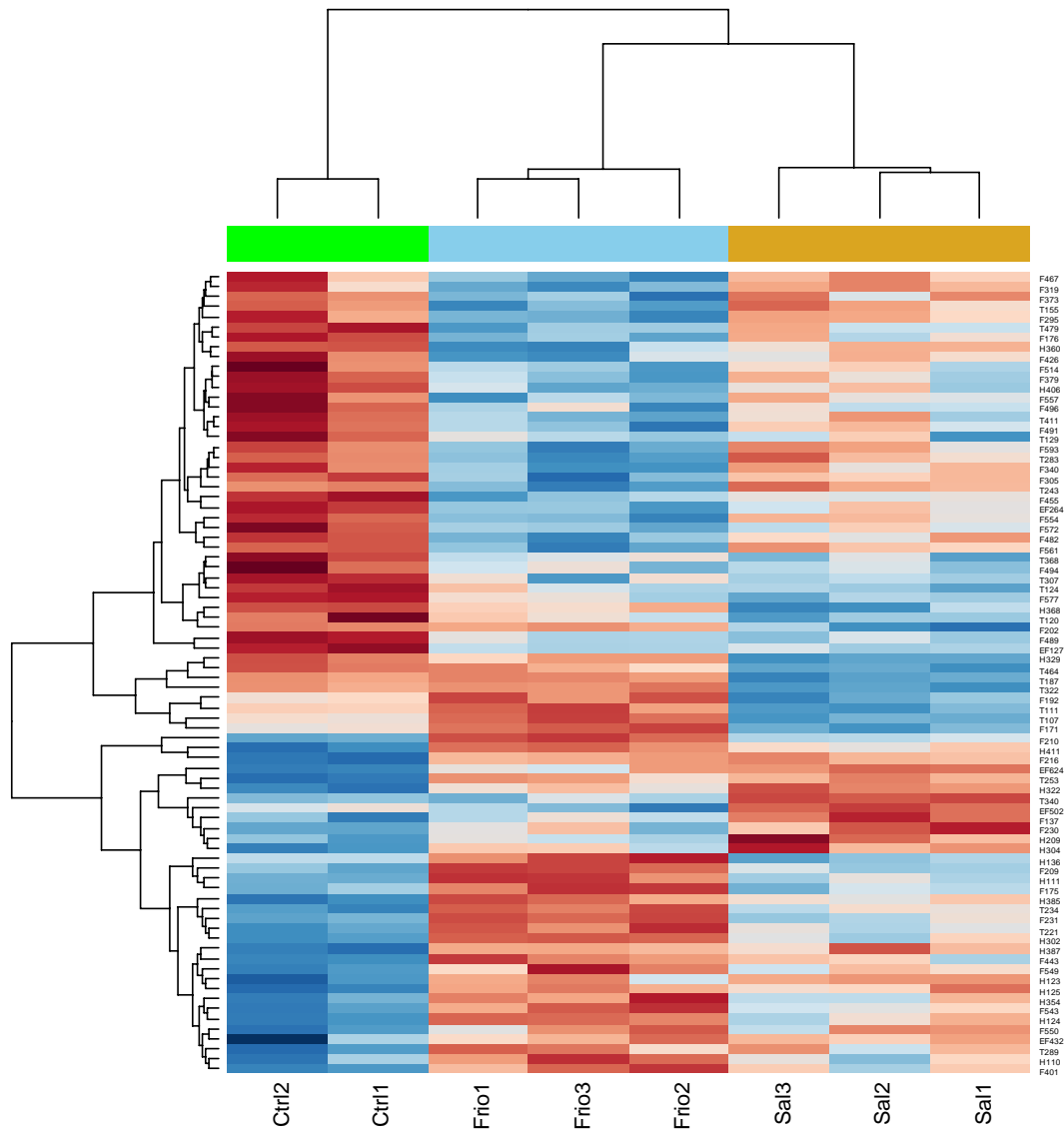


Figura 32: Gráfico de conglomerados de los 80 genes candidato
C1 = control hoja 1, C2 = control hoja 2, F1= frío hoja 1, F2 = frío hoja 2, F3 = frío hoja 3, S1:
salinidad hoja 1, S2 = salinidad hoja 2, S3 = salinidad hoja 3

IV.3.4. Validación por técnicas de PCR cuantitativa (qRT-PCR)

La PCR cuantitativa es, básicamente, una PCR convencional en la que los equipos de amplificación (termocicladores) llevan incorporados un sistema de detección de fluorescencia (fluorímetro), basándose la tecnología en la utilización de unas moléculas específicas denominadas fluoróforos e interferentes (*quenchers*). Ambos factores permiten monitorear, en tiempo real, lo que está ocurriendo dentro de cada tubo en cada ciclo de amplificación sustituyendo a los pasos de amplificación, electroforesis y análisis de imagen de una PCR tradicional.

Los factores que afectan a la PCR cuantitativa (qRT-PCR) son, entre otros:

- optimización del protocolo de reacción para maximizar su eficiencia. Hay que tener en cuenta que una pequeña diferencia en la eficiencia de reacción por ciclo, puede originar una substancial diferencia en la cantidad final de producto; por ejemplo, si una reacción es tan solo un 90% tan eficiente como otra, la relación de producto final entre ambas reacciones después de 40 ciclos será alrededor de 65 a 1
- evitar un número excesivo de ciclos de amplificación
- un adecuado procesamiento de las muestras que asegure la eliminación de los inhibidores que pueden disminuir la eficiencia de amplificación

El parámetro fundamental en una PCR cuantitativa y en función del cual se van a realizar todos los cálculos analíticos y obtención de resultados, es el denominado Ciclo Umbral (*Threshold cycle* o Ct), que se define como el ciclo a partir del cual la fluorescencia es estadísticamente significativa por encima del ruido de fondo (*background*).

IV.3.4.i) Validaciones de clones por PCR cuantitativa (PCR en tiempo real)

Se validaron por PCR cuantitativa los ESTs EF127 (BU671885), EF264 (BU671886), F171 (BU671987), F379 (BU671983), F443 (BU671999), F455 (BU672004) y H360 (BU672086) que mostraron un perfil contrastante según las

Figuras 29 y 30. Como paso previo a la utilización de ellos como templado, se probaron diluciones seriadas al décimo de los clones conteniendo los fragmentos de ADNc que actuaron como sustrato de reacciones de PCR. Las reacciones de amplificación para las muestras elegidas fueron amplificadas por duplicado así como también para el gen endógeno de actina utilizado como control de expresión basal. Los controles negativos que se incluyeron fueron reacciones sin templado de ADNc para cada par de oligonucleótidos ensayado, incluyendo actina de girasol (AN: [AAF82805](#)). Asimismo se incluyeron los controles negativos de síntesis de ADNc, es decir, utilizando ARN como templado para descartar producto de amplificación por ADN remanente en la extracción de ARN total.

Además, se hizo una titulación de los oligonucleótidos en una concentración final de 100, 200 y 300 nM para evaluar las distintas curvas de amplificación, siendo la concentración de 100 nM final la de mejor eficiencia para estos oligonucleótidos.

La cuantificación de los cambios relativos en la expresión génica se realizó utilizando el método $2^{-\Delta\Delta CT}$ según lo descrito por Livak y Schmittgen (2001).

En la Figura 33 podemos observar para el EST F171 la sobre y sub-regulación respecto del control para los dos tratamientos de estrés abiótico en un gráfico de reporte normalizado de fluorescencia (Rn) en función del ciclo, validando de esta manera experimentalmente el perfil de expresión representado en la Figura 31.

IV.3.5. Envío de datos de micromatrices a una base de datos de reposición pública

Los datos de micromatrices de ADNc generados en este trabajo fueron enviados a la base pública de reposición GEO (Gene Expression Omnibus) del NCBI (<http://www.ncbi.nlm.nih.gov/geo/>), encontrándose pronto su liberación con fecha 25 de marzo de 2007 (Fernández y col. 2006, GEO NCBI GPL: 4366).

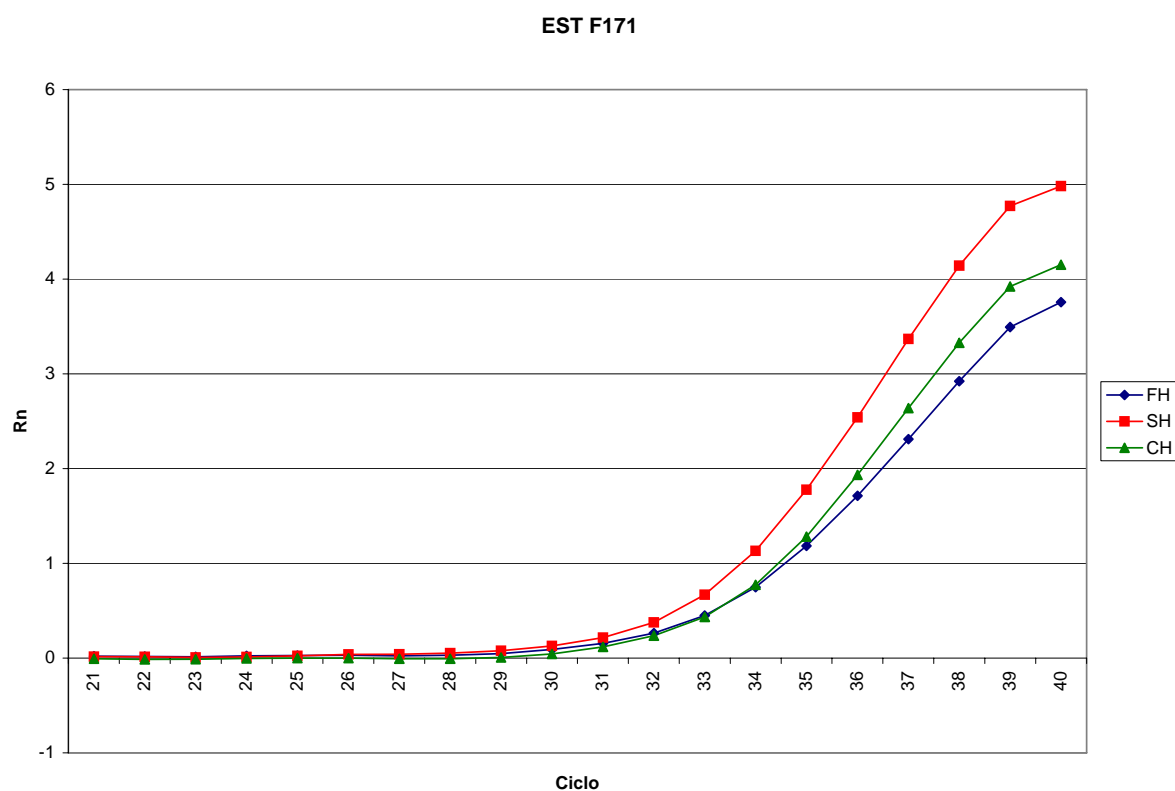


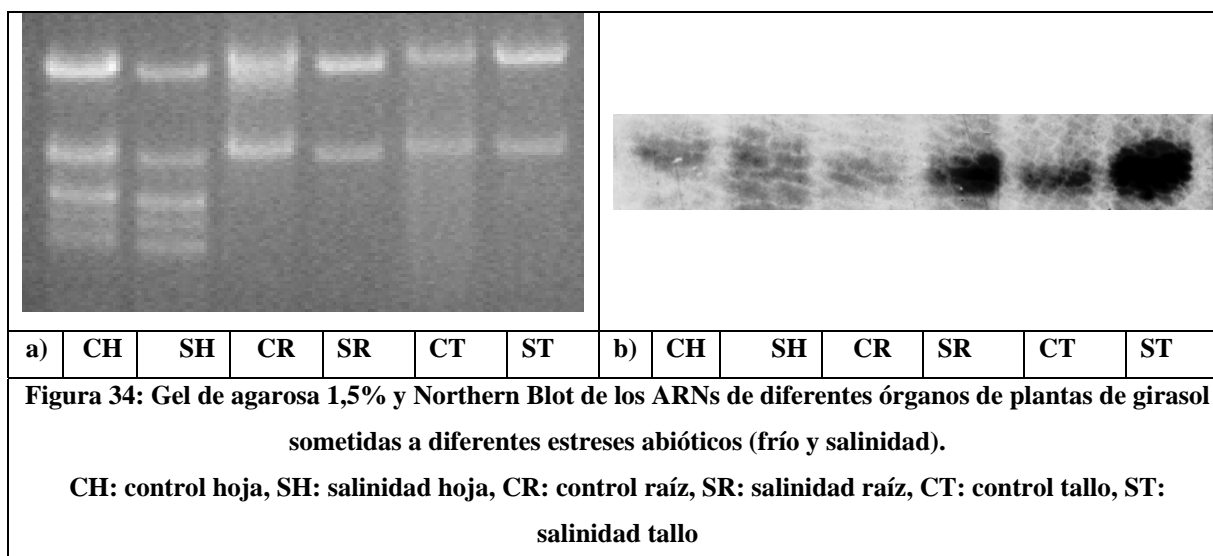
Figura 34: Gráfico de radio normalizado en función del ciclo para el gen candidato F171
 FH = frío hoja 1, CH = control hoja, SH = salinidad hoja

IV.3.6. Validación por técnica de Northern Blot

En paralelo al análisis de PCR cuantitativa, se realizaron ensayos de Northern Blot para validar algunos genes candidato.

Se validó el gen candidato T110, correspondiente a una aldolasa de localización plástica tipo I (proteína AldP de *Nicotiana sp.*), la cual se sobreexpresa en condiciones de salinidad, según fue publicado por Yamada y *col.*, 2000 como la primera demostración de un estrés abiótico asociado a una proteína de este tipo.

En la Figura 34 a) se muestra el gel de agarosa teñido con BrEt que se usó para transferir a membrana de nylon y realizar el Northern Blot representado en la Figura 34 b). En el caso de hoja se observan las bandas de ARN de cloroplastos.



IV.3.7. Discusión

La información obtenida a partir del análisis de las micromatrices de cDNA fue de valiosa utilidad no sólo para detectar posibles genes candidato sino también para estudiar y analizar el perfil de expresión de ellos a través de los diferentes estreses abióticos en estudio en este trabajo de Tesis. En particular, es interesante remarcar que dada la estrategia utilizada en el diseño de los arreglos de ADNc a partir de ESTs de clonotecas organo-especificos diferenciales, un alto porcentaje de los genes analizados (27,8%) presentaron niveles de transcritos modificados significativamente en condiciones de estrés respecto a otros trabajos que analizan micromatrices que representan ADNc de colecciones no diferenciales (Seki y col. 2002).

Analizando los resultados precedentes correspondientes a esta información y sus gráficos aclaratorios, se observa que los niveles de expresión de los genes analizados presentaron una tendencia marcada y definida que permite clasificarlos en dos grandes grupos: aquellos que mantuvieron su nivel de expresión estable y aquellos que modificaron su nivel de expresión en presencia de un estrés abiótico, siendo *a priori*, difícil separar los genes asociados a respuestas a estrés por frío de aquellos asociados a respuestas a estrés por salinidad. Esto es consistente con lo publicado por Zhu J.K., 2001b, donde se menciona que algunas de las respuestas

inducidas por estrés salino están íntimamente relacionadas al estrés por sequía y por frío, siendo el estudio de los mecanismos de tolerancia a salinidad, por lo tanto, necesarios para entender los mecanismos de tolerancia cruzada frente a otros estreses abióticos. Tanto el estrés por sequía, salinidad o bajas temperaturas reducen la disponibilidad de agua para las células lo que genera una disminución en el potencial agua. Las plantas reaccionan con una serie de respuestas a esta disminución del potencial agua que incluyen la acumulación de solutos y la modificación de componentes de la pared celular entre otros para evitar deshidratación, síntesis de proteínas protectoras e inducción de mecanismos de tolerancia a deshidratación ocasionada por cualquiera de los tres estreses mencionados, evitando o reparando el daño celular ocasionado (Verslues y col 2006). La activación de estas respuestas requiere de una compleja red de señalización y transducción de señales algunas de ellas comunes a varios estreses y otras específicas de un tipo de estrés en particular como aquellas asociadas con las secuencias DREB/CBF (Chinnusamy y col 2003; Zhu, 2001). Las plantas en el campo no se encuentran sometidas a un solo tipo de estrés sino a una suma de ellos (bióticos y abióticos) lo que genera respuestas de tipo complejas (Agarwal y col. 2006; Fujita y col. 2006).

En el presente trabajo se desarrolló un primer trabajo de análisis exploratorio utilizando el algoritmo *k-centroides* de clasificación no supervisada que permitió determinar la existencia de tres grandes grupos de genes conformados por aquellos que no modificaron su expresión respecto del control (CH) (Grupo 2: 126 genes), aquellos que se sobre-expresaron respecto del control cuando las plantas fueron sometidas a estrés por frío (FH) o salinidad (SH) (Grupo 1: 112 genes) y el resto de los genes formado por aquellos que se sub-expresaron respecto del control cuando fueron sometidos a estrés por FH o SH (Grupo 3: 49 genes). Este primer análisis exploratorio de los datos mediante biplot, donde cada gen fue representado por un vector del gráfico con diferente dirección de acuerdo al tratamiento de estrés abiótico aplicado, permitió abrir el gran grupo de genes asociado a estreses abiótico, detectándose respuestas específicas de cada gen frente a cada tratamiento. A partir de aquí, y mediante el análisis de componentes principales (ACP) se logró detectar

el número final de genes candidato que presentaban cambios frente a estos estreses ambientales (Figura 30). Esta metodología de análisis fue aplicada con éxito al análisis de otros trabajos de micromatrices de ADNc como el reportado por Hewezi y col 2006. Con toda esta información se evaluaron los perfiles de expresión por gen candidato en función de su función hipotética predicha por análisis *in silico* obteniendo un análisis funcional preliminar. Como se infiere a partir de este detalle de análisis de mayor escala (grandes grupos) a menor (gen) podemos concluir que existen muchas herramientas de gran utilidad para la interpretación y asistencia en el análisis de micromatrices. Una de las más comúnmente utilizadas es el gráfico de conglomerados, presentado en la Figura 32, el cual consiste de pequeñas celdas a las cuales se les asigna un color que representa el nivel relativo de expresión (Allison y col. 2006). Estos gráficos, comúnmente denominados “*heatmaps*” surgen de un análisis de conglomerados entre las muestras y los genes. Frecuentemente, las filas representan niveles de expresión similares, mientras que las columnas indican las diferentes muestras biológicas. Si bien este tipo de gráficos es fácil de analizar visualmente, carece de exactitud debido a la subjetividad visual en la escala de colores. Es por ello que en la actualidad se tienden a usar los gráficos “*heatmaps*” en asociación con los gráficos de perfiles intrínsecos por gen (Lemaire-Chamley y col. 2005), ambos presentados en las Figuras 31 y 32 de III.1.2. En este trabajo en particular fue de mayor utilidad el análisis de genes por perfil funcional, ya que una vez detectados los perfiles de expresión gen a gen, se pudo detectar un gradiente en el nivel de sobre/sub-expresión y asociar cada gen candidato a una clase funcional con potencial implicancia en los mecanismo de respuestas a estreses abióticos según trabajos previos descritos para distintas especies.

IV.3.7.1. Diferencias entre frío y salinidad

Las diferencias entre el control y los tratamientos es tan amplia que mediante el uso del algoritmo *k-centroide* solo es posible detectar aquellos genes inducidos por ambos tratamientos de aquellos detectados en las plantas sin tratar, siendo aún más

desafiante detectar específicamente genes sub o sobre-expresados en frío y en salinidad. A pesar que los mecanismos de aclimatación a bajas temperaturas es aún poco conocido en girasol, estudios recientes sugirieron que las plantas de girasol no generan mecanismos de aclimatación bajo determinadas condiciones experimentales (Hewezi y col. 2006). Si bien la técnica de micromatrices permite la comparación de un gran número de genes que se inducen o reprimen bajo ciertas condiciones ambientales, se debe ser ser cauto y minucioso en la interpretación y análisis de los resultados obtenidos. Por ejemplo, el estrés salino produce una rápida caída en la tasa de fotosíntesis en minutos (Kawasaki y col., 2001). No obstante, esto no significa que los genes involucrados en la maquinaria fotosintética estén regulados *per se* por dicho tratamiento. Reymond y col. 2000 analizaron la expresión de 150 genes en respuesta a heridas, inoculación con patógenos y sequía y concluyeron que la mayoría de los genes inducidos por herida se encontraban también sobre-regulados por sequía, por lo que estos genes se encontrarían primeramente inducidos por sequía que por un dano mecánico. No obstante, el dosaje de estrés y el tiempo de exposición al mismo son factores de interés a tener en cuenta (Xiong y Zhu, 2002).

IV.3.7.2. Análisis de expresión por funcionalidad hipotética asociada

La Tabla 14 resume el perfil de los 80 genes candidato identificados en respuesta a frío y salinidad por funcionalidad hipotética de acuerdo con el criterio de mejor puntaje por similitud de secuencia utilizando el algoritmo BLAST (Altschul y col. 1990) y anotados bajo terminología GO (Gene Ontology) (Ashburner y col. 1990). Dentro de estos genes, 48 fueron sobre o sub-regulados bajo ambos estreses reportándose mecanismos de regulación comunes en respuesta a frío y salinidad, 17 genes fueron específicamente sobre-regulados mientras que 12 fueron sub-regulados en un estrés y no en el otro. De manera interesante 36 genes (43,7%) de los genes candidatos se reportaron como no clasificadas por no tener similitud a ninguna secuencia/gen público detectado en bases de datos

a) Metabolismo general/Fotosíntesis

Existen reportados en la literatura variados trabajos de diferentes senales inducidas por estreses abióticos tales como frío, salinidad y sequía que afectan el metabolismo general de las plantas (Hasegawa y col. 2000; Chen y col 2002; Hewezi y col 2006b; Sahi y col 2006; Dhanaraj y col. 2006) y como expresión primera de éste, el proceso fotosintético. En esta sección del trabajo reportamos tres genes con comportamiento diferencial: el primero de ellos sobre-regulado bajo ambos estreses (ESTs T411, BU671801, similar a una aldolasa de localización plastídica), el segundo (T340, BU671843, similar a una glutamina sintasa cloroplástica) fue sub-regulado bajo estrés por frío, mientras que el tercero (H136) no presentó cambios en su nivel de expresión bajo frío pero fue sub-regulado bajo estrés salino. Los genes asociados a aldolasas plastídicas caracterizados en plantas de género *Nicotiana* fueron agrupados en dos familias: AldP1 y AldP2. Yamada y col. 2000 reportaron por primera vez que AldP2 era sobre-regulada or estrés salino, mientras que AldP1 se reprimía ante dicha condición ambiental. Además, y de manera típica, los caminos metabólicos que llevan a la fijación de CO₂ y captura de luz se encuentran reprimidos bajo estrés abiótico, aunque hay evidencia de sobre-expresión de una glutamina sintasa para mejorar la tolerancia a salinidad en plantas (Hoshida y col. 2000). Por otro lado, el EST H136 (BU672102) fue sub-regulado para ambos estreses.

La exposición a bajas temperaturas en combinación con altos niveles de irradiancia causa una rápida inhibición en de la fotosíntesis en un amplio rango de plantas que incluyen al tomate, zapallo y maíz. Algunos de estos elementos que contribuyen a la inhibición fueron identificados por Martino-Catt y Ott, 1992. El dano producido se encuentra bien documentado (Kee y col. 1986; Percival y col 1987) mientras que para especies moderadamente sensibles tales como el maíz sería la principal causa de la abrupta caída en la tasa fotosintética. No obstante, en las plantas más sensibles al frío como tomate, la reducción en la activación de las bifosfatasa de estroma parece ser el factor dominante que limita la asimilación de carbono en estas condiciones (Sassenrath y col., 1991). Por otro lado, las bajas temperaturas durante

la noche pueden causar severas reducciones en la fijación de CO₂ durante el día luego del estrés por frío. Es así que con toda esta información podemos explicar por qué los genes que codifican productos asociados a metabolismo general se encontraban sub-regulados bajo frío y salinidad en este análisis, siendo los más afectados entre ellos las proteínas fotosintéticas, las proteínas de pegado a clorofila, la enzima Rubisco y las proteínas asociadas a la captura de luz. De acuerdo con Hewezi y col. 2006, y dada la característica del girasol de ser una planta medianamente tolerante a estrés por frío (Kratsch y col. 2000), se sugiere que la caída en el metabolismo general y fotosintético es uno de los procesos celulares asociados a la baja de temperaturas en esta compuesta, además que la exposición por cortos períodos de especies tolerantes y sensibles al frío resulta en una disminución de la fotosíntesis y una sub-regulación de los genes asociados a dicho proceso (Martino-Catt y Ott, 1992; Kreps y Simon, 1997). En contraposición, cuando las hojas de *Arabidopsis* y otras plantas herbáceas tolerantes al frío desarrollan a una temperatura de crecimiento de 5 °C, mostraron una recuperación marcada en la actividad fotosintética. En girasol demostramos que este proceso no se encuentra solo sub-regulado bajo condiciones de frío sino también bajo estrés salino. El EST H385 (BU672090), similar a una fructose-1,6 bifosfatasa, enzima clave asociada al ciclo de reducción del carbono por fotosíntesis y extremadamente sensible al peróxido de hidrógeno generado ante estrés abiótico, fue también sub-regulado para ambos estreses (Charles y Halliwell, 1981). Ghosh (2001) también reportó a esta enzima como uno de los posibles sitios de inhibición de la actividad fotosintética bajo estreses ambientales, ya que una reducción de ella conlleva a una disminución en la tasa de dicho proceso (Zrenner y col. 1996). Poormohammad Kiani. y col. 2006 publicaron una sub-regulación de fructose-1,6 bifosfatasa bajo estrés por sequía en girasol debido fundamentalmente al cierre de estomas post-déficit hídrico. En adición, Seeman y Sharkey (1982) observaron que NaCl reduce la actividad fotosintética en *Phaseolus vulgaris* independientemente del cierre de estomas por reducción del pool de Rubisco, a pesar que la disminución en la fotosíntesis en respuesta a la salinidad ha sido atribuida a un efecto de control e inhibición de apertura de estomas (Walker y col. 1982). El ciclo de Calvin es el

principal camino afectado por deshidratación y estrés salino en plantas sensibles a como el girasol, no estanto aún elucidados los mecanismos que afectan este ciclo metabólico bajo condiciones de frío. Resulta interesante que la misma enzima pero aislada del arroz halofítico salvaje, *P. coarctata*, no se encuentra inhibida por salinidad pudiendo deberse a propiedades estructurales intrínsecas que sobrepasan dicha inhibición (Ghosh y col 2001). Lundmark y col. 2006 reportaron que durante la aclimatación al frío existe un aumento en la disponibilidad de fósforo inorgánico (Pi) e intermediarios fosforilados tanto en ciclo de Calvin como en la vía de síntesis de sucrose, lo que genera un aumento en las capacidades enzimáticas en ambos procesos. Una consecuencia de estos cambios ocurridos en la disponibilidad del Pi de citosol y en la capacidad de síntesis de la sucrosa sería el bombeo de muchas cantidades de carbono or fuera del cloroplasto vía un transportador de triosa-fosfato, reduciendo la capacidad del ciclo de Calvin de regenerar Rubisco e inhibir fotosíntesis. Sin embargo, las especies tolerantes al frío como *Arabidopsis thaliana* y los cereales de invierno son capaces de recuperar la capacidad fotosintética y recuperar su crecimiento luego de varios días o semanas a bajas temperaturas a través del proceso de aclimatación (Levitt 1980; Hurry y Huner 1991; Strand y col. 1997), existiendo en la actualidad poca evidencia acerca de cómo el girasol reacciona los días siguientes al estrés por frío.

b) Mecanismos de transcripción y transducción de señales

La activación de la maquinaria de transcripción que regula la expresión de genes salinidad-específicos requiere la inducción de factores de transcripción (TFs) específicos así como también de ARN polimerasas (Estruch, 2000). Existe un número amplio de proteínas regulatorias, identificadas principalmente a través del chip de *Arabidopsis thaliana* (Chen y col. 2002) que mostraron cambios en TFs bajo determinados estreses ambientales. En este análisis, y de acuerdo con lo publicado por Hewezi y col. 2006, se pudo identificar un fragmento de una proteína perteneciente a la familia *zinc finger* inducido en bajas temperaturas. Rodriguez Milla y col. 2006 también reportaron una proteína de este tipo sobre-regulada asociada a estrés salino en *Arabidopsis* diferente de DREB1A publicada por

Maruyama y col. 2004. Muchos genes candidato fueron identificados como TFs o receptores para senales de calcio por analisis *in silico* que mostraron la abundancia relativa de genes asociados a estos procesos en clonotecas de ADNc órgano-específicas desarrolladas por nuestro grupo (Fernández y col. 2003).

Sumado a ello y considerando la enorme cantidad de TFs presentes en *Arabidopsis* posible causa de la complejidad del metabolismo secundario en plantas ([Szathmáry y col., 2001](#)) podríamos explicar la magnitud de la implicancia de dichas proteínas en las respuestas a estrés abiótico más allá de la interacción entras las plantas y el ambiente junto al nivel de duplicaciones que se encuentren en el genoma de *Arabidopsis* ([Arabidopsis Genome Initiative, 2000](#)).

De acuerdo con Mahalingam y col. 2003, las colecciones de EST analizadas por la aproximación de *northern digital* representaron un 13% de los genes asociados a procesos de transducción de senales. Sumado a ello, las quinasas también fueron reportadas como altamente expresadas bajo aclimatación por frío principalmente. Esta información fue presentada para plantas de *Medicago sativa* analizadas 10 minutos después de exposición a bajas temperaturas (Thomashow, 2001). Zhu (2001) reportó en *Arabidopsis thaliana* la identificación de una proteina quinasa calcio-dependiente activada durante estrés salino. No obstante las respuestas a frío no están relacionadas con las respuestas a salinidad, sequía o regulación por ABA, se pudo detectar una cascada transcripcional que incluyen genes activados por TFs sometidos a bajas temperaturas: CBF y DREB1 (Zhu 2001, Maruyama y col. 2004). Esta respuesta fue también reportada en arroz por Dubouzet y col. 2003).

c) Mecanismos de traducción

De manera general los estreses bióticos y/o abióticos regulan este paso (Tabla 14) como un mecanismo de protección contra enzimas claves de procesos metabólicos. La regulación de la maquinaria de traducción es considerada un componente integrall en la respuesta celular a los estreses (Wood y Oliver, 1999; Dhanaraj y col. 2006). Wool (1996) demostró que las proteínas ribosomales no son solo centrales en lo que a eficiencia de traducción compete sino que también poseen una function extra-ribosomal. En el presente trabajo pudimos confirmar esto por

encontrarse dichas proteínas sobre-reguladas. También pudimos identificar factores de elongación que no se sobre-regulaban por estrés salino según fue reportado previamente en varios sistemas (Sahi y col. 2006; Harding y col. 2002).

d) Degradación de proteínas/Plegamiento/Interacción proteína-proteína

La degradación de proteínas durante un proceso de estrés es un fenómeno altamente conservado y regulado en todos los organismos reportados a la actualidad (Vierstra y Callis 1999). En este análisis, el EST F443 (AN: BU671999) similar a una chaperona de cobre aislada de tomate fue sub-regulada bajo frío y salinidad, según lo reportado previamente por Cui y col. 2005 por un análisis MALDI-TOF realizado en plántulas de arroz sometidas a frío. Dicha chaperona parece jugar un rol muy importante en la movilización del cobre desde órganos vegetativos hacia órganos reproductivos, contribuyendo al desarrollo (Mira y col. 2001). Esto explicaría la sub-regulación reportada en girasol en este análisis para hoja. Sumado a ello, el EST F231 (AN: BU671928) similar a una ciclofilina fue también sub-regulada bajo ambos tratamientos.

Jiang y Deyholos (2006) identificaron en plantas de *Arabidopsis* tratadas con NaCl en solución tres peptidil-prolil-isomerasas (PPI), siendo una de ellas muy similar a una ciclofilina sub-regulada bajo estrés salino. Contrariamente, el EST T124 (AN: BU671806) similar a una proteína de shock térmico (Hsps) fue sobre-regulada. Estas Hsps, también llamadas proteínas del estrés, son actualmente reconocidas dentro de un rango de funciones celulares y procesos metabólicos bajo situaciones normales de crecimiento y bajo condiciones de estreses ambientales diferentes del shock térmico (Grove 2002). El concepto de chaperonas moleculares que fue expuesto a mediados de los años 80, evolucionó a partir del trabajo de bioquímicos y biólogos celulares, siendo varias Hsps actualmente adjudicadas al grupo funcional de chaperonas (chaperoninas). Renaut y col. 2004, identificaron por análisis proteómico en álamo varias Hsps sobre-reguladas bajo estrés por frío, siendo una de ellas similar a una de *Helianthus annuus*.

e) Mecanismos de estrés oxidativo

El estrés por frío, salinidad y sequía, combinados con condiciones de alta luminosidad, resultan en una alta producción de especies reactivas de oxígeno (ROS) mediante el aparato fotosintético ya que las condiciones que limitan la disponibilidad de CO₂ para la reacción en oscuridad dejan al oxígeno como uno de los principales productos reducidos de la fotosíntesis (Mittler R. 2002). Sequía, salinidad y frío inducen la acumulación de especies reactivas de oxígeno tales como superóxido, peróxido de hidrógeno y radicales hidroxilos (Hasegawa y col., 2000). El peróxido de hidrógeno H₂O₂ se genera en los peroxisomas por la acción enzimática de la glicolato oxidasa (Mittler y col. 2004). En este estudio, el EST T120 (AN: BU671805) detectado como sobre-regulado bajo ambos estreses se encuentra probablemente involucrado en la generación de especies reactivas de oxígeno. Dichos ROS podrían ser señales que inducen ROS scavengers y otros mecanismos de protección, así como agentes de dano, contribuyendo al estrés por heridas en plantas (Prasad y col., 1994). Los ESTs generados a partir de hoja y tallo de las clonotecas de ADNc que codifican para peroxidasas, tioredoxinas, catalas y mostraron cambios transcripcionales en respuesta a los estreses estudiados. La mayor parte de estas proteínas fueron sobre-reguladas en ambos estreses debido a una acumulación de esos productos a lo largo del estrés oxidativo. Sin embargo, una NADH-plastoquinona reductasa y una hidrolasa catalítica fueron sub-reguladas. De acuerdo con lo expuesto por Kawasaki y col. 2001, los genes que codifican proteínas asociadas a homeóstasis celular (respiración, génesis celular y reparación de DNA) muestran una caída bajo la exposición a estreses abióticos.

f) Transporte

La homeostasis iónica durante el estrés salino se encuentra afectada por el flujo de sodio, transporte y compartimentalización (Sahi y col. 2006). En este trabajo, el EST H360 (AN: BU672086) similar a una ATP sintetasa fue sobre-regulado en ambos estreses, como ocurre con el EST F557 (AN: BU672042) con similitud a una

proteína transportadora. Esto sería consecuencia de la gran actividad generada por el flujo de iones transportadores durante la tolerancia a salinidad y nutrición del potasio (Blumwald y Grover, 2006). Los transportadores de potasio podrían funcionar en el transporte del catión, el cual es un cofactor esencial de muchas enzimas (Hasegawa y *col.* 2000); o controlar la absorción de K^+ y Na^+ , pudiendo ser un determinante importante de la tolerancia a la salinidad (Bray, 1997). Sumado a ello, se hipotetiza que las proteínas transportadoras, las proteínas de canal y los transportadores de azúcar funcionan a través de la membrana plasmática y tonoplasto de modo de ajustar la presión osmótica bajo condiciones de salinidad (Seki y *col.* 2002). De manera contraria, las proteínas de transferencia de lípidos (LTPs) son otro grupo de proteínas de transporte asociadas al metabolismo de los ácidos grasos, pudiendo tener una función en la reparación del dano en membrana o cambios en la composición lipídica de la pared como un posible mecanismo de regulación de la permeabilidad contra iones tóxicos y mantenimiento de la fluidez de la membrana (Holmberg y Bülow, 1998; Torres-Schumann y *col.* 1992). Un gran número de LTPs mostraron tener incumbencia en la extensibilidad de la pared celular, pudiendo algunas ser secretadas en condiciones de alta concentración de NaCl en solución (Nieuwland y *col.* 2005). Aproximadamente la mitad de los transcriptos asociados a LTPs detectables en un experimento de micromatrices de raíz en *Arabidopsis* fueron sub-reguladas por tratamientos de salinidad (Jiang y *col.* 2006), mientras que Hewezi y *col.* 2006 también detectaron LTPs de *Arabidopsis* sub-reguladas en girasol en tratamientos de frío.

g) Metabolismo secundario

El EST H123 (AN: BU672069) muestra una alta identidad con una proteína fosfato sintetasa de mio-inositol (proteínas MIPS), isomerasa involucrada en el metabolismo del inositol (EC 5.5.1.4.). Dicha proteína fue sub-regulada bajo los tratamientos de frío y salinidad. El inositol es un osmoprotector natural de la pared celular sintetizado subcelularmente a fosfatidil inositol y luego reciclado dentro del ciclo de este metabolito bajo las exposiciones a estreses abióticos dentro una cascada de señalización compleja. Das-Chatterjee y *col.* 2006 reportaron la

existencia de un tabaco transgénico tolerante a salinidad en *Porteresia coarctata* (arroz salvaje halofítico) mediado por la introducción de una proteína de tipo MIPS. De manera contrastante, Chun y col. 2003 publicaron en sésamo una sub-regulación bajo estrés salino en semillas durante la germinación en dicha especie vegetal de la proteína SeMIPS1, proteína MIPS aislada en sésamo. Sumado a ello, Nelson y col. 1998, 1999 demostraron que la transcripción de los genes MIPS es afectada por la salinidad durante la biosíntesis de mio-inositol y sus derivados, mientras que Ishitani y col. (1996) también demostraron evidencia de la sobre-expresión de los genes MIPS durante condiciones de salinidad en plantas mediantemente tolerantes, mientras que evidenciaron una sub-expresión en los niveles de transcripción de ésta en plantas sensibles a la salinidad (e.g.: *Arabidopsis thaliana*). Dichos resultados explicarían la sub-regulación del EST H123 en girasol cultivado (especie sensible a la salinidad) el cual presenta identidad con un alto valor de puntaje a una proteína MIPS de otra especie oleaginosa no reportada (GO:0004512) (Fernandez y col. 2003).

V. CONCLUSIONES

1. Se creó un banco local de secuencias tipo ESTs de girasol cultivado conformado por 318 unigenes órgano específicos reunidos en 919 secuencias, utilizando la técnica de SSH como estrategia para la identificación de nuevas secuencias expresadas de girasol por una estrategia genómica económicamente eficiente para proyectos de pequeña escala.
2. La creación de este banco contribuyó con un 60% de secuencias públicas nuevas para girasol al momento de su publicación, y con un 14,5 % de secuencias nuevas en la actualidad. El número de genes novedosos predichos varió según la clonoteca analizada, en un rango que va de 56% en las clonotecas de flor estadio R4 al 16% en las de flor R1.
3. El desarrollo y aplicación de una rutina bioinformática denominada BioPipeline® y la creación de un servidor local en INTA denominado INTA01 permitió el análisis de las secuencias de ESTs generadas en forma automatizada y eficiente, con disminución del error en la manipulación de datos.
4. Se creó una base de datos en Access® que permitió organizar las secuencias por clonoteca, por nombre, por tamaño de inserto y por anotación funcional, quedando disponible para la generación de consultas e informes. Los campos y registros fueron programados en Visual Basic generando interrelaciones entre ellos para su uso integral actual o con las secuencias obtenidas a futuro.
5. Se generó una micromatriz de ADNc a partir de la base de ESTs local con un diseño de tres réplicas biológicas para la evaluación de respuestas a dos estreses abióticos relevantes para el cultivo de girasol: frío y salinidad.

6. El análisis de la matriz transcripcional permitió detectar 3 grupos de genes, los grupos 1 y 3 (integrados por 112 y 49 genes, respectivamente) mostraron sobre-y sub expresión -en uno y otro caso respecto del control- cuando las plantas fueron sometidas a estrés (por frío o por salinidad), mientras que Grupo 2, integrado por 126 genes, no mostró variación en sus niveles medios a través de los tratamientos.
7. Del análisis estadístico realizado a partir de la matriz transcripcional, aplicando la prueba de comparación de medias por análisis de la varianza y Análisis de Componentes Principales (ACP) surgieron 80 genes candidatos con expresión diferencial ante frío y/o salinidad o en ambos.
8. A partir del análisis conjunto de perfiles transcripcionales por métodos de agrupamiento y del análisis de perfiles individuales de los 80 genes candidatos, se detectaron 48 sobre o sub expresados frente a uno u otro estrés, indicando la implicancia de mecanismos regulatorios comunes a ambos estreses. Asimismo, se detectaron 17 y 12 genes sobre o sub expresados respectiva y específicamente frente a un estrés y no frente al otro.
9. De acuerdo al análisis funcional predicho por comparación con bases públicas, los genes candidatos específicamente regulados en situaciones de estrés formarían parte de mecanismos de regulación involucrando factores de transcripción y traducción, proteínas relacionadas con procesos de degradación/plegado o interacción entre proteínas y enzimas asociadas a la síntesis de metabolitos secundarios y generación y procesamiento de especies reactivas de oxígeno.

10. Se validó experimentalmente la expresión diferencial de siete genes candidatos mediante la técnica de PCR cuantitativa (PCR en tiempo real) y northern.

11. Dentro del grupo de genes candidato específicos de estrés, 12 corresponden a secuencias con funcionalidad desconocida y sólo 3 de las secuencias analizadas en este trabajo mostraron patrones transcripcionales opuestos, incluyendo una con similitud a un transportador de lípidos.

BIBLIOGRAFÍA

- Adams, M.D., Kelley, J.M., Gocayne J.D., Dubnick M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F. (1991) **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science*, 252, 1651-1656
- Agarwal, P.K, Agarwal, P., Reddy, M.K. and Sopory, S.: **Role of DREB transcription factors in abiotic and biotic stress tolerance in plants.** *Plant Cell Rep.* (2006), DOI 10.1007/s00299-006-0204-8
- Ali S, Holloway B, Taylor WC: **Normalization of cereal endosperm EST libraries for structural and functional genomic analysis.** *Plant Mol Biol Rep* 2002, 18:123-132.
- Allison D., Xiangqin C., Grier P.P., and Sabripour M.: **Microarray data analysis: from disarray to consolidation and consensus (2006).** *Nature Reviews Genetics*, 7:55-65
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool (1990)** *J Mol Bio* 1990, **215**:208-218
- Amaratunga, D. and Cabrera, J.: **“Exploration and análisis of DNA Microarray and Protein Array Data”**, 2004, Wiley-Interscience
- Apse MP., Aharon GS., Snedden WS, Blumwald, E.: **Salt tolerance conferred by overexpression of a vacuolar Na⁺/H⁺ antiport in *Arabidopsis*** (1999). *Science* 285:1256-1258
- Arabidopsis Genome Initiative. **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* (2000).** *Nature* 408:796–815.
- Arumuganathan, K. and Earle, E.D.: **Nuclear DNA content of some important plant species (1991).** *Plant Molecular Biology Reporter*, 9(3): 211-215
- Asagir S.A., Asociación Argentina de Girasol, **“Girasol: la especie y su historia”** <http://www.asagir.org.ar>

-
- Asamizu E, Nakamura Y, Sato S, Tabata S: **A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12028 non-redundant expressed sequence tags from normalized and size-selected cDNA library (2002).** *DNA Res*, **7**:175-180.
 - Ashburner M, Ball CA, Blake JA, Botstein D., Buttler H, Cherry JM, Davis AP, Dolinsky K, Dwight SS, Eppig JT et al: **Gene ontology: tool for the unification of biology (2000).** The gene Ontology Consortium. *Nature Genet.*, **25**:25-29
 - Ausbel, F.; Brent, R.; Kingston, R.; Moore, D.; Seidman, J.; Smith, J.; Struhl, K. (Eds.) (1997). **Current Protocols in Molecular Biology.**
 - Bartels D., Sunkar R.: **Drought and salt tolerance in plants (2005).** *Crit. Rev. Plant Science* **24**:23:58
 - Bello, R., Colombini M. y Takeda E. **Análisis de datos de microarreglos de dos canales. Un acercamiento teórico-práctico.** Taller práctico. Curso de Postgrado IC-FCEyN- UBA. Guía de Trabajos Prácticos 2006.
 - Berry S.T., Allen R.J., Barnes S.R. and Caligari P.D.S.: **Molecular analysis of *Helianthus annuus* L. 1. Restriction fragment length polymorphism between inbred lines of cultivated sunflower (1994).** *Theor. Appl. Genet.* **1994**, **89**:435-441
 - Berry S.T., Leon A.J., Hanfrey C.C., Challis P., Burkholz A., Barness S., Rufener G.K., Lee M., Caligari P.D.S: **Molecular marker analysis of *Helianthus annuus* L.: 2. Construction of an RFLP linkage map for cultivated sunflower (1995).** *Theor. Appl. Genet.*, **91**:195–199
 - Blumwald E., Aharon GS., Apse, P.: Sodium transport in plant cells (2000). *Biochem. Biophys. Acta* **1456**:140-151
 - Blumwald E., Grover A.: **Salt tolerance. In: Halford NG (ed), 2006. Plant biotechnology: current and future uses of genetically modified crops.** John Wiley and Sons Ltd., UK, pp 206-224
-

-
- Bonaldo M.F., Lennon G. and Soares M.B.: **Normalization and subtraction: two approaches to facilitate gene discovery (1996)**. *Genome Res.*, 6:791-806
 - Bohnert H.J, Jensen R.G. **Metabolic engineering for increased salt tolerance - the next step (1996)**. *Australian Journal of Plant Physiology*, 23: 661-667.
 - Boyer, J.S.: *Plant productivity and environment*. Science, 1982, Washington, v.218, p. 443-448
 - Brandt S., Kloska S., Altmann T., Kehr J.: **Using array hybridization to monitor gene expression at the single cell level (2000)**. *Journal of Experimental Botany*, Vol. 53, 379, 2315-2323
 - Bray E.A.: **Plant responses to water deficit (1997)**. *Trends in Plant Science* 2,48:54
 - Camon E, Barrell D, Brooksbank C, Magrane M , Apweiler R: **The Gene Ontology Annotation (GOA) project-application of GO in Swiss-Prot, TrEMBL and InterPro (2003)**. *Comp. and Funct. Genom.*, 4:71-74
 - De Castro C. y Bouças Farias JR. en:**Girassol no Brasil**, Villas Bôas de Campos Leite, RM., Brighenti AM. y de Castro, C., 2005, Embrapa
 - Cato S.A., Gardner, R.C., Kent J. and Richardson, T.E.: **A rapid PCR-based method for genetically mapping ESTs (2001)**. *Theor. Appl. Genet.*, 102:296–306
 - Charles SA, Halliwell B.: **Light activation of fructose bisphosphatase in isolated spinach chloroplasts and deactivation by hydrogen peroxide (1981)**. *Planta* 151:242–246
 - Chen W., Provart NJ., Glazebrook J., Katagiri F., Chang H., Eulgem T., Mauch F., Luan S., Zou G., Whitham SA., Budworth PR., Tao Y., Xie Z., Chen X., Lam S., Kreps JA., Harper JF. Si-Ammour A., Mauch-Mani B., Heinlein M., Kobayashi K., Hohn T., Dangl JF., Wang X. and Zhu T.: **Expression profile matrix of Arabidopsis transcription factor genes suggests their**
-

-
- putative functions in response to environmental stresses** (2002). *The Plant Cell*, Vol. 14, 559-574
- Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK.: **ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis** (2003). *Genes y Dev.* 17(8):1043-54
 - Chinnusamy V, Jagendorf A., Zhu JK: **Understanding and improving salt tolerance in plants** (2005). *Crop Science* 45:437-448.
 - Chuaqui R.F., Bonner, R.F., Best C.J., Gillespie, J.W., Flaig M.J., Hewitt S.M., Phillips J.L., Krizman D.B., Tangrea M.A., Ahram M., Linehan W.M., Knezevic V., Emmert-Buck M.R.: **Post-analysis follow-up and validation microarray experiments** (2002), *Nat. Genet.* 32, 509– 514
 - Chun, J-A., Jin U-H., Lee J-W., Yi, Y-B., Hyung N-I., Kang M-H, Pyee J-H., Suh MC, Kang C-W., Seo H-Y., Lee S-W., Chung C-H.: **Isolation and characterization of a myo-inositol 1-phosphate synthase cDNA from developing sesame (*Sesamum indicum L.*) seeds: functional and differential expression, and salt-induced transcription during germination** (2003). *Planta*, 216:874-880
 - Chung CH., Yee YJ., Kim DH., Chung DS.: **Changes of lipid, protein, RNA and fatty acid composition in developing sesame (*Sesamum indicum L.*) seeds** (1995). *Plant Sci* 109:237-243
 - Cui S., Huang F., Wang J., Ma X., Cheng Y. and Liu J.: **A proteomic analysis of cold stress reponse in rice seedlings** (2005). *Proteomics*, 5, 3162–3172
 - Dalmaso C., Broët P., Moreau T. **A simple procedure for estimating the false discovery rate.** *Bioinformatics* 2005, 21(5) 660-668
 - Das-Chatterjee A., Goswami L., Maitra S., Ghosh Dastidar K., Ray S., Majumder A.L.: **Introgression of a novel salt-tolerant L-myo-inositol 1-phosphate synthase from *Porteresia coarctata* (Roxb.) Tateoka (*PcINOI*) confers salt tolerance to evolutionary diverse organisms.** *FEBS Letter* 580 (2006), 3980-3988
-

-
- Dezar C.A. **Caracterización funcional de genes pertenecientes a la familia Hd-Zip de girasol. Estudio particular de *Hahb-4*, un miembro de la familia involucrado en la respuesta al estrés hídrico.** Tesis doctoral 2005. Facultad de Bioquímica y Ciencias Biológicas
 - Dezar C.A., Gago G.M., González D.H., Chan, R.L. **Hahb-4, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to *Arabidopsis thaliana* plants (2005).** Transgenic Research, 14:429-440
 - Dezar C.A., Fedrigo G.V., Chan R.L. **The promoter of the sunflower HD-Zip protein gene *Hahb-4* directs tissue-specific expression and is inducible by water stress, high salt concentrations and ABA (2005).** Plant Science, 169:447-459.
 - Cátedra de Biología Celular y Molecular. Universidad Nacional del Litoral.
 - Dhanaraj AL, Alkharouf NW, Beard HS, Chouikha IB, Matthews BF, Wei H, Arora R. and Rowland LJ: **Major differences observed in transcript profiles of blueberry during cold acclimation under field and cold room conditions (2006).** Planta, doi 10.1007/s00425-006-0382-1
 - Dhand R. **Functional genomics (2000).** Nature; 405:819.
 - Di Rienzo J.A., Fabro G. Alvarez M.E. **2004 Selección de genes en un experimento factorial con “microarrays” mediante técnicas estadísticas multivariadas.** Basic & Applied Genetics XVI (Supplement) 2004 47.
 - Di Rienzo J., Gonzalez L., Tablada M., Bruno C. and Balzarini M. **A cluster based technique to group random effects from empirical BLUPs: An application to genotype selection (2005).** Journal of Crop Improvement: 15(1):133 – 146
 - Di Rienzo J.A, **Comunicación Personal 2006**
 - Diatchenko L, Lau YF, Campbell AP, Chenchick A, Moqaddam F, Huang B, Lukyanov S, Konstantin L., Gurskaya N, Sverdlov E, Siebert PD: **Suppression subtractive hybridization: A method for generating**
-

-
- differentially regulated or tissue-specific cDNA probes and libraries (1996).** *Proc Natl Sci USA*, 93:6025-6030
- Diatchenko L., Lukyanov S., Lau, YF. And Siebert, PD: **Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes (1999).** *Meth. Enzymol.*, 303:349-380
 - Dopazo J. y Valencia A. en **Genómica y Mejora Vegetal**, Grupo Mundi Prensa, (Ed. 2002)
 - Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K.: **OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expresión (2003).** *Plant J.* 33(4):751-63
 - Dunn M.A.; Hughes M.A.; Zang L.; Pearce R.S.; Quigley A.S.; Jack P.S.: **Nucleotide sequence and molecular analysis of the low temperature induced gene, BLT4.** (1991) *Mol. Gen. Genet.* 229(3): 389-94.
 - Eyster K.M., Lindahl R. **Molecular medicine: a primer for clinicians. Part XII: DNA microarrays and their application to clinical medicine (2001).** *S D J Med.*;54(2):57-61
 - Eberwine J, Spencer C, Miyashiro K, Mackler S, Finnell R. **Complementary DNA synthesis in situ: methods and applications (1992).** *Methods Enzymol.*, 216:80-100
 - Estruch F.: **Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast (2000).** *FEMS Microbiol Rev* 25:469-486
 - Fedorova, M., van de Mortel, J., Matsumoto, P.A., Cho, J.,Town, C.D., VandenBosch, K.A., Gantt, J.S. and Vance,C.P.: **Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula* (2002).** *Plant Physiol.* 130(2): 519–537
-

-
- Fei Z., Tang X., Alba RM., White JA., Ronning C., Martin GB., Tanksley SD. y Giovannoni, J.J.: **Comprehensive EST analysis of tomato and comparative genomics on fruit ripening** (2004) *The Plant Journal*, 40, 47-59
 - Fernández P., Paniego N., Lew S., Esteban Hopp H. and Heinz R.A.: **Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project (2003)**. *BMC Genomics*, 4:40
 - Fernández, P., Peluffo L., Hopp H.E., Paniego N. y Heinz R.A.: **“Identification and characterization of candidate genes for resistance to Sclerotinia head rot in sunflower”**, RedBIO 2005, 7-11 de junio de 2005, UCA, Buenos Aires, Argentina
 - Fernández P., Di Rienzo J., Hopp H.E., Paniego N. y Heinz R.A.: **“Análisis de expresión y búsqueda de genes candidato mediante la técnica de micromatrices de ADNc en girasol cultivado (*Helianthus annuus*)”**. 52 Congreso Sociedad Brasileira de Genética. 3-6 de septiembre de 2006, Foz de Iguazu, Brasil
 - Fick, G.N. and Zimmer, D.E.: **Yield stability of sunflower hybrids and open pollinated varieties. In "Proc. 7TH Int. Sunflower Conf."** (1976), Krasnadar, Russia, June 27-July 3, Int. Sunflower Assoc., Paris, France, pp 253-258.
 - Fields S, Hohara Y, Lockhart D. **Functional genomics (1999)**. *Proc. Natl. Acad. Sci.*; 96:8825-8826.
 - Firestein G.S. and Pisetsky D.S.: **DNA microarrays: boundless technology or bound by technology? Guidelines for studies using microarray technology (2002)**, *Arthritis Rheum.* 46, 859–861
 - Flowers T.J and Yeo A.R. **Breeding for salinity resistance in crop plants: where next?** (1995) *Australian Journal of Plant Physiology*, 22:875-884.
-

-
- Franco D.: Aceite de girasol: Análisis de la cadena alimentaria. Cámara de la Industria Aceitera de la República Argentina (CIARA), Julio de 2005
 - Fujita M., Fujita Y., Noutoshi Y., Takahashi F., Narusaka Y., Yamaguchi-Shinozaki K. and Shinozaki K.: **Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks (2006).** Current Opinion in Plant Biology, 9:436-442
 - Garzón A. y Jiménez, J. en: **Genómica y Mejora Vegetal**, Grupo Mundi Prensa, (Ed. 2002)
 - Gay, C., Corbineau, F. and Côme, D.: **Effects of temperatura and oxigen on seed germination and seedling growth in sunflower (*Helianthus annuus L.*) (1991).** Environmental and Experimental Botany, 31:193-200
 - Gedil, M.A., Wye C., Berry S., Segers B., Peleman J., Jones R., Leon A., Slabaugh M.B., Knapp S.J. **An integrated restriction fragment length polymorphism-amplified fragment length polymorphism linkage map for cultivated sunflower (2001).** Genome, 44:213–221
 - Gentzbittel L., Mestries E., Mouzeyar S., Mazeyrat F., Badaoui S., Vear F., De Labrouhe D.T., Nicolas P.: **A composite map of expressed sequences and phenotypic traits of the sunflower *Helianthus annuus L.* genome (1999).** Theor. Appl. Genet., 99:218–234
 - Gentzbittel L, Vear F, Zhang YX, Berville: **A development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus L.*) (1995).** Theor. Appl. Genet., 90:1079–1086
 - Gentzbittel, L., Y.X. Zhang, F. Vear, B. Griveau, and P. Nicolas.: **RFLP studies of genetic relationships among inbred lines of the cultivated sunflower *Helianthus annuus L.*: evidence for distinct restorer and maintainer germplasm pools (1994).** Theor. Appl. Genet., 89:419–425.
 - Gepstein, S., Grover A., Blumwald E.: **Producing biopharmaceuticals in the dessert: building an abiotic stress tolerance in plants for salt, heat and**
-

-
- drought.** In: Knablein J., Muller RH. (eds), 2006. Modern biopharmaceuticals. Wiley-VCH Verlag GmbH, Weinhaum, pp 967-994
- Ghosh S., Bagchi S. and Majumder AL.: **Chloroplast fructose-1,6-bisphosphatase from *Oryza* differs in salt tolerance property from the *Porteresia* enzyme and is protected by osmolytes** (2001). Plant Science 160 (2001) 1171–1181
 - Gibson S.; Arondel V.; Iba K.; Somerville C: **Cloning of a temperatura-regulated gene encoding a chloroplast amega-3 desaturase from *Arabidopsis thaliana*** (1994). Plant Physiol. 106: 1615-1621
 - Gilmour S.J.; Artus N.N.; Thomashow M.F: **cDNA sequence analysis and expression of two cold-regulated genes of *Arabidopsis thaliana*** (1992) Plant Mol. Biol. 18: 13-21.
 - Grover A.: **Molecular biology of stress responses** (2002). Cell Stress & Chaperones 1: 1–5
 - Guy C.L: **Cold acclimation and freezing stress tolerance: role of protein metabolism** (1990) Annu. Rev. Plant Physiol. Plant Mol. Biol. 41: 187-223
 - Harding HP., Calfon M., Urano F., Novoa I., Ron D.: **Transcriptional and translational control in the mammalian unfolded protein response** (2002). Annu Rev Cell Dev Biol 18:575-599
 - Hasegawa, PM, Bressan RA, Zhu, JK, Bohnert HJ: **Plant cellular and molecular responses to high salinity** (2000). Annual Rev Plant Physiol Plant Mol Biol 51:463-499
 - Heiser, C. B., and D. M. Smith: **New chromosome numbers in *Helianthus* and related genera.** Proc. Indiana Acad. Sci. 1955, 64: 250–253
 - Heiser, C.B. Jr.: **The Sunflower.** 1976. Univ. of Oklahoma Press, Norman 198 pp.
 - Henning S, Groth D, and Lehrach H: **Automated Gene Ontology annotation for anonymous sequence data** (2003). Nucleic Acids Res., 31: 3712-3715
-

-
- Hewezi T., Petitprez M. and Gentzbittel L.: **Primary metabolic pathways and signal transduction in sunflower (*Helianthus annuus L.*): comparison of transcriptional profiling in leaves and immature embryos using cDNA microarrays.** (2006) *Planta* , 223: 948–964
 - Hewezi T, Lèger M., El Kayal W. and Gentzbittel L.: **Transcriptional profiling of sunflower plants growing unde low temperatures reveals an extensive SUB-regulation of gene expression associated with chilling sensitivity** (2006). *Journal of Experimental Botany* doi:10.1093/jxb/erl080
 - Holmberg, N. and Bülow L.: **Improving stress tolerance in plants by gene transfer** (1998). *Trends in Plant Science* 3,61-66
 - Horvath D, Schaffer R, West M, Wisman E: **Arabidopsis microarrays identify conserved and differentially expressed genes involved in shoot growth and development from distantly related plant species (2003).** *Plant J.* 34:125–124
 - Hoshida H., Tanaka Y, Hibino T, Hayashi Y, Tanaka A, Takabe T, Takabe T.: **Enhanced tolerance to salt stress in transgenic rice that overexpresses chloroplast glutamine synthetase** (2000) *Plant Molecular Biology*, 43(1):103-11
 - Houde M.; Dhindsa R.S.; Sarhan F: **A molecular marker to select for freezing tolerance in Gramineae** (1992). *Mol. Gen. Genet.* 234:43-48
 - Houde M., Belcaid M., Ouellet, F., Danyluk J., Monroy AF., Dryanova A., Gulick P., Bergeron A., Laroche A., Links M., MaCarthy L., Crosby, WL. And Sarhan F.: **Wheat EST resources for functional genomics of abiotic stress** (2006). *BMC Genomics*, 7:149
 - Hu W., Wang Y., Bowers C. y Ma H.: **Isolation, sequence analysis, and expression studies of florally expressed cDNAs in Arabidopsis** (2003). *Plant Molecular Biology*, 53:545-563
 - Huang L., Ye Z., Bell, R. and Dell, B: **Boron nutrition and chilling tolerance of warm climate crop species** (2005) *Annals of Botany*, 96: 755-767
-

-
- Huang X, Adams MD, Zhou H and Kerlavage AR: **A tool for analyzing and annotating genomic sequences** (1997) *Genomics*, 46:37-45.
 - Hurry VM and Huner NPA.: **Low growth temperature effects a differential inhibition of photosynthesis in spring and winter-wheat** (1991). *Plant Physiology* 96, 491–497.
 - Ishitani M., Majumder AL., Bornhouser A., Michalowski CB., Jensen RG., Bohnert HJ.: **Coordinate transcriptional induction of myo-inositol metabolism during environmental stress** (1996). *Plant J* 9:537:548
 - Jan C.C., Vick B.A., Miller J.F., Kahler A.L., Butler E.T., III.: **Construction of an RFLP linkage map for cultivated sunflower** (1998) *Theor Appl Genet.*, 96:15–22
 - Jiang Y. and Deyholos MK: **Comprehensive transcriptional profiling of NaCl-stressed *Arabidopsis* roots reveals novel classes of responsive genes** (2006), *BMC Plant Biology* 6:25
 - Kanesaki Y., Suzuki I., Allakhverdiev SI, Mikami K., Murata N.: **Salt stress and hyperosmotic stress regulate the expression of different sets of genes in *Synechocystis sp.*** (2002). PCC 6803. *Biochem Biophys Res Commun* 290:339-348
 - Kawasaki S., Borchert C., Deyholos M., Wang H., Brazille S., Kawai K., Galbraith D., Bohnert HJ: **Gene expression profiles during the initial phase of salt stress in rice** (2001). *Plant Cell* 13:889-905
 - Kee SC, Martin B. and Ort DR: **The effects of chilling in the dark and in the light on photosynthesis of tomato: electron transfer reactions** (1986). *Photosynth. Res.* 8, 41-51.
 - Kesseli, R. V. and Michelmore, R. W. **The Compositae: systematically fascinating but specifically neglected. In: Genome Mapping in Plants**, (1997), A.H. Paterson ed. R.G. Landes Co. Georgetown, TX. pp179 - 191.
 - Kinman, M.L.: **Proc 4th Intl Sunflower Conference**, 1970, Memphis, Tennessee, USA. pp. 181-183.
-

-
- Kore-Eda S., Cushman MA, Akselrod I, Bufford D. Fredrickson M., Clark E, Cushman JC: **Transcript profiling of salinity stress responses by large-scale expressed sequence tag analysis in *Mesembryanthemum crystallinum*** (2004). *Gene* 341:83-92
 - Kratsch HA, Wise RR: **The ultrastructure of chilling stress.** (2000) *Plant, Cell and Environment* 23: 337–350
 - Kreps JA, Simon AE: **Environmental and genetic effects of circadian clock-regulated gene expression in *Arabidopsis*** (1997). *The Plant Cell* 8, 489-503
 - Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF: **Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress** (2002) *Plant Physiology* 130, 2129–2141
 - Leclercq P.: **Une stérilité mâle cytoplasmique chez le tournesol** (1969). *Ann. Amélior. plantes*, 19: 99-106.
 - Lemaire-Chamley M., Petit J., Garcia V., Just D., Baldet P., Germain V., Fagard M., Mouassite M., Cheniclet C. and Rothan C.: **Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato** (2005). *Plant Physiology* on line publication, www.plantphysiol.org/cgi/doi/10.1104/pp.105.063719
 - Levitt J.: **Chilling, Freezing, and High Temperature Stresses** (1980) Vol. 1, 2nd edn. Academic Press, New York, USA
 - Lew S. y Guida G.: **Desarrollo y puesta en marcha del servidor INTA01,** versión beta. Mayo de 2002.
 - Liu X. and Baird, W.V.: **Differential expression of genes regulated in response to drought or salinity stress in sunflower** (2003). *Crop Science*, 43:678–687
 - Livak K.J., Schmittgen T.D. : **Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method** (2001). *Methods*, 25:402-408
-

-
- Lockhart D.J., Winzeler EA. **Genomics, gene expression and DNA arrays** (2000) *Nature*, 405 (6788):827-36.
 - Lundmark M., Cavaco AM., Trevanion S., Hurry V.: **Carbon partitioning and export in transgenic *Arabidopsis thaliana* with altered capacity for sucrose synthesis grown at low temperature: a role for metabolite transporters** (2006). *Plant, Cell and Environment* 29:1703-1714
 - Marin K., Kanasaki Y., Los DA, Murata N. Suzuki I, Hagemann M: **Gene expression profiling reflects physiological processes in salt acclimation of *Synechocystis sp. strain PCC 6803*** (2004). *Plant Physiology* 136:3290-3300
 - Macchia, M., Benvenuti, A., Baldanzi, M.: **Temperature requirements during germination in sunflower**. In: **Conferencia Internacional de Girasol**, 11, 1985, Mar del Plata. Actas. Mar del Plata: ASAGIR/ISA, 1985. T1, p.93-97
 - Mahajan, S. and Tuteja, N.: **Cold, salinity and drought stresses: An overview**. (2005) *Archives of Biochemistry and Biophysics*, 444:139–158
 - Mahalingam R., Gomez-Buitrago A., Eckardt N., Shah N., Guevara-Garcia A., Day P., Raina R. and Fedoroff Nina V.: **Characterizing the stress/defense transcriptome of *Arabidopsis*** (2003). *Genome Biology*, 4:R20.
 - Manavella P., Arce A., Dezar C., Bitton F., Renour, JP., Crespi M. and Chan, R.: **Analysis of Transcriptional Networks Regulated by the Sunflower Hahb-4 Transcription Factor in *Arabidopsis* Revealed a Conserved Role in Ethylene Signalling** (2006). *The Plant Journal*, 48:125-137
 - Martino-Catt S., Ort DR: **Low temperature interrupts circadian regulation of transcriptional activity in chilling-sensitive plants** (1992). *Proceedings of the National Academy of Sciences, USA* 89, 3731-3735
 - Maruyama K., Sakuma Y, Kasuga M, Ito Y, Seki M, Goda H, Shimada Y, Yoshida S, Shinozaki K, Yamaguchi-Shinozaki K.: **Identification of cold-inducible SUBstream genes of the *Arabidopsis* DREB1A/CBF3**
-

-
- transcriptional factor using two microarray systems** (2004). *Plant. J.*, 38(6):982-93.
- Merrien, A.: **Physiologie du tournesol**, 1992, Paris: CETIOM, 66 p.
 - Michiels, S.. **Prediction of cancer outcome with microarrays: a multiple random validation strategy** (2005). *Lancet*, 365, 488-492
 - Mira H., Martínez-García F., Peñarrubia L.: **Evidence for the plant-specific intercellular transport of the Arabidopsis copper chaperone CCH** (2001) *Plant J.* 25:521–528
 - Mittler, R: **Oxidative stress, antioxidants, and stress tolerance** (2002). *Trends Plant Sci.* 7, 405–410
 - Mittler, R.: **Abiotic stress, the field environment and stress combination** (2006). *Trends in Plant Science* 11,1:15-19
 - Nelson DE., Rammesmayer G., Bohnert HJ.: **Regulation of cell-specific inositol metabolism and transport in plant salinity tolerance** (1998). *Plant Cell* 10:753-764
 - Nelson DE., Koukoumanos M., Bohnert HJ.: **Myo-inositol-dependent sodium uptake in ice plant** (1999). *Plant Physiol* 119:165-172
 - Nieuwland J, Feron R, Huisman BAH, Fasolino A, Hilbers CW, Derksen J, Mariani C: **Lipid transfer proteins enhance cell wall extension in tobacco** (2005). *Plant Cell* 17 (7):2009-2019
 - Nordin K.; Heino P.; Palva E.T: **Separate signal pathways regulate the expression of a low-temperature-induced gene in *Arabidopsis thaliana* (L.) Heynh** (1991) *Plant Mol. Boil.* 21:641-53
 - Nuez, F., Carrillo, J.M. y Lozano R. en **Genómica y Mejora Vegetal**, Grupo Mundi Prensa, (Ed. 2002)
 - Oullet F.; Vazquez-Tello A.; Sarhan F: **The wheat wcs120 promoter is cold-inducible in both monocotyledonous and dicotyledonous species** (1998) *FEBS Lett.* 423:324-28
 - Paniego N, Echaide M, Muñoz M, Fernández L, Torales S, Faccio P, Fuxan I, Carrera M, Zandomeni R, Suárez EY, Hopp HE: **Microsatellite isolation**
-

-
- and characterization in sunflower (*Helianthus annuus* L.)** (2002). Genome, 45:34–43
- Paniego N., Heinz R., Fernandez P. Talia P., Nishi V. and Hopp HE: **Sunflower. En: "Genome Mapping & Molecular Breeding"** (2006) Vol. II (Oilseeds), Ed.Springer-Verlag- Germany VoII .
 - Patanjali, S.R., Parimoo S. and Weissman S.M.: **Construction of a uniform-abundance (normalized) cDNA library** (1991). Proc. Natl. Acad. Sci. USA, 88:1943-1947
 - Pearson, W.R.: **Rapid and sensitive sequence comparison with FASTP and FASTA** (1990). Methods Enzymol., 183, 63-98.
 - Peluffo L., Fernández P., Hopp E., Paniego N. y Heinz R.A: **“Identificación y caracterización de genes candidato para resistencia a la podredumbre húmeda del capítulo de girasol (*Sclerotinia sclerotiorum*)”**. Tercer Congreso Argentino de Girasol, 30-31 de mayo de 2005, Hotel Hilton, Buenos Aires, Argentina.
 - Percival, MP, Bradbury, M.; Hayden, DB.; Baker, NR. (1987): **Modification of the photochemical apparatus in maize by photoinhibitory stress at low temperature. In: Biggins J. editor. Progress in Photosynthesis Research, Vol 4.** Dordrecht, The Netherlands: Martinus Nijhoff; 1987. pp. 47–50. .
 - Poormohammad Kiani S., Grieu P., Maury P., Hewezi T., Gentzbittel L., Sarrafi A.: **Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.)** (2006). Theor Appl Genet DOI 10.1007/s00122-006-0419-7
 - Poormohammad Kiani S., Talia P., Nishinakamasu V., Maury P., Grieu P., Perrault A., Hopp, E.; Heinz R.; Gentzbittel L., Paniego N. and Sarrafi A.: **Development of a new AFLP and SSR genetic linkage map and identification of QTLs for photosynthesis related traits under well-watered and water-stressed conditions in sunflower (*Helianthus annuus* L.)** (2006). Unpublished.
-

-
- Prasad, TK., Anderson MD., Martin BA and Stewart CR.: **Evidence for chilling-induced oxidative stress in maize seedlings and a regulatory role for hydrogen peroxide** (1994). *The Plant Cell*, Vol 6, 1:65-74
 - Puebla AF. y Del Viso F.: **Tolerancia a factores bióticos**. En: **“Biotecnología y Mejoramiento Vegetal”**, capítulo 8:355-365. Eds. Echenique V., Rubinstein C., Mroginski L. (2004). Ediciones INTA.
 - Raven, P. y D. Axelrod.. **Angiosperm biogeography and past continental movements**. *Ann. Missouri Bot. Gard.* 1974, 61: 539-673
 - Reddy A.R., Ramakrishna A.C., Sekhar A.C., Nagablushana I., RavindraBabu P., Bonaldo M.F., Soares M.B. and Bennetzen J.L.: **Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa L. subsp. Indica cv.Nagina 22*)**. *Genome* 2002, 45:204-211
 - Renaut J., Lutts S., Hoffmann L. and Hausman J-F.: **Responses of poplar to chilling temperatures: proteomic and physiological aspects** (2004). *BMC Plant Biology* 6:81-90
 - Rensink, W.A. and Buell, C. Robin: **Microarray expression profiling resources for plant genomics** (2005). *Trends in Plant Science*, Vol. 10, No. 12, 603-609
 - Robinson, R.G.: **Production and culture** In: Carter. JF. (Ed.) **Sunflower science and technology**. Madison: ASA, 1978, p.89-95
 - Robinson R.G., Bernat, L.A., Geise, H.A., Jonson, F.K., Kinman, M.L., Mader, E.L., Oswalt, R.M., Putt E.D., Swallers, C.M and Williams, J.H.: **Sunflower development at latitudes ranging from 31 to 39 degrees** (1967). *Crop Science*, Madison, 7:134-136
 - Rocco, M.: **Citogenética clásica y molecular en girasol** (2002). Tesis de Licenciatura, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires
-

-
- Rodriguez Milla MA., Townsend J., Feng Chang I. and Cushman JC: **The Arabidopsis AtDi19 gene family encodes a novel type of Cys2/His2 zinc-finger protein implicated in ABA-independent dehydration, high-salinity stress and light signaling pathways** (2006) *Plant Molecular Biology* 61:13–30
 - Romero M de C., Di Rienzo J.A. **Análisis de perfiles genéticos: Selección de atributos**. .IV Congreso Latinoamericano de Biología Matemática, Tandil, Bs. As, Noviembre 2005.
 - Rueda E.C., Dezar C.A., Gonzalez D.H., Chan R.L.: **Hahb-10, a sunflower homeobox-leucine zipper gene, is involved in the response to dark/light conditions and promotes a reduction of the life cycle when expressed in Arabidopsis** (2005). *Plant & Cell Physiology*, 46(12):1954-1963
 - Sadras, V.O. and Hall, A.J.: **Quantification of temperature, photoperiod and population effects on plant leaf area in sunflower crops**. *Field Crops Research* 1988, Amsterdam, 18:185-196
 - Sahi, C., Agarwal M., Reddy MK, Sopory SK, Grover A: **Isolation and expression analysis of salt stress-associated ESTs from contrasting rice cultivars using a PCR-based subtraction method** (2003). *Theor Appl Genet* 106:620-628
 - Sahi, C., Singh, A., Kumar K., Blumwald E. and Grover A.: **Salt stress response in rice: genetics, molecular biology, and comparative genomics**. (2006) *Funct. Integr. Genomics*, 6:263-284
 - Sambrook J., Fritsch E.F., Maniatis T.: ***Molecular Cloning: A Laboratory Manual*, 2nd edn** (1989). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
 - Sassenrath GF, Ort DR, Portis AR: **Impaired reductive activation of stromal bisphosphatases in tomato leaves following low-temperature exposure at high light** (1991) *Arch Biochem Biophys* 282: 302-308
-

-
- Schneiter A.A., Miller J.F.: **Description of Sunflower Growth Stages** (1981). Crop Science, 11:635-638
 - Seeman JR., Sharkey TD.: **Salinity and nitrogen effects on photosynthesis, ribulose-1,5-bisphosphate carboxylase and metabolite pool size in *Phaseolus vulgaris*** (1982) PlantPhysiol. 82 (1982) 555–560
 - Seiler J.G., Rieseberg L.H. (1997). **Systematics, origin, resources of the wild and domesticated sunflower**. En: Sunflower Technology and Production, Agronomy series Agronomy, Inc., Madison, Wisconsin, pp. 21-65.
 - Seki, M., Narusaka M., Abe H., Kasuga M., Yamaguchi-Shinozaki K., Carninici P., Hayashizaki Y, Shinozaki K.: **Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray** (2001), Plant Cell 13:61-72
 - Seki, M., Narusaka M., Ishida J., Nanjo T., Fujita M., Oono Y., Kamiya A., Nakajima M., Enju A., Sakurai T., Satou M., Akiyama K., Taji, T., Yamaguchi-Shinozaki K., Carninici P, Kawai J., Hayashizaki Y. and Shinozaki K.: **Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray** (2002). The Plant Journal, 134:279-292
 - Shi H., Ishitani M., Kim C., Zhu, JK.: **The *Arabidopsis thaliana* salt tolerance gene SOS1 encodes a putative Na⁺/K⁺ antiporter** (2000). Proc. Natl. Acad. Sci. USA:97:6896-6901
 - Shiozaki N., Yamada M., Yoshiba Y.: **Analysis of salt-stress-inducible ESTs isolated by PCR-subtraction in salt-tolerant rice** (2005). Theor Appl Genet 110:1177-1186
 - Shy C., Ingvaridsen C., Thümmler, F., Melchinger A., Wenzel G. y Lübberstedt T.: **Identification by suppression subtractive hybridization of genes that are differentially expressed between near-isogenic maize lines in association with sugarcane mosaic virus resistance** (2005). Mol. Gen. Genomics, 273:450-461
-

-
- Smol J, Astriab M, Dudzinska-Madej B, Twardowski T.: **Stress conditions applied to the interpretation of translation machinery** (2001). Acta Biol Hung. 2001;52(1):161-70
 - Soares M.B., Bonaldo M.F., Jelene P., Su L., Lawton L. and Efstratiadis A.: **Construction and characterization of a normalized cDNA library** (1991). Proc. Natl. Acad. Sci. USA, 27:9228-9232
 - Sottosanto, JB, Gelli A, Blumwald E.: **DNA array analysis of *Arabidopsis thaliana* lacking a vacuolar Na/H antiporter: impact of AtNHX1 on gene expression** (2004). Plant Journal 40:752-771
 - Staudt L.M, Brown P.O. **Genomic views of the immune system** (2000). Annu Rev Immunol; 18:829-59.
 - Steponkus P.L.: **Role of the plasma membrane in freezing injury and cold acclimation** (1984) Annu. Rev. Plant Physiol. 35: 543-584
 - Steponkus P.L.; Uemura M.; Webb M.S: **A contrast of the cryostability of the plasma membrane of winter rye and spring oat – two species that widely differ in their freezing tolerance and plasma membrane lipid composition.** (1993) En Advances in Low-Temperature Biology, ed. P.L. Steponkus, 2: 211-312. London, JAI Press
 - Strand Å., Hurry V., Gustafsson P. and Gardeström P.: **Development of *Arabidopsis thaliana* leaves at low temperatures releases the suppression of photosynthesis and photosynthetic gene expression despite the accumulation of soluble carbohydrates** (1997). Plant Journal 12, 605–614.
 - Suh M, Kim M.J., Hur C., Bae J.M., Park Y.I., Chung C., Kang C., Ohlrogge J.B.: **Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds** (2003). Plant Mol Biol., 52:1107–1123
 - Szathmáry E., Jordán F., Pál C. **Can genes explain biological complexity?** (2001) Science. 292:1315–1316.
-

-
- Taiz, L. and Zeiger E. **Plant Physiology**, 1998. Sinauer Associates, Inc., Publishers Second Edition.
 - Taji T., Seki M., Satou M., Sakurai T., Kobayashi M., Ishiyama K., Narusaka Y., Narusaka M., Zhu JK, Shinozaki K.: **Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using *Arabidopsis* microarray** (2004). *Plant Physiology* 135:1697-1709
 - Talia P. **Aislamiento y caracterización de análogos de genes de resistencia (RGAs) a patógenos en girasol.** (2002) Tesis de grado para optar a la Licenciatura en Ciencias Biológicas.
 - Tang S., Yu J.K., Slabaugh M.B., Shintani D.K., Knapp S.J. **Simple sequence repeat map of the sunflower genome** (2002). *Theor. Appl. Genet.*, 105:1124-1136
 - .Tanji KK. **Agricultural salinity assessment and management.** NY.USA: Irrigation and Drainage Division, American Society of Civil Engineers. 1990.
 - Takemura M, Fujishige K, Hyodo H, Ohashi Y, Kami C, Nishii A, Ohyama K and Kohchi T: **Systematic isolation of genes expressed at low levels in inflorescence apices of *Arabidopsis thaliana*.** (1999) *DNA Res*, 6:275-82.
 - Tamborindeguy C., Ben C., Liboz T., Gentzbittel L.: **Sequence evaluation of four specific cDNA libraries for developmental genomics of sunflower** (2004). *Mol Genet. Genom* 271:367-375
 - Thomashow MF: **Plant cold acclimation: freezing tolerance genes and regulatory mechanisms** (1999) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50: 571-599
 - Thomashow, MF: **So What's New in the Field of Plant Cold Acclimation? Lots!** (2001). *Plant Physiology*, 125:89-93
 - Thompson J.D., Higgins D.G. and Gibson T.J.: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence**
-

-
- weighting, position-specific gap penalties and weight matrix choice** (1994). *Nucleic Acids Res.*, 22:4673-80.
- Torres-Schumann S., Godoy JA and Pintor-Toro JA: **A probable lipid transfer protein gene is induced by NaCl stems of tomato plants** (1992). *Plant Mol. Biol.* 18,749-757
 - Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton G.G. et al: **The sequence of the human genome** (2001). *Science*, 291, 1304-1351
 - Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu JK.: **Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status** (2006). *Plant J.* 45(4):523-39
 - Vierstra RD, Callis J: **Polypeptide tags, ubiquitous modifiers for plant protein regulation** (1999). *Plant Mol Biol* 41:435-442
 - Villalobos, F.J. and Ritchie J.T.: **The effect of temperature on leaf emergence rates of sunflower genotypes** (1992). *Field Crops Research*, Amsterdam, 29:37-46
 - Vucmirovic O, Tilghman S. **Exploring genome space** (2000). *Nature*;405: 820-822
 - Vodkin L.O., Khanna A., Shealy R., Clough S., Orlando González D., Philip R., Zabala G., Thibaud-Nissen F., Sidarous M., Strömvik M.V., Shoop E., Schmidt C., Retzel E., Erpelding J., Shoemaker R., Rodriguez-Huete, A.M., Polacco J., Coryell V., Keim P., Gong G., Liu L., Pardinás J. and Schweitzer P.: **Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant** (2004). *BMC Genomics*, 5:73
 - Walker R., Torofalvy E., SUBton WJS.: **Photosynthetic responses of Citrus varieties Rangpur lime and Etrog Citron to salt treatment** (1982). *Aust. J. Plant Physiol.* 9: 783–790.
-

-
- Wilson C., Shannon MC: **Salt induced Na⁺/H⁺ antiport in root plasma membrane of a glycophytic and halophytic species of tomato** (1995) *Plant Sci.* 107:147-157
 - Winicov I, Bastola D.R.: **Salt tolerance in crop plants: new approaches through tissue culture and gene regulation** (1997). *Acta Physiologiae Plantarum* ,19: 435-449.
 - Wolfraim L.A.; Langis R.; Tyson H. , Dhindsa R.S: **cDNA sequence, expression and transcriptibility of a cold-acclimation-specific gene, cas18, of alfalfa (*Medicago fulcata*) cells** (1993) *Plant Physiol.* 1101: 1275-1282
 - Wood AJ, Oliver MJ: **Translational control in plant stress: formation of messenger ribonucleoprotein complexes (mRNPs) in *Tortula ruralis* in response to desiccation** (1999). *Plant J.* 18:359-370
 - Wool, IG: **Extraribosomal functions of ribosomal proteins** (1996). *Trends Biochem Sci* 21:164-165
 - Yamada, S., Toshiyuki K., Akiko H., Kuwata S., Hidemasa I. and Tomoaki K. **Differential expression of plastidic aldolase genes in *Nicotiana* plants under salt stress** (2000). *Plant Science*, 154:61-69
 - Yamaguchi-Shinozaki K.; Shinozaki K.: **Characterization of the expression of a desiccation-responsive rd29 gene of *Arabidopsis thaliana* (L.) and the analysis of its promoter in transgenic plants** (1993) *Mol. Gen. Genet.* 236: 331-340
 - Zepeda Garcia, **Tesis de Maestría en Ingeniería de Sistemas Empresariales** (2003), Universidad Ibero Americana, México.
 - Zhang HX. And Blumwald E.: **Transgenic salt.tolerant tomato plants accumulate salt in foliage but not in fruit** (2001). *Nat. Biotechnol.* 19:756-768
 - Zhang HX, Hodson JN, Williams JP, Blumwald E.: **Engineering salt-tolerant *Brassica* plants: characterization of yield and seed oil quality in**
-

transgenic plants with increased vacuolar sodium accumulation (2001).
Proc. Natl. Acad. Sci. USA 98:12832-12836

- Zhu, J.K: **Plant salt tolerance**. Trends in Plant Science (2001) (a), 6, 2, 66-71
- Zhu, J.K: **Cell signalling under salt, water and cold stresses**. Current Opinion in Plant Biology (2001) (b), 4:401-406.
- Zrenner R., Krause KP., Apel P., Sonnewald U.: **Reduction of the cytosolic fructose-1,6-bisphosphatase in transgenic potato plants limits photosynthesis sucrose biosynthesis with no impact on plant growth and tuber yield** (1996). Plant J 9:671–681