



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Departamento de Biotecnología

*Análisis transcripcional del desarrollo vegetativo
de cítricos y su regulación por giberelinas*

Memoria presentada por

LAURA HUERTA MARTÍNEZ

para optar al grado de

DOCTOR INGENIERO AGRÓNOMO

Directores

José Luis García Martínez
Miguel Ángel Pérez Amador

Valencia, febrero 2008

*A mis padres,
A Luis*

RESUMEN

El cultivo de los cítricos representa la mayor producción de frutos en el mundo y España, en particular, ocupa una posición relevante siendo el quinto productor y primer exportador mundial de cítricos para su consumo en fresco. La obtención de variedades de alta calidad comercial y elevada producción, capaces de satisfacer las exigencias del mercado, constituye un objetivo esencial de la citricultura actual. El desarrollo vegetativo determina la forma y el comportamiento general de la planta, de modo que tiene una influencia decisiva en la cantidad y calidad del fruto, así como en las prácticas culturales. Sin embargo, el conocimiento de las señales que regulan el desarrollo vegetativo en cítricos es limitado. En este trabajo se han desarrollado nuevas herramientas de genómica funcional, en particular colecciones de ESTs y micromatrices de DNA, con las que hemos estudiado en detalle diversos aspectos del desarrollo vegetativo en cítricos. Mediante el análisis funcional de una colección de ESTs generada a partir de tejidos vegetativos hemos aislado los genes relacionados con la función principal de estos tejidos, la fotosíntesis. Adicionalmente, a partir de la colección de ESTs del Proyecto de Genómica Funcional de Cítricos (CFGP) hemos identificado genes implicados en el metabolismo de isoprenoides y flavonoides en cítricos. La mayoría de los genes de la biosíntesis de fitohormonas derivadas de isoprenoides, como giberelinas (GAs), brasinosteroides y ácido abscísico no habían sido aislados antes en cítricos. La identificación de los genes de biosíntesis de monoterpenos y sesquiterpenos, responsables del aroma tanto de las hojas como del fruto, ha mostrado que la familia de terpeno sintasas en cítricos es una de las más amplias y diversas en plantas. El análisis de la expresión génica mediante micromatrices de DNA nos ha permitido elaborar mapas de expresión génica en los tejidos reproductivos y en los tejidos vegetativos, incluida la raíz. Mediante estos mapas hemos podido comprender mejor las funciones de cada uno de los tejidos analizados. Finalmente, hemos estudiado la regulación del desarrollo vegetativo mediada por GAs mediante el análisis de la expresión génica de plantas transgénicas del patrón de cítricos citrange Carrizo que sobre-expresan el gen *CcGA20ox1* (implicado en la biosíntesis de GAs). Los resultados muestran que la sobre-expresión de dicho gen conlleva una profunda modulación de la expresión génica. En concreto, hemos encontrado, por primera vez en plantas con un contenido elevado de GAs, una inducción global de los genes implicados en la fotosíntesis y en el ciclo de Calvin y una represión general de los genes relacionados con la síntesis de proteínas y con la biogénesis de ribosomas, lo que sugiere que dichos procesos están regulados por GAs. Además, hemos podido determinar que la activación transcripcional de los genes

relacionados con la fotosíntesis conduce a un aumento de la capacidad de fijación fotosintética de CO₂ de estas plantas. La expresión de genes que pertenecen a grupos funcionales no descritos previamente como grupos regulados por GAs (como estrés de origen biótico y abiótico o biosíntesis de la cutícula), así como de genes relacionados con la arquitectura de la pared celular se encuentra asimismo modificada en las plantas que sobre-expresan el gen *CcGA20ox1*. Las herramientas de genómica funcional generadas en este trabajo con su potencial para aislar genes de interés agronómico, para generar marcadores moleculares y para caracterizar colecciones de germoplasma, entre otros, podrían contribuir al desarrollo de nuevos cultivares de cítricos adaptados a las condiciones ambientales y cuya producción pueda satisfacer las características deseadas por los consumidores.

RESUM

El cultiu dels cítrics representa la major producció de fruits en el món i Espanya, en particular, ocupa una posició rellevant sent el cinqué productor i primer exportador mundial de cítrics per al seu consum en fresc. L'obtenció de varietats d'alta qualitat comercial i elevada producció, que s'adaptin a les exigències del mercat, constituïx un objectiu essencial de la citricultura actual. El desenvolupament vegetatiu determina la forma i el comportament general de la planta, i té una influència decisiva en la quantitat i qualitat del fruit, així com en les pràctiques culturals. Així i tot, el coneixement dels senyals que regulen el desenvolupament vegetatiu en cítrics és limitat. En aquest treball s'han desenvolupat noves eines de genòmica funcional, en particular col·leccions d'ESTs i micromatrius de ADN, amb les que hem estudiat en detall diferents aspectes del desenvolupament vegetatiu en cítrics. Mitjançant l'anàlisi funcional d'una col·lecció d'ESTs generada a partir de teixits vegetatius hem aïllat els gens relacionats amb la funció principal d'aquests teixits, la fotosíntesi. A més, a partir de la col·lecció d'ESTs del Projecte de Genòmica Funcional de Cítrics (CFGP) hem identificat gens implicats en el metabolisme d'isoprenoids i flavonoids en cítrics. La majoria dels gens de la biosíntesi de fitohormones derivades d'isoprenoids, com giberelines (GAs), brasinosteroids i àcid abscísic no havien sigut aïllats abans en cítrics. La identificació dels gens de biosíntesi de monoterpens i sesquiterpens, responsables de l'aroma de fulles i fruits, ha mostrat que la família de terpeno sintasas en cítrics és una de les més àmplies i diverses en plantes. L'anàlisi de l'expressió gènica per mitjà de micromatrius de ADN ens ha permès elaborar mapes d'expressió gènica en teixits reproductius i vegetatius, inclosa l'arrel. Mitjançant aquests mapes hem pogut comprendre millor les funcions de cadascun dels teixits analitzats. Per últim, hem estudiat la regulació del desenvolupament vegetatiu per GAs per mitjà de l'anàlisi de l'expressió gènica de plantes transgèniques del patró de cítrics citrange Carrizo que sobre-exprésen el gen *CcGA20ox1* (implicat en la biosíntesi de GAs). Els resultats mostren que la sobre-expressió d'aquest gen suposa una profunda modulació de l'expressió gènica. En concret, hem trobat, per primera vegada en plantes amb un contingut elevat de GAs, una inducció global dels gens implicats en la fotosíntesi i en el cicle de Calvin i una repressió general dels gens relacionats amb la síntesi de proteïnes i amb la biogènesi de ribosomes, suggerint que aquests processos estan regulats per GAs. A més, hem pogut determinar que l'activació transcripcional dels gens relacionats amb la fotosíntesi conduïx a un augment de la capacitat de fixació fotosintètica de CO₂ d'aquestes plantes. L'expressió de gens que pertanyen a grups funcionals no descrits prèviament com a grups regulats per GAs (com a estrès d'origen

biòtic i abiòtic o biosíntesi de la cutícula), així com de gens relacionats amb l'arquitectura de la paret cel·lular es troba així mateix modificada en les plantes que sobre-expressen el gen *CcGA20ox1*. Les eines de genòmica funcional generades en aquest treball, amb el seu potencial per a aïllar gens d'interès agronòmic, per a generar marcadors moleculars i per a caracteritzar col·leccions de germoplasma, entre altres, podrien contribuir al desenvolupament de nous cultivars de cítrics adaptats a les condicions ambientals i amb una producció capaç de satisfer les característiques desitjades pels consumidors.

ABSTRACT

Citrus are the most important fruit tree crop in the world and Spain is the fifth producer and the first world-wide exporter of citrus for fresh consumption. A main objective in citrus research is the improvement of citrus quality and production. Vegetative development is important in determining the shape and the general performance of trees, and has a decisive influence in the amount and quality of fruit production, as well as in culture practices. However, very little is known about the regulation of vegetative development in citrus. In this work, we have developed new genomic tools for large-scale studies of citrus vegetative growth, including the generation of an EST collection and the construction of cDNA microarrays. Genes involved in photosynthesis were identified by functional analysis of an EST collection generated from vegetative tissues. In addition, we have analyzed 85,965 ESTs from the Citrus Functional Genomic Project to characterize the isoprenoid and flavonoid metabolic pathways in citrus. Genes involved in gibberellin (GA), brassinosteroid and abscisic acid biosynthesis were identified for the first time in citrus. Identification of monoterpene and sesquiterpene synthases showed that citrus terpene synthase genes represents one of the largest and diverse terpene synthase family in plants. Expression maps of vegetative and reproductive organs were generated using cDNA microarrays. Major functions of each organ were characterized using these expression maps. The effect of GA on internode transcriptome was investigated in transgenic Carrizo citrange plants (a citrus hybrid widely used as a rootstock) overexpressing endogenous CcGA20ox1 (encoding a GA biosynthetic enzyme), using a citrus cDNA microarray. Substantial modulation of gene expression was found in these plants. Extensive upregulation of genes involved in photosynthesis and carbon utilization, and downregulation of those involved in protein synthesis and ribosome biogenesis was shown for the first time in plants with higher GA content, suggesting that these processes are regulated by GA. Importantly, increase of net photosynthesis in attached leaves was also demonstrated. In addition, the expression of other genes belonging to functional groups not reported previously to be regulated by GA (diverse abiotic and biotic stresses, and cuticle biosynthesis), as well as genes involved in cell wall architecture were also differentially expressed. Functional genomic tools developed in this work, with their potential to: i) isolate genes of agronomical relevance, ii) generate molecular markers and iii) characterize germplasm collections, will be of great help in the development of varieties that are well adapted to environmental conditions and produce fruits exhibiting those characteristics desired by most consumers.

INTRODUCCIÓN

1. Los cítricos.....	1
1.1. El género <i>Citrus</i> y su importancia económica.....	1
1.2. Desarrollo vegetativo de los cítricos	3
1.2.1. Aspectos generales.....	3
1.2.2. Control del desarrollo vegetativo.....	4
1.2.3. Influencia del patrón.....	5
1.3. Mejora genética en cítricos	6
2. Genómica funcional de plantas	7
2.1. Colecciones de ESTs: herramientas para la identificación de genes, el análisis de la expresión génica y el desarrollo de marcadores moleculares	8
2.2. Micromatrices de DNA: la revolución del análisis de la expresión génica	11
2.3. El Proyecto de Genómica Funcional de Cítricos (CFGP)	13
3. Aproximación genómico-funcional a diversos aspectos del desarrollo vegetativo en cítricos	15
3.1. Biosíntesis de isoprenoides y flavonoides	15
3.1.1. Biosíntesis de isoprenoides	15
3.1.2. Biosíntesis de flavonoides.....	18
3.2. Mapas de expresión génica en plantas	20
3.3. Control del desarrollo vegetativo: giberelinas	21
3.3.1. Metabolismo de giberelinas (GAs).....	21
3.3.2. Modificación de la arquitectura de las plantas mediante cambios en los niveles de GA 20-oxidasas.....	23
3.3.3. Identificación de genes de respuesta temprana a GAs	25
OBJETIVOS	29

MATERIALES Y MÉTODOS

1. Material vegetal y tratamientos hormonales	31
1.1. Material de campo	31
1.2. Material de invernadero	32
2. Extracción de RNA y RNA poli(A ⁺) de plantas.....	32
2.1. Protocolo general de extracción de RNA total	33
2.1.1. Purificación adicional del RNA para los experimentos con micromatrices.....	33

2.1.2. Purificación adicional del RNA para los experimentos de RT-PCR.....	33
2.2. Extracción de RNA poli(A ⁺).....	34
3. Construcción y manipulación de una genoteca de cDNA	34
3.1. Genoteca primaria	34
3.2. Genoteca amplificada	35
3.3. Escisión masiva	36
4. Generación de una colección de ESTs	36
4.1. Obtención de DNA plasmídico y secuenciación	36
4.1.1. Cultivos bacterianos.....	36
4.1.2. Obtención de DNA plasmídico	37
4.1.3. Secuenciación.....	37
4.2. Pre-procesado y ensamblaje de ESTs	38
4.3. Anotación funcional de los unigenes	39
4.3.1. Metodología para la anotación básica de los unigenes: base de datos EST2uni.....	39
4.3.2. Métodos para la anotación adicional de los unigenes	39
4.4. Análisis de expresión génica según la frecuencia de ESTs	41
5. Construcción de la micromatriz de cDNA de cítricos CIT1	42
6. Experimentos de micromatrices de cDNA.....	43
6.1. Diseño de experimentos	44
6.2. Preparación de muestras de cDNA marcadas.....	46
6.2.1. Marcaje mediante el método indirecto	46
6.2.2. Caracterización de las propiedades del cDNA marcado.....	47
6.3. Hibridación de las micromatrices	47
6.3.1. Protocolo de hibridación A	48
6.3.2. Protocolo de hibridación B	49
6.4. Adquisición y análisis de la imagen	50
6.5. Pre-procesado y normalización de los datos	51
6.6. Evaluación de la calidad de los datos obtenidos con micromatrices	53
6.6.1. Parámetros de calidad de GenePix	53
6.6.2. Validación técnica de la micromatriz CIT1	54
6.7. Identificación de los genes expresados diferencialmente.....	55
6.7.1. Test de la <i>t</i> de Student.....	55
6.7.2. Análisis SAM.....	55
6.7.3. El problema de las pruebas múltiples	56
6.8. Análisis de agrupamiento y análisis funcional de los datos	56

7. Análisis de expresión de mRNA	57
7.1. Análisis Northern.....	57
7.1.1. Electroforesis en gel de agarosa desnaturalizante	57
7.1.2. Transferencia de RNA a membranas de nylon	57
7.1.3. Obtención de las sondas de DNA	58
7.1.4. Hibridación con sondas radiactivas.....	58
7.2. RT-PCR semicuantitativa.....	59
8. Medidas de fotosíntesis	60

RESULTADOS Y DISCUSIÓN I: Análisis funcional de ESTs de cítricos

1. Genes implicados en el desarrollo vegetativo en cítricos.....	61
1.1. Obtención de una genoteca de tejidos vegetativos	61
1.2. Colección de ESTs: secuenciación y agrupamiento.....	62
1.3. Anotación funcional de los unigenes	64
1.4. Ontologías génicas enriquecidas en Veg1	65
1.4.1. Fotosíntesis.....	69
1.4.2. Fotosíntesis: proteínas de la antena.....	70
1.4.3. Fijación de carbono.....	73
2. Genes implicados en el metabolismo de isoprenoides y flavonoides en cítricos	74
2.1. Biosíntesis de isoprenoides	75
2.1.1. Biosíntesis de isopentenil difosfato y dimetilalil difosfato.....	75
2.1.2. Prenildifosfatos: precursores de los distintos tipos de terpenoides	79
2.1.3. Biosíntesis de monoterpenos y sesquiterpenos: aceites esenciales	83
2.1.4. Biosíntesis de triterpenos: brasinosteroides y limonoides	88
2.1.5. Biosíntesis de diterpenos: giberelinas.....	92
2.2. Biosíntesis de flavonoides	94

RESULTADOS Y DISCUSIÓN II: Mapas transcripcionales en cítricos

1. Caracterización de la micromatriz de cDNA de cítricos, CIT1, generada en el CFGP	99
1.1. Caracterización técnica de la micromatriz CIT1	99
1.1.1. Rango dinámico y niveles de sensibilidad	100
1.1.2. Reproducibilidad <i>intra-array</i>	101
1.1.3. Reproducibilidad <i>inter-array</i>	103

1.2. Validación de la micromatriz CIT1 como herramienta para identificar cambios en la expresión génica en cítricos	104
2. Caracterización de diferentes tejidos de cítricos mediante el análisis global de la expresión génica.....	108
2.1. Identificación de los genes comúnmente expresados en diversos tejidos de cítricos.....	108
2.2. Relación entre los tejidos estudiados según los perfiles de expresión génica	110
2.3. Identificación de los genes preferentemente expresados en la raíz.....	111
2.4. Identificación de los genes preferentemente expresados en los tejidos reproductivos	114
2.5. Identificación de los genes preferentemente expresados en los tejidos vegetativos.....	118

RESULTADOS Y DISCUSIÓN III: Regulación del transcriptoma de cítricos por giberelinas

1. Cambios en la expresión génica de entrenudos de plantas transgénicas de cítricos que sobre-expresan una GA 20-oxidasa	121
1.1. La sobre-expresión en sentido y en antisentido del gen <i>CcGA20ox1</i> modifica los niveles de transcritos endógenos de <i>CcGA20ox1</i>	121
1.2. La sobre-expresión en sentido del gen <i>CcGA20ox1</i> causa un remodelado substancial del transcriptoma	122
1.3. La sobre-expresión en sentido del gen <i>CcGA20ox</i> implica la sobre-representación de diversas ontologías génicas.....	127
1.3.1. Fotosíntesis y Utilización de Carbono.....	129
1.3.2. Biosíntesis de Proteínas	131
1.4. La activación transcripcional de los genes relacionados con la fotosíntesis conduce a un aumento en la capacidad fotosintética de las plantas <i>CcGA20ox</i>	132
1.5. La sobre-expresión en sentido del gen <i>CcGA20ox1</i> provoca cambios en los niveles de expresión de genes específicos.....	134
1.5.1. Metabolismo de GAs.....	134
1.5.2. Biosíntesis y modificación de la pared celular	134
1.5.3. Biosíntesis de lignina	136
1.5.4. Estrés de origen abiótico.....	137
1.5.5. Estrés de origen biótico.....	140

1.5.6. Factores de transcripción.....	140
2. Cambios en la expresión génica de entrenudos de explantes tratados con GA ₃ ...	142
2.1. La aplicación de GA ₃ a tiempos cortos altera el transcriptoma de los entrenudos de explantes de cítricos	142
CONCLUSIONES	147
BIBLIOGRAFÍA	149
ANEJOS	
Anejo I.....	173
Anejo II.....	193
Anejo III.....	201



Introducción

1. LOS CÍTRICOS

1.1. El género *Citrus* y su importancia económica

Las principales especies comerciales de cítricos son árboles de hoja perenne, con orígenes tropicales y subtropicales y que pertenecen a la familia *Rutaceae*, subfamilia *Aurantioideae*. Las plantas de esta subfamilia se caracterizan por un fruto en forma de baya en hesperidio procedente de un ovario multicarpelar, que contiene vesículas de zumo y una corteza coriácea con alta densidad de glándulas esenciales. Una segunda característica de muchas de las especies de esta subfamilia es la presencia de semillas poliembriónicas que contienen tanto embriones cigóticos como nucelares, producidos sin fecundación previa a partir de la nucela, mediante el fenómeno de la apomixis (Davies y Albrigo, 1994). La ordenación de la subfamilia *Aurantioideae* resulta compleja debido a que los cítricos hibridan con bastante facilidad, incluso entre diferentes géneros, tienen una elevada tasa de mutaciones espontáneas y, como consecuencia de la apomixis, las variaciones genéticas tienden a perpetuarse (Davies y Albrigo, 1994). Los tres géneros de cítricos cultivados en la actualidad son: 1) *Fortunella* (kumquat), 2) *Poncirus*, formado por especies de hoja trifoliada caduca, de gran resistencia al frío y utilizados como portainjertos, y 3) *Citrus*, que incluye las especies más importantes desde el punto de vista agronómico como las naranjas dulces (*Citrus sinensis* [L.] Osb.), mandarinas (*Citrus reticulata* Blanco y *Citrus unshiu* Marc.), pomelos (*Citrus paradisi* Macf.), limones (*Citrus limon* Burm. f.) y limas (*Citrus aurantifolia* L.). La sistemática del género *Citrus* también es controvertida ya que según Swingle y Reece (1967) este género está formado por 16 especies mientras que según Tanaka (1977) comprende 162 especies. Recientemente, diversos trabajos tratan de evaluar la variabilidad genética del género *Citrus* mediante marcadores moleculares (Moore et al, 2001; Barkley et al 2006; Ancillo et al, 2007). La gran mayoría de las especies de los géneros *Citrus*, *Fortunella* y *Poncirus* son diploides, con un número básico de cromosomas $x = 9$ ($2x = 2n = 18$), excepto *Fortunella hindsii* que es tetraploide ($4x = 2n = 36$). El tamaño estimado del genoma de cítricos es de 382 Mb (Arumuganathan et al, 1991), inferior al de arroz (419 Mb) (Goff et al, 2002; Yu et al, 2002) y unas tres veces superior al tamaño del genoma de *Arabidopsis* (125 Mb) (*Arabidopsis* Genome Initiative, 2000).

El área general de origen de los cítricos se localiza en el sureste de Asia, en una zona que abarca desde Arabia Oriental hasta Filipinas, y desde el Himalaya hasta Indonesia. Dentro de esta gran región, el noreste de la India, el norte de Birmania y la provincia de Yunnan (en el centro-sur de China), se consideran los tres centros de

origen más importantes (Gmitter y Hu, 1990). Actualmente el cultivo de los cítricos se extiende por la mayor parte de las regiones tropicales y subtropicales comprendidas entre las latitudes 44° N y 41° S (Agustí, 2000). Los cítricos se cultivan en más de un centenar de países, en una superficie total de 6.7 millones de hectáreas y alcanzan una producción de 108 millones de toneladas, cifra muy superior a la de otras frutas como los plátanos (70 millones de toneladas) o las uvas (69 millones) (Tabla Suplementaria I) (FAO, 2006; <http://www.fao.org>). España, con una producción de poco más de 6 millones de toneladas, ocupa el quinto lugar en la producción mundial de cítricos, precedido de Brasil (20 millones de toneladas), China (supera los 17 millones de toneladas), EEUU (11 millones de toneladas) y México (6.5 millones de toneladas). La producción de naranjas es la mayoritaria a nivel mundial (60%), seguida de la de mandarinas (24%), limones (12%) y pomelo (4%) (Tabla Suplementaria I) (FAO, 2006). Las exportaciones mundiales de cítricos superan los 12 millones de toneladas (FAO, 2005) y España es el país con el mayor volumen de exportación de cítricos para su consumo en fresco del mundo con más de 3 millones de toneladas, seguido por Suráfrica y EEUU con 2 y 1 millón de toneladas, respectivamente (Tabla Suplementaria I) (FAO, 2005). Esto significa que España puede canalizar en torno al 25% del comercio mundial de cítricos. España es el primer exportador mundial de mandarinas y limones (Tabla Suplementaria I) y su mercado más importante es la UE, principalmente Alemania y Francia, países a los que destina el 50% de la exportación (Tabla Suplementaria I) (FAO, 2005). La producción de cítricos en España se localiza en cuatro comunidades autónomas: la Comunidad Valenciana, que constituye la zona de producción más importante (62% de la producción nacional), Andalucía (22%), Región de Murcia (13.5%) y Cataluña (2.5%) (MAPA, 2006; <http://www.mapa.es>). Los cítricos de mayor producción en España son las naranjas (53%), seguidas de las mandarinas (33%) y los limones (14%) y la Comunidad Valenciana lidera la producción de naranjas y mandarinas (MAPA, 2006). En el grupo de las mandarinas, destaca la producción de clementinas (82%), especialmente de la variedad clementino de Nules (*Citrus clementina* Hort. ex Tan.).

La importancia de la citricultura en el mundo, y en particular en España, está fuera de toda duda. Sin embargo, no se halla exenta de serios problemas como techos en el consumo y producciones crecientes que conducen a una caída de precios. El avance de la citricultura implica una mejora de la rentabilidad de las explotaciones que pasa por la reducción de los costes de cultivo, la reducción de la estacionalidad de la oferta o la mejora de la calidad del fruto (Agustí, 2000). La obtención de variedades de alta calidad comercial y elevada producción, capaces de satisfacer las exigencias del mercado constituye un objetivo esencial ligado a la mejora del material vegetal.

1.2. Desarrollo vegetativo de los cítricos

1.2.1. Aspectos generales

En condiciones subtropicales los cítricos presentan un crecimiento anual discontinuo con periodos de crecimiento efectivo que se alternan con periodos de inactividad vegetativa. Estos periodos de crecimiento e inactividad vegetativa están ligados a los cambios climáticos anuales como las condiciones de humedad, temperatura, fotoperiodo e irradiación. En las condiciones de clima mediterráneo los cítricos entran en latencia, aunque sin perder sus hojas, hacia mediados de otoño, concluyendo dicho periodo de reposo invernal a finales de invierno o principios de primavera (Davies y Albrigo, 1994). Durante el periodo de actividad vegetativa, el desarrollo de la parte aérea tampoco se produce de forma continua sino en ciclos definidos denominados brotaciones. En nuestras condiciones climáticas se producen tres brotaciones: una en primavera, otra en verano y otra en otoño. La brotación de primavera tiene lugar en marzo o abril tras el periodo de latencia y es muy importante puesto que coincide con el principal periodo de floración, por lo que determina la cosecha. Los brotes que surgen en primavera se clasifican de acuerdo con el número de hojas y flores que los forman de modo que pueden distinguirse: 1) ramos de flor o brotes multiflorales sin hojas, 2) brotes mixtos o brotes multiflorales con varias hojas, 2) flores solitarias o brotes uniflorales sin hojas, 4) brotes campaneros o brotes uniflorales con hojas y 5) brotes vegetativos o brotes que únicamente están formados por hojas (Agustí, 2000). La segunda brotación, a principios de verano (junio), y la tercera (septiembre), que precede a un nuevo periodo de inactividad vegetativa, son exclusivamente vegetativas en condiciones normales y determinan la envergadura del árbol. Estas dos brotaciones difieren de la de primavera no sólo por la ausencia de flores, sino porque las hojas son de mayor tamaño y los brotes, aunque en menor número, son de mayor longitud y vigor. El crecimiento vegetativo depende, por tanto, de la segunda y tercera brotación (Davies y Albrigo, 1994). Los factores que determinan la brotación no son muy bien conocidos. No parece existir requerimientos de frío y las yemas brotan en cualquier época del año cuando la temperatura del suelo supera los 12°C, con independencia de la del aire (Agustí, 2000).

El desarrollo de los árboles de cítricos pasa por dos etapas: el periodo juvenil y el adulto. El periodo juvenil es relativamente prolongado (puede oscilar entre 5 y 10 años) y su duración varía para cada especie y según factores ambientales como la temperatura y la humedad. El periodo juvenil se define como aquél en el que la planta presenta una capacidad de crecimiento vegetativo elevada, lo que se traduce en un

mayor vigor, es incapaz de inducir el proceso de floración, por lo que se produce una entrada en producción tardía, y desarrolla espinas, estructuras morfológicas características de este periodo (Agustí, 2000). En la fase juvenil el hábito de crecimiento del árbol es muy erecto, sin apenas ramificaciones. Durante el crecimiento del árbol, el meristemo apical se desprende y surgen yemas laterales a lo largo del eje central. Los nuevos tallos surgen de yemas vegetativas laterales y, a su vez, el meristemo apical de estos tallos se desprenderá y emergerán otras yemas laterales. Este patrón de crecimiento en zig-zag, llamado crecimiento simpodial, es característico de la mayoría de las especies de cítricos (Davies y Albrigo, 1994). La mayor tasa de crecimiento vegetativo durante el periodo juvenil se ha relacionado con una ausencia de competencia entre puntos de desarrollo de modo que cuando estos puntos son numerosos la competencia por nutrientes aumenta, aparece una gradual pérdida de la dominancia apical y una orientación geotrópica de los tallos que marca el final del periodo juvenil (Agustí, 2000). Dado que la elongación de los tallos vegetativos termina con la abscisión de la yema terminal, los árboles tienen un hábito de crecimiento determinado (Davies y Albrigo, 1994). Las hojas de los cítricos comienzan su expansión antes de que finalice la elongación del tallo. Las hojas en desarrollo son generalmente importadoras de carbono, hasta su total expansión. La asimilación neta de CO₂ aumenta durante el desarrollo de la hoja, alcanza un máximo a los 3-4 meses de edad y comienza a descender a partir de los 6 meses. Pese a ello, las hojas continúan activas hasta que caen, lo que puede suceder tras dos años. Además de su papel de asimilación de carbono, las hojas funcionan como órganos de reserva de nutrientes. El principal carbohidrato de reserva es el almidón. Mientras que las hojas de las plantas anuales muestran un claro ciclo diurno en el que todo el almidón se elimina durante la noche, las hojas de los cítricos conservan la mayor parte del almidón de modo que sólo se observan pequeñas fluctuaciones diurnas (Spiegel-Roy y Goldschmidt, 1996).

1.2.2. Control del desarrollo vegetativo

Las condiciones ambientales, de suma importancia en el desarrollo vegetativo, operan a través de su influencia sobre el balance hormonal y la movilización de reservas. La temperatura es, probablemente, la variable climática de mayor influencia en el desarrollo de los cítricos y determina los periodos de inactividad vegetativa. La temperatura del suelo influye sobre el número de nuevos brotes y la temperatura del aire durante el día influye sobre la longitud de los brotes (Agustí, 2000). El desarrollo

óptimo de los brotes se produce entre los 22 y 34°C y no se relaciona con la alternancia de temperaturas entre el día y la noche (Ben-Cheikh, 1997a).

La intensidad de la luz también afecta al crecimiento y desarrollo vegetativo. La luz se ha relacionado con el desarrollo de nuevos brotes de modo que en las zonas internas del árbol, con una intensidad luminosa limitada, la brotación es menor. La intensidad de la luz tiene un efecto directo sobre la asimilación neta de CO₂ por las hojas y un efecto indirecto sobre la temperatura de las mismas. El crecimiento vegetativo de los cítricos está estrechamente relacionado con la asimilación neta de CO₂, suponiendo que otros factores como la temperatura, nutrición y agua no sean limitantes. La asimilación neta de CO₂ aumenta de modo lineal con el aumento de la radiación fotosintética activa (PAR) hasta alcanzar un máximo de eficiencia fotosintética entre 700 y 1000 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ y las mesetas por encima de este nivel se conocen como nivel de saturación de luz (Syvertsen, 1984). La asimilación neta de CO₂ será potencialmente mayor cuanto más tiempo permanezca el árbol en condiciones de saturación de luz. Por lo tanto, el crecimiento vegetativo de los cítricos suele ser mayor cuando la duración del día es de moderada a larga (>12 h), cuando se logran las mayores unidades de calor y cuando el agua no es un factor limitante. Sin embargo, intensidades de luz extremadamente altas pueden provocar una disminución de la asimilación neta de CO₂ debido a que conllevan un aumento de la temperatura de la hoja. Las temperaturas óptimas de asimilación neta de CO₂ en los cítricos oscilan entre 25 y 30°C. Las temperaturas superiores a 35°C reducen la actividad fotosintética porque limitan la actividad de la ribulosa 1,5-bisfosfato carboxilasa/ oxidasa (RuBisCO) (Davies y Albrigo, 1994).

La disponibilidad, movilización y suministro de nutrientes y agua al brote en desarrollo son esenciales para la síntesis de los constituyentes celulares y, por tanto, para su crecimiento. Así pues, la elongación y la acumulación de materia seca en los brotes, en condiciones no limitantes de disponibilidad de agua y nutrientes, se ven favorecidas por días largos y temperaturas medias moderadamente altas, tanto durante el día como por la noche (Davies y Albrigo, 1994). La correlación entre la síntesis y el transporte de hormonas es la responsable del crecimiento periódico de brotes y ramas. En particular, la aplicación de giberelinas al inicio de la brotación estimula el crecimiento de los brotes iniciados y de sus hojas (Agustí, 2000).

1.2.3. Influencia del patrón

Los árboles de cítricos cultivados en la actualidad están formados por dos partes, el patrón y la variedad, la segunda injertada sobre el primero, de modo que

combinen entre sí las mejores características posibles. La elección del patrón es fundamental porque: 1) aporta a la planta el sistema radicular, que ancla la planta al suelo, absorbe agua y nutrientes, acumula carbohidratos y sintetiza algunas hormonas, 2) permite la adaptación de la variedad a las condiciones particulares del suelo confiriendo tolerancia a factores desfavorables como caliza, sequía y salinidad y 3) confiere tolerancia a ciertas enfermedades. El patrón influye en muchas de las características de una variedad como el vigor y el tamaño del árbol, el desarrollo y la profundidad de las raíces, la cosecha y el tamaño y la calidad del fruto (Agustí, 2000).

Uno de los patrones más extendidos en la citricultura española actual es citrange Carrizo de modo que un 90% de las nuevas plantaciones están injertadas sobre este patrón. El patrón citrange Carrizo (*Citrus sinensis* L. Osbeck x *Poncirus trifoliata* L. Raf.) es un híbrido obtenido por Savage en 1909 polinizando flores de 'Washington Navel' con polen de *Poncirus trifoliata* (Savage y Gadner, 1965). Se trata de un patrón tolerante a tristeza, psoriasis y xiloporosis, pero sensible a exocortis. Presenta sensibilidad al hongo *Armillaria mellea* y una resistencia moderada a *Phytophthora* spp. En cuanto a su comportamiento frente a las condiciones del medio es sensible a la caliza y a la salinidad. Posee buena compatibilidad con todas las variedades de naranjo dulce, mandarino y pomelo cultivadas en España y las variedades injertadas sobre citrange Carrizo presentan excelentes cosechas de elevada calidad comercial (Agustí, 2000).

Los patrones de cítricos presentan distintos efectos en relación con el tamaño que inducen en las variedades injertadas sobre ellos de modo que algunos patrones reducen el tamaño final del árbol. La utilización de patrones enanizantes o semi-enanizantes puede aportar ciertas ventajas como el aumento de la producción por unidad de superficie mediante plantaciones de alta densidad, la reducción de los costes del cultivo derivado de la poda, los tratamientos fitosanitarios y la recolección, la adaptación a pequeñas explotaciones y la posibilidad de mecanización del trabajo. Dada la difusión mayoritaria del portainjerto citrange Carrizo, la posibilidad de disponer de plantas de citrange Carrizo enanizantes o semi-enanizantes constituye un objetivo de gran interés.

1.3. Mejora genética en cítricos

La mejora genética de los cítricos, tanto de la variedad como del patrón, es muy costosa, principalmente por la superficie y el tiempo (20 años o más) necesarios para la evaluación en campo de las características agronómicas de las plantas mejoradas. Además, las características de la biología reproductiva de los cítricos,

como la apomixis, la incompatibilidad sexual y la elevada heterocigosis dificultan la mejora del material vegetal. La apomixis es un fenómeno frecuente en los cítricos mediante el cual se producen embriones sin fecundación previa. Estos embriones se forman a partir de la nucela (embriones nucelares) y las semillas que los contienen originan plantas cuya constitución genética es idéntica a la de la madre. En las especies o variedades apomícticas se producen en el mismo rudimento seminal los dos procesos, sexual y asexual, formándose semillas que contienen un embrión zigótico y uno o varios nucelares. Frecuentemente, los embriones nucelares son más vigorosos que el zigótico, que en muchos casos no completa su desarrollo y aborta. La apomixis complica la selección de híbridos. Sin embargo, es un carácter apreciado para la propagación de patrones uniformes e idénticos a los parentales (Davies y Albrigo, 1994). Además, muchas especies presentan esterilidad total o parcial en óvulos y/o polen, lo que impide su uso como parentales en programas de mejora mientras que otras especies son sexualmente incompatibles, dificultando la obtención de híbridos. Finalmente, la mayoría de las especies de cítricos son muy heterocigóticas y producen una progenie sexual muy variable (Herrero et al, 1996).

Esta combinación de factores adversos ha dificultado los esfuerzos de mejora de los cítricos mediante los métodos de mejora tradicional, de modo que la mayoría de los cítricos de interés comercial son el resultado de mutaciones espontáneas en los árboles de campo. En este contexto, los abordajes de genómica funcional, con su potencial para aislar genes de interés agronómico, para generar marcadores moleculares y para caracterizar colecciones de germoplasma, entre otros, podrían contribuir al desarrollo de nuevos cultivares de cítricos adaptados a las condiciones ambientales y cuya producción pueda satisfacer las características deseadas por los consumidores (Forment et al, 2005).

2. GENÓMICA FUNCIONAL DE PLANTAS

La secuencia genómica de una planta permite comprender la base genética de las diferencias entre las plantas y otros eucariotas, y sienta los cimientos para la caracterización detallada de los genes que constituyen una planta. El primer genoma secuenciado en plantas fue el de *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), seguido por el de arroz (Goff et al, 2002; Yu et al, 2002; International Rice Genome Sequencing Project, 2005) y, recientemente, por el de *Populus trichocarpa* (Tuskan et al, 2006) y *Vitis vinifera* (Jaillon et al, 2007; Velasco et al, 2007). La secuenciación del genoma de otras especies vegetales de interés agronómico como trigo, maíz, soja, tomate, patata o naranjo, entre otras, se está llevando a cabo en la

actualidad. En particular, el *DOE Joint Genome Institute* (JGI) ha iniciado un proyecto preliminar de secuenciación a baja cobertura de naranjo dulce (*Citrus sinensis* L. Osbeck). La interpretación de un genoma pasa por la anotación tanto a nivel estructural, con la identificación de los genes que lo integran, como a nivel funcional, con la asignación de funciones a dichos genes. La anotación de un genoma es un proceso dinámico e iterativo. Por ejemplo, el genoma de *Arabidopsis* continua siendo re-anotado por *The Institute for Genomic Research* (TIGR) desde que se completó (Haas et al, 2003; Wortman et al, 2003). Un punto crítico en la anotación de un genoma es la definición de la estructura de los modelos génicos, por ejemplo, del límite intrón-exón y de las regiones no traducibles. La construcción de los modelos génicos depende de la información procedente de las colecciones de ESTs (*expressed sequence tags*) y de secuencias de cDNA de longitud completa. Por otro lado, la anotación funcional de un genoma depende del uso de la bioinformática para inferir la función génica ya que sólo se dispone de resultados experimentales que pueden utilizarse en esta asignación para una proporción pequeña de los genes de un genoma (Rensink y Buell, 2005).

La secuenciación y anotación, tanto estructural como funcional, de un genoma es un proceso costoso lo que explica que dicho esfuerzo se haya llevado a cabo sólo en algunas especies vegetales. Sin embargo, la secuenciación de genotecas de cDNA y el análisis de las ESTs resultantes constituye el modo más directo de obtener información de un genoma a nivel del RNA, de modo que una de las aplicaciones más extendida de la genómica en plantas ha sido la generación de colecciones de ESTs.

2.1. Colecciones de ESTs: herramientas para la identificación de genes, el análisis de la expresión génica y el desarrollo de marcadores moleculares

Las ESTs constituyen la información de secuencia más abundante de los genomas de plantas, tanto en términos de número de secuencias como de cantidad total de nucleótidos. Las ESTs son secuencias parciales obtenidas de la secuenciación del extremo 5' y/o 3' de clones de cDNA aislados al azar (Adams et al, 1991). Una genoteca de cDNA contiene decenas de miles de clones que representan la porción transcrita de un genoma en una condición definida (un tejido, un estadio del desarrollo concreto y en respuesta a un determinado estímulo ambiental o una situación de estrés biótico o abiótico). A pesar de que una EST representa sólo una porción de la secuencia codificante (aproximadamente de 500 pb), la obtención masiva de estas secuencias parciales es de gran utilidad. La generación de una colección de ESTs constituye un método relativamente rápido y potente para la identificación de genes,

para la confirmación de las regiones codificantes en secuencias genómicas, para el estudio de relaciones filogenéticas, para la estimación de los niveles de expresión de los genes de los que proceden y para el desarrollo de micromatrices de DNA. Adicionalmente, los proyectos de secuenciación de ESTs pueden llegar a identificar una proporción significativa del contenido génico de un organismo de modo que pueden suponer el inicio de proyectos de secuenciación de un genoma (Alba et al, 2004). Actualmente, la base de datos de ESTs del NCBI (*National Center for Biotechnology Information*) contiene más de 49 millones de ESTs, de las que más de 12 millones derivan de plantas (dbEST 11 enero 2008; <http://www.ncbi.nlm.nih.gov/dbEST/>; Boguski et al, 1993). Las colecciones de ESTs más exhaustivas en plantas corresponden a la especie modelo *Arabidopsis* con 1.5 millones de ESTs, seguida de varios cereales fundamentales para la alimentación humana como arroz, maíz o trigo, con más de un millón de ESTs en cada caso.

Una de las aplicaciones iniciales de las colecciones de ESTs fue la identificación de nuevos genes implicados en rutas metabólicas específicas de las plantas, como genes relacionados con algunas modificaciones de los ácidos grasos, genes de la biosíntesis de isoprenoides o de la biosíntesis de la pared celular (revisado por Ohlrogge y Benning, 2000). En este sentido, una cuestión típica es cuáles son los genes que se expresan en una condición definida (tejido, estadio del desarrollo concreto y en respuesta a un determinado estímulo).

Aunque el primer propósito de la secuenciación de ESTs suele ser la caracterización de los genes de un organismo, una colección de ESTs puede aportar información sobre la expresión génica. Este aspecto del análisis de ESTs, llamado expresión digital, se basa en que si un gen X tiene un nivel de expresión elevado (lo que implica niveles elevados de mRNA_X), el número de cDNAs obtenido a partir del mRNA_X será alto en la genoteca realizada con el material vegetal en el que se expresa el gen X. Tras la secuenciación al azar de un gran número de clones de la genoteca de cDNA, la proporción de ESTs que corresponde al gen X supondrá una estimación de la abundancia del mRNA_X en la población original (Ohlrogge y Benning, 2000). Los análisis de expresión digital se apoyan en el desarrollo de métodos estadísticos robustos que permiten valorar la fiabilidad de la información obtenida (Audic y Claverie, 1997; Romualdi et al, 2003). El agrupamiento de los genes en función de su abundancia relativa en diferentes genotecas de cDNA permite obtener patrones de expresión génica de modo que los genes con perfiles de expresión similares en las diferentes condiciones estudiadas quedan agrupados. Además, los propios tejidos y/o condiciones estudiadas pueden agruparse en función de la expresión génica. En plantas, se han identificado genes diferencialmente expresados en distintos tejidos

mediante el análisis de colecciones de ESTs en arroz (Ewing et al, 1999), patata (Ronning et al, 2003), trigo (Ogihara et al, 2003) y cebada (Zang et al, 2004). Asimismo, esta metodología se ha desarrollado para la identificación de genes relacionados con la maduración del fruto en tomate (Fei et al, 2004) y en manzano (Park et al, 2006) y con el desarrollo de la flor y del fruto en vid (da Silva et al, 2005). Además, la caracterización del gran número actual de colecciones de ESTs disponibles para multitud de plantas hace posible la comparación entre diferentes especies.

Las colecciones de ESTs constituyen también una valiosa herramienta para el desarrollo de marcadores moleculares, en particular en especies de interés agronómico para las que todavía no se dispone de la información derivada de la secuenciación del genoma. La selección asistida por marcadores moleculares constituye una estrategia clave en la mejora genética de plantas. Entre los marcadores moleculares que pueden desarrollarse a partir de ESTs destacan los polimorfismos de un solo nucleótido (SNPs, *single-nucleotide polymorphisms*) y las repeticiones de secuencia simple (SSRs, *simple sequence repeats*). Las posibles funciones de los genes en los que se encuentran los marcadores derivados de ESTs pueden deducirse por similitud de secuencia con bases de datos de proteínas (por ejemplo, Swissprot) (Varshney et al, 2005). Los marcadores moleculares derivados de sitios polimórficos en los genes responsables de la variación de los rasgos fenotípicos se denominan marcadores funcionales (FMs) (Andersen y Lübberstedt, 2003). Los FMs se han desarrollado en muchas especies vegetales para las que se dispone de datos de ESTs (Gupta y Rustgi, 2004). Rudd et al (2005) demostraron la viabilidad de la predicción de marcadores moleculares (SSRs, SNPs) en más de 50 especies vegetales a partir de los datos de ESTs (PlantMarkers, <http://markers.btk.fi/>). Recientemente, se han identificado posibles SSRs y SNPs mediante el análisis de colecciones de ESTs en manzano (Newcomb et al, 2006) y en melón (González-Ibeas et al, 2007). La mayor ventaja de los FMs sobre otros marcadores generados a partir de regiones anónimas del genoma (RMs, *random markers*) es que están completamente ligados al rasgo de interés puesto que derivan del gen responsable de dicho carácter (Varshney et al, 2005).

Una colección de ESTs presenta ciertas limitaciones. El hecho de que las ESTs representen el conjunto de genes transcritos en unas determinadas condiciones implica que los genes con bajos niveles de expresión estarán poco representados en la colección de ESTs. La presencia de una EST constituye un atributo fiable mientras que su ausencia sólo permite concluir que en las condiciones a partir de las que se ha generado la colección de ESTs no ha sido posible detectar dicho transcrito

(Rudd, 2003). Esta limitación de la representación génica de una colección de ESTs se ha abordado mediante estrategias de normalización y/o sustracción de las genotecas de cDNA, que permiten equilibrar la proporción relativa de transcritos comunes y raros (Bonaldo et al, 1996). Asimismo, la secuenciación de clones procedentes de diferentes genotecas de cDNA generadas en distintas condiciones (tejidos, estadios de desarrollo, respuesta a estrés) permite maximizar el conjunto de genes muestreados, mientras que la secuenciación exhaustiva de una sola genoteca de cDNA permitirá obtener transcritos con bajos niveles de expresión. Otras propiedades inherentes a las ESTs como la calidad de secuencia, la redundancia o la naturaleza parcial (aproximadamente de 500 pb) implican que su manejo dependa de numerosas herramientas bioinformáticas que, en último término, aportan un valor añadido a las colecciones de ESTs (revisado por Dong et al, 2005). En los últimos años se han desarrollado numerosas bases de datos de ESTs de plantas que han añadido la anotación funcional de las ESTs y que permiten seleccionar y obtener información sobre grupos de secuencias que satisfacen criterios de búsqueda específicos. Recientemente se ha creado la base de datos TA (*TIGR Plant Transcript Assemblies*, <http://plantta.tigr.org>), que se nutre de las secuencias de ESTs y cDNAs disponibles en la base de datos de nucleótidos GenBank del NCBI para la generación de un conjunto de transcritos ensamblados y anotados (Childs et al, 2007). La versión actual de la base de datos TA (versión 5) contiene más de 11 millones de ESTs procedentes de 254 especies vegetales.

2.2. Micromatrices de DNA: la revolución del análisis de la expresión génica

Los mayores avances tecnológicos para la medida de la expresión génica se han producido en la última década. En particular, la tecnología de las micromatrices ha revolucionado la biología molecular, de modo que actualmente es la más utilizada en experimentos de expresión génica de gran escala puesto que permite la detección simultánea de miles de transcritos a un coste razonable (Busch y Lohmann, 2007). La capacidad de medir la expresión génica de todos los genes de un organismo constituye un potente sistema analítico y la disponibilidad de la tecnología necesaria para ello ha abierto nuevas oportunidades a la comunidad científica (Meyers et al, 2004). La tecnología de las micromatrices se apoya en la existencia de colecciones de ESTs o datos de secuencia genómica (y, por tanto, está limitada por la disponibilidad de los mismos), en instrumentación robótica con capacidad de trabajo a muy pequeña escala y en compuestos fluorescentes para la detección simultánea de poblaciones de transcritos derivadas de diferentes muestras (Alba et al, 2004). Además, el conjunto

masivo de datos generado en un experimento de análisis de la expresión génica mediante micromatrices ha presentado nuevos problemas estadísticos y analíticos que han supuesto la convergencia de la biología, las matemáticas y la bioinformática (Meyers et al, 2004).

Una micromatriz de DNA consiste en un conjunto de secuencias de DNA (sondas) inmovilizado sobre una superficie sólida y organizado en cuadrículas microscópicas mediante un sistema de robótica (Alba et al, 2004). Los datos de expresión génica se obtienen mediante la hibridación de la micromatriz con una población de RNA o cDNA, marcado con fluorescencia, que procede de la muestra problema, de modo que se obtiene una señal de fluorescencia para cada sonda, cuya intensidad representa el nivel de expresión de dicho gen en la muestra problema (Busch y Lohmann, 2007). Actualmente, se pueden distinguir tres tipos de micromatrices de DNA según la naturaleza de las sondas y las técnicas de hibridación utilizadas: 1) productos de PCR y 2) oligonucleótidos largos (50-70 bases), ambos hibridados con dos poblaciones de cDNA, marcadas diferencialmente, que derivan de dos condiciones experimentales (por ejemplo, muestra problema y muestra control) y 3) oligonucleótidos cortos (25 bases) sintetizados *in situ* e hibridados con una única población de cDNA (Busch y Lohmann, 2007). Desde el primer trabajo de análisis de la expresión génica mediante micromatrices (Schena et al, 1995), las basadas en productos de PCR han sido muy utilizadas debido a su flexibilidad y a su razonable coste de fabricación (Busch y Lohmann, 2007). Estas micromatrices se generan a partir de genotecas de cDNA por lo que dependen de la disponibilidad de colecciones de clones. Sin embargo, el hecho de que en la última década se hayan creado numerosas colecciones de clones de cDNA en plantas ha permitido el desarrollo de este tipo de micromatrices (Alba et al, 2004), sobre todo en especies de interés agronómico como arroz (Rabbani et al, 2003; Lian et al, 2006), maíz (Fernandes et al, 2002), trigo (Clarke y Rahman, 2005), soja (Vodkin et al, 2004; Ainsworth et al, 2006), patata (Ducreux et al, 2005; Stupar et al, 2007), tomate (Alba et al, 2005) y fresa (Aharoni et al, 2004b). Sin embargo, una limitación importante es que las diferentes sondas de la micromatriz presentan, con frecuencia, secuencias de DNA similares, lo que conlleva problemas de hibridación cruzada (Alba et al, 2004). Recientemente, el programa CATMA de Arabidopsis (Complete Arabidopsis Transcriptome Microarray) ha generado un conjunto estandarizado de amplicones diseñado para minimizar la similitud con cualquier otra secuencia y, por tanto, el problema de la hibridación cruzada. La micromatriz CATMA permite analizar la expresión de 18,981 genes mediante sondas específicas de 150-500 pb de longitud (Allemeersch et al, 2005). Por otro lado, las micromatrices de oligonucleótidos largos contienen un conjunto

estandarizado de sondas de unas 60 bases diseñadas para representar genes únicos, minimizando el problema de la hibridación cruzada. Las principales limitaciones de estas micromatrices son la complejidad del proceso de diseño de los oligonucleótidos y el coste derivado de la síntesis de los mismos (Alba et al, 2004). La micromatriz de Agilent para *Arabidopsis*, que permite analizar la expresión de más de 40,000 unidades de transcripción, constituye un ejemplo de este tipo de micromatrices (<http://www.chem.agilent.com>). Por otro lado, algunas empresas, como Operon, producen conjuntos de oligonucleótidos para diferentes especies de plantas como *Arabidopsis*, arroz, maíz, vid o tomate (<http://www.operon.com>). Finalmente, las micromatrices de oligonucleótidos cortos sintetizados *in situ*, contienen conjuntos de oligonucleótidos de 25 bases de longitud sintetizados mediante fotolitografía. Actualmente, Affymetrix (<http://www.affymetrix.com>), con la patente de esta tecnología, produce micromatrices para un gran número de plantas como *Arabidopsis*, arroz, cebada, maíz, trigo, vid, tomate, *Medicago*, *Populus* y *Citrus*, entre otros (revisado por Rensink y Buell, 2005). El análisis de la expresión génica mediante micromatrices en plantas de interés agronómico permite identificar conjuntos de genes cuya expresión está correlacionada con el rasgo agronómico de interés estudiado.

2.3. El Proyecto de Genómica Funcional de Cítricos (CFGP)

El Proyecto de Genómica Funcional de Cítricos (CFGP; <http://bioinfo.ibmcp.upv.es>) surge fruto de la necesidad de dotar a la Citricultura actual de nuevas herramientas biotecnológicas que, por un lado, permitan profundizar en la comprensión de los mecanismos moleculares subyacentes a la biología de los cítricos, en particular de los relacionados con rasgos agronómicos de interés, y que, por otro lado, constituyan un sólido punto de apoyo en el desarrollo de futuros programas de mejora para la obtención de nuevas variedades. El CFGP se inicia en el año 2001 como una colaboración trilateral entre el Instituto de Biología Molecular y Celular de Plantas (IBMCP), el Instituto Valenciano de Investigaciones Agrarias (IVIA) y el Instituto de Agroquímica y Tecnología Alimentaria (IATA). La presente Tesis Doctoral se integra en el CFGP desde los orígenes del mismo.

En la actualidad existen diversas iniciativas de genómica funcional en cítricos, destacando los esfuerzos de secuenciación del genoma de cítricos (Roose et al, 2007; Chen y Gmitter, 2008; Terol y Talón, 2008); la secuenciación masiva de ESTs de cítricos (Machado et al, 2006; 2007); el desarrollo de la base de datos HarvEST (<http://harvest.ucr.edu/>), que recoge ESTs de diversas especies de interés agronómico, entre ellas cítricos (Close et al, 2007) y el de la micromatriz de

oligonucleótidos cortos de Affymetrix (GeneChip® Citrus Genome Array), que representa 33,879 transcritos de cítricos (Close et al, 2006). Sin embargo, cuando se inició el CFGP la base de datos GenBank del NCBI contaba con apenas 2,000 secuencias de cítricos: 1,300 ESTs de frutos en desarrollo de *Citrus unshiu*, 300 secuencias de flavedo y albedo de frutos de pomelo y 200 secuencias de *Citrus sinensis*. Esto implica que el CFGP fue la primera aproximación genómica a gran escala iniciada en cítricos. El objetivo general del CFGP fue el desarrollo de herramientas genómicas para el estudio de procesos biológicos relevantes para la citricultura, desde la biología del desarrollo (desarrollo vegetativo, floración, cuajado, desarrollo y maduración del fruto, calidad del fruto, senescencia y abscisión), a la respuesta al estrés de origen tanto biótico (infección por virus y hongos) como abiótico (salinidad, sequía y deficiencia férrica), y a los procesos de post-cosecha (respuesta al frío y al ataque por hongos). Este objetivo general se concretó, inicialmente, con la construcción de 25 genotecas de cDNA que cubren un amplio rango de tejidos, estadios de desarrollo y condiciones de estrés biótico y abiótico, así como de una genoteca genómica de *Citrus clementina*. A partir de los clones generados se aislaron un total de 22,635 ESTs, que representaban 11,836 posibles transcritos únicos o unigenes de cítricos. Como un primer paso en la caracterización del transcriptoma de cítricos, 12,672 clones de cDNA, que representaban 6,875 posibles unigenes, se utilizaron para generar una micromatriz de cDNA de cítricos (Forment et al, 2005). Este trabajo fue pionero en genómica funcional de cítricos puesto que abordó, por primera vez, la secuenciación masiva de ESTs y la generación de una micromatriz de cDNA en cítricos. Posteriormente, se hicieron públicas otras colecciones de ESTs de cítricos (Fujii et al, 2006; Machado et al, 2006, 2007) y se puso al servicio de la comunidad científica la micromatriz de oligonucleótidos cortos de cítricos (Close et al, 2006). La micromatriz de cDNA generada por el CFGP (Forment et al, 2005) permitió el análisis de la expresión génica durante el desarrollo y la maduración del fruto en cítricos (Cercós et al, 2006) así como en relación a la respuesta a la infección del virus de la tristeza de los cítricos (Gandía et al, 2007). Recientemente, se ha generado una genoteca de cDNA normalizada de longitud completa y 9 genotecas de cDNA estándar a partir de tejidos y tratamientos de variedades y patrones seleccionados en relación a la calidad del fruto, la resistencia a la abscisión y la tolerancia a la salinidad. A partir de estas últimas genotecas se han aislado y anotado funcionalmente 52,626 ESTs de cítricos, que representan 15,664 unigenes. El análisis funcional de los unigenes de cítricos se ha centrado en rutas metabólicas tanto del metabolismo primario como secundario, con especial detalle en la biosíntesis de lignina en cítricos y en distintos aspectos de genómica comparada entre cítricos y *Arabidopsis* como la identificación

de posibles ortólogos y la caracterización de familias génicas (Terol et al, 2007). En la actualidad el CFGP dispone de un total de 52 genotecas de cDNA (30 genotecas estándar, 6 genotecas de sustracción y 16 genotecas de longitud completa) generadas a partir de un amplio rango de condiciones (tejidos, estadios de desarrollo, condiciones de estrés biótico y abiótico) con el objeto de que la colección de ESTs derivada de estas genotecas maximice la representación del transcriptoma de los cítricos. En estos momentos, la colección de ESTs del CFGP consta de 85,965 ESTs agrupadas en 27,551 unigenes (11,844 *contigs*, o conjuntos de secuencias formados por dos o más ESTs y 15,707 *singletons* o secuencias únicas) que se gestionan en la base de datos EST2uni (Forment et al, 2008) (<http://bioinf.comav.upv.es/est2uni>), herramienta bioinformática desarrollada en el CFGP, que permite el pre-procesado, el ensamblaje y la anotación de las ESTs. Los *contigs* formados por un elevado número de ESTs, aunque aumentan la redundancia de una colección de ESTs, son de gran utilidad para la identificación de posibles marcadores moleculares, en particular SSRs y SNPs. El análisis de la colección de ESTs de cítricos del CFGP, mediante EST2uni, ha permitido identificar 1,533 posibles SSRs y 6,139 posibles SNPs en cítricos (<http://bioinf.comav.upv.es/est2uni>). La identificación de este conjunto de posibles marcadores moleculares en cítricos constituye una herramienta fundamental a partir de la cual se pueden desarrollar posteriores estudios que aborden los problemas concretos de la taxonomía, las relaciones filogenéticas y la variabilidad genética del género *Citrus*, cuestiones críticas en la caracterización del germoplasma, en el control de la erosión genética, en el registro de nuevas variedades y en el establecimiento de nuevos programas de mejora genética.

3. APROXIMACIÓN GENÓMICO-FUNCIONAL A DIVERSOS ASPECTOS DEL DESARROLLO VEGETATIVO EN CÍTRICOS

3.1. Biosíntesis de isoprenoides y flavonoides

3.1.1. Biosíntesis de isoprenoides

Los isoprenoides (o terpenoides) constituyen el grupo funcional y estructuralmente más variado de metabolitos en las plantas (Croteau et al, 2000). Los esteroides (isoprenoides con un esqueleto C30) son componentes esenciales de las membranas; carotenoides (C40) y clorofilas (que contienen una cadena lateral C20) actúan como pigmentos fotosintéticos; plastoquinona y ubiquinona (que contienen largas cadenas laterales de isoprenoides) participan en el transporte electrónico; el

isopreno (C5) es un producto volátil que liberan los tejidos fotosintéticamente activos y que participa en la formación del ozono troposférico; giberelinas (C20), brasinosteroides (C30) y ácido abscísico (C15) son fitohormonas derivadas de isoprenoides y algunos monoterpenos (C10), sesquiterpenos (C15) y diterpenos (C20) están implicados en procesos de defensa (Lange y Ghassemian, 2003). La biosíntesis de los isoprenoides comienza con la formación de isopentenil difosfato (IPP, C5) y dimetilalil difosfato (DMAPP, C5). En plantas, la síntesis de estas unidades C5 se produce mediante dos rutas localizadas en diferentes compartimentos celulares: la ruta citosólica del mevalonato (MVA) y la ruta plastidial del metileritritol fosfato (MEP) (Fig. 1). El IPP derivado de la ruta del MVA puede ser transportado a la mitocondria para producir isoprenoides mitocondriales. Los isoprenoides derivados de la ruta del MVA son ubiquinonas, esteroides (incluidos los brasinosteroides), sesquiterpenos, politerpenos y dolicoles. A partir de la ruta plastidial del MEP se sintetizan otros isoprenoides como plastoquinonas, filoquinonas, clorofilas, tocoferoles, giberelinas, carotenoides, ácido abscísico y monoterpenos (Fig. 1).

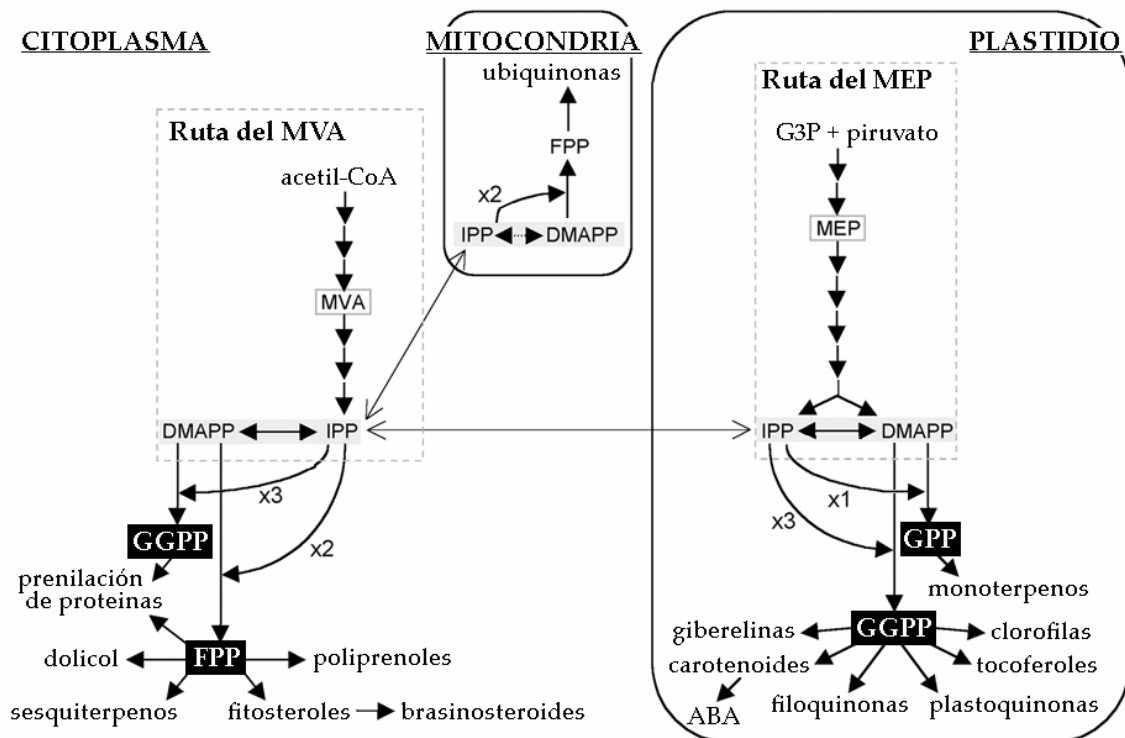


Fig. 1. Biosíntesis de isoprenoides en plantas. La formación de IPP y DMAPP se produce mediante la ruta citosólica del MVA y la ruta plastidial del MEP. Además, se señalan los metabolitos derivados de la ruta del MVA, cuya síntesis se produce en el citosol, y los derivados del MEP que se sintetiza, al menos en parte, en el plastidio.

Los cítricos constituyen una fuente importante de metabolitos derivados de isoprenoides, como carotenoides, relacionados con el color del fruto y con propiedades nutricionales como precursores de vitamina A y antioxidantes; limonoides, implicados en el sabor del fruto y con valor como anticancerígenos; y mono- y sesquiterpenos, relacionados con el sabor de los frutos y con el aroma, de frutos y de hojas. Los carotenoides se acumulan en el fruto durante su maduración, especialmente en el flavedo (o parte más externa de la corteza) que es una de las fuentes más ricas y complejas de carotenoides en plantas (revisado por Rodrigo y Zacarías, 2006). En particular, el fruto de los cítricos constituye un material de gran interés para el estudio de la acumulación de xantofilas (carotenoides oxigenados) pues esta acumulación masiva no sucede en otras especies como tomate o *Arabidopsis*. La relación entre la acumulación de carotenoides y la expresión de los genes implicados en su biosíntesis ha sido estudiada con detalle durante la maduración del fruto en tres variedades de cítricos con diferente composición en carotenoides por Kato et al (2004). Los resultados obtenidos demostraron que la acumulación de carotenoides durante la maduración de los frutos es un proceso específicamente regulado por la expresión coordinada de los diferentes genes de biosíntesis de carotenoides (Kato et al, 2004). Los aceites esenciales son las fracciones líquidas volátiles que contienen las sustancias responsables del aroma de las plantas. En los cítricos constituyen uno de los principales productos y se obtienen tanto de la corteza del fruto, mediante presión en frío, como de las hojas, por destilación en corriente de vapor. Los aceites esenciales, además de ser responsables en gran medida del aroma y del sabor percibido por los consumidores, son valiosos desde el punto de vista industrial para la obtención de fragancias, fármacos y productos agroquímicos (revisado por Rodrigo y Zacarías, 2006). Los constituyentes volátiles de los aceites esenciales de cítricos son, fundamentalmente, mono- y sesquiterpenos. La composición de los aceites esenciales varía mucho entre especies, variedades e, incluso para un mismo cultivar, entre la piel del fruto y las hojas. Así, la fracción volátil de los aceites esenciales de la piel de frutos de *Citrus clementina* (especie a partir de la que se han generado la mayor parte de ESTs del CFGP) está compuesta por limoneno (90%), sabineno (<4%) y mirceno (<2%), mientras que los aceites esenciales más importantes de las hojas son sabineno (45%) y linalool (20%) (Lota et al, 2001). Esta variabilidad parece deberse principalmente a la expresión de diferentes terpeno sintasas, enzimas clave en la biosíntesis de mono-, sesqui- y diterpenos. En los últimos años se han aislado y caracterizado algunos genes que codifican terpeno sintasas del género *Citrus*, tanto monoterpeno sintasas, como limoneno sintasa, β -pineno y γ -terpineno sintasa de *Citrus limon* (Lücker et al, 2002) y de *Citrus unshiu* (Shimada et al, 2004), como

sesquiterpeno sintasas, como β -farneseno sintasa de *Citrus junus* (Maruyama et al, 2001) y valenceno sintasa de *Citrus sinensis* (Sharon-Asa et al, 2003). Sin embargo, dado que los aceites esenciales de los cítricos presentan una gran variedad de mono- y sesquiterpenos, cabe esperar que la síntesis de estos compuestos esté catalizada por otras terpeno sintasas todavía no identificadas en cítricos.

A pesar de los trabajos realizados en la caracterización de la biosíntesis de carotenoides (Kato et al, 2004) y de diversas terpeno sintasas en cítricos (Lücker et al, 2002; Maruyama et al, 2001; Sharon-Asa et al, 2003; Shimada et al, 2004), la mayor parte de las rutas de biosíntesis de metabolitos derivados de isoprenoides todavía no se han estudiado en este género. La colección de ESTs de cítricos generada por el CFGP constituye una herramienta de gran interés para la identificación de los genes implicados en las diferentes rutas del metabolismo de isoprenoides, incluyendo la búsqueda de genes que codifican otras terpeno sintasas en cítricos.

3.1.2. Biosíntesis de flavonoides

Los flavonoides constituyen un grupo importante de metabolitos secundarios en las plantas que están implicados en diversas funciones como la protección frente a la radiación UV, la pigmentación de las flores para atraer a los polinizadores, la germinación del polen, la resistencia al ataque por microorganismos y la defensa frente a insectos (Croteau et al, 2000). Además, los flavonoides poseen un amplio rango de actividades bioquímicas y farmacológicas en mamíferos y otros sistemas biológicos, por lo que constituyen un grupo de potenciales agentes terapéuticos (Middleton et al, 2000). Los cítricos acumulan grandes cantidades de flavonoides, tanto en los frutos como en las hojas, especialmente flavanonas que son parcialmente responsables del aroma y del sabor del fruto. Además, los cítricos son un componente importante de la dieta por sus propiedades antioxidantes, antiinflamatorias, antialérgicas y supresoras de mutagénesis, de modo que constituyen una fuente importante de flavonoides para las industrias agroalimentarias y farmacéuticas (Manthey et al, 2001).

La biosíntesis de flavonoides ha sido muy estudiada en plantas y los genes implicados en esta ruta han sido aislados y caracterizados en diversas especies vegetales (revisado por Winkel-Shirley, 2001; Ververidis et al, 2007). La biosíntesis de flavonoides comienza con la formación de la flavanona naringenina en dos pasos consecutivos catalizados por las enzimas chalcona sintasa (CHS) y chalcona isomerasa (CHI) (Fig. 2). Posteriormente, la naringenina puede ser hidroxilada para formar el flavanol dihidroxikaempferol, reacción catalizada por la enzima flavanona

3-hidroxilasa (F3H). En cítricos se han aislado los genes que codifican estos tres pasos iniciales de la biosíntesis de flavonoides (Moriguchi et al, 1999; Moriguchi et al, 2001). El análisis de la expresión génica de estos genes mostró una mayor expresión en las etapas iniciales del desarrollo del fruto, correlacionada con la acumulación de flavonoides (Moriguchi et al, 2001). La conversión de flavanol (como el dihidrokaempferol) en flavonol (como el kaempferol) está catalizada por flavonol sintasa (FLS). En cítricos, se ha aislado el gen que codifica FLS (Moriguchi et al, 2002) y se ha caracterizado su actividad enzimática (Wellmann et al, 2002). Por otro lado, la flavanona naringenina puede sufrir dos glucosilaciones sucesivas: la primera catalizada por 7-O-glucosiltransferasa y la segunda por una rhamnosiltransferasa (Fig. 2). En cítricos, se ha aislado y caracterizado funcionalmente el gen *Cm1,2RhaT*, que codifica una 1,2 rhamnosiltransferasa (Frydman et al, 2004).

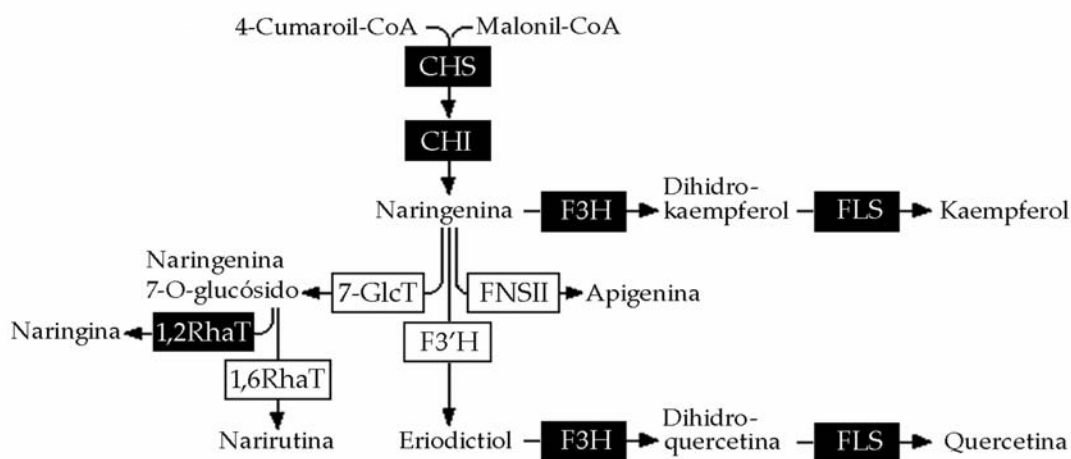


Fig. 2. Biosíntesis de flavonoides en cítricos. Las cajas negras señalan los pasos para los que se han aislado los correspondientes genes en cítricos y las cajas blancas aquellos pasos que todavía continúan sin caracterización molecular en cítricos.

En los últimos años se han caracterizado diversos genes implicados en la biosíntesis de flavonoides en cítricos. Sin embargo, los genes que codifican algunos pasos fundamentales, como la formación de flavonas (ej. apigenina) todavía no han sido aislados. La identificación de estos genes a partir de la colección de ESTs del CFGP debe aportar un mayor conocimiento de la biosíntesis de flavonoides en cítricos y, adicionalmente, supone una herramienta básica con la que abordar la fisiología, la bioquímica y la regulación genética del metabolismo de flavonoides en este género.

3.2. Mapas de expresión génica en plantas

Todos los organismos eucariotas complejos, incluidas las plantas superiores, están formados por distintos tipos celulares organizados en tejidos y órganos. El conjunto de tejidos y órganos de un organismo concreto se genera a lo largo de su ciclo de vida a través de un proceso, regulado temporal y espacialmente, de expresión selectiva de fracciones específicas del mismo genoma en diferentes células (Meyerowitz 2002). Uno de los objetivos de la biología del desarrollo es definir el subconjunto de genes expresado y su abundancia relativa en cada órgano y tipo de tejido. Las plantas superiores siguen un proceso de desarrollo relativamente simple, con menos de 25 tejidos y tipos celulares principales, por lo que suponen un buen modelo para definir los patrones de expresión génica específicos de cada órgano y tipo de tejido (Ma et al, 2005). El análisis de la expresión génica en multitud de tejidos y estadios de desarrollo puede conducir a la creación de un atlas de expresión génica con el que describir los patrones de expresión de los genes de un organismo. La generación de dicho atlas transcripcional constituye una herramienta de gran utilidad pues permite conocer tanto el patrón de expresión de un grupo de genes de interés como el conjunto de genes que se ajusta a un patrón de expresión definido. En este sentido, resulta posible inferir la función de un grupo de genes bajo la suposición de que los genes con un patrón de expresión similar están implicados en procesos similares (Borevitz y Ecker, 2004). El desarrollo de la tecnología de micromatrices de DNA, que permiten medir la expresión de miles de genes simultáneamente, proporciona un método eficaz para analizar el conjunto de genes que se expresa en un órgano o en un tipo de tejido. En los últimos años, diversos trabajos han abordado el análisis de los patrones de expresión génica en diferentes órganos mediante micromatrices, especialmente en *Arabidopsis*. Ma et al (2005) analizaron la expresión génica en 18 órganos o tipos de tejidos que representaban diferentes aspectos del ciclo de vida de *Arabidopsis*. Sin embargo, el conjunto de datos de expresión más completo es el de Schmid et al (2005) que forma parte del consorcio AtGenExpress (<http://www.arabidopsis.org/info/expression/ATGenExpress.jsp>). Este trabajo analizó la expresión de 22,000 genes de *Arabidopsis*, utilizando la micromatriz ATH1 (Redman et al, 2004), en 79 muestras que cubren diversos estadios de desarrollo, desde la embriogénesis hasta la senescencia, y diversos órganos. Los resultados obtenidos representan el estado transcripcional subyacente al programa de desarrollo de la planta y revelan ciertos aspectos de interés. Por ejemplo, los perfiles de expresión génica muestran que, en general, la identidad de órgano domina sobre el estado de desarrollo o las condiciones ambientales, que constituyen componentes menores.

Además, la anotación funcional mediante ontologías génicas muestra que los órganos que son morfológicamente similares contienen ontologías génicas similares (Schmid et al, 2005). Contrariamente a otros trabajos previos (Czechowski et al, 2004), los resultados obtenidos por Schmid et al (2005) muestran que la expresión de los genes que codifican factores de transcripción y componentes de señalización es similar al nivel de expresión promedio del genoma completo.

Recientemente, diversos trabajos han abordado la obtención de mapas transcripcionales a un nivel de resolución superior mediante el análisis de la expresión génica en tipos celulares específicos (revisado por Galbraith y Birnbaum, 2006). Los trabajos del grupo de Philip Benfey (Birnbaum et al, 2003; Nawy et al, 2005; Brady et al, 2007) han aportado un exhaustivo mapa transcripcional para casi todos los tipos celulares de la raíz de *Arabidopsis*. Los distintos tipos celulares se han aislado obteniendo diferentes líneas transgénicas en las que cada población celular se encuentra marcada con fluorescencia. Se aíslan los protoplastos de la raíz de estas líneas y las células marcadas con fluorescencia se separan utilizando *fluorescence-activated cell sorter* (FACS) (Birnbaum et al, 2005). Por otro lado, la microdissección por láser permite separar células individuales o grupos celulares a partir de tejidos intactos y esta técnica se ha utilizado para analizar la expresión génica durante el proceso de embriogénesis en *Arabidopsis* (Casson et al, 2005; Spencer et al, 2007).

La elaboración de mapas de expresión génica de diversos órganos, tejidos o células en plantas define e identifica el estado de dicho órgano, tejido o célula a nivel molecular lo que permite profundizar en la comprensión de sus funciones. Adicionalmente, las regiones promotoras de los genes diferencialmente expresados en cada órgano, tejido o célula constituyen una herramienta biotecnológica de interés para dirigir la expresión génica a los mismos.

3.3. Control del desarrollo vegetativo: giberelinas

3.3.1. Metabolismo de giberelinas

Las giberelinas (GAs) constituyen un grupo de hormonas vegetales que controlan diversos aspectos del desarrollo como la germinación, la expansión de las hojas, la elongación del tallo y de la raíz, el tiempo de floración y el desarrollo de la flor y del fruto (Sponsel y Hedden, 2004; Swain y Singh, 2005). La aplicación de GAs y de inhibidores de su biosíntesis es una práctica ampliamente utilizada en la agricultura, por ejemplo para estimular el crecimiento de los frutos en las variedades de uva sin semillas, para retrasar la senescencia del fruto en naranjas y limones, para aumentar

el cuajado en mandarinas, manzanas y peras, para aumentar la elongación del tallo en la caña de azúcar o para disminuir el crecimiento de diversas especies como algodón y manzano (Hedden y Phillips, 2000a). El metabolismo de GAs ha sido estudiado en profundidad y está bien caracterizado (Hedden y Phillips, 2000b; Olszewski et al, 2002; Sponsel y Hedden, 2004). Las GAs se sintetizan a partir del precursor común geranylgeranyl difosfato mediante la acción de tres grupos de enzimas (Fig. 3): 1) dos diterpeno sintasas, copalil difosfato sintasa (CPS) y *ent*-kaureno sintasa (KS), conducen a la formación de *ent*-kaureno; 2) dos monooxigenasas dependientes de citocromo P450, *ent*-kaureno oxidasa (KO) y ácido *ent*-kaurenoico oxidasa (KAO), catalizan la conversión de *ent*-kaureno en GA₁₂ y GA₅₃; y 3) dioxigenasas dependientes de 2-oxoglutarato, metabolizan GA₁₂ y GA₅₃ en diferentes GAs.

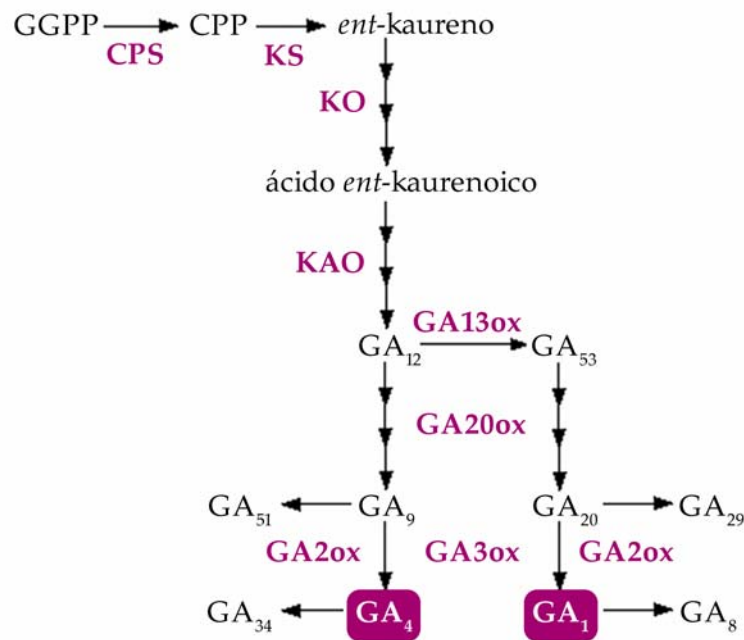


Fig. 3. Esquema del metabolismo de GAs. GGPP, geranylgeranyl difosfato; CPP, copalil difosfato; CPS, copalil difosfato sintasa; KS, *ent*-kaureno sintasa; KO, *ent*-kaureno oxidasa; KAO, ácido *ent*-kaurenoico oxidasa; GA13ox, GA 13-oxidasa; GA20ox, GA 20-oxidasa; GA3ox, GA 3-oxidasa; GA2ox, GA 2-oxidasa. Las GAs activas están marcadas en cajas.

El grupo de las dioxigenasas dependientes de 2-oxoglutarato incluye las enzimas GA 20-oxidasa (GA20ox) y GA 3-oxidasa (GA3ox), que actúan consecutivamente para producir las GAs bioactivas (GA₁ y GA₄) a través de dos rutas paralelas: la ruta de la hidroxilación temprana en C-13 (que conduce a la formación de GA₁) y la ruta de la no hidroxilación en C-13 (mediante la que se sintetiza GA₄) (Fig. 3). Una tercera clase de dioxigenasas, GA 2-oxidasa (GA2ox), inactivan tanto las GAs bioactivas (GA₁ y GA₄) como sus precursores (GA₂₀, GA₉, GA₁₂ y GA₅₃). En cítricos, la

principal ruta de síntesis de GAs es la ruta de la hidroxilación temprana en C-13, tanto en los tejidos vegetativos (Poling y Maier, 1988; Turnbull, 1989) como en los reproductivos (Goto et al, 1989; Talón et al, 1990; Talón et al, 1992; Ben-Cheikh et al, 1997b) de modo que GA₁ es la GA bioactiva mayoritaria en cítricos. La mayoría de los genes que codifican las enzimas del metabolismo de giberelinas han sido aislados en diversas especies (Hedden y Kamiya, 1997; Hedden y Phillips, 2000b; Olszewski et al, 2002; Sponsel y Hedden, 2004). Las dioxigenasas constituyen pequeñas familias multigénicas cuyos miembros muestran distintos patrones de expresión regulados por diversos factores, tanto endógenos como ambientales (Kamiya y García-Martínez, 1999; Yamaguchi y Kamiya, 2000). La expresión de los genes que codifican estas dioxigenasas está sujeta a una regulación negativa (GA20ox y GA3ox) o positiva (GA2ox) por las GAs activas (Hedden y Phillips, 2000b).

3.3.2. Modificación de la arquitectura de las plantas mediante cambios en los niveles de GA 20-oxidasas

El importante incremento de la producción agrícola que se produjo durante la llamada 'Revolución Verde' fue posible por la introducción de variedades semi-enanas de alto rendimiento de trigo y arroz, en combinación con la aplicación de grandes cantidades de fertilizantes nitrogenados y pesticidas (Hedden, 2003). La fertilización nitrogenada es esencial para el aumento de la producción de grano pero promueve la elongación del tallo y el consecuente encamado de los cereales. Sin embargo, las variedades semi-enanas de la 'Revolución Verde' mostraron un mayor rendimiento y menores problemas de encamado bajo condiciones de elevada fertilización (Sakamoto y Matsuoka, 2004). La identificación de las mutaciones responsables del hábito de crecimiento semi-enano de las variedades de trigo y arroz de la 'Revolución Verde' ha mostrado que los genes afectados están implicados en la transducción de señal (proteínas DELLA) y en la ruta de biosíntesis de GAs (Hedden, 2003). En particular, la mutación *sd1*, responsable del carácter semi-enano de los cultivares de arroz utilizados durante la 'Revolución Verde' (como IR8), afecta a un gen que codifica una GA 20-oxidasa que se expresa en el tallo (Monna et al, 2002; Sasaki et al, 2002; Spielmeier et al, 2002). Estos trabajos ponen de manifiesto el importante papel de las GAs en la regulación de procesos del desarrollo críticos para la agricultura.

En numerosas especies vegetales se ha demostrado que la sobre-expresión de los genes que codifican GA20ox implica una alteración de la concentración de GAs bioactivas, lo que indica que estos genes tienen un papel muy importante en la regulación de los últimos pasos de la ruta de biosíntesis de GAs. En este sentido, la

generación de plantas transgénicas que sobre-expresan *GA20ox* ha sido descrita en diversas especies vegetales como *Arabidopsis* (Huang et al, 1998; Coles et al, 1999), álamo (Eriksson et al, 2000), tabaco (Vidal et al, 2001), manzano (Bulley et al, 2005) y, más recientemente, en el patrón de cítricos citrange Carrizo (Fagoaga et al, 2007). En *Arabidopsis*, la sobre-expresión de *GA20ox* condujo a la formación de hipocotilos más largos, elongación del tallo y adelanto de la floración, asociado con un aumento del contenido de GAs bioactivas respecto al control (Huang et al, 1998; Coles et al, 1999). Por el contrario, la expresión en antisentido de *GA20ox* dio como resultado plantas de *Arabidopsis* de altura y contenido en GA_4 reducidos (Coles et al, 1999). Estos resultados demostraron que los niveles de GAs y, consecuentemente, el crecimiento y desarrollo de las plantas, podía ser modulado mediante la modificación de la expresión de los genes que codifican *GA20ox*, y abrieron un camino para la manipulación biotecnológica de la arquitectura de las plantas mediante la alteración del metabolismo de GAs. En el caso del álamo, Eriksson et al (2000) describieron que la sobre-expresión del gen de *Arabidopsis AtGA20ox1* no sólo producía niveles elevados de GAs activas y un mayor crecimiento de las plantas, sino que, además, estas plantas tenían más biomasa y un mayor número y longitud de fibras xilemáticas, caracteres deseables en la producción de papel. En manzano, se ha aislado un fragmento de un gen que codifica *GA20ox* (*MpGA20ox1*) y se ha sobre-expresado en sentido y en antisentido en la variedad Greensleeves (Bulley et al, 2005). En este caso, la co-supresión del gen endógeno produjo plantas enanas, con un número y longitud de entrenudos reducido y un área de la hoja reducida, como resultado de un menor contenido en GA_1 (Bulley et al, 2005). En cítricos se ha aislado un gen que codifica *GA20ox* (*CcGA20ox1*) a partir del portainjertos citrange Carrizo (Vidal et al, 2003). Recientemente, se han generado plantas transgénicas de citrange Carrizo que sobre-expresan el gen *CcGA20ox1* en sentido y en antisentido (Fagoaga et al, 2007). Las líneas de sobre-expresión del gen *CcGA20ox1* en sentido mostraron una altura mayor debido a entrenudos más largos, mayor longitud de las espinas (órganos típicos de los cítricos en el estadio juvenil), menor área foliar y menor diámetro de los tallos, debido a una disminución en el área del xilema, caracteres asociados con un aumento de 2-3 veces del contenido de GA_1 en brotes jóvenes en desarrollo (Fagoaga et al, 2007). Por otro lado, la sobre-expresión en antisentido del gen *CcGA20ox1* dio lugar a plantas con una altura ligeramente inferior, menor longitud de las espinas y un porte arbustivo, más ramificado, relacionado con un menor contenido en GA_1 (Fagoaga et al, 2007). Todos estos trabajos han demostrado que la actividad GA 20-oxidasa es determinante para la producción de GAs en las plantas y que, por tanto, constituye una buena diana para la manipulación genética de la biosíntesis de GAs. En particular, en

frutales como manzano y cítricos, la posibilidad de obtener portainjertos con un efecto enanizante, mediante la reducción del contenido de GAs bioactivas, aportaría considerables beneficios al permitir una mayor densidad de plantación, un manejo del cultivo más sencillo y al facilitar la recogida mecánica de los frutos, reduciendo con todo ello los costes de producción.

La percepción de GAs y la ruta de transducción de señal convierten la señal de GAs en cambios en la expresión génica y en la morfología de la planta (Sun y Gubler, 2004; Schwechheimer, 2008). Por tanto, cabe esperar que el fenotipo observado en las plantas transgénicas con el metabolismo de GAs modificado esté asociado con cambios importantes en la expresión génica. La tecnología de micromatrices permite examinar los cambios en los niveles de expresión de miles de genes en un solo experimento. Sin embargo, existe poco trabajo de este tipo en la literatura utilizando transgénicas con el metabolismo de GAs modificado. En este sentido, se ha descrito el análisis del transcriptoma del xilema en desarrollo de plantas transgénicas de álamo (*Populus tremula* L. x *tremuloides* Michx.) que sobre-expresan el gen *AtGA20ox1* de *Arabidopsis* (Israelsson et al, 2003). En este caso, se observa una inducción de los genes implicados en la formación y extensión de la pared celular (como pectin metil esterasa, pectato liasa, celulasa, endoglucanasas y expansinas) asociado con el mayor número y longitud de fibras xilemáticas observado en estas plantas transgénicas (Eriksson et al, 2000). Sin embargo, este estudio transcripcional examinó sólo un grupo de genes pues utilizó una micromatriz de DNA que contiene 2,995 ESTs procedentes de una genoteca de cDNA de floema, cambium y xilema en desarrollo de *Populus tremula* L. x *tremuloides* Michx. (Sterky et al, 1998).

3.3.3. Identificación de genes de respuesta temprana a GAs

La mejora del conocimiento del papel de las GAs en la regulación del crecimiento y desarrollo de la planta requiere la identificación de los genes de respuesta temprana a GAs, es decir, aquellos cuya expresión está alterada antes de que se produzca ningún cambio morfológico. La identificación de genes regulados por GAs se ha abordado con diversas estrategias, como la hibridación substractiva (Phillips y Huttlu, 1994) o *differential display* (van der Knaap et al, 1997). Recientemente, se han descrito algunos trabajos en los que se analizan los cambios producidos en la expresión génica tras la aplicación de GAs en arroz y *Arabidopsis* utilizando micromatrices. En arroz, la respuesta transcripcional a GAs se ha estudiado en callos (Yazaki et al, 2004) y en plántulas (Yang et al, 2004). El trabajo de Yazaki et al (2004) identificó genes regulados por GAs tras el cultivo de callos de arroz

durante 3 días en GAs (Yazaki et al, 2004). Entre los genes inducidos se encontraron genes implicados en la división y diferenciación celular y en defensa (Tabla Suplementaria II). Yang et al (2004) estudiaron los cambios de expresión génica en plántulas de arroz tras la aplicación durante 24 horas de GA₃. Los resultados obtenidos por ambos grupos no coinciden totalmente. Sin embargo, la expresión de los genes que codifican xiloglucano endotransglicosilasa (XET) apareció inducida por GAs en los dos trabajos, sugiriendo que las GAs tienen un papel en la regulación de la elongación y la organización de la pared celular (Tabla Suplementaria II). En *Arabidopsis*, se ha examinado el efecto de las GAs sobre la expresión génica en semillas en germinación (Ogawa et al, 2003; Cao et al, 2006), en plántulas (Zentella et al, 2007) y en flores (Cao et al, 2006). Los experimentos realizados en *Arabidopsis* han aprovechado la ventaja de disponer de mutantes con defectos en la ruta de biosíntesis de GAs. En concreto, el mutante *ga1-3*, que muestra una deficiencia severa de GAs, ya que el gen *GA1* codifica una copalil difosfato sintasa (CPS) que cataliza el primer paso de la biosíntesis de GAs (Sun y Kamiya, 1994) (Fig. 3). Este mutante es extremadamente enano y presenta un retraso en la floración y esterilidad masculina (Koorneef y van der Veen, 1980; Wilson et al, 1992), defectos que se rescatan completamente tras la aplicación exógena de GA. El trabajo de Ogawa et al (2003) identificó 357 genes regulados por GAs durante la germinación (230 inducidos y 127 reprimidos), analizando semillas del mutante *ga1-3* tras el tratamiento con GA durante 3, 6 y 12 horas (Tabla Suplementaria II). Los genes inducidos por GAs incluyeron aquéllos implicados en el metabolismo de la pared celular (elongación celular) y en la división celular. Asimismo se encontraron modulados por GAs diversos genes relacionados con otras hormonas vegetales (Ogawa et al, 2003). Posteriormente, Cao et al (2006) compararon los patrones de expresión génica en semillas embebidas y flores en desarrollo del mutante *ga1-3* y del control, obteniendo 1,112 genes expresados diferencialmente en las semillas embebidas y 1,248 en las flores en desarrollo (Tabla Suplementaria II). Los resultados mostraron que las GAs regulan diferentes conjuntos de genes con funciones similares, o diferentes miembros de la misma familia génica, en el control de la germinación y del desarrollo floral. El análisis de ontologías de los genes expresados diferencialmente en cada tejido mostró la implicación de los mismos en los procesos de elongación celular, biogénesis y modificación de la pared celular, fosforilación y defosforilación de proteínas, señalización por ABA y auxinas y estrés oxidativo (Cao et al, 2006). Además, una proporción elevada (56%) de los genes regulados por GAs según Ogawa et al (2003), fueron también identificados por Cao et al (2006) (Tabla Suplementaria II). Finalmente, Zentella et al (2007) identificaron 67 genes de respuesta temprana a GAs en tejidos

vegetativos (tallos) de plántulas del mutante *ga1-3* de *Arabidopsis* de 8 días de edad en presencia o ausencia de un tratamiento de GAs (1h con 2 μ M GA₄). En este caso, el número de genes de respuesta de GAs identificado fue mucho menor que en los trabajos anteriores (Ogawa et al, 2003; Cao et al, 2006). Además, los cambios de expresión observados tras 1h de tratamiento con GAs son muy sutiles, inferiores a 2 veces (Zentella et al, 2007). Estas diferencias pueden deberse a las condiciones experimentales o a los tejidos estudiados, semillas en germinación (Ogawa et al, 2003; Cao et al, 2006) y flores en desarrollo (Cao et al, 2006) frente a tallos vegetativos (Zentella et al, 2007).



Objetivos

OBJETIVOS

El trabajo presentado en esta memoria de Tesis Doctoral se ha centrado en el desarrollo de nuevas herramientas de genómica funcional que, por un lado, permitan profundizar en la comprensión de los mecanismos moleculares subyacentes a la biología de los cítricos, en particular de los relacionados con el desarrollo vegetativo, y que, por otro lado, constituyan un sólido punto de apoyo en el diseño de futuros programas de mejora para la obtención de nuevas variedades de cítricos. Los objetivos específicos planteados en este trabajo son:

- I) Identificar genes implicados en el desarrollo vegetativo en cítricos mediante la generación y el análisis funcional de una colección de ESTs de tejidos vegetativos.
- II) Identificar genes relacionados con el metabolismo de isoprenoides y flavonoides en cítricos mediante el análisis funcional de la colección de ESTs generada por el Proyecto de Genómica Funcional de Cítricos (CFGP).
- III) Caracterizar la micromatriz de cDNA generada por el CFGP con el fin de valorar la fiabilidad de los datos obtenidos.
- IV) Elaborar mapas de expresión génica de los tejidos reproductivos y vegetativos mediante el análisis de micromatrices de cDNA con el fin de profundizar en la comprensión de las funciones de cada uno de los tejidos estudiados.
- V) Estudiar la regulación del desarrollo vegetativo mediada por GAs, por un lado, mediante el análisis del transcriptoma de plantas transgénicas del patrón de cítricos citrange Carrizo que contienen niveles elevados de GAs y, por otro, mediante el análisis de la expresión génica de brotes vegetativos de citrange Carrizo incubados con ácido giberélico.



*Materiales
y métodos*

1. Material vegetal y tratamientos hormonales

1.1. Material de campo

La genoteca de cDNA de tejidos vegetativos se realizó a partir de hojas y entrenudos de brotes en desarrollo de la brotación de otoño procedentes de tres árboles adultos de la variedad de mandarina Clementines (*Citrus clementina* Hort. ex Tan.) injertados sobre el patrón mandarino Cleopatra (*Citrus reshni* Hort. ex Tan.), del Banco de Germoplasma de cítricos del Instituto Valenciano de Investigaciones Agrarias (IVIA), localizado en Moncada (Valencia). Los árboles presentaban buen vigor y estaban libres de plagas. Los tratamientos hormonales consistieron en la pulverización directa, sobre brotes jóvenes de estos árboles de soluciones de GA₃ (50 ppm) y de paclobutrazol (500 ppm). El RNA procedente de brotes jóvenes tratados con GA₃, brotes jóvenes tratados con paclobutrazol y hojas y entrenudos procedentes de brotes jóvenes sin tratar se utilizó para construir una genoteca de cDNA.

En los análisis de expresión génica diferencial entre distintos tejidos de cítricos, los tejidos reproductivos estudiados se obtuvieron de estos mismos árboles. Se recogieron flores y frutos en los estadios fenológicos, según la escala BBCH (Agustí et al, 1995), de botón blanco (56), bola hueca (59) y antesis (61) y frutos con estilo (69) y sin estilo (71), respectivamente. El muestreo se realizó al azar en diferentes zonas de los tres árboles formando tres lotes o réplicas biológicas.

Para el análisis del efecto de GAs sobre el transcriptoma de cítricos, se utilizaron entrenudos en desarrollo procedentes de árboles transgénicos de citrange Carrizo (un híbrido de *Citrus sinensis* L. Osb. var. Washington Navel x *Poncirus trifoliata* L. Raf) que sobre-expresan el gen *CcGA20ox1*, tanto en sentido como en antisentido (Fagoaga et al, 2007). Las líneas seleccionadas para este trabajo fueron la línea sentido S23 y la línea antisentido A4, que muestran una expresión elevada del transgén *CcGA20ox1* y niveles modificados de GAs en tejidos vegetativos (Fagoaga et al, 2007). Estos árboles fueron producidos a partir de estaquillas enraizadas y cultivados en condiciones estándar en una parcela experimental del IVIA. Como controles se utilizaron árboles generados por el mismo procedimiento pero transformados únicamente con el vector vacío. La caracterización del transcriptoma de estas plantas se realizó utilizando entrenudos procedentes de brotes vegetativos en desarrollo de la brotación de otoño.

En todos los casos, incluyendo el material de invernadero, las muestras se congelaron en N₂ líquido y se conservaron a -80°C hasta la extracción de RNA.

1.2. Material de invernadero

En la caracterización técnica de la micromatriz de cDNA de cítricos del CFGP se utilizaron hojas en desarrollo procedentes de plantas de *Citrus clementina* de 2 años de edad cultivadas en los invernaderos del Instituto de Biología Molecular y Celular de Plantas (IBMCP), localizado en la Universidad Politécnica de Valencia. En los experimentos de análisis de expresión génica diferencial entre distintos tejidos de cítricos se utilizaron las cuatro últimas hojas apicales de brotes vegetativos. Se definieron 5 estadios de desarrollo del brote en función del tamaño de estas últimas hojas: 1) estadio 1, brotes con hojas de 1-2 cm, 2) estadio 2, brotes con hojas de 2-3 cm, 3) estadio 3, brotes con hojas de 3-4 cm, 4) estadio 4, brotes con hojas de 4-6 cm y 5) estadio 5, brotes con hojas de 6-8 cm.

El efecto de GAs a tiempos cortos se estudió utilizando brotes jóvenes de unos 10 cm de longitud, con 5-6 hojas en desarrollo, de plantas de citrange Carrizo procedentes de semilla de unos 3 meses de edad. Los brotes se cortaron y se incubaron durante 24 h en viales con 40 ml de medio de Murashige y Skoog con o sin 10 μ M GA₃. La incubación de los explantes se realizó a 24°C en condiciones de fotoperiodo de 16 h de luz y 8 h de oscuridad. Se realizaron cuatro réplicas biológicas, con 5 brotes por réplica, para cada uno de los tratamientos: tiempo 0 y 24 h sin y con GA₃. El análisis transcriptómico se realizó utilizando los entrenudos procedentes de los explantes de cada uno de los tratamientos descritos.

2. Extracción de RNA total y RNA poli(A⁺) de plantas

La extracción de RNA se llevó a cabo utilizando el protocolo general que se describe más adelante. En los experimentos con micromatrices, el RNA obtenido se purificó tal como se describe en el apartado 2.1.1 puesto que este paso adicional mejoró notablemente la eficiencia del marcaje de las muestras de RNA. En todos los casos, la pureza del RNA obtenido se evaluó mediante medidas de absorbancia a 230, 260 y 280 nm obtenidas utilizando el espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies). La relación 260:280 es indicativa de posible contaminación por proteínas mientras que el cociente 260:230 indica el grado de contaminación por carbohidratos y otros metabolitos. En general, se trabajó con RNA con valores superiores a 1.9 para ambos cocientes.

2.1. Protocolo general de extracción de RNA total

La extracción de RNA total de cítricos se realizó siguiendo el protocolo descrito por Malmberg et al (1985). Unos 500 mg de material vegetal triturado en mortero con N₂ líquido se homogeneizaron en 0.5 ml de tampón TCES (0.2 M Tris-HCl pH 8, 0.2 M NaCl, 50 mM EDTA, 2% SDS), 0.5 ml de fenol ácido (Prolabo) y 5 µl de β-mercaptoetanol y se incubaron 5 min a 50°C. El extracto se centrifugó 10 min a 2,000 rpm y el sobrenadante se transfirió a un tubo nuevo al que se añadió 0.8 ml de cloroformo:alcohol isoamílico (24:1), agitando con vórtex durante 2 min. La fase acuosa obtenida tras centrifugar 10 min a 2,000 rpm se transfirió a un microtubo que se volvió a centrifugar 15 min a 13,000 rpm. El sobrenadante se precipitó con LiCl 3 M a 4°C durante toda la noche y se recogió por centrifugación durante 15 min a 13,000 rpm. El precipitado se lavó tres veces consecutivas con 1 ml de etanol al 70%, se secó en vacío durante 5 min y se resuspendió en 40 µl de agua tratada con dietil pirocarbonato (DEPC). El rendimiento obtenido con este procedimiento fue de 200-300 µg de RNA total por 500 mg de peso fresco.

2.1.1. Purificación adicional del RNA para los experimentos con micromatrices

El RNA total obtenido con el método descrito se purificó, adicionalmente, utilizando el sistema *RNeasy Mini Kit* (Qiagen), siguiendo las instrucciones del fabricante. Para ello, 100 µg de RNA total (en un volumen de 100 µl) se mezclaron con 350 µl de tampón de lisis (contiene tiocianato de guanidinio y β-mercaptoetanol) y con 250 µl de etanol (100%). La muestra se transfirió a una columna *RNeasy*, donde el RNA se une a una membrana de gel de sílice. La columna se lavó dos veces con un tampón que contiene etanol, se secó por centrifugación a 13,000 rpm durante 1 min y el RNA se eluyó en 30 µl de agua libre de RNasa (Qiagen). El rendimiento obtenido con el procedimiento descrito fue de 75-90 µg de RNA purificado por 100 µg de RNA total.

2.1.2. Purificación adicional del RNA para los experimentos de RT-PCR

En los experimentos de RT-PCR, el RNA total obtenido mediante el método general se trató con DNasa I libre de RNasa (Qiagen) para eliminar la contaminación de las muestras con DNA genómico. El procedimiento seguido fue similar al descrito en el apartado 2.1.1 pero la muestra de RNA, fijada a la membrana de la columna

RNeasy, se digirió con 3 U de DNasa I, siguiendo las instrucciones del fabricante, antes de proceder al lavado de la columna y a la elución del RNA.

2.2. Extracción de poli(A⁺) RNA

El poli(A⁺) RNA se extrajo con el sistema *QuickPrep micro mRNA Purification Kit* (Amersham Biosciences) a partir de 250 µg de RNA total obtenidos según el protocolo descrito por Malberg et al (1985). A la solución acuosa de RNA total se añadió 400 µl del tampón de extracción (contiene tiocianato de guanidinio y N-lauroil sarcosinato) y 800 µl de tampón TE (10 mM Tris-HCl pH 7.5, 1 mM EDTA) y la mezcla se centrifugó 1 min a 13,000 rpm. La fase superior se transfirió al lecho de oligo(dT)-celulosa obtenido previamente por centrifugación y se resuspendió. Dicha suspensión se centrifugó 10 s a 13,000 rpm y se lavó cinco veces con 1 ml de tampón TEN (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5 M NaCl), dos veces más con 1 ml de tampón TEN 0.1X (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1 M NaCl), se resuspendió en 0.3 ml de este último tampón y se transfirió a una micro-columna de polipropileno. La columna se centrifugó 5 s a 13,000 rpm y se lavó tres veces con 0.5 ml de tampón TEN 0.1X. Finalmente, el poli(A⁺) RNA se eluyó dos veces con 0.2 ml de tampón TE a 65°C. Este RNA poli(A⁺) se precipitó con 0.1 volúmenes de acetato potásico 2.5 M pH 5, 2.5 volúmenes de etanol y 10 µl de glucógeno (5 mg/ml) durante toda la noche a 4°C. El poli(A⁺) se recogió por centrifugación a 13,000 rpm durante 10 min a 4°C, se lavó con etanol al 70% y se resuspendió en 20 µl de agua libre de RNasa (Amersham Biosciences). El rendimiento obtenido en las extracciones de poli(A⁺) RNA fue alrededor del 2%.

3. Construcción y manipulación de una genoteca de cDNA

3.1. Genoteca primaria

La genoteca direccional de cDNA de tejidos vegetativos en desarrollo se construyó a partir de una mezcla a partes iguales de RNA procedente de brotes jóvenes tratados con GA₃, brotes jóvenes tratados con paclobutrazol y hojas y entrenudos procedentes de brotes jóvenes sin tratar. Para ello se utilizó el sistema *λZAP-cDNA Synthesis Kit and Gigapack III Gold Cloning Kit* (Stratagene), siguiendo las instrucciones del fabricante (<http://www.stratagene.com/manuals/200401.pdf>). La primera cadena de cDNA se sintetizó a partir de 5 µg de RNA poli (A⁺), tratado 5 min a 65°C para relajar la estructura secundaria, utilizando transcriptasa reversa (50 U/µl),

oligo (dT) que contiene un sitio de restricción *Xho*I como cebador y en presencia de 5-metil-dCTP, de modo que se obtuvo un cDNA hemimetilado. En paralelo se realizó una síntesis control que además incorpora [α - 32 P] dATP (800 Ci/mmol). Antes de la síntesis de la segunda cadena de cDNA, el cDNA se incubó con RNasa H (1.5 U/ μ l), que degrada la cadena de RNA molde utilizado para la síntesis de la primera cadena. La segunda cadena de cDNA se sintetizó con DNA polimerasa I (9 U/ μ l), en presencia de [α - 32 P] dATP (800 Ci/mmol) para comprobar la calidad del cDNA obtenido en un gel de agarosa alcalino. Tras formar extremos romos en el cDNA de doble cadena mediante *Pfu* DNA polimerasa (2.5 U/ μ l), se procedió a la ligación, mediante T4 DNA ligasa (4 U/ μ l), de unos adaptadores que incluyen un extremo cohesivo *Eco*RI y a la posterior fosforilación de los mismos mediante T4 polinucleótido kinasa (10 U/ μ l). El cDNA modificado con los adaptadores *Eco*RI se digirió con la enzima de restricción *Xho*I (40 U/ μ l) de modo que se obtuvo un cDNA con un sitio *Eco*RI en el extremo 5' y un sitio *Xho*I en el extremo 3'. Este cDNA se fraccionó por tamaño con una columna Sephacryl S-500 (Life Technologies) y se seleccionaron las fracciones correspondientes a cDNAs de longitud mayor a 500 pb. El cDNA (100 ng) se ligó de forma direccional con el vector Uni-ZAP XR y los productos de la ligación se empaquetaron utilizando el extracto de empaquetamiento Gigapack III Gold (Stratagene). El número de fagos existentes en la genoteca primaria se calculó mediante la titulación de la misma. Para ello, una colonia de la cepa XL1-Blue MRF' se creció a 37°C en medio líquido YT 2X (16 g/l triptona, 5 g/l NaCl, 10 g/l extracto de levadura) suplementado con 0.2% maltosa y 10 mM MgSO₄ hasta una OD₆₀₀ de 0.5. Las células XL1-Blue MRF' (200 μ l) se incubaron 15 min a 37°C con varias diluciones de la genoteca primaria y se crecieron a 37°C durante toda la noche en placas de agar YT 2X (16 g/l triptona, 5 g/l NaCl, 10 g/l extracto de levadura, 15 g/l agar bacteriológico) suplementadas con 10 mM MgSO₄. El número de calvas o pfu (*plaque forming units*) obtenidas con las distintas diluciones permitió calcular la titulación de la genoteca.

3.2. Genoteca amplificada

La amplificación de la genoteca se realizó en 20 alícuotas de unos 5×10^4 pfu aproximadamente con 600 μ l de células XL1-Blue MRF' (Stratagene). Las células infectadas se sembraron en placas de 150 mm con agar NZY (5 g/l NaCl, 2 g/l MgSO₄, 5 g/l extracto de levadura, 10 g/l hidrolizado de caseína, 15 g/l agar) y se incubaron a 37°C hasta que comenzaron a diferenciarse las calvas. Se añadió 8 ml de solución SM (5.8 g/l NaCl, 2 g/l MgSO₄, 50 ml Tris-HCl 1 M, 5 ml gelatina 2%) a cada placa y se incubó con agitación suave a 4°C durante toda la noche para facilitar la difusión de los

fagos. La suspensión de fagos se recogió, se añadió cloroformo 5% y se centrifugó para eliminar los restos celulares. El sobrenadante se recogió y se añadió cloroformo hasta una concentración final de 0.3% (v/v). La genoteca amplificada, denominada Veg1, se almacenó a -80°C en alícuotas con dimetil sulfóxido (DMSO) al 7% (v/v) y se tituló siguiendo el mismo procedimiento descrito para la genoteca primaria.

3.3. Escisión masiva

El vector Uni-ZAP XR permite la escisión *in vivo* del fagémido pBluescript de modo que los insertos pueden ser caracterizados en un sistema de plásmidos. La escisión masiva de la genoteca amplificada se realizó a partir de 10^7 pfu de fagos, 10^8 pfu de células XL1-Blue MRF' y 10^9 pfu de fago *ExAssist* (Stratagene). Las células coinfectadas se incubaron en medio LB (10 g/l triptona, 10 g/l NaCl, 5 g/l extracto de levadura) a 37°C durante 2-3 h con agitación, se lisaron por incubación a 65-70°C durante 20 min, se centrifugaron y se recogió el sobrenadante que contiene los fagémidos con los clones en forma de DNA de cadena simple. La infección de 200 μ l de células SOLR (OD_{600} de 1) con 1 μ l de este sobrenadante se produjo por incubación a 37°C durante 15 min y permitió obtener DNA de doble cadena en forma de plásmidos. Las células SOLR infectadas con los fagémidos (10 μ l) se sembraron en placas de agar LB (10 g/l triptona, 10 g/l NaCl, 5 g/l extracto de levadura, 20 g/l agar bacteriológico) con ampicilina (50 μ g/ml) y se incubaron a 37°C durante toda la noche para obtener colonias bacterianas aisladas.

4. Generación de una colección de ESTs

4.1. Obtención de DNA plasmídico y secuenciación

4.1.1. Cultivos bacterianos

Las colonias bacterianas aisladas generadas por escisión masiva de la genoteca amplificada se seleccionaron al azar y se cultivaron en placas de 96 pocillos (Eppendorf) con 1.5 ml de medio LB 2X (20 g/l triptona, 10 g/l extracto de levadura, 10 g/l NaCl) y ampicilina (50 mg/l). Los cultivos en placa se incubaron a 37°C con agitación (300 rpm) durante 20 h. Una parte del cultivo (150 μ l) se transfirió a una nueva placa de 96 pocillos (Axygen) a la que se añadió glicerol hasta una concentración final de 15%. Estas placas, que suponen un duplicado de cada clon, se sellaron con láminas estériles de aluminio y se almacenaron a -80°C.

4.1.2. Obtención de DNA plasmídico

En este trabajo se han utilizado dos métodos para la obtención de DNA plasmídico en placas de 96 pocillos: el sistema *PerfectPrep* (Eppendorf) y el método *Montage* (Millipore).

Sistema *PerfectPrep* (Eppendorf)

El DNA plasmídico se obtuvo por lisis alcalina mediante el sistema *PerfectPrep Plasmid 96 Vac Direct Bind* (Eppendorf) siguiendo las instrucciones del fabricante. Las placas con el cultivo bacteriano se centrifugaron a 3000 rpm durante 5 min, se eliminó el sobrenadante y los sedimentos bacterianos se resuspendieron completamente en la solución 1 que contiene RNasa A (150 µl) utilizando el vórtex. A continuación se añadió la solución 2 de lisis alcalina (150 µl), la solución 3 de neutralización (150 µl) y este lisado bacteriano se transfirió a una nueva placa que contiene un filtro. Se aplicó vacío durante 5 min de modo que los restos celulares quedaron retenidos en el filtro y los lisados pasaron a la placa inferior, que contiene una membrana. En esta segunda placa se añadió 300 µl de la solución 4 que contiene guanidinio, se aplicó vacío durante 2 min, se añadió 400 µl de tampón salino y se aplicó vacío durante 5 min más. El DNA plasmídico retenido en la membrana de esta segunda placa se eluyó en 30 µl de agua estéril por aplicación de vacío durante 5 min.

Método *Montage* (Millipore)

El método de purificación de DNA plasmídico *Montage* (Millipore) consiste en una versión modificada del método de lisis alcalina. El procedimiento es muy similar al descrito en el apartado anterior salvo que, en este caso, todos los pasos se realizaron de forma automatizada utilizando el dispensador de líquidos MultiProbe II (Packard).

4.1.3. Secuenciación

La cantidad de DNA plasmídico obtenido se estimó mediante electroforesis en geles de agarosa 1% y TAE 1X de una alícuota del mismo, obteniendo una concentración de unos 50-100 ng/µl. Una alícuota del DNA plasmídico se utilizó para la secuenciación de los clones desde el extremo 5', que se llevó a cabo en el Servicio de Secuenciación del IBMCP. La secuenciación se basó en el uso de dideoxinucleótidos marcados con compuestos fluorescentes, utilizando el sistema *ABI PRISM Big Dye*

Terminator Cycle Sequencing Kit (Applied Biosystems). Las reacciones de secuenciación se realizaron en un termociclador *ABI GeneAmp PCR System 9700* (Applied Biosystems) y los productos marcados se resolvieron y analizaron en un secuenciador capilar *ABI 3100* (Applied Biosystems). Finalmente, otra alícuota del DNA plasmídico se utilizó para la construcción de la micromatriz de cDNA del CFGP, tal como se describe en el apartado 5.

4.2. Pre-procesado y ensamblaje de ESTs

El pre-procesado y ensamblaje de las secuencias obtenidas se realizó con EST2uni (Forment et al, 2008; <http://bioinf.comav.upv.es/est2uni>), herramienta bioinformática desarrollada en el CFGP. El pre-procesado de las secuencias consta de los siguientes pasos: 1) lectura de la secuencia nucleotídica definida en cada cromatograma y asignación de un valor de calidad a cada base, mediante el programa Phred (Ewing et al, 1998); 2) eliminación de las regiones de baja calidad y de las regiones procedentes del vector con el programa Lucy (Chou y Holmes, 2001); 3) enmascaramiento de elementos repetitivos y regiones de baja complejidad con los programas RepeatMasker (<http://www.repeatmasker.org>) y SeqClean (<http://compbio.dfci.harvard.edu/tgi/software>), respectivamente; y 4) detección y eliminación de secuencias contaminantes con el programa SeqClean utilizando la base de datos UniVec del NCBI (<http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html>). Tras el pre-procesado de las secuencias, todas aquéllas que no superaron una longitud de 100 bases no se incluyeron en los análisis posteriores.

Las secuencias de alta calidad, obtenidas mediante el procedimiento de pre-procesado descrito, se agruparon con el doble objetivo de eliminar redundancia y obtener un conjunto de posibles genes únicos. El agrupamiento o ensamblaje de las ESTs de alta calidad se realizó con la herramienta TGICL (Perteza et al, 2003) que, en un primer paso, agrupa las secuencias basándose en la comparación por parejas de todas ellas utilizando el algoritmo Megablast (Zhang et al, 2000) y, en un paso posterior, ensambla cada grupo de secuencias utilizando el algoritmo CAP3 (Huang y Madan, 1999), manteniendo los parámetros que establece el algoritmo por defecto: 1) solapamiento entre dos secuencias >30 pb; 2) identidad de las secuencias en la región solapada >94%; y 3) número de bases que no solapan <30 pb. Tras el agrupamiento, para cada grupo de secuencias o *contig* (conjunto de secuencias formado por dos o más ESTs), se obtuvo una secuencia consenso. Las ESTs que, según los parámetros anteriores, no forman parte de ningún grupo constituyen los *singletons* (ESTs que no muestran similitud con ninguna otra secuencia de la

colección). Así pues, la colección de ESTs generada queda agrupada en un conjunto no redundante de posibles genes únicos (o unigenes) formado tanto por los *contigs* como por los *singletons*. La secuencia deducida de aminoácidos de cada unigen se obtuvo utilizando ESTScan (Iseli et al, 1999).

4.3. Anotación funcional de los unigenes

4.3.1. Metodología para la anotación básica de los unigenes: base de datos EST2uni

La anotación funcional de los unigenes de cítricos se llevó a cabo utilizando la base de datos EST2uni (Forment et al, 2008; <http://bioinf.comav.upv.es/est2uni>). Esta herramienta realiza búsquedas automáticas con el algoritmo BLAST (Altschul et al, 1997) frente a la base de datos UniRef90 (Wu et al, 2006; <http://www.ebi.ac.uk/uniref>; UniProtKB versión 8.9; 17 Octubre 2006) y al conjunto completo de proteínas de Arabidopsis obtenido del TAIR (Rhee et al, 2003; <http://www.arabidopsis.org>; TAIR6 Noviembre 2005), utilizando los parámetros por defecto y un criterio no estricto de 10^{-5} para el valor E . Los unigenes fueron anotados con la descripción del conjunto de proteínas de UniRef90 más similar o, con la descripción de la proteína de Arabidopsis más similar. Los unigenes se anotaron como 'altamente similares' cuando el primer resultado del BLAST presentaba un valor E menor que 10^{-15} . Los unigenes cuyo primer resultado del BLAST mostraba un valor E mayor que 10^{-5} no se consideraron para la anotación. Adicionalmente, se ha llevado a cabo la anotación funcional de los unigenes de cítricos basada en motivos funcionales utilizando el algoritmo de búsqueda HMMER (Eddy, 1998) frente a la base de datos de familias de proteínas Pfam (<http://pfam.janelia.org>; Bateman et al, 2004).

4.3.2. Métodos para la anotación adicional de los unigenes

La anotación básica de los unigenes realizada con EST2uni, en particular de los unigenes procedentes de la genoteca de tejidos vegetativos y de los unigenes implicados en el metabolismo de isoprenoides y flavonoides, se suplementó con información adicional completando así la anotación existente. Esta anotación adicional se realizó de forma automática cuando fue posible; sin embargo, en algunos casos, exigió una revisión manual de la anotación obtenida con EST2uni. Los métodos que se utilizaron para desarrollar esta anotación adicional se describen a continuación.

Categorías funcionales

Los unigenes de cítricos representados en la genoteca de tejidos vegetativos se anotaron según dos modelos de anotación adicionales: la ontología génica (GO; Ashburner et al, 2000) organizada en tres categorías generales ('proceso biológico', 'función molecular' y 'componente celular') y la anotación funcional del MIPS (Munich Information Centre for Protein Sequences; <http://mips.gsf.de>; MIPS v2.1). La asignación de GO se realizó utilizando el programa FatiGO (Al-Shahrour et al, 2004; <http://www.babelomics.org>) mientras que la anotación funcional del MIPS se obtuvo mediante la herramienta FunCat (Ruepp et al, 2004). Para ello, se utilizó el código AGI (*Arabidopsis Genome Initiative*, 2000) correspondiente al gen de *Arabidopsis* más similar a cada unigen de cítricos.

El programa FatiGO se utilizó, adicionalmente, para llevar a cabo los análisis de enriquecimiento funcional puesto que permite encontrar términos GO significativamente sobre-representados en un conjunto de genes respecto a un grupo de referencia. En este análisis se tuvo en cuenta la corrección por experimentos múltiples, pues de no ser así se produce un aumento en la proporción de falsos positivos, es decir, términos GO identificados como sobre-representados cuando, realmente, no lo están. El método de corrección utilizado fue el de FDR (*False Discovery Rate*) (Benjamini y Hochberg, 1995).

Rutas metabólicas

Con el objetivo de obtener una anotación más integrada de los unigenes de cítricos, se utilizaron tres bases de datos adicionales: MetaCyc (Krieger et al, 2004; <http://metacyc.org>), AraCyc (Mueller et al, 2003; <http://www.arabidopsis.org/biocyc>) y KEGG (Kanehisa et al, 2004; <http://genome.ad.jp/kegg>). MetaCyc es una base de datos que contiene rutas metabólicas verificadas experimentalmente e información sobre enzimas corregida a partir de la literatura científica. Una ruta metabólica en MetaCyc consiste en el conjunto de reacciones, enzimas y metabolitos que la componen y los genes que codifican dichas enzimas, así como la información sobre su regulación en diferentes especies, desde microorganismos a plantas y humanos. Las rutas de *Arabidopsis* se exportan a AraCyc, la base de datos del metabolismo de *Arabidopsis*, de modo que MetaCyc constituye la base de datos de referencia. La base de datos del KEGG (Kyoto Encyclopedia of Genes and Genomes) contiene información sobre redes de interacción molecular como rutas y complejos, sobre genes y proteínas y sobre compuestos y reacciones químicas. La metodología seguida

en este caso fue partir del código AGI correspondiente al gen de Arabidopsis más similar a cada unigen de cítricos, obtener la ruta metabólica en la que participa, extraer las secuencias de otros genes que forman parte de esa ruta y realizar búsquedas en la base de datos del CFGP, utilizando la herramienta EST2uni, para tratar de localizar otros elementos de cítricos de la misma ruta no anotados previamente.

Localización subcelular

La determinación de la localización subcelular de una proteína constituye un paso importante hacia la comprensión de su función. Los péptidos que dirigen las proteínas a la ruta secretora (SP), a mitocondria (mTP) o a cloroplastos (cTP) se encuentran en la región N-terminal de la secuencia. La secuencia deducida de aminoácidos de cada unigen de cítricos estudiado se obtuvo mediante la herramienta Translate (<http://www.expasy.org/tools/dna.html>) del *Swiss Institute of Bioinformatics* (SIB) y se comprobó que contenía la región N-terminal completa mediante BLASTP frente a la base de datos de proteínas no redundantes GenBank del NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>). La predicción de la localización subcelular de diversas proteínas de cítricos se realizó siguiendo la metodología descrita por Emanuelsson et al (2007). La secuencia de aminoácidos deducida de cada unigen de cítricos se analizó con TargetP v1.1 (Emanuelsson et al, 2000; <http://www.cbs.dtu.dk/services/TargetP/>) que asigna una posible localización, en plantas, para cloroplasto, mitocondria, ruta secretora y otras, y proporciona un coeficiente de fiabilidad (RC) de la predicción. Cuando TargetP predijo un posible péptido señal de cloroplasto (cTP), la secuencia se analizó con ChloroP v1.1 (Emanuelsson et al, 1999; <http://www.cbs.dtu.dk/services/ChloroP/>) para obtener un informe detallado de las puntuaciones asignadas a lo largo de la secuencia. El valor límite de la predicción es 0.5, de modo que puntuaciones superiores a 0.55 se consideran bastante sólidas. Cuando la predicción de TargetP fue 'otra localización' se utilizaron otras herramientas de predicción como PSORT (Nakai y Horton, 1999; <http://www.psort.org>) y BaCellO (Pierleoni et al, 2006; <http://gpcr.biocomp.unibo.it/bacello>).

4.4. Análisis de expresión génica según la frecuencia de ESTs

La frecuencia de ESTs procedentes de la genoteca Veg1 se utilizó para detectar genes expresados diferencialmente durante el desarrollo vegetativo. Así, para cada unigen de cítricos representado en la genoteca de tejidos vegetativos se calculó tanto la frecuencia de ESTs procedentes de dicha genoteca como la frecuencia de

ESTs procedentes del resto de la colección del CFGP. La identificación de unigenes expresados diferencialmente en Veg1 se realizó con el programa IDEG6 (Romualdi et al, 2003), utilizando el test estadístico de Audic y Claverie (1997) y la corrección por experimentos múltiples de Bonferroni.

5. Construcción de la micromatriz de cDNA de cítricos CIT1

Los análisis de expresión génica se realizaron utilizando la micromatriz de cDNA de cítricos, construida por el Laboratorio de Genómica del CFGP, que contiene 12,672 clones que corresponden a 6,875 posibles unigenes (Forment et al, 2005). Los 12,672 clones contenidos en la micromatriz, que llamaremos CIT1, proceden de 18 genotecas de cDNA (Tabla 1).

Tabla 1. Procedencia de los clones que componen la micromatriz CIT1			
Genoteca	Nº clones (%)	Especie	Tejido/ estado de desarrollo/ tratamientos
Veg1	1,152 (9.1)	<i>C. clementina</i>	Hojas, entrenudos y brotes apicales: aplicaciones foliares de GA ₃ y paclobutrazol
IF1	1,920 (15.2)	<i>C. clementina</i>	Flores en distintos estadios de desarrollo, desde yemas florales hasta flores en anthesis
OF1	192 (1.5)	<i>C. clementina</i>	Ovarios senescentes y frutos de 5 semanas tratados con GA ₃
OF2	1,344 (10.6)	<i>C. clementina</i>	Ovarios de yemas florales a frutos de 1 y 3 semanas tratados con GA ₃
AbsAOv1	128 (1.0)	<i>C. clementina</i>	Zona de abscisión A de ovarios en desarrollo tratados con etileno
CTVLeafMc1	1,248 (9.8)	<i>C. macrophylla</i>	Hojas infectadas con el <i>Virus de la Tristeza de los Cítricos</i>
DroRLeaf1	177 (1.4)	<i>C. clementina</i>	Hojas de plantas rehidratadas tras 24 h de estrés hídrico
Drought1	672 (5.3)	<i>C. clementina</i>	Hojas de plantas tras 5, 10, 24 h de estrés hídrico
Drought2	864 (6.8)	<i>C. reshni</i>	Raíces de plantas tras 5, 10, 24 h de estrés hídrico
FlavFr1	1,344 (10.6)	<i>C. clementina</i>	Flavedo de frutos maduros almacenados a 2°C durante 7 y 24 días
FlavRip1	215 (1.7)	<i>C. clementina</i>	Flavedo de frutos en desarrollo
FlavSen1	288 (2.3)	<i>C. clementina</i>	Flavedo de frutos senescentes
LongSaltL1	207 (1.6)	<i>C. sinensis</i>	
NTLeaf1	384 (3.0)	<i>C. clementina</i>	Hojas de plantas cultivadas en condiciones estándar
PhyRootSr1	960 (7.6)	<i>C. aurantium</i>	Raíces infectadas con <i>Phytophthora citrophthora</i>
PhyRootSw1	864 (6.8)	<i>C. sinensis</i>	Raíces infectadas con <i>Phytophthora citrophthora</i>
Roots1	137 (1.1)	<i>citrange Carrizo</i>	Raíces de plántulas cultivadas 1-4 días con paclobutrazol o auxinas
SaltLeaf1	576 (4.5)	<i>C. clementina</i>	Hojas de plantas tratadas con 150 mM NaCl durante 6, 24 h y 7 días

La amplificación por PCR de los 12,672 clones para la posterior impresión de los mismos fue realizada por el Laboratorio de Genómica del IBMCP. Las reacciones de PCR se realizaron a partir de 4 ng de DNA plasmídico utilizando 1.5 U de *Taq* polimerasa, 200 μ M dNTPs, 400 nM de cebador T7 modificado (5'-CGACTCACTATA GGGCGAATTGG-3') y 400 nM de cebador T3 modificado (5'-GGAAACAGCTATGAC CATGATTAC-3'), en un volumen final de 100 μ l. Las reacciones de PCR se realizaron, en placas de 96 pocillos, a 94°C durante 3 min, seguido de 5 ciclos de amplificación (94°C durante 30 s, 66°C durante 30 s, 72°C durante 2 min), 30 ciclos más (94°C durante 30 s, 60°C durante 30 s, 72°C durante 2 min) e incubadas finalmente a 72°C durante 10 min. Una alícuota de los productos de PCR se analizó en geles de agarosa 1%, mientras que el resto se purificó utilizando el sistema *Multiscreen-PCR 96-well Filtration System* (Millipore) y se resuspendió en agua hasta una concentración final de 200-400 ng/ μ l. Los productos de PCR purificados se transfirieron a placas de 384 pocillos (Corning) hasta una concentración de 100-200 ng/ μ l en 50% (v/v) DMSO. Además, se incluyeron los controles *Lucidea Universal ScoreCard Kit* (Amersham Biosciences) diluidos a 100 ng/ μ l en 50% (v/v) DMSO para evaluar la precisión de los datos de expresión obtenidos. Los productos de PCR purificados y los controles se imprimieron en cristales de vidrio recubiertos con γ -aminopropilsilano (UltraGAPS, 25 x 75 mm) (Corning) utilizando un robot de impresión MicroGrid II (Biorobotics) en un formato de 48 bloques con 16 x 17 puntos por bloque con una humedad relativa constante del 45%. Tras la impresión, el DNA se fijó a los cristales de vidrio modificados mediante 150 mJ de irradiación con luz ultravioleta y las micromatrices se almacenaron en un desecador (Aname).

6. Experimentos de micromatrices de cDNA

Los experimentos con micromatrices de cDNA comienzan con el diseño del experimento, es decir, la definición de los objetivos, la selección y la obtención del material biológico apropiado para probar la hipótesis de interés a nivel transcripcional y la decisión sobre qué muestras deben hibridarse juntas. La siguiente fase consiste en la generación de los datos e incluye la extracción de RNA, el marcaje con cada uno de los compuestos fluorescentes, la hibridación con la micromatriz y la adquisición y el análisis de la imagen. El análisis de los datos generados incluye dos etapas: 1) el pre-procesado y normalización de los mismos, y 2) la identificación de los genes candidatos y la interpretación biológica de los datos obtenidos (Fig. 4). En este apartado se describe la metodología seguida en este trabajo en las diferentes etapas.

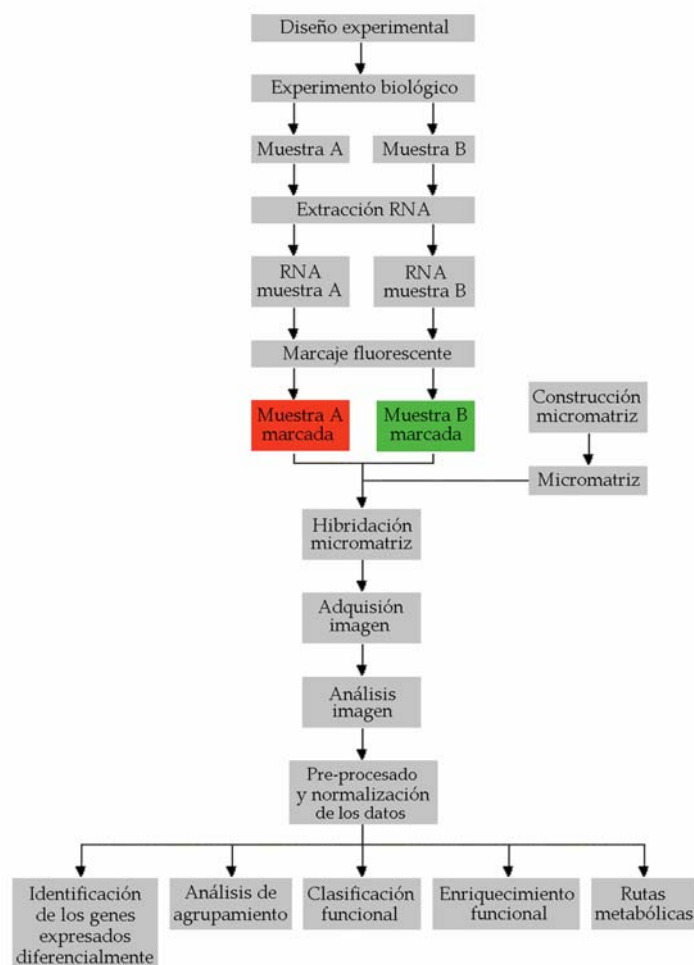


Fig. 4. Diagrama de flujo para un experimento tipo de análisis de la expresión génica mediante micromatrices de cDNA. La idea subyacente a un experimento de micromatrices de cDNA es que tras la hibridación competitiva entre la muestra marcada con el compuesto fluorescente Cy5 (rojo) y la marcada con Cy3 (verde), las medidas de fluorescencia obtenidas indican la abundancia relativa de cada transcrito correspondiente a cada sonda de la micromatriz en ambas muestras.

6.1. Diseño de experimentos

El principal objetivo del diseño de experimentos es que tanto el análisis de los datos como la interpretación de los resultados sea lo más simple y potente posible. La propiedad principal de los datos generados en un experimento de micromatrices de cDNA es que son, inherentemente, comparativos. El aspecto más importante de los principios de diseño para estos experimentos comparativos es determinar qué muestras de RNA se marcarán con cada compuesto fluorescente, cuáles se hibridarán juntas en la misma micromatriz y cuántas réplicas se realizarán.

Las cuestiones que nos hemos planteado en estos experimentos han sido, principalmente, de dos tipos: 1) ¿qué genes cambian su expresión en la condición *x* respecto a la condición control? y 2) ¿qué genes responden de modo similar en dos o

más condiciones cuando se comparan con la condición control? Los diseños de experimentos que se planteraron para abordar este tipo de cuestiones responden al esquema representado en la Figura 5.

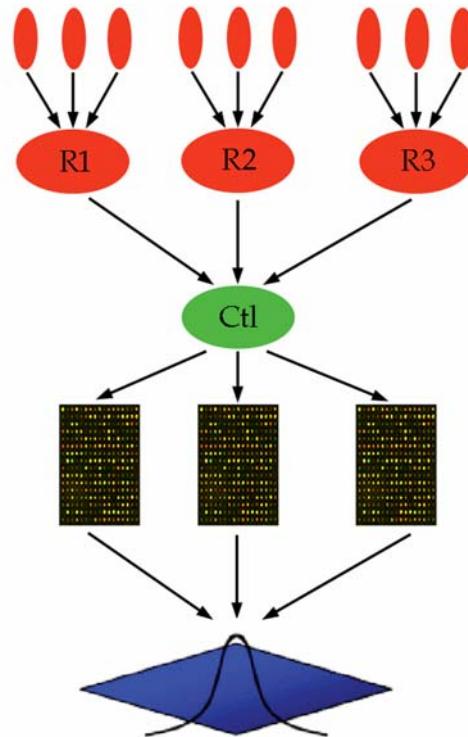


Fig. 5. Diseño de experimentos de micromatrices de cDNA. Las diferentes muestras problema (ej. brotes de un mismo individuo tratados con GA₃; círculos rojos pequeños) se agrupan constituyendo una réplica biológica (R1, R2, R3) y se marcan con el compuesto fluorescente Cy5. El control (Ctl) se genera a partir de diferentes muestras y constituye la muestra de referencia, que se marca con el compuesto fluorescente Cy3. Cada una de las réplicas biológicas se compara con la referencia mediante la co-hibridación en la misma micromatriz. Los datos obtenidos se promedian (rombo azul) para generar una distribución (media, desviación) que permite estimar la variación biológica.

Las características del diseño de experimentos utilizado en este trabajo son:

- 1) La condición control (Ctl) se convierte en una referencia *de facto* con significado biológico.
- 2) Todas las hibridaciones implican la comparación de cada una de las diferentes condiciones estudiadas con la condición control (Ctl). Se trata, por lo tanto, de comparaciones directas que generan una varianza inferior a la derivada de un diseño indirecto (Yang y Speed, 2002).
- 3) No se realizó marcaje reverso (*dye-swap*), es decir, hibridaciones por duplicado asignando los compuestos fluorescentes de forma opuesta en la segunda hibridación. La proporción de genes afectados por un sesgo en la incorporación de los compuestos fluorescentes se ha estimado en un 0.1% (Causton et al,

2003). Este problema es más frecuente cuando el marcaje del cDNA se realiza con el método directo. En todos los experimentos se utilizó el método indirecto de marcaje del cDNA (ver apartado 6.2.1) de modo que la proporción de genes potencialmente afectados por este problema es baja.

- 4) En los experimentos de micromatrices de cDNA se realizaron, al menos, tres réplicas de acuerdo con Lee et al (2000). Las repeticiones se realizaron partiendo de diferentes muestras biológicas (Fig. 5) por lo que se trata de réplicas biológicas (R1, R2, R3). Este tipo de réplicas implica un mayor grado de variación en los resultados obtenidos pero permite generalizar las conclusiones extraídas del experimento.

6.2. Preparación de muestras de cDNA marcadas

6.2.1. Marcaje mediante el método indirecto

El marcaje de las muestras de RNA para la hibridación de las micromatrices se realizó mediante el método indirecto. Este método utiliza nucleótidos dUTP modificados con un grupo aminoalilo [5-(3-aminoalilo)-2-deoxi-UTP, aa-dUTP] que se incorporan en la cadena de cDNA sintetizada durante la transcripción reversa. Por otro lado, los compuestos fluorescentes Cy3 y Cy5 utilizados contienen un grupo reactivo N-hidroxi-succinimida (NHS) de modo que el cDNA queda marcado por la reacción de estos grupos NHS con los grupos aminoalilo. El método indirecto que se utilizó es básicamente el descrito por *The Institute for Genomic Research* (TIGR; <http://www.arabidopsis.org>). La síntesis de cDNA se realizó utilizando 30 µg de RNA total (ver apartado 2.1.1), tratado 10 min a 70°C para relajar la estructura secundaria, y 400 U de transcriptasa reversa SuperScript III (Invitrogen), en presencia de 334 µM aa-UTP (Sigma), 6 µg oligo(dT)₂₄, 500 µM de dATP, dCTP y dGTP, 166 µM dTTP y 10 µM DTT en un volumen final de 30 µl, durante 3 h a 50°C. Las muestras se incubaron con 250 µM NaOH y 125 µM EDTA durante 15 min a 65°C y se neutralizaron con 250 µM HCl. La reacción de cDNA se purificó en una columna Qiaquick (Qiagen) para eliminar aa-dUTP no incorporado. El cDNA modificado con aa-dUTP se secó, se resuspendió en 10 µl NaHCO₃ 100 mM pH 9 y se incubó 1 h en oscuridad con los compuestos fluorescentes Cy3 ó Cy5 CyDye NHS-éster (Amersham Biosciences), con objeto de que los grupos NHS-éster de los compuestos fluorescentes reaccionen formando enlaces covalentes con los grupos aminoalilo del cDNA modificado. El cDNA marcado con los compuestos fluorescentes se purificó nuevamente con una columna Qiaquick (Qiagen) siguiendo las instrucciones del fabricante.

6.2.2. Caracterización de las propiedades del cDNA marcado

La calidad del cDNA marcado, tanto el tamaño de los fragmentos obtenidos como la eficiencia del marcaje, se estimó mediante electroforesis en geles de agarosa 1% y TAE 1X de una alícuota de cada muestra, utilizando el sistema de electroforesis BioRad Miniprotein (Bio-Rad Laboratories). Para ello, el cDNA marcado se secó, se resuspendió en 10 μ l de agua y una alícuota (1/10 del total) se cargó junto a 5 μ l de glicerol 15%. Tras la electroforesis, el gel se colocó sobre un portaobjetos de cristal, se deshidrató y se escaneó utilizando el escáner GenePix 4000B y el programa asociado GenePix 4.1 (Molecular Devices) (Fig. 6A). La cantidad de compuesto fluorescente incorporado se determinó con un espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies), midiendo la absorbancia de las muestras de cDNA marcadas a las longitudes de onda de 550 nm y 650 nm, que corresponden con la máxima absorción para Cy3 y Cy5, respectivamente (Fig. 6B).

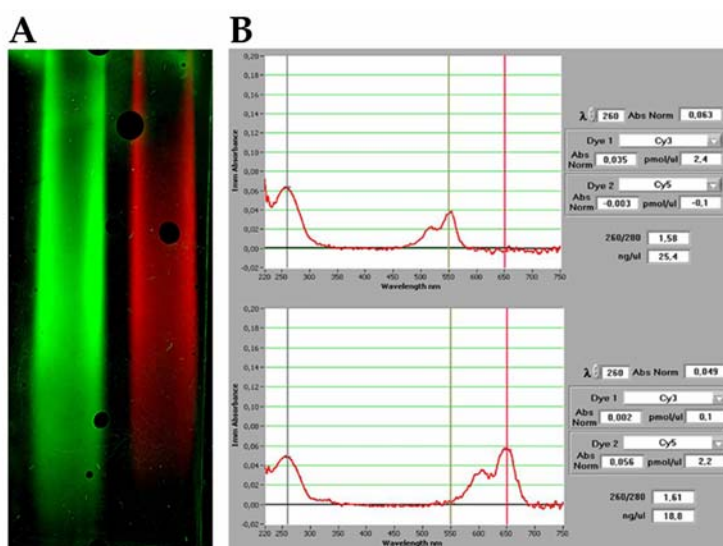


Fig. 6. Caracterización de las propiedades del cDNA marcado. La electroforesis en geles de agarosa 1% (A) permitió estimar la eficiencia del marcaje con los compuestos fluorescentes Cy3 (carril izquierdo) y Cy5 (carril derecho). La cantidad de cDNA marcado se determinó por espectrofotometría (B), tanto para Cy3 (panel superior) como para Cy5 (panel inferior).

6.3. Hibridación de las micromatrices

La hibridación de las micromatrices se realizó siguiendo los protocolos de hibridación recomendados para *UltraGAPS Coated Slides* (Corning), que es el soporte que se utilizó para la construcción de la micromatriz CIT1 (ver apartado 5).

6.3.1. Protocolo de hibridación A

Pre-hibridación

Las micromatrices se pre-hibridan con el objetivo doble de bloquear la superficie libre de la micromatriz para minimizar el ruido de fondo y de eliminar las sondas de DNA mal adheridas para maximizar la señal de hibridación. La pre-hibridación se realizó en solución SSC 3X (Sigma), 0.1% SDS (Sigma) y 0.1 mg/ml BSA durante 60 min a 50°C. Las micromatrices pre-hibridadas se lavaron a temperatura ambiente durante 15 s en agua y 2 min en isopropanol. Las micromatrices se secaron por centrifugación a 1,500 rpm durante 5 min.

Hibridación

Las micromatrices se hibridaron manualmente utilizando las cámaras de hibridación ArrayIt (Telechem International). Para ello, las muestras de cDNA marcadas, bien con el compuesto fluorescente Cy3 o bien con Cy5, se secaron por separado, se resuspendieron en la solución de hibridación compuesta por SSC 3X (Sigma), 0.1% SDS (Sigma) y 0.1 mg/ml DNA de esperma de salmón (Stratagene) en un volumen final de 80 µl, y se incubaron 1 min a 95°C. Las muestras desnaturalizadas permanecieron a temperatura ambiente protegidas de la luz directa. Las cámaras de hibridación se precalentaron a 50°C y se añadió 50 µl de agua en cada uno de los pocillos de la cámara y otros 50 µl repartidos en tres puntos en la parte central. En la cámara precalentada se depositó la micromatriz que se cubrió con un cristal LifterSlip (Erie Scientific). Las muestras se inyectaron desde la parte inferior de la micromatriz de modo que el espacio de la hibridación se rellena por capilaridad. Una vez aplicadas las muestras a las cámaras de hibridación, éstas se cerraron y la hibridación se realizó en un baño de agua a 50°C durante toda la noche en oscuridad.

Lavados

Tras la hibridación, las micromatrices se lavaron secuencialmente, con agitación, en SSC 2X (Sigma), 0.1% SDS (Sigma) a 42°C durante 5 min, en SSC 0.1X, 0.1% SDS durante 10 min a temperatura ambiente, 5 veces en SSC 0.1X durante 1 min a temperatura ambiente, y finalmente en SSC 0.01X durante 10 s. Las micromatrices se secaron centrifugando a 1,500 rpm durante 5 min.

6.3.2. Protocolo de hibridación B

Este protocolo de hibridación presenta diversas modificaciones respecto al protocolo A, entre ellas la incorporación de 50% (v/v) formamida en la solución de hibridación. En la realización de este trabajo se ha comprobado que la utilización de este protocolo supone una mejora substancial de la hibridación, evaluada mediante los parámetros de calidad obtenidos durante el análisis de los datos (ver apartado 6.6.1). De hecho, las recomendaciones del fabricante de *UltraGAPS Coated Slides* (Corning) actualmente sólo incorporan este protocolo de hibridación.

Pre-hibridación

La pre-hibridación se realizó en tampón SSC 5X (Sigma), 0.1% SDS (Sigma) y 0.1 mg/ml BSA durante 60 min a 42°C. Las micromatrices pre-hibridadas se lavaron dos veces en SSC 0.1X a temperatura ambiente durante 5 min y una vez en agua a temperatura ambiente durante 30 s. Las micromatrices se secaron centrifugando a 1,600 rpm durante 2 min.

Hibridación

Las micromatrices se hibridaron manualmente utilizando las cámaras de hibridación ArrayIt (Telechem International). Para ello, las muestras de cDNA marcadas, bien con el compuesto fluorescente Cy3 o bien con Cy5, se resuspendieron en una solución de hibridación fresca que contiene 50% formamida (v/v), SSC 5X (Sigma), 0.1% SDS (Sigma) y 0.1 mg/ml DNA de esperma de salmón (Stratagene) en un volumen final de 80 µl, y se incubaron 2 min a 95°C. Las muestras desnaturalizadas permanecieron a temperatura ambiente protegidas de la luz directa. La hibridación se realizó tal como se ha descrito en el protocolo A con la única modificación de que la temperatura de hibridación fue de 42°C.

Lavados

Tras la hibridación, las micromatrices se lavaron secuencialmente, con agitación, en SSC 2X (Sigma), 0.1% SDS (Sigma) a 42°C durante 5 min, dos veces en SSC 0.1X, 0.1% SDS durante 5 min a temperatura ambiente, 5 veces en SSC 0.1X durante 1 min a temperatura ambiente, y finalmente en SSC 0.01X durante 10 s. Las micromatrices se secaron por centrifugación a 1,600 rpm durante 2 min.

6.4. Adquisición y análisis de la imagen

El primer paso en el análisis de los datos es el escaneado de la micromatriz y la extracción de los datos de intensidad de las imágenes. En el proceso de adquisición y análisis de la imagen hay cuatro pasos básicos: 1) escaneado, 2) reconocimiento de cada punto o *gridding*, 3) segmentación y 4) extracción de la intensidad y cálculo del cociente de expresión.

Las micromatrices se escanearon con un escáner GenePix 4000B (Molecular Devices) utilizando dos versiones del programa asociado: GenePix 4.1 y GenePix 6.0 (empleado en los experimentos de micromatrices de la última parte de este trabajo) y con unos parámetros de excitación a 532 y 635 nm para Cy3 y Cy5, respectivamente, resolución de 10 μm y potencia del láser del 100%. Cuando se utilizó el programa GenePix 4.1, los valores de ganancia del tubo fotomultiplicador (PMT) se ajustaron manualmente de modo que la intensidad global de la señal adquirida en cada canal (Cy3 y Cy5) fuera similar. Este ajuste se consiguió cuando los histogramas de distribución de los píxel (Fig. 7) en cada canal solapan, de modo que el cociente entre ambos es, aproximadamente, igual a 1. Al mismo tiempo, el número de elementos con píxel saturados se minimizó (unos 10 elementos de los 12,672 contenidos en la micromatriz CIT1) con el objetivo de maximizar el rango dinámico del escáner. El programa GenePix 6.0 permite realizar un ajuste automático de los valores de ganancia de PMT una vez fijada la proporción de píxel saturados. En nuestro caso, se utilizó un valor de saturación del 0.04%.

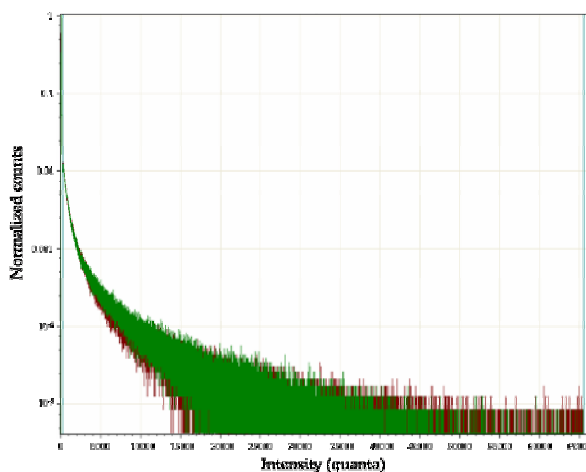


Fig. 7. Histograma de la distribución de píxel de cada canal, Cy3 (verde) y Cy5 (rojo). El ajuste de los histogramas de distribución de los píxel se consigue cuando éstos solapan.

El reconocimiento de los puntos o *gridding* es el proceso computacional que permite localizar cada punto de la micromatriz en la imagen obtenida. Para ello, un fichero GAL (*GenePix Array List*) o gradilla de reconocimiento queda superpuesto a la imagen. Este fichero se genera durante el proceso de impresión de la micromatriz (ver apartado 5), está organizado en bloques y contiene la información referente a la posición de cada punto, un identificador (ID) único para cada punto y la descripción de cada ID, entre otros campos. Se realizó un ajuste general de la gradilla de forma automática y un ajuste más preciso de cada bloque de forma manual que permitió, adicionalmente, marcar los puntos de mala calidad (intensidad heterogénea, morfología aberrante), quedando así eliminados de posteriores análisis.

La segmentación es el proceso que permite diferenciar, para cada punto de la micromatriz, los píxel que forman parte de la señal verdadera de intensidad (*foreground*: F635, F532) de los píxel que constituyen la señal de fondo (*background*: B635, B532). El algoritmo utilizado en el proceso de segmentación fue el de segmentación circular adaptativo de modo que a cada punto de la micromatriz se le aplicó un círculo de diámetro variable que contenía los píxel que forman parte de la señal de intensidad. La estimación de la señal de fondo se realizó mediante el método local que utiliza, para cada punto, la intensidad de los píxel que rodean el círculo establecido en la segmentación. Tras el proceso de segmentación, se calcularon dos valores de intensidad para cada punto de la micromatriz y para cada canal, la intensidad de la señal (como la mediana de la intensidad de los píxel que constituyen la señal: F635Median y F532Median) y la intensidad del fondo (como la mediana de la intensidad de los píxel que constituyen el fondo: B635Median, B532Median). La intensidad neta o corregida de cada punto se obtuvo mediante la sustracción de la intensidad del fondo a la intensidad de la señal en cada canal. Los cocientes de expresión entre ambos canales se calcularon utilizando las intensidades netas o corregidas, es decir,

$$\text{Ratio of Medians} = (\text{F635Median} - \text{B635Median}) / (\text{F532Median} - \text{B532Median})$$

6.5. Pre-procesado y normalización de los datos

Los datos extraídos durante el análisis de la imagen deben ser pre-procesados para excluir los puntos de baja calidad y normalizados para eliminar los errores sistemáticos generados en el proceso. El pre-procesado de los datos consistió en eliminar todos aquellos puntos de la micromatriz cuya intensidad neta fue menor a dos veces el valor medio del fondo global de la micromatriz en los dos canales. Es decir,

$$Flag = [(F635Mediana - B635Mediana) \leq 2 \times FondoGlobal_{635}] \text{ y} \\ [(F532Mediana - B532Mediana) \leq 2 \times FondoGlobal_{532}]$$

Los puntos de la micromatriz que cumplieron la citada condición quedaron eliminados para los análisis posteriores. Los cocientes de expresión se transformaron en logaritmo en base dos (\log_2), de modo que la interpretación biológica de los valores de expresión es más sencilla ya que esta transformación permite comparar, en la misma escala, los valores de inducción y los de represión. La normalización es el proceso por el que la variación sistemática debida a causas no biológicas se minimiza y se estandariza (Fig. 8). Este tipo de variación tiene que ver con factores técnicos como las diferencias en la cantidad de la muestra marcada o en la incorporación de los dos compuestos fluorescentes (Cy3, Cy5), la sensibilidad de los mismos a la luz o al calor, así como los parámetros del escaneado de ambos canales (Leung y Cavalieri, 2003).

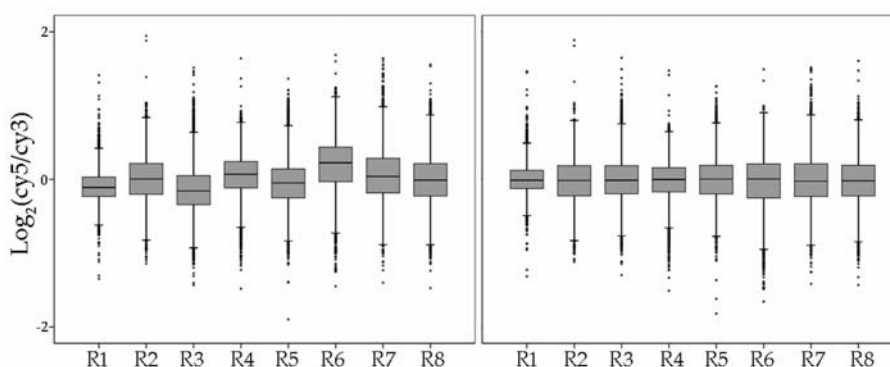


Fig. 8. Normalización global. El panel de la izquierda muestra los diagramas de caja correspondientes a los datos de expresión, \log_2 (Cy5/Cy3), en las diferentes réplicas de uno de los experimentos realizados en este trabajo, antes de la normalización. El panel de la derecha muestra los diagramas de caja correspondientes a los mismos datos tras la normalización global.

En un experimento de micromatrices cabe esperar que la mayoría de los genes analizados no cambien su expresión, es decir, que la distribución de los valores de expresión, \log_2 (Cy5/Cy3), tenga por media el valor cero. La distribución se centró en cero mediante una normalización lineal global. Así pues, todos los puntos se corrigieron por el mismo factor de modo que la forma de la distribución no varió (Fig. 8). La normalización permite comparar los resultados obtenidos en las diferentes micromatrices de un experimento. Los valores correspondientes a \log_2 (Cy5/Cy3) pueden tener una dependencia sistemática según la intensidad, que se manifiesta como una desviación respecto al cero para los valores de baja intensidad (Causton

et al, 2003). La transformación de los datos mediante la función *Lowess* (*Locally weighted polynomial regression*) ha sido propuesta como un método de normalización para los experimentos con micromatrices que corrige este efecto (Yang et al, 2002). Los datos de micromatrices obtenidos en este trabajo se normalizaron mediante la función *Lowess* utilizando el programa Acuity (Molecular Devices) (Fig. 9).

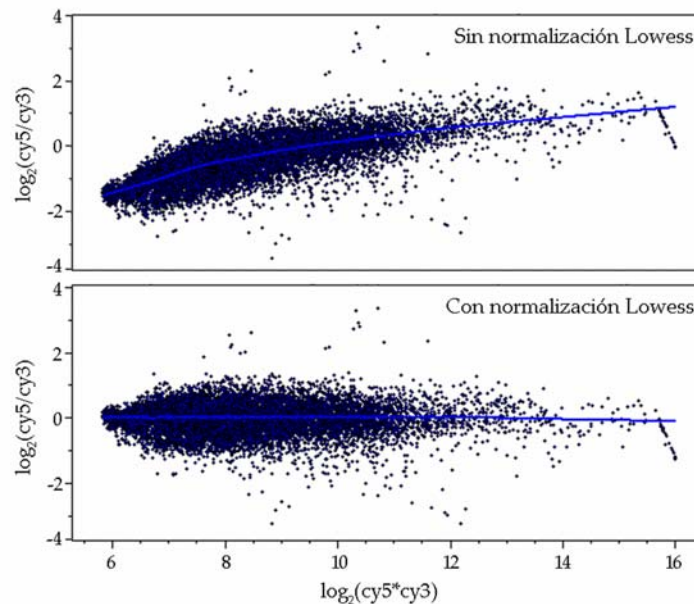


Fig. 9. Normalización *Lowess*. La representación de los datos de expresión, $\log_2(\text{Cy5}/\text{Cy3})$, respecto a la suma de las intensidades de ambos canales, $\log_2 \text{Cy5} + \log_2 \text{Cy3}$, permite observar la desviación de la distribución respecto a cero a bajas intensidades (panel superior). Este efecto es minimizado mediante la transformación *Lowess* de los datos (panel inferior).

6.6. Evaluación de la calidad de los datos obtenidos con micromatrices

6.6.1. Parámetros de calidad de GenePix

El programa de análisis de imágenes de micromatrices, GenePix (Molecular Devices), permite realizar un 'control de calidad de cada micromatriz' mediante un análisis estadístico de los datos y el cálculo de diversos parámetros que permiten estimar la calidad de los datos obtenidos en cada micromatriz. En la realización de este trabajo, se observó que los valores obtenidos para los parámetros de calidad varían substancialmente según el protocolo de hibridación empleado (Tabla 2). Las diferencias más notables se refieren a una reducción de la señal de fondo (procedente de la unión inespecífica de la sonda), especialmente en el canal de Cy3 (532 nm), de modo que el valor medio del fondo se redujo, al menos, a la mitad utilizando el protocolo de hibridación B (ver apartado 6.3.2). Esta reducción conllevó un aumento

en la relación señal-fondo y una reducción en la proporción de puntos cuya señal no se detecta, es decir, que no superan el criterio establecido en el apartado 6.5. La mejora en la calidad de los datos obtenidos con el protocolo de hibridación B nos llevó a utilizar este método en los sucesivos experimentos de micromatrices.

Tabla 2. Control de calidad de los datos obtenidos con micromatrices

Protocolo A					
	Cy5 (635)	Cy3 (532)	Threshold	Results	
				635	532
Median signal-to-background	10.3	6.5	> 10	Pass	Fail
Mean of median background	130.054	220.07	< 500	Pass	Pass
Median signal-to-noise	35.5	17.3	> 10	Pass	Pass
Feature variation	0.289	0.295	< 0.5	Pass	Pass
Background variation	0.436	0.508	< 0.5	Pass	Fail
Features with saturated píxel	1.7	0.63	< 1%	Fail	Fail
Not Found features	2127/12767 (16.6%)		< 10%	Fail	
Bad features	7/12767 (0.1%)		< 7%	Pass	
Protocolo B					
	Cy5 (635)	Cy3 (532)	Threshold	Results	
				635	532
Median signal-to-background	15	15.3	> 10	Pass	Pass
Mean of median background	60.002	57.448	< 500	Pass	Pass
Median signal-to-noise	57.3	59.6	> 10	Pass	Pass
Feature variation	0.298	0.302	< 0.5	Pass	Pass
Background variation	0.301	0.428	< 0.5	Pass	Pass
Features with saturated píxel	0.22%	0.06%	< 1%	Pass	Pass
Not Found features	1148/12767 (9%)		< 10%	Pass	
Bad features	8/12767 (0.1%)		< 7%	Pass	

6.6.2. Validación técnica de la micromatriz CIT1

Los controles *Lucidea Universal ScoreCard Kit* (Amersham Biosciences), distribuidos en la micromatriz CIT1 (Tabla 3), permitieron estimar el rango dinámico, los niveles de sensibilidad y la reproducibilidad *intra-array*.

Tabla 3. Controles *Lucidea* distribuidos en la micromatriz CIT1

Tipo de control	Nombre	ID cDNA	mRNA en la muestra (pg/ μ l)		Cociente de expresión
			Test	Referencia	
Calibración	Cal2	cYIR02	5000	5000	1:1
Calibración	Cal3	cYIR03	1500	1500	1:1
Calibración	Cal4	cYIR04	500	500	1:1
Calibración	Cal6	cYIR05	50	50	1:1
Calibración	Cal7	cYIR07	15	15	1:1
Calibración	Cal8	cYIR08	5	5	1:1
Calibración	Cal9	cYIR09	1.5	1.5	1:1
Ratio	Ratio1	rYIR1	50	150	1:3 Bajo
Ratio	Ratio2	rYIR2	150	50	3:1 Bajo
Ratio	Ratio3	rYIR3	500	1500	1:3 Alto
Ratio	Ratio4	rYIR4	1500	500	3:1 Alto
Negativo	Negativo1	nYIR1	0	0	0

Estos controles son genes artificiales diseñados a partir de regiones intergénicas de levadura de tres tipos: 1) controles de calibración que permiten evaluar el rango dinámico y la sensibilidad del sistema; 2) controles de ratio que representan niveles de expresión bajos y altos; y 3) controles negativos que permiten evaluar los niveles de intensidad correspondientes a hibridación no específica. Los datos de expresión correspondientes a estos genes se obtuvieron añadiendo en cada reacción de marcaje una solución de mRNA (1 μ l) que contiene las secuencias complementarias a los cDNAs depositados en la micromatriz CIT1.

6.7. Identificación de los genes expresados diferencialmente

La identificación de los genes expresados diferencialmente en los experimentos de micromatrices se realizó mediante dos métodos según el número de réplicas realizadas en cada experimento: 1) el test de la t de Student en experimentos con tres réplicas biológicas y 2) el análisis SAM (*Significance Analysis of Microarrays*) (Tusher et al, 2001), en experimentos de dos clases de más de tres réplicas biológicas o en experimentos de múltiples clases con tres réplicas por clase.

6.7.1. Test de la t de Student

Para cada uno de los genes que componen la micromatriz CIT1 se calculó el estadístico t para comprobar la hipótesis nula 'no existen diferencias de expresión entre las condiciones comparadas' utilizando el programa Acuity (Molecular Devices). El estadístico t se define como $t = R_g / SE_g$, siendo R_g la media de los cocientes de expresión (expresados en \log_2) de las réplicas y SE_g el error estándar. Una vez calculado el estadístico t para cada uno de los genes, se convirtió en un valor de probabilidad p según la tabla de la distribución t de Student. Un inconveniente de este método es el de la estabilidad, es decir, que genes cuyo SE sea, por azar, pequeño obtendrán valores del estadístico t elevados, incluso cuando el valor de R_g sea bajo. Este método se utilizó para identificar los genes expresados diferencialmente en el análisis del transcriptoma de las plantas transgénicas que sobre-expresan el gen *CcGA20ox1*.

6.7.2. Análisis SAM

El análisis SAM constituye una versión modificada del test de la t de Student en el que se añade una constante en el denominador de modo que el estadístico se

define como $s = R_g / c + SE_g$, siendo c la constante. La adición de dicha constante reduce el problema de la estabilidad anteriormente señalado. En este método el estadístico s se convierte en un valor q mediante un análisis de permutaciones (Tusher et al, 2001) lo que exige un número suficiente de réplicas, al menos cuatro en el caso de comparaciones de dos clases. Este método se utilizó para identificar los genes expresados diferencialmente en el resto de los análisis transcriptómicos realizados en este trabajo.

6.7.3. El problema de las pruebas múltiples

Cada vez que se realiza un test estadístico se pueden cometer dos tipos de errores. El error de tipo I, o falso positivo, ocurre cuando se declara un gen como diferencialmente expresado cuando en realidad su expresión no es diferencial. El error de tipo II, o falso negativo, sucede cuando un gen que verdaderamente se expresa diferencialmente no se declara como tal. En un experimento de micromatrices se llevan a cabo miles de pruebas estadísticas (una por cada gen) con lo que se pueden acumular un número elevado de falsos positivos. Este problema se conoce como el problema de las pruebas múltiples. La corrección del valor p obtenido en el test t se realizó mediante el método de FDR (Benjamini y Hochberg, 1995) utilizando el programa Acuity (Molecular Devices). El FDR se define como la proporción esperada de falsos positivos entre los genes identificados como diferencialmente expresados. El valor q obtenido con el análisis SAM constituye una estimación del FDR. Así pues, se consideraron genes expresados diferencialmente aquéllos con un FDR inferior al 5%. Adicionalmente, se aplicó un límite en el valor de cambio de expresión de al menos 1.6 veces ($|\log_2(\text{Cy5/Cy3})| \geq 0.678$) de modo que el análisis fue más estricto.

6.8. Análisis de agrupamiento y análisis funcional de los datos

El análisis de agrupamiento jerárquico de los diferentes tejidos de cítricos, en función de los perfiles de expresión génica, se realizó en Acuity (Molecular Devices) con el método UPGMA (*Unweighted Pair-Group Method with Arithmetic Mean*) utilizando los parámetros de ligamiento completo y correlación de Pearson centrada aplicados a los valores de expresión normalizados con la función *Lowess* y expresados en \log_2 .

El conjunto de los unigenes expresados diferencialmente, identificados en cada uno de los experimentos de micromatrices realizados en este trabajo, se analizó funcionalmente para facilitar la interpretación biológica de los resultados obtenidos. El

procedimiento seguido fue similar al descrito en el apartado 4.3.2. Para cada unigen de cítricos expresado diferencialmente se obtuvo el código AGI correspondiente al gen de Arabidopsis más similar y, a través de dicho código, la anotación funcional del MIPS mediante la herramienta FunCat (Ruepp et al, 2004) y las categorías GO con el programa FatiGO (Al-Shahrour et al, 2004). Además, el programa FatiGO se utilizó en los análisis de enriquecimiento funcional, comparando cada subconjunto de genes expresado diferencialmente con el conjunto de los genes representado en la micromatriz CIT1. El análisis de las rutas metabólicas representadas en el subconjunto de los unigenes expresado diferencialmente se realizó, asimismo a través del código AGI correspondiente al gen de Arabidopsis más similar. Estos códigos se utilizaron para obtener las rutas metabólicas en las que participa cada gen, se extrajeron los códigos AGI de otros genes implicados en cada ruta metabólica identificada y se comprobó si dichos genes están incluidos en la micromatriz CIT1 y si se encuentran expresados diferencialmente.

7. Análisis de expresión de mRNA

7.1. Análisis Northern

7.1.1. Electroforesis en gel de agarosa desnaturizante

Las muestras de RNA total (10-20 µg), en un volumen de 4 µl de agua tratada con DEPC, se mezclaron con 10 µl de formamida desionizada, 3.5 µl de formaldehído 37%, 1 µl de tampón MOPS 20X (0.4 M MOPS pH 7, 0.16 M NaAc, 20 mM EDTA), 1 µl glicerol 80% y 0.3 µl de una solución saturada de azul de bromofenol. La mezcla se incubó 10 min a 65°C y se enfrió en hielo. Las muestras se cargaron en un gel desnaturizante de agarosa al 1.5% que contenía tampón MOPS 1X, formaldehído 1.2% y 10 µg de BrEt, por cada 100 ml de gel preparado. La electroforesis se realizó a un voltaje constante de 90 V, en tampón MOPS 1X, hasta que el azul de bromofenol recorrió aproximadamente $\frac{2}{3}$ del gel.

7.1.2. Transferencia de RNA a membranas de nylon

El gel se lavó durante 30 min en solución SSC 20X (3 M NaCl, 0.3 M citrato sódico). Sobre el gel se colocó un fragmento de membrana de nylon neutra (Hybond-N, Amersham Biosciences) del tamaño del gel y, sobre ella, cuatro papeles Whatmann 3MM, una capa de unos 5 cm de papel absorbente y un ligero peso. La

transferencia por capilaridad se llevó a cabo durante toda la noche. Al finalizar, el RNA se fijó a la membrana mediante irradiación con luz ultravioleta en un UV Stratalinker 1800 (Stratagene).

7.1.3. Obtención de las sondas de DNA

Las sondas de DNA se obtuvieron mediante reacciones de PCR a partir de 4 ng de DNA plasmídico utilizando 5 U de polimerasa *Expand High Fidelity* (Roche Applied Science), 200 μ M dNTPs, 400 nM de cebador T7 modificado (5'-CGACTCACT ATAGGGCGAATTGG-3') y 400 nM de cebador T3 modificado (5'-GGAAACAGCTAT GACCATGATTAC-3'), en un volumen final de 100 μ l. Las condiciones de las reacciones de PCR fueron de 94°C durante 3 min, seguido de 5 ciclos de amplificación (94°C durante 30 s, 66°C durante 30 s, 72°C durante 2 min), 30 ciclos más (94°C durante 30 s, 60°C durante 30 s, 72°C durante 2 min) e incubadas a 72°C durante 10 min. Los productos de PCR se analizaron en geles de agarosa 1%, obteniéndose bandas únicas del tamaño esperado, se purificaron utilizando el sistema *QIAquick PCR Purification Kit* (Qiagen) y se resuspendieron en agua hasta una concentración final de 200 ng/ μ l.

7.1.4. Hibridación con sondas radiactivas

El marcaje de las sondas de DNA con el isótopo radiactivo se realizó utilizando el sistema *Ready-To-Go DNA Labelling Beads (-dCTP)* (Amersham Biosciences), basado en el método de anillamiento de cebadores al azar. Alícuotas de 50 ng de DNA se diluyeron en 45 μ l de agua estéril y se incubaron a 95°C durante 3 min para su desnaturalización. La solución de DNA se añadió al tubo que contenía los componentes liofilizados necesarios para la síntesis de la sonda, al que también se añadió 5 μ l de [α^{32} P] dCTP 10 μ Ci/ μ l (Amersham Biosciences) y se incubó a 37°C durante 1 h. El DNA marcado se purificó mediante cromatografía de exclusión molecular en Sephadex G-50 (Amersham Biosciences). La sonda purificada se incubó a 95°C durante 3 min antes de añadirla a la solución de hibridación.

Las membranas se pre-hibridaron a 42°C durante 30 min con tampón de hibridación (0.25 M Na₂HPO₄ / Na₂PO₄ pH 7.2, 7% SDS, 1 mM EDTA, 8% PEG 6000, 0.2 M NaCl, 30% formamida) y se hibridaron a 42°C durante una noche con el tampón de hibridación conteniendo la sonda marcada y desnaturalizada. Tras la hibridación, las membranas se lavaron secuencialmente durante 15 min en SSC 3X, 0.5% SDS a 42°C y a 55°C, y en SSC 1X, 0.1% SDS a 42°C y a 55°C. A continuación se lavaron

durante 5-10 min en SSC 1X, 0.1% SDS a 60°C y a 65°C, respectivamente. Las membranas se expusieron a auto-radiografía con películas Hyperfilm (Amersham Biosciences) a -80°C en presencia de pantallas intensificadoras. Las películas se revelaron en un procesador automático M35 X-Omat (Kodak).

7.2. RT-PCR semicuantitativa

La síntesis de la primera cadena de cDNA se realizó utilizando el sistema *First Strand cDNA synthesis kit* (Amersham Biosciences). El RNA total tratado con DNasa (3 µg) (ver apartado 2.1.2) se llevó a 8 µl con agua tratada con DEPC, y se incubó a 65°C durante 10 min. Este RNA desnaturalizado se transfirió a otro tubo con 5 µl de la solución que contiene la transcriptasa reversa, 1 µl de DTT y 1 µl de oligonucleótido *Not I-d(T)₁₈* como cebador. La mezcla se incubó a 37°C durante 1 h. Alícuotas (1 µl) del heterodúplex RNA-DNA resultante se emplearon como molde en las reacciones posteriores de PCR, en presencia de 2.6 U de polimerasa *Expand High Fidelity* (Roche Applied Sciences), 200 µM dNTPs y 0.6 µM de cada cebador específico (Tabla 4), en un volumen final de 50 µl. El número de ciclos en el que se produce la amplificación exponencial se determinó, para cada gen de cítricos, realizando las reacciones de PCR a 20, 23, 26 y 29 ciclos (Tabla 4).

Tabla 4. Secuencia de los cebadores específicos utilizados en el análisis de RT-PCR semicuantitativa

Unigen de cítricos	Secuencia de los cebadores específicos	Nº ciclos PCR	Tamaño esperado (pb)
aCL8Contig9	S 5'-GGAGGACAGGACAAGATTGGA R 5'-ACATTCATGGCCACATCTCAA	21	400
aCL172Contig2	S 5'-GGCAGCCGACAAATCAGTAG R 5'-TTGGGAACCCAGACAATGTG	21	250
aCL48Contig1	S 5'-CATGATTGTGCAAGAGCAAGAG R 5'-CTAAAGCGTGTATTGGCAGCTT	21	250
aCL3307Contig1	S 5'-CGACCGGCTACTTTTCTTGA R 5'-TGGGCACATCATCCACATAG	20	450
aCL43Contig3	S 5'-CGGCCTCAAGTCTTCTTCTG R 5'-CTGCACTTGACGCTTGTTGT	20	450
aCL960Contig1	S 5'-ATTGTTGCTGCTGCTGCTTT R 5'-GCAACCGACCTTCCTTTTTC	26	400
aCL563Contig1	S 5'-CAGCCATCTCTCATCGGAAT R 5'-CCTGTGGACAATGGATGGAC	21	350

S, cebador sentido; R, cebador reverso

Los parámetros de las reacciones de PCR fueron de 94°C durante 3 min, seguido de un número variable de ciclos (dependiendo de cada gen, Tabla 4) de 94°C durante 30, 60°C durante 45 s y 72°C durante 45 s y, finalmente, 72°C durante 10 min. Alícuotas de los productos de PCR (20 µl) se separaron por electroforesis en geles de

agarosa al 1%, TAE 1X y 0.5 µg/ml BrEt. La cuantificación de los productos de PCR se realizó sobre la imagen del gel utilizando el programa GeneSnap (SynGene). El unigen de cítricos aCL563Contig1 que codifica una actina (CX289161) se utilizó como control interno.

8. Medidas de fotosíntesis

La tasa neta de fotosíntesis (PN, $\mu\text{mol CO}_2\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) se midió en hojas completamente expandidas de la brotación de otoño, seleccionadas al azar, de las plantas transgénicas de citrange Carrizo que sobre-expresan el gen *CcGA20ox1* y de las plantas control, utilizando un medidor portátil de fotosíntesis CIRAS-2 (PP Systems). La tasa neta de fotosíntesis se midió a una radiación fotosintética activa (PAR) de 600, 800 y 1000 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. El flujo de aire a través de la cámara fue de 195 ml/min y la temperatura del aire se mantuvo entre 25-27°C en todas las medidas realizadas. En estas mismas condiciones se midió la conductancia estomática (G_s , $\text{mmol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) y la transpiración (E , $\text{mmol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$). El análisis estadístico de los datos de PN, G_s y E se realizó utilizando el paquete estadístico SPSS (Norusis, 1993), mediante el análisis de la varianza (ANOVA) para determinar el efecto del genotipo y las interacciones entre el genotipo y el aporte de PAR.



*Resultados y
discusión I:
Análisis funcional de ESTs de cítricos*

1. GENES IMPLICADOS EN EL DESARROLLO VEGETATIVO EN CÍTRICOS

1.1. Obtención de una genoteca de tejidos vegetativos

El desarrollo vegetativo determina la forma y el comportamiento fisiológico general de la planta, de modo que tiene una influencia decisiva en la cantidad y calidad de la producción de fruta, así como en las prácticas culturales. Con objeto de obtener información sobre los genes implicados en el desarrollo vegetativo en cítricos se construyó una genoteca de cDNA, denominada Veg1, tal como se ha descrito en Materiales y Métodos, a partir de los siguientes tejidos vegetativos de la variedad mandarina Clemenules (*Citrus clementina* Hort. Ex Tan.) injertada sobre el patrón mandarino Cleopatra (*Citrus reshni* Hort. Ex Tan.): brotes jóvenes tratados con GA₃, brotes jóvenes tratados con paclobutrazol y hojas y entrenudos procedentes de brotes jóvenes sin tratar. Las titulaciones de las genotecas primaria y amplificada fueron de 2.3×10^6 y 1.0×10^{10} pfu/ml, respectivamente. El análisis por PCR de 1,152 clones procedentes de la genoteca amplificada Veg1 permitió estimar que la longitud media de los fragmentos de cDNA fue de 1,400 pb (Tabla Suplementaria III). En la distribución por tamaño de los insertos (Fig. 10) se observó que la longitud de la mayor parte de los mismos (63%) superó las 1,000 pb. El rango de tamaños observado, entre 500 y 4,000 pb, se corresponde con la distribución esperada para la mayoría de los mRNAs (Sambrook et al, 1989).

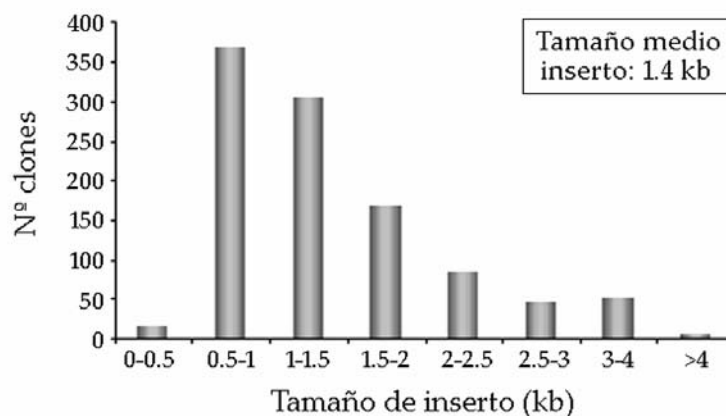


Fig. 10. Distribución por tamaños de los insertos de la genoteca Veg1.

Una genoteca de cDNA constituye una herramienta útil en genómica funcional pues permite identificar y caracterizar genes relevantes de un organismo. En este trabajo se ha construido una genoteca de cDNA a partir de diversos tejidos vegetativos de *Citrus clementina*. Según la fórmula de Clarke-Carbon, una genoteca de cDNA

debe contener teóricamente al menos 3.3×10^5 clones independientes para poder aislar un clon derivado de mRNA de baja abundancia con un 99% de probabilidad (Sambrook et al, 1989). Debido al limitado volumen y a la insuficiente estabilidad de las genotecas primarias es necesario amplificar la genoteca. La capacidad obtenida, tanto de la genoteca primaria como de la amplificada, 2.3×10^6 pfu/ml y 1.0×10^{10} pfu/ml, respectivamente, cumple los requerimientos para aislar un clon de cDNA derivado de mRNA de baja abundancia. El tamaño medio de los transcritos de plantas es de 1,500 pb (Alexandrov et al, 2006). Los resultados muestran que la mayor parte de los insertos de la genoteca Veg1 tiene un tamaño superior a 1,000 pb, con un tamaño medio de 1,400 pb. La genoteca Veg1 constituye pues una buena herramienta para posteriores análisis como la secuenciación masiva de ESTs (*expressed sequence tags*, o secuencias parciales de cDNA aisladas al azar a partir de una población de transcritos), el aislamiento de genes completos y el análisis de perfiles de expresión génica.

1.2. Colección de ESTs: secuenciación y agrupamiento

La generación de ESTs es una aproximación eficiente para identificar el conjunto de genes que se expresa en un organismo (por ejemplo, da Silva et al, 2005; Newcomb et al, 2006; Terol et al, 2007). Con la finalidad de identificar genes expresados durante el desarrollo vegetativo en cítricos se generó una colección de ESTs a partir de la genoteca amplificada de cDNA procedente de tejidos vegetativos de *Citrus clementina*. Un total de 1,824 clones independientes de cDNA se aislaron al azar a partir de la genoteca Veg1 y se secuenciaron desde el extremo 5'. Tras el pre-procesado de las secuencias, 1,689 ESTs de alta-calidad que, adicionalmente, superaron 100 pb de longitud, se utilizaron para posteriores análisis (Tabla 5, Tabla Suplementaria IV). La longitud de dichas ESTs, comprendida entre 101 y 810 bases, con una longitud media de 542 bases, fue relativamente superior a los valores obtenidos en otros proyectos de secuenciación de ESTs como el de *Populus* con 470 pb (Sterky et al, 2004) o el de manzano con 468 pb (Newcomb et al, 2006). La distribución por tamaños de las ESTs mostró que la mayoría de ellas (82%) tenían una longitud igual o superior a 400 bases (Tabla Suplementaria IV).

El número combinado de *contigs* y *singletons* que resulta de un proceso computacional de agrupamiento o ensamblaje de ESTs constituye una estimación del número de genes únicos de un organismo de modo que a medida que el número de ESTs crece, el número de genes únicos deberá ser redefinido. Los 1,824 clones de la genoteca Veg1 fueron agrupados inicialmente en 703 *contigs* y 719 *singletons* cuando

la colección de ESTs de cítricos del CFGP contaba con 22,635 ESTs (Forment et al, 2005). Dado que la colección de ESTs de cítricos del CFGP ha crecido considerablemente, en la actualidad contiene 85,965 ESTs agrupadas en 27,551 unigenes (11,844 *contigs* y 15,707 *singletons*), los resultados de este trabajo han sido redefinidos dado que consideramos que, pese al esfuerzo adicional que ello conlleva, suponen una mejor estimación del número de genes únicos. Así pues, las ESTs procedentes de la genoteca Veg1 se ensamblaron junto con el conjunto total de ESTs obtenidas en el CFGP (85,965 ESTs) de modo que del total de 1,689 ESTs de alta-calidad, 1,262 ESTs quedaron agrupadas en 971 *contigs* o conjuntos de secuencias formados por dos o más ESTs, mientras que las restantes 427 ESTs se clasificaron como *singletons*, pues no mostraron similitud con ninguna otra secuencia de la colección (Tabla 5, Tabla Suplementaria IV).

Tabla 5. Resumen de la colección de ESTs de la genoteca Veg1	
<i>Resultados de la secuenciación</i>	
Número de clones de cDNA aislados	1,824
Número de clones de cDNA secuenciados	1,771
Número de ESTs de alta-calidad	1,689
Longitud media de las ESTs	542 bases
<i>Resultados del agrupamiento</i>	
Número de ESTs ensambladas	1,689
Número de <i>contigs</i>	971
Número de <i>singletons</i>	427
Número de unigenes	1,398
Redundancia	17%
Unigenes únicos	439
Novedad	31%
La redundancia se ha estimado como $[1 - (\text{Unigenes}/\text{ESTs}) \times 100]$ y la novedad como $[\text{Unigenes únicos}/\text{Unigenes}] \times 100$	

La generación de una colección de ESTs constituye un método relativamente rápido y potente para la identificación de genes, para la confirmación de las regiones codificantes en secuencias genómicas, para el estudio de relaciones filogenéticas, para la estimación de los niveles de expresión de los genes de los que proceden y para el desarrollo de micromatrices de cDNA (Alba et al, 2004). Sin embargo, una colección de ESTs presenta ciertas limitaciones. Una de ellas es la generación de ESTs redundantes derivadas de los transcritos más comunes, lo que puede reducir la eficiencia de esta aproximación. Las 1,689 ESTs obtenidas de la genoteca Veg1 representaron 1,398 unigenes o posibles transcritos únicos (971 *contigs* y 427 *singletons*) de modo que la redundancia de esta genoteca fue del 17%. Adicionalmente, la genoteca Veg1 aportó un total de 439 unigenes únicos a la colección del CFGP (valor obtenido añadiendo al número de *singletons* el número de

contigs formados sólo por ESTs de la genoteca Veg1) lo que supone una novedad del 31%. Este valor es una estimación de la capacidad de la genoteca para aportar nuevos genes a la colección. El bajo nivel de redundancia obtenido, junto al porcentaje relativamente elevado de novedad, indican que el potencial de esta genoteca para aislar genes de cítricos no ha sido agotado en este trabajo, de modo que la genoteca Veg1 puede secuenciarse con más profundidad permitiendo la identificación de nuevos genes de *Citrus clementina*.

1.3. Anotación funcional de los unigenes

Muchos de los unigenes pueden ser asignados a una función en base a la similitud de secuencia con proteínas o genes, bien con una función demostrada o bien con posibles funciones basadas asimismo en la similitud de secuencia con otros genes conocidos. La asignación de una posible función a cada uno de los unigenes obtenidos a partir de la genoteca Veg1 se realizó comparando la secuencia consenso de cada unigen, mediante BLAST (Altschul et al, 1997), con las bases de datos de proteínas: TAIR pep y UniRef90 pep. Además se identificaron los motivos funcionales mediante una búsqueda HMMER utilizando la base de datos Pfam (ver Materiales y Métodos). En la Tabla 6 se muestra el número de unigenes con secuencias similares en las diferentes bases de datos según los distintos niveles de significatividad (valores E) considerados. La comparación con el conjunto completo de proteínas de Arabidopsis (TAIR pep) permitió asignar una posible función (basado en un valor $E < 10^{-5}$) al 83.1% (1162) de los unigenes (Tabla 6, Anejo I). Los resultados de las búsquedas por similitud de secuencia obtenidos en cada base de datos consultada, se recogen, para cada unigen de la genoteca Veg1, en la Tabla Suplementaria V.

Tabla 6. Número de unigenes con secuencias similares en las tres bases de datos consultadas a diferentes valores E

Base de Datos	Nº total de unigenes con secuencia similar				Sin secuencia similar
	$E < 10^{-5}$	$E < 10^{-20}$	$E < 10^{-50}$	$E < 10^{-100}$	
TAIR pep	1,162	1,044	722	321	236
UniRef90 pep	1,184	1,061	743	339	214
Pfam	619	510	281	281	779

El análisis de los resultados obtenidos en la base de datos UniRef90pep nos permitió estimar que de los 1,398 unigenes de la genoteca Veg1, un 72.5% (1,013 unigenes) presenta una similitud de secuencia significativa (valor $E < 10^{-5}$) con proteínas cuya función es conocida, un 12.2% (171 unigenes) con proteínas de función

desconocida y un 15.3% (214 unigenes) no tiene similitud de secuencia con ninguna proteína de la base de datos utilizada (Fig. 11, Tabla Suplementaria V).

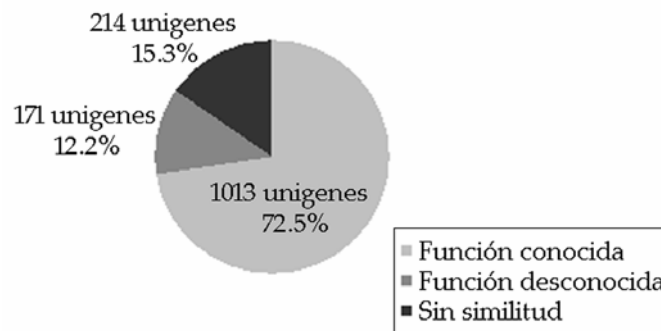


Fig. 11. Clasificación funcional general de los unigenes procedentes de la genoteca de cDNA de tejidos vegetativos según los resultados obtenidos en la base de datos UniRef90 pep.

Este último grupo probablemente incluye genes específicos de cítricos, aunque también puede incluir secuencias no codificantes (UTRs), secuencias contaminantes de DNA o artefactos. La proporción de genes específicos de cítricos se estimó comparando las secuencias de estos 214 unigenes, mediante BLASTX, con la base de datos PlantTA (Childs et al, 2007; <http://plantta.tigr.org>), excluyendo de ella los datos procedentes de los géneros *Citrus* y *Poncirus*. Los resultados de este análisis (Tabla Suplementaria V) indican que, de los 1,398 unigenes obtenidos de Veg1, 172 (12.3%) no muestran similitud significativa con ninguna secuencia de plantas, por lo que podrían ser considerados específicos de cítricos.

1.4. Ontologías génicas enriquecidas en Veg1

El principal objetivo de un experimento de escala genómica es obtener una interpretación funcional de los datos de modo que sea posible relacionar dicha información con la hipótesis que originó el experimento. La anotación funcional de los genes permite estructurarlos según su función. En este trabajo se utilizaron dos fuentes de información funcional para la anotación de los unigenes de cítricos procedentes de la genoteca Veg1: la Ontología Génica (GO; Ashburner et al, 2000) organizada en tres categorías generales ('proceso biológico', 'función molecular' y 'componente celular') y el Catálogo Funcional (FunCat; Ruepp et al, 2004) del MIPS (*Munich Information Centre for Protein Sequences*; <http://mips.gsf.de>). En primer lugar, para cada unigen de cítricos de la genoteca Veg1 se obtuvo el código AGI (*Arabidopsis Genome Initiative*, 2000) del gen de *Arabidopsis* más similar (Anejo I). Posteriormente, a cada uno de estos genes (1,162 unigenes) se le asignó una

categoría funcional MIPS utilizando la herramienta FunCat y las ontologías génicas mediante el programa FatiGO (Al-Shahrour et al, 2004). De este modo, un total de 981 unigenes fueron adscritos a una o más categorías funcionales según FunCat (Tabla Suplementaria VI). En el caso de la anotación funcional según GO, 598 unigenes se anotaron con una o más ontologías en la componente 'proceso biológico' (Tabla Suplementaria VI), 644 unigenes en 'función molecular' y 718 unigenes en 'componente celular'. En la Tabla 7 se muestra un resumen de la anotación funcional de los unigenes de Veg1 siguiendo el Catálogo Funcional del MIPS. La categoría que presentó un mayor número de unigenes fue la de 'localización subcelular' (23%), seguida de 'proteínas con función de unión' (14.77%), 'metabolismo' (10.38%), 'procesado de proteínas' (7.61%), 'transporte celular' (5.60%) e 'interacción con el ambiente' (4.70%).

Tabla 7. Clasificación funcional según FunCat de los unigenes de Veg1

CATEGORÍA FUNCIONAL	Citrus	
	Nº	%
01 Metabolismo	232	10.38%
02 Energía	48	2.15%
04 Proteínas de Reserva	2	0.09%
10 Ciclo Celular y Procesamiento del ADN	48	2.15%
11 Transcripción	76	3.40%
12 Síntesis de Proteínas	48	2.15%
14 Procesado de Proteínas (plegamiento, modificación, destino)	170	7.61%
16 Proteínas con Función de Unión o Necesidad de Cofactores	330	14.77%
18 Regulación del Metabolismo	30	1.34%
20 Transporte Celular	125	5.60%
30 Comunicación Celular - Transducción de Señal	53	2.37%
32 Defensa	98	4.39%
34 Interacción con el Ambiente	105	4.70%
36 Interacción Sistémica con el Ambiente	39	1.75%
38 Elementos transponibles, Proteínas Virales y Plasmídicas	1	0.04%
40 Procesado Celular	21	0.94%
41 Desarrollo (Sistémico)	37	1.66%
42 Biogénesis de Componentes Celulares	57	2.55%
43 Diferenciación de Tipos Celulares	7	0.31%
45 Diferenciación de Tejidos	3	0.13%
47 Diferenciación de Órganos	7	0.31%
70 Localización Subcelular	514	23.01%
73 Localización Celular	1	0.04%
75 Localización Tisular	1	0.04%
99 Proteínas sin Clasificación	181	8.10%

El análisis de la información funcional disponible para los genes analizados constituye un paso necesario, pero no suficiente, para la interpretación funcional de un experimento a escala genómica. De hecho, un error conceptual típico consiste en asignar importancia a las funciones más abundantes (Dopazo y Al-Shahrour, 2007). Es decir, los grupos funcionales más abundantes en el conjunto de los unigenes procedentes de la genoteca Veg1 pueden ser simplemente aquéllos más abundantes

de modo general en la planta. Con objeto de caracterizar la colección de unigenes obtenida en relación al desarrollo vegetativo, se realizó un análisis de enriquecimiento funcional mediante el cual se comparó el conjunto de unigenes obtenidos de la genoteca Veg1 con el genoma completo de Arabidopsis. La cuestión subyacente a este análisis es: ¿existe alguna función representada por una proporción de genes significativamente mayor en el subconjunto de unigenes procedentes de tejidos vegetativos de lo que cabría esperar en el genoma completo de Arabidopsis?. La herramienta FatiGO (Al-Shahrour et al, 2004) permite encontrar términos GO significativamente sobre-representados en un conjunto de genes respecto a un grupo de referencia. Los resultados obtenidos para la ontología 'proceso biológico' se recogen en la Figura 12 y en la Tabla Suplementaria VII.

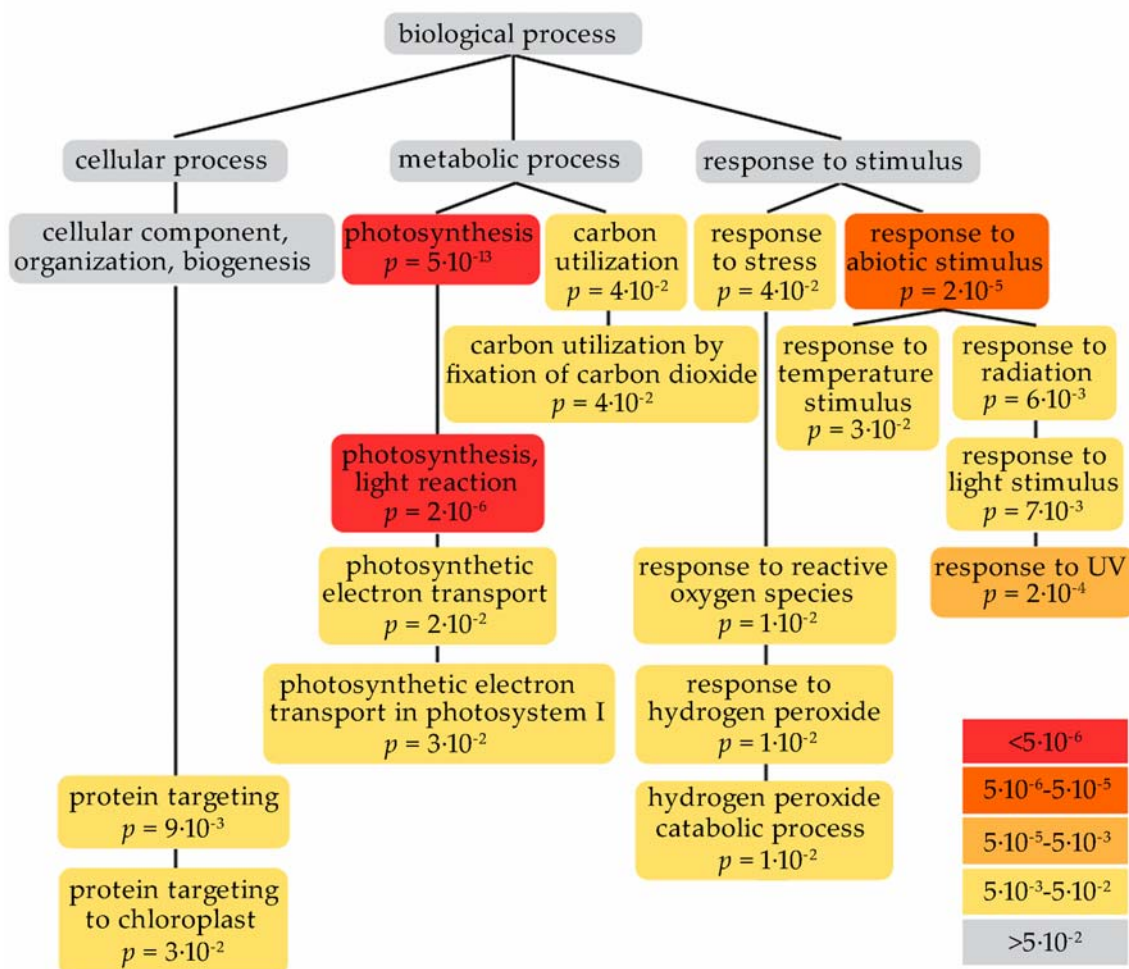


Fig. 12. Ontologías génicas de la componente 'proceso biológico' sobre-representadas en el conjunto de unigenes de la genoteca Veg1. La escala de color muestra el nivel de significatividad, considerando ontologías enriquecidas aquellas cuyo valor p corregido es inferior a 0.05.

El conjunto de unigenes de cítricos generados a partir de la genoteca Veg1 mostró una sobre-representación significativa (valor p corregido < 0.05) de aquellos unigenes pertenecientes a las ontologías génicas no redundantes de 'transporte fotosintético de electrones en el fotosistema I', 'utilización de carbono por fijación de CO_2 ', 'proteínas dirigidas al cloroplasto', 'catabolismo del peróxido de hidrógeno', 'respuesta a temperatura' y 'respuesta a ultravioleta' (Fig. 12, Tabla Suplementaria VII). Las dos primeras ontologías están relacionadas con una de las funciones fisiológicas clave del desarrollo vegetativo, la fotosíntesis, de modo que se analizaron en detalle y los resultados obtenidos se presentan en los siguientes apartados.

Durante la fotosíntesis se producen y se destruyen distintas especies reactivas del oxígeno de un modo controlado que contribuye a la regulación del transporte de electrones (Noctor y Foyer, 1998). La sobre-representación de unigenes relacionados con el 'catabolismo del peróxido de hidrógeno' en la genoteca Veg1 como catalasa, que convierte el peróxido de hidrógeno en oxígeno y agua, y monodehidroascorbato reductasa, que regenera ascorbato, indica la importancia de los sistemas de defensa frente a las especies de oxígeno reactivo en los tejidos fotosintéticos. Las ontologías génicas sobre-representadas de 'respuesta a temperatura' y 'respuesta a UV' están relacionadas con dos de los estímulos ambientales más relevantes para el desarrollo vegetativo, la temperatura y la luz (Fig. 12, Tabla Suplementaria VII). La genoteca Veg1 se construyó a partir de tejidos vegetativos procedentes de la brotación de otoño de árboles adultos de campo (apartado 1.1 de Materiales y Métodos). Es posible que la sobre-representación de dichas ontologías génicas esté relacionada con las condiciones climáticas. Los organismos fotosintéticos se aclimatan a las condiciones de luz del ambiente previniendo los posibles daños del exceso de irradiación lo que resulta coherente con la sobre-representación de la ontología de 'respuesta a UV'.

1.4.1. Fotosíntesis

La fotosíntesis oxigénica es el principal proceso por el que la luz solar se convierte en energía química y está catalizado por cuatro complejos multi-proteicos de membrana (Fig. 13): fotosistema I (PSI), fotosistema II (PSII), citocromo b/f y ATP sintasa. La fotosíntesis oxigénica requiere la actividad del PSII para la ruptura del agua. Los electrones generados en esta oxidación son transferidos a plastoquinona (PQ), unida al PSII, que se reduce a plastoquinol (PQH_2) en la membrana tilacoidal. El complejo citocromo b/f oxida el plastoquinol y reduce la plastocianina (PC) en el lumen tilacoidal. El PSI oxida la PC y reduce la ferredoxina (Fd) en el estroma. La reducción

del NADP^+ por la ferredoxina está catalizada por la enzima ferredoxina- NADP^+ reductasa (FNR). Durante este transporte de electrones del agua al NADP^+ , se produce la transferencia de protones del estroma al lumen en varios puntos: 1) la oxidación del agua libera H^+ en el lumen, 2) la reducción de PQ por el complejo citocromo b/f está acompañada por una translocación de protones del estroma al lumen. Este gradiente de protones al lumen es utilizado por ATP sintasa para generar ATP (Scheller et al, 2005).

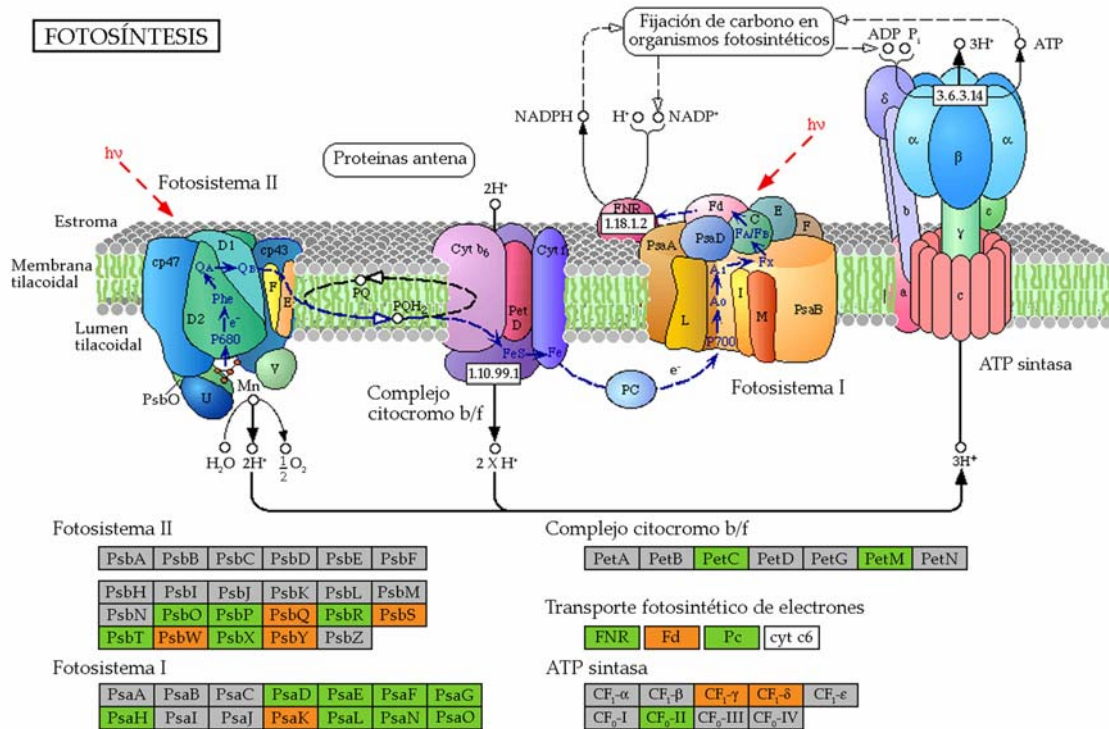


Fig. 13. Esquema del transporte de electrones en las membranas tilacoidales mostrando los cuatro complejos de membrana: fotosistema I y II, complejo citocromo b/f y ATP sintasa. En la parte inferior se indican las subunidades que forman cada uno de los complejos en plantas superiores; en gris las subunidades codificadas por genes cloroplásticos, en color las codificadas por genes nucleares, verde cuando se aislaron unigenes a partir de la genoteca Veg1 y naranja cuando se identificaron unigenes sólo de otras genotecas del CFGP.

Los cuatro componentes de la cadena de transporte electrónico están constituidos por múltiples subunidades que, en plantas superiores, están codificadas por genes cloroplásticos y por genes nucleares (Fig. 13). A partir del conjunto de ESTs del CFGP se obtuvieron los unigenes que codifican cada una de estas subunidades en cítricos. Además, se identificaron los unigenes de cítricos que codifican ferredoxina- NADP^+ reductasa (FNR), ferredoxina (Fd) y plastocianina (PC), proteínas implicadas en el transporte de electrones (Tabla 8, Tabla Suplementaria VIII). En la genoteca Veg1, preparada a partir de tejidos vegetativos (hojas y entrenudos procedentes de brotes vegetativos en desarrollo) se identificaron 45 ESTs (2.7%) que corresponden a

la mayoría de los genes que codifican proteínas implicadas en la cadena de transporte electrónico (Tabla 8).

Tabla 8. Unigenes de cítricos que codifican las subunidades proteicas de los componentes de la cadena de transporte electrónico

Proteína	Nº ESTs resto CFGP (‰)	Nº ESTs Veg1 (‰)	Enriquecimiento en Veg1	Significatividad
Fotosistema II				
PsbO	35 (0.42)	6 (3.55)	9	>0.999
PsbP	22 (0.20)	5 (2.96)	15	>0.999
PsbQ	15 (0.18)	–		
PsbR	17 (0.20)	3 (1.78)	9	>0.99
PsbS	13 (0.15)	–		
PsbT	10 (0.12)	2 (1.18)	10	>0.95
PsbW	20 (0.24)	–		
PsbX	9 (0.11)	5 (2.96)	28	>0.999
PsbY	12 (0.14)	–		
Fotosistema I				
PsaD	13 (0.15)	2 (1.18)	8	>0.95
PsaE	8 (0.09)	1 (0.59)		
PsaF	24 (0.28)	2 (1.18)	4	>0.95
PsaG	7 (0.08)	4 (2.37)	29	>0.999
PsaH	3 (0.04)	6 (3.55)	100	>0.999
PsaK	7 (0.08)	–		
PsaL	7 (0.08)	1 (0.59)		
PsaN	6 (0.07)	2 (1.18)	12	>0.99
PsaO	6 (0.07)	1 (0.59)		
Complejo citocromo b/f				
PetC	12 (0.14)	1 (0.59)		
PetM	68 (0.81)	–		
ATP sintasa				
AtpA	1 (0.01)	–		
AtpD	11 (0.13)	–		
AtpG	3 (0.04)	1 (0.59)	17	>0.95
Transporte fotosintético de electrones				
PetH (FNR)	8 (0.09)	1 (0.59)		
PetF (Fd)	32 (0.38)	–		
PetE (Pc)	16 (0.19)	2 (1.18)	6	>0.95

La colección de ESTs del CFGP se ha generado a partir de 52 genotecas de cDNA que cubren un amplio rango de tejidos, estadios de desarrollo y condiciones de estrés biótico y abiótico. Sin embargo, los datos de expresión obtenidos como la frecuencia de ESTs asignada a los genes que codifican proteínas de los complejos estudiados indican una expresión diferencial en los tejidos vegetativos (Tabla 8), lo que es consistente con su importancia en el transporte fotosintético de electrones.

1.4.2. Fotosíntesis: proteínas de la antena

Las proteínas Lhc (*light-harvesting chlorophyll a/b binding*) participan en la maximización y regulación de la captación de luz en los fotosistemas. Seis de las

proteínas Lhc (Lhcb1 a 6) están asociadas con el fotosistema II y las seis restantes (Lhca1 a 6) forman parte del fotosistema I (Fig. 14). Recientemente, se ha sugerido la existencia de un gen que codifica una nueva proteína Lhc (*Lhcb7*) (Klimmek et al, 2006).

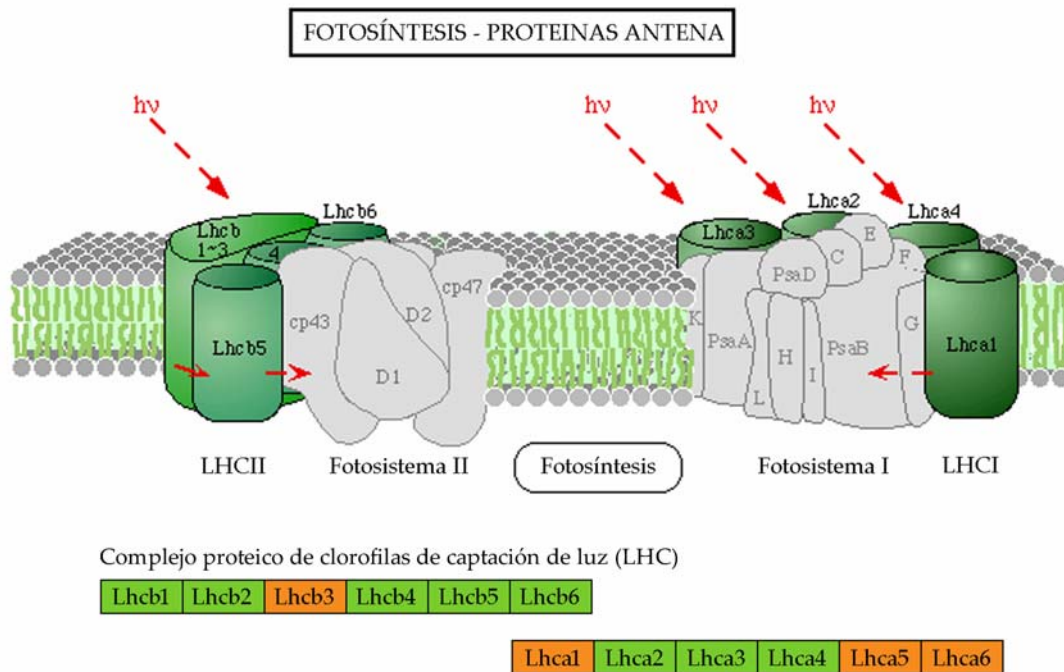


Fig. 14. Proteínas Lhc en cítricos. En cajas verdes se indican los tipos de proteínas Lhc para los que se aislaron unigenes a partir de la genoteca Veg1, y en cajas naranjas los tipos de proteínas Lhc para los que se identificaron unigenes sólo de otras genotecas del CFGP.

En la colección de ESTs generada a partir de la genoteca Veg1 se identificaron unigenes correspondientes a Lhcb1, -2, -4, -5 y -6 y Lhca2, -3 y -4 (Fig. 14). Adicionalmente, utilizando las ESTs de cítricos depositadas en la base de datos del CFGP (<http://bioinfo.ibmcp.upv.es/genomics/cfgpDB/>) se extrajeron todas las secuencias significativamente similares a los genes *Lhc* de *Arabidopsis*, de modo que en el conjunto de ESTs de cítricos del CFGP se encontraron representados los 12 tipos diferentes de proteínas Lhc: Lhcb1 a 6 y Lhca 1 a 6 (Fig. 14, Tabla Suplementaria IX). En cuanto a *Lhcb7*, en la colección de ESTs del CFGP, se identificaron tres ESTs de cítricos que corresponden a este gen. El alineamiento de la secuencia de aminoácidos deducida de *Lhcb7* de *Arabidopsis*, arroz, chopo y *Citrus* (Fig. 15) mostró que la secuencia de cítricos tiene conservados los motivos Lhc conocidos, incluyendo el motivo específico Lhcb7 y el motivo de trimerización presente en Lhcb1, Lhcb2 y Lhcb3 (Klimmek et al, 2006). La existencia de Lhcb7 no ha sido demostrada a nivel de proteína, pero el hecho de que el gen *Lhcb7* se transcribe y, por

tanto, está representado por ESTs de diferentes especies, entre ellas, cítricos, y que se encuentra conservado (Fig. 15) indica que codifica una proteína funcional.

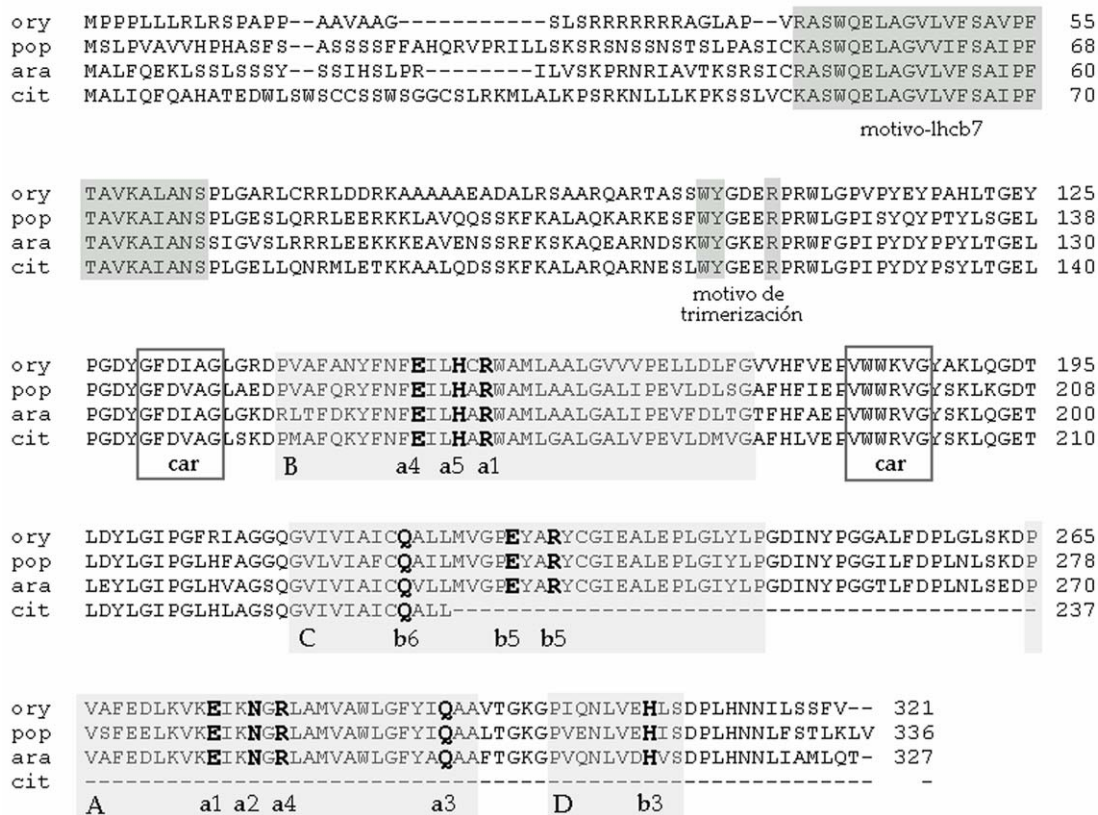


Fig. 15. Alineamiento de las secuencias deducidas de aminoácidos de Lchb7 de Arabidopsis (ara), chopo (pop), arroz (ory) y *Citrus* (cit) en las que se indican las hélices transmembrana (A-D), los sitios de unión de carotenoides (car), ligandos de clorofila (letras en negrita), el motivo específico Lhcb7 y el motivo de trimerización (de acuerdo con Klimmek et al, 2006).

Un total de 390 ESTs de cítricos correspondieron a genes que codifican los 12 tipos diferentes de proteínas Lhc, siendo Lhcb1 la más representada (136 ESTs). Los datos de expresión calculados según la frecuencia de ESTs asignada a los genes que codifican las 13 proteínas Lhc (Tabla 9) permitieron distinguir dos grupos. El primero contiene los genes que codifican proteínas Lhc abundantes: Lhca1 a 4 y Lhcb1 a 6 (excepto Lhcb3 para la que sólo se obtuvieron 4 ESTs). Una característica interesante de este grupo es que la mayoría de sus miembros (todos salvo Lhcb3 y Lhca1) aparecieron significativamente sobre-representados en la genoteca generada a partir de brotes vegetativos en desarrollo (Tabla 9; Tabla Suplementaria IX). El segundo grupo incluye los genes que codifican proteínas Lhc poco abundantes: Lhca5 (1 EST), Lhca6 (1 EST) y Lhcb7 (3 ESTs). Una característica común de estos genes fue su ausencia de expresión en brotes jóvenes (Tabla 9). Este patrón de expresión coincide

con el descrito en chopo (Klimmek et al, 2006) y parece indicar una regulación distinta de los dos grupos de genes *Lhc*.

Tabla 9. Unigenes de cítricos que codifican proteínas Lhc

Proteína	Nº ESTs resto CFGP (‰)	Nº ESTs Veg1 (‰)	Enriquecimiento en Veg1	Significatividad
Lhcb1	123 (1.46)	13 (7.70)	5	>0.999
Lhcb2	49 (0.58)	3 (1.78)	3	>0.95
Lhcb3	4 (0.05)	–		
Lhcb4	49 (0.58)	1 (0.59)		
Lhcb5	39 (0.46)	5 (2.96)	6	>0.999
Lhcb6	12 (0.14)	2 (1.18)	8	>0.99
Lhcb7	3 (0.04)	–		
Lhca1	17 (0.20)	–		
Lhca2	26 (0.31)	3 (1.78)	6	>0.99
Lhca3	26 (0.31)	3 (1.78)	6	>0.99
Lhca4	6 (0.07)	4 (2.37)	33	>0.999
Lhca5	1 (0.01)	–		
Lhca6	1 (0.01)	–		

Los resultados obtenidos muestran que los complejos de captación de luz (antena) de *Citrus* contienen los mismos tipos de proteínas que los de *Arabidopsis* y *Populus*. Además, los datos de expresión de los genes *Lhc* en *Citrus* son similares a los obtenidos en estas especies, mostrando que los genes *Lhca5*, *Lhca6* y *Lhcb7* presentan un patrón de expresión diferente al de los genes que codifican las proteínas Lhc abundantes. Es de interés señalar que las 5 ESTs de cítricos que corresponden a los genes *Lhca5*, *Lhca6* y *Lhcb7* proceden de genotecas preparadas a partir de hojas procedentes de plantas sometidas a diferentes condiciones de estrés: salino, hídrico, deficiencia férrica e infección por el virus de la tristeza de los cítricos. Aunque estos datos de expresión no permiten obtener una conclusión sobre la función específica de las proteínas codificadas, sugieren una posible función de las mismas en condiciones de estrés.

1.4.3. Fijación de carbono

Los cítricos pertenecen al grupo de plantas C₃ de modo que la fijación fotosintética de carbono se produce a través del ciclo de Calvin (Spiegel-Roy y Goldschmidt, 1996). A partir del conjunto de ESTs del CFGP se obtuvieron los unigenes que codifican cada una de las enzimas del ciclo de Calvin en cítricos. Un total de 529 ESTs de cítricos correspondieron a genes que codifican enzimas del ciclo de Calvin, siendo fructosa-bifosfato aldolasa la más representada (157 ESTs) (Tabla 10). En la genoteca Veg1 se identificaron 28 ESTs (1.6%) relacionadas con el ciclo de

Calvin. Sin embargo, RuBisCO (18 ESTs) y gliceraldehído-3-P deshidrogenasa (5 ESTs) fueron las más abundantes en este caso (Tabla 10).

Tabla 10. *Unigenes de cítricos que codifican las enzimas del ciclo de Calvin*

Enzima	EC	Nº Total ESTs (‰)	Nº ESTs Veg1 (‰)	Enriquecimiento en Veg1	Significatividad
RuBisCO	4.1.1.39	69 (0.80)	18 (10.66)	13	>0.999
Fosfoglicerato kinasa	2.7.2.3	50 (0.58)	1 (0.59)	1	
Gliceraldehído-3-P deshidrogenasa	1.2.1.13	32 (0.37)	5 (2.96)	8	>0.999
Triosa fosfato isomerasa	5.3.1.1	69 (0.80)	–		
Fructosa-bifosfato aldolasa	4.1.2.13	156 (1.81)	1 (0.59)		
Fructosa-1,6-bifosfatasa	3.1.3.11	16 (0.19)	1 (0.59)	3	
Transquetolasa	2.2.1.1	25 (0.29)	–		
Sedoheptulosa-1,7-bifosfatasa	3.1.3.37	2 (0.02)	–		
Ribosa 5-fosfato isomerasa	5.3.1.6	50 (0.58)	1 (0.59)	1	
Ribulosa-fosfato 3-epimerasa	5.1.3.1	11 (0.13)	–		
Fosforibuloquinasa	2.7.1.19	21 (0.24)	1 (0.59)	2	

La sobre-representación de ESTs (13 veces) correspondientes con RuBisCO observada en la genoteca Veg1 refleja el papel crucial que desempeña esta enzima en la bioquímica de los tejidos vegetativos.

2. GENES IMPLICADOS EN EL METABOLISMO DE ISOPRENOIDES Y FLAVONOIDES EN CÍTRICOS

Los cítricos constituyen una fuente importante de metabolitos derivados de isoprenoides, como carotenoides, limonoides y mono- y sesquiterpenos. Por otro lado, los cítricos acumulan grandes cantidades de flavonoides, tanto en los frutos como en las hojas que, además de ser responsables del aroma, son un componente importante de la dieta por sus propiedades antioxidantes, antiinflamatorias, antialérgicas y supresoras de mutagénesis (Manthey et al, 2001). En los últimos años, se han caracterizado algunos de los genes que participan en el metabolismo de isoprenoides y flavonoides en cítricos como los de biosíntesis de carotenoides (Kato et al, 2004) o algunos genes que codifican monoterpeneo (Lücker et al, 2002) o sesquiterpeno sintasas (Sharon-Asa et al, 2003). A pesar de ello, la mayor parte de las rutas de biosíntesis de metabolitos derivados de isoprenoides todavía no se han estudiado en cítricos. En este trabajo se han identificado los genes potencialmente implicados en estas rutas en cítricos, combinando búsquedas de similitud y motivos de secuencia con el conocimiento sobre el metabolismo de isoprenoides y flavonoides en otras especies. Los resultados obtenidos aportan un mayor conocimiento del metabolismo de isoprenoides y flavonoides en cítricos y mejoran substancialmente la anotación génica disponible actualmente. La identificación de estos genes constituye,

adicionalmente, una herramienta de base con la que abordar la fisiología, la bioquímica y la regulación genética del metabolismo de isoprenoides y flavonoides en cítricos.

2.1. Biosíntesis de isoprenoides

2.1.1. Biosíntesis de isopentenil difosfato y dimetilalil difosfato

Todos los isoprenoides derivan de un precursor común de cinco carbonos (C5), isopentenil difosfato (IPP), y de su isómero dimetilalil difosfato (DMAPP). Las plantas sintetizan estas unidades C5 mediante dos rutas localizadas en diferentes compartimentos celulares (Fig. 16). En el citosol, la ruta del mevalonato (MVA) produce IPP y DMAPP mientras que en los plastidios se sintetiza IPP y DMAPP por una vía independiente de la del MVA llamada ruta del metileritritol fosfato (MEP) (Rodríguez-Concepción y Boronat, 2002).

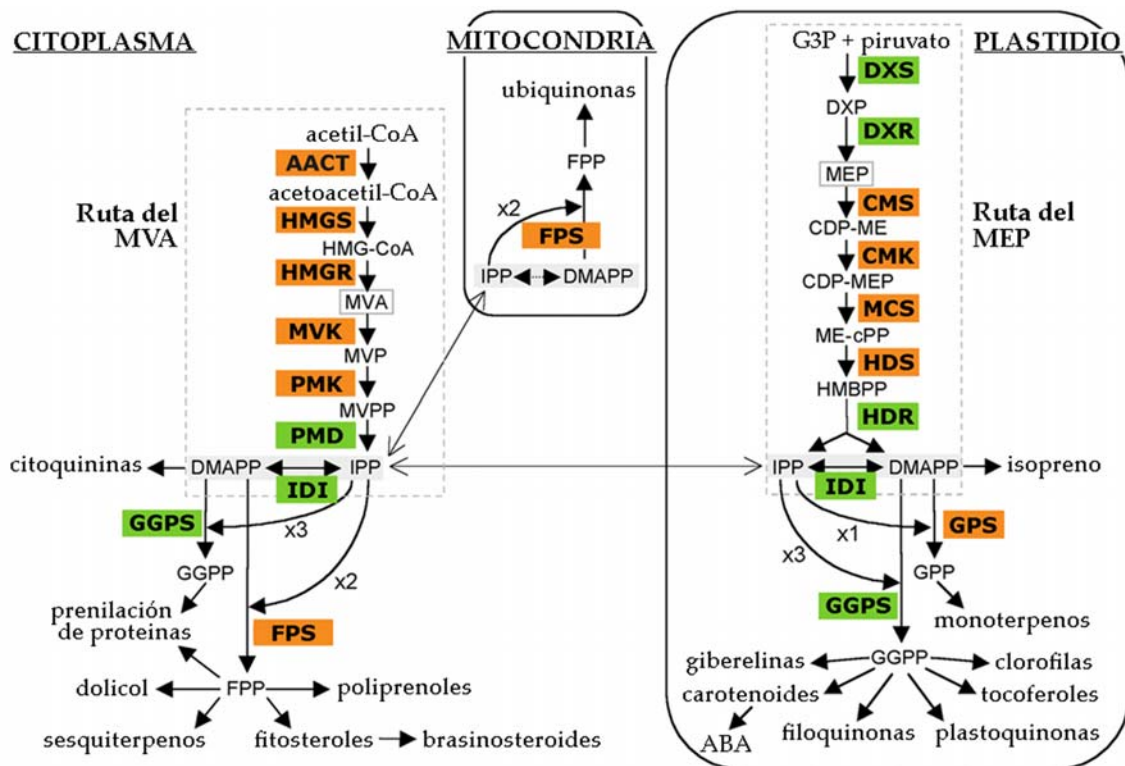


Fig. 16. Biosíntesis de IPP y DMAPP en las células vegetales mediante la ruta citosólica del MVA y la ruta plastidial del MEP. Además, se señalan los metabolitos derivados de la ruta del MVA, cuya síntesis se produce en el citosol, y los derivados del MEP que se sintetizan, al menos en parte, en el plastidio. Las cajas representan los pasos para los que se identificaron los unigenes correspondientes en el conjunto de ESTs del CFGP (en naranja) o para los que se encontró algún representante procedente de la genoteca Veg1 (en verde).

Biosíntesis de IPP y DMAPP en los plastidios: ruta del MEP

Las enzimas responsables de la biosíntesis de IPP y DMAPP a través de la ruta del MEP están localizadas en los plastidios (Fig. 16). En el primer paso, 1-deoxi-D-xilulosa 5-fosfato sintasa (DXS, EC 2.2.1.7) convierte piruvato y gliceraldehido-3-fosfato en 1-deoxi-D-xilulosa 5-fosfato (DXP). La colección de ESTs del CFGP contiene 5 unigenes con secuencias significativamente similares al gen *DXS* (Tabla 11, Tabla Suplementaria X). En el siguiente paso, DXP reductoisomerasa (DXR, EC 1.1.1.267; 7 unigenes en cítricos) cataliza la formación de metileritritol 4-fosfato (MEP) a partir de DXP. El MEP se convierte en IPP y DMAPP en pasos consecutivos catalizados por 4-difosfocitidil metileritritol sintasa (CMS, EC 2.7.7.60; 1 unigen), 4-difosfocitidil metileritritol kinasa (CMK, EC 2.7.1.148; 1 unigen), metileritritol 2,4-ciclodifosfato sintasa (MCS, EC 4.6.1.12; 3 unigenes), hidroximetilbutenil 4-difosfato sintasa (HDS, EC 1.17.4.3; 3 unigenes) y, finalmente, hidroximetilbutenil 4-difosfato reductasa (HDR, EC 1.17.1.2; 7 unigenes). En el conjunto de ESTs del CFGP fue posible identificar, por primera vez en cítricos, los unigenes que posiblemente codifican cada una de las enzimas de la ruta del MEP (Tabla 11; Tabla Suplementaria X).

Tabla 11. Unigenes de cítricos que codifican enzimas de biosíntesis de IPP y DMAPP

Enzimas	EC	Contigs	Singletons	Nº total ESTs
Ruta del metileritritol fosfato				
1-deoxi-D-xilulosa 5-fosfato sintasa (DXS)	2.2.1.7	2	3	31
1-deoxi-D-xilulosa 5-fosfato reductoisomerasa (DXR)	1.1.1.267	2	5	32
4-difosfocitidil metileritritol sintasa (CMS)	2.7.7.60	0	1	1
4-difosfocitidil metileritritol kinasa (CMK)	2.7.1.148	1	0	2
Metileritritol 2,4-ciclodifosfato sintasa (MCS)	4.6.1.12	2	1	14
Hidroximetilbutenil 4-difosfato sintasa (HDS)	1.17.4.3	2	1	31
Hidroximetilbutenil 4-difosfato reductasa (HDR)	1.17.1.2	5	2	54
Isopentenil difosfato d-isomerasa (IDI)	5.3.3.2	2	0	34
Ruta del mevalonato				
Acetoacetyl-CoA tiolasa (AACT)	2.3.1.9	3	6	19
Hidroximetilglutaril-CoA sintasa (HMGS)	2.3.3.10	3	3	28
Hidroximetilglutaril-CoA reductasa (HMGR)	1.1.1.34	3	0	13
Mevalonato kinasa (MVK)	2.7.1.36	2	1	6
Fosfomevalonato kinasa (PMK)	2.7.4.2	0	1	1
Mevalonato difosfato decarboxilasa (PMD)	4.1.1.33	1	0	6

Las ESTs más abundantes fueron las correspondientes a los genes que codifican HDR e IDI (1‰ de las ESTs de la colección) seguidos por DXR, DXS y HDS (1.1‰) (Tabla 11). Estas ESTs aparecieron ampliamente distribuidas en la colección de ESTs del CFGP (generada a partir diferentes tejidos y estadios de desarrollo), lo que sugiere que dichos genes se expresan en toda la planta. En la genoteca Veg1 se

identificaron ESTs correspondientes a los genes que codifican DXS, DXR, HDR e IDI (Fig. 16) lo que indica la importancia de esta ruta en tejidos vegetativos.

La mayor parte del trabajo sobre la caracterización de las enzimas de la ruta del MEP se ha llevado a cabo en *Escherichia coli*. Esta ruta constituye un buen ejemplo de cómo, usando información de secuencia, la bioinformática y la genómica comparada han hecho posible identificar de un modo relativamente rápido y simple los genes potencialmente implicados en una ruta metabólica en diferentes organismos (Rodríguez-Concepción y Boronat, 2002). La diferencia más obvia entre las enzimas de la ruta del MEP de plantas y *E. coli* es la presencia de extensiones N-terminales de secuencia y longitud variable en las primeras que funcionan como péptidos señal de plastidios (Rodríguez-Concepción y Boronat, 2002). El análisis de la posible localización subcelular (ver Materiales y Métodos) de las secuencias de aminoácidos deducidas de los unigenes de cítricos que integran la ruta del MEP mostró que todas las secuencias que contienen la región codificante N-terminal completa presentan un péptido señal de plastidio (Tabla 12, Tabla Suplementaria XI), lo que es consistente con su probable función en la biosíntesis plastidial de isoprenoides.

Tabla 12. Unigenes de cítricos implicados en la ruta MEP

Proteína	Unigen de Cítricos	Gen de Arabidopsis más similar	N-terminal completa	Longitud (cTP)
DXS	aCL303Contig1	At4g15560	Sí	716 (36)
DXR	aCL315Contig1	At5g62790	Sí	473 (50)
CMS	aC34104E04EF_c	At2g02500	No	-
CMK	aCL7203Contig1	At2g26930	No	-
MCS	aCL1267Contig1	At1g63970	Sí	237 (55)
HDS	aCL234Contig1	At5g60600	Sí	741 (48)
HDR	aCL2Contig14	At4g34350	Sí	464 (38)
IDI	aCL72Contig2	At3g02780	Sí	303 (51)

La longitud de la secuencia de aminoácidos deducida de los unigenes de cítricos así como la longitud del posible péptido señal de plastidio (cTP) se indica en número de aminoácidos.

La enzima IPP isomerasa (IDI, EC 5.3.3.2; 2 unigenes) cataliza la conversión reversible de IPP a DMAPP. Esta enzima se ha localizado en plantas tanto en el citosol como en los plastidios (Rodríguez-Concepción y Boronat, 2002); pero, el análisis de las secuencias de los genes de plantas que codifican IDI no permite establecer los dos grupos diferentes de secuencias correspondientes a enzimas citosólicas y plastidiales. Se ha propuesto que un sitio alternativo de iniciación de la traducción podría generar enzimas citosólicas a partir de los mismos genes que codifican las plastidiales (Cunningham y Gantt, 2000). El unigen de cítricos aCL72Contig2 presentó una elevada similitud de secuencia con IPP isomerasa (Fig. 17) y su secuencia deducida de aminoácidos mostró, en su región N-terminal, un

posible péptido señal de plastidio (Tabla 12, Fig. 17). En la Figura 17 se señalan además dos posibles sitios alternativos al anterior para la traducción, de modo que las proteínas de cítricos obtenidas a partir de estas dos metioninas mostraron una localización subcelular citosólica (ver Materiales y Métodos, Tabla Suplementaria XI) de acuerdo con la hipótesis de Cunningham y Gantt (2000).

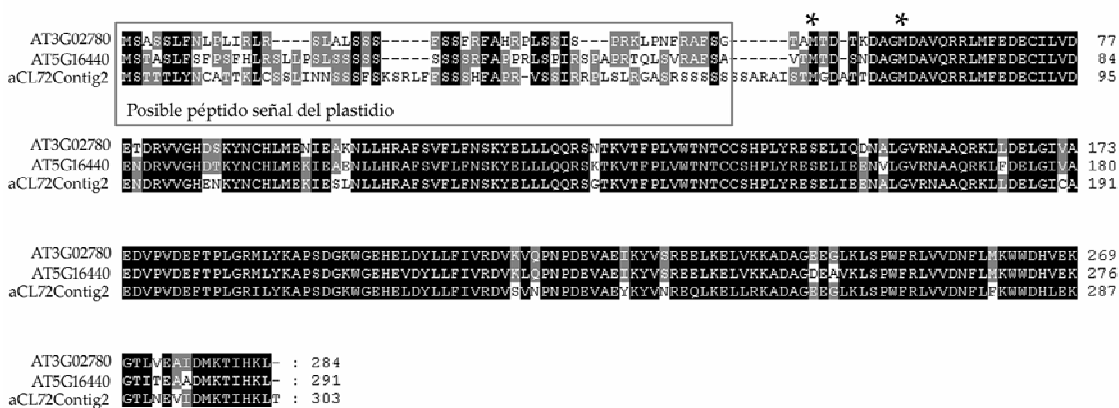


Fig. 17. Alineamiento de las secuencias deducidas de aminoácidos de IDI de Arabidopsis (At3g02780, At5g16440) y de *Citrus* (aCL72Contig2). El alineamiento se creó con ClustalX 1.83 y se editó en GeneDoc. La caja de la zona N-terminal señala el posible péptido señal de plastidio y las metioninas marcadas constituyen posibles sitios alternativos para la traducción.

Biosíntesis de IPP y DMAPP en el citosol: ruta del MVA

En el citosol se produce IPP a partir de acetil-CoA mediante la ruta del mevalonato (MVA). El IPP derivado de esta ruta puede ser transportado a la mitocondria para producir isoprenoides mitocondriales (Fig. 16). Todas las enzimas implicadas en la ruta del MVA se encontraron representadas en la colección de ESTs de cítricos del CFGP (Tabla 11, Tabla Suplementaria X). Los primeros pasos de la ruta conducen a MVA e incluyen las enzimas acetoacetil-CoA tiolasa (AACT, EC 2.3.1.9; 9 unigenes), hidroximetilglutaril-CoA sintasa (HMGS, EC 2.3.3.10; 6 unigenes) e hidroximetilglutaril-CoA reductasa (HMGR, EC 1.1.1.34; 3 unigenes). Dos fosforilaciones sucesivas catalizadas por mevalonato kinasa (MVK, EC 2.7.1.36; 3 unigenes) y fosfomevalonato kinasa (PMK, EC 2.7.4.2; 1 unigen) convierten el mevalonato en mevalonato 5-difosfato (MVPP). Finalmente, mevalonato difosfato descarboxilasa (PMD, EC 4.1.1.33; 1 unigen) cataliza la descarboxilación del MVPP para formar IPP (Fig. 16). Para cada unigen de cítricos potencialmente implicado en la ruta del MVA se realizó la predicción de la localización subcelular de su secuencia deducida de aminoácidos (ver Materiales y Métodos, Tabla Suplementaria XI). Este análisis permitió asignar una localización citosólica, coherente con su participación en la ruta del MVA, para AACT, HMGS, HMGR y MVK. Las secuencias de cítricos

correspondientes a PMK y PMD no contenían la región N-terminal completa por lo que no fue posible la predicción de su localización subcelular (Tabla Suplementaria XI).

Los resultados descritos muestran que las dos rutas de síntesis de IPP y DMAPP están completamente representadas por ESTs de cítricos y, que para la mayor parte de las enzimas implicadas en estas rutas se ha obtenido un cierto grado de redundancia, tanto en términos de ESTs (expresión) como en el número de *contigs* y *singletons* (Tabla 11). Esta redundancia sugiere que algunas de las enzimas de biosíntesis de IPP y DMAPP están codificadas por familias génicas en cítricos.

2.1.2. Prenildifosfatos: precursores de los distintos tipos de terpenoides

El DMAPP constituye la unidad inicial para las reacciones de condensación secuencial, catalizadas por preniltransferasas, que permiten la síntesis de una serie de prenildifosfatos, intermediarios en la síntesis de numerosos isoprenoides (Fig. 18). La adición de una molécula de IPP a DMAPP conduce a la formación de geranil difosfato (GPP), reacción catalizada por la enzima geranil difosfato sintasa (GPS, EC 2.5.1.1; 12 unigenes). La adición de una segunda molécula de IPP al GPP conduce a la formación de farnesil difosfato (FPP), reacción catalizada por farnesil difosfato sintasa (FPS, EC 2.5.1.10; 3 unigenes). La elongación de DMAPP con un total de tres moléculas de IPP forma geranilgeranil difosfato (GGPP) por la acción de la geranilgeranil difosfato sintasa (GGPS, EC 2.5.1.29; 8 unigenes). Los prenildifosfatos GPP, FPP y GGPP son intermediarios en la síntesis de monoterpenos (C10), sesquiterpenos (C15) y diterpenos (C20), respectivamente. La condensación de dos moléculas de FPP da lugar a escualeno (C30), reacción catalizada por escualeno sintasa (EC 2.5.1.21; 1 unigen). La síntesis de triterpenos (C30) como fitosteroles (incluidos los brasinosteroides) y limonoides, deriva de escualeno (Fig. 18). Los tetraterpenos (C40) como carotenoides y ABA se producen a partir de fitoeno, que se sintetiza en dos pasos consecutivos catalizados por fitoeno sintasa (EC 2.5.1.32; 4 unigenes) a partir de GGPP. En el conjunto de ESTs del CFGP fue posible identificar los unigenes de cítricos correspondientes a todas estas preniltransferasas (Tabla 13).

Tabla 13. Unigenes de cítricos que codifican preniltransferasas

Enzima	EC	Contigs	Singletons	Nº total ESTs	Localización subcelular
Geranil difosfato sintasa (GPS)	2.5.1.1	6	6	20	P
Farnesil difosfato sintasa (FPS)	2.5.1.10	1	2	4	C
Geranilgeranil difosfato sintasa (GGPS)	2.5.1.29	7	1	50	P
Escualeno sintasa (SQS)	2.5.1.21	0	1	1	-
Fitoeno sintasa (PYS)	2.5.1.32	2	2	7	P

Abreviaturas: P, plastidial; C, citoplasmática; -, región N-terminal incompleta

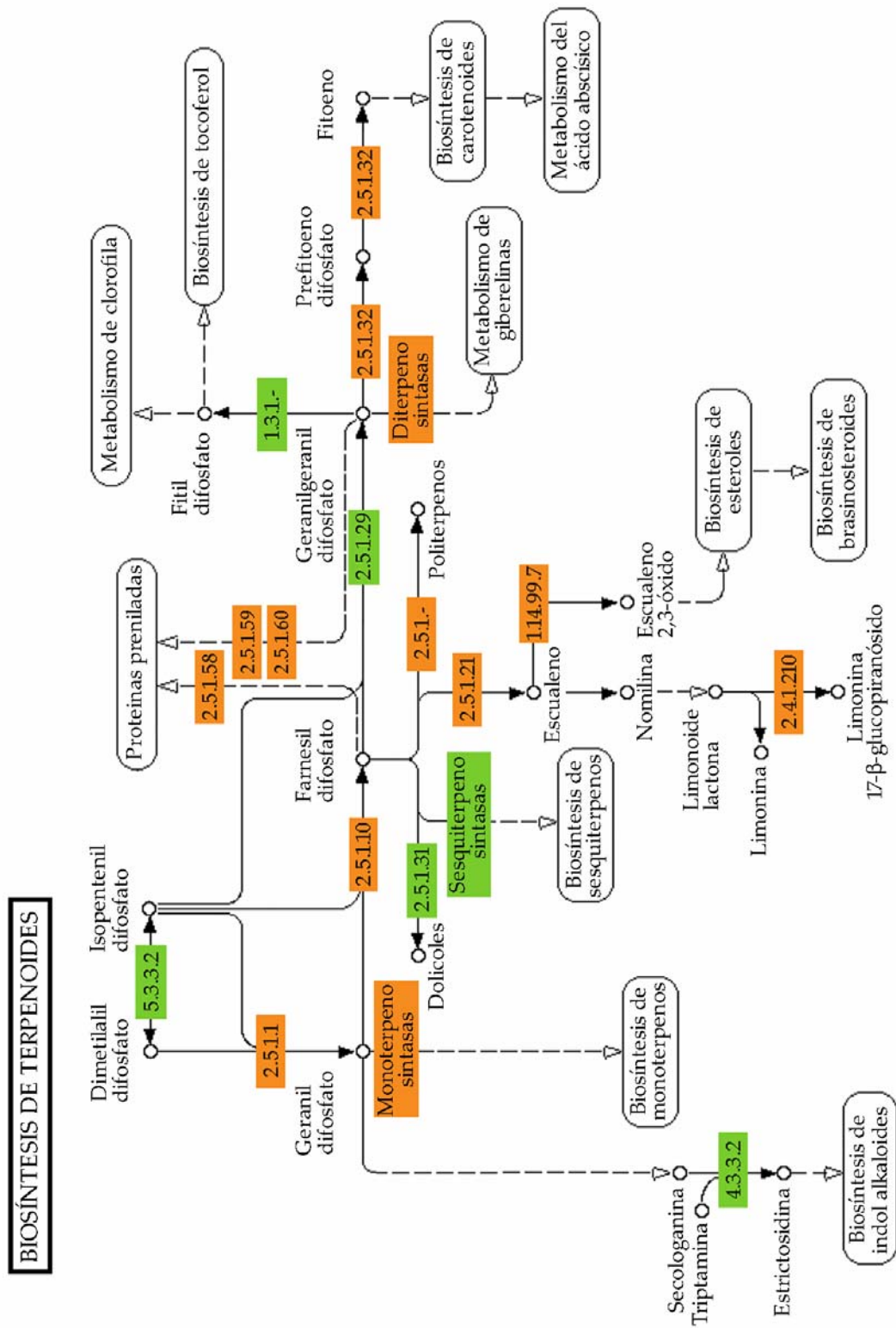


Fig. 18. Biosíntesis de los distintos tipos de terpenoides en cítricos. Las cajas representan los pasos para los que se identificaron los candidatos correspondientes en el conjunto de ESTs del CFGP (en naranja) o para los que se encontró algún representante procedente de la genoteca Veg1 (en verde).

Un total de 4 ESTs procedentes de la genoteca Veg1 mostraron similitud de secuencia con GGPS, mientras que para el resto de preniltransferasas no se identificaron ESTs en Veg1 (Tabla Suplementaria X). La predicción de la localización subcelular de la secuencia deducida de aminoácidos de los unigenes de cítricos identificados como preniltransferasas permitió asignar, tal como cabía esperar, una localización plastidial para GPS, GGPS y PYS, y citosólica para FPS (Tabla 13, Tabla Suplementaria XI). Las enzimas GPS, GGPS y FPS actúan en los puntos en los que el metabolismo de isoprenoides se ramifica (Fig. 18) por lo que pueden tener un papel clave en el control del flujo de IPP hacia la síntesis de diferentes familias de terpenoides (Lücker et al, 2007). En este sentido se ha demostrado que una reducción en la expresión de *GPS* en *Arabidopsis* y en tomate provoca enanismo en las plantas debido a un descenso en el contenido de giberelinas, mientras que los niveles de carotenoides y clorofila no se alteran (van Schie et al, 2007). Por otro lado, la sobreexpresión de *PYS* en tomate provoca un aumento del contenido de carotenoides en el fruto y un descenso en el contenido de clorofila (Fraser et al, 2007). Estos trabajos muestran que la modificación de la expresión de los genes que codifican preniltransferasas en plantas constituye una importante herramienta biotecnológica para la obtención de ciertos caracteres de interés como una altura reducida de la planta o el aumento en el contenido de carotenoides del fruto.

Además, se identificaron los unigenes de cítricos relacionados con la biosíntesis de diferentes tipos de terpenoides (Tabla 14, Fig. 18). La enzima estrictosidina sintasa (EC 4.3.3.2) está implicada en la biosíntesis de alcaloides derivados de terpenoides catalizando la condensación de triptamina y del monoterpeno secologanina para dar lugar al alcaloide estrictosidina (Ma et al, 2006). Un total de 34 ESTs de cítricos (dos de Veg1), agrupadas en 14 unigenes, presentaron secuencias similares a estrictosidina sintasa (Tabla 14, Tabla Suplementaria X). La enzima dehidrodolicol difosfato sintasa (EC 2.5.1.31) cataliza la síntesis de dehidrodolicol difosfato, prenildifosfato de cadena larga precursor de dolicol difosfato, implicado en la glucosilación de las proteínas (Cunillera et al, 2000). En cítricos se identificaron 18 ESTs (una de ellas procedente de Veg1), agrupadas en dos unigenes, similares a dehidrodolicol difosfato sintasa (Tabla 14, Tabla Suplementaria X). El solanesil difosfato es el precursor de las cadenas laterales de plastoquinona (implicada en el transporte fotosintético de electrones) y ubiquinona (implicada en las reacciones respiratorias mitocondriales). En *Arabidopsis*, la formación del politerpeno solanesil difosfato (C45), tanto a partir de GGPP como de FPP, está catalizada por la enzima solanesil difosfato sintasa (Hirooka et al, 2003). Dos ESTs de cítricos mostraron secuencias similares a solanesil difosfato sintasa (Tabla 14, Tabla Suplementaria X).

La prenilación de proteínas es una modificación post-traducciona que implica la unión covalente del grupo prenilo del FPP o del GGPP a las proteínas diana y que tiene un papel importante en rutas de transducción de señal y de tráfico intracelular (Crowell, 2000). La transferencia del grupo prenilo está catalizada por las enzimas farnesiltransferasas (EC 2.5.1.58) y geranilgeraniltransferasas I (EC 2.5.1.59) y II (EC 2.5.1.60). En la colección de ESTs generada en el CFGP se identificaron un total de 35 ESTs cuya secuencia es similar a este grupo de transferasas (Tabla 14, Tabla Suplementaria X).

Tabla 14. Unigenes de cítricos implicados en la biosíntesis de diversos terpenoides

Enzima	EC	Contigs	Singletons	Nº total ESTs
Terpenoides derivados de GPP				
Monoterpeno sintasas		Apartado 2.1.3		
Estrictosidina sintasa	4.3.3.2	6	8	34
Terpenoides derivados de FPP				
Dehidrodolicol difosfato sintasa	2.5.1.31	1	1	18
Sesquiterpeno sintasas		Apartado 2.1.3		
Biosíntesis de esteroides		Anejo II		
Biosíntesis de brassinosteroides		Apartado 2.1.4		
Biosíntesis de limonoides		Apartado 2.1.4		
Terpenoides derivados de FPP y GGPP				
Solanecil difosfato sintasa	2.5.1.-	1	0	2
Biosíntesis de plastoquinona		Anejo II		
Biosíntesis de ubiquinona		Anejo II		
Farnesiltransferasas	2.5.1.58	3	0	11
Geranilgeraniltransferasas I	2.5.1.59	1	0	2
Geranilgeraniltransferasas II	2.5.1.60	3	1	22
Terpenoides derivados de GGPP				
Biosíntesis de tocoferol		Anejo II		
Biosíntesis de carotenoides		Anejo II		
Metabolismo de ácido abscísico		Anejo II		
Metabolismo de giberelinas		Apartado 2.1.5		

En el Anejo II se representan las rutas de biosíntesis de tocoferol, biosíntesis de carotenoides, metabolismo de ácido abscísico y biosíntesis de plastoquinona y de ubiquinona, así como los unigenes identificados en este trabajo que codifican las enzimas que catalizan cada uno de los pasos de estas rutas en cítricos. Los resultados muestran que casi todos los pasos de estas rutas están representados por ESTs de cítricos y que, en general, el número de unigenes de cítricos que corresponde a enzimas implicadas en estas rutas es elevado. La posible presencia de familias génicas que codifican enzimas del metabolismo de isoprenoides facilita, probablemente, una ajustada regulación de la expresión de isoenzimas específicas en ciertos tejidos, estadios de desarrollo y/o condiciones ambientales. Es posible que la secuenciación adicional de ESTs de cítricos permita agrupar algunos *singletons* en *contigs*, reduciendo así el tamaño del conjunto de unigenes obtenido en el CFGP y,

por tanto, una parte de la redundancia génica observada en las rutas del metabolismo de isoprenoides. Sin embargo, recientemente se ha estudiado con detalle la ruta de biosíntesis de lignina en cítricos y los resultados muestran que algunas enzimas están codificadas por un mayor número de genes en cítricos que en *Arabidopsis* (Terol et al, 2007). Estos resultados apoyan la existencia de una 'expansión' de los genes implicados en las rutas derivadas del metabolismo de isoprenoides.

2.1.3. Biosíntesis de monoterpenos y sesquiterpenos: aceites esenciales

En cítricos, los aceites esenciales constituyen uno de los principales productos y se obtienen tanto de la piel del fruto mediante presión en frío como de las hojas por destilación en corriente de vapor. La composición de los aceites esenciales varía mucho entre especies, variedades e, incluso para un mismo cultivar, entre la piel del fruto y las hojas. Esta variabilidad parece deberse principalmente a la expresión de diferentes terpeno sintasas, enzimas clave en la biosíntesis de mono-, sesqui- y diterpenos (Fig. 18). En los últimos años se han aislado y caracterizado algunos genes que codifican terpeno sintasas del género *Citrus*, tanto monoterpeno sintasas (Lücker et al, 2002; Shimada et al, 2004) como sesquiterpeno sintasas (Maruyama et al, 2001; Sharon-Asa et al, 2003). Con objeto de identificar nuevos genes que codifican terpeno sintasas en cítricos, se utilizaron las secuencias previamente caracterizadas como motores de búsqueda en la base de datos del CFGP. Este método permitió identificar un total de 54 ESTs, agrupadas en 11 unigenes cuya secuencia es similar a monoterpeno sintasas, 12 unigenes similares a sesquiterpeno sintasas (Tabla 15, Tabla Suplementaria XII) y 5 unigenes similares a terpeno sintasas (Tabla Suplementaria XII).

Tabla 15. Posibles mono- y sesquiterpeno sintasas de cítricos

Posible función	Unigen	Nº ESTs	Posible función	Unigen	Nº ESTs
Monoterpeno sintasas			Sesquiterpeno sintasas		
Limoneno sintasa	aCL2450Contig1	3	β-Farneseno sintasa	aC01017A08	1
	aCL2450Contig2	6		aC06010C07	1
	aC02013A08	1		aCL9256Contig1	2
	aIC0AAA64DH02	1		aKN0AAB3AE04	1
γ-Terpineno sintasa	aCL3564Contig1	5	aKN0AAL1DG10	1	
β-Pineno sintasa	aKN0AAI2CC06	1	Valenceno sintasa	aCL7292Contig1	2
Linalol sintasa	aCL4530Contig1	4	aC06019E08	1	
	aCL4629Contig1	4	Germacreno D sintasa	aC32011H02	1
α-Terpineol sintasa	aKN0AAM2BC07	1	aKN0AAP6YP01	1	
	aIC0AAA51DF11	1	aC32102B11	1	
	aC06019H04	1	aCL6701Contig1	2	
			aCL4874Contig1	4	

Monoterpeno sintasas de cítricos

Las monoterpeno sintasas convierten el geranil difosfato (GPP) en diferentes monoterpenos (C10) en los plastidios, por lo que la región N-terminal de todas las monoterpeno sintasas caracterizadas en plantas contiene un péptido señal de plastidio. En este trabajo se identificaron un total de 28 ESTs, agrupadas en 11 unigenes (5 *contigs* y 6 *singletons*), que codifican posibles monoterpeno sintasas en cítricos (Tabla 15, Tabla Suplementaria XII). Cuatro unigenes de cítricos (11 ESTs) mostraron secuencias deducidas de aminoácidos muy similares a limoneno sintasa de *Citrus limon* (Lücker et al, 2002) y de *Citrus unshiu* (Shimada et al, 2004) (Fig. 19).

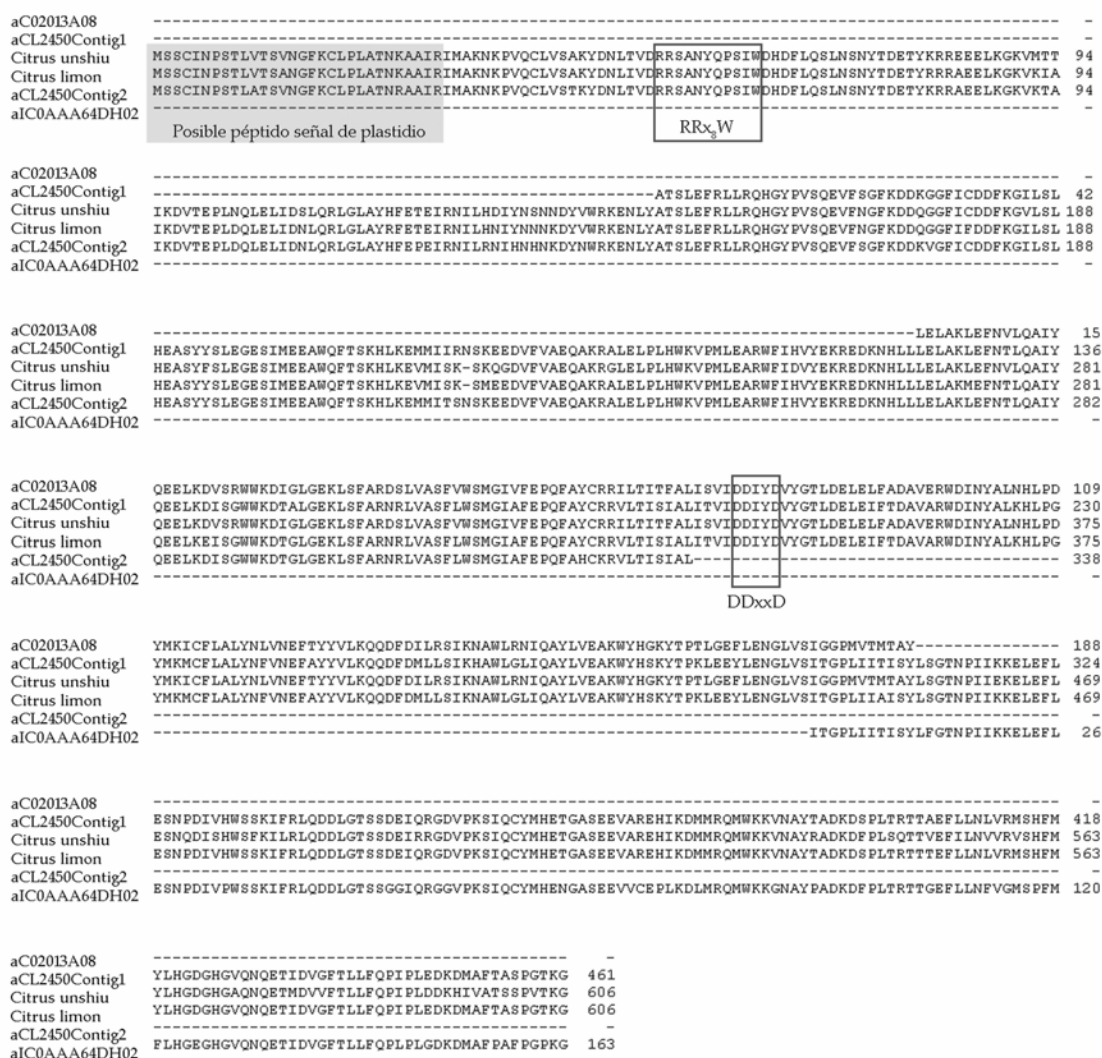


Fig. 19. Alineamiento de las secuencias de aminoácidos deducidas de los unigenes de cítricos identificados en este trabajo con las limoneno sintasas de *Citrus limon* (AF514289) y *Citrus unshiu* (AB110636). El alineamiento se creó con ClustalX 1.83 y se editó en GeneDoc. La caja sombreada señala el posible péptido señal de plastidio y las cajas huecas los motivos RR_xW y DDxD característicos de monoterpeno sintasas.

En la región N-terminal de la secuencia deducida de aminoácidos del unigen aCL2450Contig2 se detectó un posible péptido señal de plastidio de 30 aminoácidos aguas arriba del motivo RRx₈W conservado en monoterpene sintasas (Bohlmann et al, 1998). Aunque los unigenes aCL2450Contig1 y aC02013A08 carecían de la región N-terminal completa fue posible identificar el motivo DDxxD conservado en terpeno sintasas (Starks et al, 1997) (Fig. 19). La secuencia de aminoácidos deducida del unigen aCL3564Contig1 fue muy similar a γ -terpineno sintasa (97% de identidad, 99% de similitud) de *Citrus limon* (Lücker et al, 2002) y presentó un posible péptido señal de plastidio de 40 aminoácidos aguas arriba del motivo conservado RRx₈W. El unigen aKN0AAI2CC06 fue muy similar a β -pineno sintasa de *Citrus limon* (97% de identidad, 98% de similitud) (Lücker et al, 2002) y de *Citrus unshiu* (94% de identidad, 96% de similitud) (Shimada et al, 2004). La región N-terminal de la secuencia de aminoácidos de este unigen mostró un posible péptido señal de plastidio de 22 aminoácidos y el motivo RRx₈W.

Este trabajo permitió identificar, adicionalmente, unigenes cuyas secuencias son similares a monoterpene sintasas que todavía no se han caracterizado en cítricos (Tabla Suplementaria XII). La secuencia deducida de aminoácidos de los unigenes aCL4530Contig1 y aCL4629Contig1 fue similar a linalol sintasa de *Arabidopsis*, At1g61680 (Chen et al, 2003b) (59% de identidad, 75% de similitud) y At1g61120 (44% de identidad, 62% de similitud), respectivamente. Por otro lado, el unigen aLC0AAA51DF11 fue similar a α -terpineol sintasa (42% de identidad, 66% de similitud) de *Vitis vinifera* (Martin y Bohlmann, 2004).

Sesquiterpene sintasas de cítricos

Las sesquiterpene sintasas convierten el farnesil difosfato (FPP) en diferentes sesquiterpenos (C15) en el citosol, por lo que la región N-terminal carece de péptido señal de plastidio. En este trabajo se identificaron un total de 18 ESTs, agrupadas en 12 unigenes (4 *contigs* y 8 *singletons*), que codifican posibles sesquiterpene sintasas en cítricos (Tabla 15, Tabla Suplementaria XII). Cinco unigenes de cítricos fueron muy similares a β -farneseno sintasa de *Citrus junus* (Maruyama et al, 2001): aCL9256Contig1 (99% de identidad, 99% de similitud), aKN0AAB3AE04 (97% de identidad, 98% de similitud), aC01017A08 (95% de identidad, 98% de similitud) y aKN0AAL1DG10 y aC06010C07 (que mostraron menor grado de similitud) (Tabla Suplementaria XII). La secuencia deducida de aminoácidos del unigen aCL9256Contig1 presentó el motivo RRx₈W, a 21 aminoácidos del extremo N-terminal,

y el motivo DDxxD. La corta región N-terminal aguas arriba del motivo RRx₈W resultó indicadora de la ausencia de un péptido señal de plastidio, consistente con la localización citosólica de sesquiterpeno sintasas.

El unigen aCL7292Contig1 presentó una similitud elevada (99% de identidad, 99% de similitud) con valenceno sintasa de *Citrus sinensis* (Sharon-Asa et al, 2003). La región N-terminal de la secuencia deducida de aminoácidos del mismo mostró el motivo RRx₈W conservado en sesquiterpeno sintasas. La secuencia de aminoácidos del unigen aC06019E08 resultó idéntica a valenceno sintasa de *Citrus sinensis*.

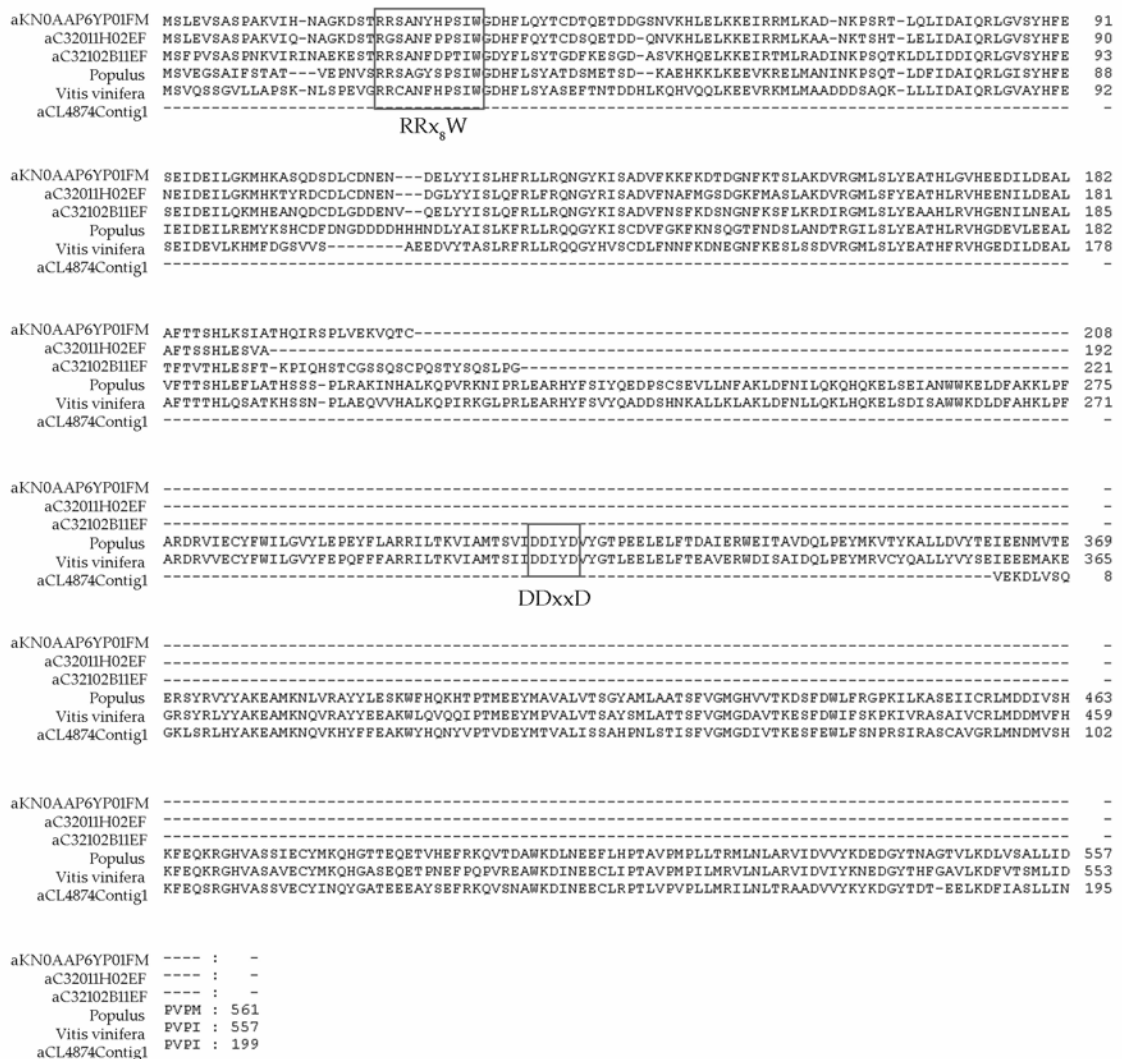


Fig. 20. Alineamiento de las secuencias de aminoácidos deducidas de los unigenes de cítricos identificados en este trabajo con germacreno D sintasa de *Populus* (AAR99061) y de *Vitis vinifera* (AAS66357). El alineamiento se creó con ClustalX 1.83 y se editó en GeneDoc. Las cajas huecas señalan los motivos RRx₈W y DDxxD característicos de sesquiterpeno sintasas.

En el análisis de los unigenes de cítricos que mostraron similitud de secuencia con diversas sesquiterpeno sintasas también fue posible identificar cinco unigenes con secuencias similares a germacreno sintasa (una sesquiterpeno sintasa) que, hasta el momento, no ha sido descrita ni caracterizada en cítricos (Tabla 15, Tabla Suplementaria XII). Las secuencias de los unigenes aC32011H02EF, aC32102B11, aKN0AAP6YP01FM y aCL4874Contig1 resultaron similares (Fig. 20) a germacreno D sintasa de chopo (Arimura et al, 2004) y de vid (Lücker et al, 2004). Los tres primeros presentaron el motivo RRx₈W conservado en sesquiterpeno sintasas y su región N-terminal no mostró péptidos señal de plastidio. El unigen aCL4874Contig1 fue similar a germacreno D sintasa (61% de identidad, 79% similitud) de vid (Fig. 20).

Otras terpeno sintasas de cítricos

Adicionalmente a las terpeno sintasas descritas en los apartados anteriores, en el conjunto de ESTs de cítricos del CFGP se identificaron 8 ESTs, agrupadas en 5 unigenes, que presentaron similitud con terpeno sintasas (Tabla Suplementaria XII). Los unigenes aCL4500Contig1 y aLC0AAA81DA02 mostraron similitud con terpeno sintasas de *Medicago truncatula* (Gómez et al, 2005) y los unigenes aLC0AAA1CC04 y aLC0AAA53DC12 fueron similares a terpeno sintasas de *Citrus junus* (AF288465). La secuencia deducida de aminoácidos de aLC0AAA1CC04, pese a contener la región N-terminal completa, no mostró péptido señal de plastidio por lo que podría clasificarse como sesquiterpeno sintasa.

Los análisis de homología de terpeno sintasas (TPS) muestran que, aunque el grado de conservación a nivel de secuencia no es elevado entre TPS de distintas especies vegetales, existen algunos dominios conservados que sugieren una similitud significativa en estructura y función (Steele et al, 1998). Estos dominios conservados han sido la base para aislar genes que codifican TPS en diversas especies vegetales utilizando RT-PCR con oligonucleótidos degenerados (por ejemplo Bohlmann et al, 1998; Sharon-Asa et al, 2003). Actualmente, la existencia de bases de datos de ESTs de un número creciente de especies vegetales facilita la búsqueda de genes que codifican TPS. Esta aproximación ha permitido identificar diversas TPS en vid (Martin y Bohlmann, 2004; Lücker et al, 2004). En este trabajo se ha utilizado la colección de ESTs de cítricos generada por el CFGP junto con la información de secuencia de diferentes TPS de otras especies vegetales para identificar un total de 28 unigenes de cítricos que codifican posibles TPS. La identificación de estos unigenes muestra que la familia de TPS en *Citrus* es una de las más amplias y diversas en plantas. Sin embargo, resulta necesario señalar que la función atribuida a estos unigenes es una

predicción que debe confirmarse mediante ensayos de actividad de las enzimas codificadas por los correspondientes unigenes.

La caracterización de los unigenes identificados en este trabajo permitirá abordar, con herramientas moleculares, la regulación de la formación del sabor y del aroma durante el desarrollo del fruto en cítricos. Estos genes pueden constituir, adicionalmente, marcadores moleculares para asistir la mejora de variedades de cítricos superiores en sabor, aroma y propiedades nutricionales. Finalmente, el aislamiento de genes que codifican terpeno sintasas en cítricos tiene interés en aplicaciones biotecnológicas como la producción de aromas en *E. coli* o en sistemas de fermentación de levaduras manipulados para la producción de terpenos (Martin et al, 2003; Reiling et al, 2004).

2.1.4. Biosíntesis de triterpenos: brasinosteroides y limonoides

Biosíntesis de brasinosteroides

Los brasinosteroides (BRs) son un grupo de hormonas vegetales que regulan el crecimiento y desarrollo de las plantas (Fujioka y Yokota, 2003). Los BRs pertenecen al grupo de triterpenos y constituyen un grupo de esteroides modificados, por lo que su biosíntesis puede dividirse en dos partes: 1) síntesis de campesterol (CR) a partir de escualeno y 2) síntesis de brasinólido (BL), BR biológicamente activo, a partir de CR (Fig. 21). La oxidación del escualeno, catalizada por escualeno monooxigenasa (EC 1.14.99.7; 4 unigenes), conduce a la formación de 2,3-oxidoescualeno (Fig. 18), que posteriormente es modificado para producir fitosteroles como sitosterol y campesterol. En el Anejo II se presenta la ruta de biosíntesis de esteroides en plantas así como los unigenes de cítricos identificados en este trabajo como candidatos a codificar las enzimas que catalizan cada uno de los pasos de dicha ruta.

La ruta específica de biosíntesis de BL comienza con la conversión de campesterol en campestanol mediante varios pasos de reducción del doble enlace del C5 (Fig. 21). Uno de estos pasos ha sido caracterizado en detalle en *Arabidopsis* y corresponde a una 5 α reductasa codificada por el gen *DET2* (At2g38050). La secuencia deducida de aminoácidos del unigen aC31805H09 de cítricos fue similar a *DET2* (64% de identidad, 81% de similitud) de *Arabidopsis* (Li et al, 1996). Los niveles endógenos de campestanol en el mutante *det2* de *Arabidopsis* son del 10% respecto al control (Fujioka et al, 1997) lo que sugiere la existencia de otras 5 α reductasas que catalizan la misma reacción (Clouse, 2002). En *Arabidopsis* se han identificado

homólogos funcionales a *DET2* como At3g55360 y At5g16010 (*AraCyc*), para los que también se encontraron unigenes similares en cítricos (Tabla 16). A partir de campestanol se sintetiza BL mediante dos rutas paralelas: la ruta de la oxidación temprana en C-6 y la ruta de la oxidación tardía en C-6 (Fig. 21).

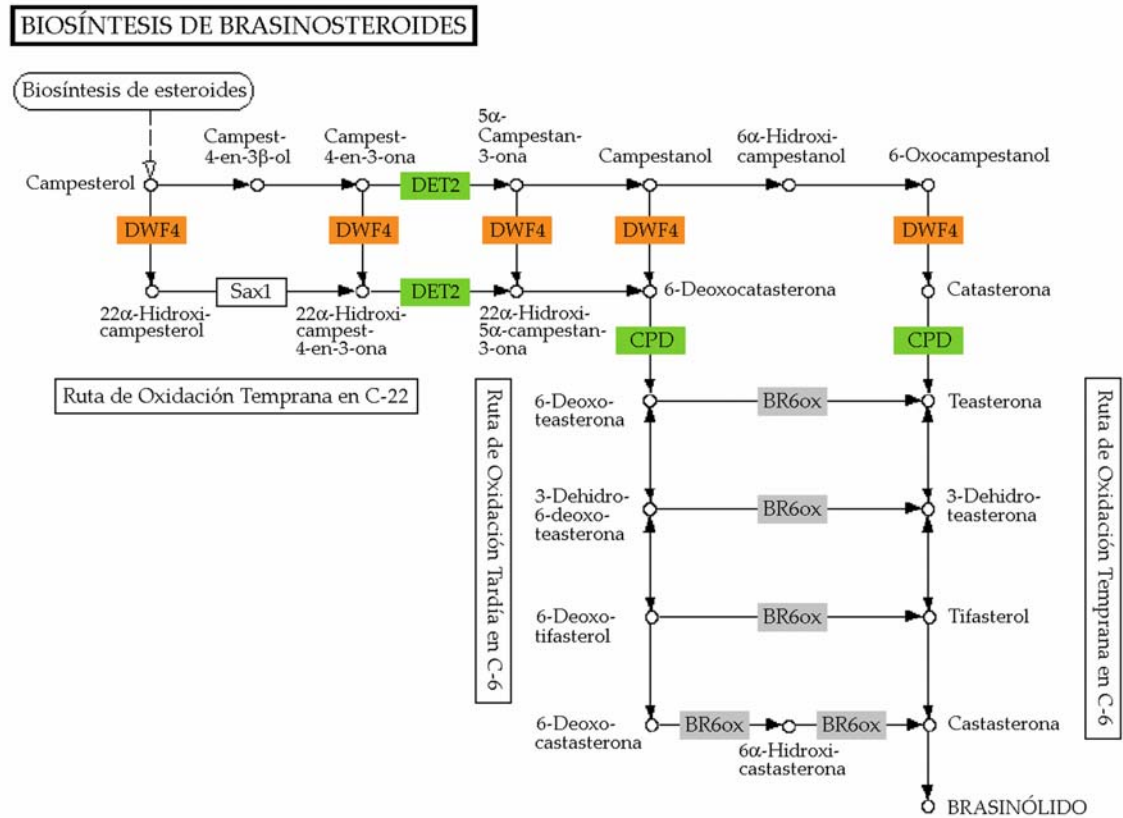


Fig. 21. Biosíntesis de brasinosteroides en cítricos. Las cajas señalan los pasos para los que se identificaron unigenes en el conjunto de ESTs del CFGP (en naranja), para los que se encontró algún representante procedente de la genoteca Veg1 (en verde) o para los que ninguna EST de cítricos mostró similitud significativa de secuencia (en gris). El gen *sax1* todavía no se ha clonado en ninguna especie vegetal.

La conversión de campestanol a 6-deoxocatasterona (ruta de la oxidación tardía en C-6) y de 6-oxocampestanol a catasterona (ruta de la oxidación temprana en C-6) está catalizada por una 22 α hidroxilasa, función codificada en *Arabidopsis* por el gen *DWF4* (At3g50660) (Choe et al, 1998). En la base de datos del CFGP se identificaron dos unigenes (*aCL2095Contig1* y *aCL2095Contig2*), cuyas secuencias de aminoácidos fueron similares (66% de identidad, 76% de similitud y 74% de identidad, 86% de similitud, respectivamente) a *DWF4* de *Arabidopsis* (Tabla 16). El siguiente paso, en ambas ramas de la ruta, implica la hidroxilación del C-23, que en *Arabidopsis* está catalizada por el producto del gen *CPD* (At5g05690) (Szekeres et al, 1996). Recientemente se han caracterizado dos nuevas C-23 hidroxilasas en *Arabidopsis*

(CYP90C1 y CYP90D1) codificadas por los genes At4g36380 y At3g13730, respectivamente (Ohnishi et al, 2006). Los unigenes de cítricos cuyas secuencias mostraron similitud con las C-23 hidroxilasas descritas en Arabidopsis se recogen en la Tabla 16.

Tabla 16. Unigenes de cítricos que codifican enzimas de la biosíntesis de BRs

Enzimas	Unigen de cítricos	Gen de ATH más similar	Valor E	Nº total ESTs
Esteroides 5 α reductasa (DET2)	aC31805H09	At2g38050	2E-62	1
	aCL5912Contig1	At3g55360	1E-145	3
	aC01011F03	At5g16010	7E-41	1
Esteroides C-22 α hidroxilasa (DWF4, CYP90B1)	aCL2095Contig1	At3g50660	0	4
	aCL2095Contig2	At3g50660	1E-154	6
Esteroides C-23 α hidroxilasa (CPD, CYP90A1)	aCL1811Contig1	At5g05690	0	8
	aCL1811Contig2	At5g05690	1E-100	2
	aC08025D05	At5g05690	7E-74	1
	(CYP90D1) aIC0AAA98BE11	At3g13730	2E-78	1
	(CYP90C1) –	At4g36380	–	–
Esteroides C-6 oxidasa (BR6ox)	–	At5g38970	–	–
	–	At3g30180	–	–

En tomate, Arabidopsis y arroz se han caracterizado las enzimas que catalizan la conversión de 6-deoxocastasterona a castasterona mediante C-6 oxidación (Bishop et al, 1999; Shimada et al, 2001; Hong et al, 2002) (Fig. 21). En Arabidopsis, existen en dos isoformas (codificadas por los genes At5g38970 y At3g30180) para la enzima esteroide C-6 oxidasa (Tabla 16). La colección de ESTs del CFGP contiene 85,965 ESTs generadas a partir de 52 genotecas de cDNA que cubren un amplio rango de tejidos, estadios de desarrollo y condiciones de estrés biótico y abiótico. Sin embargo, no se identificó ninguna EST de cítricos cuya secuencia fuera similar a esteroide C-6 oxidasa (Tabla 16). En Arabidopsis, guisante y tomate abundan los intermediarios 6-deoxo (Nomura et al, 2001), lo que indica que la ruta de la oxidación tardía en C-6 es la principal. No obstante, la ausencia de ESTs de cítricos con similitud de secuencia a esteroide C-6 oxidasa puede indicar que la ruta de la oxidación temprana en C-6 es la ruta de biosíntesis de BRs predominante en cítricos.

Los BRs constituyen un grupo de compuestos polihidroxilados derivados de esteroides que se han identificado en un amplio rango de especies vegetales (Bajguz y Tretyn, 2003). Sin embargo, hasta la fecha, los BRs no han sido estudiados en cítricos. En nuestro trabajo hemos identificado, por primera vez, los unigenes de cítricos que codifican esteroide 5 α reductasa, esteroide C-22 α hidroxilasa y esteroide C-23 α hidroxilasa (Tabla 16); aunque resulta necesario señalar que la función atribuida a

estos genes se basa en similitud de secuencia y debe confirmarse mediante ensayos de actividad de las correspondientes enzimas.

Biosíntesis de limonoides en cítricos

Los limonoides constituyen un grupo de triterpenos altamente oxigenados presentes sólo en el orden *Rutales*, principalmente en las familias *Rutaceae* (a la que pertenece el género *Citrus*) y *Meliaceae* (Moriguchi et al, 2003). En cítricos se dan dos formas químicas: agliconas (amargas) y glucósidos (no amargos). El precursor inicial de los limonoides conocidos en cítricos es la nomilina, que se sintetiza en el floema a partir de escualeno (Fig. 22). Este precursor migra a las hojas, frutos y semillas donde se sintetizan diversos limonoides, fundamentalmente obacunona y limonina. Estas agliconas amargas se convierten en 17- β -D-glucopiranosído derivados no amargos, como limonina 17- β -D-glucopiranosído, el limonoide glucósido más abundante en cítricos, durante la maduración del fruto (Moriguchi et al, 2003).

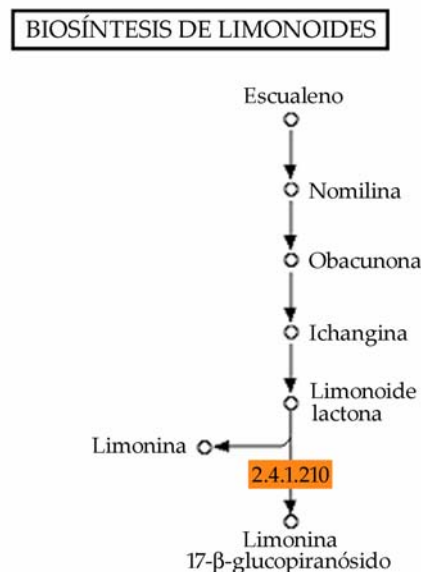


Fig. 22. Biosíntesis de limonoides en cítricos. El gen que codifica limonoide UDP-glucosiltransferasa (EC 2.4.1.210) está representado en la colección de ESTs del CFGP.

Este proceso natural de pérdida de amargor está catalizado por la enzima UDP-D-glucosa:limonoide glucosiltransferasa (LGT, EC 2.4.1.210). El gen que codifica esta enzima ha sido clonado y caracterizado a partir de frutos de *Citrus unshiu* (Kita et al, 2000). El unigen aCL770Contig1 del CFGP muestra una similitud de secuencia elevada (97% de identidad, 98% de similitud) con este gen. Este unigen está formado por 17 ESTs, ninguna de ellas procedente de genotecas generadas a partir de tejidos

vegetativos, lo que es consistente con que el sitio de metabolización de limonoides agliconas a glucósidos está limitado a la semilla y al fruto maduro (Kita et al, 2000).

2.1.5. Biosíntesis de diterpenos: giberelinas

Las giberelinas (GAs) constituyen un grupo de hormonas vegetales que controlan diversos aspectos del desarrollo como la germinación, la expansión de las hojas, la elongación del tallo y de la raíz, el tiempo de floración y el desarrollo de la flor y del fruto (Sponsel y Hedden, 2004; Swain y Singh, 2005). Las GAs forman un grupo de diterpenos tetracíclicos que se sintetizan a partir del precursor común geranilgeranil difosfato (Fig. 23).

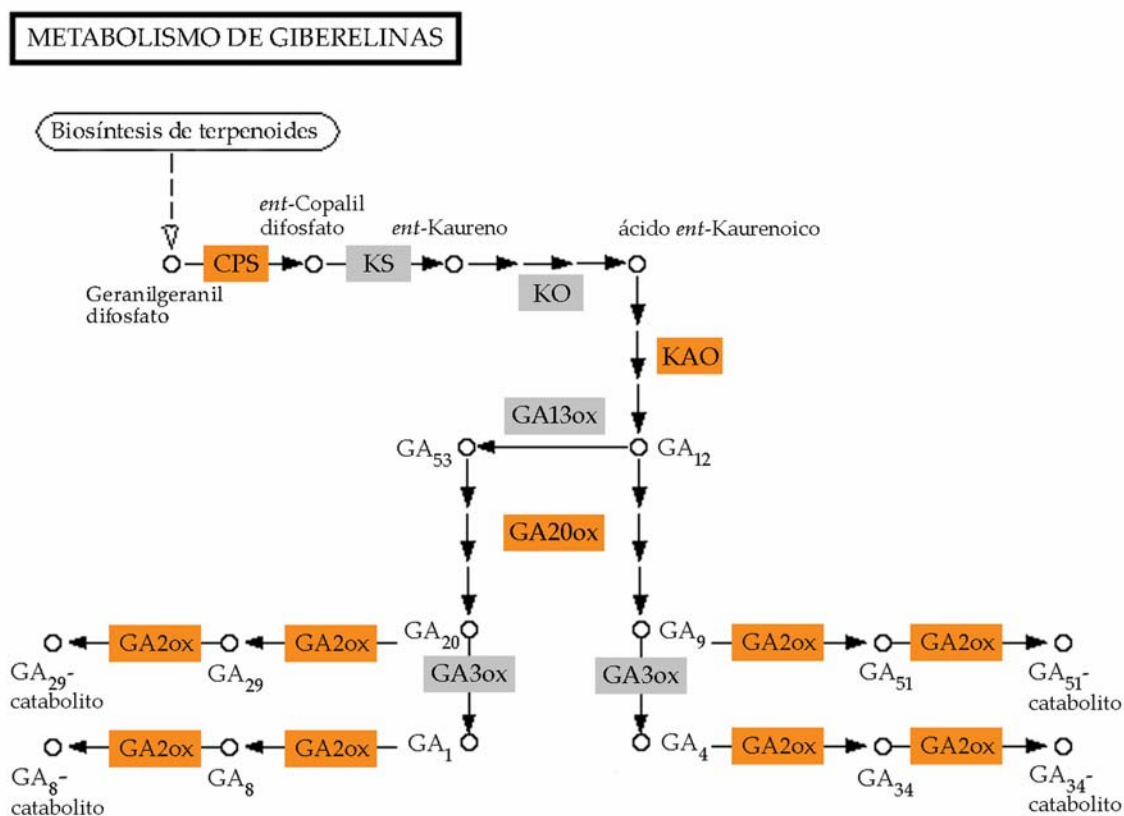


Fig. 23. Metabolismo de giberelinas en cítricos. Las cajas naranjas señalan los pasos para los que se identificaron unigenes en el conjunto de ESTs del CFGP y las grises los pasos para los que ninguna ESTs de cítricos mostró similitud significativa de secuencia.

Los dos primeros pasos de la biosíntesis de GAs están catalizados por dos diterpeno sintasas, copalil difosfato sintasa (CPS, EC 5.5.1.13) y *ent*-kaureno sintasa (KS, EC 4.2.3.19) y conducen a la formación de *ent*-kaureno. A continuación, dos monooxigenasas dependientes de citocromo P450, *ent*-kaureno oxidasa (KO, EC 1.14.13.78) y ácido *ent*-kaurenoico oxidasa (KAO, EC 1.14.13.79), catalizan la

conversión de *ent*-kaureno en GA₁₂ y GA₅₃. Finalmente, dos dioxigenasas dependientes de 2-oxoglutarato, GA 20-oxidasas (GA20ox, EC 1.14.11.12) y GA 3-oxidasas (GA3ox, EC 1.14.11.15), actúan consecutivamente para producir las GAs bioactivas (GA₁ y GA₄) a través de dos rutas paralelas: la ruta de la hidroxilación temprana en C-13 (que conduce a la formación de GA₁) y la ruta de la no hidroxilación en C-13 (mediante la que se sintetiza GA₄). Una tercera clase de dioxigenasas, GA 2-oxidasas (GA2ox, EC 1.14.11.13), inactivan tanto las GAs bioactivas (GA₁ y GA₄) como sus precursores inmediatos (GA₂₀ y GA₉) (Fig. 23).

En el conjunto de ESTs generado a partir de la genoteca Veg1 de tejidos vegetativos no se identificó ninguna EST relacionada con el metabolismo de GAs. Sin embargo, en el conjunto de ESTs del CFGP se identificaron un total de 11 ESTs, agrupadas en 9 unigenes, que codifican enzimas de cítricos potencialmente implicadas en la biosíntesis de giberelinas. Dos unigenes de cítricos (aC08014B02 y aC18022B11) mostraron similitud de secuencia con CPS, cuatro unigenes (aC08012B11, aKN0AAL1BE10, aKN0AAP12YO22, aCL8938Contig1) con KAO, un unigen (aCL8642Contig1) con GA20ox y dos unigenes con GA2ox (aC08003D01 y aC18016F11) (Tabla 17).

Tabla 17. Unigenes de cítricos que codifican enzimas del metabolismo de giberelinas

Enzimas	Unigen de cítricos	Gen de ATH más similar	Valor E	Nº total ESTs
Copalil difosfato sintasa (CPS)	aC08014B02	At4g02780	1E-39	1
	aC18022B11	At4g02780	5E-27	1
Ácido <i>ent</i> -kaurenoico oxidasa (KAO)	aC08012B11	At1g05160	4E-56	1
	aKN0AAL1BE10	At2g32440	5E-73	1
	aKN0AAP12YO22	At2g32440	2E-48	1
	aCL8938Contig1	At2g32440	1E-119	2
Giberelina 20-oxidasa (GA20ox)	aCL8642Contig1	At4g25420	1E-148	2
Giberelina 2-oxidasa (GA2ox)	aC08003D01	At1g30040	6E-49	1
	aC18016F11	At1g30040	6E-13	1

La colección de ESTs del CFGP consta de 85,965 ESTs generadas a partir de 52 genotecas de cDNA que cubren un amplio rango de tejidos, estadios de desarrollo y condiciones de estrés biótico y abiótico. Sin embargo, la proporción de ESTs relacionadas con el metabolismo de GAs es realmente baja (0.12 %) lo que indica que el nivel de expresión de estos genes en el amplio rango de condiciones estudiadas es bajo. Aunque el número de ESTs relacionadas con la biosíntesis de GAs no es suficiente para analizar la expresión génica según la frecuencia de ESTs, resulta interesante destacar que 8 de las 11 ESTs (73%) identificadas proceden de tres genotecas de cDNA generadas a partir de las raíces lo que puede indicar que, en cítricos, las raíces son órganos fundamentales en la síntesis de giberelinas.

2.2. Biosíntesis de flavonoides

Los cítricos acumulan grandes cantidades de flavonoides, tanto en los frutos como en las hojas, especialmente flavanonas. La biosíntesis de flavanonas se produce en dos pasos: el primero, catalizado por chalcona sintasa (CHS, EC 2.3.1.74; 10 unigenes), consiste en la condensación de 4-cumaroil-CoA con tres moléculas de malonil-CoA para formar naringenina-chalcona, y el segundo, mediado por chalcona isomerasa (CHI, EC 5.5.1.6; 5 unigenes), conduce a la formación de naringenina (Fig. 24). En el conjunto de ESTs de cítricos del CFGP se identificaron 119 ESTs con elevada similitud de secuencia a CHS previamente caracterizada en *Citrus sinensis* (Moriguchi et al, 1999) y 12 ESTs similares a CHI de *Citrus unshiu* (Moriguchi et al, 2001) (Tabla 18, Tabla Suplementaria XIII).

Tabla 18. Unigenes de cítricos que codifican enzimas de biosíntesis de flavonoides

Enzimas	EC	Contigs	Singletons	Nº Total ESTs
Chalcona sintasa (CHS)	2.3.1.74	5	5	119
Chalcona isomerasa (CHI)	5.5.1.6	2	3	12
Flavanona 3-hidroxilasa (F3H)	1.14.11.9	9	7	50
Flavona sintasa II (FNS II)	-	2	0	8
Flavonoide 3'-hidroxilasa (F3'H)	1.14.13.21	5	2	44
Flavonol sintasa (FLS)	1.14.11.23	1	3	29
Dihidroflavonol 4-reductasae (DFR)	1.1.1.219	6	2	23

En cítricos, la naringenina puede sufrir las siguientes transformaciones:

- 1) Hidroxilación para formar el flavanol dihidroxikaempferol, reacción catalizada por flavanona 3-hidroxilasa (F3H, EC 1.14.11.9; 16 unigenes) (Fig. 24, Tabla Suplementaria XIII). Un total de 50 ESTs del CFGP mostraron similitud de secuencia con el gen de *Citrus unshiu* que codifica F3H (Moriguchi et al, 2001).
- 2) Dos glucosilaciones sucesivas: la primera está catalizada por 7-O-glucosiltransferasa (EC 2.4.1.185) cuya actividad ha sido estudiada en *Citrus paradisi* (McIntosh et al, 1990) aunque todavía no se ha aislado el gen correspondiente. La segunda glucosilación está mediada bien por 1-2 rhamnosiltransferasa (EC 2.4.1.236) para formar naringina o bien por 1-6 rhamnosiltransferasa para formar narirutina (Fig. 24). El gen que codifica 1-2 rhamnosiltransferasa ha sido aislado y caracterizado en *Citrus maxima* (Frydman et al, 2004) pero no se identificó ninguna EST en el CFGP cuya secuencia fuera similar a este gen. La actividad 1-6 rhamnosiltransferasa se ha descrito en cítricos pero todavía no se ha purificado dicha enzima ni se ha caracterizado el gen correspondiente.

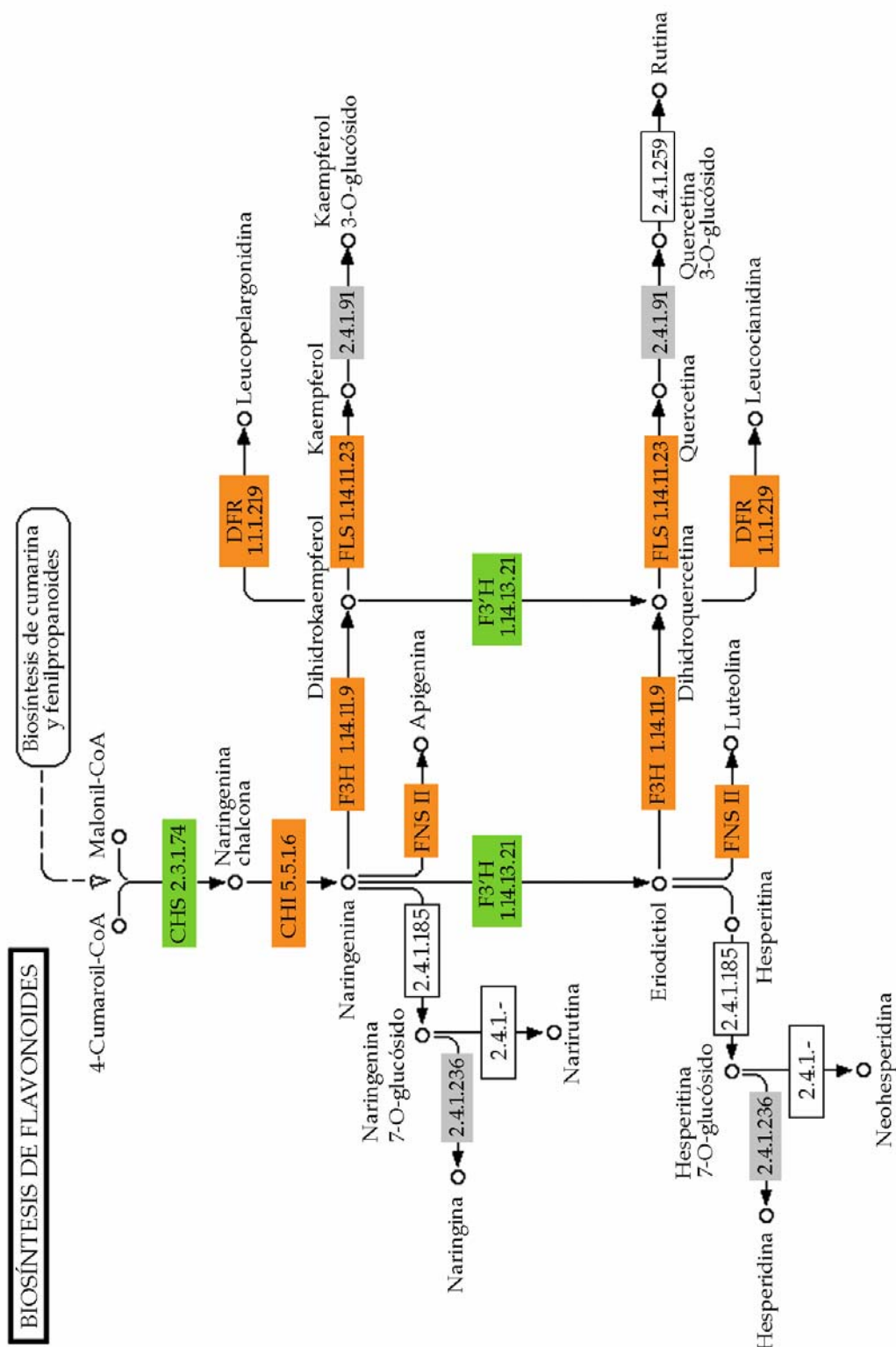


Fig. 24. Biosíntesis de flavonoides en cítricos. Las cajas coloreadas representan los pasos para los que se identificaron unigenes en el conjunto de ESTs del CFGP (en naranja), para los que se encontró algún representante procedente de la genoteca Veg1 (en verde) o para los que ninguna EST de cítricos mostró similitud significativa de secuencia (en gris). Las cajas vacías indican los pasos para los que todavía no se han clonado los genes implicados.

- 3) Oxidación para formar la flavona apigenina, catalizada por flavona sintasa II (FNS II). Las plantas superiores han desarrollado dos sistemas independientes para catalizar la síntesis de flavona: flavona sintasa I (FNS I) presente sólo en la familia *Apiaceae* y flavona sintasa II (FNS II) más extendida en las plantas. El gen que codifica FNS II ha sido aislado y caracterizado en *Gerbera* (Martens y Forkmann, 1999) y en otras especies vegetales. En cítricos, este paso de la biosíntesis de flavonoides todavía no ha sido caracterizado. En el conjunto de ESTs de cítricos del CFGP se identificaron dos unigenes, aCL2638Contig1 y aCL2638Contig2, cuyas secuencias deducidas de aminoácidos fueron similares a FNS II, 54% identidad, 76% similitud y 55% identidad, 76% similitud, respectivamente (Tabla 18, Tabla Suplementaria XIII).
- 4) Hidroxilación para formar eriodictiol, reacción catalizada por flavonoide 3'-hidroxilasa (F3'H, EC 1.14.13.21; 7 unigenes). La actividad enzimática de F3'H ha sido estudiada en cítricos (Doostdar et al, 1995) pero, hasta el momento, no se ha caracterizado el gen de cítricos correspondiente. En la colección de ESTs del CFGP se identificaron 44 ESTs, agrupadas en 7 unigenes, con similitud significativa de secuencia con el gen que codifica F3'H en *Arabidopsis* (At5g07990) (Tabla 18, Tabla Suplementaria XIII). En la colección de ESTs generada a partir de la genoteca Veg1 se encontró representado este gen. Este resultado es consistente con el hecho de que la actividad enzimática de F3'H de cítricos está presente en hojas jóvenes (Doostdar et al, 1995).

La naringina es la flavanona diglucosilada que más se acumula en frutos y hojas de especies como *Citrus paradisi*. Sin embargo, en otras especies de cítricos como la objeto de nuestro trabajo, *Citrus clementina*, la flavanona diglucosilada más abundante es hesperidina (Tabla 19, Kawaii et al, 1999; 2000).

Tabla 19. Contenido en flavonoides en frutos y hojas (tomado de Kawaii et al, 1999, 2000)

Especie	µg/100 mg peso seco			
	Naringina	Narirutina	Hesperidina	Rutina
Frutos				
<i>C. paradisi</i>	1459	500	5	0
<i>C. clementina</i>	0	51.1	852	0
Hojas				
<i>C. paradisi</i>	402	80.5	0	0
<i>C. clementina</i>	0	0	763	885

La hesperidina, análogamente a la naringina, procede de la doble glucosilación de la hesperitina. La biosíntesis de hesperitina no se conoce totalmente pues el paso de eriodictiol a hesperitina todavía no está caracterizado. El eriodictiol puede ser

oxidado para formar la flavona luteolina o hidroxilado formando el flavanol dihidroquercetina, de forma análoga a la naringenina (Fig. 24). Ambos flavanoles, dihidrokaempferol y dihidroquercetina, son convertidos en los flavonoles kaempferol y quercetina, respectivamente, mediante la acción de flavonol sintasa (FLS, EC 1.14.11.23; 4 unigenes) (Fig. 24). Un total de 29 ESTs de cítricos del CFGP, agrupadas en 4 unigenes, mostraron similitud de secuencia con flavonol sintasa de *Citrus unshiu* (Moriguchi et al, 2002) (Tabla 18, Tabla Suplementaria XIII). Algunas especies de cítricos, como *Citrus clementina*, acumulan, especialmente en las hojas, elevadas cantidades de rutina (Tabla 19, Kawaii et al, 1999; 2000). La rutina es un flavonol glucósido con propiedades farmacológicas como antiinflamatorio (Manthey et al, 2001). La conversión de quercetina en rutina se produce en dos pasos (Fig. 24): 1) síntesis de quercetina 3-O-glucósido mediante flavonol 3-O-glucosiltransferasa (EC 2.4.1.91), para la que no se identificó ninguna EST en el CFGP y 2) conversión de 3-O-glucósido en rutina mediante flavonol 3-O-glucósido L-ramnosiltransferasa (EC 2.4.1.259), actividad que ha sido purificada parcialmente en *Vigna radiata* (Barber, 1962) pero para la que no se ha identificado el correspondiente gen.

La utilización de secuencias de genes implicados en la biosíntesis de flavonoides previamente caracterizados como motores de búsqueda en la base de datos de ESTs de cítricos del CFGP ha permitido caracterizar esta ruta en cítricos. Junto a genes ya descritos en cítricos como los correspondientes a CHS (Moriguchi et al, 1999) y CHI y F3H (Moriguchi et al 2001), este trabajo ha permitido identificar, por primera vez en cítricos, los genes potencialmente correspondientes a flavona sintasa II (FNSII) y flavonoide 3'-hidroxilasa (F3'H). La identificación de estos genes constituye el paso inicial para completar nuestro conocimiento actual sobre la síntesis de flavonoides en cítricos. Resulta de especial interés la caracterización de FNSII, enzima clave en la biosíntesis de flavonas, para comprender el papel de este grupo de flavonoides en cítricos.



*Resultados y
discusión II:*

Mapas transcripcionales en cítricos

1. CARACTERIZACIÓN DE LA MICROMATRIZ DE cDNA DE CÍTRICOS, CIT1, GENERADA EN EL CFGP

1.1. Caracterización técnica de la micromatriz CIT1

Los análisis de expresión génica se realizaron utilizando la micromatriz de cDNA de cítricos construida por el Laboratorio de Genómica del CFGP que contiene 12,672 clones que corresponden a 6,875 posibles unigenes (Forment et al, 2005). Tal como se explicó anteriormente, el número de unigenes que resulta de un proceso computacional de agrupamiento o ensamblaje de ESTs varía a medida que el número de ESTs crece. Los 12,672 clones que contiene la micromatriz CIT1 correspondían a 6,875 posibles unigenes cuando la colección de ESTs de cítricos del CFGP contaba con 22,635 ESTs. Actualmente dicha colección contiene 85,965 ESTs, de modo que los 12,672 clones de la micromatriz CIT1 corresponden a 6,034 posibles unigenes. La anotación funcional de los unigenes de cítricos contenidos en la micromatriz CIT1 se recoge en la Tabla Suplementaria XIV. Todos los datos de expresión presentados en este trabajo se obtuvieron utilizando la micromatriz CIT1 y se redefinieron teniendo en cuenta los resultados del último ensamblaje realizado en el CFGP.

Los 12,672 clones contenidos en la micromatriz proceden de 18 genotecas de cDNA (ver Tabla 1 en Materiales y Métodos), entre ellas Veg1. De los 1,824 clones obtenidos a partir de la genoteca Veg1, un total de 1,152 clones fueron incluidos en la micromatriz CIT1 (9.1%) (Tabla Suplementaria XV). El análisis de los productos de PCR mostró que un 90% de los 1,152 clones de la genoteca Veg1 incluidos en la micromatriz CIT1 producían una sola banda en la reacción de PCR (Tabla Suplementaria III). Una proporción elevada de los 6,034 posibles unigenes contenidos en la micromatriz CIT1 fueron impresos entre 2 y 5 veces, lo que permitió la confirmación *intra-array* de los datos de expresión. Por otro lado, alrededor de 4,600 unigenes estaban representados sólo una vez en la micromatriz CIT1.

Con objeto de valorar la fiabilidad de los datos generados utilizando la micromatriz CIT1 se llevó a cabo una caracterización técnica de la misma. Para ello, se realizaron 8 réplicas técnicas utilizando hojas jóvenes procedentes de plantas de invernadero de *Citrus clementina*. En cada una de las réplicas se extrajo el poli(A⁺) RNA y se marcó con el compuesto fluorescente Cy5. El diseño del experimento consistió en la hibridación de cada una de las réplicas técnicas con una muestra de referencia, marcada con el compuesto fluorescente Cy3, formada por hojas procedentes de 40 plantas. Las hibridaciones se realizaron con micromatrices procedentes de un mismo lote. Los puntos que mostraron una intensidad neta menor a

dos veces el valor medio del fondo en los dos canales no se consideraron en los análisis posteriores. Los controles negativos siempre se encontraron en este grupo (Tabla Suplementaria XVI). Los datos de expresión se normalizaron utilizando la función *Lowess* que corrige la dependencia sistemática según la intensidad (Yang et al, 2002; apartado 6.5 de Materiales y Métodos). La Figura 25 muestra los diagramas de caja (*box plot*) del logaritmo en base dos de los cocientes de expresión tras la normalización de los datos con la función *Lowess* (Yang et al, 2002; apartado 6.5 de Materiales y Métodos) para cada una de las ocho réplicas técnicas. Los diagramas de caja mostraron una amplitud o dispersión homogénea de los datos en todas las réplicas, lo que refleja una buena reproducibilidad entre éstas.

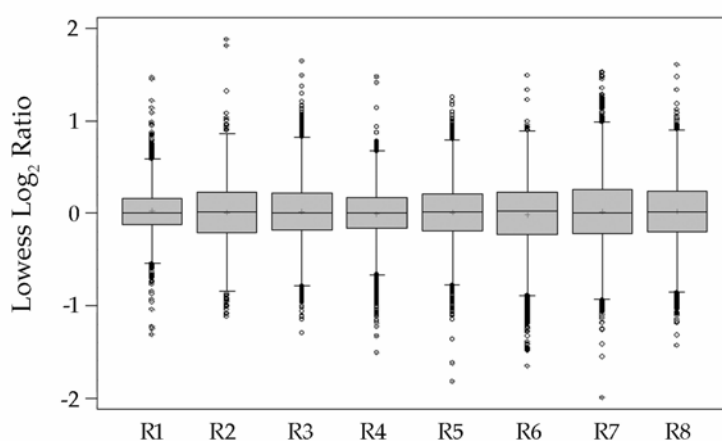


Fig. 25. Diagramas de caja de los datos de expresión. Para cada una de las réplicas técnicas (R1 a R8) se representan los valores correspondientes al logaritmo en base dos de los cocientes de expresión (Cy5/Cy3) tras la normalización de los datos con la función *Lowess* (Yang et al, 2002).

1.1.1. Rango dinámico y niveles de sensibilidad

El rango dinámico y los niveles de sensibilidad se determinaron utilizando los controles depositados en la micromatriz CIT1 (apartado 6.6.2 de Materiales y Métodos). Aunque en algunos casos fue posible detectar intensidades correspondientes al control de calibración 7, es decir 15 pg de mRNA (Tabla 3 en Materiales y Métodos), se obtuvieron resultados reproducibles cuando se incluyeron, al menos, 50 pg mRNA en las muestras (Tabla Suplementaria XVI), lo que representa alrededor de 10 copias de mRNA por célula (Ruan et al, 1998). El rango de intensidades detectadas de modo consistente fue de 200 a 50,000, lo que corresponde a un rango dinámico de 2.5 órdenes de magnitud.

1.1.2. Reproducibilidad *intra-array*

La reproducibilidad *intra-array*, es decir, la reproducibilidad de los datos procedentes de elementos depositados más de una vez en una micromatriz, se estimó, en primer lugar, comparando los niveles de intensidad y los cocientes de expresión de los controles distribuidos en la micromatriz. Los valores de los cocientes de expresión Cy5/Cy3 fueron bastante consistentes, de modo que, para la mayoría de los controles, los valores del coeficiente de variación (CV) no superaron el 14% (Tabla 20, Tabla Suplementaria XVI).

Tabla 20. Valores del coeficiente de variación de los cocientes de expresión obtenidos para los controles de ratio distribuidos en la micromatriz CIT1

	Réplica1	Réplica2	Réplica3	Réplica4	Réplica5	Réplica6	Réplica7	Réplica8
Control Ratio 1	0.20	0.15	0.13	0.13	0.14	0.12	0.08	0.12
Control Ratio 2	0.09	0.10	0.15	0.11	0.12	0.17	0.15	0.12
Control Ratio 3	0.09	0.14	0.13	0.20	0.14	0.11	0.09	0.09
Control Ratio 4	0.09	0.11	0.12	0.16	0.11	0.17	0.08	0.09

El análisis de los valores de expresión de distintos clones de un mismo *contig* permitió obtener otra estimación de la reproducibilidad *intra-array*. Para ello se evaluaron 10 unigenes, seleccionados al azar, representados en la micromatriz CIT1 por 3, 4 ó 5 clones (Tabla Suplementaria XVII). Este análisis también se basó en la agrupación de las ESTs de la colección según el último ensamblaje realizado en el CFGP, de modo que los resultados que se presentan difieren sensiblemente de los obtenidos por Forment et al (2005). Para cada una de las 8 réplicas técnicas realizadas se calcularon los coeficientes de variación de los cocientes de expresión correspondientes a cada uno de los 10 unigenes estudiados (Tabla Suplementaria XVII). Los coeficientes de variación calculados oscilaron entre el 5% y el 16%, y la mayoría de ellos no superaron el 13% (Tabla Suplementaria XVII). La Figura 26 muestra los valores de los cocientes de expresión (Cy5/Cy3) obtenidos para 6 de los unigenes estudiados (aCL920Contig2 y aCL2162Contig1, representados por 3 clones en CIT1; aCL2738Contig1 y aCL436Contig1, por 4 clones; y aCL217Contig1 y aCL36Contig2, por 5 clones) en cada una de las 8 repeticiones técnicas realizadas. Los valores de los cocientes de expresión fueron similares para los diferentes miembros de un *contig* en cada una de las réplicas (Fig. 26). Los unigenes estudiados están representados por clones de diferente longitud y origen, es decir, proceden de diferentes genotecas generadas a partir de diversas condiciones ambientales y de desarrollo. Sin embargo, los resultados obtenidos indican que los niveles de expresión relativos entre las muestras marcadas con Cy5 y con Cy3 fueron consistentes entre los

miembros del mismo *contig* independientemente, por tanto, de la longitud y del origen de los clones que representan cada *contig*.

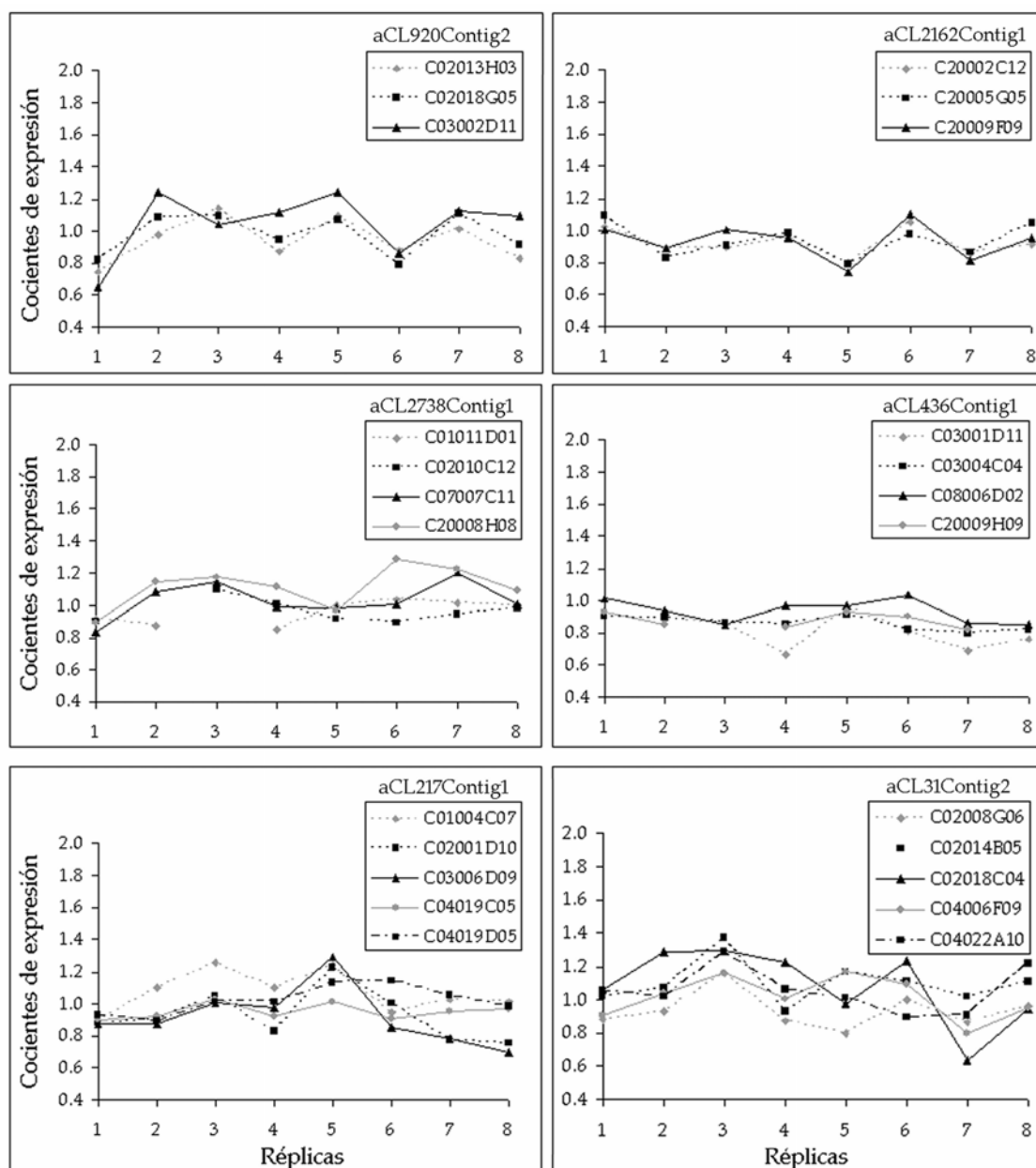


Fig. 26. Representación gráfica de los cocientes de expresión (Cy5/Cy3) correspondientes a cada uno de los clones que representan cada unigen en la micromatriz CIT1, en cada una de las réplicas realizadas. Los unigenes aCL920Contig2 y aCL2162Contig1 codifican, respectivamente, 9-cis-epoxicarotenoide dioxigenasa y beta-ketoacil-CoA sintasa; aCL2738Contig1 y aCL436Contig1 corresponden a la cadena β -1 de tubulina y aldehído dehidrogenasa y los unigenes aCL217Contig1 y aCL36Contig2 codifican la subunidad β de chaperonina de 60 kDa y peroxidasa, respectivamente.

1.1.3. Reproducibilidad *inter-array*

La reproducibilidad *inter-array*, es decir, la reproducibilidad de los datos procedentes de un mismo elemento en cada una de las micromatrices, se evaluó comparando los resultados obtenidos en las 8 réplicas técnicas realizadas. La proporción de puntos cuya intensidad neta superó en, al menos, dos veces a la intensidad del fondo fue del 74%, 75%, 66%, 68%, 79%, 79%, 78% y 70%, para cada una de las ocho réplicas técnicas. Además, se compararon los puntos cuya intensidad neta superaba el criterio establecido en las diferentes micromatrices. Es decir, se calculó cuántos de los puntos (74%) que superaron el criterio de intensidad en la réplica 1 (R1) también los superaban en las demás réplicas. Los resultados mostraron que el mismo conjunto de puntos fue consistentemente seleccionado en cada una de las ocho réplicas técnicas (Tabla 21, Tabla Suplementaria XVIII).

Tabla 21. Comparación de los puntos cuyas intensidades netas superaron el límite establecido en las diferentes réplicas

Réplicas técnicas	Réplicas técnicas							
	R1	R2	R3	R4	R5	R6	R7	R8
R1	-	92	82	84	94	91	93	85
R2	91	-	85	86	95	92	95	89
R3	92	96	-	92	95	93	97	93
R4	91	94	90	-	94	93	96	90
R5	88	90	79	81	-	90	92	84
R6	85	87	78	80	91	-	91	83
R7	87	91	82	83	93	92	-	87
R8	90	95	88	88	96	94	98	-

Los valores corresponden al porcentaje de puntos seleccionados para cada una de las réplicas de la izquierda que, además, se encontraron en cada una de las restantes réplicas.

Adicionalmente, la reproducibilidad *inter-array* se determinó mediante el cálculo del coeficiente de variación de los cocientes de expresión obtenidos, para cada uno de los puntos de la micromatriz CIT1, en las ocho réplicas técnicas (Tabla Suplementaria XIX). La representación gráfica de la distribución de frecuencias del coeficiente de variación mostró que, para la mayoría de los puntos, el coeficiente de variación fue del 8-15%, con un valor promedio del 14%, de modo que el 86% de puntos mostró un coeficiente inferior al 20% (Tabla Suplementaria XIX, Fig. 27A). Finalmente, para cada punto de la micromatriz CIT1 se representaron los valores del coeficiente de variación obtenido en las ocho réplicas frente a la intensidad total, o suma de las intensidades de los dos canales, en función logarítmica (Tabla Suplementaria XIX, Fig. 27B). El promedio del coeficiente de variación observado fue inferior al 14% para los puntos con una intensidad total en \log_{10} inferior a 4, grupo que incluyó el 96% de los puntos. El coeficiente de variación obtenido a partir de réplicas técnicas utilizando

micromatrices de cDNA suele ser del 10-12% (Yue et al, 2001), lo que indica que la plataforma de micromatrices desarrollada en el CFGP permite obtener resultados reproducibles.

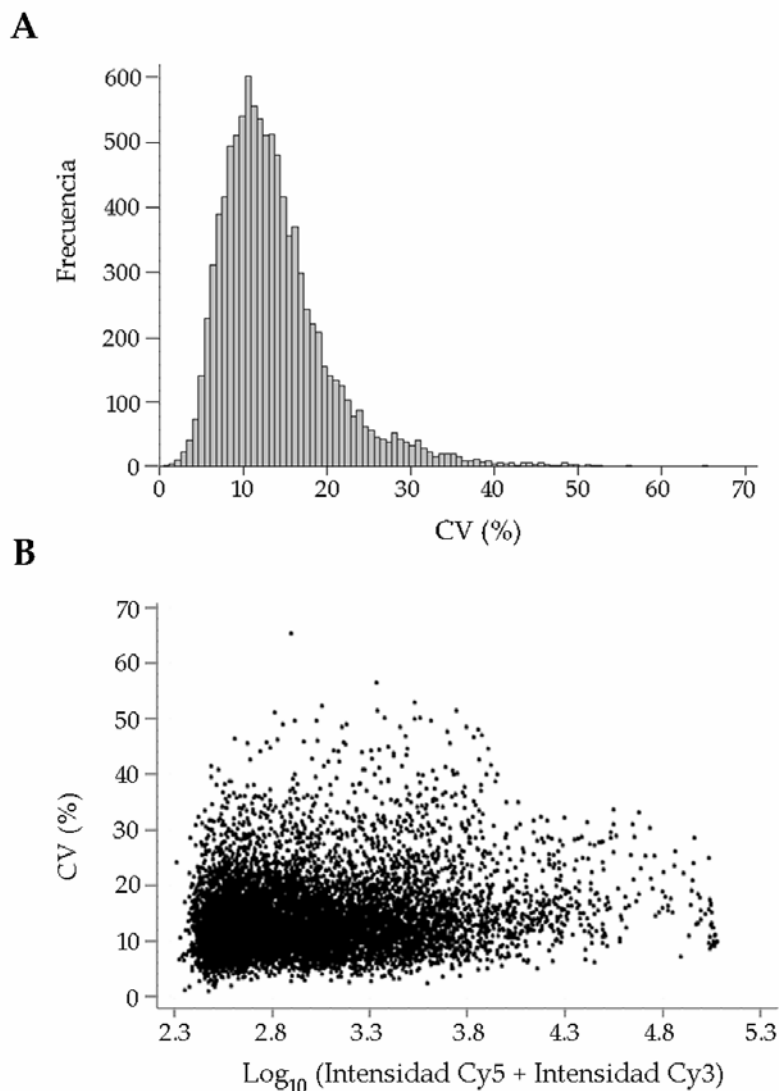


Fig. 27. Coeficiente de variación (CV) de cada uno de los puntos de la micromatriz CIT1 en las ocho réplicas técnicas realizadas. A, Distribución de frecuencias o histograma del coeficiente de variación. B, Representación del coeficiente de variación en función de la intensidad total, como $\log_{10}(\text{Intensidad Cy5} + \text{Intensidad Cy3})$.

1.2. Validación de la micromatriz CIT1 como herramienta para identificar cambios en la expresión génica en cítricos

El análisis de la expresión génica de hojas y entrenudos procedentes de brotes jóvenes de *Citrus clementina* se llevó a cabo con un doble objetivo de validar la utilización de la micromatriz CIT1 para identificar cambios en la expresión génica por comparación de los resultados obtenidos con análisis *Northern* y, por otro lado, como

un experimento preliminar en la búsqueda de genes expresados diferencialmente en diferentes tejidos en cítricos. Las hojas y los entrenudos se recogieron a partir de brotes vegetativos de *Citrus clementina* en dos estadios del desarrollo vegetativo: estadio 1, con hojas de 1-2 cm y estadio 5, con hojas de 6-8 cm (Fig. 28A). El RNA total procedente de las hojas de brotes en el estadio 1 y en el estadio 5 se mezcló a partes iguales y se comparó con el RNA total procedente de los entrenudos, obtenido del mismo modo. El diseño del experimento de micromatrices consistió en cuatro comparaciones directas, dos de ellas con marcaje reverso (Fig. 28B).

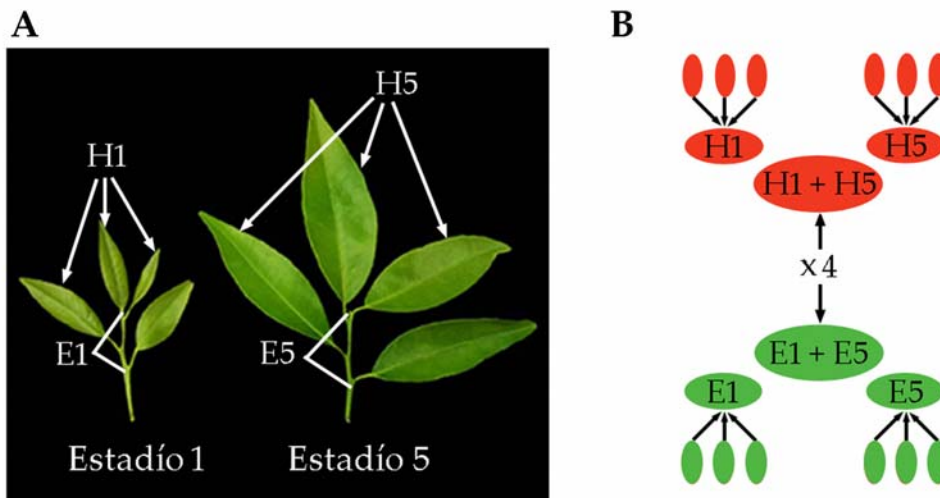


Fig. 28. Análisis de la expresión génica de hojas y entrenudos de brotes vegetativos de *Citrus clementina*. A, Los brotes vegetativos estudiados corresponden a dos estadios del desarrollo vegetativo diferentes: estadio 1, brotes con hojas de 1-2 cm y estadio 5, brotes con hojas de 6-8 cm. Las muestras de cada estadio proceden de, al menos, tres brotes vegetativos. B, El RNA total de las hojas de los brotes del estadio 1 y del estadio 5 se mezcló a partes iguales y se comparó con el RNA de entrenudos de los brotes del estadio 1 y del estadio 5 mezclados, asimismo, a partes iguales. Se realizaron cuatro hibridaciones, dos de ellas con marcaje reverso, es decir, las muestras marcadas con Cy5 se marcan con Cy3 en la hibridación reversa.

Los genes expresados diferencialmente en hojas o en entrenudos se identificaron mediante el análisis estadístico SAM (*Significance Analysis of Microarrays*) (Tusher et al, 2001) aplicando un criterio para delta de modo que el valor de FDR fuera menor al 5% ($\Delta = 0.337$). De este modo se identificaron un total de 696 clones con cambios significativos en su expresión (Fig. 29; Tabla Suplementaria XX). Además de cumplir este criterio, los genes expresados diferencialmente se definieron como aquéllos cuyo valor de cambio de expresión fue de, al menos, 1.6 veces ($|\log_2(\text{Cy5/Cy3})| \geq 0.678$). Estas condiciones permitieron identificar un total de 262 unigenes diferencialmente expresados (4.3% del total de unigenes de la micromatriz CIT1). De ellos, 159 unigenes estaban expresados diferencialmente en hojas y 103 unigenes se encontraron expresados diferencialmente en entrenudos (Tabla Suplementaria XX).

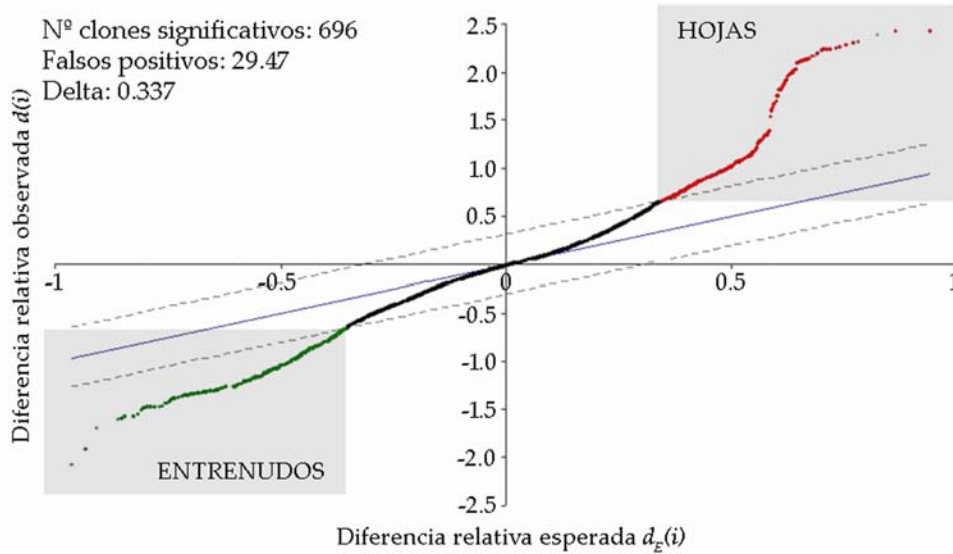


Fig. 29. Identificación de los clones con cambios significativos en su expresión mediante el análisis SAM. El gráfico representa las diferencias relativas observadas $d(i)$ frente a las diferencias relativas esperadas $d_E(i)$. La línea sólida indica la línea $d(i) = d_E(i)$ en la que las diferencias relativas observadas son idénticas a las diferencias relativas esperadas. Las líneas punteadas se encuentran a una distancia $\Delta = 0.337$ de modo que los puntos que se encuentran fuera de esta franja corresponden a clones expresados diferencialmente con un FDR < 5%.

La anotación funcional de los unigenes expresados diferencialmente (Tabla Suplementaria XX) se desarrolló mediante el análisis de la similitud de dichas secuencias, tal como se describe en el apartado 4.3 de Materiales y Métodos. En el siguiente apartado de este capítulo se presentan los resultados obtenidos en la caracterización de diversos tejidos de cítricos mediante el análisis global de la expresión génica. Dado que en el diseño de dicho experimento se incluyeron los tejidos vegetativos comparados en este experimento preliminar (hojas y entrenudos) junto a otros tejidos de cítricos (como raíces y tejidos reproductivos), la interpretación funcional de los resultados obtenidos en ambos experimentos de expresión génica se realizará conjuntamente más adelante (apartado 2.5).

Los datos de expresión génica obtenidos, en la comparación hojas-entrenudos, utilizando la micromatriz de cDNA de cítricos, CIT1, se confirmaron mediante el análisis *Northern* de seis de los unigenes de cítricos diferencialmente expresados. Se seleccionaron tres unigenes diferencialmente expresados en hojas: aCL11Contig1 (proteína relacionada con lectina), aCL1819Contig1 (fosfoetanolamina N-metiltransferasa) y aC20003A04 (unigen sin proteína similar en las bases de datos consultadas; ver apartado 4.3 de Materiales y Métodos) y tres unigenes diferencialmente expresados en entrenudos: aCL8Contig9 (unigen sin proteína similar), aCL3307Contig1 (xiloglucano:xiloglucosil transferasa) y aCL7Contig6 (metalotioneína). En el análisis *Northern* se incluyeron los mismos tejidos (hojas y

entrenudos) en los mismos estadios de desarrollo (estadio 1 y 5), pero además se añadieron hojas y entrenudos procedentes de los estadios de desarrollo intermedios (estadios 2, 3, 4; ver apartado 1.2 de Materiales y Métodos) junto al peciolo y el nervio central de hojas procedentes del estadio de desarrollo 5. Los seis unigenes de cítricos analizados mostraron patrones de expresión similares a los obtenidos utilizando la micromatriz CIT1 (Fig. 30), de modo que la expresión de los unigenes aCL11Contig1, aCL1819Contig1 y aC20003A04 fue claramente superior en hojas y la de los unigenes aCL8Contig9, aCL3307Contig1 y aCL7Contig6 en entrenudos (Fig. 30). La confirmación cualitativa de los patrones de expresión obtenidos mediante análisis *Northern* indica que los datos de expresión obtenidos utilizando la micromatriz CIT1 son consistentes.

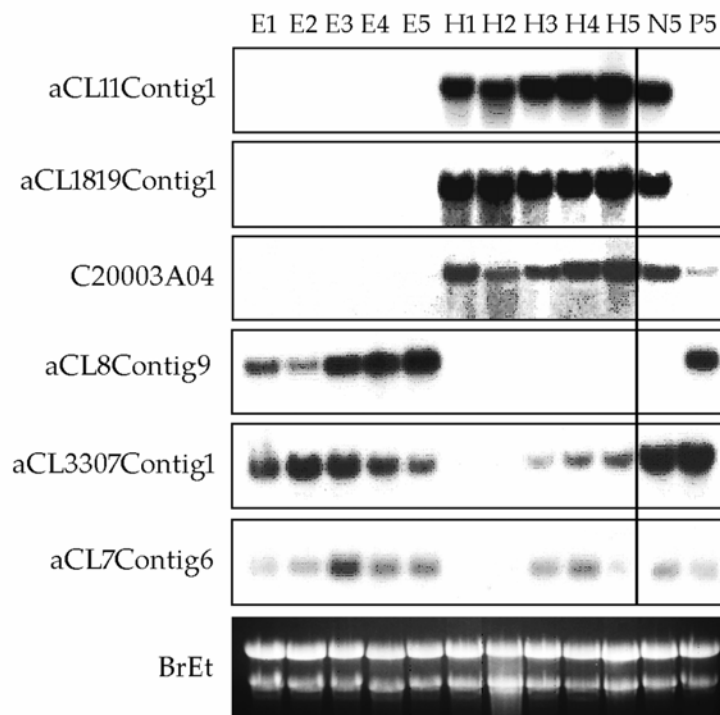


Fig. 30. Análisis *Northern* de los niveles de expresión de tres unigenes de cítricos diferencialmente expresados en hojas (H): aCL11Contig1 (proteína relacionada con lectina), aCL1819Contig1 (fosfoetanolamina N-metiltransferasa) y aC20003A04 (unigen sin proteína similar) y tres unigenes diferencialmente expresados en entrenudos (E): aCL8Contig9 (unigen sin proteína similar), aCL3307Contig1 (xiloglucano:xiloglucosil transferasa) y aCL7Contig6 (metalotioneina) procedentes de brotes vegetativos de *Citrus clementina* en cinco estadios del desarrollo (apartado 1.2 de Materiales y Métodos). La tinción con bromuro de etidio (BrEt) se utilizó como control de carga.

2. CARACTERIZACIÓN DE DIVERSOS TEJIDOS DE CÍTRICOS MEDIANTE EL ANÁLISIS GLOBAL DE LA EXPRESIÓN GÉNICA

2.1. Identificación de los genes comúnmente expresados en diversos tejidos de cítricos

Durante el ciclo de vida de las plantas se forman órganos vegetativos, como raíces, tallos y hojas, así como órganos reproductivos, como flores, frutos y semillas, de modo que cada órgano está especializado en llevar a cabo funciones biológicas concretas. En este trabajo, se estudiaron los patrones de expresión génica de diferentes órganos o tipos de tejidos en cítricos, utilizando la micromatriz CIT1. Para ello se seleccionaron muestras procedentes de diferentes tejidos vegetativos: raíces (RA), lámina foliar (LA), nervio central (NE) y peciolo (PE) de la hoja y entrenudos (EN) de brotes vegetativos en desarrollo; así como muestras procedentes de diferentes tejidos reproductivos: flores en los estadios de desarrollo de botón blanco, bola hueca y antesis y frutos con y sin estilo de *Citrus clementina* (Fig. 31A). Dado que el objetivo general de este trabajo fue la caracterización en detalle del desarrollo vegetativo en cítricos, el diseño del experimento de micromatrices consistió en la comparación directa de RNA procedente de cada uno de los tejidos vegetativos descritos frente a una mezcla, a partes iguales, de RNA procedente de cada uno de los tejidos reproductivos. Se realizaron tres réplicas biológicas (Fig. 31B).

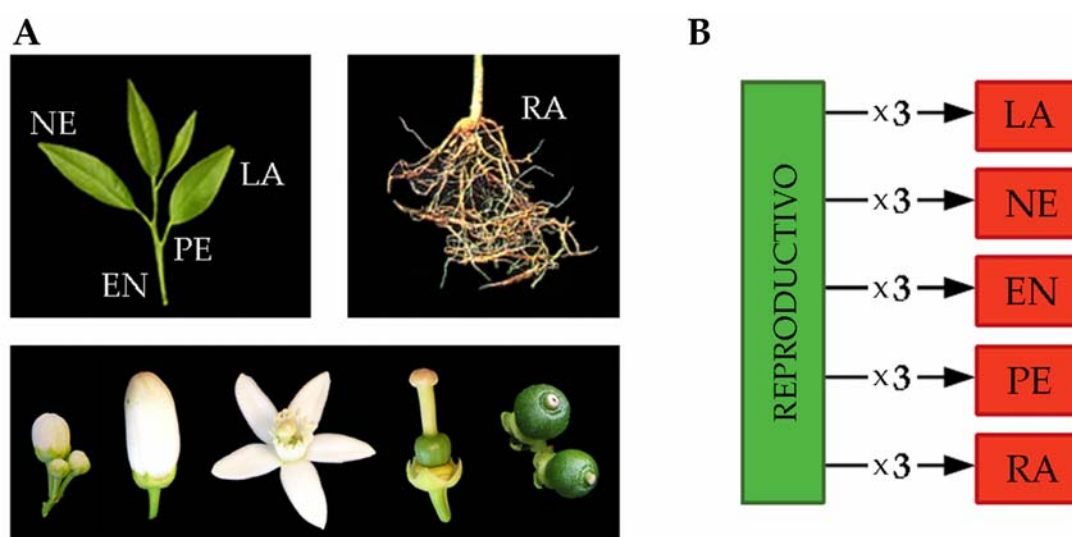


Fig. 31. Diseño del experimento de micromatrices. A, Tejidos vegetativos: raíces (RA), lámina foliar (LA), nervio central (NE) y peciolo (PE) de la hoja y entrenudos (EN) y tejidos reproductivos: flores en diferentes estadios de desarrollo de *Citrus clementina*, utilizados en los experimentos con micromatrices. B, El RNA procedente de cada uno de los tejidos vegetativos se marcó con Cy5 y se comparó con una mezcla de RNA procedente de todos los tejidos reproductivos marcado con Cy3. Se realizaron tres réplicas biológicas para cada tejido.

El número de unigenes de cítricos detectado en los diferentes tejidos estudiados fue similar, desde un 72% a un 84% de los clones de cDNA de la micromatriz CIT1. El número similar de mRNAs encontrado en cada muestra refleja un gran solapamiento en los transcritos presentes en los diferentes tejidos. Con objeto de estimar cuántos unigenes de cítricos presentan una expresión similar en todos los tejidos estudiados se determinó el subconjunto de unigenes cuyos valores de expresión se encontraban comprendidos entre -1.6 y 1.6 ($|\log_2(\text{Cy5/Cy3})| \leq 0.678$) en las 15 comparaciones realizadas. Un total de 2,725 unigenes (45% de los contenidos en la micromatriz CIT1) mostraron niveles de expresión similares en los distintos tejidos estudiados (Tabla Suplementaria XXI). El análisis de enriquecimiento funcional, realizado con el programa FatiGO (Al-Shahrour et al, 2004), de los 2,725 unigenes comúnmente expresados en los distintos tejidos estudiados respecto al conjunto de los unigenes de cítricos representados en la micromatriz CIT1 mostró una sobre-representación de los genes que codifican proteínas implicadas en los procesos de 'traducción' y de 'biogénesis y ensamblaje de ribosomas' (Fig. 32). Este grupo de unigenes comúnmente expresados parece ser esencial para los procesos celulares fundamentales y podría considerarse como genes *housekeeping* en cítricos.

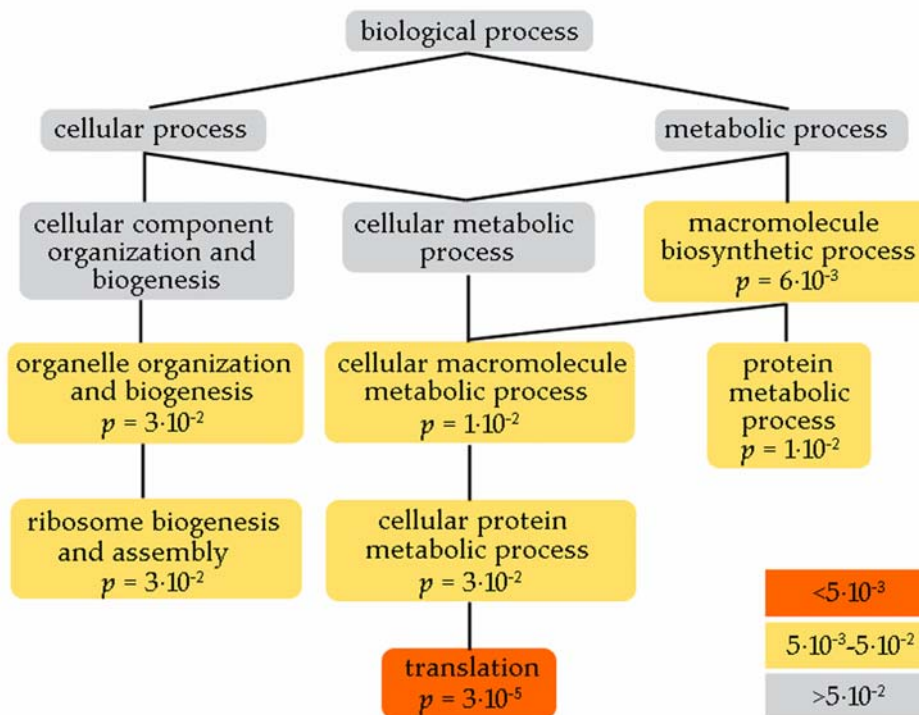


Fig. 32. Ontologías génicas de la componente 'proceso biológico' sobre-representadas en el conjunto de unigenes comúnmente expresados en los distintos tejidos analizados. La escala de color muestra el nivel de significatividad, considerando ontologías enriquecidas aquellas cuyo valor p corregido es inferior a 0.05.

2.2. Relación entre los tejidos estudiados según los perfiles de expresión génica

Los valores de expresión obtenidos para cada gen, en cada uno de los 5 tejidos vegetativos estudiados respecto al conjunto de los tejidos reproductivos, se utilizaron para examinar la relación existente entre los patrones de expresión génica en los diferentes tejidos vegetativos mediante el método de agrupamiento jerárquico (Eisen et al, 1998) utilizando el coeficiente de correlación de Pearson y ligamiento completo como parámetros. El subconjunto de los unigenes cuya expresión no cambia substancialmente en ninguno de los tejidos estudiados (ver apartado anterior) se excluyó del análisis para obtener un agrupamiento más robusto. Los algoritmos de agrupamiento jerárquico construyen un árbol o dendrograma que representa la estructura jerárquica de los datos, en este caso en base a los valores de expresión génica, de modo que la longitud de las ramas refleja el grado de similitud entre ellos (Fig. 33).

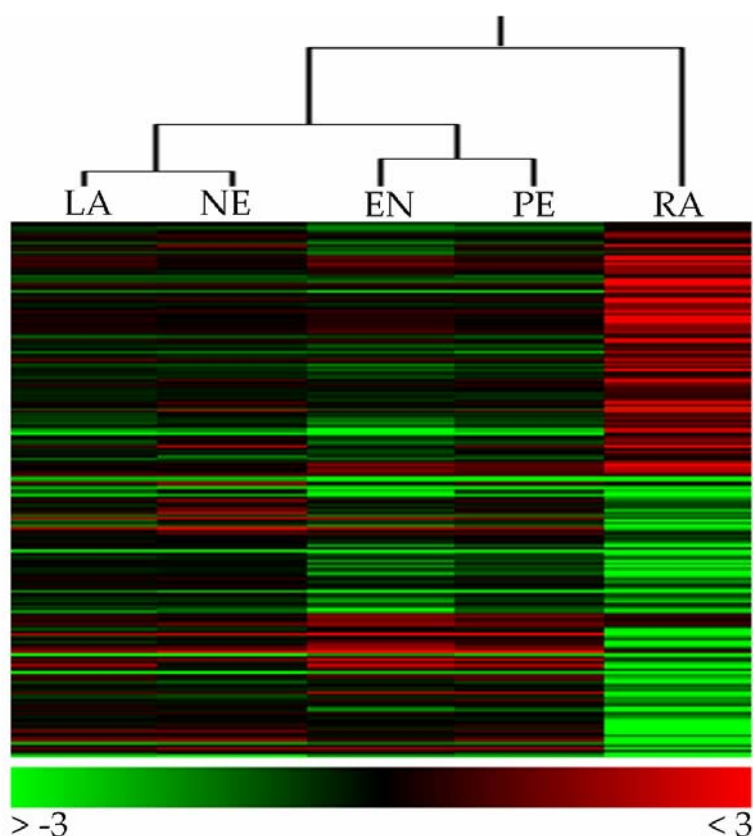


Fig. 33. Relación en los patrones de expresión génica entre los cinco tejidos vegetativos analizados, raíces (RA) y lámina foliar (LA), nervio central (NE) y peciolo (PE) de la hoja y entrenudo (EN) de brotes vegetativos en desarrollo. El agrupamiento jerárquico se realizó y visualizó con el programa Acuity (Axon) utilizando el coeficiente de Pearson y ligamiento completo como parámetros sobre los datos de expresión normalizados por *Lowess* y expresados en \log_2 . El dendrograma refleja la estructura jerárquica de los datos y el mapa de color representa los valores de expresión para cada gen (fila) y en cada uno de los tejidos (columna).

El análisis de agrupamiento jerárquico mostró que los 4 tejidos vegetativos procedentes de brotes en desarrollo, lámina foliar, nervio central y peciolo de la hoja y entrenudos, tienen patrones de expresión génica similares por lo que quedan agrupados en una rama. En esta rama fue posible distinguir dos subgrupos; el formado por la lámina y el nervio central de la hoja y el formado por el peciolo de la hoja y el entrenudo. Sin embargo, el patrón de expresión génica de las raíces diverge considerablemente del grupo anterior (Fig. 33). Estos resultados indican que los tejidos más relacionados desde el punto de vista morfológico y fisiológico tienen perfiles de expresión génica similares que los distinguen de otros tejidos. Las principales diferencias en los perfiles de expresión génica se encuentran entre la raíz y las partes aéreas y entre los órganos vegetativos y los reproductivos, lo que indica la importancia de la identidad de tejido en la elaboración de los mapas transcripcionales. Finalmente, los resultados obtenidos confirman que la utilización de los datos globales de expresión génica constituye una potente herramienta para inferir jerarquías o relaciones entre los diferentes órganos o tejidos de la planta.

2.3. Identificación de los genes preferentemente expresados en la raíz

El diseño del experimento de micromatrices planteado (Fig. 31B) para la identificación de genes de cítricos diferencialmente expresados en los diferentes tejidos estudiados (Fig. 31A) implica que los genes preferentemente expresados en la raíz serán aquellos con expresión diferencial tanto frente al conjunto de los tejidos reproductivos como frente a los diferentes tejidos vegetativos estudiados. La identificación de los genes de cítricos con expresión diferencial en raíz se realizó mediante dos tipos de análisis SAM (Tusher et al, 2001): 1) análisis SAM de una clase para la comparación de la raíz con el conjunto de tejidos reproductivos y 2) análisis SAM de dos clases para la comparación de la raíz con el resto de los tejidos vegetativos, manteniendo un valor de FDR menor al 5% en ambos casos. Adicionalmente se aplicó un límite de expresión de al menos 1.6 veces de modo que el análisis fue más estricto (ver apartado 6.7 de Materiales y Métodos). En estas condiciones se identificaron un total de 734 unigenes de cítricos (12% de los unigenes de CIT1) expresados diferencialmente en la raíz (Tabla Suplementaria XXII). De ellos, 622 unigenes (85% del total de unigenes diferencialmente expresados en raíz) fueron anotados según la función de la proteína de Arabidopsis más similar (Tabla Suplementaria XXII). Con objeto de comprender los procesos biológicos relacionados con los unigenes diferencialmente expresados en la raíz se obtuvo su anotación funcional, basada en GO (Ashburner et al, 2000) (Tabla Suplementaria XXII) y se

realizó un análisis de enriquecimiento funcional con la herramienta FatiGO (Al-Shahrour et al, 2004). El análisis de enriquecimiento funcional mostró que no existía ninguna ontología génica sobre-representada en ninguno de los tres componentes de 'proceso biológico', 'componente celular' y 'función molecular'. Sin embargo, la anotación GO mostró que la proporción de genes implicados en la respuesta a estrés fue relativamente superior en el conjunto de los genes diferencialmente expresados en la raíz (Tabla 22).

Tabla 22. Genes diferencialmente expresados en la raíz implicados en la respuesta a estrés		
	Raíz	CIT1
Unigenes de cítricos con proteína similar en Arabidopsis	622	4465
Unigenes de cítricos con anotación en 'proceso biológico'	357	1634
Respuesta a estrés	50 (14%)	150 (9%)
respuesta a estrés osmótico	10 (2.8%)	25 (1.5%)
respuesta a estrés oxidativo	14 (4%)	25 (1.5%)
respuesta a estrés salino	8 (2.2%)	20 (1.2%)
respuesta a estrés hídrico	11 (3%)	17 (1%)
Respuesta a toxinas	6 (1.7%)	6 (0.4%)

El conjunto de genes expresados diferencialmente en raíz en respuesta a estrés osmótico o salino incluyó genes que codifican MAP kinasas, implicadas en la transducción de señal por estrés osmótico; genes que codifican proteínas de dedo de zinc y genes correspondientes a factores de transcripción de la familia MYB. Entre los genes de respuesta a estrés oxidativo se encontraron diversas peroxidasas, catalasas y proteínas LEA. En particular, se identificaron diversos unigenes de cítricos similares a la proteína LEA5 (At4g02380) que mostraron una expresión elevada en raíces ($\log_2 > 3$; Tabla Suplementaria XXII). En Arabidopsis, la expresión del gen At4g02380 es, asimismo, muy elevada en raíz (Fizames et al, 2004). El grupo de genes de respuesta a estrés hídrico incluyó dos unigenes que codifican acuaporinas o proteínas PIP, implicadas en el transporte de agua a través de las membranas celulares. La secuencia de uno de los unigenes de cítricos presentó una similitud elevada con *PIP2.2* (At2g37170). En Arabidopsis, se ha demostrado que el gen *PIP2.2* se expresa predominantemente en raíz y que la proteína PIP2B es una de las isoformas PIP más abundantes en la raíz (Javot et al, 2003). El grupo de genes de respuesta a toxinas diferencialmente expresado en raíz incluyó, principalmente unigenes que codifican glutatión S-transferasas (GSTs), que catalizan la conjugación de glutatión con una variedad de sustratos electrofílicos, y están implicadas en la detoxificación tanto de compuestos xenobióticos como de compuestos endógenos tóxicos para las células vegetales. En Arabidopsis, la GST codificada por el gen At2g29420 muestra una mayor expresión en raíces (DeRidder y Goldsbrough, 2006).

Adicionalmente, el conjunto de unigenes de cítricos expresados diferencialmente en la raíz incluyó genes implicados en el desarrollo de la raíz. La secuencia del unigen aCL503Contig1 es muy similar al gen *RHD2* de Arabidopsis (At5g51060), que está implicado en la elongación de las raíces y los pelos radiculares, proceso esencial en la absorción de minerales y agua del suelo. En Arabidopsis, *RHD2* se expresa en la epidermis y en la zona de elongación de la raíz y en los pelos radiculares en elongación (Foreman et al, 2003). El unigen aCL8376Contig1 corresponde al gen *SOS4* de Arabidopsis (At5g37870), que controla la formación de los pelos radiculares y se expresa en dichas estructuras (Shi y Zhu, 2002). Otros unigenes de cítricos similares a *SUR2* (At4g31500), relacionado con la formación de raíces adventicias (Sorin et al, 2006) y *PGP1* (At2g36910), que interviene en el transporte de auxinas y se localiza en el ápice de la raíz (Geisler et al, 2005) se encontraron en el conjunto de los unigenes diferencialmente expresados en la raíz.

Los factores de transcripción (TFs) controlan muchos aspectos del desarrollo de las plantas pues especifican cuándo y dónde se expresa un conjunto de genes. Dado que muchos TFs están regulados a nivel de la transcripción, conocer dónde se expresan los genes que codifican TFs resulta de interés para comprender el papel biológico de dichos TFs (Czechowski et al, 2004). La micromatriz CIT1 contiene 197 unigenes de cítricos que codifican TFs, lo que supone un 3% del total de unigenes representados en CIT1 (Anejo III). En el conjunto de unigenes expresados diferencialmente en raíces de cítricos se identificaron 32 unigenes que codifican TFs (16% de los representados en CIT1) (Tabla Suplementaria XXIII). En la Figura 34 se muestran los valores de expresión, representados como porcentaje de expresión en cada una de las partes aéreas estudiadas (lámina foliar, nervio central y peciolo de la hoja, entrenudo y conjunto de tejidos reproductivos) respecto a la expresión obtenida en la raíz, para 11 unigenes de cítricos que codifican TFs. Una de las familias más representadas entre los unigenes de cítricos diferencialmente expresados en raíz fue la familia MYB (Fig. 34A). En Arabidopsis, la expresión de la mayoría de los genes *MYB* aumenta en respuesta a estrés, por ejemplo, At1g72650, At5g67300 y At4g21440 responden a estrés salino (Yanhui et al, 2006). Adicionalmente, se identificaron unigenes diferencialmente expresados en la raíz que codifican TFs pertenecientes a otras familias (Fig. 34B). Hasta donde sabemos, no existen trabajos publicados acerca de la expresión de la mayoría de los genes que codifican los TFs identificados en este trabajo como diferencialmente expresados en raíz (Tabla Suplementaria XXIII). Sin embargo, el hecho de que estos genes se expresen diferencialmente en la raíz en cítricos puede indicar un papel en la regulación del

desarrollo o de otros procesos específicos de la raíz. Esta información puede constituir la base para el estudio de las funciones de dichos genes.

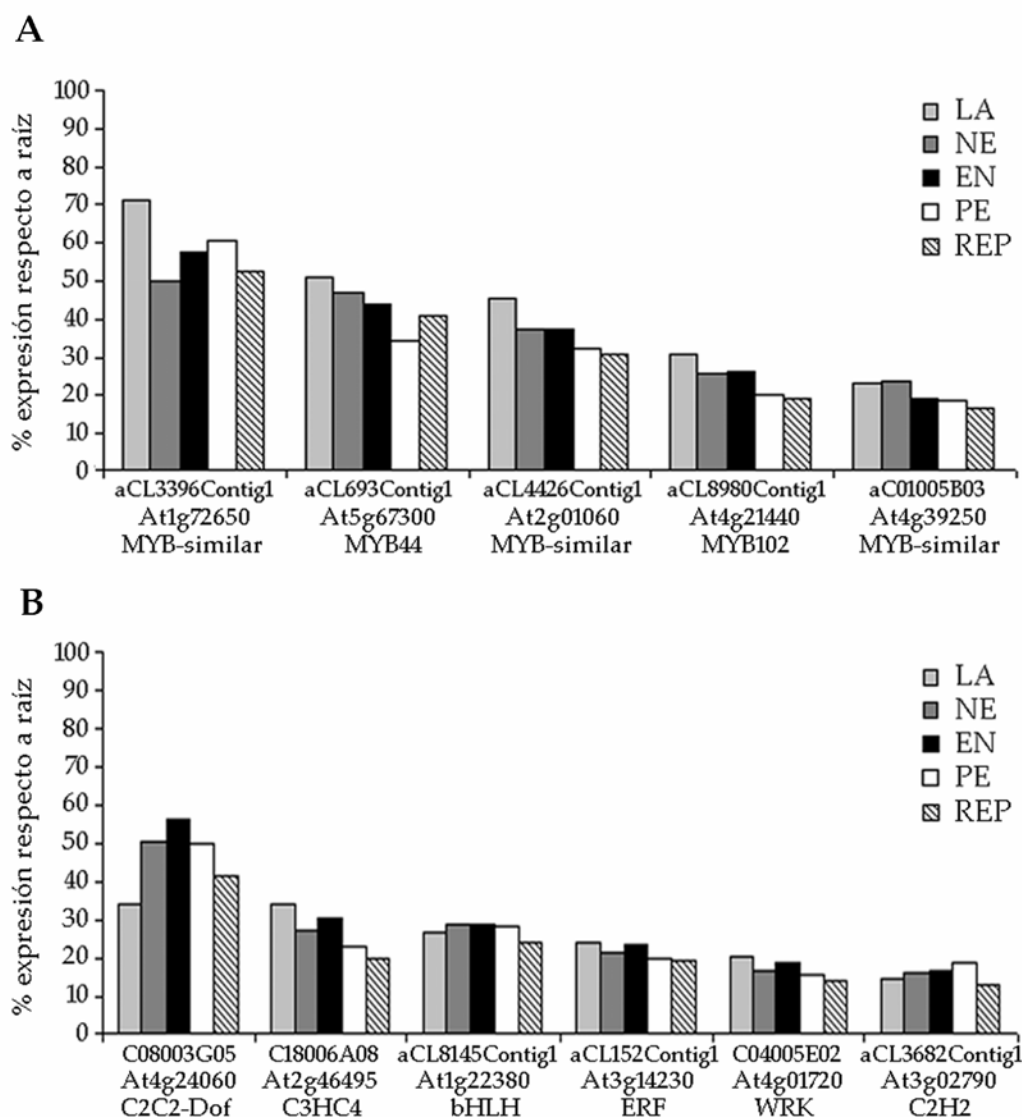


Fig. 34. Porcentaje de expresión en cada una de las partes aéreas estudiadas (lámina foliar, LA; nervio central de la hoja, NE; entrenudo, EN; peciolo, PE y conjunto de tejidos reproductivos, REP) respecto a la expresión obtenida en la raíz para los siguientes genes de cítricos que codifican factores de transcripción de la familia MYB (A) y de las familias C2C2-Dof, C3HC4, bHLH, ERF, WRK y C2H2 (B).

2.4. Identificación de los genes preferentemente expresados en los tejidos reproductivos

La identificación de los genes diferencialmente expresados en los tejidos reproductivos se llevó a cabo siguiendo la misma metodología descrita en el apartado anterior. Los análisis SAM realizados permitieron identificar un total de 238 unigenes de cítricos (4% de los unigenes de CIT1) expresados diferencialmente en el conjunto

de los tejidos reproductivos de cítricos estudiados (Tabla Suplementaria XXIV). El análisis de enriquecimiento funcional, realizado con el programa FatiGO (Al-Shahrour et al, 2004), de este grupo de genes respecto al conjunto de los unigenes representados en la micromatriz CIT1 mostró que las GO no redundantes de ‘desarrollo del carpelo’ (GO:0048440, $p = 9 \cdot 10^{-3}$) ‘biosíntesis de ácidos grasos’ (GO:0006633, $p = 4 \cdot 10^{-2}$) y ‘transporte de lípidos’ (GO:0006869, $p = 2 \cdot 10^{-3}$) estaban significativamente sobre-representadas en el conjunto de unigenes diferencialmente expresados en los tejidos reproductivos (Fig. 35).

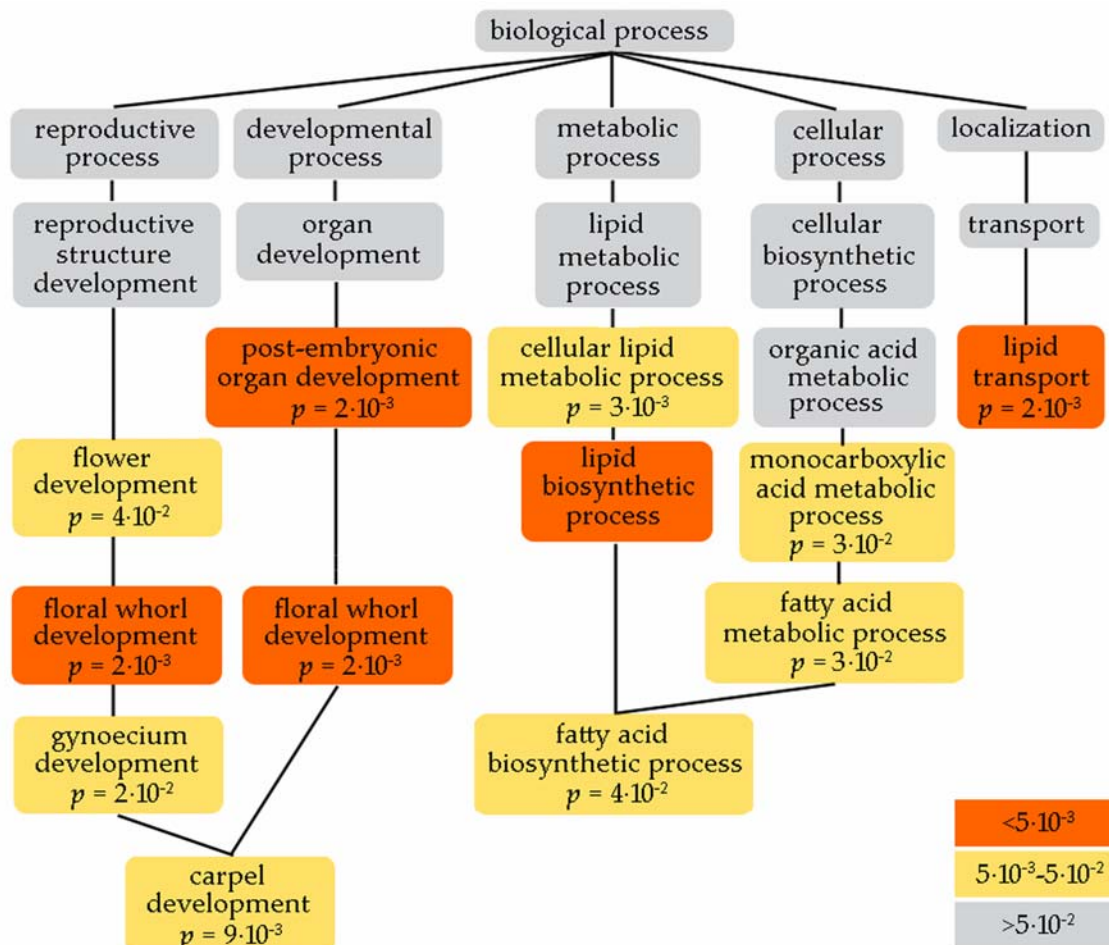


Fig. 35. Ontologías genéticas de la componente ‘proceso biológico’ sobre-representadas en el conjunto de unigenes expresados diferencialmente en los tejidos reproductivos. La escala de color muestra el nivel de significatividad, considerando ontologías enriquecidas aquellas cuyo valor p corregido es inferior a 0.05.

Las GO relacionadas con el desarrollo floral (Fig. 35) incluyeron los unigenes de cítricos implicados en especificar la identidad de los órganos florales como *agamous* (AG, aCL2625Contig1), *pistillata* (PI, aC20007B02), *sepallata 3* (SEP3, aCL176Contig1) y los relacionados con el desarrollo del carpelo como *shatterproof 1*

(*SHP1*, aCL4207Contig1) y *crabs claw* (*CRC*, C20019G10) (Tabla Suplementaria XXIV). Los patrones de expresión de estos genes en los distintos tejidos estudiados mostraron, como era de esperar, una expresión diferencial en el conjunto de los tejidos reproductivos (Fig. 36). Sin embargo, los patrones de expresión de otros genes implicados en el desarrollo floral (como *apetala*, *leafy*, *cauliflower* o *terminal flower 1*, entre otros) no pudieron ser evaluados porque estos genes no se encuentran representados en la micromatriz CIT1.

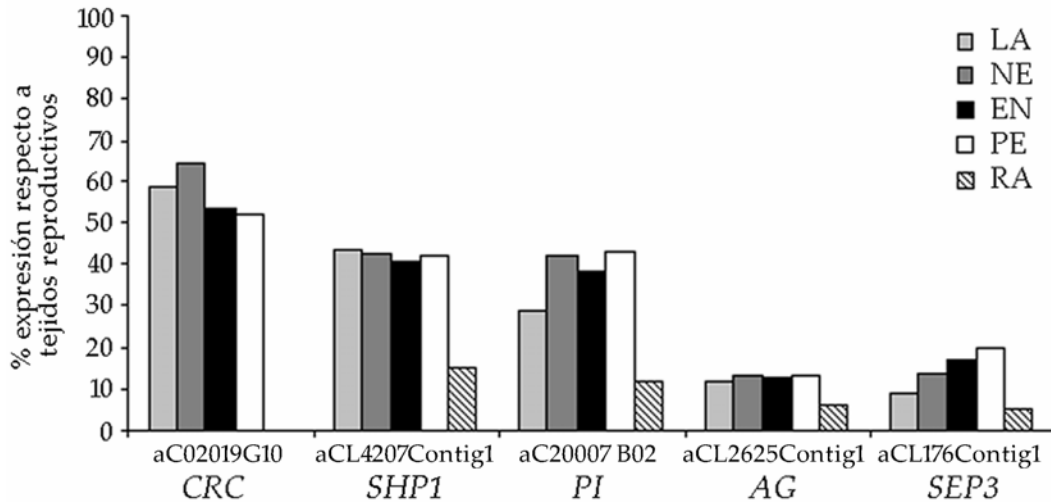


Fig. 36. Porcentaje de expresión en cada uno de los tejidos vegetativos analizados (lámina foliar, NA; nervio central de la hoja, NE; entrenudo, EN; peciolo, PE y raíz, RA) respecto al conjunto de los tejidos reproductivos para los siguientes genes de cítricos implicados en el desarrollo floral: *AG* (aCL2625Contig1), *PI* (aC20007B02), *SEP3* (aCL176Contig1), *SHP1* (aCL4207Contig1) y *CRC* (C20019G10).

Recientemente, se ha estudiado la expresión de los siguientes genes de cítricos implicados en la floración: *leafy*, *apetala1*, *terminal flower 1*, y diversos homólogos a *flowering locus T* (Nishikawa et al, 2007). Sin embargo, los genes implicados en especificar la identidad de los órganos florales como *agamous*, *pistillata*, *sepallata 3* y los relacionados con el desarrollo del carpelo como *shatterproof 1* y *crabs claw* no habían sido identificados previamente en cítricos.

Adicionalmente, el análisis de enriquecimiento funcional mostró que las GO correspondientes a la biosíntesis de ácidos grasos y al transporte de lípidos estaban significativamente sobre-representadas en los tejidos reproductivos (Fig. 35; Tabla Suplementaria XXIV). La ontología génica de biosíntesis de ácidos grasos incluyó los genes *CER1*, que en *Arabidopsis* está relacionado con la fertilidad del polen (Aarts et al, 1995) y *JMT*, cuya expresión es elevada en flores de *Arabidopsis* (Seo et al, 2001). La síntesis y el almacenamiento de lípidos son procesos de conocida importancia durante la microsporogénesis y el desarrollo del polen (Ferreira et al,

1997), lo que resulta coherente con la sobre-representación de los genes implicados en la biosíntesis de ácidos grasos en los tejidos reproductivos de cítricos. La categoría de ‘transporte de lípidos’ incluyó, principalmente, genes que codifican proteínas de transferencia de lípidos (LTPs) (Tabla Suplementaria XXIV). Las LTPs constituyen un grupo de proteínas abundantes en plantas. La capacidad de unión de las LTPs con ácidos grasos y de transferencia de fosfolípidos entre membranas ha sido demostrada *in vitro* pero su función biológica todavía no se conoce. Sin embargo, se ha sugerido un posible papel en el ensamblaje de la cutina y las ceras y en la defensa de las plantas frente al ataque por patógenos (Arondel et al, 2000). La caracterización de la expresión de los miembros de la familia LTP en *Arabidopsis* ha mostrado una expresión elevada en flores (Clark y Bohnert, 1999; Arondel et al, 2000). La identificación de los unigenes de cítricos que codifican LTPs como diferencialmente expresados en tejidos reproductivos resulta consistente con los estudios de expresión realizados previamente en *Arabidopsis*.

Finalmente, en el conjunto de unigenes diferencialmente expresados en los tejidos reproductivos de cítricos se identificaron diversos TFs (Tabla Suplementaria XXIII). En la Figura 37 se muestran los valores de expresión, representados como porcentaje de expresión en cada uno de los tejidos estudiados (lámina foliar, nervio central y peciolo de la hoja, entrenudo y raíz) respecto a la expresión obtenida en el conjunto de los tejidos reproductivos, para 6 unigenes de cítricos que codifican TFs.

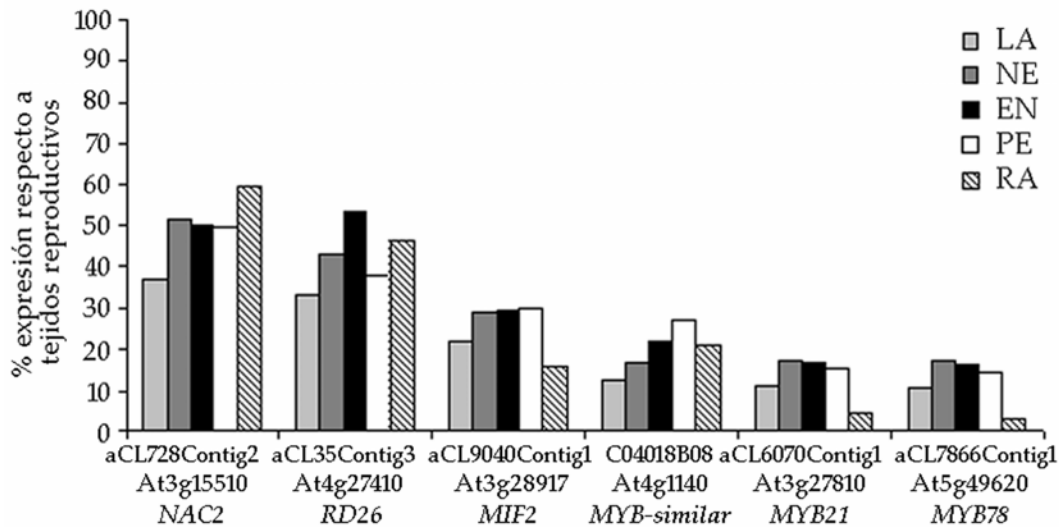


Fig. 37. Porcentaje de expresión en cada uno de los tejidos vegetativos analizados (lámina foliar, NA; nervio central de la hoja, NE; entrenudo, EN; peciolo, PE y raíz, RA) respecto a la expresión obtenida en el conjunto de los tejidos reproductivos para los siguientes genes de cítricos que codifican factores de transcripción de la familia NAC: aCL728Contig2 y aCL35Contig3; de la subfamilia MIF (aCL9040Contig1) y de la familia MYB: aC04018B08, aCL6070Contig1 y aCL7866Contig1.

La familia MYB de TFs es una de las más extensas en plantas, con 198 miembros en Arabidopsis. En los últimos años, los genes de dicha familia MYB han sido implicados en diversos procesos fisiológicos y bioquímicos (Yanhui et al, 2006). Sin embargo, para muchos de los miembros de la familia todavía no se dispone de información funcional. La secuencia del unigen de cítricos aCL6070Contig1 fue similar a la del gen *MYB21* (At3g27810) de Arabidopsis, especie en la que se ha demostrado que dicho gen codifica un TF específico de flor (Shin et al, 2002) y regula importantes aspectos de la respuesta a jasmonato durante el desarrollo de los estambres (Mandaokar et al, 2006). El unigen aCL7866Contig1 mostró una elevada similitud de secuencia con el gen *MYB78* de Arabidopsis. La función de dicho gen aún no ha sido estudiada, pero los patrones de expresión obtenidos en nuestro trabajo muestran una clara expresión diferencial en tejidos reproductivos de cítricos. La secuencia del unigen aCL9040Contig1 fue similar al gen *MIF2* de Arabidopsis, cuya expresión es elevada en flores y frutos (Hu y Ma, 2006). La identificación de nuevos TFs diferencialmente expresados en los tejidos reproductivos supone la base para el estudio de su posible papel en la regulación del desarrollo reproductivo en cítricos.

2.5. Identificación de los genes preferentemente expresados en los tejidos vegetativos

La identificación de los genes de cítricos con expresión diferencial en tejidos vegetativos se realizó mediante dos tipos de análisis SAM (Tusher et al, 2001): 1) análisis SAM de una clase para la comparación del conjunto de tejidos vegetativos con el conjunto de tejidos reproductivos y 2) análisis SAM de dos clases para la comparación del conjunto de tejidos vegetativos con la raíz, manteniendo un valor de FDR menor al 5% en ambos casos. Adicionalmente se aplicó un límite de expresión de al menos 1.6 veces de modo que el análisis fue más estricto (ver apartado 6.7 de Materiales y Métodos). En estas condiciones se identificaron un total de 185 unigenes de cítricos (3% de los unigenes de CIT1) expresados diferencialmente en tejidos vegetativos (Tabla Suplementaria XXV). El análisis de enriquecimiento funcional, realizado con el programa FatiGO (Al-Shahrour et al, 2004), mostró una sobre-representación significativa de los genes implicados en la GO no redundante de 'transporte fotosintético de electrones en el fotosistema I' (GO:0009773, $p = 1 \cdot 10^{-2}$) (Fig. 38). Los genes relacionados con la cadena de transporte electrónico, como los que codifican las diferentes subunidades de los fotosistemas I y II, ferredoxina o plastocianina, junto a los que codifican diferentes proteínas de unión a clorofila

mostraron, como cabía esperar, una expresión diferencial en los tejidos vegetativos (Tabla Suplementaria XXV).

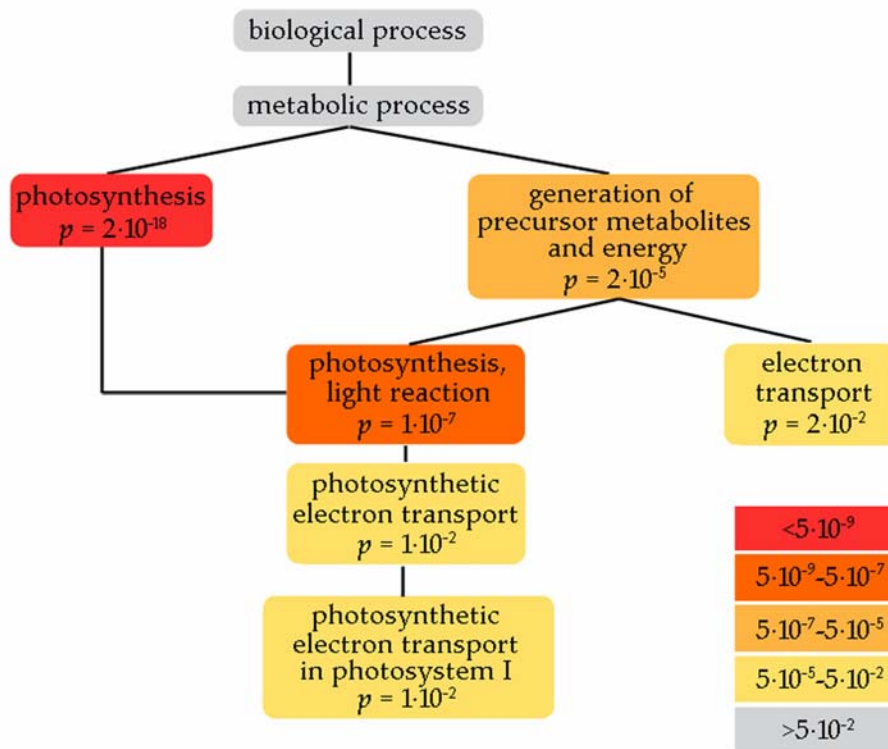


Fig. 38. Ontologías genéticas de la componente 'proceso biológico' sobre-representadas en el conjunto de unigenes expresados diferencialmente en los tejidos vegetativos. La escala de color muestra el nivel de significatividad, considerando ontologías enriquecidas aquellas cuyo valor p corregido es inferior a 0.05.

En el conjunto de unigenes expresados diferencialmente en tejidos vegetativos se identificaron 4 unigenes que codifican TFs y reguladores de la transcripción (2% de los TFs representados en CIT1) (Tabla Suplementaria XXIII). La secuencia del unigen aCL2656Contig1 fue similar al gen *CCA1*. En *Arabidopsis*, la mayor expresión de este gen se encuentra en hojas (Wang et al, 1997). El unigen aCL6460Contig1 mostró una secuencia similar al gen *ICU4*, que está implicado en la morfogénesis de las hojas en *Arabidopsis* (Serrano-Cartagena et al, 2000). El unigen aCL198Contig1 corresponde con el gen *XERICO*, un regulador del metabolismo de ABA cuya expresión es mayor en el meristemo apical del brote (Ko et al, 2006). En cuanto a los genes expresados diferencialmente en entrenudos o en hojas (ver apartado 1.2), cabe señalar que ninguno de ellos, excepto el unigen aCL11Contig1, que codifica una proteína relacionada con lectina, mostró una expresión diferencial en tejidos vegetativos.



*Resultados y
discusión III:*

Regulación del transcriptoma por GAs

1. CAMBIOS EN LA EXPRESIÓN GÉNICA DE ENTRENUDOS DE PLANTAS TRANSGÉNICAS DE CÍTRICOS QUE SOBRE-EXPRESAN UNA GA 20-oxidasa

1.1. La sobre-expresión en sentido y en antisentido del gen *CcGA20ox1* modifica los niveles de transcritos endógenos de *CcGA20ox1*

Las plantas transgénicas de citrange Carrizo, un importante patrón de cítricos, que sobre-expresan un gen de cítricos que codifica una GA 20-oxidasa (*CcGA20ox1*), tanto en sentido como en antisentido, muestran una arquitectura modificada que se correlaciona con mayores (sentido) o menores (antisentido) niveles de GAs (Fagoaga et al, 2007). En este trabajo se seleccionaron dos líneas representativas; la línea S23, de sobre-expresión de *CcGA20ox1* en sentido y la línea A4, que sobre-expresa dicho gen en antisentido. El contenido de GA₁ en estas líneas está significativamente alterado en brotes en desarrollo, de modo que mientras los niveles de GA₁ en el control son de 8.0 ng/g peso fresco, en las plantas S23 y A4 son de 22.3 ng/g peso fresco y 4.6 ng/g peso fresco, respectivamente (Fagoaga et al, 2007). En concreto, se estudiaron los entrenudos procedentes de brotes jóvenes de estas plantas porque se sabe que el crecimiento de estos órganos depende de GAs (Reid y Howell, 1995; Israelsson et al, 2004; Sun, 2004).

Previamente a la caracterización del transcriptoma de estas plantas, se confirmó que las líneas transgénicas seleccionadas en este trabajo contenían niveles elevados de transcritos correspondientes al transgén *CcGA20ox1* (Fig. 39A). Los niveles de transcritos de *CcGA20ox1* en la línea S23 fueron 11 veces superiores a los obtenidos en las plantas control. Asimismo, se observó un aumento de 13 veces en la línea A4 respecto a los niveles de transcritos de *CcGA20ox1* obtenidos en las plantas control. Por otro lado, se analizaron mediante RT-PCR los niveles endógenos de transcritos de *CcGA20ox1* utilizando una secuencia de la región 3' no codificante del gen que no está incluida en el transgén (Fig. 39B). En la línea S23 no fue posible detectar los transcritos correspondientes al gen *CcGA20ox1* endógeno; probablemente como resultado del mecanismo de regulación por retroalimentación negativa que, previamente se ha mostrado que regula la transcripción de este gen utilizando aplicaciones de GA₃ y paclobutrazol (Vidal et al, 2003). En el caso de la línea A4, la reducción observada en los niveles endógenos de transcritos de *CcGA20ox1* fue variable entre las réplicas biológicas, con una reducción promedio del 50% (Fig. 39B). El mismo tipo de material que el descrito en este análisis se utilizó para evaluar los cambios de expresión génica en las líneas transgénicas seleccionadas.

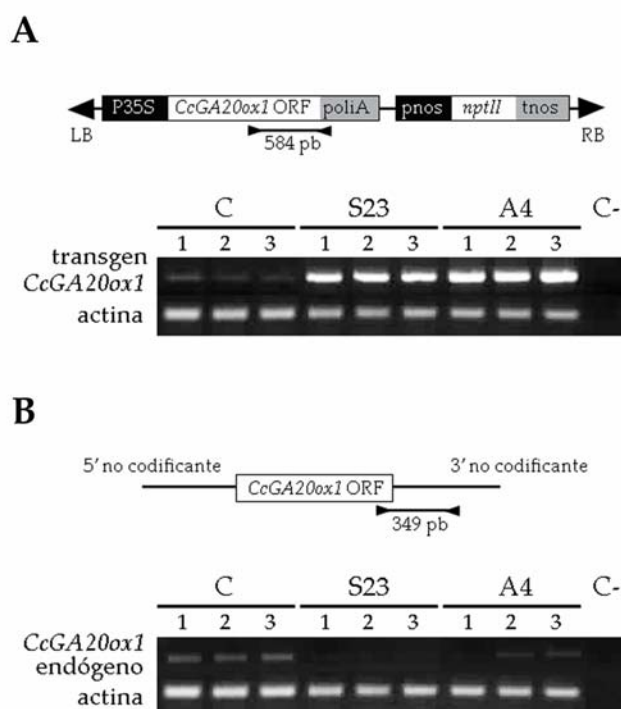


Fig. 39. Niveles de transcritos del transgén *CcGA20ox1* (A) y del gen *CcGA20ox1* endógeno (B) en tres réplicas biológicas de las líneas representativas de sobre-expresión en sentido (S23), en antisentido (A4) y en la línea control (C). Los niveles de transcritos se determinaron mediante RT-PCR semicuantitativa tal como se describe en el apartado 7.2 de Materiales y Métodos, utilizando una actina de cítricos (CX289161) como control interno.

1.2. La sobre-expresión en sentido del gen *CcGA20ox1* causa un remodelado substancial del transcriptoma

La caracterización del transcriptoma de plantas transgénicas de cítricos que presentan niveles modificados de GA (por sobre-expresión en sentido y en antisentido del gen *CcGA20ox1*), y por tanto una arquitectura de la planta modificada (Fig. 40A) (Fagoaga et al, 2007), se llevó a cabo utilizando la micromatriz CIT1 (Forment et al, 2005; ver apartado 5 de Materiales y Métodos). El diseño del experimento consistió en comparaciones directas, es decir, las muestras procedentes de las plantas que sobre-expresan el gen *CcGA20ox1* se compararon con los controles, plantas transformadas con el vector vacío, en la misma micromatriz (Fig. 40B). Se realizaron tres réplicas biológicas, cada una de ellas compuesta de tres entrenudos procedentes de brotes jóvenes en desarrollo de la línea representativa de sobre-expresión en sentido (S23) o en antisentido (A4). El RNA de entrenudos de brotes jóvenes en desarrollo de plantas transformadas con el vector vacío (plantas control) se juntó y se utilizó como referencia con significado biológico. Los criterios de diseño se especifican con más detalle en el apartado 6 de Materiales y Métodos.

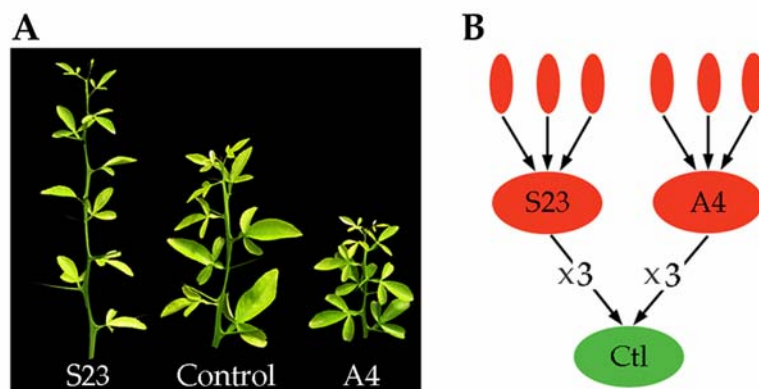


Fig. 40. Diseño del experimento. A, El transcriptoma de las plantas transgénicas que sobre-expresan el gen *CcGA20ox1* se analizó a partir de los entrenudos de brotes en desarrollo de las líneas S23 y A4. B, Las muestras procedentes de las líneas transgénicas se marcaron con el compuesto fluorescente Cy5 y se compararon con las procedentes de plantas control, marcadas con Cy3, mediante la co-hibridación en la misma micromatriz. Se realizaron tres réplicas biológicas para cada comparación directa.

La Figura 41A muestra los diagramas de caja del \log_2 de los cocientes de expresión (Cy5/Cy3) tras la normalización de los datos con la función *Lowess* (Yang et al, 2002; ver apartado 6.5 de Materiales y Métodos) para cada una de las tres réplicas correspondientes a la línea de sobre-expresión en sentido (S1, S2, S3) y en antisentido (A1, A2, A3). Los diagramas de caja mostraron una dispersión muy similar en cada grupo de datos, lo que refleja una buena reproducibilidad en cada uno de los dos grupos de réplicas. Las réplicas de la línea A4 presentaron una amplitud menor que las réplicas de la línea S23, lo que indica que los cambios en el transcriptoma correspondientes a la línea en antisentido son menores que los observados para la línea en sentido. La representación de los datos de expresión correspondientes a ambos grupos de réplicas mediante 'gráficos de volcán' confirmó esta observación (Fig. 41B). Los 'gráficos de volcán' representan, para cada uno de los genes de la micromatriz, la relación entre los cocientes de expresión y los valores de probabilidad de que las diferencias observadas ocurran por casualidad (valor p), de modo que los sectores superior izquierdo y derecho incluyen los cambios de expresión estadísticamente significativos. El gráfico correspondiente a las réplicas de la línea en sentido mostró un elevado número de datos en dichos sectores mientras que aparecieron casi vacíos en el caso de las réplicas correspondientes a la sobre-expresión en antisentido de *CcGA20ox1* (Fig. 41B). Por tanto, como era de esperar, se observó un elevado número de genes expresados diferencialmente en la línea S23 mientras que no se encontraron diferencias significativas entre el transcriptoma de la línea A4 y del control. La ausencia de cambios en el transcriptoma de la línea A4 puede deberse a la relativamente baja reducción del contenido de GAs en estas

plantas (8.0 ng/g peso fresco en el control frente a 4.6 ng/g peso fresco en la línea A4, Fagoaga et al, 2007). Por lo tanto, sólo se describirán los resultados obtenidos con las plantas que sobre-expresan el gen *CcGA20ox1* en sentido, en adelante, plantas *CcGA20ox*.

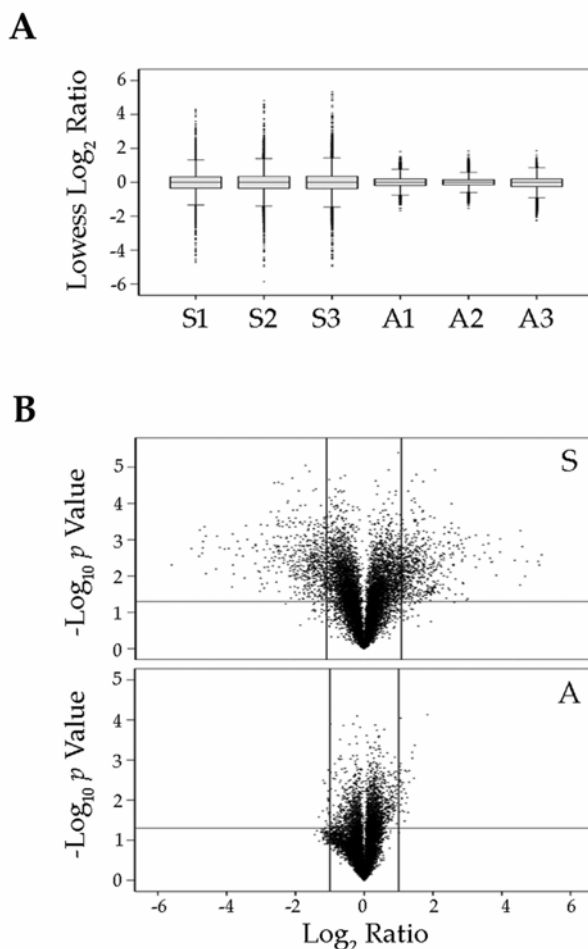


Fig. 41. Cambios globales en la expresión génica de entrenudos de citrange Carrizo que sobre-expresan el gen *CcGA20ox1* en sentido y en antisentido. A, Los diagramas de caja muestran la distribución del log₂ del cociente de intensidades tras la normalización *Lowess* para cada una de las tres réplicas en sentido (S1, S2, S3) y de las tres en antisentido (A1, A2, A3). B, Los gráficos de volcán representan el efecto de la sobre-expresión del gen *CcGA20ox1* en sentido (S) y en antisentido (A). El eje x muestra el promedio de la expresión relativa entre las plantas transgénicas y los controles en las tres réplicas, expresado en log₂. El eje y muestra la significatividad de la *t* de Student. La línea horizontal representan el umbral de $p=0.05$ y la barra vertical los cambios de expresión de dos veces.

En las plantas *CcGA20ox*, se identificaron un total de 1,228 ESTs que corresponden a 726 unigenes (12% del total de unigenes de la micromatriz CIT1) con expresión diferencial (p valor del test *t* de Student corregido según Benjamini-Hochberg ≤ 0.05 y un valor de cambio de expresión de al menos 1.6 veces). Adicionalmente, se detectaron otros 249 clones de cDNA de los cuales no se dispone

de información de secuencia (es decir, no presentan más de 100 pb de secuencia de alta calidad) como diferencialmente expresados, pero no se consideraron en los análisis posteriores. Estos resultados indican que la modificación genética del metabolismo de GAs provoca una remodelación considerable del transcriptoma. En el total de los genes expresados diferencialmente, 336 (46.3%) correspondieron a genes inducidos (Tabla Suplementaria XXVI) y 390 (53.7% del total) a genes reprimidos (Tabla Suplementaria XXVII). El hecho de que ambas proporciones sean similares sugiere que ambos modos de regulación, inducción y represión, tienen una importancia similar en la respuesta a la sobre-expresión del gen *CcGA20ox1*. Resulta de interés destacar que un 14.6% de los genes expresados diferencialmente (106) no mostraron similitud de secuencia significativa con ninguna proteína de Arabidopsis. Este grupo probablemente incluye los genes específicos de cítricos y el valor relativamente elevado obtenido refleja la importancia de la secuenciación de ESTs en diferentes especies vegetales. Algunos de estos genes podrían estar implicados en funciones o rutas metabólicas específicas de cítricos.

La anotación de los unigenes expresados diferencialmente (Tablas Suplementarias XXVI y XXVII) se desarrolló mediante análisis de similitud de dichas secuencias, tal como se describe en el apartado 4.3 de Materiales y Métodos. En el análisis funcional y en la discusión de los unigenes de cítricos expresados diferencialmente, se utilizó la proteína de Arabidopsis más similar para cada unigen, de modo que fue posible realizar esta asignación en un 85.4% de los genes expresados diferencialmente (620 unigenes). Con el objetivo de obtener una visión más integrada de los cambios de expresión génica observados y para explorar los procesos biológicos en los que los genes expresados diferencialmente se encuentran implicados, se obtuvo la anotación funcional del MIPS mediante la herramienta FunCat (Ruepp et al, 2004; ver Tabla Suplementaria XXVIII para los unigenes inducidos y Tabla Suplementaria XIX para los unigenes reprimidos). En la mayoría de las categorías funcionales, la proporción de genes inducidos fue similar a la proporción de genes reprimidos (Tabla 23). Sin embargo, se observó una tendencia hacia la inducción en las categorías de 'Energía' (4.4% frente a 1.8%) e 'Interacción con el Ambiente' (7.0% frente a 3.9%). Por otro lado, se observó una tendencia hacia la represión en la categoría de 'Síntesis de Proteínas' (1.4% frente a 7.9%).

Tabla 23. Clasificación funcional según FunCat de los unigenes expresados diferencialmente en las plantas CcGA20ox

CATEGORÍA FUNCIONAL	Inducidos		Reprimidos	
	No.	%	No.	%
01 Metabolismo	73	14.6	66	10.6
02 Energía	22	4.4	11	1.8
04 Proteínas de Reserva	1	0.2	–	0.0
10 Ciclo Celular y Procesamiento del DNA	5	1.0	11	1.8
11 Transcripción	10	2.0	9	1.4
12 Síntesis de Proteínas	7	1.4	49	7.9
14 Procesado de proteínas (plegamiento, modificación, destino)	23	4.6	35	5.6
16 Proteínas con Función de Unión o Necesidad de Cofactores	60	12.0	99	15.9
18 Regulación del Metabolismo	5	1.0	7	1.1
20 Transporte Celular	32	6.4	32	5.1
30 Comunicación Celular – Transducción de Señal	8	1.6	9	1.4
32 Defensa	32	6.4	32	5.1
34 Interacción con el Ambiente	35	7.0	24	3.9
36 Interacción Sistémica con el Ambiente	11	2.2	14	2.2
40 Procesado Celular	3	0.6	6	1.0
41 Desarrollo (Sistémico)	8	1.6	13	2.1
42 Biogénesis de Componentes Celulares	8	1.6	14	2.2
43 Diferenciación de Tipos Celulares	2	0.4	3	0.5
47 Diferenciación de Órganos	–	–	3	0.5
70 Localización Subcelular	111	22.2	142	22.8
77 Localización en Órganos	3	0.6	–	–
99 Proteínas sin Clasificación	40	8.0	44	7.1
TOTAL	499	100.0	623	100.0

Entre los 726 genes expresados diferencialmente, 293 pertenecen a múltiples categorías.

Los resultados obtenidos en los análisis de expresión génica con la micromatriz CIT1 se confirmaron mediante el análisis de la expresión, por RT-PCR con oligonucleótidos específicos, de 6 unigenes expresados diferencialmente en las plantas CcGA20ox1 y en los controles. Se seleccionaron unigenes relacionados con la fotosíntesis (aCL48Contig1, RuBisCo activasa; y aCL43Contig3, subunidad pequeña de RuBisCo), con estrés de origen abiótico (aCL172Contig2, RD22), con la pared celular (aCL3307Contig1, xiloglucano endotransglucosilasa/hidrolasa), con el metabolismo secundario (aCL960Contig1, geranilgeranil difosfato sintasa) y un unigen (aCL8Contig9) cuya secuencia no muestra similitud significativa con ninguna secuencia de las bases de datos consultadas (ver apartado 4.3 de Materiales y Métodos). Los niveles de transcritos de los 6 genes examinados (Fig. 42A) mostraron un alto grado de concordancia con los cambios de expresión génica detectados en el análisis transcriptómico, que se expresa mediante la buena correlación ($R = 0.96$) obtenida entre ambos métodos (Fig. 42B). Así pues, los resultados obtenidos mediante RT-PCR apoyan los generados con la aproximación transcriptómica utilizando la micromatriz CIT1, demostrando la reproducibilidad de los datos obtenidos en nuestras condiciones experimentales.

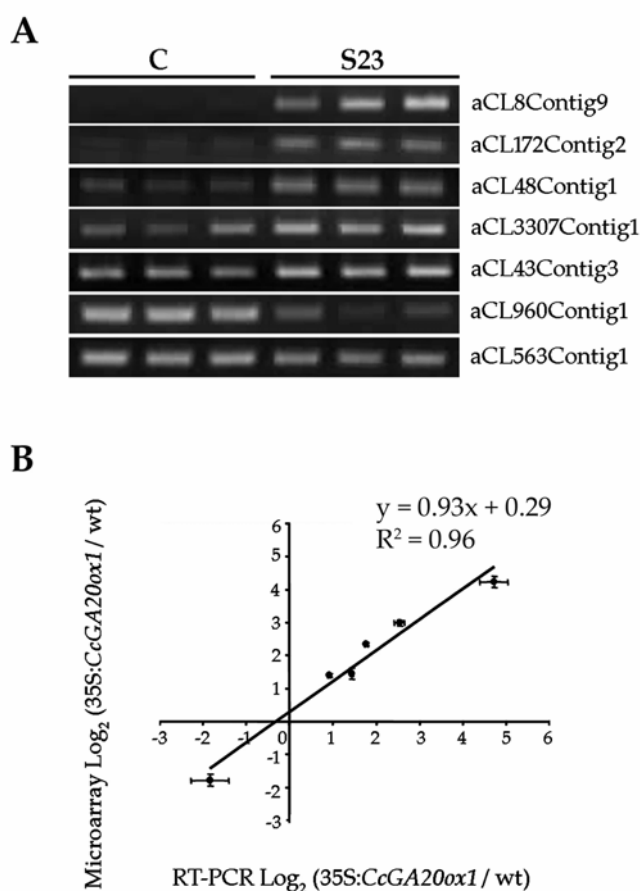


Fig. 42. Confirmación de los datos obtenidos con la micromatriz CIT1 mediante RT-PCR. A, La expresión de seis unigenes expresados diferencialmente en las plantas *CcGA20ox* se ha estudiado por RT-PCR semicuantitativa utilizando oligonucleótidos específicos. El gen utilizado como control interno fue una actina de cítricos (CX289161). aCL8Contig9, sin anotación; aCL172Contig2, RD22; aCL48Contig1, RuBisCo activasa; aCL3307Contig1, xiloglucano:xiloglucosil transferasa; aCL43Contig3, RuBisCo subunidad pequeña; aCL960Contig1, geranilgeranil difosfato sintasa; aCL563contig1, actina. B, Correlación entre los resultados obtenidos mediante RT-PCR y análisis transcriptómico para seis genes expresados diferencialmente en las plantas *CcGA20ox*. Los valores representados corresponden a la media de tres réplicas biológicas \pm SE.

1.3. La sobre-expresión del gen *CcGA20ox1* implica la sobre-representación de diversas ontologías génicas

Con el objetivo de identificar procesos clave alterados en las plantas *CcGA20ox* realizamos un análisis de enriquecimiento funcional en el conjunto de genes expresados diferencialmente utilizando la herramienta FatiGO (Al-Shahrour et al, 2004), que está basada en Ontologías Génicas (GO) (Ashburner et al, 2000), tal como se describe en el apartado 4.3 Materiales y Métodos.

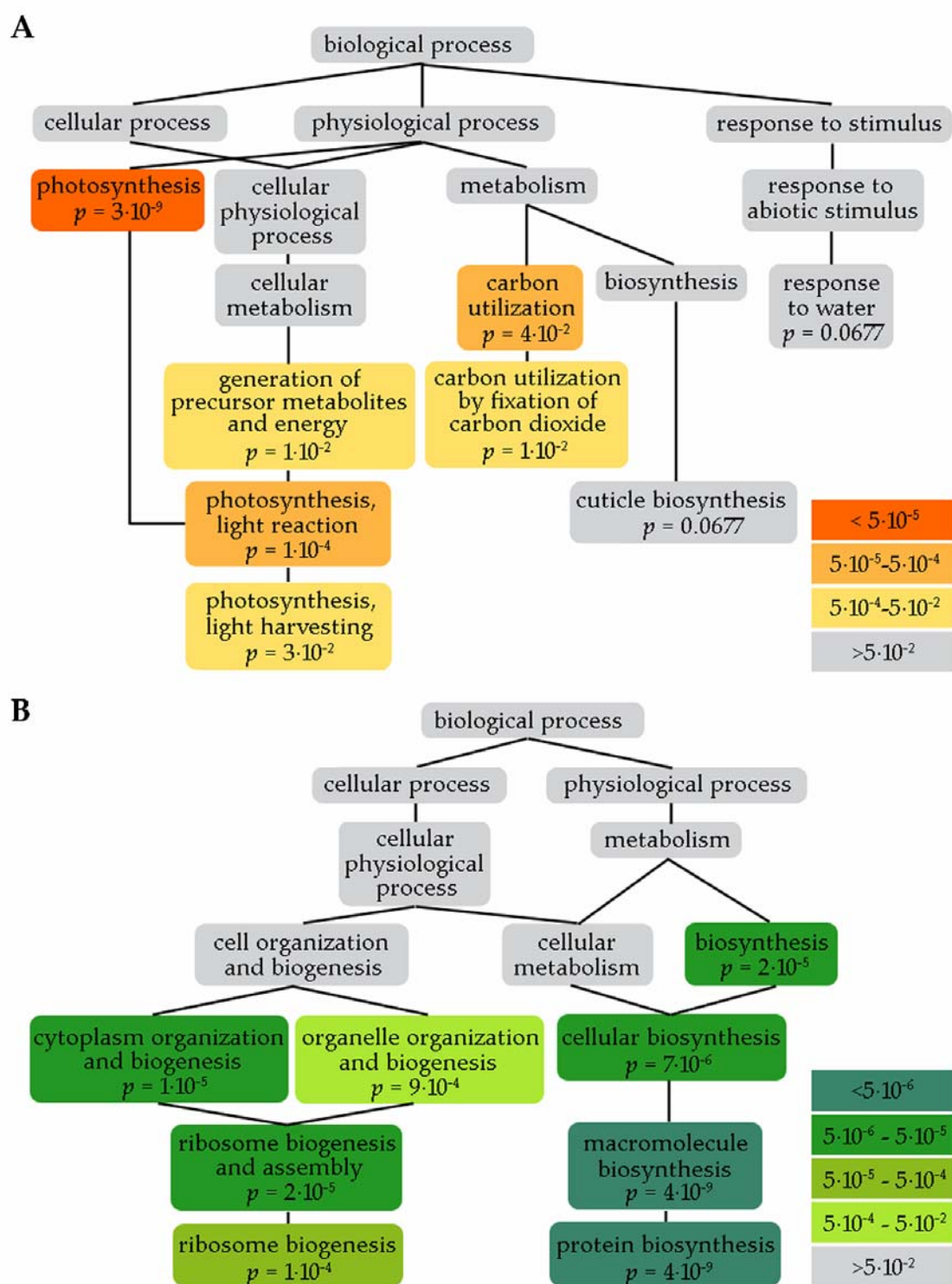


Fig. 43. Representación jerárquica de las ontologías génicas de la componente 'proceso biológico' significativamente sobre-representados en el conjunto de genes inducidos (A) y reprimidos (B) realizada a partir de los resultados obtenidos utilizando la herramienta FatiGO. Las categorías significativas (valor p de un test de Fisher corregido por test múltiples ≤ 0.05) se muestran utilizando una escala de colores de acuerdo a su nivel de significatividad. Las categorías necesarias para completar la jerarquía se muestran en gris.

Las categorías GO identificadas como significativamente sobre-representadas en el conjunto de genes inducidos (Fig. 43A, Tabla Suplementaria XXX) fueron 'fotosíntesis, captación de luz' (GO:0009765, $p = 0.03$) y 'utilización de carbono por fijación de dióxido de carbono' (GO:0015977, $p = 0.01$), mientras que en el conjunto de genes reprimidos encontramos sobre-representadas las categorías de 'biosíntesis de proteínas' (GO:0006412, $p = 4 \cdot 10^{-9}$) y 'biogénesis de ribosomas' (GO:0007046, $p = 1 \cdot 10^{-4}$) (Fig. 43B, Tabla Suplementaria XXXI). Las categorías GO correspondientes a 'respuesta a agua' (GO:0009414, $p = 0.0677$) y 'biosíntesis de la cutícula' (GO:0042335, $p = 0.0677$) (Tabla Suplementaria XXX), se encontraron cerca del límite de significatividad establecido ($p \leq 0.05$) por lo que se incluyeron en la Fig. 43A y se discutirán más adelante.

1.3.1. Fotosíntesis y Utilización de Carbono

En las plantas *CcGA20ox* se encontraron inducidos diversos genes incluidos en las categorías de 'captación de luz' (ej. proteínas de unión a clorofila a/b) y en las reacciones de transferencia de electrones (ej. ferredoxina y fotosistemas I y II), así como en las categorías de 'generación de metabolitos precursores y energía' (ej. glicolato oxidasa, fructosa-bifosfato aldolasa, tioredoxinas y ferredoxinas) y en la utilización de carbono (ej. anhidrasas carbónicas, RuBisCo y gliceraldehido 3-fosfato deshidrogenasas) (Fig. 43A, Tabla Suplementaria XXX). Dado que la categoría de 'utilización de carbono por fijación de dióxido de carbono' apareció significativamente sobre-representada, quisimos saber si otros genes del ciclo de Calvin-Benson, además de los presentes en la categoría enriquecida, estaban también inducidos. Tal como se muestra en la Figura 44, la mayoría de los genes implicados en el ciclo de Calvin-Benson estaban inducidos en las plantas *CcGA20ox*, incluyendo RuBisCo (2.7 veces, $p < 0.05$), GADPH (2.6 veces, $p < 0.05$), fructosa-bifosfato aldolasa (3.9 veces, $p < 0.05$) y fructosa-bifosfatasa (1.9 veces, $p < 0.05$). Los genes que codifican fosforibuloquinasa (3.7 veces, $p = 0.059$) y transquetolasa (1.5 veces, $p < 0.05$) se consideraron inducidos, aunque no cumplían los valores establecidos para identificarlos como diferencialmente expresados. Adicionalmente, dos unigenes de cítricos que codifican RubisCo activasa, enzima que regula la actividad de la RuBisCo, se encontraron asimismo inducidos (4.8 veces, $p < 0.05$) en las plantas *CcGA20ox*. Los posibles cambios en la expresión de otros genes implicados en dicha ruta, que codifican fosfoglicerato kinasa, sedoheptulosa-bifosfatasa y ribulosa-5-fosfato isomerasa, no pudieron ser evaluados porque estos genes no se encuentran

representados en la micromatriz CIT1. Estos resultados sugieren que la sobreexpresión de *CcGA20ox1* provoca un aumento en la capacidad de fijación de carbono de las plantas transgénicas.

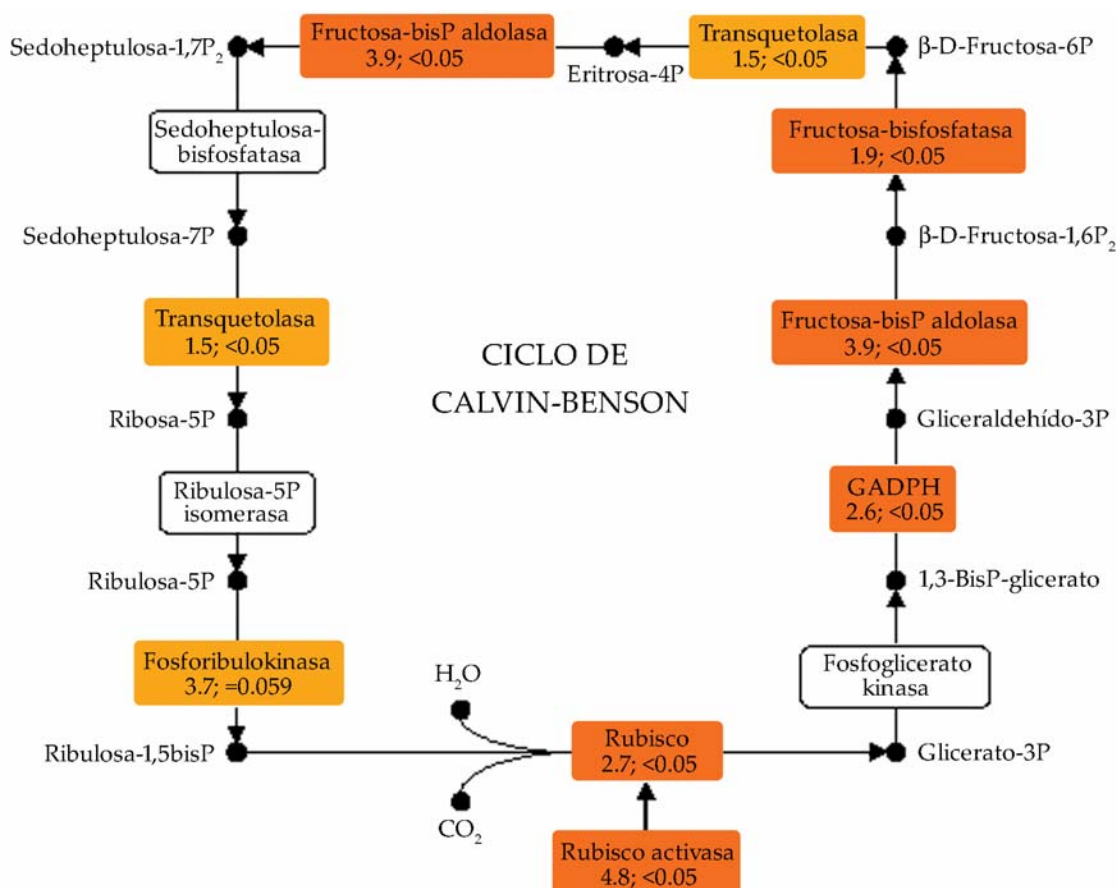


Fig. 44. Cambios en los niveles de transcritos de los genes implicados en el ciclo de Calvin-Benson en entrenados de citrange Carrizo entre las plantas *CcGA20ox* y las plantas control. Los genes diferencialmente expresados (al menos un cambio de expresión de 1.6 veces y un valor $p \leq 0.05$) se muestran en cajas de color rojo oscuro. Las cajas de color rojo más claro corresponden a genes inducidos aunque no cumplen uno de los dos criterios.

Hasta el momento, no existen evidencias previas que muestren una activación transcripcional de los genes relacionados con la fotosíntesis y la fijación de carbono en plantas transgénicas con el metabolismo de GAs modificado. Sin embargo, diversos trabajos han señalado que el contenido de algunas proteínas implicadas en el ciclo de Calvin-Benson y en la utilización de carbono aumenta tras la aplicación de GA_3 . Los niveles de RuBisCo en hojas de haba y soja son mayores tras 1 h de tratamiento con GA_3 como resultado de un aumento en la traducción (Yuan y Xu, 2001). Los niveles de fructosa-bifosfato aldolasa aumentan en raíces de arroz 24 h después de la aplicación de GA_3 (Konishi et al, 2004). Asimismo, la anhidrasa carbónica (que cataliza la hidratación reversible de CO_2 y, por tanto, la disponibilidad de CO_2 para RuBisCo)

aumenta en hojas de *Brassica juncea* tras el tratamiento con GA₃ (Hayat et al, 2001). Además, se ha encontrado que RuBisCo activasa tiene propiedades de unión a GAs (Komatsu et al, 1996) y que su sobre-expresión en arroz promueve un mayor crecimiento de la planta (Sharma and Komatsu, 2002).

1.3.2. Biosíntesis de Proteínas

Las categorías de 'biosíntesis de proteínas' y 'biogénesis de ribosomas' incluyeron, principalmente, genes que codifican proteínas ribosomales (al menos 25 pertenecientes a la subunidad ribosomal 60S y 15 a la subunidad 40S) (Fig. 43B, Tabla Suplementaria XXXI). Adicionalmente, en la categoría de 'biosíntesis de proteínas' se encontraron tres factores de elongación (Tabla Suplementaria XXXI). La amplia represión de genes que corresponden a estas categorías funcionales sugiere que la maquinaria de síntesis de proteínas puede estar afectada en las plantas *CcGA20ox*. Sin embargo, la menor expresión de los genes incluidos en la 'biosíntesis de proteínas' y en la 'biogénesis de ribosomas' no parece afectar a la síntesis de todas las proteínas porque al menos las implicadas en la fotosíntesis y en la fijación de carbono probablemente están aumentadas. La reducción en la expresión de los genes englobados en estas categorías puede interpretarse como un cambio en el patrón de síntesis de proteínas. De acuerdo con esta hipótesis, se ha mostrado que, en células de aleurona de cebada tratadas con GA₃, la síntesis proteica global es modificada para producir importantes cantidades de α -amilasa y otras proteínas inducidas por GA₃, mientras que la síntesis de otras proteínas queda reducida o detenida (Mozer, 1980; Jacobsen and Beach, 1985).

La categoría de 'biosíntesis celular' también se encontró sobre-representada en el conjunto de genes reprimidos en las plantas *CcGA20ox* (Fig. 43B), e incluyó genes que codifican limoneno sintasa (que cataliza la ciclación de geranyl difosfato en la biosíntesis de monoterpenos), sesquiterpeno sintasa (relacionada con la biosíntesis de sesquiterpenos), y flavonol 3-O-metiltransferasa, chalcona sintasa, y una oxidoreductasa 2-oxoglutarato-Fe(II) oxigenasa (estos tres últimos genes están implicados en la biosíntesis de flavonoides) (Tabla Suplementaria XXXI). La represión de todos estos genes es interesante debido a la relevancia de las rutas de biosíntesis de isoprenoides y flavonoides en cítricos (Rodrigo y Zacarías, 2006).

1.4. La activación transcripcional de los genes relacionados con la fotosíntesis conduce a un aumento en la capacidad fotosintética de las plantas *CcGA20ox*

La inducción global observada en los genes que codifican proteínas de los fotosistemas y de unión a clorofila (Fig. 43A, Tabla Suplementaria XXX), así como en los genes implicados en la fijación de carbono (Fig. 44), nos condujo a la hipótesis de que estos cambios en la expresión génica podrían estar asociados a un aumento en la capacidad fotosintética de las plantas *CcGA20ox*. Para confirmar esta hipótesis, se midió la absorción neta de CO₂ en hojas jóvenes de plantas *CcGA20ox* y en plantas control. Como muestra la Tabla 24, la absorción fotosintética de CO₂ en hojas jóvenes de plantas *CcGA20ox* fue significativamente mayor que en las plantas control a una radiación fotosintética activa (PAR) de 800 y 1,000 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ mientras que no se encontraron diferencias significativas a 600 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ (Tabla 24). Estos resultados apoyan la conclusión de que, aunque no tenemos datos sobre la producción de biomasa de estas plantas, la inducción global de los genes que corresponden a las categorías GO sobre-representadas de 'fotosíntesis' y 'utilización de carbono' en las plantas *CcGA20ox* ciertamente tiene un efecto fisiológico.

El efecto observado de la sobre-expresión de *CcGA20ox1* sobre la fotosíntesis puede estar relacionado con que el mesófilo (parénquima en empalizada y esponjoso) de las hojas de estas plantas es más compacto (Fagoaga et al, 2007). Resulta interesante señalar que, en tabaco, se ha observado un efecto positivo de las GAs sobre la fotosíntesis neta cuando ésta se midió en plantas enteras que sobre-expresan una GA20ox, pero no cuando se midió en hojas individuales (Biemelt et al, 2004). En este caso, sin embargo, el mesófilo de las plantas transgénicas no presentaba diferencias respecto al control. El efecto de las GAs sobre la fotosíntesis ha sido ciertamente controvertido, de modo que algunos autores han encontrado que la aplicación de GA₃ tiene un efecto positivo (Yuan y Xu, 2001; Hayat et al, 2001), negativo (Dijkstra et al, 1990) o ningún efecto (Cramer et al, 1995). Estos resultados aparentemente contradictorios pueden ser debidos a los diferentes sistemas experimentales y métodos utilizados por los autores para determinar la fotosíntesis (Nagel y Lambers, 2002). La sobre-expresión de *CcGA20ox1* también incrementa significativamente tanto la conductancia estomática (g_s) como la tasa de transpiración (E), con independencia del valor de PAR (Tabla 24). Sin embargo, la eficiencia en el uso del agua, estimada como el cociente entre Pn/E, fue similar en las plantas *CcGA20ox1* y en los controles. Únicamente a un valor de PAR de 1000 este parámetro fue algo inferior en las plantas *CcGA20ox1*.

Tabla 24. Efecto de la sobre-expresión de CcGA20ox1 en la tasa fotosintética (PN, mmol CO₂ m⁻² s⁻¹), conductancia estomática (Gs, mmol m⁻² s⁻¹), transpiración (E, mmol m⁻² s⁻¹) y eficiencia en la utilización de agua (PN/E, μmol CO₂ /mmol H₂O)

Genotipo	Radiación fotosintética activa (PAR, μmol CO ₂ m ⁻² s ⁻¹)											
	600				800				1000			
	PN	Gs	E	PN/E	PN	Gs	E	PN/E	PN	Gs	E	PN/E
control	10.33 ± 0.84a,c	82.20 ± 4.33a,c	1.80 ± 0.08a,c	5.70 ± 0.29a,c	11.84 ± 1.04a,cd	70.60 ± 4.3a,c	1.57 ± 0.08a,c	7.44 ± 0.36a,d	14.45 ± 1.20a,d	80.1 ± 5.81a,c	1.65 ± 0.09a,c	8.67 ± 0.36a,e
CcGA20ox1	12.11 ± 1.53a,c	144.90 ± 20.27b,c	2.79 ± 0.28b	4.67 ± 0.68a,c	17.14 ± 1.06b,d	114.70 ± 13.07b,c	2.26 ± 0.18b,c	7.74 ± 0.36a,d	20.37 ± 1.24b,d	176.10 ± 21.94b,c	2.89 ± 0.25b,c	7.29 ± 0.43b,d

Las medidas se realizaron en hojas jóvenes totalmente expandidas en las plantas CcGA20ox y en el control. Cada valor representa la media ± error estándar de 12 medidas independientes. Los valores de cada columna y de cada fila señalados con una letra diferente (antes d e/a coma para las columnas, tras la coma para las filas) son significativamente diferentes ($p \leq 0.05$)

1.5. La sobre-expresión en sentido del gen *CcGA20ox1* provoca cambios en los niveles de expresión de genes específicos

Con el propósito de identificar otros procesos metabólicos y fisiológicos relacionados con el fenotipo de las plantas *CcGA20ox*, resultó de interés examinar los genes expresados diferencialmente en dichas plantas (Tablas Suplementarias XXVI y XXVII) que no formaban parte de las categorías enriquecidas descritas anteriormente.

1.5.1. Metabolismo de GAs

Respecto a los genes implicados en el metabolismo de GAs, la micromatriz CIT1 contiene dos unigenes, aC08003D01 y aC18016F11, cuya secuencia es similar a GA 2-oxidasa2 de *Arabidopsis* (*GA2ox2*, At1g30040). Sin embargo, ninguno de ellos mostró una expresión diferencial ni en las plantas transgénicas ni en los explantes tratados con GA₃ comparados con el control. El hecho de que estos unigenes de cítricos no parezcan inducirse por GAs indica que no están sujetos a la regulación por retroalimentación positiva, contrariamente a lo que sucede con el gen *GA2ox2* de *Arabidopsis* (Thomas et al, 1999). El único gen del metabolismo de GAs de cítricos del que se conoce su regulación por retroalimentación negativa es *CcGA20ox1* (Vidal et al, 2003), pero este gen no se encuentra en la micromatriz de cDNA de cítricos utilizada en este análisis transcriptómico. Sin embargo, los datos obtenidos mediante RT-PCR (Fig. 39B) confirmaron que en las plantas *CcGA20ox* la expresión del gen endógeno *CcGA20ox1* es muy reducida; probablemente como resultado del mecanismo de regulación por retroalimentación negativa.

1.5.2 Biosíntesis y modificación de la pared celular

La pared celular vegetal es una entidad compleja, diversa y dinámica, puesto que cambia durante los procesos de división, crecimiento y diferenciación celular. La dinámica de la pared celular depende de numerosas enzimas y proteínas estructurales, la mayoría de ellas codificadas por familias multigénicas que, con frecuencia, exhiben perfiles de expresión y funciones fisiológicas diferentes (Imoto et al, 2005). Las GAs inducen la elongación y/o división celular y se ha sugerido que las plantas *CcGA20ox* presentan un aumento en las divisiones celulares en los entrenudos en elongación (Fagoaga et al, 2007). Por lo tanto, cabía esperar que los niveles de transcritos correspondientes a las enzimas implicadas en la división celular estuvieran modificados en esas plantas. Sin embargo, en el conjunto de genes

expresados diferencialmente en las plantas *CcGA20ox* no se observaron genes relacionados con la división celular sino con la biosíntesis y modificación de la pared celular. Así pues, en las plantas *CcGA20ox* se observó la inducción de los unigenes que codifican una endo-1,4- β -D-glucanasa, requerida para la formación normal de celulosa, una proteína similar a celulosa sintasa, implicada en la síntesis de polisacáridos de la matriz (Cosgrove, 2005), y cuatro β -1,3-glucanasas, que rompen la callosa, polisacárido abundante en la placa celular de las células en división y que también se deposita tras la exposición a estrés de tipo abiótico o biótico (Scheible y Pauly, 2004) (Tabla 25). Además, se observó la inducción de tres unigenes que codifican xiloglucano endotransglucosilasa-hidrolasa (XTH), enzima que regula la extensión de la pared celular (Tabla 25). Previamente, se ha descrito un efecto positivo del tratamiento con GA₃ tanto en la actividad de XTH (Potter y Fry, 1993; Smith et al, 1996) como en la expresión de XTH (Uozu et al, 2000; Jan et al, 2004). Los resultados obtenidos indican que la elongación de los entrenudos de las plantas *CcGA20ox* es consecuencia de procesos de elongación celular, resultado tanto de la relajación de la pared celular (inducción de genes XTH) como de la síntesis de celulosa (inducción de endo-1,4- β -D-glucanasa y proteína similar a celulosa sintasa).

Tabla 25. Genes expresados diferencialmente en las plantas *CcGA20ox* implicados en la biosíntesis y modificación de la pared celular

Enzimas	Unigen de cítricos	Cambio de expresión	Gen de ATH más similar
Celulosa			
Celulosa sintasa-like (GT2)	aCL1355Contig1	2.58	At2g32540
Endo-1,4- β -glucanasa, KORRIGAN (GH9)	aCL5359Contig1	1.72	At5g49720
Hemicelulosa			
Xiloglucano endotransglucosilasa-hidrolasa (GH16)	aCL156Contig1	4.44	At1g14720
	aCL3307Contig1	3.68	At5g65730
	aCL6772Contig1	2.75	At4g37800
	aCL6641Contig1	-2.51	At5g13870
Callosa			
β -1,3-glucanasa (GH17)	aC04017A07	5.17	At4g16260
	aCL20Contig2	3.56	At3g57270
	aCL20Contig7	2.62	At3g57270
	aCL168Contig3	1.62	At2g27500
	aCL2349Contig1	-2.46	At2g16230
Pectina			
Pectin metil esterasa (CE8)	aCL3567Contig1	3.48	At2g26440
	aCL1215Contig1	2.16	At5g20860
	aCL120Contig3	-3.20	At3g14310
	aCL120Contig2	-2.68	At3g14310
	aCL120Contig1	-2.57	At3g14310
	aCL1691Contig1	-2.68	At1g11580
Pectato liasa (PL1)	aCL3827Contig2	-2.79	At1g04680
	aCL4999Contig1	-2.01	At1g04680

Finalmente, diversos genes que codifican enzimas que modifican pectinas como pectin metil esterasa (PME) y pectato liasa (PL) se expresaron diferencialmente (Tabla 25), lo que sugiere que el metabolismo de pectinas pudiera estar alterado en las plantas *CcGA20ox*. En el xilema en desarrollo de plantas transgénicas de álamo que sobre-expresan una GA 20-oxidasa de Arabidopsis (*AtGA20ox1*), se ha descrito la inducción de los genes que codifican PME y PL (Israelsson et al, 2003), lo que indica que el aumento observado en el crecimiento secundario de esas plantas requiere, probablemente, una mayor actividad de las enzimas que modifican pectinas. De acuerdo con esta idea, la represión de los genes *PME* y *PL* de cítricos podría explicar la reducción del desarrollo del xilema observada en las plantas *CcGA20ox* (Fagoaga et al, 2007).

1.5.3. Biosíntesis de lignina

La sobre-expresión del gen *AtGA20ox1* de Arabidopsis en tabaco produce un aumento en la expresión de diversos genes implicados en la biosíntesis de lignina, acompañado de un mayor contenido de lignina (Biemelt et al, 2004). Por otro lado, las plantas transgénicas de álamo que sobre-expresan el gen *AtGA20ox1* presentan más fibras de xilema y de mayor longitud, lo que sugiere que los niveles elevados de GA estimulan el crecimiento secundario (Eriksson et al, 2000), probablemente asociado a la mayor deposición de lignina en la pared celular. En el caso de cítricos, no se ha observado un aumento en el desarrollo del xilema, al menos en brotes jóvenes en desarrollo (Fagoaga et al, 2007). Además, entre los genes expresados diferencialmente en las plantas *CcGA20ox* encontramos 7 unigenes implicados en la biosíntesis de lignina, 3 inducidos y 4 reprimidos (Tabla 26), lo que no nos permite concluir si estos cambios de expresión conllevan una alteración en el contenido de lignina en cítricos.

Tabla 26. Genes expresados diferencialmente en la plantas *CcGA20ox* implicados en la biosíntesis de lignina

Enzimas	Unigen de Cítricos	Cambio de expresión	Gen de ATH más similar
Fenilalanina amonio liasa (PAL)	aC02002A11	2.28	At3g10340
Cinamato 4-hidroxilasa (C4H)	aCL959Contig1	2.04	At2g30490
4-cumarato-CoA ligasa (4CL)	aCL5408Contig1	-1.66	At1g62940
Hidroxicinamoil-CoA siquimato/quinato	aCL4633Contig1	-3.07	At5g48930
Hidroxicinamoil transferasa (HCT)	aCL6878Contig1	-2.13	At5g48930
	aCL602Contig1	-1.89	At5g48930
Cafeoil-CoA O-metiltransferasa (CCoAMT)	aCL139Contig2	1.92	At4g34050

1.5.3. Estrés de origen abiótico

Diversos trabajos sugieren que las GAs podrían estar implicadas en el estrés originado por diferentes tipos de factores abióticos (Achard et al, 2006; Magome et al, 2004). Puesto que las categorías GO de 'respuesta a agua' y 'biosíntesis de la cutícula' (proceso alterado por la sequía; Aharoni et al, 2004a) estaban casi significativamente sobre-representadas en el conjunto de los genes inducidos (Fig. 43A; Tabla Suplementaria XXX), buscamos genes expresados diferencialmente en las plantas *CcGA20ox* implicados en la respuesta originada por estrés por desecación, osmótico, salino y oxidativo (Tabla 27). En el conjunto de genes expresados diferencialmente en las plantas *CcGA20ox* (Tabla 27) encontramos que 23 unigenes (15 de ellos relacionados con la 'respuesta a agua') estaban inducidos mientras que 11 aparecieron reprimidos. Así pues, en las plantas *CcGA20ox* se observó una inducción de los genes que codifican LEA5, dehidrina, delta 1-pirrolina-5-carboxilato sintasa (P5CS) y proteínasas tipo cisteína (RD19, RD21). Con anterioridad a este trabajo, no se habían descrito cambios en la expresión de genes implicados en estrés por déficit hídrico en plantas con el metabolismo de GAs modificado. La acumulación de miembros de la familia de dehidrinas y de proteínas LEA en arroz (Xu et al, 1996), girasol (Cellier et al, 1998), cebada (Zhu et al, 2000) y trigo (Sivamani et al, 2000; López et al, 2001) así como la sobre-expresión de delta 1-pirrolina-5-carboxilato sintasa en tabaco (Kavi Kishor et al, 1995) aumenta la tolerancia a la sequía. Así pues, los resultados obtenidos en este trabajo sugieren la posibilidad de que las plantas *CcGA20ox* también puedan mostrar una mayor tolerancia al estrés originado por déficit hídrico.

La superficie de las plantas está protegida por la cutícula, una compleja estructura lipídica compuesta por una capa externa de cera (epicuticular), que constituye una barrera protectora frente al estrés debido a diferentes factores ambientales, principalmente a sequía y patógenos (Chen et al, 2003a). De acuerdo con esta idea, el análisis transcriptómico de *Populus* en respuesta a estrés originado por déficit hídrico está asociado a una regulación positiva no sólo de la ontología génica de 'respuesta a agua' sino también de 'biosíntesis de cera' y 'biosíntesis de la cutícula' (Street et al, 2006). En el caso de las plantas de cítricos *CcGA20ox* se encontraron inducidos, además de los genes implicados en la respuesta a agua (Tabla 27), otros genes relacionados con lípidos (ej. GDSL lipasas/hidrolasas y acil-ACP tioesterasa), ceras (*Wax2*) y biosíntesis de la cutícula como la enzima de condensación de ácidos grasos de cadena muy larga (VLCFA) y proteínas de

transferencia de lípidos (LTP) (Tabla 28), lo que sugiere que en estas plantas probablemente exista una modificación de la cutícula para evitar la pérdida de agua.

Tabla 27. Genes expresados diferencialmente en las plantas CcGA20ox relacionados con el estrés abiótico

Inducidos				
Descripción	Términos GO	Unigen de Cítricos	Cambio expresión	Gen de ATH más similar
Familia de proteínas dehidrina	Respuesta a agua	aCL6Contig16	3.93	At1g54410
		aCL6Contig7	3.41	At1g54410
		aCL6Contig21	3.38	At1g54410
Proteína LEA (LEA5)	Respuesta a estrés hídrico	aCL9Contig8	2.96	At4g02380
	Respuesta a estrés oxidativo	aCL6Contig22	2.61	At4g02380
Cisteína proteinasa (RD21)	Respuesta a ROS	aCL9Contig19	2.40	At4g02380
	Respuesta a estrés hídrico	aCL23Contig3	2.30	At1g47128
		aCL23Contig1	2.13	At1g47128
Delta 1-pirrolina-5-carboxilato sintasa (P5CS1)	Respuesta a estrés hídrico	aCL174Contig2	2.47	At2g39800
	Respuesta a desecación			
	Respuesta a estrés salino			
	Respuesta hiperosmótica			
9-cis-epoxycarotenoide dioxigenasa zeaxantina epoxidasa (ABA1)	Respuesta a estrés hídrico	aCL920Contig2	2.16	At3g63520
	Respuesta a estrés hídrico	aCL1551Contig1	2.18	At5g67030
	Respuesta a estrés osmótico	aCL3421Contig1	1.79	At5g67030
Proteína COR413	Respuesta a estrés hídrico	aCL5208Contig1	2.04	At1g29395
	Respuesta a ABA			
Glutación S-transferasa (GST8)	Respuesta a estrés hídrico	aCL87Contig1	1.97	At1g78380
	Respuesta a estrés oxidativo			
Cisteína proteinasa (RD19)	Respuesta a estrés hídrico	aCL96Contig1	1.68	At4g39090
	Respuesta a desecación			
	Respuesta a estrés osmótico			
Proteína RD22	Respuesta a estrés salino			
	Respuesta a desecación	aCL172Contig1	5.84	At5g25610
	Respuesta a estrés salino	aCL172Contig2	5.81	At5g25610
Proteína de dedo de zinc	Respuesta a ABA			
	Respuesta a estrés osmótico	aCL198Contig1	2.18	At2g04240
Sinaptobrevina	Respuesta a estrés salino			
	Respuesta a estrés osmótico	aCL1312Contig1	1.65	At5g22360
Isoflavona reductasa	Respuesta a estrés oxidativo	aCL4218Contig1	2.90	At1g75280
	Respuesta a estrés oxidativo			
Peroxidasa 42 (PER42)	Respuesta a estrés oxidativo	aCL36Contig3	1.73	At4g21960
		aCL36Contig2	1.63	At4g21960
Chalcona sintasa (CHS)	Respuesta a estrés oxidativo	aCL1023Contig1	2.17	At5g13930
Reprimidos				
Factor de transcripción RD26	Respuesta a estrés hídrico	aCL35Contig3	-1.69	AT4G27410
	Respuesta a ABA			
Lipoxigenasa (LOX2)	Respuesta a estrés hídrico	aCL241Contig1	-3.83	AT3G45140
	Respuesta a desecación	aCL622Contig2	-3.36	AT1G05260
Peroxidasa 3 (RCI3)	Respuesta hiperosmótica			
	Respuesta a estrés osmótico	aCL2951Contig1	-2.11	AT1G77120
Alcohol deshidrogenasa (ADH1)	Respuesta a estrés salino	aCL2656Contig1	-1.82	AT1G01060
	Respuesta a ABA			
Factor de transcripción LHY	Respuesta a estrés salino			
	Respuesta a ABA			
Factor de transcripción MYB78	Respuesta a estrés salino	aCL7866Contig1	-2.33	AT5G49620
	Respuesta a ABA			
Proteína TIP	Respuesta a estrés salino	aCL62Contig1	-2.04	AT2G36830
L-ascorbato peroxidasa 3 (APX3)	Respuesta a estrés oxidativo	aCL7975Contig1	-1.74	AT4G35000
	Respuesta a estrés oxidativo			
Peroxidasa	Respuesta a estrés oxidativo	aCL622Contig1	-2.16	AT5G15180
Chalcona sintasa (CHS)	Respuesta a estrés oxidativo	aCL27Contig2	-1.92	AT5G13930
Calreticulina 1 (CRT1)	Respuesta a estrés oxidativo	aCL1164Contig1	-2.34	AT1G56340

Nuestros resultados apoyan la conclusión de que la sobre-expresión del gen CcGA20ox1 en cítricos produce una re-programación transcripcional que coincide parcialmente con la originada por la sequía (Street et al, 2006). Sin embargo, resulta

importante señalar que una diferencia considerable entre la respuesta a sequía y a la sobre-expresión de *GA20ox* es la reducción en la fotosíntesis observada en el primer caso (Street et al, 2006), contrariamente al aumento encontrado en el segundo.

Tabla 28. Genes expresados diferencialmente en las plantas *CcGA20ox* implicados en las rutas de ácidos grasos y lípidos

Descripción	Unigen de Cítricos	Cambio de expresión	Gen de ATH más similar
GDSL lipasa/hidrolasa	aCL2438Contig1	2.40	At2g04570
	aCL273Contig1	1.67	At5g33370
	aCL3133Contig2	1.23	At3g16370
	aCL3133Contig1	1.17	At3g16370
	aC02007G06SK_c	0.84	At3g16370
	aC05014A08SK_c	-0.95	At5g55050
Acil-tioesterasa	aCL2357Contig1	0.71	At1g08510
Biosíntesis de glicerolípidos			
Glicerol-3-fosfato aciltransferasa	aCL288Contig1	1.48	At2g38110
Diacilglicerol aciltransferasa	aC04022C03SK_c	0.85	At3g51520
Biosíntesis de fosfolípidos			
Colina-fosfato citidililtransferasa	aCL3347Contig1	-0.70	At2g32260
Inositol-3-fosfato sintasa	aCL25Contig1	-1.34	At2g22240
Fosfoetanolamina N-metiltransferasa	aCL1819Contig1	-2.26	At3g18000
Biosíntesis de glicolípidos			
UDP-galactosa:MGDG galactosiltransferasa	aC02006F12SK_c	-0.76	At4g00550
Biosíntesis de ácidos grasos de cadena muy larga (VLCFA)			
Enzima de condensación de VLCFA (CUT1)	aCL2743Contig1	1.75	At1g68530
Beta-ketoacil-CoA sintasa (FID)	aCL2162Contig1	1.45	At2g26250
Beta-ketoacil-CoA sintasa	aC07007G06SK_c	0.94	At5g04530
Transporte de lípidos			
Proteína de transferencia de lípidos 1 (LTP1)	aC02009H06SK_c	0.69	At2g38540
Proteína de transferencia de lípidos 4 (LTP4)	aCL3Contig9	1.45	At5g59310
Proteína de transferencia de lípidos 4 (LTP4)	aCL3Contig25	1.44	At5g59310
Proteína de transferencia de lípidos 6 (LTP6)	aCL2327Contig1	-1.46	At3g08770
Proteína de transferencia de lípidos (LTP)	aCL12Contig7	3.54	At3g18280
	aCL12Contig9	2.99	At3g18280
	aCL12Contig8	2.67	At3g18280
	aCL369Contig2	2.06	At3g18280
	aCL2301Contig1	-1.27	At3g18280
	aCL595Contig2	-0.79	At2g10940
Biosíntesis de ceras			
Wax2	aCL2973Contig1	0.99	AT5G57800

El ácido abscísico (ABA) regula el desarrollo vegetativo en respuesta al estrés causado por diversos factores ambientales como condiciones de sequía o de elevada salinidad. Bajo condiciones de estrés por déficit hídrico, el contenido en ABA aumenta e induce el cierre de los estomas, para evitar la pérdida de agua por transpiración, a través de una compleja cascada de señalización (Blatt, 2000). Aunque el contenido en ABA en las plantas *CcGA20ox* se desconoce, la posibilidad de un aumento del mismo no parece consistente con la observación de que tanto la apertura estomática (estimada mediante la conductancia estomática) como la tasa de transpiración se encuentran aumentadas en las hojas de las plantas transgénicas de cítricos

(Tabla 24). Además la expresión de dos genes que codifican las enzimas clave en la biosíntesis de ABA (NCED3 y ABA2; Nambara y Marion-Poll, 2005), presentes en la micromatriz CIT1, no se vió afectada (datos no presentados); mientras que tres factores de transcripción, que responden a ABA en *Arabidopsis* (uno de tipo NAC y dos MYB) se encontraron reprimidos en las plantas *CcGA20ox* (Tabla 27). Por lo tanto, los cambios de expresión observados en los genes relacionados con estrés por déficit hídrico no parecen estar mediados por ABA en las plantas *CcGA20ox*.

1.5.5. Estrés de origen biótico

Diversos tipos de genes relacionados con el estrés producido por factores bióticos (ej. 6 quitinasas y 10 inhibidores de proteasas de tipo Kunitz) aparecieron reprimidos en las plantas *CcGA20ox* (Tabla Suplementaria XXVII). Las quitinasas son conocidas proteínas PR (*pathogenesis-related*) codificadas por familias multigénicas (Neuhaus, 1999), y su sobre-expresión en plantas transgénicas pueden aumentar la protección frente a hongos fitopatógenos (Kim et al, 2003). Nuestros resultados sugieren que una posible función de la producción de GAs en plantas de arroz infectadas con *Gibberella fujikuroi*, que causa una enfermedad llamada 'bakanae' que condujo al descubrimiento de las GAs (Sponsel y Hedden, 2004), podría ser la de facilitar la infección de dicho hongo por una disminución en la expresión de las quitinasas de la planta.

1.5.6 Factores de transcripción

Los genes que codifican factores de transcripción constituyen una parte importante del genoma de los organismos eucariotas, y la mayoría de estos genes pueden agruparse en diferentes familias génicas según el tipo de dominio de unión al DNA (Riechmann y Ratcliffe, 2000). Los factores de transcripción integran la expresión de otros genes en respuesta a diversos estímulos ambientales y estados de desarrollo, aunque se desconocen muchos de los genes regulados por los factores de transcripción. La Tabla 29 resume los cambios de expresión de los genes de cítricos que codifican posibles factores de transcripción y que muestran una expresión diferencial en las plantas *CcGA20ox*, 11 de ellos inducidos y 8 reprimidos.

Muchos de los factores de transcripción codificados por genes expresados diferencialmente en las plantas *CcGA20ox* parecen regular procesos relacionados con las alteraciones fenotípicas observadas en dichas plantas (por ejemplo, cambio en la arquitectura de la planta y resistencia a estrés). Así pues, entre los genes inducidos

que codifican factores de transcripción encontramos los correspondientes a *BEL1-like homeodomain1* (BLH1) (la mutación de este gen causa una fenotipo enano en *Arabidopsis*; Bhatt, 2004), a WRKY (muchos WRKY tienen funciones de regulación en respuesta a la infección por patógenos y otros tipos de estrés; Eulgem et al, 2000), a bHLH (en concreto, bHLH062, que responde a diversos tipos de estrés; Heim et al, 2003) y a YABBY.

Tabla 29. Genes expresados diferencialmente en las plantas *CcGA20ox* que codifican posibles factores de transcripción

Descripción	Unigen de cítricos	Cambio de expresión	Gen de ATH más similar
Factor de transcripción de la familia IIB (TFIIB)	aCL656Contig3	2.41	At4g36650
Proteína con un grupo B de alta movilidad (HMGB4)	aCL138Contig4	-1.60	At2g17560
Subfamilia relacionada con kinasa CDC2 (AFC1)	aCL3136Contig1	-6.77	At3g53570
Familia de Factores de Transcripción MYB			
Factor de transcripción de la familia Myb (MYB52)	aCL5017Contig1	1.94	At1g17950
Factor de transcripción de la familia Myb (MYB121)	aCL2843Contig1	-3.13	At3g30210
Factor de transcripción de la familia Myb (MYB78)	aCL7866Contig1	-2.33	At5g49620
Factor de transcripción relacionado con Myb (CCA1)	aCL2656Contig1	-1.82	At1g01060
Familia de Factores de Transcripción CCAAT-HAP5			
Heme activated protein (HAP5c)	aCL665Contig2	1.92	At1g08970
Familia de Factores de Transcripción C2C2-YABBY			
Factor de transcripción de la familia YABBY	aCL4648Contig1	1.78	At2g26580
Familia de Factores de transcripción Homeobox			
BEL1-like homeodomain 1 (BLH1)	aCL157Contig1	2.81	At2g35940
BEL1-like homeodomain 1 (BLH1)	aCL1577Contig1	1.94	At2g35940
Class II knotted1-like homeobox (KNAT3)	aCL1472Contig1	1.94	At5g25220
Familia de Factores de transcripción bHLH			
Basic helix-loop-helix (bHLH) family protein	aCL9380Contig1	1.68	At3g07340
Familia de Factores de transcripción WRKY			
Factor transcripción de la familia WRKY (WRKY31)	aCL1201Contig1	1.64	At4g22070
Familia de Factores de transcripción ABI3VP1			
Factor de transcripción de la familia B3 (VRN1)	aCL7325Contig1	1.70	At3g18990
Familia de Factores de transcripción bZIP			
Factor de transcripción de la familia bZIP	aCL6889Contig1	-5.32	At1g08320
Familia de Factores de transcripción C2H2			
Factor transcripción familia dedo de zinc (SUF4)	aCL335Contig1	-1.62	At1g30970
Familia de Factores de transcripción NAC			
Factor de transcripción de la familia NAC (RD26)	aCL35Contig3	-1.69	At4g27410
Response Regulador Gene Family			
pseudo-response regulator 5 (APRR5)	aCL5406Contig1	1.60	At5g24470

En el conjunto de genes reprimidos que codifican factores de transcripción encontramos una proteína HMGB (*high mobility group B*) (cuya sobre-expresión en *Arabidopsis* altera el crecimiento de la plántula bajo diferentes condiciones de estrés; Kwak et al, 2007), un factor de transcripción de tipo NAC (RD26) (implicado en la respuesta a la desecación en *Arabidopsis*; Yamaguchi-Shinozaki et al, 1992) y varios miembros de la familia de factores de transcripción de tipo MYB (CCA2 y MYB78 y MYB121, que responden a estrés salino y a ABA en *Arabidopsis*, respectivamente; Yanhui et al, 2006). Recientemente, se ha demostrado que un gen que codifica el

factor de transcripción tipo YABBY (*YAB1*) de arroz, con elevada similitud de secuencia al inducido en las plantas *CcGA20ox* (Tabla 29), está implicado en la regulación negativa de la biosíntesis de GA (Dai et al, 2007).

2. CAMBIOS EN LA EXPRESIÓN GÉNICA DE ENTRENUDOS DE EXPLANTES TRATADOS CON GA₃

2.1. La aplicación de GA₃ a tiempos cortos altera el transcriptoma de los entrenudos de explantes de cítricos

Con objeto de comprender mejor cómo controlan las GAs el desarrollo de los brotes vegetativos en citrange Carrizo, se diseñó un experimento para conocer cómo se ve alterado el transcriptoma tras una aplicación de GA₃ a tiempos cortos. El tratamiento con GA₃ regula negativamente la expresión del gen *CcGA20ox1* en cítricos, mientras que el tratamiento con paclobutrazol (PCB) tiene el efecto contrario (Vidal et al, 2003). La expresión de dicho gen se analizó, mediante RT-PCR, en entrenudos procedentes de brotes jóvenes de citrange Carrizo, en el mismo estadio de desarrollo que los utilizados en el análisis del transcriptoma anteriormente descrito, cultivados en presencia de 10 μ M GA₃ y recogidos 6, 12 y 24 h después del tratamiento con GA₃. Los resultados mostraron una clara reducción de los niveles de transcritos de *CcGA20ox1* tras 24 h de tratamiento con GA₃ (Fig. 45), lo que indica que GA₃ había sido transportado y metabolizado eficientemente en los explantes tratados con GA₃. Por otro lado, el tratamiento con PCB aumentó ligeramente los niveles de expresión del gen *CcGA20ox1* (Fig. 45), aunque el efecto no fue tan claro como el observado con GA₃. Por este motivo, en el análisis del transcriptoma se utilizaron entrenudos procedentes de explantes recogidos tras 24 h de la aplicación de GA₃.

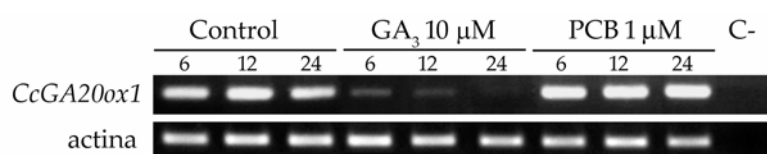


Fig. 45. Análisis de la expresión del gen *CcGA20ox1* en entrenudos de citrange Carrizo procedentes de explantes recogidos tras 6, 12 y 24 h del tratamiento con 10 μ M de GA₃ o con 1 μ M de PCB. Los niveles de transcritos se determinaron mediante RT-PCR semicuantitativa, tal como se describe en Materiales y Métodos, utilizando una actina de cítricos (CX289161) como control interno.

El diseño del experimento de análisis del transcriptoma de entrenudos de citrange Carrizo procedentes de explantes recogidos a las 24 h del tratamiento con $10 \mu\text{M}$ GA_3 consistió en dos comparaciones directas, tiempo 24 h frente a tiempo 0 h y tiempo 24 h con tratamiento de GA_3 frente a tiempo 0 h, de modo que fuera posible estimar el efecto tiempo y el efecto tratamiento (Fig. 46). Se realizaron 4 réplicas biológicas, cada una de ellas formada por entrenudos procedentes de 5 explantes, para cada una de las tres condiciones: tiempo 0 h, tiempo 24 h y tiempo 24 h con tratamiento de GA_3 . El RNA de las réplicas biológicas correspondientes al tiempo 0 h se juntó y se utilizó como material de referencia (Fig. 46).

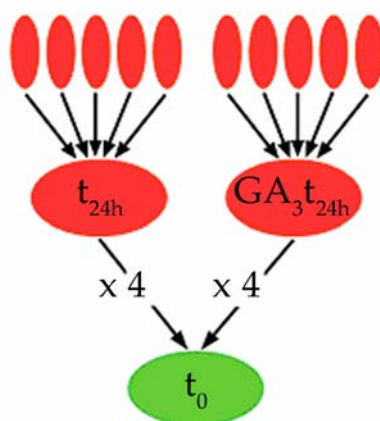


Fig. 46. Diseño del experimento. El efecto de la aplicación de GA_3 en el transcriptoma de entrenudos procedentes de brotes en desarrollo de citrange Carrizo se obtuvo de la doble comparación 'muestras procedentes de entrenudos recogidos a las 24 h de la aplicación de GA_3 , marcadas con Cy5, con muestras control, marcadas con Cy3' y 'muestras procedentes de entrenudos recogidos a las 24 h, marcadas con Cy5, con muestras control, marcadas con Cy3'.

La identificación de los genes expresados diferencialmente se realizó mediante el análisis SAM (Tusher et al, 2001). Los genes regulados por GAs se identificaron utilizando un límite de FDR del 5% y un valor de cambio de expresión de al menos 1.6. De este modo, se identificaron 123 ESTs significativamente reguladas en respuesta al tratamiento de GA_3 que correspondieron a 75 unigenes, 26 inducidos y 49 reprimidos (Tabla 30). Tal como sucedía en el análisis transcriptómico de las plantas transgénicas, un 14.7% de los genes expresados diferencialmente (11 genes) no mostraron similitud de secuencia significativa con ninguna proteína de Arabidopsis (Tabla 30). En este caso también se realizó un análisis de enriquecimiento funcional con FatiGO (Al-Shahrour et al, 2004) pero no se encontró ninguna categoría significativamente sobre-representada. Los genes expresados diferencialmente tras el tratamiento con GA_3 se clasificaron en categorías MIPS con la intención de

comprender su posible función biológica (Tabla Suplementaria XXXII para los genes inducidos y Tabla Suplementaria XXXIII para los reprimidos).

Tal como sucedía en las plantas *CcGA20ox* (Tabla Suplementaria XXVII), diversos genes implicados en la síntesis de proteínas (dos que codifican proteínas ribosomales y uno que codifica un factor de iniciación de la traducción) también aparecieron reprimidos tras la aplicación de GA₃ (Tabla 30). Estos resultados apoyan la idea de que las GAs parecen inducir una traducción diferencial de las proteínas. Sin embargo, en contraste con las plantas *CcGA20ox*, no se detectó un aumento en los transcritos de RuBisCo tras 24h de la aplicación de GA₃ utilizando explantes de cítricos cultivados *in vitro*, así como de ningún otro gen implicado en la fijación de carbono. Esto significa que el aumento de la expresión de los genes relacionados con la fotosíntesis y la fijación de carbono en las plantas *CcGA20ox* fue, probablemente, resultado de un efecto a largo plazo de la modificación de los niveles de GAs en las plantas transgénicas.

Tabla 30. Genes expresados diferencialmente tras la aplicación de GA₃

Inducidos					
Unigen de Cítricos	Cambio expresión	Valor q (%)		Gen de Arabidopsis más similar	Valor E
aCL32Contig2	4.16	0.211	At1g19640	Ácido jasmónico carboxil metiltransferasa (JMT)	3E-79
aCL32Contig1	2.99	0.211-1.137	At1g19640	Ácido jasmónico carboxil metiltransferasa (JMT)	2E-90
aCL25Contig1	2.79	0.211	At2g22240	Inositol-3-fosfato sintasa (IPS 2)	0
aCL76Contig1	2.55	0.211-3.399	At4g11650	Proteína similar a osmotina (OSM34)	3E-78
aCL2301Contig1	2.42	0.211	At3g18280	Proteína de transferencia de lípidos (LTP)	7E-27
aCL5030Contig1	2.29	1.628	At4g11650	Proteína similar a osmotina (OSM34)	3E-86
aCL5853Contig1	2.12	0.600	At1g30820	CTP sintasa	1E-136
aCL3705Contig1	1.87	0.211	At2g01570	Proteína de respuesta a GAs (RGA1)	0
aCL3827Contig2	1.86	0.211	At1g04680	Pectato liasa	1E-102
aCL39Contig3	1.80	0.211			
aCL39Contig2	1.79	0.211			
aCL23Contig1	1.75	0.211-1.137	At1g47128	Cisteína proteinasa (RD21A)	0
aCL37Contig2	1.74	0.211	At2g32210	Proteína de función desconocida	2E-11
aCL39Contig1	1.70	0.211-1.628			
aCL23Contig3	1.68	0.211	At1g47128	Cisteína proteinasa (RD21A)	0
aCL12Contig4	1.67	0.211	At5g26330	Plastocianina	2E-20
aCL5Contig2	1.66	0.211			
aCL6Contig7	1.65	1.628-2.236	At1g54410	Dehidrina	8E-09
aCL5Contig3	1.64	0.597-0.600	At5g47560	Transportador de malato	0
aCL2773Contig1	1.64	0.211	At2g42590	Proteína 14-3-3 (GRF12)	1E-113
aCL6Contig16	1.64	0.958-3.399	At1g54410	Dehidrina	8E-09
aCL5Contig4	1.62	1.628	At5g05580	Omega-3 ácido graso desaturasa (FAD8)	6E-60
aCL129Contig1	1.61	1.138	At2g03440	Proteína relacionada con nodulina	6E-07
aCL5Contig7	1.61	0.852	At1g05870	Proteína de función desconocida	2E-49
aCL6Contig21	1.61	1.628	At1g54410	Dehidrina	8E-09
aCL4194Contig1	1.60	1.138	At1g77280	Proteína kinasa	0

Reprimidos						
aC08007H11	-4.54	0.211				
aCL190Contig4	-3.78	0.211	At1g17860	Proteína de la familia Kunitz		4E-26
aCL10Contig1	-2.04	0.211	At3g12500	Endoquitinasa básica		3E-87
aCL251Contig1	-2.02	0.211				
aCL10Contig3	-1.95	0.211	At3g12500	Endoquitinasa básica		1E-100
aCL10Contig9	-1.94	0.211-2.356	At3g12500	Endoquitinasa básica		3E-82
aC08001B11	-1.86	0.211	At1g09630	Proteína Ras de unión a GTP		5E-25
aCL1Contig33	-1.81	1.483	At1g17860	Proteína de la familia Kunitz		3E-26
aCL11Contig1	-1.81	0.211-1.483	At5g06740	Proteína lectina kinasa		8E-14
aC08008G03	-1.81	0.211	At4g17380	Proteína MSH2 (MSH2)		3E-18
aC08001H03	-1.79	0.211	At4g09340	Proteína SPRY		2E-20
aCL10Contig7	-1.77	0.363-1.628	At3g12500	Endoquitinasa básica		1E-100
aCL2152Contig1	-1.77	0.211	At1g10840	Factor 3 de iniciación de la traducción		1E-164
aC08015B11	-1.74	0.211				
aCL11Contig4	-1.73	0.211	At5g03140	Proteína lectina kinasa		3E-08
aCL10Contig4	-1.72	0.211-1.483	At3g12500	Endoquitinasa básica		5E-87
aCL4076Contig1	-1.72	0.211	At2g20340	Tirosina descarboxilasa		1E-170
aCL5231Contig1	-1.71	0.211				
aCL7529Contig1	-1.71	0.211	At3g29450	Proteína de función desconocida		4E-09
aCL2797Contig2	-1.68	0.211	At1g04300	Proteína mefrina		3E-57
aCL1Contig10	-1.68	1.628	At1g17860	Proteína de la familia Kunitz		9E-24
aCL4111Contig1	-1.68	0.211	At2g42710	Proteína ribosomal L1		2E-64
aC03002C05	-1.68	0.211	At1g78280	Factor de transcripción jumonji (jmiC)		6E-12
aC08013E08	-1.67	0.211				
aCL7581Contig1	-1.67	0.211	At1g62710	Enzima vacuolar (beta-VPE)		2E-32
aC02013A07	-1.67	0.211	At5g55760	Regulador transcripcional Sir2		1E-103
aCL6631Contig1	-1.66	0.211	At2g42730	Proteína F-box		2E-19
aCL829Contig1	-1.66	0.211	At1g70540	Inhibidor de invertasa		4E-20
aCL8391Contig1	-1.66	0.211	At5g52590	Proteína RabGAP/TBC		1E-129
aCL7155Contig1	-1.66	0.211	At3g52500	Aspartil proteasa		1E-07
aCL7099Contig1	-1.66	0.211	At3g47160	Proteína de función desconocida		4E-46
aCL705Contig1	-1.65	0.597	At3g26740	Proteína de respuesta a luz CCR		6E-30
aC02009D08	-1.65	0.363	At2g26790	Pentatricopéptido (PPR)		1E-26
aCL1656Contig1	-1.65	0.211	At1g49320	Proteína con dominio BURP		1E-57
aC16011H07	-1.65	0.211	At1g01490	Proteína de unión a metales pesados		9E-08
aC02019F03	-1.64	0.211				
aC02002A01	-1.64	0.211	At1g21410	Proteína F-box		1E-07
aCL9011Contig1	-1.64	0.211	At5g24560	Proteína F-box		2E-36
aC05004G01	-1.64	0.211				
aCL5247Contig1	-1.64	0.211	At3g05590	Proteína ribosomal 60S (RPL18B)		1E-41
aCL16Contig1	-1.64	0.211-0.363	At5g54770	Tiazol (ARA6)		1E-151
aC02019D11	-1.64	0.211	At4g22720	Glicoproteasa M22		2E-20
aCL3781Contig1	-1.63	0.852	At5g42740	Glucosa-6-fosfato isomerasa (PGIC)		1E-131
aC16011E01	-1.62	0.726	At5g65690	Fosfoenolpiruvato carboxikinasa (ATP)		2E-11
aC01009G11	-1.62	0.211	At5g48790	Proteína de función desconocida		8E-56
aCL3172Contig1	-1.62	0.211	At5g41990	Proteína kinasa		1E-60
aCL7137Contig1	-1.62	0.211	At5g59380	Proteína con dominio de unión a metil-CpG		3E-06
aCL7667Contig1	-1.61	0.211	At2g35680	Fosfatasa		1E-18
aC08013F08	-1.61	0.211	At3g14290	20S proteasoma (PAE2)		2E-69

El resultado más destacado obtenido con los entrenudos tratados con GA₃ fue que la mayoría de los genes expresados diferencialmente estaban implicados en estrés de origen tanto abiótico como biótico (Tabla 30), efecto que también se observó en el análisis transcriptómico de las plantas *CcGA20ox*. En el caso de los genes inducidos, se encontró que de un total de 22 genes, 14 de ellos codifican proteínas implicadas en la respuesta a estrés hídrico (dehidrinas y cisteína proteinasa), en la percepción y en la respuesta a la temperatura (inositol-3-fosfato sintasa y omega-3 ácido graso desaturasa), y en la respuesta de la planta a estímulos bióticos y defensa (proteína similar a osmotina). Además, los genes correspondientes a dos grupos de proteínas implicadas en los mecanismos de defensa de la planta (endoquitinasas básicas y proteínas de la familia Kunitz) estaban reprimidos, tal como sucedía en las plantas *CcGA20ox*.

Las proteínas DELLA están localizadas en el núcleo y son reguladores negativos de la transducción de señal de GAs cuya degradación está promovida por las propias GAs (Fleet and Sun, 2005). Recientemente, se ha propuesto que las proteínas DELLA son integradores de la respuesta a señales ambientales en *Arabidopsis* (Achard et al, 2006). De acuerdo con esta idea, sugerimos que la sobre-representación de genes de respuesta a estrés hídrico (Tabla Suplementaria XXX y Fig. 43A) así como la expresión diferencial de otros genes relacionados con el estrés de origen biótico (Tabla Suplementaria XXVII y XXXIII) y abiótico (Tabla 27) puede ser consecuencia de un descenso en el contenido de proteínas DELLA causado por los niveles elevados de GAs. No obstante, esta hipótesis debe ser confirmada comprobando los niveles de proteínas DELLA en las plantas *CcGA20ox*. Desafortunadamente, todavía no se conoce la complejidad de la familia DELLA en cítricos, ni se dispone de herramientas moleculares para evaluar los niveles de proteínas DELLA en las plantas *CcGA20ox* o en respuesta a estrés en cítricos. En este sentido, resulta interesante destacar que el unigen de cítricos aCL3705Contig1, que codifica una proteína similar a RGA1 de *Arabidopsis* (At2g01570), no mostró cambios de expresión en las plantas *CcGA20ox* pero se observó una inducción significativa en los entrenudos tratados con GA₃ (Tabla Suplementaria XXXII). Así pues, el gen *RGA1* de cítricos parece estar regulado positivamente a nivel de mRNA tras la aplicación a tiempos cortos de GAs, de acuerdo con los resultados obtenidos previamente con el gen *OsGAI* de arroz (Ogawa et al, 2000).



Conclusiones

CONCLUSIONES

1. La construcción de la genoteca de cDNA Veg1, la generación de 1,689 ESTs y el análisis funcional de dicha colección permitió caracterizar los componentes de la cadena de transporte electrónico y los complejos antena de captación de luz en cítricos. El análisis de los datos de expresión puso de manifiesto dos grupos de genes *Lhc*; uno que codifica proteínas Lhc abundantes, cuya expresión fue diferencial en los tejidos vegetativos, y otro que codifica proteínas Lhc poco abundantes, cuya expresión sólo se detectó en los tejidos vegetativos bajo condiciones de estrés. Este patrón de expresión parece indicar una regulación distinta de los dos grupos de genes *Lhc*.
2. El análisis funcional de la colección de 85,965 ESTs del Proyecto de Genómica Funcional de Cítricos (CFGP) permitió identificar la mayoría de los genes del metabolismo de isoprenoides y flavonoides en cítricos. La mayor parte de los genes de biosíntesis de fitohormonas derivadas de isoprenoides, como giberelinas, brasinosteroides y ácido abscísico no habían sido caracterizados antes en cítricos. Por otro lado, la identificación de los genes de biosíntesis de monoterpenos y sesquiterpenos, responsables del aroma tanto de las hojas como del fruto, mostró que la familia de terpeno sintasas en cítricos es una de las más amplias y diversas en plantas.
3. Los datos de expresión obtenidos en la caracterización técnica de la micromatriz CIT1 demostraron que los resultados derivados de la plataforma de micromatrices generada en el CFGP son reproducibles.
4. El análisis global de la expresión génica de diferentes tejidos de cítricos mediante la micromatriz CIT1 mostró que la raíz es el órgano con un mayor número de genes expresado diferencialmente, seguida de los tejidos reproductivos y los vegetativos, lo que posiblemente refleja la naturaleza más especializada de la raíz y de los órganos reproductivos.
5. El aumento del contenido de GAs, producido por la sobre-expresión del gen *CcGA20ox1* produjo una reorganización del transcriptoma en los entrenudos de citrange Carrizo. En este trabajo, se describe por primera vez que, plantas con un contenido elevado de GAs poseen una inducción global de los genes implicados en la fotosíntesis y en el ciclo de Calvin. Esta manipulación genética

del metabolismo de GAs está relacionada con un aumento en la capacidad de fijación fotosintética de CO₂ de las plantas transgénicas. La represión general de los genes relacionados con la síntesis de proteínas y con la biogénesis de ribosomas, que no había sido descrita previamente, sugiere que dichos procesos también están regulados por GAs.

6. El análisis del transcriptoma de entrenudos tras 24 h de aplicación de GA₃ mostró pocos cambios. La aplicación a corto plazo de GA₃ no tuvo efecto sobre la fotosíntesis, por lo que la activación transcripcional observada en las plantas transgénicas es, probablemente, el resultado de una adaptación a largo plazo a la modificación del contenido de GAs. Por el contrario, la inducción de genes relacionados con estrés de origen abiótico y la represión de genes implicados en estrés de origen biótico fue observada en ambos casos, lo que sugiere que la regulación de la expresión de estos genes debe ser un efecto temprano del modo de acción de GAs en cítricos.



Bibliografía

BIBLIOGRAFÍA

- Aarts MG, Keijzer CJ, Stiekema WJ, Pereira A** (1995) Molecular characterization of the *CER1* gene of Arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* **7**: 2115-2127
- Achard P, Cheng H, De Grauwe L, Decat J, Schoutteten H, Moritz T, Van Der Straeten D, Peng J, Harberd NP** (2006) Integration of plant responses to environmentally activated phytohormonal signals. *Science* **5757**: 91-94
- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al** (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**: 1651-1656
- Agustí M** (2000) Citricultura. Ed. Mundi-Prensa, Madrid, España
- Agustí M, Zaragoza S, Bleiholder H, Buhr L, Hack H, Klose R, Stauß R** (1995) Escala BBCH para la descripción de los estadios fenológicos del desarrollo de los agrios (Gén. *Citrus*). *Levante Agrícola* **332**: 189-199
- Aharoni A, Dixit S, Jetter R, Thoenes E, van Arkel G, Pereira A** (2004a) The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in Arabidopsis. *Plant Cell* **16**: 2463-2480
- Aharoni A, Giri AP, Verstappen FW, Berteaux CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ** (2004b) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell* **16**: 3110-3131
- Ainsworth EA, Rogers A, Vodkin LO, Walter A, Schurr U** (2006) The effects of elevated CO₂ concentration on soybean gene expression. An analysis of growing and mature leaves. *Plant Physiol* **142**: 135-147
- Alba R, Fei Z, Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'Ascenzo M, Gordon JS, Rose JK, et al** (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J* **39**: 697-714
- Alba R, Payton P, Fei Z, McQuinn R, Debbie P, Martin GB, Tanksley SD, Giovannoni J** (2005) Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. *Plant Cell* **17**: 2954-2965
- Alexandrov NN, Troukhan ME, Brover VV, Tatarinova T, Flavell RB, Feldmann KA** (2006) Features of Arabidopsis genes and genome discovered using full-length cDNAs. *Plant Mol Biol* **60**: 69-85

- Allemeersch J, Durinck S, Vanderhaeghen R, Alard P, Maes R, Seeuws K, Bogaert T, Coddens K, Deschouwer K, Van Hummelen P, et al** (2005) Benchmarking the CATMA microarray. A novel tool for Arabidopsis transcriptome analysis. *Plant Physiol* **137**: 588-601
- Al-Shahrour F, Díaz-Uriarte R, Dopazo J** (2004) FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* **20**: 578-580
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402
- Ancillo G, Gadea J, Forment J, Guerri J, Navarro L** (2007) Class prediction of closely related plant varieties using gene expression profiling. *J Exp Bot* **58**: 1927-1933
- Andersen JR, Lübberstedt T** (2003) Functional markers in plants. *Trends Plant Sci* **8**: 554-560
- Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796-815
- Arimura G, Huber DP, Bohlmann J** (2004) Forest tent caterpillars (*Malacosoma disstria*) induce local and systemic diurnal emissions of terpenoid volatiles in hybrid poplar (*Populus trichocarpa x deltoides*): cDNA cloning, functional characterization, and patterns of gene expression of (-)-germacrene D synthase, *PtdTPS1*. *Plant J* **37**: 603-616
- Arondel V, Vergnolle C, Cantrel C, Kader JC** (2000) Lipid transfer proteins are encoded by a small multigene family in *Arabidopsis thaliana*. *Plant Sci* **157**: 1-12
- Arumuganathan K, Earle ED** (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* **9**:208-218
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al** (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25-29
- Audic S, Claverie JM** (1997) The significance of digital gene expression profiles. *Genome Res* **7**: 986-995
- Bajguz A, Tretyn A** (2003) The chemical characteristic and distribution of brassinosteroids in plants. *Phytochemistry* **62**: 1027-1046
- Barber GA** (1962) Enzymatic glycosylation of quercetin to rutin. *Biochemistry* **25**: 463-468

- Barkley NA, Roose ML, Krueger RR, Federici CT** (2006) Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor Appl Genet* **112**: 1519-1531
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al** (2004) The Pfam protein families database. *Nucleic Acids Res* **1**: 138-141
- Ben-Cheikh W** (1997a) Giberelinas y regulación del desarrollo en cítricos. Influencia de la polinización y brotación sobre la abscisión de frutos. Efecto del viroide de la exocortis y la temperatura sobre el crecimiento del tallo. Tesis Doctoral, Escuela Técnica Superior de Ingenieros Agrónomos, Universidad Politécnica de Valencia.
- Ben-Cheikh W, Pérez-Botella J, Tadeo FR, Talón M, Primo-Millo E** (1997b) Pollination increases gibberellin levels in developing ovaries of seeded varieties of *Citrus*. *Plant Physiol* **114**: 557-564
- Benjamini Y, Hochberg Y** (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc* **57**: 289-300
- Bhatt AM, EtcHELLS JP, Canales C, Lagodienko A, Dickinson H** (2004) VAAMANA-a BEL1-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in *Arabidopsis*. *Gene* **17**: 103-111
- Biemelt S, Tschiersch H, Sonnewald U** (2004) Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. *Plant Physiol* **135**: 254-265
- Birnbaum K, Jung JW, Wang JY, Lambert GM, Hirst JA, Galbraith DW, Benfey PN** (2005) Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat Methods* **2**: 615-619
- Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN** (2003) A gene expression map of the *Arabidopsis* root. *Science* **302**: 1956-1960
- Bishop GJ, Nomura T, Yokota T, Harrison K, Noguchi T, Fujioka S, Takatsuto S, Jones JDG, Kamiya Y** (1999) The tomato DWARF enzyme catalyses C-6 oxidation in brassinosteroid biosynthesis. *Proc Natl Acad Sci USA* **96**: 1761-1766
- Blatt MR** (2000) Cellular signaling and volume control in stomatal movements in plants. *Annu Rev Cell Dev Biol* **16**: 221-241
- Boguski MS, Lowe TM, Tolstoshev CM** (1993) dbEST-database for expressed sequence tags. *Nat Genet* **4**: 332-333

- Bohlmann J, Meyer-Gauen G, Croteau R** (1998) Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc Natl Acad Sci USA* **95**: 4126-4133
- Bonaldo MF, Lennon G, Soares MB** (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res* **6**: 791-806
- Borevitz JO, Ecker JR** (2004) Plant genomics: the third wave. *Annu Rev Genomics Hum Genet* **5**: 443-477
- Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN** (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**: 801-806
- Bulley SM, Wilson FM, Hedden P, Phillips AL, Croker SJ, James DJ** (2005) Modification of gibberellin biosynthesis in the grafted apple scion allows control of tree height independent of the rootstock. *Plant Biotechnol J* **3**: 215-223
- Busch W, Lohmann JU** (2007) Profiling a plant: expression analysis in Arabidopsis. *Curr Opin Plant Biol* **10**: 136-141
- Cao D, Cheng H, Wu W, Soo HM, Peng J** (2006) Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in Arabidopsis. *Plant Physiol* **142**: 509-525
- Casson S, Spencer M, Walker K, Lindsey K** (2005) Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. *Plant J* **42**: 111-123
- Causton HC, Quackenbush J, Brazma A** (2003) Microarray gene expression data analysis: a beginner's guide. Blackwell Pub. Oxford, UK. 24-25.
- Cellier F, Conejero G, Breitler JC, Casse F** (1998) Molecular and physiological responses to water deficit in drought-tolerant and drought-sensitive lines of sunflower. Accumulation of dehydrin transcripts correlates with tolerance. *Plant Physiol* **116**: 319-328
- Cercós M, Soler G, Iglesias DJ, Gadea J, Forment J, Talon M** (2006) Global analysis of gene expression during development and ripening of citrus fruit flesh. A proposed mechanism for citric acid utilization. *Plant Mol Biol* **62**: 513-527
- Chen C, Gmitter FG** (2008) *Citrus* EST characterization and integrated genome mapping. In *Plant & Animal Genomes XVI Conference*. Town & Country Convention Center. San Diego, CA. USA
- Chen X, Goodwin SM, Boroff VL, Liu X, Jenks MA** (2003a) Cloning and characterization of the *WAX2* gene of Arabidopsis involved in cuticle membrane and wax production. *Plant Cell* **15**: 1170-1185

- Chen F, Tholl D, D'Auria JC, Farooq A, Pichersky E, Gershenzon J** (2003b) Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell* **15**: 481-494
- Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP** (2007) The TIGR Plant Transcript Assemblies database. *Nucleic Acids Res* **35**: 846-851
- Choe S, Dilkes BP, Fujioka S, Takatsuto S, Sakurai A, Feldmann KA** (1998) The *DWF4* gene of *Arabidopsis* encodes a cytochrome P450 that mediates multiple 22 α -hydroxylation steps in brassinosteroid biosynthesis. *Plant Cell* **10**: 231-243
- Chou HH, Holmes MH** (2001) DNA sequence quality trimming and vector removal. *Bioinformatics* **17**: 1093-1104
- Clark AM, Bohnert HJ** (1999) Cell-specific expression of genes of the lipid transfer protein family from *Arabidopsis thaliana*. *Plant Cell Physiol* **40**: 69-76
- Clarke B, Rahman S** (2005) A microarray analysis of wheat grain hardness. *Theor Appl Genet* **110**: 1259-1267
- Close TJ, Wanamaker S, Lyon M, Mei G, Davies C, Roose ML** (2006) A GeneChip® for *Citrus*. In *Plant & Animal Genomes XIV Conference*. Town & Country Convention Center. San Diego, CA. USA
- Close TJ, Wanamaker S, Lyon M, Roose ML** (2007) HarvEST: A portable and companion online browser for genome research. In *Plant & Animal Genomes XV Conference*. Town & Country Convention Center. San Diego, CA. USA
- Clouse SD** (2002) Brassinosteroids. In: Somerville CR, Meyerowitz EM (ed) *The Arabidopsis Book*. American Society of Plant Biologists, Rockville, MD, USA
- Coles JP, Phillips AL, Croker SJ, García-Lepe R, Lewis MJ, Hedden P** (1999) Modification of gibberellin production and plant development in *Arabidopsis* by sense and antisense expression of gibberellin 20-oxidase genes. *Plant J* **17**: 547-556
- Cosgrove DJ** (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol* **6**: 850-861
- Cramer MD, Nagel OW, Lips SH, Lambers H** (1995) Reduction, assimilation and transport of N in normal and gibberellin-deficient tomato plants. *Physiol Plant* **95**: 347-354
- Croteau R, Kutchan T, Lewis N** (2000) Natural Products (Secondary Metabolism). In: Buchanan RB, Grissem W, Jones RL (eds) *Biochemistry and Molecular Biology of Plants*, American Society of Plant Biologists, Rockville, MD, USA, 1250-1318

- Crowell DN** (2000) Functional implications of protein isoprenylation in plants. *Prog Lipid Res* **39**: 393-408
- Cunillera N, Arro M, Fores O, Manzano D, Ferrer A** (2000) Characterization of dehydrodolichol diphosphate synthase of *Arabidopsis thaliana*, a key enzyme in dolichol biosynthesis. *FEBS Lett* **477**: 170-174
- Cunningham FX Jr, Gantt E** (2000) Identification of multi-gene families encoding isopentenyl diphosphate isomerase in plants by heterologous complementation in *Escherichia coli*. *Plant Cell Physiol* **41**: 119-123
- Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK** (2004) Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J* **38**: 366-379
- da Silva FG, Iandolino A, Al-Kayal F, Bohlmann MC, Cushman MA, Lim H, Ergul A, Figueroa R, Kabuloglu EK, Osborne C, et al** (2005) Characterizing the grape transcriptome. Analysis of expressed sequence tags from multiple *Vitis* species and development of a compendium of gene expression during berry development. *Plant Physiol* **139**: 574-597
- Dai M, Zhao Y, Ma Q, Hu Y, Hedden P, Zhang Q, Zhou DX** (2007) The rice *YABBY1* gene is involved in the feedback regulation of gibberellin metabolism. *Plant Physiol* **144**: 121-133
- Davies FS, Albrigo LG** (1994) *Cítricos*. Ed Acribia, Zaragoza, España.
- DeRidder BP, Goldsbrough PB** (2006) Organ-specific expression of glutathione S-transferases and efficacy of herbicide safeners in *Arabidopsis*. *Plant Physiol* **140**: 167-175
- Dijkstra P, Ter Reegen H, Kuiper PJC** (1990) Relation between relative growth rate, endogenous gibberellins and the response to applied gibberellic acid for *Plantago major*. *Physiol Plant* **79**: 629-634
- Dong Q, Kroiss L, Oakley FD, Wang BB, Brendel V** (2005) Comparative EST analyses in plant systems. *Methods Enzymol* **395**: 400-418
- Doostdar H, Shapiro JP, Niedz R, Burke MD, McCollum TG, McDonald RE, Mayer RT** (1995) A cytochrome P450 mediated naringenin 3'-hydroxylase from sweet orange cell cultures. *Plant Cell Physiol* **36**: 69-77
- Dopazo J, Al-Shahrour F** (2007) Functional annotation of microarray experiments. In: Falciani F (ed) *Microarray Technology Through Applications*. Taylor & Francis Group, New York, USA, 155-171
- Ducreux LJ, Morris WL, Hedley PE, Shepherd T, Davies HV, Millam S, Taylor MA** (2005) Metabolic engineering of high carotenoid potato tubers containing enhanced levels of beta-carotene and lutein. *J Exp Bot* **56**: 81-89

- Eddy SR** (1998) Profile hidden Markov models. *Bioinformatics* **14**: 755-763
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H** (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**: 953-971
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G** (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**: 1005-1016
- Emanuelsson O, Nielsen H, von Heijne G** (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* **8**: 978-984
- Eriksson ME, Israelsson M, Olsson O, Moritz T** (2000) Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. *Nat Biotechnol* **18**: 784-788
- Eulgem T, Rushton PJ, Robatzek S, Somssich IE** (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci* **5**: 199-206
- Ewing RM, Ben Kahla A, Poirot O, Lopez F, Audic S, Claverie JM** (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res* **9**: 950-959
- Ewing B, Hillier L, Wendl MC, Green P** (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175-185
- Fagoaga C, Tadeo FR, Iglesias DJ, Huerta L, Lliso I, Vidal AM, Talón M, Navarro L, García-Martínez JL, Peña L** (2007) Engineering of gibberellin levels in citrus by sense and antisense overexpression of a *GA 20-oxidase* gene modifies plant architecture. *J Exp Bot* **58**: 1407-1420
- Fei Z, Tang X, Alba RM, White JA, Ronning CM, Martin GB, Tanksley SD, Giovannoni JJ** (2004) Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. *Plant J* **40**: 47-59
- Fernandes J, Brendel V, Gai X, Lal S, Chandler VL, Elumalai RP, Galbraith DW, Pierson EA, Walbot V** (2002) Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. *Plant Physiol* **128**: 896-910
- Ferreira MA, Engler JD, Miguens FC, VanMontagu M, Engler G, deOliveira DE** (1997) Oleosin gene expression in *Arabidopsis thaliana* tapetum coincides with accumulation of lipids in plastids and cytoplasmic bodies. *Plant Physiol Biochem* **35**: 729-739

- Fizames C, Muños S, Cazettes C, Nacry P, Boucherez J, Gaymard F, Piquemal D, Delorme V, Commes T, Doumas P, et al** (2004) The Arabidopsis root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence. *Plant Physiol* **134**: 67-80
- Fleet CM, Sun TP** (2005) A DELLAcate balance: the role of gibberellin in plant morphogenesis. *Curr Op Plant Biol* **8**: 77-85
- Foreman J, Demidchik V, Bothwell JH, Mylona P, Miedema H, Torres MA, Linstead P, Costa S, Brownlee C, Jones JD, et al** (2003) Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature* **422**: 442-446
- Forment J, Gadea J, Huerta L, Abizanda L, Agustí J, Alamar S, Alós E, Andrés F, Arribas R, Beltrán JP, Berbel A, Blázquez MA, Brumós J, Cañas LA, Cercós M, Colmenero-Flores JM, Conesa A, Estables B, Gandía M, García-Martínez JL, Gimeno J, Gisbert A, Gómez G, González-Candelas L, Granell A, Guerra J, Lafuente MT, Madueño F, Marcos JF, Marqués MC, Martínez F, Martínez-Godoy MA, Miralles S, Moreno P, Navarro L, Pallás V, Pérez-Amador MA, Pérez-Valle J, Pons C, Rodrigo I, Rodríguez PL, Royo C, Serrano R, Soler G, Tadeo F, Talón M, Terol J, Trenor M, Vaello L, Vicente O, Vidal C, Zacarías L, Conejero V** (2005) Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies. *Plant Mol Biol* **57**: 375-391
- Forment J, Gilabert F, Robles A, Conejero V, Nuez F, Blanca JM** (2008) EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration. *BMC Bioinformatics* **9**: 5
- Fraser PD, Enfissi EM, Halket JM, Truesdale MR, Yu D, Gerrish C, Bramley PM** (2007) Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids, and intermediary metabolism. *Plant Cell* **19**: 3194-3211
- Frydman A, Weisshaus O, Bar-Peled M, Huhman DV, Sumner LW, Marin FR, Lewinsohn E, Fluhr R, Gressel J, Eyal Y** (2004) Citrus fruit bitter flavors: isolation and functional characterization of the gene *Cm1,2RhaT* encoding a 1,2-rhamnosyltransferase, a key enzyme in the biosynthesis of the bitter flavonoids of citrus. *Plant J* **40**: 88-100
- Fujii H, Shimada T, Endo T, Shimizu T, Omura M** (2006) 29,228 *Citrus* ESTs-collection and analysis toward the functional genomics phase. In: *Plant & Animal Genomes XIV Conference*. Town & Country Convention Center. San Diego, CA. USA

- Fujioka S, Li J, Choi YH, Seto H, Takatsuto S, Noguchi T, Watanebe T, Kuriyama H, Yokota T, Chory J, Sakurai A** (1997) The *Arabidopsis deetiolated2* mutant is blocked early in brassinosteroid biosynthesis. *Plant Cell* **9**: 1951-1962
- Fujioka S, Yokota K** (2003) Biosynthesis and metabolism of brassinosteroids. *Annu Rev Plant Biol* **54**: 137-164
- Galbraith DW, Birnbaum K** (2006) Global studies of cell type-specific gene expression in plants. *Annu Rev Plant Biol* **57**: 451-475
- Gandía M, Conesa A, Ancillo G, Gadea J, Forment J, Pallás V, Flores R, Duran-Vila N, Moreno P, Guerri J** (2007) Transcriptional response of *Citrus aurantifolia* to infection by *Citrus tristeza virus*. *Virology* **367**: 298-306
- Geisler M, Blakeslee JJ, Bouchard R, Lee OR, Vincenzetti V, Bandyopadhyay A, Titapiwatanakun B, Peer WA, Bailly A, Richards EL, et al** (2005) Cellular efflux of auxin catalyzed by the *Arabidopsis* MDR/PGP transporter AtPGP1. *Plant J* **44**: 179-194
- Gmitter FG, Hu X** (1990) The possible role of Yunnan, China, in the origin of contemporary *Citrus* species (*Rutaceae*). *Econ Bot* **44**: 267-277
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al** (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92-100
- González-Ibeas D, Blanca J, Roig C, González-To M, Picó B, Truniger V, Gómez P, Deleu W, Caño-Delgado A, Arús P, et al** (2007) MELOGEN: an EST database for melon functional genomics. *BMC Genomics* **8**: 306
- Goto A, Yamane H, Takahashi N, Hirose K** (1989) Identification of nine gibberellins from young fruit of Satsuma mandarin (*Citrus unshiu* Marc.) *Agric Biol Chem* **53**: 2817-2818
- Gupta PK, Rustgi S** (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct Integr Genomics* **4**: 139-162
- Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al** (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**: 5654-5666
- Hayat S, Ahmad A, Mobin M, Fariduddin Q, Azam ZM** (2001) Carbonic anhydrase, photosynthesis, and seed yield in mustard plants treated with phytohormones. *Photosynthetica* **39**: 111-114
- Hedden P** (2003) The genes of the Green Revolution. *Trends Genet* **19**: 5-9
- Hedden P, Kamiya Y** (1997) Gibberellin biosynthesis: enzymes, genes and their regulation. *Annu Rev Plant Physiol Plant Mol Biol* **48**: 431-460

- Hedden P, Phillips AL** (2000a) Manipulation of hormone biosynthetic genes in transgenic plants. *Curr Opin Biotechnol* **11**: 130-137
- Hedden P, Phillips AL** (2000b) Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci* **5**: 523-530
- Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC** (2003) The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol Biol Evol* **20**: 735-747
- Herrero R, Asins MJ, Carbonell EA, Navarro L** (1996) Genetic diversity in the orange subfamily *Aurantioideae*. I. Intraspecies and intragenus genetic variability. *Theor Appl Gen* **92**: 599-609
- Hirooka K, Bamba T, Fukusaki E, Kobayashi A** (2003) Cloning and kinetic characterization of *Arabidopsis thaliana* solanesyl diphosphate synthase. *Biochem J* **370**: 679-686
- Hong Z, Ueguchi-Tanaka M, Shimizu-Sato S, Inukai Y, Fujioka S, Shimada Y, Takatsuto S, Agetsuma M, Yoshida S, Watanabe Y, et al** (2002) Loss-of-function of a rice brassinosteroid biosynthetic enzyme, C-6 oxidase, prevents the organized arrangement and polar elongation of cells in the leaves and stem. *Plant J* **32**: 495-508
- Hu W, Ma H** (2006) Characterization of a novel putative zinc finger gene *MIF1*: involvement in multiple hormonal regulation of *Arabidopsis* development. *Plant J* **45**: 399-422
- Huang X, Madan A** (1999) CAP3: A DNA sequence assembly program. *Genome Res* **9**: 868-877
- Huang S, Raman AS, Ream JE, Fujiwara H, Cerny RE, Brown SM** (1998) Overexpression of 20-oxidase confers a gibberellin-overproduction phenotype in *Arabidopsis*. *Plant Physiol* **118**: 773-781
- Imoto K, Yokoyama R, Nishitani K** (2005) Comprehensive approach to genes involved in cell wall modifications in *Arabidopsis thaliana*. *Plant Mol Biol* **58**: 177-192
- International Rice Genome Sequencing Project** (2005) The map-based sequence of the rice genome. *Nature* **436**: 793-800
- Iseli C, Jongeneel CV, Bucher P** (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol*: 138-148
- Israelsson M, Eriksson ME, Hertzberg M, Aspeborg H, Nilsson, Moriz T** (2003) Changes in the wood-forming tissue of transgenic hybrid aspen with increased secondary growth. *Plant Mol Biol* **52**: 893-903

- Israelsson M, Mellerowicz E, Chono M, Gullberg J, Moritz T** (2004) Cloning and overproduction of gibberellin 3-oxidase in hybrid aspen trees. Effects on gibberellin homeostasis and development. *Plant Physiol* **135**: 221-230
- Jacobsen JV, Beach LR** (1985) Control of transcription of α -amylase and rRNA genes in barley aleurone protoplasts by gibberellin and abscisic acid. *Nature* **316**: 275-277
- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al** (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463-467
- Jan A, Yang G, Nakamura H, Ichikawa H, Kitano H, Matsuoka M, Matsumoto H, Komatsu S** (2004) Characterization of a xyloglucan endotransglucosylase gene that is up-regulated by gibberellin in rice. *Plant Physiol* **136**: 3670-3681
- Javot H, Lauvergeat V, Santoni V, Martin-Laurent F, Güçlü J, Vinh J, Heyes J, Franck KI, Schöffner AR, Bouchez D, Maurel C** (2003) Role of a single aquaporin isoform in root water uptake. *Plant Cell* **15**: 509-522
- Kamiya Y, García-Martínez JL** (1999) Regulation of gibberellin biosynthesis by light. *Curr Opin Plant Biol* **2**: 398-403
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M** (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**: D277-D280
- Kato M, Ikoma Y, Matsumoto H, Sugiura M, Hyodo H, Yano M** (2004) Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. *Plant Physiol* **134**: 824-837
- Kavi Kishor PB, Hong Z, Miao GH, Hu CAA, Verma DPS** (1995) Over-expression of delta-pyrroline-5-carboxylate synthetase increases proline production and confers osmotolerance in transgenic plants. *Plant Physiol* **108**: 1387-1394
- Kawaii S, Tomono Y, Katase E, Ogawa K, Yano M** (1999) Quantitation of flavonoid constituents in citrus fruits. *J Agric Food Chem* **47**: 3565-3571
- Kawaii S, Tomono Y, Katase E, Ogawa K, Yano M, Koizumi M, Ito C, Furukawa H** (2000) Quantitative study of flavonoids in leaves of citrus plants. *J Agric Food Chem* **48**: 3865-3871
- Kim JK, Jang IC, Wu R, Zuo WN, Boston RS, Lee YH, Ahn IP, Nahm BH** (2003) Co-expression of a modified maize ribosome-inactivating protein and a rice basic chitinase gene in transgenic rice plants confers enhanced resistance to sheath blight. *Transgenic Res* **12**: 475-484

- Kita M, Hirata Y, Moriguchi T, Endo-Inagaki T, Matsumoto R, Hasegawa S, Suhayda CG, Omura M** (2000) Molecular cloning and characterization of a novel gene encoding limonoid UDP-glucosyltransferase in *Citrus*. *FEBBS Lett* **469**: 173-178
- Klimmek F, Sjödin A, Noutsos C, Leister D, Jansson S** (2006) Abundantly and rarely expressed Lhc protein genes exhibit distinct regulation patterns in plants. *Plant Physiol* **140**: 793-804
- Ko JH, Yang SH, Han KH** (2006) Upregulation of an Arabidopsis RING-H2 gene, *XERICO*, confers drought tolerance through increased abscisic acid biosynthesis. *Plant J* **47**: 343-355
- Komatsu S, Masuda T, Hirano H** (1996) Rice gibberellin-binding phosphoprotein structurally related to ribulose-1,5-bisphosphate carboxylase/oxygenase activase. *FEBBS Lett* **384**: 167-171
- Koornneef M, van der Veen JH** (1980) Induction and analysis of gibberellin-sensitive mutants in *Arabidopsis thaliana* (L.) Heynh. *Theor Appl Gen* **58**: 257-263
- Konishi H, Yamane H, Maeshima M, Komatsu S** (2004) Characterization of fructose-bisphosphate aldolase regulated by gibberellin in roots of rice seedling. *Plant Mol Biol* **56**: 839-848
- Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, Pick J, Rhee SY, Karp PD** (2004) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* **32**: D438-D442
- Kwak KJ, Kim JY, Kim YO, Kang H** (2007) Characterization of transgenic Arabidopsis plants overexpressing high mobility group B proteins under high salinity, drought or cold stress. *Plant Cell Physiol* **48**: 221-231
- Lange BM, Ghassemian M** (2003) Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant Mol Biol* **51**: 925-948
- Lee ML, Kuo FC, Whitmore GA, Sklar J** (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* **97**: 9834-9839
- Leung YF, Cavalieri D** (2003) Fundamentals of cDNA microarray data analysis. *Trends Genet* **19**: 649-659
- Li J, Nagpal P, Vitart V, McMorris TC, Chory J** (1996) A role for brassinosteroids in light-dependent development of Arabidopsis. *Science* **272**: 398-401
- Lian X, Wang S, Zhang J, Feng Q, Zhang L, Fan D, Li X, Yuan D, Han B, Zhang Q** (2006) Expression profiles of 10,422 genes at early stage of low nitrogen stress in rice assayed using a cDNA microarray. *Plant Mol Biol* **60**:617-630

- López CG, Banowetz G, Peterson CJ, Kronstad WE** (2001) Differential accumulation of a 24-kd dehydrin protein in wheat seedlings correlates with drought stress tolerance at grain filling. *Hereditas* **135**: 175-181
- Lota M, de Rocca Serra D, Tomi F, Casanova J** (2001) Chemical variability of peel and leaf essential oils of 15 species of mandarins. *Biochem Syst Ecol* **29**: 77-104
- Lücker J, Bouwmeester HJ, Aharoni A** (2007) Metabolic Engineering of Terpenoid Biosynthesis in Plants. In: Verpoorte R, Alfermann AW, Johnson TS (ed) *Applications of Plant Metabolic Engineering*. Springer Pub, The Netherlands, 219-236
- Lücker J, Bowen P, Bohlmann J** (2004) *Vitis vinifera* terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (-)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. *Phytochemistry* **65**: 2649-2659
- Lücker J, El Tamer MK, Schwab W, Verstappen FW, van der Plas LH, Bouwmeester HJ, Verhoeven HA** (2002) Monoterpene biosynthesis in lemon (*Citrus limon*). cDNA isolation and functional analysis of four monoterpene synthases. *Eur J Biochem* **269**: 3160-3171
- Ma X, Panjkar S, Koepke J, Loris E, Stöckigt J** (2006) The structure of *Rauvolfia serpentina* strictosidine synthase is a novel six-bladed beta-propeller fold in plant proteins. *Plant Cell* **18**: 907-920
- Ma L, Sun N, Liu X, Jiao Y, Zhao H, Deng XW** (2005) Organ-specific expression of Arabidopsis genome during development. *Plant Physiol* **138**: 80-91
- Machado MA, Amaral AM, Freitas-Astua J, Berger IJ, Boscariol-Camargo R, Carlos EF, Filho HC, Locali EC, Lucheta AR, Stach-Machado DR, et al** (2007) Analysis of *Citrus* transcriptome: CitEST in Brazil. In *Plant & Animal Genomes XV Conference*. Town & Country Convention Center. San Diego, CA. USA
- Machado MA, Souza AA, Targon ML, Takita MA, Freitas-Astua J, Filho HC, Amaral AM, Palmieri DA, Boscariol-Camargo R, Cristofani M, et al** (2006) Current situation of *Citrus* genome project in Brazil (CitEST). In *Plant & Animal Genomes XIV Conference*. Town & Country Convention Center. San Diego, CA. USA

- Magome H, Yamaguchi S, Hanada A, Kamiya Y, Oda K** (2004) *dwarf and delayed-flowering 1*, a novel Arabidopsis mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. *Plant J* **37**: 720-729
- Malmberg R, Messing J, Sussex I** (1985) *Molecular biology of plants. A laboratory course manual*. New York: Cold Spring Harbor Laboratory Press, 41
- Mandaokar A, Thines B, Shin B, Lange BM, Choi G, Koo YJ, Yoo YJ, Choi YD, Choi G, Browse J** (2006) Transcriptional regulators of stamen development in Arabidopsis identified by transcriptional profiling. *Plant J* **46**: 984-1008
- Manthey JA, Grohmann K, Guthrie N** (2001) Biological properties of citrus flavonoids pertaining to cancer and inflammation. *Curr Med Chem* **8**: 135-153
- Martens S, Forkmann G** (1999) Cloning and expression of flavone synthase II from *Gerbera* hybrids. *Plant J* **20**: 611-618
- Martin DM, Bohlmann J** (2004) Identification of *Vitis vinifera* (-)-alpha-terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry* **65**: 1223-1229
- Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD** (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol* **21**:796-802
- Maruyama T, Ito M, Honda G** (2001) Molecular cloning, functional expression and characterization of (*E*)-beta farnesene synthase from *Citrus junos*. *Biol Pharm Bull* **24**: 1171-1175
- McIntosh CA, Latchinian L, Mansell RL** (1990) Biosynthesis of naringin in *Citrus paradisi*: UDP-glucosyl-transferase activity in grapefruit seedlings. *Phytochemistry* **29**: 1533-1538
- Meyerowitz EM** (2002) Plants compared to animals: the broadest comparative study of development. *Science* **295**: 1482-1485
- Meyers BC, Galbraith DW, Nelson T, Agrawal V** (2004) Methods for transcriptional profiling in plants. Be fruitful and replicate. *Plant Physiol* **135**: 637-652
- Middleton E Jr, Kandaswami C, Theoharides TC** (2000) The effects of plant flavonoids on mammalian cells: implications for inflammation, heart disease, and cancer. *Pharmacol Rev* **52**: 673-751
- Monna L, Kitazawa N, Yoshino R, Suzuki J, Masuda H, Maehara Y, Tanji M, Sato M, Nasu S, Minobe Y** (2002) Positional cloning of rice semidwarfing gene, *sd-1*: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res* **28**: 11-17

- Moore GA** (2001) Oranges and lemons: clues to the taxonomy of *Citrus* from molecular markers. *Trends Genet* **17**: 536-540
- Moriguchi T, Kita M, Hasegawa S, Omura M** (2003) Molecular approach to citrus flavonoid and limonoid biosynthesis. *Food, Agriculture and Environment* **1**: 22-25
- Moriguchi T, Kita M, Ogawa K, Tomono Y, Endo T, Omura M** (2002) Flavonol synthase gene expresión during citrus fruit development. *Physiol Plant* **114**: 251-258
- Moriguchi T, Kita M, Tomono Y, Endo-Inagaki T, Omura M** (1999) One type of chalcone synthase gene expressed during embryogenesis regulates the flavonoid accumulation in citrus cell cultures. *Plant Cell Physiol* **40**: 651-655
- Moriguchi T, Kita M, Tomono Y, Endo-Inagaki T, Omura M** (2001) Gene expression in flavonoid biosynthesis: correlation with flavonoid accumulation in developing citrus fruit. *Physiol Plant* **111**: 66-74
- Mozer TJ** (1980) Control of Protein Synthesis in Barley Aleurone Layers by the Plant Hormones Gibberellic Acid and Abscisic Acid. *Cell* **20**: 479-485
- Mueller LA, Zhang P, Rhee SY** (2003) AraCyc: a biochemical pathway database for Arabidopsis. *Plant Physiol* **132**: 453-460
- Nagel OW, Lambers H** (2002) Changes in the acquisition and partitioning of carbon and nitrogen in the gibberellin-deficient mutants *A70* and *W335* of tomato (*Solanum lycopersicum* L.). *Plant Cell Environ* **25**: 883-891
- Nakai K, Horton P** (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**: 34-36
- Nambara E, Marion-Poll A** (2005) Abscisic Acid Biosynthesis and Catabolism. *Annu Rev Plant Biol* **56**: 165-185
- Nawy T, Lee JY, Colinas J, Wang JY, Thongrod SC, Malamy JE, Birnbaum K, Benfey PN** (2005) Transcriptional profile of the Arabidopsis root quiescent center. *Plant Cell* **17**: 1908-1925
- Neuhaus JL** (1999) Plant chitinase (PR-3, PR-4, PR-8, PR-11). In: Datta S, Muthukrishnan S (eds) *Pathogenesis-related Proteins in Plants*. Boca Raton, CRC Press, 77-105
- Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EH, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ, et al** (2006) Analyses of expressed sequence tags from apple. *Plant Physiol* **141**: 147-166

- Nishikawa F, Endo T, Shimada T, Fujii H, Shimizu T, Omura M, Ikoma Y** (2007) Increased *CiFT* abundance in the stem correlates with floral induction by low temperature in Satsuma mandarin (*Citrus unshiu* Marc.). *J Exp Bot* **58**: 3915-3927
- Noctor G, Foyer CH** (1998) Ascorbate and glutathione: keeping active oxygen under control. *Annu Rev Plant Physiol Plant Mol Biol* **49**: 249-279
- Nomura T, Sato T, Bishop GJ, Kamiya Y, Takatsuto S, Yokota T** (2001) Accumulation of 6-deoxocathasterone and 6-deoxocastasterone in Arabidopsis, pea and tomato is suggestive of common rate-limiting steps in brassinosteroid biosynthesis. *Phytochemistry* **57**: 171-178
- Norusis MJ** (1993) Documentation SPSS for Windows, SPSS, Inc., Chicago
- Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S** (2003) Gibberellin biosynthesis and response during Arabidopsis seed germination. *Plant Cell* **15**: 1591-1604
- Ogawa M, Kusano T, Katsumi M, Sano H** (2000) Rice gibberellin-insensitive gene homolog, *OsGAI*, encodes a nuclear-localized protein capable of gene activation at transcriptional level. *Gene* **245**: 21-29
- Ogihara Y, Mochida K, Nemoto Y, Murai K, Yamazaki Y, Shin-I T, Kohara Y** (2003) Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant J* **33**: 1001-1011
- Ohlrogge J, Benning C** (2000) Unraveling plant metabolism by EST analysis. *Curr Opin Plant Biol* **3**: 224-228
- Ohnishi T, Szatmari AM, Watanabe B, Fujita S, Bancos S, Koncz C, Lafos M, Shibata K, Yokota T, Sakata K, et al** (2006) C-23 hydroxylation by Arabidopsis CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis. *Plant Cell* **18**: 3275-3288
- Olszewski N, Sun TP, Gubler F** (2002) Gibberellin signalling, biosynthesis, catabolism, and response pathways. *Plant Cell* **14**: S61-S80
- Park S, Sugimoto N, Larson MD, Beaudry R, van Nocker S** (2006) Identification of genes with potential roles in apple fruit development and biochemistry through large-scale statistical analysis of expressed sequence tags. *Plant Physiol* **141**: 811-824
- Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J** (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* **19**: 651-652

- Phillips AL, Huttly AK** (1994) Cloning of two gibberellin-regulated cDNAs from *Arabidopsis thaliana* by subtractive hybridization: expression of the tonoplast water channel, γ -TIP, is increased by GA₃. *Plant Mol Biol* **24**: 603-615
- Pierleoni A, Martelli PL, Fariselli P, Casadio R** (2006) BaCellLo: a balanced subcellular localization predictor. *Bioinformatics* **22**: 408-416
- Poling SM, Maier VP** (1988) Identification of endogenous gibberellins in navel orange shoots. *Plant Physiol* **88**: 639-642
- Potter I, Fry SC** (1993) Xyloglucan endotransglycosylase activity in pea internodes. Effects of applied gibberellic acid. *Plant Physiol* **103**: 235-24
- Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K** (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiol* **133**: 1755-1767
- Redman JC, Haas BJ, Tanimoto G, Town CD** (2004) Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *Plant J* **38**: 545-561
- Reid J, Howell SH** (1995) Hormone mutants and plant development. In: Davies P (ed) *Plant Hormones: Physiology, Biochemistry and Molecular Biology*, 2^d ed. Kluwer Acad Pub, Dordrecht, The Netherlands, 448-485
- Reiling KK, Yoshikuni Y, Martin VJ, Newman J, Bohlmann J, Keasling JD** (2004) Mono and diterpene production in *Escherichia coli*. *Biotechnol Bioeng* **87**: 200-212
- Rensink WA, Buell CR** (2005) Microarray expression profiling resources for plant genomics. *Trends Plant Sci* **10**: 603-609
- Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, García-Hernández M, Huala E, Lander G, Montoya M, et al** (2003) The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res* **31**: 224-228
- Riechmann JL, Ratcliffe OJ** (2000) A genomic perspective on plant transcription factors. *Curr Op Plant Biol* **3**: 423-434
- Rodríguez-Concepción M, Boronat A** (2002) Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol* **130**: 1079-1089
- Rodrigo MJ, Zacarías L** (2006) Horticultural and quality aspects of *Citrus* fruits. In: Hiu YH (ed) *Handbook of Fruits Processing and Technology*. Blackwell Publishing, Iowa, USA, 293-308

- Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA** (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol Genomics* **12**: 159-162
- Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, et al** (2003) Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol* **131**: 419-429
- Roose ML, Niedz RP, Gmitter FG, Close TJ, Dandekar AM, Cheng JF, Rokhsar DS** (2007) Analysis of a 1.2x whole genome sequence of *Citrus sinensis*. In Plant & Animal Genomes XV Conference. Town & Country Convention Center. San Diego, CA. USA
- Ruan Y, Gilmore J, Conner T** (1998) Towards Arabidopsis genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J* **15**: 821-833
- Rudd S** (2003) Expressed sequence tags: alternative or complement to whole genome sequence? *Trends Plant Sci* **8**: 321-329
- Rudd S, Schoof H, Mayer K** (2005) PlantMarkers – a database of predicted molecular markers from plants. *Nucleic Acids Res* **33**: D628-D632
- Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Guldener U, Mannhaupt G, Munsterkötter M, Mewes HW** (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res* **32**: 5539-5545
- Sakamoto T, Matsuoka M** (2004) Generating high-yielding varieties by genetic manipulation of plant architecture. *Curr Opin Biotechnol* **15**: 144-147
- Sambrook J, Fritsch EF, Maniatis T** (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York
- Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M** (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* **416**: 701-702
- Savage EM, Gardner FE** (1965) The origin and history of Troyer and Carrizo citranges. *The Citrus Industry* **26**: 5
- Scheible WR, Pauly M** (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. *Curr Opin Plant Biol* **7**: 285-295
- Scheller HV, Lunde C, Haldrup A, Jensen PE** (2005) Functional characterization of the photosynthetic apparatus in *Arabidopsis thaliana*. In: Leister D (ed) *Plant Functional Genomics*, Food Product Press, New York, USA, 393-429

- Schena M, Shalon D, Davis RW, Brown PO** (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-470
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU** (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37**: 501-506
- Schwechheimer C** (2008) Understanding gibberellic acid signaling-are we there yet? *Curr Opin Plant Biol* **11**: 9-15
- Seo HS, Song JT, Cheong JJ, Lee YH, Lee YW, Hwang I, Lee JS, Choi YD** (2001) Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proc Natl Acad Sci USA* **98**: 4788-4793
- Serrano-Cartagena J, Candela H, Robles P, Ponce MR, Pérez-Pérez JM, Piqueras P, Micol JL** (2000) Genetic analysis of incurvata mutants reveals three independent genetic operations at work in Arabidopsis leaf morphogenesis. *Genetics* **156**: 1363-1377
- Sharma A, Komatsu S** (2002) Involvement of a Ca(2+)-dependent protein kinase component downstream to the gibberellin-binding phosphoprotein, RuBisCO activase, in rice. *Biochem Biophys Res Commun* **290**: 690-695
- Sharon-Asa L, Shalit M, Frydman A, Bar E, Holland D, Or E, Lavi U, Lewinsohn E, Eyal Y** (2003) Citrus fruit flavor and aroma biosynthesis: isolation, functional characterization, and developmental regulation of *Cstps1*, a key gene in the production of the sesquiterpene aroma compound valencene. *Plant J* **36**: 664-674
- Shi H, Zhu JK** (2002) *SOS4*, a pyridoxal kinase gene, is required for root hair development in Arabidopsis. *Plant Physiol* **129**: 585-593
- Shimada T, Endo T, Fujii H, Hara M, Ueda T, Kita M, Omura M** (2004) Molecular cloning and functional characterization of four monoterpene synthase genes from *Citrus unshiu* Marc. *Plant Sci* **166**: 49-58
- Shimada Y, Fujioka S, Miyauchi N, Kushiro M, Takatsuto S, Nomura T, Yokota T, Kamiya Y, Bishop GJ, Yoshida S** (2001) Brassinosteroid-6-oxidases from Arabidopsis and tomato catalyze multiple C-6 oxidations in brassinosteroid biosynthesis. *Plant Physiol* **126**: 770-779
- Shin B, Choi G, Yi H, Yang S, Cho I, Kim J, Lee S, Paek NC, Kim JH, Song PS, Choi G** (2002) *AtMYB21*, a gene encoding a flower-specific transcription factor, is regulated by COP1. *Plant J* **30**: 23-32

- Sivamani E, Bahieldin1 A, Wraith JM, Al-Niemi T, Dyer WE, Ho TD, Qu R** (2000) Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA1* gene. *Plant Sci* **155**: 1-9
- Smith RC, Matthews PR, Schünmnn PHD, Chandler PM** (1996) The regulation of leaf elongation and xyloglucan endotransglycosylase by gibberellin in 'Himalaya' barley (*Hordeum vulgare* L.) *J Exp Bot* **47**: 1395-1404
- Sorin C, Negroni L, Balliau T, Corti H, Jacquemot MP, Davanture M, Sandberg G, Zivy M, Bellini C** (2006) Proteomic analysis of different mutant genotypes of Arabidopsis led to the identification of 11 proteins correlating with adventitious root development. *Plant Physiol* **140**: 349-364
- Spencer MW, Casson SA, Lindsey K** (2007) Transcriptional profiling of the Arabidopsis embryo. *Plant Physiol* **143**: 924-940
- Spiegel-Roy P, Goldschmidt EE** (1996) Biology of citrus. Cambridge University Press. Cambridge, GB
- Spielmeier W, Ellis MH, Chandler PM** (2002) Semidwarf (*sd-1*), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proc Natl Acad Sci USA* **99**: 9043-9048
- Sponsel V, Hedden V** (2004) Gibberellin biosynthesis and inactivation. In: Davies P (ed) *Plant Hormones: Biosynthesis, Signal Transduction, Action*. 3d ed. Kluwer Acad Pub, Dordrecht, The Netherlands, 63-94
- Starks CM, Back K, Chappell J, Noel JP** (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-*epi*-aristolochene synthase. *Science* **277**: 1815-18520
- Steele CL, Crock J, Bohlmann J, Croteau R** (1998) Sesquiterpene synthases from grand fir (*Abies grandis*). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase. *J Biol Chem* **273**: 2078-2089
- Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Lindvall JJ, Tandre K, Strauss SH, et al** (2004) A *Populus* EST resource for plant functional genomics. *Proc Natl Acad Sci USA* **101**: 13951-13956
- Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, et al** (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proc Natl Acad Sci USA* **95**: 13330-13335

- Street NR, Skogström O, Sjödin A, Tucker J, Rodríguez-Acosta M, Nilsson P, Jansson S, Taylor G** (2006) The genetics and genomics of the drought response in *Populus*. *Plant J* **48**: 321-341
- Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S, Veilleux RE, Busse JS, Erhardt RJ, Buell CR, Jiang J** (2006) Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics* **176**: 2055-2067
- Sun TP** (2004) Gibberellin signal transduction in stem elongation and leaf growth. In: Davies P (ed) *Plant Hormones: Biosynthesis, Signal Transduction, Action*. 3d ed. Kluwer Acad Pub, Dordrecht, The Netherlands, 304-320
- Sun TP, Gubler E** (2004) Molecular mechanism of gibberellin signaling in plants. *Annu Rev Plant Biol* **55**: 197-223
- Sun TP, Kamiya Y** (1996) The Arabidopsis *GA1* locus encodes the cyclase *ent*-kaurene synthase A of gibberellin biosynthesis. *Plant Cell* **6**: 1509-1518
- Swain SM, Singh DP** (2005) Tall tales from sly dwarves: novel functions of gibberellins in plant development. *Trends Plant Sci* **10**: 123-129
- Swingle WT, Reece PC** (1967) The botany of *Citrus* and its relatives. In: Reuther W, Webber H, Batchelor L (eds) *The Citrus Industry*. University of California, USA
- Syvertsen JP** (1984) Light acclimation in *Citrus* leaves. II. Assimilation and light, water and nitrogen efficiency. *J Amer Soc Hort Sci* **109**: 812-817
- Szekeres M, Németh K, Koncz-Kálmán Z, Mathur J, Kauschmann A, Altmann T, Rédei GP, Nagy F, Schell J, Koncz C** (1996) Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in Arabidopsis. *Cell* **85**: 171-182
- Talón M, Hedden P, Primo-Millo E** (1990) Gibberellins in *Citrus sinensis*: a comparison between seeded and seedless varieties. *J Plant Growth Regul* **9**: 201-206
- Talón M, Zacarías L, Primo-Millo E** (1992) Gibberellins and parthenocarpic ability in developing ovaries of seedless mandarins. *Plant Physiol* **99**: 1575-1581
- Tanaka T** (1977) Fundamental discussion of *Citrus* classification. *Stud Citrol* **14**: 1-6
- Terol J, Conesa A, Colmenero JM, Cercós M, Tadeo F, Agustí J, Alós E, Andrés F, Soler G, Brumós J, et al** (2007) Analysis of 13000 unique *Citrus* clusters associated with fruit quality, production and salinity tolerance. *BMC Genomics* **8**: 31
- Terol J, Talón M** (2008) Analysis of the repetitive content of the genome *Citrus clementina* based on BAC-end sequencing. In *Plant & Animal Genomes XVI Conference*. Town & Country Convention Center. San Diego, CA. USA

- Thomas SG, Phillips AL, Hedden P** (1999) Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proc Natl Acad Sci USA* **96**: 4698-4703
- Turnbull CGN** (1989) Identification and quantitative analysis of gibberellins in *Citrus*. *J Plant Growth Regul* **8**: 273-282
- Tusher VG, Tibshirani R, Chu G** (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98**: 5116-5121
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al** (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596-1604
- Uozu S, Tanaka-Ueguchi M, Kitano H, Hattori K, Matsuoka M** (2000) Characterization of XET-related genes of rice. *Plant Physiol* **122**: 853-859
- van der Knaap E, Jagoueix S, Kende H** (1997) Expression of an ortholog of replication protein A1 (RPA1) is induced by gibberellin in deepwater rice. *Proc Natl Acad Sci USA* **94**: 9979-9983
- van Schie CC, Ament K, Schmidt A, Lange T, Haring MA, Schuurink RC** (2007) Geranyl diphosphate synthase is required for biosynthesis of gibberellins. *Plant J* **52**: 752-762
- Varshney RK, Graner A, Sorrells ME** (2005) Genomics-assisted breeding for crop improvement. *Trends Plant Sci* **10**: 621-630
- Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, et al** (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* **2**: e1326
- Ververidis F, Trantas E, Douglas C, Vollmer G, Kretschmar G, Panopoulos N** (2007) Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health. *Biotechnol J* **2**: 1214-1234
- Vidal AM, Ben-Cheikh W, Talón M, García-Martínez JL** (2003) Regulation of gibberellin 20-oxidase gene expression and gibberellin content in citrus by temperature and citrus exocortis viroid. *Planta* **217**: 442-448
- Vidal AM, Gisbert C, Talón M, Primo-Millo E, López-Díaz I, García-Martínez JL** (2001) The ectopic overexpression of a citrus gibberellin 20-oxidase enhances the non-13-hydroxylation pathway of gibberellin biosynthesis and induces an extremely elongated phenotype in tobacco. *Physiol Plant* **112**: 251-260

- Vodkin LO, Khanna A, Shealy R, Clough SJ, Gonzalez DO, Philip R, Zabala G, Thibaud-Nissen F, Sidarous M, Strömvik MV, et al** (2004) 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. *BMC Genomics* **5**: 73
- Wang ZY, Kenigsbuch D, Sun L, Harel E, Ong MS, Tobin EM** (1997) A Myb-related transcription factor is involved in the phytochrome regulation of an Arabidopsis *Lhcb* gene. *Plant Cell* **9**: 491-507
- Wellmann F, Lukacin R, Moriguchi T, Britsch L, Schiltz E, Matern U** (2002) Functional expression and mutational analysis of flavonol synthase from *Citrus unshiu*. *Eur J Biochem* **269**: 4134-4142
- Wilson RN, Heckman JW, Somerville CR** (1992) Gibberellin is required for flowering in *Arabidopsis thaliana* under short days. *Plant Physiol* **100**: 403-408
- Winkel-Shirley B** (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* **126**: 485-493
- Wortman JR, Haas BJ, Hannick LI, Smith RK Jr, Maiti R, Ronning CM, Chan AP, Yu C, Ayele M, Whitelaw CA, et al** (2003) Annotation of the Arabidopsis genome. *Plant Physiol* **132**: 461-468
- Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, López R, et al** (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* **1**: 187-191.
- Xu D, Duan X, Wang B, Hong B, Ho T, Wu R** (1996) Expression of a Late Embryogenesis Abundant Protein Gene, *HVA1*, from Barley Confers Tolerance to Water Deficit and Salt Stress in Transgenic Rice. *Plant Physiol* **110**: 249-257
- Yamaguchi S, Kamiya Y** (2000) Gibberellin biosynthesis: its regulation by endogenous and environmental signals. *Plant Cell Physiol* **41**: 251-257
- Yamaguchi-Shinozaki K, Koizumi M, Urao S, Shinozaki K** (1992) Molecular cloning and characterization of 9 cDNAs for genes that are responsive to desiccation in *Arabidopsis thaliana*: Sequence analysis of one cDNA clone that encodes a putative transmembrane channel protein. *Plant Cell Physiol* **33**: 217-224
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP** (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**: e15

- Yang GX, Jan A, Shen SH, Yazaki J, Ishikawa M, Shimatani Z, Kishimoto N, Kikuchi S, Matsumoto H, Komatsu S** (2004) Microarray analysis of brassinosteroids- and gibberellin-regulated gene expression in rice seedlings. *Mol Gen Genomics* **271**: 468-478
- Yang YH, Speed T** (2002) Design issues for cDNA microarray experiments. *Nat Rev Genet* **3**: 579-588
- Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, et al** (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol* **60**: 107-124
- Yazaki J, Shimatani Z, Hashimoto A, Nagata Y, Fujii F, Kojima K, Suzuki K, Taya T, Tonouchi M, Nelson C, et al** (2004) Transcriptional profiling of genes responsive to abscisic acid and gibberellin in rice: phenotyping and comparative analysis between rice and Arabidopsis. *Physiol Genomics* **17**: 87-100
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al** (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**: 79-92
- Yuan L, Xu DQ** (2001) Stimulation effect of gibberellic acid short-term treatment on leaf photosynthesis related to the increase in Rubisco content in broad bean and soybean. *Photosynth Res* **68**: 39-47
- Yue H, Eastman PS, Wang BB, Minor J, Doctolero MH, Nuttall RL, Stack R, Becker JW, Montgomery JR, Vainer M, Johnston R** (2001) An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. *Nucleic Acids Res* **29**: e41
- Zentella R, Zhang ZL, Park M, Thomas SG, Endo A, Murase K, Fleet CM, Jikumaru Y, Nambara E, Kamiya Y, Sun TP** (2007) Global analysis of DELLA direct targets in early gibberellin signaling in Arabidopsis. *Plant Cell* **19**: 3037-3057
- Zhang Z, Schwartz S, Wagner L, Miller W** (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**: 203-214
- Zhang H, Sreenivasulu N, Weschke W, Stein N, Rudd S, Radchuk V, Potokina E, Scholz U, Schweizer P, Zierold U, et al** (2004) Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant J* **40**: 276-290
- Zhu B, Choi DW, Fenton R, Close TJ** (2000) Expression of the barley dehydrin multigene family and the development of freezing tolerance. *Mol Gen Genet* **264**: 145-153



Anejos

Anejo I. Anotación funcional de los unigenes de Veg1 según la proteína de Arabidopsis más similar

Unigen de cítricos	Código AGI	Descripción	Valor E
aCL43Contig3	At5g38410	RuBisCO small subunit 3B (RBCS-3B)	3E-77
aCL705Contig1	At3g26740	light responsive protein-related	6E-30
aCL428Contig1	At4g24220	expressed protein	2E-85
aCL63Contig4	At4g35090	catalase 2	0
aCL14Contig6	At2g34430	chlorophyll A-B binding protein / LHCII type I (LHB1B1)	1E-132
aCL7Contig10	At3g15353	metallothionein protein	1E-17
aCL48Contig2	At2g39730	RuBisCO activase (RA)	0
aCL11Contig1	At5g06740	lectin protein kinase family protein	8E-14
aCL93Contig1	At4g10340	chlorophyll A-B binding protein CP26 (LHCB5)	1E-130
aCL871Contig1	At3g01500	carbonic anhydrase 1 (CA1)	1E-115
aCL241Contig1	At3g45140	lipoxygenase (LOX2)	0
aCL23Contig1	At1g47128	cysteine proteinase (RD21A)	0
aCL2388Contig1	At1g52230	photosystem I reaction center subunit VI (PSI-H)	9E-47
aCL14Contig1	At1g29910	chlorophyll A-B binding protein 2 / LHCII type I (CAB-2)	1E-132
aCL1393Contig1	At2g06520	membrane protein, contains 2 transmembrane domains	1E-19
aCL129Contig1	At2g03440	nodulin-related	6E-07
aCL92Contig1	At2g02400	cinnamoyl-CoA reductase family	1E-129
aCL554Contig1	At4g32940	vacuolar processing enzyme gamma (gamma-VPE)	0
aCL3Contig25	At5g59310	lipid transfer protein 4 (LTP4)	4E-26
aCL223Contig1	At1g55670	photosystem I reaction center subunit V (PSI-G)	7E-58
aCL2238Contig2	At3g47470	chlorophyll A-B binding protein 4 / LHCI type III (CAB-4)	1E-128
aCL13Contig1	At3g43810	calmodulin-7 (CAM7)	4E-81
aCL12Contig7	At3g18280	protease inhibitor/seed storage/lipid transfer protein (LTP)	3E-25
aCL117Contig1	At3g50820	extrinsic subunit of photosystem II (PSBO-2)	1E-153
aCL105Contig2	At4g33010	glycine dehydrogenase (decarboxylating)	0
aCL922Contig1	At4g19860	lecithin:cholesterol acyltransferase family protein (LACT)	0
aCL7Contig9	At3g15353	metallothionein protein	5E-18
aCL79Contig2	At5g19140	auxin/aluminum-responsive protein	1E-108
aCL6Contig3	At1g06680	photosystem II oxygen-evolving complex 23 (OEC23)	1E-114
aCL6162Contig1	At4g29090	reverse transcriptase	7E-13
aCL5Contig7	At1g05870	expressed protein	2E-49
aCL5Contig32	At2g05100	chlorophyll A-B binding protein / LHCII type II (LHCB2)	1E-132
aCL5733Contig1	At5g23120	photosystem II stability/assembly factor (HCF136)	7E-48
aCL4630Contig1	At3g49340	cysteine proteinase	2E-91
aCL460Contig1	At1g73230	nascent polypeptide-associated complex (NAC)	1E-65
aCL3Contig22	At5g01750	expressed protein	7E-69
aCL3Contig1	At1g60710	aldo/keto reductase family protein	1E-149
aCL3244Contig1	At3g44850	protein kinase-related	0
aCL3176Contig1	At1g09340	expressed protein	0
aCL3117Contig1	At5g26742	similar to DEAD box RNA helicase	0
aCL247Contig1	At1g61520	chlorophyll A-B binding protein / LHCI type III (LHCA3.1)	1E-122
aCL242Contig1	At3g61470	chlorophyll A-B binding protein (LHCA2)	1E-120
aCL2090Contig1	At3g51880	high mobility group protein alpha (HMGalpha)	7E-32
aCL1995Contig2	At1g17745	similar to D-3-phosphoglycerate dehydrogenase (3-PGDH)	1E-176
aCL193Contig3	At1g75270	dehydroascorbate reductase	3E-95
aCL158Contig1	At2g19900	encodes the malic enzyme (NADP-ME1)	0
aCL1455Contig1	At5g10980	histone H3	5E-66
aCL12Contig10	At2g16600	peptidyl-prolyl cis-trans isomerase / cyclophilin	5E-77
aCL1215Contig1	At5g20860	pectinesterase family protein	1E-126
aCL1042Contig1	AtCg00150	ATPase a subunit (ATP1)	1E-108
aCL1030Contig1	At5g54770	thiazole biosynthetic enzyme (ARA6) (THI1) (THI4)	1E-150
aCL9Contig8	At4g02380	late embryogenesis abundant 3 family protein (LEA3)	1E-14
aCL973Contig1	At5g47390	myb family transcription factor	1E-114
aCL924Contig1	At1g06870	signal peptidase	4E-34
aCL9187Contig1	At1g03940	transferase family protein	9E-21
aCL9162Contig1	At2g06925	phospholipase A2 family protein	1E-20
aCL835Contig2	At4g13930	glycine hydroxymethyltransferase	0
aCL756Contig1	At1g20340	plastocyanin	9E-49

aCL7530Contig1	At5g50810	mitochondrial import inner membrane translocase (TIM8)	6E-28
aCL7518Contig1	At1g18070	similar to elongation factor 1-alpha (EF-1-alpha)	2E-47
aCL7268Contig1	At1g51590	similar to mannosyl-oligosaccharide 1,2- α -mannosidase	1E-146
aCL6Contig7	At1g54410	dehydrin family protein	8E-09
aCL6Contig1	At1g06680	photosystem II oxygen-evolving complex 23 (OEC23)	3E-67
aCL6782Contig1	At4g18130	phytochrome E (PHYE)	2E-14
aCL6673Contig1	At1g08820	similar to vesicle-associated membrane protein (VAMP)	1E-11
aCL655Contig1	At1g74470	geranylgeranyl reductase	0
aCL642Contig1	At3g26650	glyceraldehyde 3-phosphate dehydrogenase A (GAPA)	9E-47
aCL63Contig2	At4g35090	catalase 2	0
aCL5Contig16	At2g21660	glycine-rich RNA-binding protein (GRP7)	8E-36
aCL5703Contig1	At5g25080	expressed protein	4E-59
aCL5670Contig1	At3g57540	remorin family protein	4E-13
aCL5624Contig1	At3g19170	Zinc metalloprotease pitrilysin subfamily A	9E-60
aCL5533Contig1	At2g18690	expressed protein	1E-15
aCL551Contig1	At2g36690	oxidoreductase, 2OG-Fe(II) oxygenase family protein	4E-68
aCL5365Contig1	At5g49945	expressed protein	3E-13
aCL5084Contig1	At5g13630	magnesium-chelatase subunit chlH (CHLH)	1E-159
aCL4995Contig1	At2g04360	expressed protein	1E-94
aCL48Contig1	At2g39730	RuBisCO activase (RA)	4E-60
aCL4866Contig1	At1g06460	31.2 kDa small heat shock family protein (hsp20)	4E-42
aCL4648Contig1	At2g26580	plant-specific transcription factor YABBY family protein	6E-18
aCL4590Contig1	At5g60160	aspartyl aminopeptidase	1E-103
aCL44Contig1	At5g61820	expressed protein	1E-144
aCL43Contig4	At1g67090	RuBisCO small subunit 1A (RBCS-1A)	1E-77
aCL4074Contig1	At5g22580	expressed protein	1E-33
aCL402Contig1	At5g43940	alcohol dehydrogenase class III (ADHIII)	0
aCL3Contig2	At1g28330	dormancy-associated protein (DRM1)	7E-39
aCL3Contig11	At1g28330	dormancy-associated protein (DRM1)	1E-38
aCL3890Contig1	At3g17930	expressed protein	2E-27
aCL36Contig2	At4g21960	peroxidase 42 (PER42) (P42) (PRXR1)	1E-157
aCL369Contig2	At3g18280	protease inhibitor/seed storage/lipid transfer protein (LTP)	6E-24
aCL369Contig1	At3g18280	protease inhibitor/seed storage/lipid transfer protein (LTP)	1E-23
aCL3263Contig1	At4g37300	expressed protein	2E-24
aCL314Contig1	At5g20700	senescence-associated protein-related (SAG102)	3E-31
aCL3089Contig1	At2g17800	Rac-like GTP-binding protein ARAC1/ATGP2	5E-98
aCL2970Contig1	At4g00660	DEAD/DEAH box helicase	0
aCL289Contig1	At1g78680	gamma-glutamyl hydrolase (GGH2)	1E-139
aCL2887Contig1	At4g33380	expressed protein	1E-113
aCL281Contig3	At1g42970	glyceraldehyde-3-phosphate dehydrogenase B (GAPB)	1E-122
aCL2578Contig1	At3g56940	dicarboxylate diiron protein, putative (CRD1)	1E-141
aCL245Contig2	At1g56220	dormancy/auxin associated family protein	1E-26
aCL23Contig3	At1g47128	cysteine proteinase (RD21A)	0
aCL2337Contig1	At5g19620	outer membrane OMP85 family protein	1E-123
aCL21Contig2	At5g60390	elongation factor 1-alpha (EF-1-alpha)	0
aCL19Contig7	At1g64230	similar to ubiquitin-conjugating enzyme E2-17 kDa 9 (UBC9)	5E-84
aCL19Contig2	At1g51200	zinc finger (AN1-like) family protein	5E-58
aCL1971Contig1	At5g64040	photosystem I reaction center subunit PSI-N	3E-54
aCL189Contig1	At3g52880	monodehydroascorbate reductase (NADH)	0
aCL1811Contig1	At5g05690	cytochrome P450 90A1 (CYP90A1)	0
aCL1730Contig2	At2g27680	aldo/keto reductase family protein	1E-133
aCL172Contig1	At5g25610	dehydration-responsive protein (RD22)	6E-42
aCL164Contig2	At3g29075	glycine-rich protein	2E-23
aCL15Contig1	At1g50570	similar to C2 domain-containing protein	1E-21
aCL1583Contig1	At1g54680	expressed protein	4E-60
aCL1543Contig1	At2g21580	similar to 40S ribosomal protein S25 (RPS25E)	3E-31
aCL1513Contig1	At3g21055	photosystem II 5 kD protein (PSII-T)	5E-14
aCL148Contig2	At1g79040	photosystem II 10 kDa polypeptide (PsbR)	5E-48
aCL1448Contig1	At2g30980	shaggy-related protein kinase delta (ASK4)	0
aCL1433Contig1	At5g42820	U2 snRNP auxiliary factor small subunit	3E-12
aCL13Contig4	At5g60360	cysteine proteinase / AALP protein	1E-165

aCL139Contig1	At1g69230	expressed protein	8E-18
aCL138Contig3	At5g08410	ferredoxin-thioredoxin reductase	4E-28
aCL1328Contig1	At5g45130	Ras-related protein (RHA1)	1E-98
aCL1306Contig1	At4g36810	geranylgeranyl pyrophosphate synthase (GGPS1)	1E-113
aCL12Contig8	At3g18280	protease inhibitor/seed storage/lipid transfer protein (LTP)	2E-23
aCL1263Contig1	AtCg01310	encodes a chloroplast ribosomal protein L2	3E-75
aCL1214Contig1	At1g15820	chlorophyll A-B binding protein (LHCB6)	1E-116
aCL117Contig2	At3g50820	extrinsic subunit of photosystem II (PSBO-2)	1E-119
aCL10Contig8	At1g11430	plastid developmental protein DAG	8E-81
aCL105Contig1	At4g33010	glycine dehydrogenase (decarboxylating)	4E-32
aCL1019Contig2	At4g21580	oxidoreductase, zinc-binding dehydrogenase family protein	1E-143
aCL9Contig4	At2g32520	dienelactone hydrolase family protein	1E-111
aCL9Contig21	At4g02380	late embryogenesis abundant 3 family protein (LEA3)	1E-14
aCL9Contig20	At4g02380	late embryogenesis abundant 3 family protein (LEA3)	9E-15
aCL9Contig16	At4g02380	late embryogenesis abundant 3 family protein (LEA3)	9E-15
aCL9Contig10	AtCg00600	Cytochrome b6-f complex, subunit V	4E-13
aCL971Contig1	At4g25440	WD-40 repeat family protein / zfw1 protein (ZFWD1)	1E-171
aCL970Contig2	At5g66880	serine/threonine protein kinase	4E-67
aCL95Contig4	At5g51970	sorbitol dehydrogenase	1E-125
aCL95Contig1	At5g51970	sorbitol dehydrogenase	0
aCL956Contig3	At5g42190	E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 (At2)	9E-68
aCL955Contig2	At1g56190	phosphoglycerate kinase	1E-128
aCL9469Contig1	At3g17380	mepirin and TRAF homology domain-containing protein	1E-53
aCL9463Contig1	At5g01020	protein kinase family protein	1E-17
aCL9405Contig1	At5g59140	SKP1 family protein	6E-43
aCL9386Contig1	At5g58350	protein kinase family protein	1E-136
aCL936Contig1	At3g02230	reversibly glycosylated polypeptide-1 (RGP1)	0
aCL9367Contig1	At5g58375	expressed protein	3E-18
aCL9335Contig1	At5g18070	phosphoglucosamine mutase-related (PAGM)	3E-71
aCL932Contig1	At3g56630	cytochrome P450 (CYP94A1)	1E-109
aCL931Contig1	At4g27520	plastocyanin-like domain-containing protein	4E-39
aCL9319Contig1	At3g52150	similar to 33 kDa ribonucleoprotein	6E-72
aCL9312Contig1	At2g16710	hesB-like domain-containing protein	2E-44
aCL9295Contig1	At1g25380	mitochondrial substrate carrier family protein	1E-56
aCL9283Contig1	At1g06110	F-box family protein	1E-06
aCL926Contig3	At1g25420	expressed protein	2E-87
aCL9235Contig1	At1g79650	DNA repair protein RAD23	8E-68
aCL914Contig1	At5g13420	transaldolase	0
aCL9142Contig1	At5g61590	encodes a member of the ERF subfamily B-3	3E-37
aCL912Contig1	At1g15520	ABC transporter family	0
aCL910Contig1	At5g07090	40S ribosomal protein S4 (RPS4B)	1E-140
aCL90Contig2	At5g17920	encodes a METS protein	0
aCL9070Contig1	At2g31660	importin beta-2 subunit family protein	1E-120
aCL9056Contig1	At3g06580	galactokinase (GAL1)	1E-130
aCL9050Contig1	At2g18900	transducin family protein / WD-40 repeat family protein	9E-50
aCL9033Contig1	At5g58040	fip1 motif-containing protein	1E-12
aCL9030Contig1	At4g12710	armadillo/beta-catenin repeat family protein	9E-17
aCL89Contig1	At3g14230	encodes a member of the ERF subfamily B-3	1E-68
aCL8943Contig1	At1g06950	encodes a protein thought to be a part of the translocon	1E-112
aCL893Contig1	At5g07990	flavonoid 3'-monooxygenase (F3'H)	1E-130
aCL88Contig1	At1g18650	glycosyl hydrolase family protein 17	8E-35
aCL8865Contig1	At3g09850	D111/G-patch domain-containing protein	1E-44
aCL8829Contig1	At5g38510	rhomboid family protein	1E-08
aCL8807Contig1	At1g72390	expressed protein	1E-20
aCL877Contig2	At3g62220	serine/threonine protein kinase	1E-169
aCL8779Contig1	At5g58350	protein kinase family protein	4E-23
aCL8778Contig1	At5g06620	similar to SET domain-containing protein	2E-23
aCL8761Contig1	At2g46690	auxin-responsive family protein	3E-30
aCL8735Contig1	At2g31190	expressed protein	1E-64
aCL8734Contig1	At1g08250	prephenate dehydratase family protein	3E-57
aCL8725Contig1	At1g64770	expressed protein	1E-116

aCL8721Contig1	At2g43710	acyl-(acyl-carrier-protein) desaturase (SSI2)	1E-106
aCL8680Contig1	At3g55980	zinc finger (CCCH-type) family protein	6E-06
aCL8669Contig1	At1g78560	bile acid:sodium symporter family protein	1E-142
aCL8646Contig1	At2g36070	mitochondrial import inner membrane translocase	3E-57
aCL8570Contig1	At4g27670	25.3 kDa small heat shock protein (HSP25.3-P)	4E-25
aCL855Contig1	At1g17720	serine/threonine protein phosphatase 2A (PP2A)	0
aCL8551Contig1	AtCg00440	NADH dehydrogenase D3 subunit	1E-59
aCL8535Contig1	At4g27840	expressed protein	5E-24
aCL8531Contig1	At3g15190	chloroplast 30S ribosomal protein S20	3E-30
aCL850Contig1	At5g58770	dehydrodolichyl diphosphate synthase	6E-36
aCL8448Contig1	At1g12360	cytokinesis-related Sec1 protein (KEULE)	1E-113
aCL8320Contig1	At5g62050	OXA1 protein (OXA1)	7E-39
aCL828Contig1	At3g44110	DNAJ heat shock protein (J3)	1E-174
aCL8241Contig1	At5g49890	chloride channel protein (CLC-c)	3E-53
aCL822Contig1	At4g10040	cytochrome c	1E-60
aCL8225Contig1	At5g13630	magnesium-chelatase subunit chlH (CHLH)	1E-153
aCL8210Contig1	At1g19660	similar to wound-responsive protein-related	7E-61
aCL817Contig4	At4g20260	similar to DREPP plasma membrane polypeptide-related	2E-41
aCL8170Contig1	At3g54970	expressed protein	2E-49
aCL8127Contig1	At1g74190	leucine-rich repeat family protein	3E-53
aCL808Contig1	At3g60820	20S proteasome beta subunit F1 (PBF1)	1E-109
aCL8082Contig1	At3g49800	BSD domain-containing protein	9E-21
aCL8059Contig1	AtMg00710	hypothetical protein	1E-07
aCL8050Contig1	At3g17450	hAT dimerisation domain-containing protein	1E-107
aCL804Contig2	At5g62640	similar to proline-rich family protein	1E-29
aCL8025Contig1	At3g15352	cytochrome c oxidase copper chaperone-related	2E-22
aCL8023Contig1	At1g35710	leucine-rich repeat transmembrane protein kinase	8E-30
aCL8013Contig1	At1g29690	encodes a protein with a domain similar to MACPF	1E-127
aCL8005Contig1	At4g24350	phosphorylase family protein	3E-37
aCL79Contig3	At5g19140	auxin/aluminum-responsive protein	9E-97
aCL7920Contig1	At1g26230	chaperonin	5E-73
aCL78Contig4	At5g21090	leucine-rich repeat protein	5E-97
aCL7877Contig1	At5g47760	phosphoglycolate phosphatase	1E-137
aCL7853Contig1	At5g19630	expressed protein	3E-85
aCL784Contig1	At4g20840	FAD-binding domain-containing protein	0
aCL7849Contig1	At1g47960	invertase/pectin methylesterase inhibitor family protein	4E-25
aCL7837Contig1	At5g52960	expressed protein	1E-50
aCL781Contig2	At2g39990	eukaryotic translation initiation factor 3 subunit 5	1E-120
aCL7813Contig1	At4g13850	glycine-rich RNA-binding protein (GRP2)	2E-34
aCL7793Contig1	At4g22120	early-responsive to dehydration protein-related (ERD)	1E-101
aCL777Contig1	At1g17370	oligouridylate-binding protein	1E-162
aCL7756Contig1	At5g14610	similar to DEAD box RNA helicase (DRH1)	4E-47
aCL7748Contig1	At2g02050	NADH-ubiquinone oxidoreductase B18 subunit	1E-42
aCL7736Contig1	At3g06620	protein kinase family protein	1E-94
aCL7724Contig1	At2g35360	ubiquitin family protein	1E-57
aCL7718Contig1	At5g41940	RabGAP/TBC domain-containing protein	1E-110
aCL7709Contig1	At1g76730	5-formyltetrahydrofolate cyclo-ligase family protein	2E-30
aCL7697Contig1	At1g30500	CCAAT-binding transcription factor (CBF-B/NF-YA)	7E-41
aCL7696Contig1	At1g63855	expressed protein	5E-68
aCL7682Contig1	At2g29060	scarecrow transcription factor family protein	7E-43
aCL7670Contig1	At3g56160	expressed protein	3E-27
aCL766Contig1	At5g50920	similar to ATP-dependent Clp protease ATP-binding subunit	0
aCL7651Contig1	At3g59040	pentatricopeptide (PPR) repeat-containing protein	1E-129
aCL7628Contig1	At3g25410	bile acid:sodium symporter family protein	2E-71
aCL7615Contig1	At3g52280	DNA-binding bromodomain-containing protein	5E-66
aCL7612Contig1	At2g37150	zinc finger (C3HC4-type RING finger) family protein	2E-09
aCL7609Contig1	At5g64700	nodulin MtN21 family protein	9E-48
aCL7605Contig1	At5g55850	similar to nitrate-responsive NOI protein	3E-19
aCL759Contig2	At1g26910	60S ribosomal protein L10 (RPL10B)	1E-114
aCL759Contig1	At1g26910	60S ribosomal protein L10 (RPL10B)	1E-114
aCL7581Contig1	At1g62710	vacuolar processing enzyme beta (beta-VPE)	2E-32

aCL7558Contig1	At5g67290	FAD-dependent oxidoreductase family protein	5E-51
aCL7524Contig1	At4g15560	1-deoxy-D-xylulose 5-phosphate synthase	2E-89
aCL7514Contig1	At2g44140	autophagy 4a (APG4a)	3E-49
aCL7513Contig1	At4g31040	proton extrusion protein-related	1E-76
aCL7507Contig1	At1g13245	expressed protein	2E-10
aCL7479Contig1	At5g48450	multi-copper oxidase type I family protein	1E-142
aCL7475Contig1	At3g44880	Rieske (2Fe-2S) domain-containing protein	1E-45
aCL7473Contig1	At4g30790	expressed protein	1E-112
aCL744Contig1	At4g02340	epoxide hydrolase	5E-53
aCL743Contig2	At5g59890	actin-depolymerizing factor 4 (ADF4)	2E-64
aCL7438Contig1	At4g28760	expressed protein	4E-11
aCL7437Contig1	At2g35795	DNAJ heat shock N-terminal domain-containing protein	3E-40
aCL7421Contig1	At5g04870	calcium-dependent protein kinase isoform AK1 (AK1)	1E-69
aCL7419Contig1	At4g16360	5'-AMP-activated protein kinase beta-2 subunit	7E-90
aCL73Contig1	At4g38970	fructose-bisphosphate aldolase	0
aCL738Contig3	At5g42050	expressed protein	7E-98
aCL738Contig1	At5g42050	expressed protein	3E-77
aCL7389Contig1	At5g43750	expressed protein	7E-64
aCL7382Contig1	At3g55070	expressed protein	1E-20
aCL7371Contig1	At3g49290	expressed protein	1E-22
aCL7364Contig1	At2g37970	SOUL heme-binding family protein	8E-67
aCL7330Contig1	At5g35695	hypothetical protein	6E-29
aCL7325Contig1	At3g18990	transcriptional factor B3 family protein	6E-40
aCL7322Contig1	At1g17020	oxidoreductase, 2OG-Fe(II) oxygenase family protein	2E-51
aCL7313Contig1	At3g52860	expressed protein	1E-24
aCL72Contig2	At3g02780	isopentenyl-diphosphate delta-isomerase II (IPP2)	1E-121
aCL7273Contig1	At1g65700	similar to small nuclear ribonucleoprotein	1E-40
aCL725Contig1	At1g66330	senescence-associated family protein	1E-129
aCL7254Contig1	At4g33510	2-dehydro-3-deoxyphosphoheptonate aldolase 2 (DHS2)	0
aCL7199Contig1	At3g26020	serine/threonine protein phosphatase 2A (PP2A) subunit B'	1E-134
aCL7164Contig1	At5g16110	expressed protein	1E-31
aCL7152Contig1	At3g27925	DegP protease	2E-47
aCL7136Contig1	At5g58330	malate dehydrogenase (NADP)	1E-150
aCL7110Contig1	At3g63030	methyl-CpG-binding domain-containing protein	8E-59
aCL70Contig1	At3g05500	rubber elongation factor (REF) family protein	3E-54
aCL7069Contig1	At1g70580	similar to glutamate:glyoxylate aminotransferase 1 (GGT1)	7E-71
aCL705Contig2	At3g26740	light responsive protein-related /ccr protein	5E-30
aCL7045Contig1	At3g46900	copper transporter (COPT1)	4E-29
aCL7044Contig1	At5g17170	rubredoxin family protein	3E-73
aCL701Contig1	At5g54110	vesicle-associated membrane family protein (VAMP)	3E-79
aCL7016Contig1	At1g74970	ribosomal protein S9 (RPS9)	3E-72
aCL7006Contig1	At3g59660	C2 domain-containing protein	6E-13
aCL6Contig22	At4g02380	late embryogenesis abundant 3 family protein (LEA3)	7E-17
aCL6Contig16	At1g54410	dehydrin family protein	8E-09
aCL698Contig2	At1g64090	reticulon family protein (RTNLB3)	4E-72
aCL6983Contig1	At1g05570	similar to glycosyl transferase family 48 protein	1E-84
aCL693Contig1	At5g67300	myb family transcription factor	7E-63
aCL6925Contig1	At1g29900	carbamoyl-phosphate synthase family protein	1E-118
aCL6902Contig1	At1g64470	ubiquitin family protein	2E-53
aCL6892Contig1	At1g20760	calcium-binding EF hand family protein	2E-89
aCL688Contig1	At3g52850	vacuolar sorting receptor	0
aCL6887Contig1	At2g35110	similar to NCK-associated protein 1	1E-141
aCL6882Contig1	At3g62240	zinc finger (C2H2 type) family protein	3E-58
aCL683Contig1	At2g30020	protein phosphatase 2C (PP2C)	1E-104
aCL6819Contig1	At1g02680	transcription initiation factor IID (TFIID)	5E-41
aCL6777Contig1	At5g08480	VQ motif-containing protein	2E-43
aCL6772Contig1	At4g37800	xyloglucan:xyloglucosyl transferase	1E-14
aCL674Contig2	At2g28000	RuBisCO subunit binding-protein alpha subunit	1E-102
aCL6690Contig1	At2g12400	expressed protein	1E-27
aCL664Contig3	At3g11940	40S ribosomal protein S5 (RPS5B)	7E-98
aCL664Contig1	At3g11940	40S ribosomal protein S5 (RPS5B)	7E-98

aCL6618Contig1	At3g11240	arginine-tRNA-protein transferase	1E-52
aCL6617Contig1	At5g41150	repair endonuclease (RAD1)	1E-126
aCL6602Contig1	At3g11920	glutaredoxin-related	1E-124
aCL6596Contig1	At5g55140	ribosomal protein L30 family protein	7E-37
aCL6567Contig1	At3g12220	serine carboxypeptidase S10 family protein	8E-63
aCL6560Contig1	At2g31880	leucine-rich repeat transmembrane protein kinase	1E-56
aCL6558Contig1	At3g50830	stress-responsive protein (WCOR413-like)	5E-77
aCL654Contig1	At2g29420	glutathione S-transferase	1E-55
aCL6546Contig1	At4g28025	expressed protein	4E-34
aCL64Contig1	At1g65260	PspA/IM30 family protein	1E-112
aCL6476Contig1	At1g79930	similar to heat shock protein 70 (HSP70)	1E-117
aCL646Contig1	At5g56600	profilin 5 (PRO5)	1E-58
aCL6447Contig1	At1g79850	30S ribosomal protein S17 (RPS17)	1E-33
aCL6424Contig1	At3g19640	magnesium transporter CorA-like family protein (MRS2-3)	4E-55
aCL641Contig1	At4g27000	RNA-binding protein 45 (RBP45)	6E-59
aCL6391Contig1	At4g26450	expressed protein	3E-46
aCL6355Contig1	At1g21170	expressed protein	1E-144
aCL6348Contig1	At3g59300	expressed protein	4E-21
aCL632Contig1	At4g38770	proline-rich family protein (PRP4)	1E-31
aCL6289Contig1	At1g27700	expressed protein	4E-55
aCL6272Contig1	At1g70210	cyclin delta-1 (CYCD1)	1E-23
aCL6259Contig1	At2g46220	expressed protein	9E-73
aCL622Contig2	At1g05260	peroxidase 3 (PER3)	2E-90
aCL6226Contig1	At4g10300	expressed protein	8E-47
aCL61Contig2	At5g23950	C2 domain-containing protein	3E-43
aCL619Contig2	At3g49910	60S ribosomal protein L26 (RPL26A)	7E-54
aCL6192Contig1	At4g02020	zeste-like protein 1 (EZA1)	1E-142
aCL6161Contig1	At1g55740	alkaline alpha galactosidase	1E-138
aCL6131Contig1	At3g12860	nucleolar protein Nop56	2E-80
aCL6104Contig1	At1g03090	methylcrotonyl-CoA carboxylase alpha chain (MCCA)	2E-36
aCL6075Contig1	At4g00710	protein kinase family protein	1E-110
aCL6031Contig1	At1g01900	subtilase family protein	5E-28
aCL5Contig3	At5g47560	encodes a tonoplast malate/fumarate transporter	0
aCL5Contig25	At2g21660	glycine-rich RNA-binding protein (GRP7)	9E-19
aCL5Contig23	At2g21660	glycine-rich RNA-binding protein (GRP7)	2E-36
aCL5Contig15	At3g22840	chlorophyll A-B binding family protein (ELIP)	5E-52
aCL5Contig1	At5g47560	encodes a tonoplast malate/fumarate transporter	3E-77
aCL59Contig4	At2g27830	expressed protein	1E-27
aCL599Contig2	At2g17390	ankyrin repeat family protein	1E-117
aCL5982Contig1	At2g41500	WD-40 repeat family protein	1E-145
aCL5964Contig1	At1g75880	family II extracellular lipase 1 (EXL1)	4E-72
aCL5961Contig1	At3g50560	short-chain dehydrogenase/reductase (SDR) family protein	1E-103
aCL5939Contig1	At1g54570	esterase/lipase/thioesterase family protein	3E-88
aCL58Contig5	At4g23400	major intrinsic family protein / MIP family protein	1E-143
aCL5894Contig1	At1g21680	expressed protein	3E-86
aCL588Contig2	At4g38460	geranylgeranyl pyrophosphate synthase	1E-103
aCL588Contig1	At4g38460	geranylgeranyl pyrophosphate synthase	6E-99
aCL5880Contig1	At5g20230	plastocyanin-like domain-containing protein	1E-18
aCL5871Contig1	At4g24620	glucose-6-phosphate isomerase	2E-89
aCL585Contig2	At1g79450	LEM3 (ligand-effect modulator 3) family protein	1E-103
aCL585Contig1	At1g79450	LEM3 (ligand-effect modulator 3) family protein	1E-65
aCL5811Contig1	At4g29330	Der1-like family protein	3E-87
aCL579Contig2	At5g39740	60S ribosomal protein L5 (RPL5B)	1E-130
aCL578Contig1	At3g12120	omega-6 fatty acid desaturase (FAD2)	0
aCL5781Contig1	At5g58250	expressed protein	7E-55
aCL5775Contig1	At4g24440	transcription initiation factor IIA gamma chain (TFIIA-S)	3E-50
aCL5773Contig1	At2g42010	phospholipase D beta 1 / PLD beta 1 (PLDBETA1)	1E-140
aCL5758Contig1	At3g22845	emp24/gp25L/p24 protein-related	8E-54
aCL5753Contig1	At4g11410	short-chain dehydrogenase/reductase (SDR) family protein	8E-52
aCL5724Contig1	At1g42440	expressed protein	1E-117
aCL5721Contig1	At1g08470	strictosidine synthase family protein	1E-119

aCL5713Contig1	At1g72160	phosphoglyceride transfer family protein (SEC14)	1E-12
aCL5709Contig1	At5g60580	zinc finger (C3HC4-type RING finger) family protein	1E-31
aCL5706Contig1	At2g14835	zinc finger (C3HC4-type RING finger) family protein	1E-141
aCL5695Contig1	At5g64840	ABC transporter family protein	1E-105
aCL5687Contig1	At1g56420	expressed protein	8E-52
aCL5671Contig1	At2g27920	serine carboxypeptidase S10 family protein	1E-139
aCL5665Contig1	At4g17390	60S ribosomal protein L15 (RPL15B)	1E-90
aCL5662Contig1	At3g03920	Gar1 RNA-binding region family protein	9E-42
aCL5654Contig1	At2g23600	hydrolase	3E-46
aCL5646Contig1	At2g37410	mitochondrial import inner membrane translocase (TIM17)	1E-70
aCL5641Contig1	At4g10330	glycine-rich protein	4E-11
aCL5634Contig1	At2g33250	expressed protein	8E-09
aCL5607Contig1	At5g25140	cytochrome P450 family protein	7E-54
aCL5601Contig1	At1g66670	ATP-dependent Clp protease proteolytic subunit (ClpP3)	5E-87
aCL55Contig2	At4g01850	S-adenosylmethionine synthetase 2 (SAM2)	0
aCL5590Contig1	At1g49320	BURP domain-containing protein	3E-34
aCL5589Contig1	At3g13080	ABC transporter family protein	1E-136
aCL5570Contig1	At2g36970	UDP-glucuronosyl/UDP-glucosyl transferase family protein	1E-161
aCL5503Contig1	At5g63320	similar to DNA-binding bromodomain-containing protein	5E-17
aCL548Contig1	At3g09770	zinc finger (C3HC4-type RING finger) family protein	1E-105
aCL5488Contig1	At3g51980	expressed protein	1E-121
aCL5453Contig1	At5g19150	carbohydrate kinase family	1E-94
aCL5448Contig1	At1g64355	expressed protein	3E-50
aCL5440Contig1	At1g70660	ubiquitin-conjugating enzyme family protein	4E-68
aCL5437Contig1	At3g07630	prephenate dehydratase family protein	1E-136
aCL5434Contig1	At3g18890	expressed protein	2E-58
aCL5422Contig1	At3g12760	expressed protein	1E-40
aCL5413Contig1	At1g30360	early-responsive to dehydration stress protein (ERD4)	4E-86
aCL5407Contig1	At1g31010	expressed protein	1E-25
aCL5405Contig1	At5g19420	zinc finger protein	1E-130
aCL5383Contig1	At5g53350	ATP-dependent Clp protease ATP-binding subunit ClpX1	1E-29
aCL5381Contig1	At2g25920	expressed protein	1E-55
aCL535Contig3	At3g10190	calmodulin	1E-46
aCL5340Contig1	At1g73940	expressed protein	4E-39
aCL5338Contig1	At3g19980	serine/threonine protein phosphatase (STPP)	1E-105
aCL5327Contig1	At5g58220	expressed protein	1E-89
aCL52Contig3	At3g01120	cystathionine gamma-synthase (CGS)	1E-79
aCL52Contig1	At3g01120	cystathionine gamma-synthase (CGS)	0
aCL5281Contig1	At5g03740	zinc finger (C2H2 type) family protein	3E-24
aCL5262Contig1	At3g22200	4-aminobutyrate aminotransferase	4E-92
aCL5261Contig1	At3g61490	glycoside hydrolase family 28 protein	1E-170
aCL5253Contig1	At1g17020	oxidoreductase, 2OG-Fe(II) oxygenase family protein	2E-51
aCL5224Contig1	At5g67130	expressed protein	1E-154
aCL5223Contig1	At5g09930	ABC transporter family protein	1E-101
aCL5213Contig1	At4g02630	protein kinase family protein	4E-83
aCL5188Contig1	At2g46820	expressed protein	1E-40
aCL517Contig2	At2g18050	histone H1-3 (HIS1-3)	9E-24
aCL5178Contig1	At1g32230	WWE domain-containing protein (CEO)	9E-57
aCL5170Contig1	At3g11730	Ras-related GTP-binding protein	3E-98
aCL5158Contig1	At3g27310	expressed protein	3E-68
aCL5148Contig1	At3g47550	zinc finger (C3HC4-type RING finger) family protein	4E-77
aCL5143Contig1	At4g15920	nodulin MtN3 family protein	7E-58
aCL5140Contig1	At2g24050	MIF4G domain-containing protein	3E-74
aCL512Contig1	At1g71880	sucrose transporter (SUC1)	1E-177
aCL5121Contig1	At1g71480	nuclear transport factor 2 (NTF2) family protein	5E-54
aCL5116Contig1	At3g11780	MD-2-related lipid recognition domain-containing protein	3E-56
aCL508Contig1	At3g58610	ketol-acid reductoisomerase	5E-26
aCL5079Contig1	At1g59950	aldo/keto reductase	2E-97
aCL5060Contig1	At3g26090	expressed protein	6E-95
aCL505Contig2	At2g02010	glutamate decarboxylase	0
aCL504Contig1	At5g58240	bis(5'-adenosyl)-triphosphatase	6E-55

aCL5023Contig1	At2g27050	ethylene-insensitive3-like1 (EIL1)	7E-22
aCL5001Contig1	At2g45550	cytochrome P450 family protein	1E-51
aCL4Contig31	At4g05320	polyubiquitin (UBQ10)	0
aCL4Contig29	At5g03240	similar to polyubiquitin (UBQ10)	1E-125
aCL4Contig25	At5g20620	polyubiquitin (UBQ4)	1E-168
aCL4962Contig1	At5g04600	RNA recognition motif (RRM)-containing protein	6E-81
aCL4949Contig1	At2g36910	multidrug resistance P-glycoprotein (PGP1)	4E-60
aCL493Contig1	At3g12390	nascent polypeptide associated complex (alpha-NAC)	7E-55
aCL4921Contig2	At3g15095	expressed protein	5E-25
aCL48Contig3	At2g39730	RuBisCO activase	1E-169
aCL4899Contig1	At1g16880	uridylyltransferase-related	4E-99
aCL4893Contig1	At5g60840	expressed protein	7E-11
aCL4880Contig1	At1g73950	zinc finger (C3HC4-type RING finger) family protein	4E-68
aCL4868Contig1	At1g28050	zinc finger (B-box type) family protein	6E-38
aCL4860Contig1	At4g01050	hydroxyproline-rich glycoprotein family protein	5E-35
aCL4838Contig1	At1g76080	thioredoxin family protein	7E-65
aCL4830Contig1	At2g06530	SNF7 family protein	8E-97
aCL4829Contig1	At5g48870	small nuclear ribonucleoprotein	1E-41
aCL4826Contig1	At3g26580	expressed protein	7E-57
aCL4823Contig1	At4g32260	ATP synthase family	2E-49
aCL480Contig1	At1g52760	esterase/lipase/thioesterase family protein	1E-162
aCL477Contig2	At4g17900	zinc-binding family protein	1E-75
aCL474Contig1	At3g19290	ABA-responsive element-binding protein 2 (AREB2)	3E-76
aCL4746Contig1	At3g46460	ubiquitin-conjugating enzyme 13 (UBC13)	1E-73
aCL473Contig1	At1g19140	expressed protein	4E-96
aCL472Contig3	At1g51090	heavy-metal-associated domain-containing protein	1E-16
aCL472Contig1	At4g16380	expressed protein	6E-12
aCL4722Contig1	At2g36880	S-adenosylmethionine synthetase	1E-142
aCL4712Contig1	At5g17920	encodes METS protein	1E-179
aCL4708Contig1	At5g63530	copper chaperone (CCH)-related	6E-36
aCL4705Contig1	At2g23940	expressed protein	3E-68
aCL46Contig3	At1g24020	Bet v I allergen family protein	4E-06
aCL4685Contig1	At5g41600	reticulon family protein (RTNLB4)	2E-84
aCL4665Contig1	At4g29090	reverse transcriptase	1E-28
aCL4657Contig1	AtCg00510	PSI I protein	8E-10
aCL4650Contig1	At1g50380	prolyl oligopeptidase family protein	2E-21
aCL4635Contig1	At4g36910	CBS domain-containing protein	4E-83
aCL462Contig1	At2g47110	ubiquitin extension protein 6 (UBQ6)	7E-68
aCL4623Contig1	At2g45670	calcineurin B subunit-related	5E-68
aCL4614Contig1	At5g22460	esterase/lipase/thioesterase family protein	2E-88
aCL4608Contig1	At3g07720	kelch repeat-containing protein	3E-50
aCL4607Contig1	At1g14510	PHD finger family protein	2E-15
aCL45Contig2	At3g48690	expressed protein	2E-96
aCL4595Contig1	At1g71090	auxin efflux carrier family protein	1E-125
aCL4593Contig1	At4g17600	lii3 protein	8E-65
aCL4589Contig1	At1g07040	expressed protein	1E-134
aCL4562Contig1	AtCg01100	NADH dehydrogenase ND1	2E-81
aCL4558Contig1	At5g26360	chaperonin	1E-130
aCL4551Contig1	At3g27890	NADPH-dependent FMN reductase family protein	2E-76
aCL4542Contig1	At4g25670	expressed protein	2E-33
aCL4540Contig1	At4g29735	expressed protein	1E-17
aCL4520Contig1	At1g49410	expressed protein	3E-14
aCL4515Contig1	At3g16500	auxin-responsive AUX/IAA family protein	3E-61
aCL4509Contig1	At4g03520	thioredoxin M-type 2 (TRX-M2)	3E-39
aCL44Contig4	At3g10260	reticulon family protein	1E-75
aCL44Contig2	At3g10260	reticulon family protein	7E-66
aCL4472Contig1	At2g31200	actin-depolymerizing factor 6 (ADF6)	1E-60
aCL4463Contig1	At3g05270	expressed protein	1E-41
aCL4460Contig1	At3g06050	encodes a mitochondrial peroxiredoxin	5E-82
aCL4446Contig1	At3g52960	peroxiredoxin type 2	9E-74
aCL4445Contig1	At2g27290	expressed protein	7E-44

aCL4424Contig1	At1g64980	expressed protein	4E-76
aCL4423Contig1	At2g29530	mitochondrial import inner membrane translocase (TIM10)	2E-36
aCL440Contig1	At1g49760	polyadenylate-binding protein (PABP)	0
aCL43Contig2	At1g67090	RuBisCO small subunit 1A (RBCS-1A)	5E-32
aCL4395Contig1	At3g27090	expressed protein	1E-113
aCL438Contig2	At3g22890	sulfate adenylyltransferase 1 / ATP-sulfurylase 1 (APS1)	1E-160
aCL4360Contig1	At5g16620	similar to chloroplast protein translocon component Tic40	1E-21
aCL4355Contig1	At3g27950	early nodule-specific protein	1E-111
aCL4354Contig1	At3g07565	expressed protein	1E-102
aCL4319Contig1	At3g61220	short-chain dehydrogenase/reductase (SDR) family protein	6E-95
aCL4318Contig1	At1g04260	prenylated rab acceptor (PRA1) family protein	8E-42
aCL4288Contig1	At5g04560	DEMETER protein (DME)	0
aCL4283Contig1	At3g12080	GTP-binding family protein	4E-61
aCL4274Contig1	At2g41430	dehydration-induced protein (ERD15)	6E-13
aCL426Contig1	At1g10830	sodium symporter-related	1E-129
aCL4261Contig1	At1g10760	starch excess protein (SEX1)	1E-146
aCL4244Contig1	At3g16920	glycoside hydrolase family 19 protein	1E-150
aCL422Contig1	At4g01610	cathepsin B-like cysteine protease	5E-79
aCL4218Contig1	At1g75280	isoflavone reductase	1E-133
aCL4187Contig1	At4g34590	bZIP transcription factor family protein	3E-34
aCL4180Contig1	At1g34210	somatic embryogenesis receptor-like kinase 2 (SERK2)	0
aCL4174Contig1	At2g43780	expressed protein	5E-19
aCL4159Contig1	At5g14240	expressed protein	4E-81
aCL4152Contig1	At5g06060	tropinone reductase	8E-71
aCL4148Contig1	At2g45850	DNA-binding family protein	9E-54
aCL4135Contig1	At5g13590	expressed protein	2E-47
aCL4116Contig2	At5g53370	pectinesterase family protein	1E-115
aCL4114Contig1	At3g54250	mevalonate diphosphate decarboxylase	1E-131
aCL410Contig2	At1g02880	thiamin pyrophosphokinase	6E-76
aCL4006Contig2	At5g24470	pseudo-response regulator 5 (APRR5)	4E-23
aCL4004Contig1	At1g69360	expressed protein	2E-90
aCL3Contig5	At5g59310	lipid transfer protein 4 (LTP4)	4E-26
aCL3Contig4	At1g60710	aldo/keto reductase family protein	1E-159
aCL3Contig35	At2g17200	ubiquitin family protein	1E-130
aCL3Contig18	At1g28330	dormancy-associated protein (DRM1)	4E-39
aCL398Contig2	At5g46110	phosphate/triose-phosphate translocator	1E-115
aCL398Contig1	At5g46110	phosphate/triose-phosphate translocator	1E-173
aCL3964Contig1	At2g26510	xanthine/uracil permease family protein	0
aCL3963Contig1	At1g08530	expressed protein	2E-49
aCL3949Contig1	At1g68090	calcium-binding protein annexin 5 (ANN5)	3E-91
aCL3934Contig1	At1g13560	aminoalcoholphosphotransferase (AAPT1)	0
aCL391Contig1	At1g70330	equilibrative nucleoside transporter family protein	1E-147
aCL3916Contig1	At3g15850	fatty acid desaturase family protein (ADS1)	5E-81
aCL390Contig2	At2g32670	synaptobrevin family protein	4E-86
aCL38Contig1	At5g54160	quercetin 3-O-methyltransferase 1(OMT1)	2E-91
aCL388Contig1	At2g18730	diacylglycerol kinase	0
aCL3889Contig2	At1g14900	high-mobility-group protein / HMG-I/Y protein	1E-20
aCL3887Contig1	At2g45740	similar to peroxisomal biogenesis factor 11 family protein	1E-106
aCL3882Contig1	At1g22620	SAC domain phosphoinositide (3,5)P2 phosphatase	1E-129
aCL386Contig1	At5g47120	Bax inhibitor-1 putative / BI-1 putative	4E-89
aCL3863Contig1	At1g78630	ribosomal protein L13 family protein	2E-72
aCL385Contig1	At1g48910	flavin-containing monooxygenase family protein	1E-110
aCL3858Contig1	At2g27040	PAZ domain-containing protein	0
aCL3856Contig1	At3g48800	sterile alpha motif (SAM) domain-containing protein	1E-31
aCL3855Contig1	At4g28770	expressed protein	1E-65
aCL384Contig1	At3g47850	expressed protein	2E-39
aCL3840Contig1	At1g23310	glutamate:glyoxylate aminotransferase 1 (GGT1)	1E-164
aCL3830Contig2	At5g61600	encodes a member of ERF/AP2 transcription factor family	4E-32
aCL3829Contig1	At2g25770	expressed protein	3E-32
aCL3818Contig1	At1g63940	monodehydroascorbate reductase	0
aCL3804Contig1	At5g22950	SNF7 family protein	9E-99

aCL3795Contig2	At3g13510	expressed protein	1E-109
aCL3788Contig1	At1g56300	DNAJ heat shock N-terminal domain-containing protein	5E-55
aCL3787Contig1	At5g40760	encodes a cytosolic glucose-6-phosphate dehydrogenase	0
aCL3776Contig1	At5g02810	pseudo-response regulator 7 (APRR7)	3E-25
aCL3764Contig1	At2g04700	ferredoxin thioredoxin reductase family protein	8E-59
aCL3762Contig1	At5g38650	proteasome maturation factor UMP1 family protein	1E-53
aCL3761Contig1	At2g29420	glutathione S-transferase	5E-48
aCL375Contig2	At2g46080	expressed protein	1E-123
aCL3758Contig2	At3g58500	serine/threonine protein phosphatase PP2A-4 (PP2A4)	7E-79
aCL3741Contig1	At2g17410	ARID/BRIGHT DNA-binding domain-containing protein	1E-40
aCL3734Contig1	At4g01150	expressed protein	2E-41
aCL3716Contig1	At4g31880	expressed protein	5E-40
aCL36Contig3	At4g21960	peroxidase 42 (PER42) (P42) (PRXR1)	1E-158
aCL369Contig3	At3g18280	protease inhibitor/seed storage/lipid transfer protein (LTP)	2E-23
aCL3697Contig1	At3g54050	fructose-1,6-bisphosphatase	1E-103
aCL3690Contig1	At2g16250	leucine-rich repeat transmembrane protein kinase	1E-105
aCL3689Contig1	At1g27950	lipid transfer protein-related	1E-31
aCL3682Contig1	At3g02790	zinc finger (C2H2 type) family protein	4E-43
aCL3650Contig1	At4g01310	ribosomal protein L5 family protein	1E-109
aCL3638Contig1	At3g59670	expressed protein	9E-49
aCL3637Contig1	At1g74270	60S ribosomal protein L35a (RPL35aC)	2E-57
aCL3625Contig1	At3g58170	Bet1-like SNARE 1-1 (BET11)	1E-52
aCL3622Contig1	At5g11760	expressed protein	7E-11
aCL3620Contig1	At4g35450	ankyrin repeat family protein (AFT)	7E-14
aCL3589Contig1	At2g33730	DEAD box RNA helicase	2E-86
aCL3567Contig1	At2g26440	pectinesterase family protein	0
aCL3558Contig1	At4g13500	expressed protein	8E-37
aCL3550Contig1	At3g01990	ACT domain-containing protein (ACR6)	1E-171
aCL3521Contig1	At1g18800	nucleosome assembly protein (NAP) family protein	7E-99
aCL3498Contig1	At1g50250	encodes an FTSH protease	0
aCL3487Contig1	At1g27650	U2 snRNP auxiliary factor small subunit	1E-95
aCL3461Contig1	At5g05000	translocate of chloroplast 34 (TOC34)	1E-119
aCL3457Contig1	At3g27925	DegP protease	1E-62
aCL3450Contig1	At3g08510	phosphoinositide-specific phospholipase C family protein	6E-97
aCL3436Contig1	At5g13030	expressed protein	1E-120
aCL3429Contig1	At2g25810	tonoplast intrinsic protein	1E-101
aCL341Contig1	At2g39670	radical SAM domain-containing protein	1E-136
aCL334Contig1	At3g16640	translationally controlled tumor family protein	2E-69
aCL3343Contig1	At5g54160	quercetin 3-O-methyltransferase 1 (OMT1)	1E-163
aCL3324Contig1	At2g04900	expressed protein	5E-43
aCL3303Contig1	At2g04030	heat shock protein	0
aCL329Contig2	At1g31330	photosystem I reaction center subunit III family protein	3E-79
aCL329Contig1	At1g31330	photosystem I reaction center subunit III family protein	4E-79
aCL3292Contig1	At3g23490	cyanate lyase family	4E-57
aCL3290Contig1	At3g15840	expressed protein	7E-65
aCL3276Contig1	At1g52740	histone H2A	1E-44
aCL3260Contig1	At1g78020	senescence-associated protein-related (SAG102)	2E-21
aCL3251Contig1	At4g08170	inositol 1,3,4-trisphosphate 5/6-kinase family protein	1E-131
aCL3250Contig1	At1g17620	expressed protein	7E-53
aCL324Contig2	At3g58180	PBS lyase HEAT-like repeat-containing protein	1E-106
aCL3241Contig1	At1g33970	similar to avirulence induced gene (AIG1)	7E-57
aCL3236Contig1	At4g36900	encodes a member of ERF/AP2 transcription factor family	9E-46
aCL3233Contig1	At5g02270	ABC transporter family protein	1E-138
aCL322Contig1	At1g03400	2-oxoglutarate-dependent dioxygenase	7E-82
aCL3214Contig1	At4g18100	60S ribosomal protein L32 (RPL32A)	2E-57
aCL3210Contig1	At4g32410	cellulose synthase, catalytic subunit	0
aCL319Contig1	At1g25550	myb family transcription factor	3E-77
aCL317Contig1	At4g30690	translation initiation factor 3 (IF-3) family protein	7E-82
aCL3172Contig1	At5g41990	protein kinase family protein	1E-60
aCL3166Contig1	At5g55310	DNA topoisomerase I	1E-173
aCL315Contig1	At5g62790	1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR)	0

aCL3157Contig1	At2g46970	basic helix-loop-helix (bHLH) protein	7E-31
aCL3155Contig1	At5g57040	lactoylglutathione lyase family protein	7E-70
aCL314Contig2	At5g20700	senescence-associated protein-related (SAG102)	3E-31
aCL3144Contig1	At5g02502	expressed protein	1E-11
aCL3140Contig1	At4g02770	photosystem I reaction center subunit II (PSI-D)	3E-78
aCL3133Contig2	At3g16370	GDSL-motif lipase/hydrolase family protein	1E-144
aCL3133Contig1	At3g16370	GDSL-motif lipase/hydrolase family protein	1E-128
aCL3121Contig1	At4g03560	two-pore calcium channel (TPC1)	8E-96
aCL3118Contig1	At3g06190	speckle-type POZ protein-related	0
aCL3113Contig1	At5g05100	expressed protein	2E-42
aCL310Contig2	At4g26850	expressed protein	0
aCL30Contig1	At2g33150	acetyl-CoA C-acyltransferase	3E-87
aCL3096Contig1	At5g65220	ribosomal protein L29 family protein	8E-41
aCL3063Contig1	At5g57970	similar to methyladenine glycosylase family protein	1E-111
aCL304Contig1	At5g15490	UDP-glucose 6-dehydrogenase (UDPGDH)	0
aCL3042Contig1	At1g60420	DC1 domain-containing protein	9E-50
aCL3041Contig1	At2g43420	3-beta hydroxysteroid dehydrogenase family protein	2E-95
aCL3030Contig1	At4g27435	expressed protein	8E-63
aCL302Contig2	At5g42520	expressed protein	3E-80
aCL3025Contig1	At4g37930	glycine hydroxymethyltransferase (SHM1)	1E-131
aCL3017Contig1	At5g11710	epsin N-terminal homology (ENTH) domain	1E-150
aCL2Contig14	At4g34350	ISPH (nonmevalonate pathway of isoprenoid biosynthesis)	0
aCL2Contig13	At4g34350	ISPH (nonmevalonate pathway of isoprenoid biosynthesis)	1E-119
aCL29Contig4	At3g04120	glyceraldehyde-3-phosphate dehydrogenase (GAPC)	1E-168
aCL2997Contig1	At1g23740	oxidoreductase, zinc-binding dehydrogenase family protein	5E-59
aCL2995Contig1	At1g76010	expressed protein	5E-60
aCL297Contig2	At1g17860	trypsin and protease inhibitor family protein	1E-15
aCL297Contig1	At1g17860	trypsin and protease inhibitor family protein	9E-14
aCL2963Contig1	At5g57300	UbiE/COQ5 methyltransferase family protein	1E-122
aCL2953Contig2	At5g66190	ferredoxin--NADP(+) reductase	1E-170
aCL2936Contig1	At5g60570	kelch repeat-containing F-box family protein	8E-63
aCL2935Contig1	At5g20160	ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	5E-54
aCL292Contig1	At5g14970	expressed protein	6E-83
aCL2924Contig1	At5g64030	dehydration-responsive protein-related (ERD3)	0
aCL2911Contig1	At5g01590	expressed protein	1E-172
aCL2903Contig1	At1g59950	aldo/keto reductase	1E-119
aCL2891Contig1	At5g24690	expressed protein	7E-48
aCL2874Contig1	At4g00755	F-box family protein	2E-97
aCL2872Contig1	At4g12800	photosystem I reaction center subunit XI (PSI-L)	1E-85
aCL2870Contig1	At1g69120	floral homeotic protein APETALA1 (AP1)	6E-63
aCL2866Contig1	At1g35420	dienelactone hydrolase family protein	1E-94
aCL285Contig2	At3g57490	40S ribosomal protein S2 (RPS2D)	1E-112
aCL2845Contig1	At4g03080	kelch repeat-containing phosphoesterase family protein	0
aCL2840Contig1	At4g36980	expressed protein	1E-110
aCL2832Contig1	At4g16380	expressed protein	4E-17
aCL2820Contig1	At1g59990	DEAD/DEAH box helicase	2E-80
aCL281Contig1	At1g42970	glyceraldehyde-3-phosphate dehydrogenase B (GAPB)	0
aCL2805Contig2	At2g42750	DNAJ heat shock N-terminal domain-containing protein	3E-85
aCL2801Contig1	At1g05070	expressed protein	8E-47
aCL27Contig2	At5g13930	chalcone synthase / naringenin-chalcone synthase	0
aCL27Contig1	At1g73990	peptidase U7 family protein	7E-38
aCL2799Contig1	At1g79590	similar to syntaxin 51 (SYP51)	6E-88
aCL2795Contig1	At5g56940	ribosomal protein S16 family protein	5E-57
aCL2785Contig1	At5g62360	invertase/pectin methylesterase inhibitor family protein	3E-40
aCL2767Contig1	At3g49680	branched-chain amino acid transaminase 3 (BCAT3)	1E-158
aCL2758Contig1	At4g14342	pre-mRNA splicing factor 10 kDa subunit	4E-41
aCL273Contig1	At5g33370	GDSL-motif lipase/hydrolase family protein	1E-147
aCL2738Contig1	At1g75780	tubulin beta-1 chain (TUB1)	0
aCL2732Contig2	At1g32210	defender against cell death 1 (DAD1)	8E-53
aCL2722Contig1	At5g35630	similar to glutamine synthetase	0
aCL26Contig9	At5g39670	calcium-binding EF hand family protein	7E-11

aCL26Contig1	At5g26990	drought-responsive family protein	3E-56
aCL269Contig1	At2g22780	malate dehydrogenase	1E-170
aCL2696Contig1	At3g22630	20S proteasome beta subunit D (PBD1)	3E-99
aCL2691Contig1	At5g12250	tubulin beta-6 chain (TUB6)	0
aCL268Contig1	At3g53020	regulated by TCP20	1E-53
aCL2668Contig1	At3g49990	expressed protein	2E-36
aCL265Contig2	At5g20720	20 kDa chaperonin (CPN21)	1E-98
aCL2657Contig1	At1g47420	expressed protein	5E-70
aCL264Contig1	At2g02180	tobamovirus multiplication protein 3 (TOM3)	1E-111
aCL2637Contig1	At1g79420	expressed protein	1E-130
aCL2630Contig1	At2g18290	anaphase-promoting complex, subunit 10 family	7E-91
aCL2623Contig1	At1g52570	phospholipase D alpha 2 (PLD2)	0
aCL261Contig1	At4g39235	expressed protein	1E-27
aCL2617Contig1	At3g61320	expressed protein	8E-68
aCL2615Contig1	At3g51630	protein kinase family protein	1E-143
aCL2591Contig2	At2g20360	expressed protein	1E-155
aCL2579Contig2	At1g03130	photosystem I reaction center subunit II (PSI-D)	2E-82
aCL2565Contig1	At1g56700	pyrrolidone-carboxylate peptidase family protein	3E-92
aCL2559Contig1	At1g65720	expressed protein	1E-28
aCL2555Contig1	At2g44100	Rab GDP dissociation inhibitor (GDI1)	0
aCL2553Contig1	At4g30930	50S ribosomal protein L21 (RPL21M)	8E-49
aCL2546Contig1	At3g63190	ribosome recycling factor (RRF)	3E-84
aCL2535Contig1	At1g18740	expressed protein	1E-101
aCL2511Contig1	At2g02760	ubiquitin-conjugating enzyme 2 (UBC2)	1E-85
aCL2501Contig1	At4g33550	protease inhibitor/seed storage/lipid transfer protein (LTP)	2E-16
aCL2500Contig2	At1g75510	transcription initiation factor IIF beta subunit (TFIIF-beta)	6E-70
aCL2500Contig1	At1g75510	transcription initiation factor IIF beta subunit (TFIIF-beta)	4E-97
aCL24Contig5	At3g52590	ubiquitin extension protein 1 (UBQ1)	6E-69
aCL248Contig2	At5g06290	2-cys peroxiredoxin	1E-102
aCL248Contig1	At5g06290	2-cys peroxiredoxin	1E-106
aCL2478Contig1	At5g52650	40S ribosomal protein S10 (RPS10C)	1E-45
aCL2471Contig1	At5g34850	calcineurin-like phosphoesterase family protein	0
aCL2470Contig2	At2g47710	universal stress protein (USP) family protein	1E-63
aCL243Contig1	At1g76020	expressed protein	6E-63
aCL2434Contig2	At3g25140	glycosyl transferase family 8 protein	1E-103
aCL241Contig2	At1g72520	lipoxygenase	7E-36
aCL2419Contig1	At5g08560	transducin family protein / WD-40 repeat family protein	0
aCL2410Contig1	At4g11150	vacuolar ATP synthase subunit E (VATE)	3E-88
aCL2404Contig1	At2g30050	transducin family protein / WD-40 repeat family protein	1E-153
aCL2398Contig1	At3g10950	60S ribosomal protein L37a (RPL37aB)	2E-46
aCL2388Contig2	At1g52230	photosystem I reaction center subunit VI (PSI-H)	2E-45
aCL2386Contig2	At4g31860	protein phosphatase 2C (PP2C)	1E-172
aCL2385Contig1	At3g06680	60S ribosomal protein L29 (RPL29B)	3E-25
aCL2370Contig1	At4g00870	basic helix-loop-helix (bHLH) family protein	2E-45
aCL236Contig1	At1g23440	pyrrolidone-carboxylate peptidase family protein	6E-92
aCL2356Contig1	At2g43750	cysteine synthase	1E-152
aCL2352Contig1	At1g07080	GILT family protein	1E-74
aCL2336Contig1	At2g20260	photosystem I reaction center subunit IV (PSI-E)	8E-25
aCL232Contig1	At5g48480	expressed protein	2E-18
aCL2305Contig1	At3g52560	ubiquitin-conjugating enzyme family protein	2E-75
aCL2302Contig1	At1g01220	GHMP kinase-related	1E-122
aCL2291Contig1	At4g38800	phosphorylase family protein	2E-94
aCL2267Contig1	At5g25754	expressed protein	0
aCL225Contig4	At5g59910	histone H2B	3E-45
aCL225Contig1	At2g27385	expressed protein	5E-32
aCL2247Contig1	At1g15780	expressed protein	6E-28
aCL2246Contig1	At4g01590	similar to expressed protein	1E-16
aCL2222Contig1	At3g54360	expressed protein	1E-140
aCL2219Contig1	At4g02050	sugar transporter	0
aCL2218Contig2	At5g42150	expressed protein	1E-77
aCL2217Contig1	At5g09420	chloroplast outer membrane translocon subunit	3E-82

aCL2213Contig1	At5g01820	CBL-interacting protein kinase 14 (CIPK14)	1E-120
aCL2203Contig2	At5g02020	expressed protein	9E-08
aCL21Contig6	At5g60390	elongation factor 1-alpha (EF-1-alpha)	1E-129
aCL2193Contig1	At5g16050	14-3-3 protein GF14 epsilon (GRF5)	1E-127
aCL2192Contig1	At1g78870	ubiquitin-conjugating enzyme	4E-84
aCL218Contig2	At1g16300	glyceraldehyde 3-phosphate dehydrogenase	1E-133
aCL217Contig1	At1g55490	RuBisCO subunit binding-protein beta subunit	0
aCL2176Contig1	At2g05790	glycosyl hydrolase family 17 protein	0
aCL2173Contig1	At2g47970	NPL4 family protein	1E-114
aCL2172Contig1	At3g24590	signal peptidase I family protein	4E-73
aCL2166Contig1	At5g43330	malate dehydrogenase	1E-166
aCL2159Contig1	At1g23400	expressed protein	3E-98
aCL2154Contig1	At3g56340	40S ribosomal protein S26 (RPS26C)	1E-30
aCL2150Contig1	At1g27340	F-box family protein	1E-180
aCL2147Contig1	At1g55915	expressed protein	9E-86
aCL2142Contig2	At4g31130	expressed protein	3E-62
aCL2139Contig1	At5g64510	expressed protein	6E-88
aCL2121Contig1	At3g03420	Ku70-binding family protein	4E-82
aCL211Contig1	At1g65930	isocitrate dehydrogenase	0
aCL2101Contig1	At3g01680	expressed protein	1E-56
aCL20Contig9	At3g57240	similar to glycosyl hydrolase family 17 protein	1E-107
aCL20Contig5	At3g57240	similar to glycosyl hydrolase family 17 protein	1E-114
aCL209Contig1	At4g24280	heat shock protein 70 (HSP70)	0
aCL2078Contig1	At4g32760	similar to VHS domain-containing protein	1E-155
aCL2046Contig1	At5g19780	tubulin alpha-3/alpha-5 chain (TUA5)	0
aCL2040Contig3	At1g19180	expressed protein	2E-26
aCL2027Contig1	At4g38580	copper chaperone (CCH)-related	5E-66
aCL200Contig1	At2g15430	RNA polymerase II subunit 3 (RPB36A)	1E-150
aCL2005Contig1	At5g61030	RNA-binding protein	2E-43
aCL1Contig33	At1g17860	trypsin and protease inhibitor family protein	3E-26
aCL1Contig22	At1g67360	rubber elongation factor (REF) family protein	5E-60
aCL1Contig12	At1g17860	trypsin and protease inhibitor family protein	7E-24
aCL199Contig1	At1g53280	DJ-1 family protein	5E-79
aCL1999Contig2	At4g25630	fibrillarin 2 (FIB2)	1E-127
aCL198Contig1	At2g04240	zinc finger (C3HC4-type RING finger) family protein	4E-44
aCL1988Contig1	At1g74050	60S ribosomal protein L6 (RPL6C)	1E-84
aCL1982Contig1	At4g25130	peptide methionine sulfoxide reductase	7E-99
aCL197Contig1	At4g16190	cysteine proteinase	1E-158
aCL1963Contig1	At5g59850	40S ribosomal protein S15A (RPS15aF)	2E-69
aCL1948Contig1	At1g64350	transducin family protein / WD-40 repeat family protein	1E-131
aCL1940Contig1	At1g12410	ATP-dependent Clp protease proteolytic subunit (ClpP2)	1E-118
aCL190Contig4	At1g17860	trypsin and protease inhibitor family protein	4E-26
aCL190Contig1	At1g17860	trypsin and protease inhibitor family protein	1E-26
aCL1903Contig1	At4g02620	vacuolar ATPase subunit F family protein	4E-51
aCL18Contig2	At5g63910	expressed protein	2E-24
aCL18Contig1	At3g51030	thioredoxin H-type 1 (TRX-H-1)	8E-42
aCL1890Contig1	At4g38600	HECT-domain-containing protein	0
aCL1887Contig1	At2g24360	serine/threonine/tyrosine kinase	1E-176
aCL1869Contig1	At4g24210	F-box family protein / SLEEPY1 protein	5E-40
aCL1861Contig1	At2g40100	chlorophyll A-B binding protein (LHCB4.3)	1E-114
aCL1806Contig1	At1g51060	histone H2A	6E-44
aCL179Contig2	At3g44110	DNAJ heat shock protein (J3)	1E-168
aCL1798Contig1	At4g37890	zinc finger (C3HC4-type RING finger) family protein	1E-158
aCL1794Contig1	At3g49250	expressed protein	1E-91
aCL1787Contig1	At1g24510	T-complex protein 1 epsilon subunit (TCP-1-epsilon)	1E-126
aCL1785Contig1	At2g01660	33 kDa secretory protein-related	3E-79
aCL177Contig1	At1g19000	myb family transcription factor	7E-49
aCL1778Contig1	At1g21750	encodes a protein disulfide isomerase-like (PDIL) protein	0
aCL1769Contig1	At3g15000	expressed protein	2E-72
aCL1765Contig2	At4g25650	Rieske (2Fe-2S) domain-containing protein	1E-101
aCL1764Contig1	At3g18080	glycosyl hydrolase family 1 protein	0

aCL175Contig1	At4g29040	26S proteasome AAA-ATPase subunit (RPT2a)	0
aCL1758Contig1	At5g23240	DNAJ heat shock N-terminal domain-containing protein	1E-127
aCL1755Contig1	At1g78700	brassinosteroid signalling positive regulator-related	6E-71
aCL1753Contig1	At1g65820	microsomal glutathione s-transferase	2E-65
aCL1749Contig1	At5g13550	sulfate transporter family protein	1E-116
aCL1737Contig1	At4g01070	glycosyltransferase (UGT72B1)	1E-162
aCL172Contig2	At5g25610	dehydration-responsive protein (RD22)	4E-93
aCL16Contig2	At3g12490	similar to cysteine protease inhibitor	4E-25
aCL16Contig1	At5g54770	thiazole biosynthetic enzyme (ARA6) (THI1) (THI4)	1E-151
aCL1683Contig2	At4g31940	cytochrome P450	7E-67
aCL1678Contig2	At5g66380	encodes a folate transporter	1E-129
aCL1671Contig1	At5g18400	expressed protein	2E-36
aCL1661Contig2	At4g17510	ubiquitin carboxyl-terminal hydrolase	1E-80
aCL165Contig1	At5g60660	major intrinsic family protein / MIP family protein	1E-130
aCL1648Contig2	At1g03900	expressed protein	1E-87
aCL163Contig1	At4g18100	60S ribosomal protein L32 (RPL32A)	5E-57
aCL1636Contig1	At1g68010	glycerate dehydrogenase	0
aCL1635Contig2	At4g13010	oxidoreductase, zinc-binding dehydrogenase family protein	3E-67
aCL161Contig1	At4g31750	protein phosphatase 2C (PP2C)	1E-150
aCL1615Contig1	At4g01000	ubiquitin family protein	3E-60
aCL1613Contig1	At3g18490	aspartyl protease family protein	1E-162
aCL1606Contig2	At3g43190	sucrose synthase	1E-96
aCL15Contig4	At2g29630	thiamine biosynthesis family protein	1E-176
aCL1588Contig1	At2g45770	signal recognition particle receptor protein (FTSY)	1E-164
aCL157Contig1	At2g35940	similar to BEL1-like homeobox 4 protein (BLH4)	9E-09
aCL1577Contig1	At2g35940	similar to BEL1-like homeobox 4 protein (BLH4)	5E-29
aCL156Contig1	At1g14720	xyloglucan:xyloglucosyl transferase (XTR2)	1E-129
aCL1524Contig1	At5g17310	UTP--glucose-1-phosphate uridylyltransferase	0
aCL1512Contig1	At1g02130	Ras-related protein (ARA-5) / small GTP-binding protein	1E-102
aCL150Contig1	At3g22440	hydroxyproline-rich glycoprotein family protein	1E-180
aCL1507Contig2	At2g29420	glutathione S-transferase	8E-50
aCL14Contig2	At5g51100	superoxide dismutase (Fe)	2E-81
aCL1498Contig1	At3g14420	(S)-2-hydroxy-acid oxidase / glycolate oxidase (GOX)	1E-177
aCL148Contig1	At1g79040	photosystem II 10 kDa polypeptide (PsbR)	2E-47
aCL1483Contig2	At5g47830	expressed protein	9E-40
aCL147Contig1	At2g37110	expressed protein	4E-94
aCL1478Contig1	At1g60000	29 kDa ribonucleoprotein	2E-67
aCL1466Contig2	At1g05850	chitinase-like protein 1 (CTL1)	1E-142
aCL1438Contig1	At3g13160	pentatricopeptide (PPR) repeat-containing protein	1E-79
aCL1432Contig1	At1g62040	autophagy 8c (APG8c)	6E-58
aCL142Contig1	At2g33040	ATP synthase gamma chain (ATPC)	1E-147
aCL1426Contig1	At1g29260	peroxisomal targeting signal type 2 receptor (PEX7)	1E-168
aCL1421Contig1	At4g20360	elongation factor Tu / EF-Tu (TUFA)	0
aCL1405Contig1	At3g54810	zinc finger (GATA type) family protein	2E-67
aCL1402Contig1	At5g08260	serine carboxypeptidase S10 family protein	1E-102
aCL13Contig7	At3g43810	calmodulin-7 (CAM7)	3E-81
aCL1396Contig1	At5g13520	peptidase M1 family protein	1E-159
aCL1386Contig2	At5g66650	expressed protein	1E-37
aCL1370Contig3	At1g32470	glycine cleavage system H protein	4E-71
aCL1370Contig2	At3g24503	aldehyde dehydrogenase (ALDH1a)	2E-21
aCL136Contig1	At3g48750	A-type cyclin-dependent kinase (CDKA)	1E-147
aCL1354Contig1	At1g73760	zinc finger (C3HC4-type RING finger) family protein	7E-39
aCL1352Contig1	At4g35110	expressed protein	2E-64
aCL134Contig1	At2g45550	cytochrome P450 family protein	1E-133
aCL1340Contig1	At4g03280	cytochrome B6-F complex iron-sulfur subunit	2E-96
aCL133Contig3	AtMg00030	hypothetical protein	2E-22
aCL1337Contig3	At4g38620	myb family transcription factor (MYB4)	2E-73
aCL132Contig2	At1g19180	expressed protein	5E-39
aCL1326Contig1	At4g36910	CBS domain-containing protein	4E-75
aCL131Contig5	At5g58240	bis(5'-adenosyl)-triphosphatase	2E-67
aCL1319Contig1	At1g32060	phosphoribulokinase (PRK)	0

aCL1317Contig1	At1g65980	peroxiredoxin type 2	9E-75
aCL1313Contig1	At5g41685	mitochondrial import receptor subunit TOM7	1E-22
aCL130Contig5	At5g40370	glutaredoxin	6E-42
aCL130Contig1	At1g26550	peptidyl-prolyl cis-trans isomerase PPIC-type family protein	2E-57
aCL12Contig9	At3g18280	protease inhibitor/seed storage/lipid transfer protein (LTP)	2E-23
aCL128Contig3	At5g22000	zinc finger (C3HC4-type RING finger) family protein	1E-107
aCL1285Contig3	At1g07350	transformer serine/arginine-rich ribonucleoprotein	2E-36
aCL1283Contig1	At2g29570	proliferating cell nuclear antigen 2 (PCNA2)	1E-135
aCL1280Contig1	At3g21175	zinc finger (GATA type) family protein	5E-57
aCL1274Contig1	At5g45775	60S ribosomal protein L11 (RPL11D)	3E-85
aCL1270Contig1	At2g28910	CAX-interacting protein 4 (CAXIP4)	1E-48
aCL126Contig1	At5g59950	RNA and export factor-binding protein	7E-61
aCL125Contig1	At2g37220	29 kDa ribonucleoprotein	3E-77
aCL1259Contig2	At3g59540	60S ribosomal protein L38 (RPL38B)	2E-32
aCL1240Contig1	At5g67420	LOB domain protein 37 (LBD37)	1E-73
aCL123Contig1	At5g54160	quercetin 3-O-methyltransferase 1 (OMT1)	6E-52
aCL1233Contig1	At2g16500	arginine decarboxylase 1 (SPE1)	1E-137
aCL1232Contig1	At3g15360	thioredoxin M-type 4 (TRX-M4)	1E-48
aCL1231Contig2	At2g24260	basic helix-loop-helix (bHLH) family protein	3E-57
aCL1205Contig1	At1g32230	WWE domain-containing protein (CEO)	2E-52
aCL1202Contig2	At4g15940	fumarylacetoacetate hydrolase family protein	2E-59
aCL1199Contig2	At5g17380	pyruvate decarboxylase family protein	0
aCL1174Contig1	At5g14780	formate dehydrogenase (FDH)	1E-180
aCL1164Contig1	At1g56340	calreticulin 1 (CRT1)	1E-173
aCL115Contig1	At3g03070	NADH-ubiquinone oxidoreductase-related	5E-42
aCL1153Contig2	At4g26840	ubiquitin-like protein (SMT3)	5E-43
aCL1153Contig1	At4g26840	ubiquitin-like protein (SMT3)	6E-43
aCL112Contig4	At3g04880	DNA-damage-repair/toleration protein (DRT102)	1E-123
aCL112Contig3	At3g04880	DNA-damage-repair/toleration protein (DRT102)	1E-115
aCL1122Contig1	At1g24120	DNAJ heat shock protein	1E-114
aCL1112Contig2	At2g47770	benzodiazepine receptor-related (PBR)	3E-41
aCL1107Contig2	At5g13450	ATP synthase delta chain	3E-53
aCL1106Contig1	At3g51370	protein phosphatase 2C (PP2C)	1E-176
aCL1104Contig1	At1g70160	expressed protein	6E-76
aCL10Contig9	At3g12500	basic endochitinase	3E-82
aCL10Contig4	At3g12500	basic endochitinase	5E-87
aCL1097Contig1	At4g38220	aminoacylase	1E-141
aCL1095Contig1	At3g52580	40S ribosomal protein S14 (RPS14C)	3E-61
aCL1091Contig1	At3g52300	ATP synthase D chain-related	8E-73
aCL1070Contig1	At4g02080	GTP-binding protein (SAR1A)	2E-85
aCL1066Contig1	At4g35220	cyclase family protein	4E-99
aCL1062Contig1	At5g09590	heat shock protein 70 (HSP70)	0
aCL1023Contig1	At5g13930	chalcone synthase / naringenin-chalcone synthase	1E-112
aCL1011Contig2	At2g39450	cation efflux family protein	1E-110
aCL1002Contig1	At5g33340	aspartyl protease family protein	2E-84
aC01020H11	At5g25780	eukaryotic translation initiation factor 3 subunit 9 (eIF-3 eta)	7E-97
aC01020H02	At2g32400	glutamate receptor family protein (GLR5)	1E-76
aC01020H01	At4g31050	lipoyltransferase (LIP2p)	1E-51
aC01020G12	At1g50180	disease resistance protein (CC-NBS-LRR class)	4E-09
aC01020G10	At2g29650	inorganic phosphate transporter	2E-37
aC01020G06	At2g47620	DNA-binding family protein	1E-07
aC01020G04	At2g38650	glycosyl transferase family 8 protein	9E-37
aC01020E12	At5g38350	disease resistance protein (NBS-LRR class)	5E-38
aC01020E11	At5g64480	expressed protein	1E-26
aC01020E09	At2g16650	similar to MATE efflux family protein	5E-91
aC01020D05	At1g09240	nicotianamine synthase	7E-74
aC01020D02	At1g11260	glucose transporter (STP1)	1E-20
aC01020D01	At2g29260	tropinone reductase	5E-69
aC01020C07	At5g40950	50S ribosomal protein L27 (RPL27)	3E-32
aC01020C04	AtCg00540	encodes cytochrome f apoprotein	4E-50
aC01020C02	At5g45530	expressed protein	1E-08

aC01020B11	At2g29420	glutathione S-transferase	1E-42
aC01020B06	At3g10630	glycosyl transferase family 4 protein	4E-85
aC01020B05	At1g75690	chaperone protein dnaJ-related	8E-25
aC01020B01	At4g13140	expressed protein	2E-46
aC01020A01	At1g79610	sodium proton exchanger (NHX6)	3E-21
aC01019H05	At3g25410	bile acid:sodium symporter family protein	1E-37
aC01019G10	At3g19900	expressed protein	4E-39
aC01019G09	At1g52140	expressed protein	1E-32
aC01019F06	At4g25340	FKBP-type peptidyl-prolyl cis-trans isomerase-related	6E-27
aC01019F05	At3g48470	expressed protein	1E-34
aC01019F02	At4g28290	expressed protein	4E-08
aC01019E12	At2g30580	zinc finger (C3HC4-type RING finger) family protein	1E-38
aC01019E06	At4g19006	26S proteasome regulatory subunit (RPN9)	2E-97
aC01019E05	At3g27930	expressed protein	8E-63
aC01019E01	At5g42090	expressed protein	6E-31
aC01019D09	At4g36920	floral homeotic protein APETALA2 (AP2)	2E-58
aC01019D07	At1g24020	Bet v I allergen family protein	2E-09
aC01019D04	At3g25950	expressed protein	1E-69
aC01019D02	At5g58005	expressed protein	6E-27
aC01019A05	At4g30810	serine carboxypeptidase S10 family protein	7E-34
aC01018H10	At1g18070	similar to elongation factor 1-alpha (EF-1-alpha)	8E-20
aC01018H07	At1g56050	GTP-binding protein-related	3E-79
aC01018H01	At1g50200	aminoacyl-tRNA synthetase family protein	5E-46
aC01018G11	At4g08850	leucine-rich repeat family protein	6E-16
aC01018G09	At2g21940	shikimate kinase	3E-37
aC01018F10	At2g47710	universal stress protein (USP) family protein	2E-16
aC01018F09	At2g33150	acetyl-CoA C-acyltransferase	7E-65
aC01018E04	At3g51290	proline-rich family protein	2E-26
aC01018E03	At4g16320	expressed protein	5E-29
aC01018E02	At3g27050	expressed protein	4E-53
aC01018D10	At5g37970	S-adenosyl-L-methionine:carboxyl methyltransferase	1E-52
aC01018D08	At4g20060	expressed protein	2E-21
aC01018D06	At3g54420	class IV chitinase (CHIV)	2E-37
aC01018D04	At5g40550	expressed protein	2E-69
aC01018D02	At2g23840	HNH endonuclease domain-containing protein	3E-59
aC01018C12	At3g53950	glyoxal oxidase-related	1E-51
aC01018C10	At5g43130	similar to transcription initiation factor IID (TFIID)	7E-50
aC01018C09	At5g27320	expressed protein	3E-54
aC01018C06	At5g06160	splicing factor-related	1E-50
aC01018C01	At3g57510	endo-polygalacturonase (ADPG1)	2E-28
aC01018B12	At1g01950	armadillo/beta-catenin repeat family protein	1E-16
aC01018A07	At2g29440	glutathione S-transferase	8E-32
aC01018A06	At4g35160	O-methyltransferase family 2 protein	1E-33
aC01018A04	At1g08380	expressed protein	9E-14
aC01018A01	At5g40440	mitogen-activated protein kinase kinase (MAPKK)	8E-66
aC01017H01	At4g23730	aldose 1-epimerase family protein	3E-35
aC01017G07	At1g05690	TAZ zinc finger family protein	4E-09
aC01017F11	At5g04590	sulfite reductase / ferredoxin (SIR)	3E-55
aC01017F06	At1g72770	protein phosphatase 2C (P2C-HA)	4E-41
aC01017E08	At5g25590	expressed protein	4E-22
aC01017E02	At1g05000	tyrosine specific protein phosphatase family protein	6E-79
aC01017D08	At3g23910	expressed protein	4E-69
aC01017D06	At2g37035	expressed protein	5E-13
aC01017C08	At1g30740	FAD-binding domain-containing protein	4E-29
aC01017C05	At3g01070	plastocyanin-like domain-containing protein	8E-25
aC01017B12	At1g60690	aldo/keto reductase family protein	4E-78
aC01017A08	At5g23960	encodes a sesquiterpene synthase	1E-30
aC01017A05	At4g12290	copper amine oxidase	1E-113
aC01016H07	At3g28345	ABC transporter family protein	4E-48
aC01016G11	At3g43740	leucine-rich repeat family protein	7E-30
aC01016G08	At3g61320	expressed protein	8E-17

aC01016G07	At1g15340	methyl-CpG-binding domain-containing protein	1E-35
aC01016G03	At1g78915	expressed protein	3E-40
aC01016G01	At3g43810	calmodulin-7 (CAM7)	1E-51
aC01016F12	At2g38330	MATE efflux family protein	3E-73
aC01016F10	At2g22780	malate dehydrogenase (PMDH1)	5E-89
aC01016E04	At5g35750	histidine kinase (AHK2)	1E-117
aC01016E03	At1g15990	cyclic nucleotide-regulated ion channel (CNGC7)	2E-37
aC01016D10	At3g26000	F-box family protein	2E-77
aC01016D06	At3g50950	disease resistance protein (CC-NBS-LRR class)	2E-77
aC01016C07	At4g29060	elongation factor Ts family protein (EF-Ts)	4E-58
aC01016C03	At3g27820	monodehydroascorbate reductase	1E-18
aC01016B08	At5g24530	oxidoreductase, 2OG-Fe(II) oxygenase family protein	2E-52
aC01016B03	At1g75220	integral membrane protein	7E-86
aC01016A01	At2g45990	expressed protein	2E-81
aC01015H12	At5g27280	zinc finger (DNL type) family protein	2E-63
aC01015H08	At3g24730	similar to yellow-leaf-specific protein 8 (YLS8)	9E-55
aC01015G12	At5g55260	serine/threonine protein phosphatase PP-X isozyme 2	1E-112
aC01015G07	At5g03455	rhodanese-like domain-containing protein	2E-56
aC01015G06	At3g19960	myosin (ATM)	3E-31
aC01015G04	At5g22320	leucine-rich repeat family protein	3E-83
aC01015F12	At3g04790	ribose 5-phosphate isomerase-related	6E-61
aC01015F02	At3g10620	diadenosine 5',5'''-P1,P4-tetraphosphate hydrolase	1E-81
aC01015E12	At1g45130	beta-galactosidase	8E-98
aC01015E01	At2g34790	FAD-binding domain-containing protein	4E-75
aC01015D10	At4g28780	GDSL-motif lipase/hydrolase family protein	1E-56
aC01015D03	At2g26150	heat shock transcription factor family protein	6E-38
aC01015D01	At1g06620	2-oxoglutarate-dependent dioxygenase	2E-68
aC01015C10	At3g21820	SET domain-containing protein	1E-46
aC01015C04	At5g58240	bis(5'-adenosyl)-triphosphatase	1E-23
aC01015B09	At1g26520	similar to PRLI-interacting factor L	8E-57
aC01015A09	At3g09050	expressed protein	9E-66
aC01014F11	At1g11410	S-locus protein kinase	7E-16
aC01014E07	At2g02240	F-box family protein / SKP1 interacting partner 3-related	1E-21
aC01014D09	At3g58610	ketol-acid reductoisomerase	4E-40
aC01014D06	At3g53180	glutamine synthetase	5E-53
aC01014C06	At2g15690	pentatricopeptide (PPR) repeat-containing protein	2E-38
aC01014B12	At4g13030	expressed protein	2E-35
aC01014B10	At1g76860	small nuclear ribonucleoprotein	1E-07
aC01014B02	At3g06330	zinc finger (C3HC4-type RING finger) family protein	8E-33
aC01014A11	At5g20060	similar to phospholipase/carboxylesterase family protein	2E-24
aC01014A07	At2g41900	zinc finger (CCCH-type) family protein	1E-54
aC01014A05	At1g61030	expressed protein	3E-12
aC01014A03	At1g69523	UbiE/COQ5 methyltransferase family protein	2E-56
aC01013H11	At4g10000	expressed protein	3E-16
aC01013H02	At5g49220	expressed protein	6E-40
aC01013F12	At1g32080	membrane protein, contains 12 transmembrane domains	6E-81
aC01013F11	At1g70780	expressed protein	8E-20
aC01013E12	At1g51760	IAA-amino acid hydrolase 3 (IAR3)	2E-57
aC01013E11	At1g10910	pentatricopeptide (PPR) repeat-containing protein	5E-18
aC01013E07	At5g46630	clathrin adaptor complexes medium subunit family protein	9E-69
aC01013E02	At1g54710	expressed protein	7E-14
aC01013D07	At1g42540	glutamate receptor family protein (GLR3.3)	1E-49
aC01013D05	At5g60620	phospholipid/glycerol acyltransferase family protein	3E-38
aC01013C06	At5g27430	signal peptidase subunit family protein	4E-44
aC01013C02	At3g49010	similar to 60S ribosomal protein L13 (RPL13D)	3E-19
aC01013B08	At5g63050	expressed protein	2E-06
aC01013B02	At1g08190	vacuolar assembly protein (VPS41)	6E-43
aC01013A09	At5g07020	proline-rich family protein	5E-33
aC01013A07	At2g40320	expressed protein	1E-57
aC01013A04	At4g33220	pectinesterase family protein	6E-06
aC01012H10	At3g57030	strictosidine synthase family protein	3E-42

aC01012H05	At5g17790	zinc finger (Ran-binding) family protein	5E-15
aC01012H01	At3g02350	glycosyl transferase family 8 protein	4E-19
aC01012G08	At4g31240	expressed protein	1E-30
aC01012G05	At5g27270	pentatricopeptide (PPR) repeat-containing protein	5E-24
aC01012G03	At3g18870	mitochondrial transcription termination factor-related	1E-20
aC01012F08	At1g42470	patched family protein	2E-44
aC01012E09	At1g20925	auxin efflux carrier family protein	5E-43
aC01012E03	At5g35080	expressed protein	1E-54
aC01012D02	At4g28088	hydrophobic protein	4E-23
aC01012C05	At1g49210	zinc finger (C3HC4-type RING finger) family protein	1E-51
aC01012C02	At5g66150	glycosyl hydrolase family 38 protein	3E-18
aC01012B07	At1g04770	male sterility MS5 family protein	3E-59
aC01012B04	At3g10550	expressed protein	8E-32
aC01012B02	At5g16990	NADP-dependent oxidoreductase	2E-19
aC01012A04	At5g36930	disease resistance protein (TIR-NBS-LRR class)	4E-31
aC01011H10	At4g39460	mitochondrial substrate carrier family protein	3E-62
aC01011H03	At3g01410	RNase H domain-containing protein	1E-33
aC01011F03	At5g16010	3-oxo-5-alpha-steroid 4-dehydrogenase family protein	7E-41
aC01011E07	At3g02020	aspartate kinase	8E-55
aC01011E04	At5g58300	leucine-rich repeat transmembrane protein kinase	1E-61
aC01011D09	At1g77740	1-phosphatidylinositol-4-phosphate 5-kinase	2E-64
aC01011D07	At3g14080	small nuclear ribonucleoprotein	3E-39
aC01011D06	At1g69780	homeobox-leucine zipper protein 13 (HB-13)	1E-58
aC01011C12	At1g47890	disease resistance family protein	2E-43
aC01011C01	At5g13070	MSF1-like family protein	8E-90
aC01011B05	At2g32230	pentatricopeptide (PPR) repeat-containing protein	3E-18
aC01011A09	At2g46220	expressed protein	4E-35
aC01011A07	At1g78020	senescence-associated protein-related (SAG102)	1E-28
aC01011A02	At3g57330	calcium-transporting ATPase	4E-07
aC01010H09	At1g21065	expressed protein	1E-33
aC01010H06	At2g30710	RabGAP/TBC domain-containing protein	7E-14
aC01010H05	At5g51970	sorbitol dehydrogenase	2E-34
aC01010H03	At1g09290	expressed protein	3E-21
aC01010G11	At3g12380	similar to actin	6E-36
aC01010F01	At3g03570	expressed protein	2E-65
aC01010D12	At4g16390	chloroplastic RNA-binding protein P67	6E-52
aC01010D10	At5g63860	UVB-resistance protein (UVR8)	2E-10
aC01010D06	At5g06240	expressed protein	1E-29
aC01010D02	At2g32580	expressed protein	5E-33
aC01010C12	At5g10250	phototropic-responsive protein (RPT2)	7E-51
aC01010C05	At2g41520	DNAJ heat shock N-terminal domain-containing protein	5E-52
aC01010B09	At5g09300	2-oxoisovalerate dehydrogenase	3E-44
aC01010A07	At5g06800	myb family transcription factor	1E-28
aC01010A01	At4g26600	nucleolar protein	1E-61
aC01009H02	At1g64720	expressed protein	4E-43
aC01009G11	At5g48790	expressed protein	8E-56
aC01009G08	At4g01650	expressed protein	2E-34
aC01009G02	At1g76080	thioredoxin family protein (TRX)	8E-20
aC01009F11	At5g10140	MADS-box protein flowering locus F (FLF)	4E-22
aC01009F07	At2g05120	expressed protein	5E-27
aC01009F05	At5g40980	expressed protein	2E-29
aC01009E09	At1g32080	membrane protein, contains 12 transmembrane domains	1E-103
aC01009E04	AtCg00120	ATPase alpha subunit	9E-83
aC01009D06	At1g76730	5-formyltetrahydrofolate cyclo-ligase family protein	1E-47
aC01009C05	At3g28050	nodulin MtN21 family protein	8E-10
aC01009C02	At5g04550	expressed protein	3E-43
aC01009B02	At5g40720	expressed protein	3E-33
aC01009A12	At4g15215	ABC transporter family protein	3E-41
aC01009A05	At2g39210	nodulin family protein	1E-18
aC01009A02	At5g59030	copper transporter 1 (COPT1)	7E-28
aC01008H08	At4g28830	expressed protein	5E-16

aC01008H04	At3g45040	phosphatidate cytidyltransferase family protein	2E-06
aC01008G12	At3g24800	PRT1 protein (PRT1)	2E-32
aC01008G09	At4g27700	rhodanese-like domain-containing protein	2E-14
aC01008G04	At2g19440	glycosyl hydrolase family 17 protein	2E-65
aC01008E03	At3g48590	CCAAT-box binding transcription factor Hap5a	2E-56
aC01008D08	At4g14570	acylaminoacyl-peptidase-related	4E-55
aC01008C04	At5g16780	SART-1 family protein	7E-23
aC01008B12	At5g20110	dynein light chain	5E-45
aC01008B07	At4g37580	N-acetyltransferase / hookless1 (HLS1)	8E-36
aC01006H06	At2g32450	calcium-binding EF hand family protein	1E-104
aC01006G12	At3g53940	mitochondrial substrate carrier family protein	7E-25
aC01006F11	At5g03940	signal recognition particle 54 kDa protein (SRP54)	3E-21
aC01006E11	At1g29800	similar to zinc finger (FYVE type) family protein	8E-88
aC01006E05	At1g55860	ubiquitin-protein ligase 1 (UPL1)	3E-69
aC01006D04	At3g48770	expressed protein	1E-52
aC01006C12	At2g41190	amino acid transporter family protein	5E-32
aC01006C11	At1g63910	myb family transcription factor (MYB103)	9E-70
aC01006B09	At3g55050	serine/threonine protein phosphatase 2C (PP2C6)	6E-15
aC01006A07	At5g52030	TraB protein-related	1E-20
aC01006A06	At4g02630	protein kinase family protein	1E-16
aC01005H09	At3g62000	O-methyltransferase family 3 protein	2E-79
aC01005F02	At2g44920	thylakoid lumenal 15 kDa protein	9E-26
aC01005F01	At4g02075	zinc finger (C3HC4-type RING finger) family protein	6E-44
aC01005E06	At2g27960	cyclin-dependent kinase (CKS1)	2E-36
aC01005E02	At1g52190	proton-dependent oligopeptide transport (POT) family protein	4E-78
aC01005D04	At2g27150	aldehyde oxidase 3 (AAO3)	4E-51
aC01005C10	At1g77800	PHD finger family protein	4E-44
aC01005C08	At2g29140	pumilio/Puf RNA-binding domain-containing protein	9E-36
aC01005C07	At4g12420	multi-copper oxidase (SKU5)	2E-06
aC01005C06	At5g67370	expressed protein	1E-70
aC01005B03	At4g39250	myb family transcription factor	1E-06
aC01005A10	At2g26830	choline/ethanolamine kinase family protein	6E-74
aC01005A05	At4g19400	expressed protein	4E-42
aC01004H07	At5g38660	expressed protein	8E-18
aC01004D12	At1g42540	glutamate receptor family protein (GLR3.3)	1E-68
aC01004D07	At5g15450	heat shock protein 100 (HSP100)	8E-73
aC01004C10	At1g29000	heavy-metal-associated domain-containing protein	2E-12
aC01004B07	At1g70320	ubiquitin-protein ligase 2 (UPL2)	2E-20
aC01003H08	At3g02330	pentatricopeptide (PPR) repeat-containing protein	1E-31
aC01003G12	At3g61200	thioesterase family protein	3E-08
aC01003G01	At5g20280	sucrose-phosphate synthase	6E-53
aC01003D11	At5g52780	expressed protein	3E-18
aC01003D04	At2g46820	expressed protein	5E-16
aC01003C11	At1g08920	sugar transporter / similar to ERD6 protein	3E-24
aC01003C04	At3g11340	UDP-glucuronosyl/UDP-glucosyl transferase family protein	2E-51
aC01003B10	At4g02510	chloroplast outer membrane protein	1E-53
aC01003B09	At5g20660	24 kDa vacuolar protein (VP24)	2E-54
aC01003B08	At3g11210	GDSL-motif lipase/hydrolase family protein	1E-24
aC01003B03	At5g42950	GYF domain-containing protein	2E-28
aC01003A09	At5g55300	DNA topoisomerase I	2E-52
aC01002H07	At1g49180	protein kinase family protein	1E-16
aC01002E01	At3g18030	halotolerance protein (HAL3A)	2E-42
aC01002D07	At4g11860	expressed protein	9E-75
aC01002A08	At5g58930	expressed protein	7E-22
aC01001H01	At5g02810	pseudo-response regulator 7 (APRR7)	2E-12
aC01001F11	At2g03780	translin family protein	5E-57
aC01001F06	At5g17910	expressed protein	1E-14
aC01001E11	At4g35980	expressed protein	5E-28
aC01001C09	At5g51350	leucine-rich repeat transmembrane protein kinase	1E-47
aC01001B11	At5g59610	DNAJ heat shock N-terminal domain-containing protein	3E-47

I. Unigenes de cítricos implicados en la biosíntesis de diversos terpenoides

I.1. Terpenoides derivados de GGPP

I.1.1. Biosíntesis de tocoferol

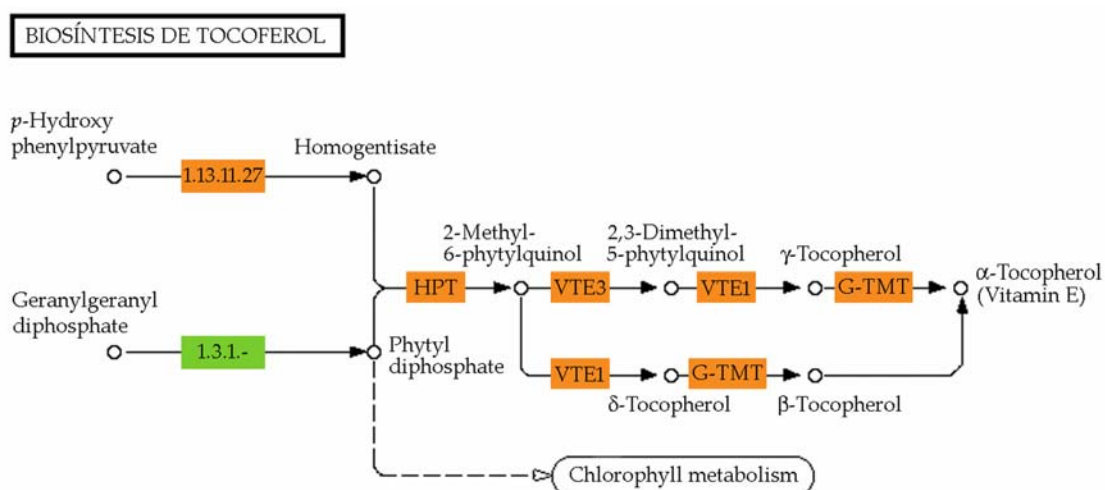


Fig. A1. Esquema de biosíntesis de tocoferoles en cítricos. Las cajas verdes indican los pasos para los que se identificaron ESTs procedentes de la genoteca Veg1 y las naranjas aquellos pasos para los que se encontraron los candidatos correspondientes en las demás genotecas del CFGP.

En la base de datos del CFGP fue posible identificar ESTs de cítricos correspondientes a cada una de las enzimas implicadas en la síntesis de tocoferol (Tabla A1).

Tabla A1. Unigenes de cítricos que codifican enzimas de biosíntesis de tocoferol

Enzimas	EC	Gen de ATH más similar	Contigs	Singletons	Total ESTs
Geranylgeranyl reductasa (GGR)	1.3.1.-	At1g74470	2	0	35
4-Hidroxifenilpiruvato dioxigenasa	1.13.11.27	At1g06570	1	1	21
Homogentisato fitilpreniltransferasa (HPT)		At2g18950	8	6	36
MPBQ/MSBQ-metiltransferasa (VTE3)	2.1.1.-	At3g63410	2	0	11
Tocopherol ciclasa (VTE1)		At4g32770	2	1	24
Gamma-tocopherol metiltransferasa (G-TMT)	2.1.1.95	At1g64970	1	1	24

1.1.2. Biosíntesis de carotenoides

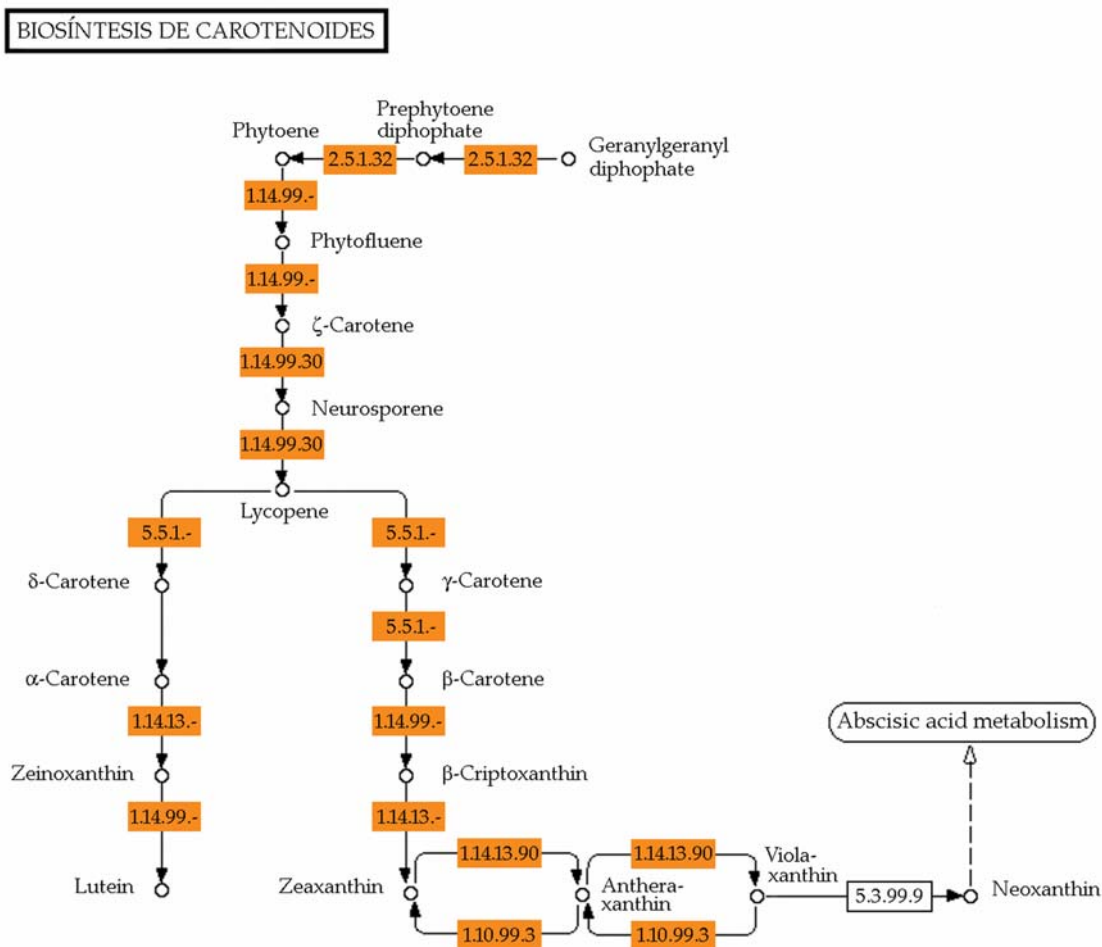


Fig. A2. Esquema de biosíntesis de carotenoides en cítricos. Las cajas naranjas representan los pasos para los que se identificaron los genes candidatos a partir de la colección de ESTs del CFGP.

En la base de datos del CFGP fue posible identificar ESTs de cítricos correspondientes a cada una de las enzimas implicadas en la biosíntesis de carotenoides (Tabla A2).

Tabla A2. Unigenes de cítricos que codifican enzimas de biosíntesis de carotenoides

Enzimas	EC	Gen de ATH más similar	Contigs	Singletons	Total ESTs
Fitoeno sintasa (PSY)	2.5.1.32	At5g17230	2	2	7
Fitoeno dehidrogenasa (PSD)	1.14.99.-	At4g14210	1	2	4
zeta-Caroteno desaturasa (ZDS)	1.14.99.30	At3g04870	3	1	14
Licopeno epsilon ciclaza (ϵ -LCY)	5.5.1.-	At5g57030	1	2	4
Carotenoide β -ring hidroxilasa (β -CHX)	1.14.13.-	At4g25700	2	0	38
Carotenoide ϵ -ring hidroxilasa (ϵ -CHX)	1.14.99.-	At3g53130	1	0	2
Licopeno beta ciclaza (β -LYC)	5.5.1.-	At3g10230	3	0	33
Zeaxantina epoxidasa (ZEP, ABA1)	1.14.13.90	At5g67030	2	3	21
Violaxantina de-epoxidasa (VDE)	1.10.99.3	At1g08550	0	2	2

1.1.3. Metabolismo del ácido abscísico

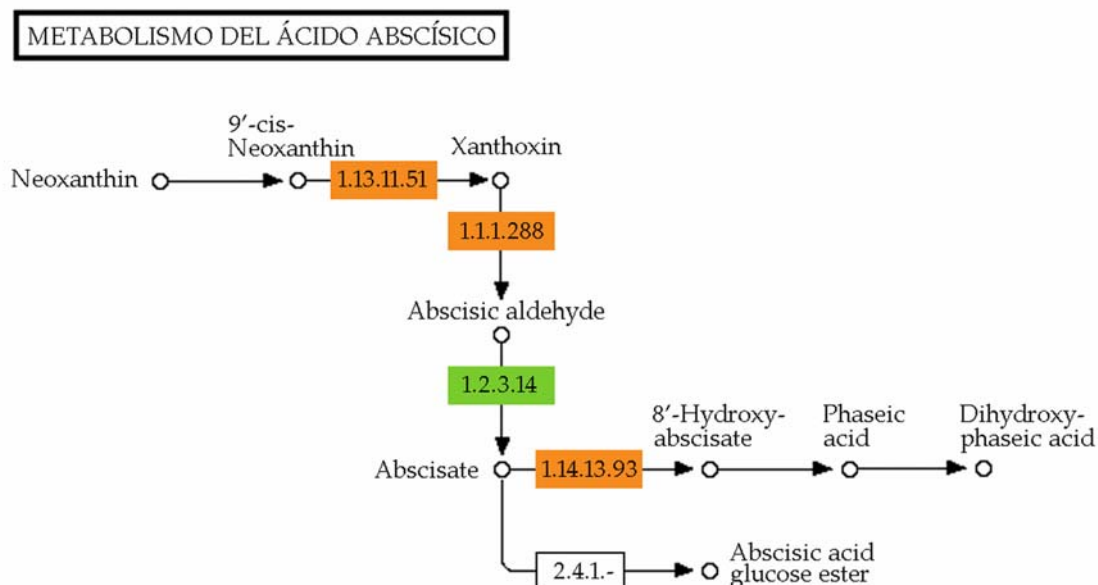


Fig. A3. Esquema del metabolismo del ácido abscísico en cítricos. Las cajas verdes indican los pasos para los que se identificaron ESTs procedentes de la genoteca Veg1 y las naranjas aquellos pasos para los que se encontraron los candidatos correspondientes en las demás genotecas del CFGP.

En la base de datos del CFGP fue posible identificar ESTs de cítricos correspondientes a cada una de las enzimas implicadas en el metabolismo del ácido abscísico (Tabla A3).

Tabla A3. Unigenes de cítricos que codifican enzimas del metabolismo del ácido abscísico

Enzimas	EC	Gen de ATH más similar	Contigs	Singletons	Total ESTs
Biosíntesis					
9-cis-epoxycarotenoide dioxigenasa	1.13.11.51	At1g30100	2	1	12
		At3g14440	1	2	11
Xantoxina dehidrogenasa (ABA2)	1.1.1.288	At1g52340	1	2	6
Abscísico aldehído oxidasa (AAO3)	1.2.3.14	At2g27150	-	2	2
Catabolismo					
Ácido abscísico 8'-hidroxilasa	1.14.13.93	At3g19270	2	2	15
		At2g29090	2	0	6
		At4g19230	3	0	11
Ácido abscísico glucosiltransferasa	2.4.1.-	At4g34131	1	0	2
		At1g05530	3	0	15
		At2g23250	1	0	24
		At2g23260	1	0	8

I.2. Terpenoides derivados de GGPP y FPP

I.2.1. Biosíntesis de plastoquinona

BIOSÍNTESIS DE UBIQUINONA

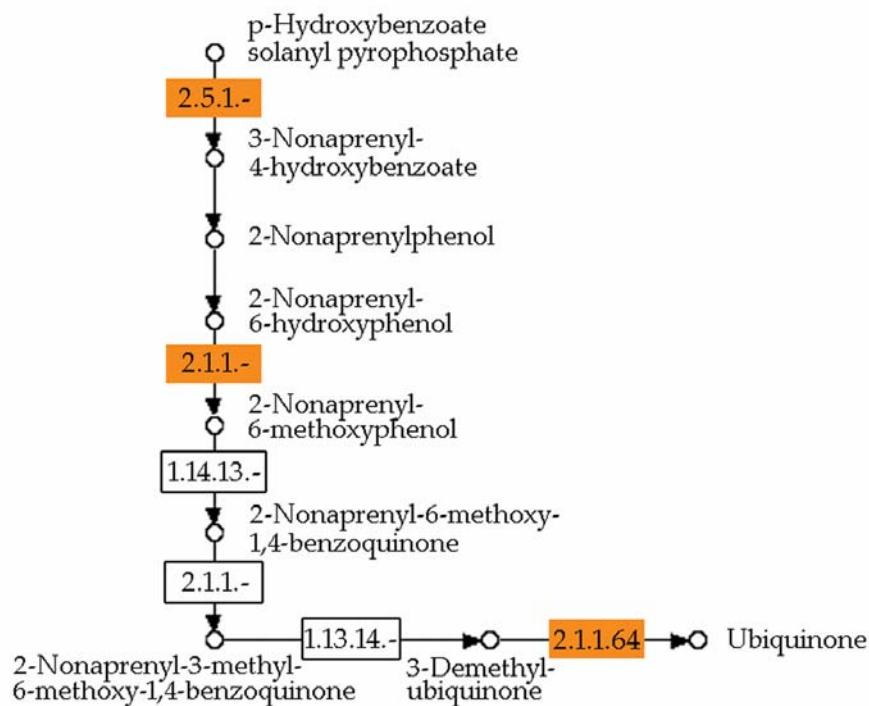


Fig. A4. Esquema de biosíntesis de ubiquinona en cítricos. Las cajas naranjas representan los pasos para los que se identificaron los genes candidatos a partir de la colección de ESTs del CFGP.

En la base de datos del CFGP fue posible identificar ESTs de cítricos correspondientes a cada una de las enzimas implicadas en la biosíntesis de ubiquinona (Tabla A4).

Tabla A4. Unigenes de cítricos que codifican enzimas de biosíntesis de ubiquinona

Enzimas	EC	Gen de ATH más similar	Contigs	Singletons	Total ESTs
Hidroxibenzoato poliprenilo difosfato transferasa (PPT)	2.5.1.-	At4g23660	1	1	15
Nonaprenildihidroxibenzoato metiltransferasa	2.1.1.-	At2g30920	1	0	2
Nonaprenildihidroxibenzoato metiltransferasa	2.1.1.64	At2g30920	1	0	2

1.2.2. Biosíntesis de plastoquinona

BIOSÍNTESIS DE PLASTOQUINONA

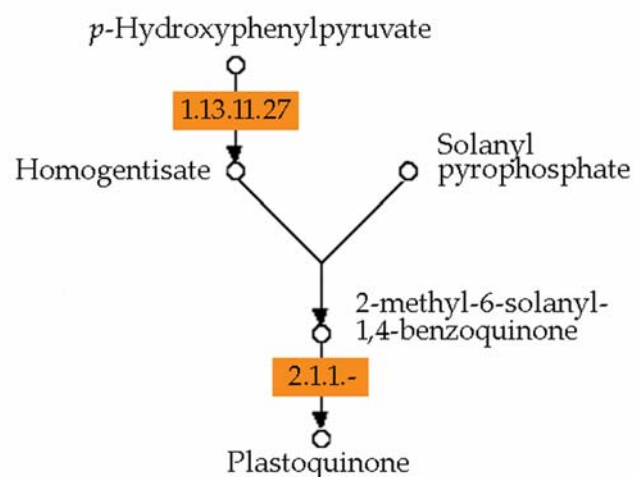


Fig. A5. Esquema de biosíntesis de plastoquinona en cítricos. Las cajas naranjas representan los pasos para los que se identificaron los genes candidatos a partir de la colección de ESTs del CFGP.

En la base de datos del CFGP fue posible identificar ESTs de cítricos correspondientes a cada una de las enzimas implicadas en la biosíntesis de plastoquinona (Tabla A5).

Tabla A5. Unigenes de cítricos que codifican enzimas de biosíntesis de plastoquinona

Enzimas	EC	Gen de ATH más similar	Contigs	Singletons	Total ESTs
4-Hidroxifenilpiruvato dioxigenasa	1.13.11.27	At1g06570	1	1	21
MPBQ/MSBQ-metiltransferasa (VTE3)	2.1.1.-	At3g63410	2	0	11

I.3. Terpenoides derivados de FPP

I.3.1. Biosíntesis de esteroides

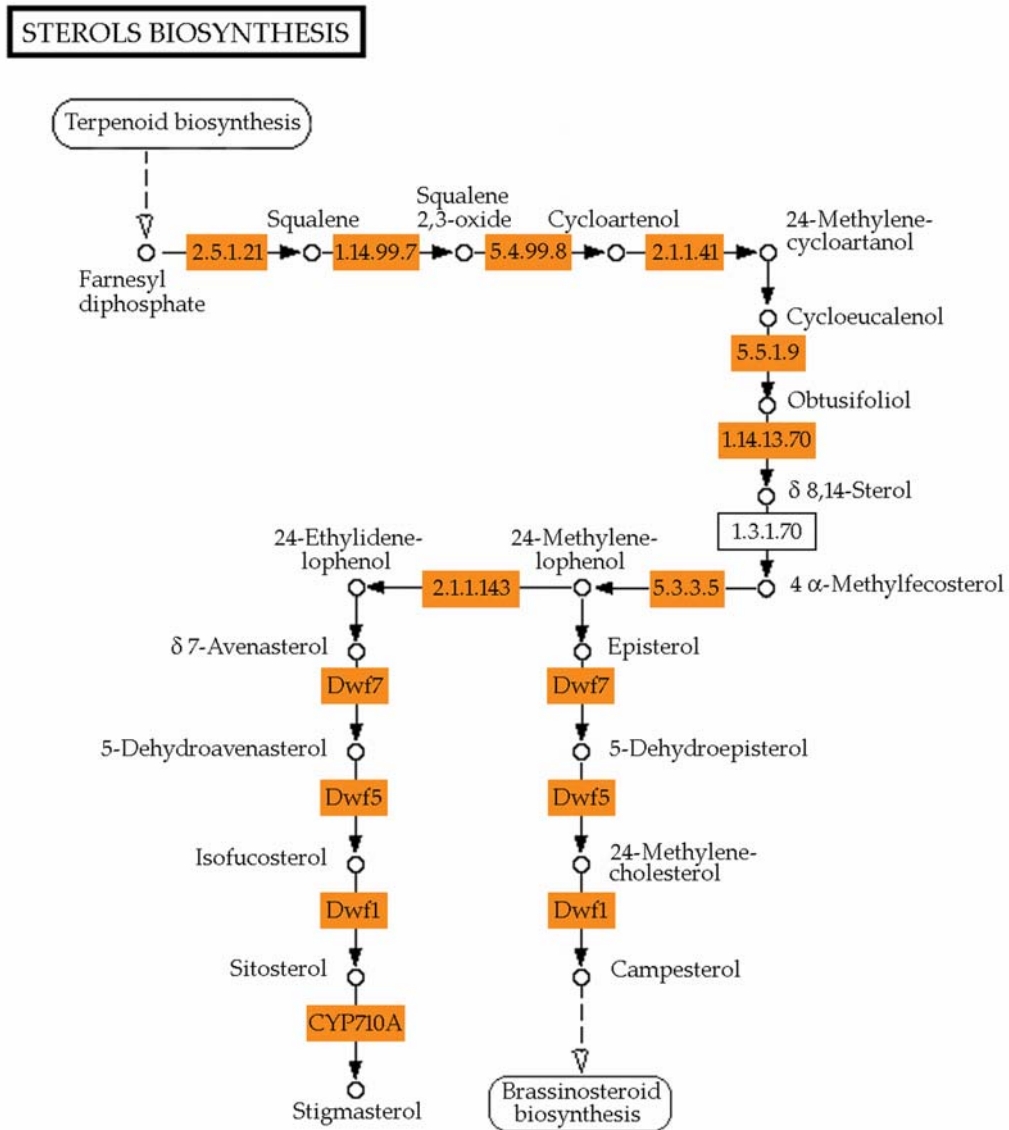


Fig. A6. Esquema de biosíntesis de esteroides en cítricos. Las cajas naranjas representan los pasos para los que se identificaron los genes candidatos a partir de la colección de ESTs del CFGP.

En la base de datos del CFGP fue posible identificar ESTs de cítricos correspondientes a cada una de las enzimas implicadas en la biosíntesis de esteroides (Tabla A6).

Tabla A6. *Unigenes de cítricos que codifican enzimas de biosíntesis de esteroides*

Enzimas	EC	Gen de ATH más similar	Contigs	Singletons	Total ESTs
Escualeno sintasa	2.5.1.21	AT4G34640	0	1	1
Escualeno monooxigenasa	1.14.99.7	AT1G58440	1	0	7
		AT4G37760	1	2	5
Cicloartenol sintasa	5.4.99.8	AT2G07050	0	3	3
Esterol 24-C-metiltransferasa	2.1.1.41	AT5G13710	2	0	5
Cicloeucaalenol cicloisomerasa	5.5.1.9	AT5G50375	1	0	4
Esterol 14-demetilasa	1.14.13.70	AT1G11680	1	2	5
Delta 14-esterol reductasa	1.3.1.70	AT3G52940	-	-	-
Colestenol delta-isomerasa	5.3.3.5	AT1G20050	1	1	3
24-metilenesterol C-metiltransferasa	2.1.1.143	AT1G20330	2	0	13
C-5 esteroles desaturasa (DWF7)	1.14.21.6	AT3G02580	0	1	1
Esterol delta 7-reductasa (DWF5)	1.3.1.21	AT1G50430	1	0	3
Delta 24-esterol reductasa (DWF1)	1.3.1.72	AT3G19820	-	-	-
C-22 esteroles desaturasa (CYP710A)	1.3.1.-	AT2G34500	1	0	7

Anejo III. Unigenes de cítricos representados en la micromatriz CIT1 que codifican factores de transcripción y/o reguladores de la transcripción

Family	Citrus unigene	AGI code	Description	E-value
ABI3VP1	aCL7219Contig1	At3g18990	transcriptional factor B3 family protein	1E-81
ABI3VP1	aCL7325Contig1	At3g18990	transcriptional factor B3 family protein	6E-40
ABI3VP1	aCL4518Contig1	At4g32010	transcriptional factor B3 family protein	2E-62
Alfin-like	aCL3306Contig1	At1g14510	PHD finger family protein	1E-101
AP2-EREBP	aCL8Contig7	At1g50640	ethylene response factor (ERF) subfamily B-1	1E-22
AP2-EREBP	aCL89Contig1	At3g14230	ethylene response factor (ERF) subfamily B-2	1E-68
AP2-EREBP	aCL152Contig1	At3g14230	ethylene response factor (ERF) subfamily B-2	5E-32
AP2-EREBP	aCL524Contig1	At3g15210	ethylene response factor (ERF) subfamily B-1	6E-21
AP2-EREBP	aC02007E03SK_c	At5g11190	ethylene response factor (ERF) subfamily B-6	2E-47
AP2-EREBP	aCL2849Contig1	At5g13330	ethylene response factor (ERF) subfamily B-4	3E-35
AP2-EREBP	aCL396Contig1	At5g13330	ethylene response factor (ERF) subfamily B-4	3E-28
AP2-EREBP	aCL1567Contig1	At5g44210	ethylene response factor (ERF) subfamily B-1	3E-23
ARF	aC03006F12SK_c	At1g30330	auxin-responsive factor (ARF6)	8E-71
ARF	aCL2271Contig1	At1g59750	auxin-responsive factor AUX/IAA-related	0
ARF	aCL3486Contig1	At5g62000	auxin-responsive factor	1E-129
bHLH	aCL8856Contig1	At1g09250	transcription factor	8E-33
bHLH	aC02020G06SK_c	At1g09530	phytochrome interacting factor 3 (PIF3)	3E-17
bHLH	aCL8145Contig1	At1g22380	basic helix-loop-helix transcription factor	5E-56
bHLH	aC03004B03SK_c	At1g22380	basic helix-loop-helix transcription factor	2E-72
bHLH	aCL5746Contig1	At1g29950	transcription factor	1E-22
bHLH	aCL3615Contig1	At1g59640	basic helix-loop-helix (bHLH) family protein	2E-72
bHLH	aC20002B11SK_c	At1g69010	basic helix-loop-helix (bHLH) family protein	3E-61
bHLH	aC20002E09SK_c	At2g42280	basic helix-loop-helix (bHLH) family protein	1E-16
bHLH	aCL3157Contig1	At2g46970	basic helix-loop-helix (bHLH) protein	7E-31
bHLH	aCL9380Contig1	At3g07340	basic helix-loop-helix (bHLH) family protein	1E-36
bHLH	aCL2290Contig2	At3g17100	transcription factor	1E-36
bHLH	aCL2564Contig1	At4g01460	basic helix-loop-helix (bHLH) family protein	9E-45
bHLH	aCL9448Contig1	At4g14410	basic helix-loop-helix (bHLH) family protein	3E-66
bHLH	aCL7683Contig1	At4g20970	basic helix-loop-helix (bHLH) family protein	2E-23
bHLH	aCL18016G11Rv_c	At4g25410	basic helix-loop-helix (bHLH) family protein	8E-22
bHLH	aC04015F06SK_c	At4g29100	ethylene-responsive family protein	1E-52
bHLH	aC08014B07SK_c	At4g29100	ethylene-responsive family protein	2E-39
bHLH	aCL488Contig3	At5g09460	transcription factor	6E-16
bZIP	aCL6889Contig1	At1g08320	bZIP family transcription factor	2E-23
bZIP	aCL3553Contig1	At1g45249	ABA-responsive element-binding protein	2E-74
bZIP	aCL474Contig1	At3g19290	ABA-responsive element-binding protein 2	3E-76
bZIP	aCL7631Contig1	At3g56850	ABA-responsive element-binding protein 3	8E-46
bZIP	aCL2910Contig1	At3g58120	bZIP transcription factor family protein	4E-73
bZIP	aC08005F12SK_c	At3g62420	bZIP transcription factor family protein	1E-12
bZIP	aCL3546Contig1	At3g62420	bZIP transcription factor family protein	4E-48
bZIP	aCL4220Contig1	At4g34590	bZIP transcription factor family protein	3E-27
bZIP	aC04015B01SK_c	At4g38900	bZIP family transcription factor	4E-31
bZIP	aCL2063Contig1	At5g44080	bZIP transcription factor family protein	4E-26
C2C2-CO-like	aC03003C09SK_c	At1g28050	zinc finger (B-box type) family protein	2E-34
C2C2-CO-like	aC05014D03SK_c	At4g27310	zinc finger (B-box type) family protein	8E-29
C2C2-CO-like	aCL1873Contig2	At4g27310	zinc finger (B-box type) family protein	4E-31
C2C2-CO-like	aCL475Contig1	At5g24930	zinc finger (B-box type) family protein	1E-115
C2C2-Dof	aC08003G05SK_c	At4g24060	Dof-type zinc finger domain-containing protein	6E-32
C2C2-Dof	aCL6670Contig1	At5g62940	Dof-type zinc finger domain-containing protein	6E-22
C2C2-Gata	aCL4896Contig1	At1g08010	zinc finger (GATA type) family protein	3E-43
C2C2-YABBY	aC02019G10SK_c	At1g69180	transcription factor CRC (CRABS CLAW)	1E-42
C2C2-YABBY	aCL4611Contig1	At2g26580	plant-specific transcription factor YABBY	8E-67
C2C2-YABBY	aCL4648Contig1	At2g26580	plant-specific transcription factor YABBY	6E-18
C2H2	aCL684Contig2	At1g04850	ubiquitin-associated (UBA)/TS-N protein	1E-73
C2H2	aCL684Contig1	At1g04850	ubiquitin-associated (UBA)/TS-N protein	1E-139

C2H2	aCL146Contig1	At1g27730	salt tolerance zinc finger protein (STZ)	4E-48
C2H2	aCL382Contig1	At1g27730	salt tolerance zinc finger protein (STZ)	1E-38
C2H2	aCL335Contig1	At1g30970	supressor of frigida 4 (SUF4)	1E-104
C2H2	aC07001D04SK_c	At1g43860	transcription factor	3E-25
C2H2	aCL5005Contig1	At1g50670	OTU-like cysteine protease family protein	1E-100
C2H2	aCL5478Contig1	At2g27100	zinc-finger (C2H2 type) protein SERRATE (SE)	1E-30
C2H2	aCL5095Contig1	At2g27760	tRNA isopentenyltransferase	1E-111
C2H2	aC08010C06SK_c	At2g28710	zinc finger (C2H2 type) family protein	3E-22
C2H2	aCL3682Contig1	At3g02790	zinc finger (C2H2 type) family protein	4E-43
C2H2	aCL5429Contig1	At3g02860	expressed protein	4E-32
C2H2	aC02021A12SK_c	At3g07940	zinc finger and C2 domain protein	1E-28
C2H2	aCL7490Contig1	At3g12270	methyltransferase family protein	2E-38
C2H2	aCL1165Contig1	At3g52800	zinc finger (AN1-like) family protein	2E-49
C2H2	aCL8194Contig1	At4g16845	vernalization 2 protein (VRN2)	5E-42
C2H2	aCL2974Contig1	At4g31420	zinc finger (C2H2 type) family protein	1E-116
C2H2	aCL5281Contig1	At5g03740	zinc finger (C2H2 type) family protein	3E-24
C2H2	aCL650Contig1	At5g64610	histone acetyltransferase	0
C2H2	aCL2178Contig1	At5g66730	zinc finger (C2H2 type) family protein	8E-24
C3H	aCL2110Contig1	At1g11020	zinc finger (C3HC4-type RING) family protein	1E-119
C3H	aCL3598Contig1	At1g15100	zinc finger (C3HC4-type RING) family protein	2E-31
C3H	aCL7031Contig1	At1g50440	zinc finger (C3HC4-type RING) family protein	3E-26
C3H	aCL466Contig1	At1g68070	zinc finger (C3HC4-type RING) family protein	1E-89
C3H	aCL1354Contig1	At1g73760	zinc finger (C3HC4-type RING) family protein	7E-39
C3H	aCL4880Contig1	At1g73950	zinc finger (C3HC4-type RING) family protein	4E-68
C3H	aCL198Contig1	At2g04240	zinc finger (C3HC4-type RING) family protein	4E-44
C3H	aCL2811Contig1	At2g31510	similar to zinc finger protein	1E-176
C3H	aCL7612Contig1	At2g37150	zinc finger (C3HC4-type RING) family protein	2E-09
C3H	aC20009E07SK_c	At2g46495	zinc finger (C3HC4-type RING) family protein	4E-14
C3H	aC18006A08Rv_c	At2g46495	zinc finger (C3HC4-type RING) family protein	4E-13
C3H	aCL7967Contig1	At3g02340	zinc finger (C3HC4-type RING) family protein	3E-31
C3H	aCL548Contig1	At3g09770	zinc finger (C3HC4-type RING) family protein	1E-105
C3H	aCL812Contig1	At3g16090	zinc finger (C3HC4-type RING) family protein	1E-177
C3H	aCL3130Contig1	At3g55530	zinc finger (C3HC4-type RING) family protein	2E-53
C3H	aCL3130Contig2	At3g55530	zinc finger (C3HC4-type RING) family protein	4E-29
C3H	aCL7954Contig1	At3g58030	zinc finger (C3HC4-type RING) family protein	2E-63
C3H	aCL3447Contig1	At4g33565	zinc finger (C3HC4-type RING) family protein	1E-41
C3H	aC02003G02SK_c	At5g08750	zinc finger (C3HC4-type RING) family protein	6E-24
C3H	aCL8387Contig1	At5g08750	zinc finger (C3HC4-type RING) family protein	2E-43
C3H	aCL4888Contig1	At5g20910	zinc finger (C3HC4-type RING) family protein	1E-102
C3H	aCL5150Contig1	At5g38895	zinc finger (C3HC4-type RING) family protein	1E-33
C3H	aCL2517Contig2	At5g41350	zinc finger (C3HC4-type RING) family protein	2E-49
CAMTA	aCL5719Contig1	At5g64220	calmodulin-binding transcription activator	2E-95
CCAAT-DR1	aCL2659Contig1	At5g23090	TATA-binding protein	5E-55
CCAAT-HAP3	aCL5582Contig1	At3g53340	CCAAT-box binding transcription factor	2E-51
CCAAT-HAP3	aCL1964Contig1	At4g14540	CCAAT-box binding transcription factor (HAP3)	2E-50
CCAAT-HAP5	aCL665Contig2	At1g08970	CCAAT-box binding transcription factor Hap5a	7E-68
CCAAT-HAP5	aCL665Contig1	At1g08970	CCAAT-box binding transcription factor Hap5a	1E-60
CCAAT-HAP5	aC01008E03SK_c	At3g48590	CCAAT-box binding transcription factor Hap5a	2E-56
EIL	aCL5023Contig1	At2g27050	ethylene-insensitive3-like1 (EIL1)	7E-22
EIL	aCL1330Contig1	At3g20770	ethylene-insensitive 3 (EIN3)	0
G2-like	aCL319Contig1	At1g25550	myb family transcription factor	3E-77
G2-like	aCL4426Contig1	At2g01060	myb family transcription factor	1E-57
G2-like	aC02013H09SK_c	At3g13040	myb family transcription factor	5E-10
G2-like	aCL4279Contig1	At4g13640	myb family transcription factor	2E-51
G2-like	aC01010A07SK_c	At5g06800	myb family transcription factor	1E-28
G2-like	aCL7275Contig1	At5g29000	myb family transcription factor	7E-49
GeBP	aCL2740Contig1	At5g14280	DNA-binding storekeeper protein-related	2E-42
GRAS	aCL7712Contig1	At1g21450	scarecrow-like transcription factor 1 (SCL1)	1E-106

GRAS	aCL3705Contig1	At2g01570	gibberellin response modulator (RGA1)	0
GRAS	aCL3779Contig1	At2g29060	scarecrow transcription factor family protein	1E-113
GRAS	aCL1853Contig1	At5g52510	scarecrow-like transcription factor 8 (SCL8)	1E-138
GRF	aCL5415Contig1	At5g53660	growth regulating factor (GRF7)	4E-07
HD-BEL	aCL4929Contig1	At2g23760	similar to BEL1-like homeobox 2 protein (BLH2)	1E-127
HD-BEL	aCL157Contig1	At2g35940	similar to BEL1-like homeobox 4 protein (BLH4)	9E-09
HD-BEL	aCL1577Contig1	At2g35940	similar to BEL1-like homeobox 4 protein (BLH4)	5E-29
HD-Zip I	aC01011D06SK_c	At1g69780	homeobox-leucine zipper protein 13 (HB-13)	1E-58
HD-Zip I	aCL6484Contig1	At5g65310	homeobox-leucine zipper protein 5 (HB-5)	1E-40
HD-Zip II	aCL997Contig1	At4g16780	homeobox-leucine zipper protein 4 (HAT4)	4E-84
HD-Zip II	aCL228Contig1	At4g37790	homeobox-leucine zipper protein 22 (HAT22)	3E-74
HD-Zip III	aCL6460Contig1	At1g52150	homeobox-leucine zipper family protein	5E-88
HD-Zip IV	aC02015G09SK_c	At3g61150	homeobox-leucine zipper family protein	7E-17
HD-HOX7 like	aC02006F10SK_c	At1g28420	homeobox transcription factor	9E-32
HD-KNOX I	aC02004E05SK_c	At1g23380	homeobox transcription factor (KNAT6)	1E-64
HD-KNOX II	aCL1472Contig1	At5g25220	similar to knotted-1 like 4 (KNAT4)	1E-151
HRT	aCL8507Contig1	At5g56780	similar to transcription factor	2E-50
HSF	aC04011G02SK_c	At4g36990	heat shock factor protein 4 (HSF4)	3E-27
HSF	aCL432Contig1	At4g36990	heat shock factor protein 4 (HSF4)	1E-82
JUMONJI	aCL7189Contig1	At4g20400	transcription factor jumonji (jmi) family protein	4E-26
LOB Class II	aCL1240Contig1	At5g67420	lateral organ boundaries protein 37 (LBD37)	1E-73
MADS-Type I	aCL8174Contig1	At5g60440	MADS-box protein (AGL62)	2E-07
MADS-Type II	aCL176Contig1	At1g24260	MADS-box protein (AGL9)	9E-93
MADS-Type II	aCL4207Contig1	At3g58780	agamous-like shatterproof 1 (SHP1)	7E-85
MADS-Type II	aCL2625Contig1	At4g18960	floral homeotic protein AGAMOUS (AG)	5E-96
MADS-Type II	aCL3917Contig1	At4g22950	MADS-box protein (AGL19)	5E-59
MADS-Type II	aC01009F11SK_c	At5g10140	MADS-box protein flowering locus F (FLF)	4E-22
MADS-Type II	aC20007B02SK_c	At5g20240	floral homeotic protein PISTILLATA (PI)	1E-20
MADS-Type II	aCL8484Contig1	At5g20240	floral homeotic protein PISTILLATA (PI)	8E-65
MADS-Type II	aCL3805Contig1	At5g62165	similar to MADS-box protein (AGL20)	1E-54
MADS-Type II	aCL5010Contig1	At5g65050	MADS-box protein (MAF2)	3E-24
MIF subfamily	aCL9040Contig1	At3g28917	zinc finger homeobox family protein	2E-21
MYB-atypical	aC04018B08SK_c	At4g11400	ARID/BRIGHT DNA-binding protein	5E-15
MYB-CCA1-like	aCL2656Contig1	At1g01060	similar to myb transcription factor (CCA1)	1E-103
MYB-CCA1-like	aCL3899Contig1	At3g09600	myb family transcription factor	1E-72
MYB-CCA1-like	aCL973Contig1	At5g47390	myb family transcription factor	1E-114
MYB-like	aC01005B03SK_c	At4g39250	myb family transcription factor	1E-06
MYB-R-R-type	aC20001D01SK_c	At2g38090	myb family transcription factor	4E-54
MYB-TBP-like	aCL3396Contig1	At1g72650	myb family transcription factor	4E-55
MYB-TBP-like	aC20010D05SK_c	At5g03780	expressed protein	2E-15
MYB-TBP-like	aC05004G07SK_c	At5g03780	expressed protein	5E-13
NAC	aCL943Contig1	At1g01720	no apical meristem (NAM) family protein	1E-112
NAC	aCL267Contig2	At1g01720	no apical meristem (NAM) family protein	1E-100
NAC	aCL7982Contig1	At1g33060	no apical meristem (NAM) family protein	2E-06
NAC	aCL1571Contig1	At1g69490	no apical meristem (NAM) family protein	2E-89
NAC	aCL157Contig2	At3g04070	no apical meristem (NAM) family protein	6E-88
NAC	aCL2766Contig1	At3g10480	no apical meristem (NAM) family protein	3E-91
NAC	aCL728Contig2	At3g15510	no apical meristem (NAM) family protein (NAC2)	3E-95
NAC	aCL35Contig3	At4g27410	no apical meristem (NAM) family protein (RD26)	1E-105
NAC	aCL8380Contig1	At4g35580	no apical meristem (NAM) family protein	1E-72
NAC	aCL4040Contig1	At5g04410	no apical meristem (NAM) family protein	1E-128
NAC	aCL3976Contig1	At5g13180	no apical meristem (NAM) family protein	4E-81
R1R2R3-MYB	aCL6786Contig1	At4g32730	myb family transcription factor	1E-21
R1R2R3-MYB	aC07001B04SK_c	At4g32730	myb family transcription factor	2E-33
R2R3-MYB	aCL5017Contig1	At1g17950	myb family transcription factor (MYB52)	2E-71
R2R3-MYB	aC01006C11SK_c	At1g63910	myb family transcription factor (MYB103)	9E-70
R2R3-MYB	aCL413Contig1	At1g68320	myb family transcription factor (MYB62)	6E-62

R2R3-MYB	aC02008E10SK_c	At2g37630	myb family transcription factor (MYB91)	5E-18
R2R3-MYB	aC20006B02SK_c	At2g37630	myb family transcription factor (MYB91)	8E-46
R2R3-MYB	aC03004B08SK_c	At3g23250	myb family transcription factor (MYB15)	5E-40
R2R3-MYB	aCL6070Contig1	At3g27810	myb family transcription factor (MYB3) (MYB21)	7E-64
R2R3-MYB	aCL2843Contig1	At3g30210	myb family transcription factor (MYB121)	1E-51
R2R3-MYB	aC18004B12Rv_c	At4g21440	myb family transcription factor (MYB102)	3E-12
R2R3-MYB	aCL8980Contig1	At4g21440	myb family transcription factor (MYB102)	2E-18
R2R3-MYB	aCL1337Contig3	At4g38620	myb family transcription factor (MYB4)	2E-73
R2R3-MYB	aCL7866Contig1	At5g49620	myb family transcription factor (MYB78)	3E-48
R2R3-MYB	aCL693Contig1	At5g67300	myb family transcription factor	7E-63
SBP	aCL6731Contig1	At2g47070	squamosa promoter-binding protein (SPL1)	1E-24
TCP	aCL5534Contig1	At2g30410	tubulin folding cofactor A (KIESEL)	2E-34
TUB	aCL1192Contig1	At1g25280	F-box family protein / tubby family protein	1E-168
TUB	aCL9120Contig1	At2g18280	tubby-like protein	7E-59
TUB	aCL7933Contig1	At2g47900	tubby-like protein	1E-175
VOZ-9	aC20003C09SK_c	At1g28520	expressed protein	5E-67
WHIRLY	aCL8737Contig1	At1g71260	expressed protein	1E-48
WRKY	aCL5056Contig1	At2g04880	WRKY family transcription factor (ZAP1)	8E-32
WRKY	aCL1506Contig1	At2g38470	WRKY family transcription factor	1E-104
WRKY	aC04005E02SK_c	At4g01720	WRKY family transcription factor	4E-11
WRKY	aCL3680Contig1	At4g18170	WRKY family transcription factor	7E-51
WRKY	aCL1201Contig1	At4g22070	WRKY family transcription factor	1E-111
WRKY	aCL3000Contig1	At4g23810	WRKY family transcription factor	4E-57
WRKY	aCL1204Contig3	At4g26640	WRKY family transcription factor	3E-37
WRKY	aC07003C03SK_c	At4g39410	WRKY family transcription factor	4E-21
ZF-HD	aCL4549Contig1	At1g69600	zinc finger homeobox family protein	1E-57
ZF-HD	aCL1228Contig1	At4g24660	zinc finger homeobox family protein	1E-58
