

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 988 846**

51 Int. Cl.:

C12P 23/00 (2006.01)

C12N 9/02 (2006.01)

C07K 14/415 (2006.01)

C12N 15/82 (2006.01)

C12N 15/52 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.07.2018 PCT/IB2018/055669**

87 Fecha y número de publicación internacional: **07.02.2019 WO19025935**

96 Fecha de presentación y número de la solicitud europea: **30.07.2018 E 18759402 (3)**

97 Fecha y número de publicación de la concesión europea: **07.08.2024 EP 3661951**

54 Título: **Genes y métodos para la preparación biotecnológica y compartimentación de apocarotenoides de alto valor añadido**

30 Prioridad:

03.08.2017 IT 201700089818

03.08.2017 IT 201700089858

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

21.11.2024

73 Titular/es:

**AGENZIA NAZIONALE PER LE NUOVE
TECNOLOGIE, L'ENERGIA E LO SVILUPPO
ECONOMICO SOSTENIBILE (ENEA) (50.0%)
Lungotevere, Thaon di Revel, 76
00196 Roma, IT y
UNIVERSITY OF ZÜRICH (50.0%)**

72 Inventor/es:

**GIULIANO, GIOVANNI;
FERRANTE, PAOLA;
DIRETTO, GIANFRANCO;
APREA, GIUSEPPE;
DEMURTAS, OLIVIA COSTANTINA;
PIETRELLA, MARCO;
MARTINOIA, ENRICO y
FRANCISCO, RITA**

74 Agente/Representante:

DURAN-CORRETJER, S.L.P

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 988 846 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Genes y métodos para la preparación biotecnológica y compartimentación de apocarotenoides de alto valor añadido

5

Sector técnico de la invención

La presente invención describe secuencias de ADN, vectores, proteínas y su uso para la producción biotecnológica de apocarotenoides de azafrán de alto valor añadido y su transporte en un compartimento subcelular o extracelular determinado. La presente invención en particular se refiere a genes, proteínas y métodos para producir apocarotenoides de azafrán en sistemas *in vitro*, en cepas modificadas genéticamente de *E. coli*, de levadura y en tejidos vegetales.

10

Estado de la técnica

15

Crocus sativus L. (Iridaceae) es una planta perenne, triploide, estéril, propagada y cultivada en zonas templadas como España o Cachemira. Sus estigmas secos constituyen la especia del azafrán, comúnmente considerada la especia más cara del mundo, con precios que oscilan entre 2.000 y 7.000 €/kg. Los elevados precios del azafrán se deben al enorme trabajo manual asociado a la recogida de estigmas: 1 kg de azafrán seco requiere la recogida manual de unas 150.000 flores (<http://www.europeansaffron.eu/>). Los estigmas del azafrán acumulan una serie de compuestos apocarotenoides: la crocetina y sus formas glicosiladas (crocinas), responsable de la coloración roja de los estigmas, la picrocrocina, responsable de su sabor amargo, y el safranal, responsable de su sabor pungente. Además de ser ingredientes alimentarios de alto valor añadido, los compuestos apocarotenoides del azafrán tienen aplicaciones cosméticas y médicas tales como, a modo de ejemplo no exhaustivo, la demostrada función protectora frente a enfermedades degenerativas de la retina de la crocetina y las crocinas. Considerando estas aplicaciones y el elevado coste de los apocarotenoides del azafrán, que actualmente se producen mediante costosos procesos de extracción, existe un interés industrial en idear métodos biotecnológicos para su producción a bajo coste, mediante la expresión de las enzimas implicadas en su biosíntesis en sistemas microbianos o en plantas altamente productivas.

20

25

30

La biosíntesis de dichos compuestos comienza con el corte simétrico, en las posiciones 7,8 y 7',8', de la zeaxantina gracias a la dioxigenasa CCD2 (ver figura 1). Los dos productos de corte de la zeaxantina, 3-OH-β-ciclocitral y crocetina dialdehído, se deshidrogenan y glicosilan para producir picrocrocina y crocinas, respectivamente. En lo que respecta al segundo paso biosintético, la deshidrogenación de crocetina dialdehído en crocina, las solicitudes de patente WO2013021261 y WO2015132411 muestran que puede catalizarse en levadura a partir de aldehído deshidrogenasa endógena de levadura (ALDH) o de deshidrogenasa exógena de diversos orígenes, incluida *Crocus sativus*.

35

40

45

Sin embargo, ninguna de las dos solicitudes de patente internacionales muestra datos sobre la eficacia con la que los ALDH descritos catalizan la reacción. Considerando que la levadura no sintetiza crocetina dialdehído ni crocetina, y que *Crocus sativus* incluye otros aldehídos, además de la crocetina dialdehído, que probablemente sean el sustrato natural de las ALDH mostradas en las solicitudes de patente, para obtener la producción biotecnológica de crocinas es necesario detectar ALDH(s) que catalizan eficazmente la conversión de crocetina dialdehído en crocetina en los estigmas de *Crocus sativus*.

Shoib Ahmad Baba et al. "Análisis integral del transcriptoma de *Crocus sativus* para el descubrimiento y expresión de genes implicados en la biosíntesis de apocarotenoides" describe el análisis del transcriptoma de *C. sativus* para identificar genes implicados en la biosíntesis de carotenoides/apocarotenoides. En el estudio se identificaron varias enzimas implicadas en la biosíntesis de apocarotenoides al comparar secuencias de proteínas presentes en el transcriptoma con aquellas codificadas por genes conocidos.

50

El estado de la técnica reconstruido anteriormente resalta entonces la necesidad de proporcionar nuevas proteínas, genes y métodos para una producción eficaz de apocarotenoides de *Crocus sativus* L., en particular para la producción de crocetina y crocina.

55

Características de la invención

La presente invención se basa en el hallazgo de métodos y materiales para mejorar la producción de compuestos de la planta de azafrán en sistemas recombinantes, así como nucleótidos y polipéptidos útiles para producir compuestos tales como crocina y crocetina.

60

Los inventores de la presente solicitud de patente han identificado y caracterizado sorprendentemente una aldehído deshidrogenasa expresada en los estigmas de *C. sativus* capaz de catalizar la conversión de crocetina dialdehído en crocetina con alta eficacia. Para ello se analizó un banco de secuencias expresadas en el estigma de *C. sativus* mediante el algoritmo *Blast* y una secuencia de ALDH derivada de *Bixa orellana*,

65

que cataliza la conversión de bixina dialdehído a norbixina (GenBank: CAD70189.1). Mediante este análisis se obtuvieron diferentes secuencias de ALDH expresada en el estigma de *C. sativus* (CsALDH) (SEQ ID NO 1-8).

- 5 Las ALDH, a nivel de aminoácidos, tienen diferentes niveles de identidad con las ALDH mostradas en las solicitudes de patente WO2013021261 y WO2015132411 y con la secuencia de Bixa orellana utilizada para su identificación (ver tabla 1 y figura 3). Sin embargo, los datos experimentales producidos por los inventores, y mostrados en los ejemplos, muestran que sólo la enzima CsALDH7 que tiene la secuencia polipeptídica SEQ ID NO 8 es capaz de convertir crocetina dialdehído en crocetina con alta eficacia.

10

Tabla 1.

	CsALDH1 (SEQ ID NO 2)	CsALDH3 (SEQ ID NO 4)	CsALDH4 (SEQ ID NO 6)	CsALDH7 (SEQ ID NO 8)	SynALDH (SEQ ID Genbank: NP_442494.1)
ALD1	78,31	37,47	27,05	25,57	26,87
ALD2	28,03	27,36	20,62	42,77	43,05
ALD3	100	38,90	26,38	27,63	26,91
ALD4	20,72	22,67	22,11	17,28	18,18
ALD5	74,90	38,49	27,94	26,54	28,34
ALD6	26,23	29,86	99,01	22,57	24,37
ALD7	74,30	37,40	27,44	27,82	26,70
ALD8	27,15	24,72	23,45	34,66	36,20
ALD9	24,60	27,31	20,36	99,79	43,42
BixaALDH	72,27	35,21	26,93	26,67	26,71

15 **Tabla 1.** Porcentaje de identidad de aminoácidos entre las ALDH descritas en la presente solicitud de patente, las ALD descritas en los documentos WO2013021261 y WO2015132411 y ALDH de *Bixa orellana* (en columna BixaALDH). Los porcentajes de identidad superiores al 95 % se muestran en pernos. SynALDH es un ALDH de *Synechocystis* capaz de convertir apocarotenal y alcanal. Dicha ALDH puede expresarse en una levadura, como *Saccharomyces cerevisiae*, o en una bacteria como *Escherichia coli*, diseñada para acumular crocetina dialdehído mediante la expresión de genes exógenos capaces de catalizar la síntesis de zeaxantina y del gen CCD2 capaz de convertir zeaxantina a crocetina dialdehído. La expresión de la proteína ALDH puede obtenerse mediante el uso de secuencias de ADN capaces de mediar altos niveles de expresión en el organismo seleccionado, según métodos conocidos por el experto en la materia.

20 Además, los inventores, basándose en los datos de actividad, los análisis de los perfiles de hidrofobicidad y las alineaciones de las secuencias de CsALDH descritas en el presente documento, lograron identificar las secuencias consenso necesarias para obtener ALDH capaz de deshidrogenar aldehídos de apocarotenoides. Más en detalle han encontrado que para obtener ALDH capaz de deshidrogenar moléculas hidrófobas como los aldehídos de los apocarotenoides es necesario que la enzima comprenda:

30 1) una región parcialmente hidrófoba situada aguas arriba del primer sitio de unión a NAD(P) (a continuación se muestra un ejemplo en fuentes subrayadas) y del primer residuo catalítico (N, en negrita) (secuencias en las que habitualmente hay al menos 90, más preferiblemente 100-150 aminoácidos aguas abajo del comienzo de la proteína), que tiene, por ejemplo, la siguiente secuencia consenso: TXXXXXXXXXEPXGXVLXISAWN (SEQ ID NO 21), en donde X es cualquier residuo de aminoácido, ejemplos de dichas regiones son las siguientes secuencias TFPSVGNIVAEPFGVVLISAWN (SEQ ID NO 22), TFPSSAQIVSEPLGVVLISAWN (SEQ ID NO 23), TYSMMNFRVKKEPLGTVLIIGYPYN (SEQ ID NO 24).

40 2) Una cola hidrófoba en el extremo carboxi terminal precedida y seguida por residuos hidrófilos (que se muestran en el perno), que tienen, por ejemplo, la siguiente secuencia consenso: KKKXXXXALLXXNIXAXXLAFFGFSKX (SEQ ID NO 28), en la que X es cualquier residuo de aminoácido, ejemplos de dicha cola hidrófoba son KKRKIIRALLAGNIIALVLAFFGFSKS (SEQ ID NO 25), KKKMVLKALLSSNIFAAILAFFGFSKDS (SEQ ID NO 26), KGGVKGALMRWLVVVAGYYLSAYMKARRA (SEQ ID NO 27).

45 Los inventores han identificado y caracterizado además la enzima CsUGT2 (SEQ ID NO 10) que cataliza los primeros pasos de glicosilación de crocetina en crocinas, conduciendo a la formación de crocina 1 y 2' (crocinas que contienen 1 o 2 moléculas de glucosa, respectivamente). Dicha enzima tiene una identidad de secuencia del 90,5 % con UGTC2, una enzima descrita en (Moraga et al., 2004, Planta, 219, 955-966) como responsable de todos los pasos de la glicosilación de la crocetina. Los inventores también identificaron y

caracterizaron dos transportadores de azafrán, altamente expresados en los estigmas, que son capaces de transportar crocinas a través de una membrana lipídica. Dichos transportadores pertenecen a la familia de transportadores de membrana ABC (cassette de unión a ATP), subfamilia C. En el estado de la técnica no se describe el uso de transportadores ABC para dirigir la acumulación de crocinas en un determinado compartimento celular.

Para identificar los transportadores mencionados anteriormente, se analizó un banco de secuencias expresadas en el estigma de *Crocus sativus* mediante el algoritmo *Blast* y una secuencia del transportador ABC, subfamilia C (ABCC), derivado de *Arabidopsis thaliana* (GenBank: AEC09006.1). Mediante este análisis se obtuvieron las secuencias de los transportadores denominados CsABCC1 y CsABCC2 (SEQ ID NO 11-14) expresados en los estigmas de *Crocus sativus*. Dichos transportadores ABCC pueden expresarse en una levadura, como *S. cerevisiae* o en una planta como *Nicotiana benthamiana* diseñada para acumular crocinas mediante la expresión de genes exógenos capaces de catalizar la síntesis de zeaxantina, del gen CCD2 capaz de convertir la zeaxantina en crocetina dialdehído, una de las ALDH caracterizadas por nosotros y CsUGT2 capaz de convertir crocetina en crocinas 1 y 2'. Cuando se expresan en *N. benthamiana*, los transportadores CsABCC1 y CsABCC2 muestran una ubicación tonoplásica (membrana de vacuola) (figura 6). De esta forma, al expresar los transportadores CsABCC1 o CsABCC2 en plantas modificadas de *N. benthamiana*, es posible obtener la acumulación de crocinas en la vacuola.

Con base en lo dicho anteriormente la presente invención se refiere a: Una secuencia de aminoácidos que tiene una secuencia seleccionada entre SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o mayor al 85% con SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2), en donde dicha secuencia de aminoácidos es una proteína transportadora apta para transportar crocinas a través de una membrana lipídica.

Una secuencia de nucleótidos que codifica una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o mayor al 85 % con SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2).

Un vector de expresión que comprende:

- una secuencia de nucleótidos que codifica una aldehído deshidrogenasa que tiene la secuencia de aminoácidos SEQ ID NO 8 (CsALDH7) o una secuencia de aminoácidos con un porcentaje de identidad igual o superior al 98 % de SEQ ID NO 8, y/o
- una secuencia de nucleótidos que codifica una enzima que tiene la secuencia de aminoácidos SEQ ID NO 10 (CsUGT2) o una secuencia de aminoácidos con un porcentaje de identidad igual o superior al 91 % de la SEQ ID NO 10, y/o una secuencia de nucleótidos que codifica una proteína de transporte que tiene secuencia de aminoácidos SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o mayor al 85 % con SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2).

Un organismo o célula huésped recombinante que expresa:

- una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o superior al 85 % con SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2).

Un método para acumular crocinas en un compartimento subcelular determinado que comprende una etapa de expresión de una secuencia de nucleótidos que codifica una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2), opcionalmente con secuencias diana específicas para abordar un compartimento celular o extracelular específico, en particular en el que dicho compartimento subcelular es la vacuola.

La presente invención se entenderá mejor mediante la siguiente descripción detallada de algunas realizaciones preferentes de la misma, proporcionada con referencia a los dibujos adjuntos, en los que las figuras relacionadas ilustran:

Descripción de las figuras

Figura 1 Ruta biosintética de apocarotenoides en azafrán. Ruta biosintética propuesta para la biosíntesis de apocarotenoides. La zeaxantina se corta en las posiciones 7,8 y 7',8' mediante la actividad de un CCD específico de *C. sativus* (CCD2, Giuliano et al. Solicitud de patente RM2014A000415, Frusciante et al., 2014, Proc Natl Acad Sci U S A, 111, 12246-12251). El producto de corte C₂₀, crocetina dialdehído, se convierte en crocetina mediante una aldehído deshidrogenasa y luego en crocinas mediante la actividad de al menos dos UDP-glicosiltransferasas (UGT).

Figura 2 Grupos de glucosa de las crocinas más abundantes de *C. sativus* (crocinas 1-4).

Figura 3 Árbol filogenético de las ALDH descritas en la presente solicitud de patente (designadas con una mancha negra). El árbol se creó utilizando las secuencias de aminoácidos con el programa MEGA7 aplicando el método Neighbor-Joining con bootstrap en 500 repeticiones. Las distancias evolutivas se calcularon utilizando el método de distancia p y se representan en número de diferencias de aminoácidos por sitio. Los números que se muestran en las ramas designan la probabilidad (en porcentaje) de que esa rama exista.

Figura 4. Modelo propuesto para la biosíntesis/compartimiento de crocinas de azafrán. La enzima CCD2 corta la zeaxantina en el cromatóforo produciendo crocetina dialdehído que se transporta en el citoplasma. La conversión de crocetina dialdehído en crocetina se realiza mediante la enzima CsALDH7 que está anclada a una membrana citoplasmática, probablemente el retículo endoplásmico (RE) mediante su cola hidrofóbica C-terminal y, junto con la enzima CsUGT2, produce crocinas 1 y 2' en el citoplasma. Una segunda UGT aún no identificada convierte las crocinas 1 y 2 en las crocinas 2, 3 y 4. Luego, las crocinas son transportadas en la vacuola, mediante la acción de los transportadores tonoplásticos CsABCC1 y ABCC2.

Figura 5. Árbol filogenético de los transportadores ABCC. El árbol fue creado con el programa MEGA7 (Kumar S et al. Mol Biol Evol. [Epub ahead of print] PubMed PMID: 27004904) aplicando el método Neighbor-Joining con bootstrap en 500 repeticiones. Las distancias evolutivas se calcularon utilizando el método de distancia p y se representan en número de diferencias de aminoácidos por sitio. Los números que se muestran en las ramas designan la probabilidad (en porcentaje) de que esa rama exista. Los transportadores de *C. sativus* que caracterizamos están designados por una mancha negra. Cs: *Crocus sativus*; Vv: *Vitis vinífera*; At: *Arabidopsis thaliana*; Os: *Oryza sativa*.

Figura 6. Imágenes de microscopio confocal de hojas de *N. benthamiana* que expresan el transportador CsABCC1 o CsABCC2 fusionado con la proteína fluorescente verde (CsABCC1:GFP; CsABCC2:GFP) y el marcador tonoplástico γ TIP fusionado con la "proteína fluorescente roja" (γ TIP:RFP). Para cada transportador se muestran las imágenes de fluorescencia de RFP y GFP. Se observa que la señal de GFP, debido a la expresión de los transportadores ABC fusionados a GFP, co-localiza con la señal de RFP, debido a la expresión del marcador tonoplástico γ TIP, mostrando entonces que los dos transportadores ABC se ubican en el tonoplasto.

Figura 7. Expresión de ALDH en *E. coli*.

(a) Representación esquemática de las construcciones expresadas en la cepa de *E. coli* que acumula zeaxantina. Cada gen se inserta en el vector pTHIO, fusionado en 5' con el gen de tiorredoxina y en 3' de la secuencia que codifica una etiqueta para la purificación (etiqueta His6). Cm^r: gen de resistencia al cloranfenicol; Amp^r: gen de resistencia a ampicilina.

(b) Western blot realizado con un anticuerpo anti-His6 de las proteínas totales extraídas de las cepas bacterianas recombinantes antes y después de la inducción con arabinosa (-) y (+). Las flechas muestran las proteínas CCD2 y ALDH.

Figura 8. Acumulación de crocetina en los clones bacterianos que expresan las enzimas CCD2 y ALDH

(a) Cromatogramas HPLC-HRMS de crocetina (masa exacta extraída de ion M+H⁺=329,1747) del estándar analítico (ESTÁNDAR) y de los clones bacterianos que sobreexpresan la enzima CCD2 sola (CCD2) o en combinación con las enzimas CsALDH7 o CsALDH1 (CCD2/CsALDH7; CCD2/CsALDH1). Se observan dos picos de crocetina (isoforma *trans* y *cis*).

(b) Diagrama de los niveles de crocetina (isómeros *trans* y *cis*) acumulados en los clones bacterianos recombinantes. Los resultados se muestran como valores promedio de tres repeticiones biológicas y las barras de error representan desviaciones estándar. Plegado IS= plegado del patrón interno (abundancia respecto al patrón interno)

Figura 9. La enzima CsUGT2 cataliza la conversión de crocetina en crocina 1 y crocina 2'.

(a) Espectros de absorción a 440 nm de los metabolitos extraídos de las reacciones de ensayos *in vitro* realizados durante 60 minutos en ausencia (pTHIO) y en presencia de la enzima CsUGT2. Incluso se muestra el espectro del extracto de azafrán. t=*trans*; c=*cis*.

(b) Cromatogramas de HPLC/HRMS (masas exactas extraídas) de *trans* crocina 1 (M+H⁺=491,2275) y *trans* crocina 2' (M+H⁺=653,2803) que se acumulan en presencia de la enzima CsUGT2. Los cuadros a la izquierda del pico de las crocinas muestran su espectro de absorción, mientras que los cuadros a la derecha muestran el espectro de masas.

(c) Ruta temporal de la conversión de crocetina en *trans* crocina 1 y *trans* crocina 2' en presencia de la enzima CsUGT2. Los resultados se muestran como valores promedio de tres repeticiones biológicas y las barras de error representan las desviaciones estándar.

Figura 10. Transporte de apocarotenoides de azafrán mediante transportadores CsABCC1 y CsABCC2 de *C. sativus*.

(a) Abundancia relacionada (expresada como abundancia con respecto al estándar interno) de

metabolitos glicosilados identificados en el extracto hidroalcohólico de azafrán mediante LC-HRMS. 1: canferol 3-O-soforosido-7 glucósido; 2: canferol 3,7,4'-triglucósido; 3: canferol 7-soforosido; 4: canferol 3-β-D-glucopiranosido; 5: canferol 3-rutinósido-7-glucósido; 6: acetilglucósido de canferol; 7: canferol -3-O-rutinósido; 8: dihidrocanferol-7-O-glucósido; 9: isorhamnetina-3-4'-diglucósido; 10: ramnetina 3-rutinósido; 11: quercetina-3-diglucósido; 12: quercetina-3-O-rutinósido; 13: naringenina-7-O-glucósido (isoforma 3); 14: picrocrocina; 15: transcrocina 5; 16: transcrocina 4; 17: cis-crocina 4; 18: transcrocina 3; 19: cis-crocina 3; 20: transcrocina 2; 21: cis crocina 2; 22: trans crocina 2'; 23: cis crocina 2'; 24: trans crocina 1; 25: cis crocina 1.

(b) Transporte neto de metabolitos de azafrán por transportadores de *C. sativus* CsABCC1 y CsABCC2 y su control pNEV (vector vacío). Los experimentos de transporte se realizaron a temperatura ambiente (RT), en hielo (ICE) y en presencia del inhibidor específico de los transportadores ABCC, Probenecid (+ Probenecid). Cabe señalar que a temperatura ambiente algunos metabolitos, principalmente crocinas, se transportan, mientras que la reacción se inhibe si se realiza en hielo o en presencia de un inhibidor de probenecid. El transporte neto se calculó deduciendo los valores medidos en ausencia de ATP de los valores medidos en presencia de ATP después de 15 minutos de transporte.

Los resultados se muestran como valores promedio de tres replicaciones biológicas y las barras de error representan desviaciones estándar.

20 Descripción detallada de la invención

La presente invención se ilustrará mejor con referencia a los siguientes ejemplos de realizaciones, que se proporcionan a modo de ilustración y no como ejemplo limitante.

En el presente documento se da a conocer, pero no forma parte de la presente invención, una secuencia de aminoácidos que tiene la SEQ ID NO 8 (CsALDH7), en la que dicha secuencia de aminoácidos SEQ ID NO 8 es una aldehído deshidrogenasa apta para catalizar la conversión de crocetina dialdehído en crocetina con alta efectividad. Se divulgan en el presente documento, pero no forman parte de la presente invención, secuencias de aminoácidos con un porcentaje de identidad igual o superior al 98 % con la SEQ ID NO 8, preferentemente igual o superior al 99 %. Dichas secuencias, por ejemplo, podrían tener mutaciones conservativas con respecto a la SEQ ID NO 8 que no alteran sustancialmente la eficacia y la especificidad catalítica de la enzima.

Se describe en este documento, pero no es parte de la presente invención, una secuencia de aminoácidos que tiene SEQ ID NO 10 (CsUGT2), en la que dicha secuencia de aminoácidos CsUGT2 es una enzima capaz de catalizar la glicosilación de crocetina en crocina 1 y 2'. También se da a conocer en el presente documento, pero no forma parte de la presente invención, secuencias de aminoácidos con un porcentaje de identidad igual o superior al 91 % de la SEQ ID NO 10, preferiblemente superior al 92 %, 95 % o al 99 % de identidad.

La presente invención se refiere además a los polipéptidos que tienen una secuencia seleccionada entre SEQ ID NO 12 (CsABCC1) y SEQ ID NO 14 (CsABCC2), en las que dichas secuencias de aminoácidos SEQ ID NO 12 (CsABCC1) y SEQ ID NO 14 (CsABCC2) son proteínas de transporte capaces de transportar crocinas a través de una membrana lipídica. La presente invención también se refiere a secuencias de aminoácidos con un porcentaje de identidad igual o superior al 85 %, preferentemente superior al 90 %, 95 % o 99 % de identidad respecto a la SEQ ID NO 12 (CsABCC1) o a la SEQ ID NO 14 (CsABCC2). Tales secuencias, por ejemplo, podrían tener mutaciones conservadoras con respecto a SEQ ID NO 12 (CsABCC1) y SEQ ID NO 14 (CsABCC2) que no alteran la función de transporte de proteínas específica.

Las secuencias de aminoácidos aquí descritas pueden incluir aminoácidos adicionales que no participan en la actividad enzimática o de transporte específica realizada por la proteína y luego dichas secuencias, por ejemplo, podrían incluir además una etiqueta de purificación (por ejemplo, etiqueta HIS o etiqueta GST) o secuencias que estabilizan la proteína (por ejemplo tiorredoxina utilizada en las construcciones para la expresión de enzimas ALDH en *E. coli*, figura 7a). En algunas realizaciones, las secuencias de aminoácidos descritas en el presente documento podrían incluir, por ejemplo, una secuencia de aminoácidos que actúa como indicador, por ejemplo GFP (proteína verde fluorescente). El porcentaje de identidad de cualquier ácido nucleico o polipéptido (secuencia de aminoácidos) con un ácido nucleico o polipéptido de referencia se puede determinar tal como lo conoce el experto en la técnica. Una secuencia de referencia (por ejemplo una secuencia de ácido nucleico o una secuencia de aminoácidos) se alinea con una o más secuencias de candidatos utilizando el programa Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>, parámetros predefinidos) que permiten realizar alineamientos de secuencias de ácido nucleico o polipéptidos a lo largo de toda su longitud (alineamiento global). Véase Chenna et al., *Nucleic Acids Res.*, 31 (13): 3497-500 (2003).

La presente invención también se refiere a las siguientes secuencias de nucleótidos: las secuencias de nucleótidos que codifican una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2), por ejemplo la secuencia de nucleótidos SEQ ID NO 11 o SEQ ID

NO 13.

Según una realización, podrían usarse secuencias de nucleótidos sintéticas, por ejemplo usando codones optimizados según la célula u organismo huésped recombinante seleccionado. Debido a la degeneración del código genético, para muchos aminoácidos existe más de un triplete de nucleótidos para codificar el mismo aminoácido. Luego, el codón en la secuencia codificante de un polipéptido determinado puede modificarse para optimizar la expresión en un organismo huésped particular, utilizando los métodos conocidos por el experto en la técnica. Según una realización, incluso todos los ADNc de las secuencias de nucleótidos aquí descritas pertenecen a la presente invención.

La presente invención se refiere además a vectores de expresión, células huésped recombinantes y organismos que expresan una o más de las secuencias de aminoácidos aquí descritas. Podrían usarse los vectores de expresión conocidos por el experto en la técnica para expresar la enzima o proteína implicada adecuadamente para el sistema de expresión específico. Los vectores de expresión se pueden encontrar comercialmente o se pueden construir ad hoc tanto para la expresión en sistemas microbianos como vegetales. Para expresar las proteínas de *C. sativus* en células bacterianas como vectores de ejemplo no exhaustivos pTHIO-Dan1 (mostrado en la figura 7a), se pueden usar vectores pQE (Qiagen), vectores para la serie pET (Invitrogen), mientras que los vectores para la expresión de levadura puede ser, por ejemplo, pNEV (usado en la figura 10), pAG (gen Add), pYES (Thermo Fisher). Para la expresión de las enzimas y de los transportadores en plantas, se pueden utilizar, por ejemplo, vectores derivados de pBI121 (Clontech) o de pCAMBIA (Gambia).

Dichos vectores podrían comprender, además de las secuencias de nucleótidos aquí descritas, una "región reguladora" (procariótica y eucariota), es decir, una secuencia de nucleótidos que influyen en la transcripción o traducción y la velocidad, la estabilidad y/o la movilidad de un producto de transcripción o traducción. Las regiones reguladoras incluyen, por ejemplo, secuencias de promotores, secuencias de potenciación, elementos de respuesta, sitios para reconocer las proteínas, elementos inducibles, secuencias de unión de proteínas, regiones no traducidas por 5' y 3' (UTR), intrones y combinaciones de los mismos. La región reguladora estará unida operativamente a las secuencias de nucleótidos que codifican una o más proteínas de la presente invención.

Las células y organismos recombinantes que incluyen las secuencias de nucleótidos o la expresión aquí descritas podrían ser, sin ninguna limitación, todos los sistemas de expresión conocidos por el experto en la materia adecuados para expresar una o más secuencias de aminoácidos involucradas, por ejemplo SEQ ID NO 8 (CsALDH7, que no forma parte de la invención), SEQ ID NO 10 (CsUGT2, que no forma parte de la invención) SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2).

Según una realización, las células y organismos recombinantes expresan una o más proteínas recombinantes que tienen SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2), junto con la enzima CCD2 que corta la zeaxantina en la posición 7,8 y 7',8'. Según una realización preferente, será una célula u organismo recombinante huésped que acumule zeaxantina.

Cualquiera de los organismos huéspedes aquí descritos puede ser un microorganismo, una planta o una célula vegetal. El microorganismo puede ser, por ejemplo, una levadura como *Saccharomyces cerevisiae* o una bacteria como *Escherichia coli*. La planta o célula vegetal puede ser, por ejemplo, *Arabidopsis thaliana*, *Nicotiana benthamiana*, *Zea mays*, *Solanum lycopersicum*, *Solanum tuberosum*, *Oryza sativa*.

Una célula huésped recombinante descrita en el presente documento puede ser una célula de levadura, una célula vegetal, una célula de mamífero, una célula de insecto o una célula bacteriana. En algunas realizaciones, la célula de levadura es una célula de *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha*, *Candida boidinii*, *Arxula adeninivorans*, *Candida albicans*.

En algunas realizaciones, los ácidos nucleicos y los polipéptidos aquí descritos se insertan en plantas. Por lo tanto, un huésped puede ser una planta o una célula vegetal que incluye al menos un gen recombinante aquí descrito. Se puede transformar una célula vegetal o una planta que tiene un gen recombinante integrado en su genoma, es decir, se puede transformar de forma estable. Las células transformadas de forma estable normalmente mantienen el ácido nucleico insertado con cada división celular. Una célula vegetal o una planta puede incluso ser transformada transitoriamente de modo que el gen recombinante no se integre en su genoma. Las células vegetales transgénicas utilizadas en los métodos aquí descritos pueden constituir una porción o una planta completa. Las técnicas para la inserción de ácidos nucleicos en plantas monocotiledóneas y dicotiledóneas son conocidas en el estado de la técnica e incluyen por ejemplo la transformación mediada por *Agrobacterium* y la electroporación.

Según una realización, las células y los organismos recombinantes se preparan para sobreexpresar enzimas que conducen a la acumulación del precursor de la ruta biosintética de los apocarotenoides en el azafrán

(zeaxantina), tales como por ejemplo crtE (Genbank ALK24264.1), crtYB (Genbank AAO73816 .1), genes crtI (Genbank ALK24268.1) de la levadura *Xanthophyllomyces dendrorhous* y crtZ de la bacteria *Erwinia uredovora* (Genbank ADD79330.1).

5 Los microorganismos hospedadores recombinantes podrían prepararse transformándolos con las secuencias de nucleótidos, por ejemplo con los vectores descritos anteriormente, de acuerdo con las técnicas conocidas, de modo de expresar las proteínas involucradas descritas aquí.

10 También se divulga, pero no forma parte de la presente invención, un método para producir crocetina y/o crocina en un organismo huésped o célula que acumula zeaxantina, que comprende una etapa de expresión en dicho organismo de la enzima que tiene la secuencia de aminoácidos SEQ ID No. 8 (CsALDH7) y/o SEQ ID N.º 10 (CsUGT2). También se divulga, pero no forma parte de la presente invención, una realización en la que el organismo o célula huésped recombinante se transforma con las secuencias de nucleótidos aquí descritas, por ejemplo SEQ ID N. 7 (CsALDH7) y/o SEQ ID N. 9 (CsUGT2), para sobreexpresar las enzimas involucradas.

15 La presente invención se refiere además a un método para acumular crocinas en un compartimento subcelular determinado de un organismo huésped recombinante que comprende una etapa de expresión de una secuencia de nucleótidos que codifica una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2).

20 En dichos métodos podría usarse cualquiera de los organismos o células recombinantes descritos en el presente documento, en particular *Saccharomyces cerevisiae* y *Nicotiana benthamiana* transformados con una de las secuencias o vectores descritos en el presente documento, para sobreexpresar las secuencias de aminoácidos implicadas. Bajo el término "sobreexpresar" se entiende que los organismos recombinantes expresan en mayor cantidad el gen exógeno con respecto al posible gen endógeno correspondiente.

25 En el presente documento se da a conocer, pero no forma parte de la presente invención, un método para deshidrogenar un aldehído de apocarotenoide, en particular para deshidrogenar crocetina dialdehído en crocina, que comprende una etapa para deshidrogenar dicho aldehído con una aldehído deshidrogenasa que comprende:

- 30
- 1) una región parcialmente hidrófoba situada después de al menos 90 residuos del extremo amino terminal, preferiblemente después de 100-150 residuos, y antes del sitio de unión de nicotinamida adenina dinucleótido fosfato (NAD(P));
 - 2) una cola hidrófoba situada en el extremo C-terminal precedida y seguida por uno o más residuos hidrófilos, por ejemplo por dos, tres, cuatro residuos.

35 También se da a conocer, pero no forma parte de la presente invención, una realización en la que dicha región parcialmente hidrófoba tiene la siguiente secuencia consenso: TXXXXXXXXXEPXGXVLX(SAWN) (SEQ ID NO 21), en la que X es cualquier residuo de aminoácido. Ejemplos de dichas regiones son las siguientes secuencias TFPSVGNIVAEPFGVLLI(SAWN) (SEQ ID NO 22), TFPSSAQIVSEPLGVVLV(SAWN) (SEQ ID NO 23), TYSMMNFRVKKEPLGTVLI(GPYN) (SEQ ID NO 24).

40 También se da a conocer, pero no es parte de la presente invención, una realización en la que dicha cola hidrófoba tiene la siguiente secuencia de consenso: KXXXXXXXXALLXXNIXAXLAFFGF(SKX) (SEQ ID NO 28), en la que X es cualquier residuo de aminoácido.

45 Ejemplos de dicha cola hidrófoba son las siguientes secuencias KKRKIIIRALLAGNIIALVLAFFGF(SKS) (SEQ ID NO 25), KKKMVLKALLSSNIFAAILAFFGF(SKDS) (SEQ ID NO 26), KGGVKGALMRWLVVAGYYLSAYMKARRA (SEQ ID NO 27).

50 También se da a conocer, pero no forma parte de la presente invención, una aldehído deshidrogenasa que tiene en el extremo carboxilo terminal la secuencia de aminoácidos KGGVKGALMRWLVVAGYYLSAYMKARRA (SEQ ID NO 27) o una secuencia de aminoácidos con 80 % de identidad, preferiblemente con 85 %, 90 %, 95 %, 99 % de identidad con SEQ ID NO 27. Además, la presente invención también se refiere a la SEQ ID NO 27 sobre sí misma y a secuencias de aminoácidos con un 80 % de identidad, preferiblemente con un 85 %, 90 %, 95 %, 99 % de identidad con esta secuencia.

55 También se da a conocer, pero no forma parte de la presente invención, un método para evaluar la capacidad de un transportador de transportar un metabolito determinado, en este documento denominado ensayo de "transportómica", que comprende los siguientes pasos:

- 60
- incubar una preparación que comprende un transportador con un extracto vegetal complejo que comprende una pluralidad de metabolitos;
 - detectar la cantidad de metabolito que fue transportado.
- 65

También se da a conocer, pero no forma parte de la presente invención, una realización en la que la preparación utilizada en el método está constituida por microsomas de levadura que incluyen el transportador a caracterizar.

5

También se describe pero no forma parte de la presente invención una realización en la que el extracto vegetal será un extracto hidroalcohólico o un extracto obtenido con disolventes apolares de órganos o tejidos, por ejemplo de *Crocus sativus* u otras plantas involucradas.

10

También se da a conocer pero no forma parte de la presente invención un método para deshidrogenar el retinal u otros apocarotenos (aldehído de apocarotenoides), con un número de átomos de carbono superior a 20 que comprende una etapa de deshidrogenar dichos aldehídos con una aldehído deshidrogenasa que tiene SEQ ID NO 8 (CsALDH7).

15

También se da a conocer pero no forma parte de la presente invención un método para la conversión de apocarotenos con un número de átomos de carbono mayor o igual a 20 en el ácido carboxílico correspondiente en un organismo o célula huésped capaz de acumular los apocarotenos antes mencionados. que comprende una etapa de expresión en dicho organismo o célula de la enzima que tiene una secuencia de aminoácidos que tiene SEQ ID NO 8 (CsALDH7). También se divulga, pero no forma parte de la presente invención, una realización en la que el organismo huésped es una bacteria, una levadura o una planta recombinante, por ejemplo seleccionada de *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha*, *Candida boidinii*, *Arxula adenivorans*, *Xanthophyllomyces dendrorhous*, *Candida albicans* especie, *Escherichia coli*, *Arabidopsis thaliana*, *Nicotiana benthamiana*, *Zea mays*, *Solanum lycopersicum*, *Solanum tuberosum*, *Oryza sativa*.

20

25

Ejemplos

Ejemplo de referencia 1: Expresión de diferentes ALDH en bacteria *E. coli*.

30

Los genes que codifican la ALDH se aislaron a partir del ARN del azafrán con oligonucleótidos específicos. Luego, los genes se clonaron en un vector para la expresión en una bacteria que tenía una resistencia a los antibióticos diferente del vector que contenía CCD2 (métodos descritos en 5,6). Dichos vectores se introdujeron en una cepa de *E. coli* capaz de producir zeaxantina y que contenía un vector que expresa CCD2, y luego capaz de producir crocetina dialdehído (Giuliano et al. Solicitud de patente RM2014A000415, Frusciante et al Proc Natl Acad Sci U S A 111 (33):12246-12251). Las cepas bacterianas así obtenidas, cada una de las cuales expresa tanto el gen CCD2 como uno de los genes ALDH (figura 7a), se cultivaron en condiciones que conducen a la expresión de las proteínas CCD2 y ALDH (Giuliano et al. Solicitud de patente RM2014A000415, Frusciante et al Proc Natl Acad Sci U.S.A 111 (33):12246-12251). Como controles se utilizaron las cepas que incluían el vector vacío o el vector que incluía el gen CCD2 (figura 7a). El análisis de la expresión proteica correcta se realizó mediante SDS-PAGE seguido de Western blot (figura 7b) realizado como se describe en Demurtas et al. Front Plant Sci.. 7:54).

35

40

El microorganismo así obtenido puede analizarse mediante uno de los métodos conocidos por el experto en la materia para detectar la presencia y cantidad de zeaxantina, crocetina dialdehído y crocetina.

45

Ejemplo de referencia 2: Identificación y cuantificación de crocetina en una bacteria *E. coli* que expresa diferentes ALDH.

50

La producción de crocetina y la cuantificación de los niveles producidos se analizaron en extractos obtenidos de los clones recombinantes obtenidos mediante cromatografía líquida acoplada a espectrometría de masas de alta resolución (LC-HRMS). Después de la inducción de las proteínas CCD2 y ALDH, las células se recogieron mediante centrifugación y se resuspendieron en disolventes específicos para la extracción de metabolitos polares y apolares. Los metabolitos polares extraídos con metanol 75 % fueron separados con una columna cromatográfica C18 Luna de fase reversa (150 × 2,1 mm, 3µm; Phenomenex) e ionizados mediante electrospray (fuente ESI) (Fasano et al. New Phytologist 210: 1382- 1394). Por el contrario, los metabolitos no polares se extrajeron con acetona, se separaron con una columna cromatográfica C30 con fase reversa (100 × 3,0 mm; YMC Europe) y se ionizaron mediante ionización química a presión atmosférica (fuente APCI).

55

60

La masa correspondiente al ion derivado de la crocetina ($M+H^+=329,1747$) se extrajo de los cromatogramas y se comparó con la del estándar auténtico (figura 8a). La crocetina producida existe en forma de 2 isómeros: trans e cis. Los niveles de crocetina trans y cis acumulados por cada cepa de bacteria recombinante se muestran en la figura 8b.

65

Ejemplo 3: Expresión de transportadores ABCC de *C. sativus* en microsomas de *S. cerevisiae*.

Los genes que codifican los transportadores CsABCC1 y CsABCC2 se aislaron a partir de ARN de azafrán con oligonucleótidos específicos. A continuación, los genes se clonaron en un vector para la expresión en *Saccharomyces cerevisiae*, pNEV-Ura, y se expresaron en una cepa de *S. cerevisiae* defectuosa para el transportador ABCC "Yeast Bile Transporter1" (YBT1). Como comprobación, se transformó la misma cepa con el vector pNEV vacío.

A partir de las cepas transformadas de levadura se realizaron preparaciones de microsomas, es decir se extrajeron las vesículas totales del sistema endomembranoso celular como se describe en Tommasini et al. Proc Natl Acad Sci USA. 93(13):6743-6748. Dichas vesículas incluyen el transportador ABCC que se va a caracterizar.

Para evaluar la integridad fisiológica de las vesículas aisladas, analizamos la capacidad de transportar el sustrato leucotrieno C4 (LTC4) (Leier et al., 1994, J Biol Chem 269(45): 27807-10).

Ejemplo 4: Desarrollo y uso de un ensayo *in vitro* para caracterizar la funcionalidad y estudiar la especificidad de los transportadores ABCC de *C. sativus*.

Se preparó un nuevo ensayo *in vitro*, a diferencia del ensayo de transporte clásico, en el que las vesículas se dejan incubar durante periodos de tiempo preestablecidos con el supuesto sustrato, se lavan, se lisan y se analiza el contenido mediante cromatografía líquida acoplada a un detector UV/visible. Con respecto a los trabajos existentes en la literatura sobre la caracterización funcional *in vitro* de transportadores, la presente invención introduce el concepto de "transportómica": el ensayo se basa en el análisis del transporte, realizado por LC-HRMS, de todas las moléculas existentes en el extracto de azafrán. Este enfoque, que prevé el uso de un extracto crudo en lugar de estándares purificados comerciales, permite caracterizar el transporte incluso de moléculas que no están disponibles comercialmente; Además, al basarse en datos obtenidos con espectrometría de masas de alta resolución, en lugar de con detectores en serie de fotodiodos (PDA), el sistema es extremadamente sensible y permite apreciar el transporte de metabolitos poco abundantes.

La figura 10a muestra el contenido de los principales metabolitos glicosilados existentes en el extracto hidroalcohólico de azafrán, mientras que la figura 10b muestra los metabolitos que son transportados específicamente por los transportadores CsABCC1 y CsABCC2 de *C. sativus*.

Con este método, entonces, se observó que los dos transportadores ABCC son capaces de transportar diferentes crocinas además de algunos flavonoides (figura 10b, barras negras). Dicho transporte no se debe a una simple extensión de los sustratos sino que se debe a la acción de una enzima ya que al realizar el ensayo en hielo (donde se reduce la actividad enzimática) se nota la reducción del transporte (figura 10b, barras blancas punteadas). Además, el transporte se debe a la actividad de los transportadores ABCC, ya que añadiendo inhibidor de Probenecid (un inhibidor de los transportadores ABC) se obtiene una reducción considerable del transporte (figura 10b, barras blancas).

Procedimientos experimentales:

Aislamiento de genes de *C. sativus*.

Los genes CsALDH, CsUGT2, CsABCC1 y CsABCC2 fueron aislados mediante ARN extraído de estigmas de azafrán recogidos el día anterior a la antesis y procedentes de la empresa Zafferanami, Varedo (MI) (<http://zafferanami.tumblr.com/>). En detalle, el ARN se retrotranscribió en ADNc mediante el kit de síntesis de ADNc Omniscrypt RT (Qiagen). Las secuencias codificantes (CDS) de los genes se obtuvieron amplificando el ADNc con la ADN polimerasa Phusion High Fidelity (NEB) con pares de oligonucleótidos específicos para cada gen:

CsALDH1: ATGGCTGCAACTAACAGCAATG(SEQ ID NO 15); TCACAGCCAAGGTGAGTTATAT (SEQ ID NO 34)

CsALDH3: ATGTCGATGCTACGCGTGG(SEQ ID NO 16); TCAAGACTCCAAGTTCCCC (SEQ ID NO 33)

CsALDH4: ATGGGATTTACCAAGGAGCAC (SEQ ID NO 35); CTAGCCAAAGTTTATTCCCTG (SEQ ID NO 36)

CsALDH7: ATGGCCTTCGATGGAGAGAAAG (SEQ ID NO 17); TCAAGATTTGGAGAAGCCAAAG (SEQ ID NO 18)

CsUGT2: ATGGAGCAGAAAGATGTGAA(SEQ ID NO 19); TCATTTGCAACACTGATGTATG (SEQ ID NO 32)

CsABCC 1: ATGTCCTCCTCTCCCTAC(SEQ ID NO 20)

; TCAAAGATCAGAAGAGCGGTT () (SEQ ID NO 31)

CsABCC2: ATGGGTTTCAAACCGCTTG(SEQ ID NO 29); CTACATTTGAGCATTGTCC) (SEQ ID NO 30)

Los amplicones obtenidos fueron luego clonados en vectores para la expresión en bacterias, levaduras y plantas.

5

Caracterización funcional de enzimas ALDH mediante ensayo *in vivo* en bacteria.

Para analizar las actividades enzimáticas de varias ALDH, los genes se clonaron en el vector pTHIO (que contiene el gen de resistencia a la ampicilina, Amp^r) (Trautmann et al., 2013, Febs j, 280(15): 3685-96) para la expresión en bacteria y las construcciones obtenidas se insertaron mediante electroporación en una cepa de *E. coli* que acumula zeaxantina (que contiene el gen de resistencia a la kanamicina, Kan^r) descrita en (Frusciante et al 2014, Proc Natl Acad Sci U S A 111 (33):12246-12251) y que porta el plásmido pTHIO-CCD2 (que contiene el gen de resistencia al cloranfenicol, Cm^r) (figura 7a). Como control, se utilizó la cepa que portaba únicamente el gen CCD2 o que portaba el plásmido pTHIO vacío. Las bacterias recombinantes así obtenidas se cultivaron a 37°C en medio líquido LB que contenía antibióticos con la mitad de concentración (25 µg/ml de kanamicina, 12,5 µg/ml de cloranfenicol y/o 50 µg/ml de ampicilina), hasta alcanzar una densidad óptica de 600 nm de 0,7. Luego se indujo la expresión de proteínas con arabinosa al 0,08 % a 20 °C durante 16 horas. Luego, las células se recogieron mediante centrifugación y los sedimentos bacterianos se resuspendieron en disolventes específicos para extraer los metabolitos polares/semipolares y no polares. Los metabolitos polares/semipolares se extrajeron como se describe en (Fasano et al. New Phytologist 210: 1382-1394) con pequeñas modificaciones: los sedimentos provenientes de 50 ml de cultivo líquido se resuspendieron en 10 ml de metanol frío, incluyendo 0,5 µg/ml de formononetina, Sigma-Aldrich), se lisaron en hielo mediante sonicación y se centrifugaron a 18.000 g durante 30 minutos. El sobrenadante se secó y se resuspendió en 200 µl de metanol al 75 %, se centrifugó adicionalmente durante 20 minutos para eliminar posibles materiales agregados y se precipitó y se sometió a análisis HPLC-HRMS. Los metabolitos no polares se extrajeron con el mismo procedimiento, utilizando acetona que incluía 0,5 µg/ml de acetato de α-tocoferol para la extracción y cloroformo para la resuspensión.

30

Caracterización funcional de la enzima CsUGT2 mediante ensayo *in vitro*.

Para analizar la actividad enzimática de CsUGT2, el gen se clonó en el vector pTHIO que contiene el gen de resistencia a la espectinomicina, Spect^r) y la construcción obtenida se insertó mediante electroporación en la cepa de *E. coli* BL21 (pGro7) descrita en (Frusciante). et al 2014, Proc Natl Acad Sci U S A 111 (33): 12246-12251). La expresión de la proteína CsUGT2 se indujo como se describe para ALDH. Luego se desarrolló un ensayo *in vitro* utilizando el lisado bruto de las bacterias que sobreexpresan la enzima CSUGT2 obtenida utilizando el tampón LEW como se describe en (Frusciante et al 2014, Proc Natl Acad Sci U S A 111 (33): 12246-12251). El ensayo se realizó en 100 µl de reacción utilizando el lisado bacteriano que contenía 40 µg de proteínas totales, el sustrato crocetina (100 µM) encapsulado en 6-O-α-maltosil-β-ciclodextrinas y sustrato UDP-glucosa (2,5 mM) como se describe en (Moraga et al., 2004, Planta, 219, 955-966). La reacción se realizó a 30°C y se detuvo en diferentes periodos de tiempo (0, 5, 10, 20, 60, 120 minutos) añadiendo 100 µl de etanol frío y congelando a -20°C. Los metabolitos polares/semipolares se extrajeron mediante sonicación y centrifugación como se describe para las ALDHs.

45

Caracterización de la localización subcelular de los transportadores ABC.

Los transportadores CsABCC1 y CsABCC2 se fusionaron en el extremo carboxi terminal con la "proteína fluorescente verde mejorada" (eGFP) y se clonaron en el vector pBI121 (Clontech). Como marcador tonoplástico se utilizó la proteína γ-TIP fusionada a "Proteína Fluorescente Roja" (RFP) (Nelson, et al, 2007 Plant J 51(6): 1126-36). Las construcciones se insertaron mediante electroporación en la cepa C58C1 de *Agrobacterium tumefaciens* y las bacterias recombinantes se usaron para transformar de forma transitoria

50

(agroinfiltración) hojas de *Nicotiana benthamiana* como se describe en (Hamilton & Baulcombe, 1999, Science 286(5441): 950-2). Las hojas fueron agroinfiltradas contemporáneamente al constructo que porta el transportador ABC y con el marcador tonoplástico. 48 horas después de la infiltración se observaron las hojas con un microscopio láser confocal (Olympus FV1000). Se utilizaron láseres de 488 nm (Ar) y 635 nm (diodo) para mostrar la fluorescencia de GFP y RFP, respectivamente.

Caracterización funcional de transportadores ABC mediante ensayo *in vitro*.

Para analizar la actividad de los transportadores CsABCC1 y CsABCC2, los genes se clonaron en el vector pNEV (Sauer & Stolz, 1994, Plant J 6(1): 67-77) para la expresión en levaduras. Las construcciones obtenidas se insertaron mediante electroporación en el mutante ybt1 de *S. cerevisiae* (MATa; ura3D::HIS3; leu2-3, 112; his3-D200; bat1D1::URA3). Como control, la cepa se transformó con el vector pNEV vacío. A partir de los transformantes obtenidos, se prepararon fracciones microsomales como se describe en (Tommasini et al. Proc Natl Acad Sci U S A. 93(13):6743-6748). La integridad de las vesículas aisladas se evaluó mediante el análisis de la capacidad de transportar el sustrato leucotrieno C4 (LTC4) (Leier et al., 1994, J Biol Chem 269(45): 27807-10). Los ensayos de transporte de los metabolitos del azafrán se realizaron mediante la técnica de filtración rápida (Tommasini et al. Proc Natl Acad Sci U S A. 93(13):6743-6748) con filtros de acetato de celulosa (tamaño de poros: 0,45 μ m).. El ensayo ideado de "transportómica" prevé, en lugar de estándares analíticos, el uso de extracto hidroalcohólico de azafrán como sustrato. Dicho extracto se preparó pulverizando 3 mg de estigmas secos (procedencia, Zafferanami), resuspiéndolos en 300 μ l de metanol al 50 % frío y homogeneizando durante 40 minutos en Mixer Mill 300 a una frecuencia de 20 Hz y finalmente centrifugando durante 20 minutos a 20000 gramos. El sobrenadante recuperado se caracterizó por HPLC-HRMS y se utilizaron cantidades estandarizadas en los ensayos de transporte (contenido total de crocinas igual a 320 μ M). La reacción de transporte se realizó en un volumen final de 650 μ L utilizando 100 μ L de preparación microsomal (que tiene una densidad óptica a 600 nm de 4 y un contenido de proteína total de 400 μ g), en tampón de transporte (glicerol 0,4 M, KCl 0,1 M, Tris-MES 20 mM, pH 7,4) y luego se añade DTT 1 mM, MgSO₄ 6 mM o 1 mM (en presencia o ausencia de MgATP, respectivamente), 100 μ g/ml de creatina quinasa y 10 mM de fosfato de creatina. El ensayo de transporte se realizó en presencia o ausencia de MgATP 4 mM. El ensayo se realizó a temperatura ambiente y se detuvo después de 30 segundos y después de 15 minutos. Los ensayos de inhibición se realizaron incubando los microsomas a temperatura ambiente en la mezcla de transporte que contenía inhibidor de Probenecid (1 mM) durante 10 minutos y luego agregando el extracto de azafrán. Los ensayos de inhibición en hielo se realizaron incubando en hielo los tubos que contenían la mezcla de reacción, inmediatamente después de la adición de los microsomas. Todas las reacciones de transporte se detuvieron mediante filtración de 100 μ l de mezcla de reacción y lavado de filtros con 2,5 ml de tampón de transporte frío tres veces. Las vesículas recolectadas en los filtros se disolvieron y los metabolitos se extrajeron incubando los filtros en 1 ml de metanol al 75 % y poniéndolos en agitación durante 20 minutos. Los lípidos de las vesículas se eliminaron recogiendo 800 μ l de muestra, agregando 400 μ l de cloroformo, agitando con vórtex y molino mezclador durante 5 minutos a una frecuencia de 20 Hz. Luego se agregaron 200 μ l de agua ultrapura y los pasos se separaron mediante agitación y centrifugación a 20000 g durante 20 minutos. La fase superior se recogió, se secó y se resuspendió en 80 μ l de metanol al 50 % que contenía 0,5 μ g/ml de formononetina y se analizó mediante HPLC-HRMS.

Análisis HPLC-HRMS.

Las muestras se analizaron mediante un espectrómetro de masas Q-exactivo (ThermoFisher Scientific), acoplado a un sistema HPLC provisto de un detector de serie de fotodiodos (Dionex). Los metabolitos polares extraídos con metanol al 75 % se separaron con una columna cromatográfica C18 Luna con fase reversa (150 x 2,1 mm, 3 μ m; Phenomenex) y se ionizaron mediante electrospray (fuente ESI) (Fasano et al. New Phytologist 210: 1382-1394). . Por el contrario, los metabolitos no polares se extrajeron con acetona, se separaron con una columna cromatográfica C30 con fase reversa (100 x 3,0 mm; YMC Europe) y se ionizaron mediante ionización química a presión atmosférica (fuente APCI).

La identificación de metabolitos se realizó basándose en la masa exacta en comparación con los estándares auténticos. Las áreas del pico de masa se normalizaron en el área del pico de masa del estándar interno (formononetina o acetato de α -tocoferol) y las abundancias de metabolitos se mostraron como "veces del estándar interno" (abundancia con respecto al estándar interno).

Se pueden usar métodos bien conocidos por los expertos en la técnica para preparar los vectores de expresión, las células recombinantes y los organismos según la presente invención. Estos métodos incluyen técnicas *in vitro* de ADN recombinante, técnicas de recombinación *in vivo* y técnicas de PCR. Véanse, por ejemplo, las técnicas descritas en Maniatis et al., 1989, MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Laboratory, Nueva York; Ausubel et al., 1989, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene Publishing Associates y Wiley Interscience, Nueva York, y PCR Protocols: A Guide to Methods and Applications (Innis et al., 1990, Academic Press, San Diego, CA).

DESCRIPCIÓN DETALLADA DE LAS SECUENCIAS

CsALDH1 (SEQ ID NO 1)
SEQ. ADN CsALDH1 (1497 pb)

5

ATGGCTGCAACTAACAGCAATGGTATCTTCAAGCTTCCAGAGATCAAGTTCACCAAGCTTTTCAT
 CAACGGAGAATTCCGTCGATTCCGTCCTCCGGGAGAACAFTTGAGACAAGGGACCCACGCAATGGC
 GACGTTATAGCCAACATCGCAGAAGGAGATAAGGAGGACGTGGATTGGCGGTGAAGGCAGCC
 AGAGAAGCCTTCGACCACGGAAAAATGGCTCGAATGTCTGGTTATGAAAGGGGGAGGATCATG
 ATGAAGTTTGAGATTGATTGAAAGCAAACATAGAAGAA TTGGCAGCACTTGACACCTTAGATG
 CAGGAAAACCTGCTCACTATGGGTAAAGGCTGTAGACATCCCAGCAGCAGTACACATGATCCGCTA
 CTACGCTGGAGCTGCAGACAAGATACATGGGGAGACACTAAAGTTGTGCGAGCGAGTTCCAAGG
 ATACACATTAAGGAGCCCATCGGGGTGGTAGGGCACATAGTTCCCTGGAAATTTCCCGACTGCG
 ATGTTCTGTCATGAAAGTGGGCCCTGCCTTGGCTGCAGGCTGCACAATGATCGTCAAGCCAGCCG
 AGCAAAACCCCGCTCAGCTCTACTACGCGCACTTGGCAAAGGAGTCTGGTATCCCGACGG
 AGTAGTCAATGTTGTTACCGGGTATGGCCCAACAGCTGGTGTCTCTTTCTTACACATGGATG
 TTGACAAGATTAGTTTACTGGCTCTACCGAAA TAGGACGGGTAGTGTGAGGCGCAGCAGCAAA
 GAGCAATCTAAAGCACGTATCGCTCGAGTTGGGTGGGAAGTCTCCCTCATTATTTTGTATGATG
 CTAATTTGGACATGGCAGTAACTTGGCAGCATGGCCATCTTCTATAACAAGGGAGAAAGTTG
 CTGTGCCGGAATCTCGTATTTATGTTTCAGGAAGGGATATAATGATGAGTTTGTGAAGAAGGCAGTG
 GAAAAGGCAAAGAGTTGGGTGGTGGAGAGCCCTTTGATCCCAATGTTTCAGAATGGGCCCTCAGG
 TCGACAAGGCACAGTTTGAAAAAGTTCTCAGCTATATTGAGCATGGTAAAAAGAGAGGGAGCCA
 CCTTGCTGGCAGGAGGAAAAGCCTGTGGTTCAGAAAAGGCTACTGCATTGAACCAACAATCTTTAC
 TGATGTCAAGGAGGATATGAAGATAGCACAAAGTGAAGATTTTTGGACCAGTCATGTCTCTCATG
 AAATTTAAGACAATTGAAGAGGCAATAGAGAAAAGCCAACACCACCAGATATGGGTTAGCAGCA
 GGAATTTGTGACGAATGACTTAAATGTGGCCAACAGCGTATCGAGATCAATTCGTGCCGGCACTG
 TTTGGATCAATTTGTTACTATGCTTTTCGATGCAGAAAACCTCCCTTTGGCGGTTACAAGATGAGTGG
 TTTGGGAAAGGATCAGGGACTCCATGCCTGGAGAAGTACCTCCAAGTGAAGTCTGTGGTCACTC
 CTATATATAACTCACCTTGGCTGTGA

SEQ. AA CsALDH1 (498 aa) (SEQ ID NO 2)

MAA TNSNGIFKLPEIKFTKLFINGEFVDSVSGRIFETRDPRNGDVIANIAEGDKEDVDLA VKAAREAF
 DHGKWPRMSGYERGRIMMKFADLIEANIEELAALDITLDAGKLL TMGKA VDIPAA VHMIRYYAGAA
 DKIHGETLKLSEFQGYTLKEPIGVVGHIPVWNPFTAMFVMK VGPALAAAGCTMIVKPAEQIPLSALY
 YAHLAKESGIPDGVVNVVVTGYGPTAGAALSSHMDVDKISFTGSTEIGRVVMEAAA KSNLKHVSLLEL
 GSKSPLIIFDDANLDMAVNLSMAIFYNKGEVCCAGSRIYVQEGHYDEFVKKA VEKAKSWVVGDPF
 DPNVQNGPQVDKAQFEK VLSYIEHGKREGATLLAGGKACGQKGYCIEPTIFTDVKEDMKIAQDEIFG
 PVMSLMKFKTIEEAIEKANTTRYGLAAGIVTNDLNVANSVRSIRAGTVWENCYAFADETPFGGYK
 MSGFGKDQGLHALEKYLQVKS VVTPPIYNPWL

10

CsALDH3 (SEQ ID NO 3)
SEQ. ADN CsALDH3 (1554 pb)

ATGTCGATGCTACGCGTGGCTCTCTGTCGTCACCGTCTGCTCGGTTCTTCTCGTCACTTCCGCCGA
 CAGATGAGTATGGACTCGAAGGGAGCAGTTTCTCAACTGAGCAATTCAGGGATTTTCAGGACTC
 AGGGGTTCAATTCAGGGAAAATGGCTTGATTCTTATGATGGGACAACCTTTGAGGTGAGAAAACCC
 TGCTACTGGGGATGTTATAGCAAAAGTGCCATCCATGGGGAGGAAGGAAACTTCAGATGCAAT
 ATCTTCGGCGTACACTACTTTGAGTTCCTGGAGCAAACTCACTGCAAAATGAAAAGGAGCAAAATGC
 CTCAGAAGATGGTACGATCTAATTACTCACAAGGAAGATCTTGCCTACTGATFACATTTGG
 AACAAGGGAAAGCCTTAACGGAGGCTCAGGGTGAGGTCAGCTATGGTGCCAGCTCATTTGAATT
 CTTTGCTGAGGAAGCAAAACGIGTGTATGGCGATATTATTCCTCCTCCTATAGCAGATCGACGA
 CTATTTGTTTTGAAACAGCCAATTTGGAGTGGTTGCAGCAATTACCCCATGGAACCTTCTCTGGC
 CATGATCGCACGAAAGGTTGGCCCTGCTCTGGCTTGTGGCTGCACAGTGGTTGTTAAGCCTTCAG
 AATTTACTCCTTTGACTGCCTGGCTGCAGCTGAGCTCGCTCTTCAAGCTGGAATCCCACCAGGT
 GCACTGAACATGGTTATGGGGAAAGGCTTCTGATTTGGAGACACTTTATAGAGAGCAGCACAGG
 TAAGGAAGCTCACTTTCACAGGTTTCGACAGCTGTAGGGAAGAAGTTAATGGCAGGATCAGCTAG
 TACAGTTAAAAAGGTA TCCCTTGAACCTGGTGGGAATGCACCTTGCATAGCTTTGATGCA
 AACTTGGATGTAGCAGTGAAAGGCTGTCTTGTGCTGCGAAGTTTCGAAAACAGTGGGCAACATGCG
 TCTGCGCAACAGAAATCCTGGTACAAGAAGGTATCTATGAAAAATTTGCTAATGCATTTGCAAA
 GGCTGTTCAAAAATCTGCAAGTCGGCAATGGCCTTGGCTGAAGGCATAACTCAGGGTCCACTAATT

15

AATGAAGCTGCCGTACAAAAGGTTGAATCGTTTTGTCAAAGATGCAACTTCAAAGGGAGCCAATG
TTATACTTGGTGGAAAGGAGGCACAGCCTCGGCATGACATTTTATGAGCCTACCGTACTCAGTAA
TGTC AAGAGTGATATGCTCATATCAAGGGAGGAAAATTTGGTCCAGTTGCACCGCTGATGCCT
TTTAAATCTGAAGAAGAAGTTATTCAGATGGCTAATGACACAAAATGCTGGTCTAGCCGCTTACA
TATTTACAAACAGCATACTTCGGTCTTGGCGTGATCTGAAGCTCTTGAAATATGGTCTAGTTGGT
GTGAACGAAGGCATAATTTCTACAGAGGTGGCACCATTTGGAGGTTGTAACAGTCTGGCCTTG
GACGAGAAGGTTCCAAATA TGGGATGGATGATTA CTTAGAACTCAAATATGTGTGCATGGGGAA
CTTGGAGTCTTGA

SEQ. AA CsALDH3 (517 aa) (SEQ ID NO 4)

MSMLRVALCRHRLGSSRHFRRQMSMDSKGA VSQLSNSGIFRTQGFIAKWLDSYDGTTFEVRNPA
TGDVIAKVPSMGRKETS DAISSAYTTLSSWSKLTANERSKCLRRWYDLIITHKEDLALLITLEQKPL
TEAQGEVSYGASFIFFAEAKRVYGDHPPPIADRRFLV LKQPIGVVAAITPWNEPLAMIARKVGPAL
ACGCTVVVKPSEFTPLIALAAEELALQAGIPPGALNMVMGKASDIDGDTLLESTQVRKLTFTGSTAAG
KKLMAGSASTVKVSVLELGGNAPCIVFDDANLDVA VKGCLA AKFRNSGQTCVCANRILVQEGIYEK
FANAFAKAVQNLQVGNGLAEGITQGPLINEAA VQKVESFVKDATSKGANVILGGRRHSLGMTFYEP
TVLSNVKSDMLISREEIFGPVAPLMPFKSEEBVIQMANDTNAGLAA YIFINSISRSWRVSEALEYGLV
GVNEGHISTEVAPFGGCKQSGLGREGSKYGMDDYLELKYVCMGNLES

5

CsALDH4 (SEQ ID NO 5)

SEQ. ADN CsALDH4 1524 pb)

ATGGGATTTACCAAGGAGCACCAGTTCCTCTCCGAGCTCGGGCTCGGCCCGAGAAATCCCGGTT
GCTACGTCGCCGGAAAGTGGAGGGGGAGCGGACCGGTAAGTCTCCTCGTCGAATCCCGCGAACA
ATCAGGTTATAGCTGAGGTGGTGGAGGCATCAATGGAAGACTACGAGGATGGCATGAAGGCAT
GTTTTGATGCTTCCAAAATA TGGATGCAGATTCCAGCTCCGAAGAGGGGAGAGATTGTTAGACA
GATTGGCGAAGCA TTGAGAAGCAA ACTCCAGCACCTTGGTAGGCTTGTCTACTAGAAAATGGGA
AAAATCTTCCCGAAGGGATTGGTGAAGTTT CAGGAAA TAGTCGACATGTGTGATTATGCTGTTG
GGTTAAGTCGGCAAT TGAATGGATCAATCATA CCGTCAGAACGTC CGAACCCACATGATGATGGA
GGTGTGGAA TCCCTTGGAA TTGTAGGTGTTATTACAGCGTTTAACTTCCCGTGTGCTGTGCTTG
GTTGGAATGCCTGCA TTGCATTAGTATGTGGAAACTGTGTTGTTTGGAAAGGTGCACCAACAAC
ACCATTAATCACAATTGCTATGACTGA ACTAGTTGCTGGGGTGTGGAGAAGAATAATTTACCA
GGTGCAATTTTACGTCATTTTGTGGAGGTGCTGAAATTTGGCCAAGCAATAAGTCATGACACAC
GGATACCTCTGGTCTCATTTACTGGAAGTTCTAAGGTTGGTCTTATGGTGCAACAAAACAGTTAGT
GAGAGATTTGGCAAATGTTTGTCTGAACTTAGTGGAAATAATGCTATAA TTGTCATGGATGATG
CTGATAATTCAACTTGTGACGTTCTGTTCTATTTGCTGCTGTTGGTACAGCTGGCCAGCGATGC
ACAACATGCCGAAGGCTGCTGGTTACGAAAAGCATA TATCAA ACTGTACTTGTATCAATTTGGTTG
GAGTCTATAAAACAAGTGCAAATTTGGAGATCCACTAGAGAAAAGGCACCTTATTGGGACCGTTGCA
CACTTCTACTTCAAAGGAGAACTTCGTGAAAAGGTGTCAGGCCATTAATCCAGGGAGGGGAAG
ATCCTCGTTGGAGGTTCTGTGATAGAA TCTGCCGGGA ACTTTGTACAACCTACAATTTGTCGAGAT
CTCTTCAGATGTACCAATTTGCAAAGAGGAGTTGTTTGGTCCGGTTCTTTATGTGATGAAAATTC
AGACTTTAAAGGAAGCAATTTGAAATAAAACA ACTCTGTACCACAAGGTTTAAGCAGCTCTATAAT
CACACGCAAACCGAAAATTATA TTTCAAGTGGCTGGGACCTCATGGCAGTGAATTGTGGAATTTGC
AATGTTAAACA TACCTACTAATGGTGTGCTGAAA TTTGGCGGTGCGTTTGGCGGTGAAAAGGCTACTG
GAGGTGGCCGAGAAGCAGGGAGTGATTCCTGGAAGCAATACATGAGGCGTTCAACTTGTACAA
TCAACTATGGGAGTGAAC TCCCTCTGGCACAGGGAAATAAAC TTTGGCTAG

10

SEQ. AA CsALDH4 (507 aa) (SEQ ID NO 6)

MGFTKEHQFLSELGLGPRNPGCYVAGKWRGSGPVVSSSNPANNQVIAEVVEASMEDYE DGMKACF
DASKIWMQIPAPKRGEIVRQIGEALRSKLQHLGRLVSLEMGKILPEGIGE VQEIVDMCDYA VGLSRQL
NGSIIPSERPNHMMMEVWNPLGIVGVITAFNFPCA VLGNACIALVCGNCV VVKGAPTTPLITIAMT
ELVAGVLEKNNLPGAIFTSFCGGA EIGQAISHDTRIPLVSFTGSSK VGLMVQQT VSERFGKCLLELSG
NNAIIVMDDADIQLA VRSVLF AAVGTAGQRCTT CRRLI VHSIYQTVLDQL VG VYKQVQIGDPLEK
GTLGPLHTSTSKENFVKGVQA IKSQGGKILVGGSVIESAGNFVQPTIVEISSDVPIVKEELFGPVLVY
MKFQTLKEAIEINNSVPQGLSSSIFTRKPEIIFKWLGP HGSDCGIVNVNIP TNGAEIGGAFGG EKATGG
GREAGSDSWKQYMRRTC TINYGSELPLAQGINFG

15

CsALDH7 (SEQ ID NO 7)
SEQ. ADN CsALDH7 (1449 pb)

ATGGCC TTCGATGGAGAGAAAAGCTAAGGAGATGGTGAAGGAGCTCAGAGAGAGCTTCAACAAA
GGGACGACCAGGAGCTACGAGTGGAGGATGAAGCAGCTGAAGGCCATGGAGAAGATGACTGA
GGAGAAGGAGAAGGATATCATGGATGCCCTCGAATCTGATCTCTCTAAGCCTCAGCTTGAGTCT
TTCC TCCATGAGATTTCAATGGCAAAAATCAGTATGTCAGTTTGC TGC TAAGAAATTTGAAACGATG
GATGAAACCAGAAAAGGTGCCAGCTCAACTTACAAC TTTTCCA TCAGTAGGAAA TATTGTTGCA

5

GAGCCCTTTGGTGTGTGCCTGATCATCTCAGCA TGGAAATTTCCCTTTTGT TATCTCTCGAACCA
GTCATAGGAGCTATTGCAGCTGGTAA TACTGTTGTTCTAAAGCCTTCGGAAAATTCGACCTGCCAC
ATCATATTGTTTGC AAGAA TTTTACTTGAAGTATGTCGATACCTCATGCGTAAGAGTTGTCGAGG
GGCGTGTTCCTGAAACAAC TGC ACTTCTAGAGCAGAAAGTGGGATAAAAATTTCTATACAGGTAA
TGGAAAAGTAGGGCGCTCGTGTATGGCTGCAGCTGCAAAAGCACCTAACACCAATGTC TTGGAA
CTTGGAGGAAAATGTC CAGTTGTTGTTGATTCAAATATCGATTTAAAAGTTGCGACCAAAAAGAG
TTGTTGTTGGGAAAGTGGGGA TGCAACAATGGACAAGCTTGCA TAGCTCCAGATTATATCATAAC
AACAAAATCA TTTGCCCAAAGCTGGTGGAGTCTCTGAAGATCAC TTTAGAAAAGGTTCTACGGC
GAGGATCCAT TGGAAAACGGAAGATCTGTCCTGTA TTTGAAAATGAAAATCATGTTGCGCGTTTGG
CAAGGCTCTTGGATGATGATATGGTTCTTGGCAAGATTA TCTATGGTGGAAAACGAGATGAGAA
CGGACTAAAGATAGCCCCAACACTGTTGTTAGATGTTCCCTGACGATTCACTGATTA TGAAGGAA
GAGATATTTGGTTCCTTGTCCAAATTA TACGGTTGACAAGATTGAAGATAGTTTTGCTGTGAT
AAACTCGAAAACGAAGCCTCTTGTGCTTACCTCTTACC AAGAACAAGAACCTAGAGAGGATG
TTCGTGGAAAACAGTATCTTCCGGAGGCATGCTAATTAACGACACGGTCTTACATGTTGCTAATCC
TTATTTGCCATTTGGTGGAGTTGGT GAGAGCGGA ACTGGATCCTACCATGGTAAA TCTCGTTCA
ATGCGTT CAGTCA TAAGAAAGCTGTTTTATCCCGTGGATTTGGAGGCGAAGT A GGGGCAAGGTA
TCCTCTTATACAGATAAGAAAGCGAAA GATTA TAAAGACATTGCTTGC GGGTAATATTA TAGCC
CTGGTTTTGGCTTCTTTGGCTTCTCCAAATCTTGA

SEQ. AA CsALDH7 (482 aa) (SEQ ID NO 8)

MAFDGEKAKEMVKELRESFNKGTTRS YEWRMKQLKAMEKMTEEKEKDIMDALES DLSK PQLESFL
HEISMAKSVQCFAAKNLKRWMKPEKVP AQLTTFPSVGNIVAEPFGVVLIIISAWNFPFLLSLEPVIGAI
AAGNTVVLKPSIAPATSSLFARILLE YVDTSCVRVVEGA VPETTALLEEQKWDKIFYTGNKGVGRVV
MAAAAKHLTPVLELGGKCPVVVDSNIDLK VAKRVVVGK WGCNNGQACIAPDYIITKSFAPKLV
ESLKITLERFYGEDPLETEDLSRIVNENHVARLARLLDDDMVSGKIIYGGKRDEKRLKIAPTLLLDVP
DDSLIMKEEIFGPLLPIITVDKIEDSFA VINSKTKPLAA YLFTKNKNLERMVETVSSGGLINDTVLH
VANPYLPFGGVGESGTSYHGKFSFNAFSHKKA VLSRFGGGEV GARYPPYTDKKRKIRALLAGNII
ALVLAFFGFSSK

10

CsUGT2 SEQ ID NO 9)
SEQ. ADN CsUGT2 (1380 pb)

ATGGAGCAGAAAGATGTGAACGGCAACAAATGCCACATCCTCCTCCTCCCCTGTCCAGCACAAG
GCCATATCAACCCCATTC TCCAATTCGGCAAGCGACTGGCCTCTCACAACCTCCTCACCACACTC
GTCAACACCCCGTTCCTCTCCA ACTCGACCAAATCCGAACCCCGTCCGGTCAACATCGAGTGCAT
TATCCGACGGATTCGATTCGGGTGGGATGAATGCAGCACCTAGTCGTCGGGCTATTTTCGACCG
ACTCGAGTCGGTCCGGT CAGAAAACACTGTTTGGCCTCATCGAGTCACTCCGGTCCAGAGGCCGA
CCGGCGCATGTTTTGGTCTACGACCCGTTCTTGGCTTGGGCAA TGAACGTAGCCGAGCGGTCCG
GGCTGCGGTCGGTTCGATTCTTACTCAGCCGTTGTCAGTGGACACAA TCTACCCACCGTGTG
GGAGGGACGAATTAAGGTTCCGGTTACCGAGCCGGTCCGGTTGCCTGGATTGCCCCCACTCGAG
CCTAGTGATCTGCCGTCGTTTCGTAACGGA TTCGGACCCAGTCGTTAACCCGGATCTCTTGCCACT
TCTAGTGAATCAGCACAAAATTTGGACAAAGCTGATA TGA TGTGATTAACTCCATA TACGAG
CTAGAGCATGAGGAATTTGATTTGGATGGAGTCCCGATTAACCA TTAACCGTGAAGTCGATTGGGC
CGACCGTGCCGTCTACTTATCTCGACAACCGAATCCCGTCCGATTCACTATGGATTCAATCTA
TACACCCCTGACACTACTCCCTACTTGGACTGGCTCGACTCCAAGGCCCGA ACTCTGTCATCTA
TGTCTCATTGGTAGTTTGTCCAGTCTGAGCCCTGATCAAACCAATGAGATTGCGTCCGGTCTGA
TCGCCACCAACAAAAGCTTTATTTGGGTGGTACGCACTTCGGAGTTAGCCAAACTCCCTGCCAA
CTTCACCCAAGAGAAATGCGAGCCGTGGGCTGGTTGTGACCTGGTGC GACCAACTCGATGTCTC
GCACACGTTGCGACCGGTTGCTTCGTGACTCACTGCGGATGGA ACTCGACTATGGAAGGTATCG
CACTTGGTGTGCCGATGGTGGGAGTGCCGCA GTGGTCCGACCAGCCTATGAATGCCAAGTATGT
AGAGGATGTGTGGAAGGTCGGCGTGAGGGCGAAGACTTATGGGAAGGATTTTGTGAGAGGGGA
GGAGTTCAA AAGATGTGTGGAGGAGGTGATGGATGGAGAGAGGAGTGGAAAGATTAGAGAGA
ATGCCGAAAGTGGTGTAAAGTTGGCCAAGGACTCTGTTAGTGAGGGTGGGAGTCTGACAAGTGA
TATCAAGGAGTTCATACATCAGTGTTCGAAAATGA

15

SEQ. AA CsUGT2 (459 aa)(SEQ ID NO 10)

MEQKDVNGNKCCHILLPCPAQGHINPILQFGKRLASHNLLTTLVNTREFLSNSTKSEPGPV
 NIECISDGFDSGGMNAAPSRRA YFDRLLES VGSETLFLGLIESLRSRGRPAHVL VYDPFLPW
 AMNVAERSGLRSVAFFTPQCAVDIYRHVWEGRIKVPVTEPVRLPGLPPEPSDLPSFVT
 DSDPVVNPDLPLLVNQHKNLDKADMMLINSIYELEHEEFDWMESRLPLPVKSIGPTVPS
 TYLDNRIPSDSHYGFNL YTPDTPYLDWLDSKAPNSVIYVSFGLSSLSPDQTNEIASGL
 LATNKSEIWVVRTSELAKLPANFTQENASRGLVVTWCDQLDVLAHVATGCFVTHCGWNST
 MEGIALGVPMVGVPQWSDQPMNAKYVEDVWKVGVRAKTYGKDFVRGEEFKRCVEEVMJGE
 RSGKIRENAARWCKLAKDSVSEGGSSDKCIKEFIHQCK

5 **CsABCC1 (SEQ ID NO 11)**
SEQ. ADN CsABCC1 (4491 pb)

ATGTCCTCCTCTCCCTACGATCTCACCTCCCTCACTTGCTCACCTCCCCCACTTCCGACCCAA TC
 CAATGGCTCCGCTTCGTCGCCCTCTCCCCCTGTCCTCAGCGTCTCCTCCTCTCCGCCGTGATATT
 CTCTTCTCCTCTCCCTCCTCTCACCTCCCTAAGAAGACTAATCTTCTTCCAAACAACAACAAAT
 AATGAGGCCAGAAAAGCCTCTCCTATACCCCGGATGAAGATGACAATGTTCCGTCAGCAAAAAGTCA
 ATACCAATCTACTCTTCAAGCTCTCCCTCGCTCTCTCCGCCGCCCTTGCCGCCGCCTACCTCGTCC
 TATTCTGCTGCTGCTTTGGTTGCAATA TCCCCACCGACTGGCGACTCGTCGAGTGGCCCTTCCATA
 TTCTTACAATTCCTTACACAATCGTCCGTCGCGATTGTCGTCGGTCACGAGAAGAAAGTTCCCGCG
 CGAGAAGCACCCGCCGGCTCTCCCGCCTTCTGGGTCGTCACCTTCGTCGTCGCTCTCCCTCTTCT
 CCGTCTCCTCCGTCGCTAGGTTCTGTCGTAAGGGTTACGTCCTCCCGACGATCCCGTCACCGTC
 GCCGTCCGTCGTCGCGAGCCTGCTCTTCTGTA TAACCGCCGTAACCGGCCGCGACCGGCATATCGA
 CGGGGACGGACCGGCCGGTAATTGCGACGCCGTCGGCGGAGAAAGGTGACCGGATACGCGACCG
 CATCGCTGCTCTCGCGGCCACGTTGGCACTGGATGAACCCGCTCATGGCGAAGGGATACCGGGA
 GCCGATACGATCGGCGACGTGCCGTCGCTCTCCCCGAGCACCGGGCGGAGCGGTTGCTGGAG
 ACGTTCCGATCCAAC TGGCCCCGAACCCGCAACCCGGTCCAACCCCGGTCCGACCCCGCTCGCTGC
 TCAAGTCTCTGCGCCCCACCTCCCTCTCACCGCCGCTCCTCTCCAATCCCTCAAGCTCACCGTCATGT
 ACGTGGGGCCCCACCTCATCCAGGGCTTCGTCGACTTCACCTCCGGGGACCGCGCTCCCCGTA
 CGAGGGATTCCGCCTCGTGTTCATTTTGC TGGTGGCGAAGTTCTTCGAGGTGCTGTGCGCGCACC
 AGTACAAC TCCAGTCGCGAAGCTCGGGATGATGATAAAGGTCTGCTGATCACGTCGCTGTA
 CCGGAAGGGCCTGCGGCTATCGTGGCGCC CAGGCAGGCGCA CGGGGTGGGGATGATCGTCAA
 CTACATGGCTGTCGACGCGCAACGTA GCCGACATGATGTTGCAAGTGCACACTACATATGGCTC
 ATGCCAGTGCAGTTAGCGTCCGCAATTCGCTGCTGTATACTTATCTGGGTCCGTCGGTTACGTC
 GGCAGTCCGTCGGGATCGCTGCGGTGCTCGTGTTCGTCGTCGTCGGGACAAAAGAGGAA TAA TAGG
 TATCAGTTTAA TTTGATGACTATGAGGGATAAGAGAA TGAAGGCAACTAATGAGATGCTTAACT
 ATATGAGGGTGATCAAGTTTCAGGCGTGGGAGGAACATTTTGATGAGAGGATCAAGAAGTTCA
 GACAAATGGAGTTTGGCTGGCTATCTAGGTTCA TACTCCATCTCCGGCAACATCATCGCCCTA
 TGGAGCGCGCCGGTGTCTCTCTCCCTCGTGTTCGGCACGTGCGTATTCGCTGGGGTCCACCT
 CGAGCCGGCCTCGTCTTCCACGCCACCTCATTCTTCAAGATCCTCCAGGAACCCCTCCGGAAC T
 TCCCCAAGCAATGATCTCCGCA TCCCAAGCGATGATCTCGCTCGACCGGCTCGACTCCTACATG
 ACCAGCGGAGAGCTCGACGAGGCGCGGTGCAGAGAA TGGACGGCTGCGATGACGACGTGGCG
 ATCGAGTTGAAGGACGGGGTTTTCGGATGGGAGGACAGGGAGGACCGGAGAAAGGAGAAGGA
 GAAGGCGTGGCTGAGGGACATAAACATGGAGGTGAAGAGGGGTTTCGCTCGCGCGCGGTGGTGGG
 GACCGTCCGGCTCGGGGAAGTCGTCGCTCATCTCGTGTCTGCTCGGCGAGATGCAGAAAGATCTCG
 GGAACGGTTAGGGTTTGTGGCACTACTGCCTATGTTGCACAGACTGCA TGGATTCAGAA TGGGA
 CTATTGAAGATAACATCCTATTTGGACTGCCGATGGACCGGGAGAAA TACAAGGAAGTAA TCCG
 TGTTTGTTGCTTACAGAAAGGACATGGAAATGATGGAGTTTGGGGACCAGACTGAGATTGGTGAG
 AGAGGAATAAATCTCAGCGGTGGTCAA AAGCAAAGGATACAGCTGGCTAGGGCTGTGTATCAA
 GATTGTGATA TCTACCTGCTCGATGATA TCTTCAGTGC TGTGATGCTCACACTGGTTCTGAAAT
 CTTCAAGGAATGTGTGAGGGGTGCACTCAGCAAGAAGACGGTCA TACTTGTGACTCATCAAGTG
 GATTICTTGCATAATGCTGATCTCA TCC TGGTAA TGAAGAAAGGGACGATCGTACAA TACGGAA
 AATACAATGATTTGCTGGGTTCTGGAACAGATTTTGCAGCTCTAGTCGCTGCTCACGACGTTCA
 ATGGAGTTAGTCGAACAGGTTCTCACAGCAA TGAATCCAATAACCCACAGGA TAAAGCATCG
 GATCAGCTTACAGGAAC TACGGCAGTCAACGGGGCAAATGGGTCGATGAACTCGCCCAAG
 AAGGAAAAGGGATCGTCAAAGCTCATAAAGGAGGAGGAGAGGGAGACCGGTCA TGTACGCTG
 GAA TGTTTACAAGAGGTACATCACCGAGGCTTGGGGTTTGTGGGGGCCGCTCACGGTTACCGCT
 ATTGCA TTTGTGTGGCAGGGTTCCGTAATGTCAAGTGATTACTGGTTGGCA TATGAGACATCCGA
 AGATAACGCTTCTCTTTCCGTCCTTCGCTATTTATTGAAGTTTATTTCTGGCA TTTGCCGTTATTT
 GATCTTTCTTGTGGCGCAAGGTCCTTTCTTA TAGCGIATTTGGGGCTGAAGACTGCTCAGATTT
 TTTCAAAGCAGATTTTGAACAGTCTCTTGCATGCTCCGATGTCGTTCTTTGACACGACACCTTCA
 GGAAGAATCTCAGTCGGGCATCATCAGATCAAAACCAACATCGATCTGTTCTTCCCTCTTTGT
 GGGGATGACTGTGTCAATGTACAT TACAGTGTCTTAGIATCATCTTTATTACGTGTACAGATGCA T

GGCCAACAATCATAGCCATTTTCCCCTAGGATGGCTGAACATTTGGTACAGGAATTACTATCTC
 GCAACATCTCGCGAGCTAACTCGAATCGACTCGATCAGAAAAGCACCCGTCATTCATCACTTTTC
 AGAGACTATTC TTGGTGTCA TGACAATAAGGTGCTTCAGAAAAAGCAAAGGTTTTCTGAGGAG
 AATTTAAAAAGAGTCAATACAAACTTAACCAATGATTTTCACAACAACGGTGC AAAATGAGTGGT
 TGGGTTTTCGITTAGAACTGATTGGTTTCACTTGTCTTGTGCATATCTGCTCTTCTCATGGTCAATGC
 TGCC TAGCAATTCATTA AACAGA ACTTGTGGTTTATCTCTTTCTTATGGTCTCTCGCTGAATT
 CTGTACTATTTTGGGCGGTA TGGA TCA GTTGCCTTCA TAGAGAATAAGATGGTGTAGTTGAAAG
 GATAAAGCAGTTACCAACATTCCTCCGAAGCAGCATGGAGGATTGAAGGCTGCC TTCTACA
 CAAA TTGGCCCA TAGAGGAGATATTACTCTGAACGATCTGAAAGGTTAGATATCGACCAATA
 CCCCTCTAGTTCTGAATGGCA TTACATTGACCATTAAGGGGGCGAGAAGGTTGGAGTAGTGGG
 TCGAACTGGAAGCGGCAAA TCGACATTGATACAAAGTTCTATTCAGAAATTGTAGAGCCTTCTGAG
 GGGAAAATAATCATCGATGGAGTAGACATTTGTACACTGGGCCTTACGATCTCCGGTACCGTT
 TTGGCA TTATTCCTCAAGA ACCGTTCTCTTTGAAGGAACTGTGAGAAGCAACATCGATCCAATT
 GGACAATATTCAGATGAAGATATATGGCAGAGCCTGGATCGTTGCCAACTCAAAGATGCCGTGG
 CTTCAAAACCTGAAAAACTCGATGCTTTGGTGGTTGATAACGGAGAGA ACTGGAGTGTGGCCA
 GAGGCAGCTTTTGTGTTTAGGCCGTGTTATGCTGAAGCGCAGTCGAATTCTATTTATGGACGAAG
 CGACTGCTTCCGTGGATTACAGACCGATGCTGTGATTCAAAA GATCATTTCGAGAAGACTTCTC
 GGCTTGTACTATTATCAGCATTGCTCATAGAA TACCACAGTCATGGATTGTGATAGAGTGTGG
 TTATCGATGCAGGACTGGCAAAAGAGTTTGCAGGCCAGCCA ACTTGATAGATCGGCCATCTCT
 CTTTGGCGCATTTGGTTCAAGGAGTATGCGAACCGCTCTTCTGATCTTTGA

5 **SEQ. AA CsABCC1 (1496 aa)(SEQ ID NO 12)**

MSSSPYDLTSLTCSPPPTSDPIQWLRVVALSPCPQRLLLSA VDILFLLSLFLTLRRLIFLPNNNNNEAR
 KPLLSPDEDDNVP SAKVNTNLLFKLSLALSAAALAAA YL VLFVVAL VRISPPTWRL VESAFLLQFLTQ
 SSVAVIVGHEKFRAEKHPPALRAFVWVTFVVVSLFVS VSSVAREF VSKG YVSPDDPVTVA VL VASLLF
 LITAVTGA TGISTGTD RPVIA TPSAEKVTGYA TASLLSRA TWHWMNPLMAKGYREPITIGDVPSL.SPE
 HRAERLLETFRSNWPEPA TRSNHPVRTALLKSFWPHLLLTAVL.SILKLTVMYVGP TLIQGFVDF TSGD
 RASPYEGFRL VFILL VAKFFEVLCAHQYNFQSQKLGMMIRSLITSL YRKGLRL.SCAARQA HGVGMI
 VNYMA VDAQQLADMMMLQVHYIWLMPVQVSVAIALLYTYLGPSVTSAVAGIAA VL VFVVLGTRKN
 NRYQFNLM
 TMRDKRMKATNEMLN YMRVIKFAWEEHF DERIKKFRQMEFGWLSRFMYSISGNIAL WSAPVLVS
 SLVFGTCVFA GVHLDAGL VFTATSFKILQEPLRNFQAMISASQAMISLDRLDYSYMTSGELDEAA V
 QRMDGCD DDVAIELKDG VFGWEDREDAEKEKEKA WLRDINMEVKRGLAA VVGTVGSGKSSLISC
 LLGEMQKISGTVRVC GTTAYVAQTAWIQNGTIEDNILFGL.PMDREKYKEVIRVCCI.QKDMEMMEFG
 DQTEIGERGINLSGGQKQRIQLARA VYQDCDIYLLDDIFSA VDAHTGSEIFKECVRGALS KKTIVLVT
 HQVDFLHNADLIL VMREGTIVQY GKYNDLLGSGTDFAAAL VAAHDSSMELVEQGS HSNESNNPQDK
 ASDQPIRNYGQSN GANGSMNSPKKEKGSSKLIKEERETGHVSWNVYKRYITEAWGLWGPLTVTAI
 AFVWQGS LMSDYWLA YETSEDNASSFRPSLFIEVYSGIA VISIFLVAARSFLIAYLGLKTAQIFFKQIL
 NSLLHAPMSFFDTTPSGRILSRASSDQTNIDLFLPFVGMTVSMYITVL.SHFITCQSAWPTHAIPLGW
 LNIWYRNYLA.TSRELTRIDSITKAPVIHIFSETILG VMTIRCERKEQRFEENLKRVTNLTNDFHNN
 GANEWLGFRI.ELIGSVL.CISALLMVML.PSNFIKPEL VGL.SLSYGL.SLNSVLFWA VWISCFIENKMVL
 VERIKQF.TNIPSEAA WRIEGCLTPPNWPNRGDITLNDLK VRYRPNTPLVLNHITLTIKGGEKVGVVGR
 TGS GKSTLIQVLFRI VEPSEGKIIDGVDICTLGLHDLRSRFGHIPQEPVLFEGTVRSNIDPIGQYSDEDIW
 QSLDRCQLKDA VASKPEKLDALVVDNGENWSVQQRQLLCLGRVMLKRSRILFMDEA TASVDSQTD
 AVIQKIREDFSACTHSIAHRIPTVMDCDRVLVIDAGLAKEFDSPANLIDRPSLFGALVQEYANRSSDL

10 **CsABCC2 (SEQ ID NO 13)
 SEQ. ADN CsABCC2 (4881 pb)**

ATGGGTTTCAAACCGCTTGGTTGGTACTGCCGGCCCGTCAAGGATGGATTGTGGTCAATGACGG
 TGCAATAATGCTCTGGGTGCC TACTCCCTGCACCATGGAGACCA TTGTGGTTGGTATCTCCCAT
 CTGGCTCTTTTTGGCACGTGCTTCTATCGTATATGGCGGATGAGTAGGGATGTA ACTGTGAAGAG
 A TACTGTTT GAGGTCTAAGTACTACCATATCC TGC TCGGACTGCTTGTCTGCTTTCTGC ACTTTGG
 AGCCTCTCTTCAGGCTGGTGTTCGGTATATCAATCATGGATA TCGATCATCACACTAGCCTTCCA
 CCATTTGAGATCCTGTCACTGATTGTTGAGGCTGTTGCTTGGTGTGTTGATGCTTGTAAATGCTTGG
 AATGGAAAGTAAGGTTTATATCCATGAATTCGGTGGTACGTCGGTTCGGTTFGTAGTCTCTATGTCT
 TGGTTGGGGAGACTGCTGTTTTTAATCTGTTCTCTCTGTTAGAGAGTACTATGAGGAATCCGTC
 TTTTATCTTTACACCAGCGTGATAGTCTGCCAGTTTTTGTGGATTCTCTCTGCTAAATTCATATG
 CCAACGCTCGATCCGTTCCCTGGATATACTCCCATTAGAGACCAAGTGTGTGCGATAACACAG
 AATACGAGTCCCTTGCTGGAGAAGAGCA TATTTGCTCTGAGAGGCATGTCAATTTATTTGTTCAA
 AATACTCTTTTCTGGATGACTCCCTCATGCAACAAGGCTATAGAAAACCTATCACTGAGAGG
 GATGTGTGGAAATTAGATTCATGGGATACTACAGAGACGTTGTATAGCAACTTCCAGAAA TGCT

GGCTTGAGGAATCCAAGAAACCCAAACCTTGGTTGTTGAGAGCACTACATAGTAGCCTTGGGGG
AAGATTTTGGCTTGGAGGCTTTTCAAGATTGGCAATGATCTCTCTCAGTTTGTGGGTCCCACGA
TTTTGAATCTCCTCTTACAATCTATGCAAGAAGGGGATCCATCCTGGAATGGCTACATCTACGCA
TTCGCAATTTTTGCTGGAGTGGTATTTGGCGTGTCTTCTGAAGCACAATACTTTCAGAATGTTAT
GCGTGTCCGGCTTTCGTTTGGAGTCTACATTTGGTTGCTGCAGTATTTTCGCAAACTACTAAGATTA
CACATGAAGGTCGGAGAAAAATTTGCATCAGGGAAGATAACTAACTTGTATGACAACCTGACGCAG
AATCACTGCAGCAAAATATGCCAACAGCTTTCACAGTTTATGGTTCAGCTCCTTTCAGAAATCAITTTG
TCTATTAATCTTCTCTACCAGGAGCTTGGTGTGCTTCACTAGCAGGCGCACTGATGCTTGTCTC
ATGTTTCCAATACAAACATTTGGTGTATAAGCAAAATGCAGAAA TTA ACTAAGGAGGGTATTCAAC
GCACTGACAAGAGAATTGGTCTAATGAATGAGTATTGGCTGCTATGGACACTGTGAAATGCTA
TGCTTGGGAGCAGAGCTTCCAGTCTAAAGTGCAAAAGTATTCGCGATGATGAACTTCTTGGTTCC
GCAAAATCCCAGCTTTTAGGATCGTTGAA TTTCTTTA TGTGAA TAGCATTCCAGTTGTGGTTCAG
GTGGTTCTTTTGGAGTTTATTCCTTACTTGGAGGGAACCTTGACACCTGCAAAGGCATTACATC
ACTTTCTGTTTGCAGTACTGCAGTTTCTCTGTTCTCATGCTTCCAAATTTGATAAATCAGGTTGT
GAACACAAAAGTCTCACTAGATAGGTTAGAGGGTCTA TTTTGGCCGAGGAGAGAATACTGTG
CCAAATCCACCTCTTGACCCAACCTTCCAGCCA TCTCAA TTAAGGATGGATAC TTTCTCGTGGGA
TCGGAAGGAGGAGAGGCCACGTTATCAAATATAACTTGGATATACCTGTTGGAAGCTTGGTT
GCCATTGTTGGAAGCACTGGAGAAAGGAAAACGTCACTTGTGTGGCAATGATTTGGTGAGCTAC

CACCAAAACCAGAAACAAAACAAAATGTTACTATTTCGAGGAAGTGTTCCTACGTTCCCCAAGT
TTCATGGATATTTCAATGCAACTGTACGGGACAACATAATTGTTTGGATCTCCATTTCAACCATTAC
GCTATGAGGAGGCAATCAAGGTGACTGCTTTGCAGCATGATCTAGATACACTTCCGGGTGGTGA
CCTCACGGAGATTGGTGAAGAGGGTGTAAA TATCAGTGGTGGACAAAAGCAAAGAGTATCCAT
GGCAAGGGCTGTGTTATCTGATTCTGATGTATA TATTTGATGATCCTTTGAGTGTCTTAGATG
CCCATGTTGGTAGAGAGGTTTTTGAACAGTGTATTGAGGATAAACTTAGAGGGAAAACACGTT
TCTTGTACAAAACCAGCTACATTTCTGCCAAA TGTGGATAGAATTTATTGTTGTTTCAATGAAGGTA
TGGTTAAAAGAGGAGGACCTTTGAAGAGCTTAGTAAACAACGTTTGGCTTTCCAGAAGTTAA
GGAAAATGCTGGGAAAATGGAGGAGCAAATAGAAGAGAAGAAAACCGAAACCAATGACAAAAG
AGAGCTCAAAAATGATTGAAAA TGGAGAACTATGAAAAC TGA AAAATGGCTTGTGAAAGAGTG
AACAAAAGGGAATAATGGAAAAGAAGGAAAGACTGTACTTATCAAACAGGAGGAACGTGAA
ACAGGTGTTATAAGTTCAAAAAGTTCTCCGCAGATACAAGAACGCATTGGGTGGGATATGGGTTG
TTGGCGTTCTCTTTACATGTTATATACTCACAGAAGTCTTAGAATCTCTAGCAGCACCTGGCTA
AGTTCTTGGACAGATCAGAGCTCAACAAAAGACCCA TGGGCCAGGATATTATAATCTGATCTATG
CACTTCTTTCTTTTGGTTCAGGTAATGGTGACTCTGGTAAACTCTTATTGGTTAA TTTAATAACAGTCT
TTCATGCTGCCAAGAGATTGCTATGATGCCATGCTGAACCTATA TTTGCGAGCTCCGATGGTCTTC
TTTTATACCAATCCACTTGGGCGGATAATAAACGATTTGCCAAAAGATTAGGTGATATTGATC
GCAATGTTGCTGTTTTTGTGAGCATGTTTCTGCAACAAGTTCATGCAGCTCCTTTCCGACTTTTGTAC
TTATAGGCTTTTGTACGACTTTATCGCTGTGGGCCA TCA TGCCACTGCTGATTTTATTCTATGCAG
CCTATTTGATTACCAGAGTACTGCACGTGAAGTAAAGCGGTTAGATCTATCACAAGATCTCCT
GTCTATGCACAAATTTGGAGAAGCATTAAATGGTCTGTCTACAATCCGTGCATACAAGGCATATG
ATAGGATGGCTCAATTAATGGGAAATCAATGGACAATAATA TCAGATA TACGCTTGTGAAT
GGTGGAAATCGATGGCTTGTCTATCCGTTTGGAAACATTTGGGGGCGATTGATTTGGCTTACA
GCAACATTTGCTGTCAATGCAAAAACACGTGCAGAAAACAGCAAGCATTGATTTGCAACAAATGG
GTCTTCTTCTTCTTATGCTTTAAAATA TACCAGCTTATTGACTGCTGTGTTGCGGCTTGAAGTA
TTGCTGAGAATAGTTTGAATGCTGTTGAGCGAATTTGGGACTTACATAGAA TTTGCC TTTGAGGCT
CCACTTGTATTGAAGATAACAAACCTCCTCCAGGTTGGCCATCATCAGGTGTTATCAAATTTCA
AGATGTAAGTTCTTCGATATAGACCAGAACTTCTCTGCTGTTCTTCATGGAATATCCTTCACAATTTG
CTGGAAGTGAAGAAGATTGGAATAGTTGGGAGGACAGGTGCTGGAAAAGTCTAGTATGCTTAAATG
ATTTGGTCTTTTGGATTTACGGAAAAGCTCTGGGGAATAATTCCTCAATCACCTGTTCTGTTTTCTGG
GACTGTCCGTTTTAATCTGGACCCCTTCAA T GAGCATAATGATGCAGATCTTTGGGAAGCTTTGG
AGAGAGCACATCTAAAAGATGTCATACGGAGGAATGC ACTGGGGCTT GATGCTGAGGTTGCAG
AAGCAGGGGAAAATTTTAGTGTGGACAAAGACAGCTACTGAGTCTTTGCTCGTGCAATTTGAG
GAGATCAAAAATCCTTGTACTTGTATGAAGCAACAGCAGCTGTTGATGTCCAAACTGATGCCCTT
ATCCAGAAAACAATCAGAGAAGAAATTC AAGTCA TGCACGATGCTCATTATAGCGCATCGTTTGA
ACACTATAATTTGATGACAAGGTTTACTCTTTGAAGGTTGGCCAGGTACTAGAAATTTGACAG
ACCAGAAAATCTCTTCGAA TGAAGATAGTCTTTCTCCAAAATGGTCCAGCACTGGAGGAGCA
GCCAATGCTCAATA TTTAAGGGGTTTGGCCGTTGCGGACCGATCAAACAAGGAAGAGACCAGC
AGGCAAGATGGCCAAAGGAGGTGGCTTGCCTCTA TTCGCTGGGCAGCAGCTGCTCAATATGCAC
TTGCCAGAGCCTAACA TCTTACATAATGACCTCCAGAGATTGGAGATTGAGGATGAGAATAG
CATTCTCAAGAAAACAAGGATGCCGTTATCACCTTGCAGAACGTTTGGAAAGGAAAAGCACGAT
TCAGTGAATGAAGAGAGCTTGAATCAGTATCAAAGTCCCAGAGAGAGATGGTGGTCACTCTCTAT
ATAAAAATGATTGAAGGCTTTCGTTGATGAGTAGATTGGCCATAACAGGCTTCAACATCCAGA
ATTTGTGTTGAAAGATCGTTCACTTGAATTTGGACAATGCTGAAAATGTAG

SEQ. AA CsABCC2(1626 aa)(SEQ ID NO 14)

MGFKPLGWYCRPVKDGLWSMTVHNALGAYTPCTMETIVVGISHLALFGTFCFYRIWRMSRDVTVKR
 YCLRSKYHILLGLLAAFCLEPLFRLVFGISIMDIDHHTSLPPFEILSLIVEA VAWCCMLVMLGMESK
 VYIHEFRWYVRFVVLVVLVGETAVFNLVLSVREY YEESVFYLYTSVIVCQFLFGFLLLIHMPTLDPFP
 GYTPIRDQVFVDNTEYESLAGEEHICPERHVNLLSKILFSWMTPLMQQG YRKPITERDVWKLDSWDT
 TETLYSNFQKCWLEESKKPKPWLLRALHSSLGGRFWLGGFFKIGNDL SQFVGPTILNLLLQSMQEGD
 PSWNGYIYAF AIFAGVVFGLSEAQYFQNVMRVGFRLRSTL VAAVTRKSLRLTHEGRRKFASGKITN
 LMTTDAESLQQICQQLHSLWSAPFRIHSIILLYQELGVASLAGALMLVLMFPPIQTLVISKMQKLTKEG
 IQRTDKRI
 GLMNELLAAMDTVKCYAWEQSFQSKVQSIRDDEL SWTRKSQLLGSLNSFM LNSIPVVVTVVSEGVY
 SLLGGNLTPAKAFTSLSLFAVLRFPFLMPLNLTQV VNTKVSLDRLEGLFLAEERILLPNPPLDPTLPAI
 SIKDGYFSWDRKEERP TLSNINLDIPVGS LVAIVGSTGEGK TSLVSAMIGELPPKPE TTNVTIRG SVA
 YVPQVSWIFNATVRDNILFGSPFQPLRYEEAIKVTALQHDLD TLPGGDLTEIGERGVNISGGQKQRVS
 MARAVYSDSDVYIFDDPLSALDAHVGREVF EQCIEDKLRGKTRVLVTNQLHFLPNVDRIIVVHEGM
 VKEEGTFEELSNNGLLFQKLMENAGKMEEQIEKKTETNDKESSKLIENGETMKTENGLL KSEQKG
 NNGKEGKTVLIKQEERETGVISSKVLRRYKNALGGIWVVA VLFTCYILTEVLRISSTWLSWTDQSS
 TKTHGPGYYNLIYALLSFGQVMVTLVNSYWLII NSLHAAKRLHDAMLNSILRAPMVFFHTNPLGRH
 NRFAKDLGDIDRNVA VFVSMFLQQVMQLLSTFVLIGFVSTLSLWAIMPLLILFYAA YLYYQSTAREV
 KRLDSITRSPVYAQFGEALNGLSTIRAYKAYDRMAAINGK SMDNNIRYTLVNMGGNRWL AIRLETL
 GGIMIWL TATFAVMQNQRAENQQAFASTMGLLLSYALNITSLTAVLRLASIAENSLNA VERIGTYIE

LPSEAPLVIEDNKPPP GWSSGVIKFQDVVLRYRPELPAVLHGISTT IAGSEKIGIVGRTGAGKSSMLN
 ALFRIVELE RGTIFIDEYDVSKFGLLDLRKALGHIPQSPVLFSGTVRFNLDPFNEHNDADLWEALERAH
 LKDVIRRNALGLDAEVAEAGENFSV GQRQLLSLARALLRRSKILVLDEATAAVDVQTDALIQKTIRE
 EFKSCTMLIIAHRLNTHDCDKVLLLEGGQVLEFDRPENLLSNEDSAFSK MVQSTGAANAQYLRGLA
 VADRSNKEETS RQDGQRRWLASIRWAAAAQYALAQSLTSSHNDLQRLEIEDENSILKKTDAVITLQ
 NVLEGGKHDSVIEESLNQYQVPRERWSSSLYK MIEGLALMSRLAHNRLQHPEFVFEDRSLDWDNAE
 M

REIVINDICACIONES

1. Secuencia de aminoácidos que tiene una secuencia seleccionada entre SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o mayor al 85 % con la SEQ ID NO 12 (CsABCC1) o la SEQ ID NO 14 (CsABCC2), en el que dicha secuencia de aminoácidos es una proteína transportadora apta para transportar crocinas a través de una membrana lipídica.
2. Secuencia de aminoácidos, según la reivindicación 1, en la que dicha secuencia es SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2).
3. Secuencia de nucleótidos que codifica una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o mayor al 85 % con la SEQ ID NO 12 (CsABCC1) o la SEQ ID NO 14 (CsABCC2) apta para transportar crocinas a través de una membrana lipídica.
4. Secuencia de nucleótidos, según la reivindicación 3, en la que dicha secuencia de nucleótidos es SEQ ID NO 11 o SEQ ID NO 13.
5. Vector de expresión que comprende:
- una secuencia de nucleótidos que codifica una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o mayor al 85 % con SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2) apta para transportar crocinas a través de una membrana lipídica.
6. Vector de expresión, según la reivindicación 5, en el que dicha secuencia de nucleótidos es SEQ ID NO 11 o SEQ ID NO 13.
7. Vector de expresión, según la reivindicación 5 o 6, en el que dicho vector se selecciona entre vectores pNEV, pAG, pYES (Thermo Fisher) para expresión en levaduras; vectores derivados de pBI121 (Clontech), o de pCAMBIA (Change) para la expresión en plantas.
8. Célula recombinante que comprende una secuencia de nucleótidos que codifica:
- una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o superior al 85 % con SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2).
9. Célula recombinante, según la reivindicación 8, en la que dicha secuencia de nucleótidos es SEQ ID NO 11 o SEQ ID NO 13.
10. Célula recombinante, según la reivindicación 8 o 9, en la que dicha célula se selecciona entre células de levadura y células vegetales.
11. Organismo huésped recombinante que comprende una secuencia de nucleótidos que codifica:
- una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o mayor al 85 % con SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2).
12. Organismo huésped recombinante, según la reivindicación 11, seleccionado entre un microorganismo o una planta.
13. Organismo huésped recombinante, según la reivindicación 11, en el que dicho organismo se selecciona de *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha*, *Candida boidinii*, *Arxula adeninivorans*, *Xanthophyllomyces dendrorhous*, especies de *Candida albicans*, *Arabidopsis thaliana*, *Nicotiana benthamiana*, *Zea mays*, *Solanum lycopersicum*, *Solanum tuberosum*, *Oryza sativa*.
14. Organismo huésped recombinante, según cualquiera de las reivindicaciones 11 a 13, en el que dicha secuencia de nucleótidos es SEQ ID NO 11 o SEQ ID NO 13.
15. Método para acumular crocinas en la vacuola de un organismo huésped que comprende una etapa de expresión de una secuencia de nucleótidos que codifica una proteína de transporte que tiene la secuencia de aminoácidos SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABCC2).

16. Método, según la reivindicación 15, en el que dicho organismo huésped es una levadura o una planta recombinante.
- 5 17. Método, según la reivindicación 15, en el que dicho organismo huésped es un organismo recombinante seleccionado de *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha*, *Candida boidinii*, *Arxula adenivorans*, *Xanthophyllomyces dendrorhous*, especies de *Candida albicans*, *Arabidopsis thaliana*, *Nicotiana benthamiana*, *Zea mays*, *Solanum lycopersicum*, *Solanum tuberosum*, *Oryza sativa*.
- 10 18. Uso de una proteína que tiene la SEQ ID NO 12 (CsABCC1) o la SEQ ID NO 14 (CsABCC2) para transportar crocinas a través de una membrana lipídica de una planta o un microorganismo, en particular para acumular crocina en la vacuola.
- 15 19. Método, según la reivindicación 15, en el que el transporte en la vacuola se controla mediante la exposición de una preparación de vacuolas purificadas de un extracto que contienen metabolitos vegetales, en el que dicha preparación de vacuolas comprende uno o más transportadores vesiculares.
- 20 20. Método, según la reivindicación 19, en el que dicha preparación está constituida por microsomas de levadura.
21. Método, según la reivindicación 19 o 20, en el que dicho extracto vegetal es un extracto obtenido con un disolvente polar o apolar.
- 25 22. Método, según cualquiera de las reivindicaciones 19 a 21, en el que dichos transportadores vesiculares tienen una secuencia seleccionada entre SEQ ID NO 12 (CsABCC1), SEQ ID NO 14 (CsABCC2) o una secuencia de aminoácidos con un porcentaje de identidad igual o superior al 85 % con SEQ ID NO 12 (CsABCC1) o SEQ ID NO 14 (CsABC2) aptas para transportar crocinas a través de una membrana lipídica.

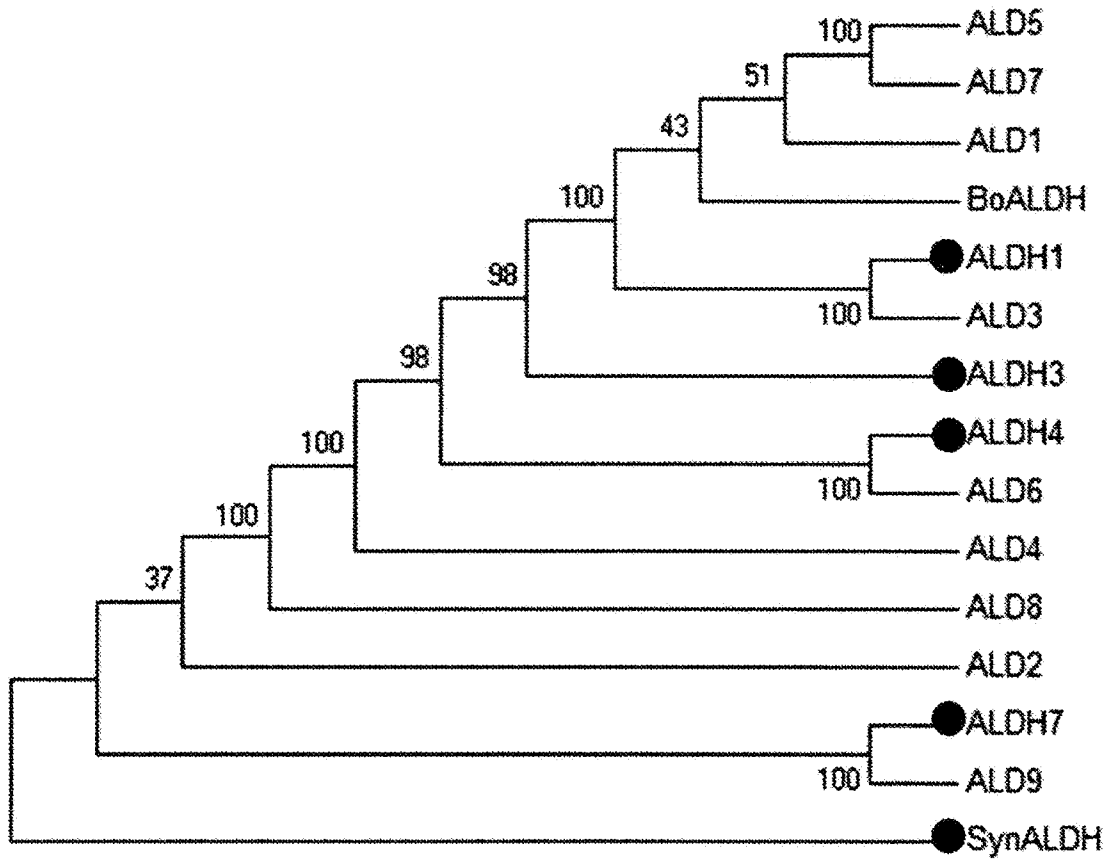


FIG. 3

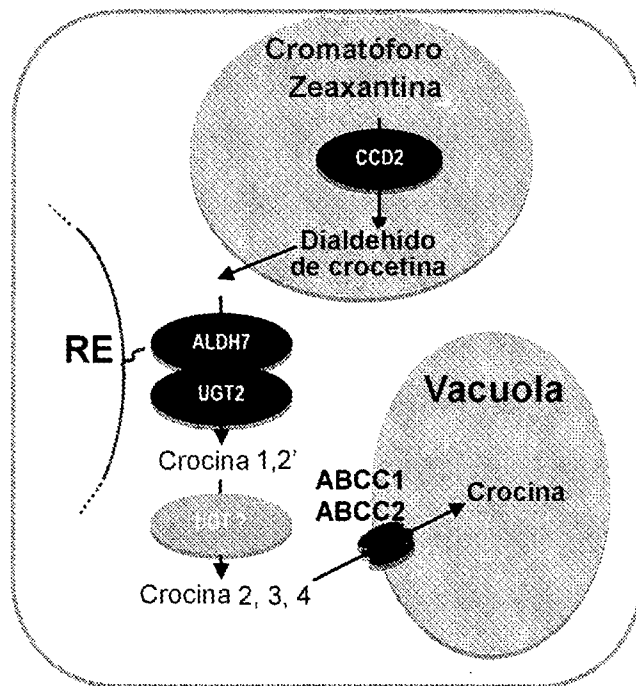


FIG. 4

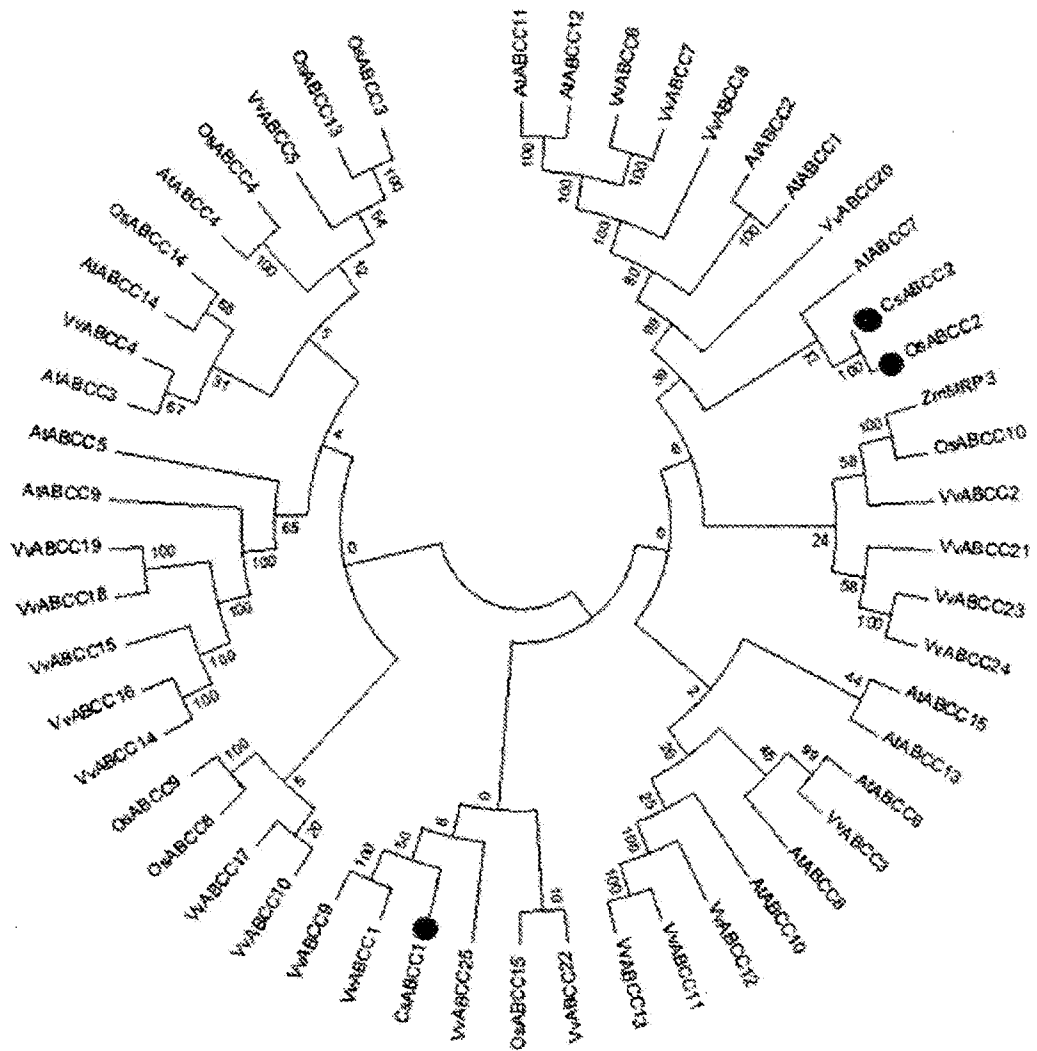


FIG. 5

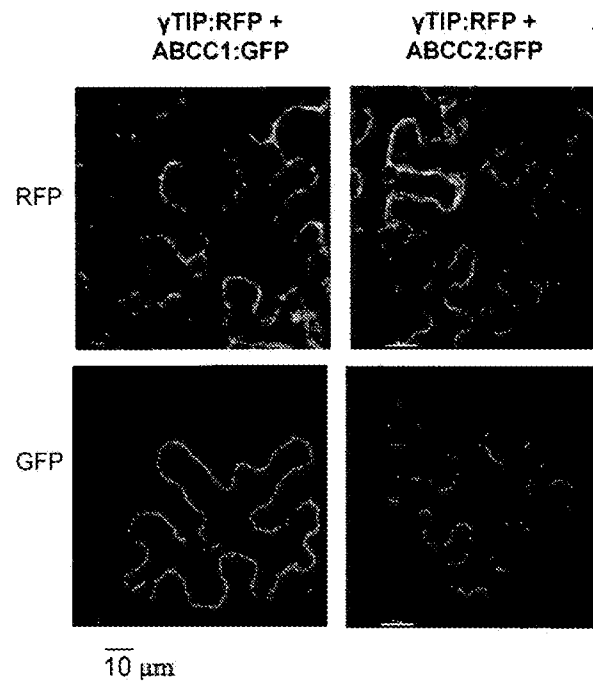


FIG. 6

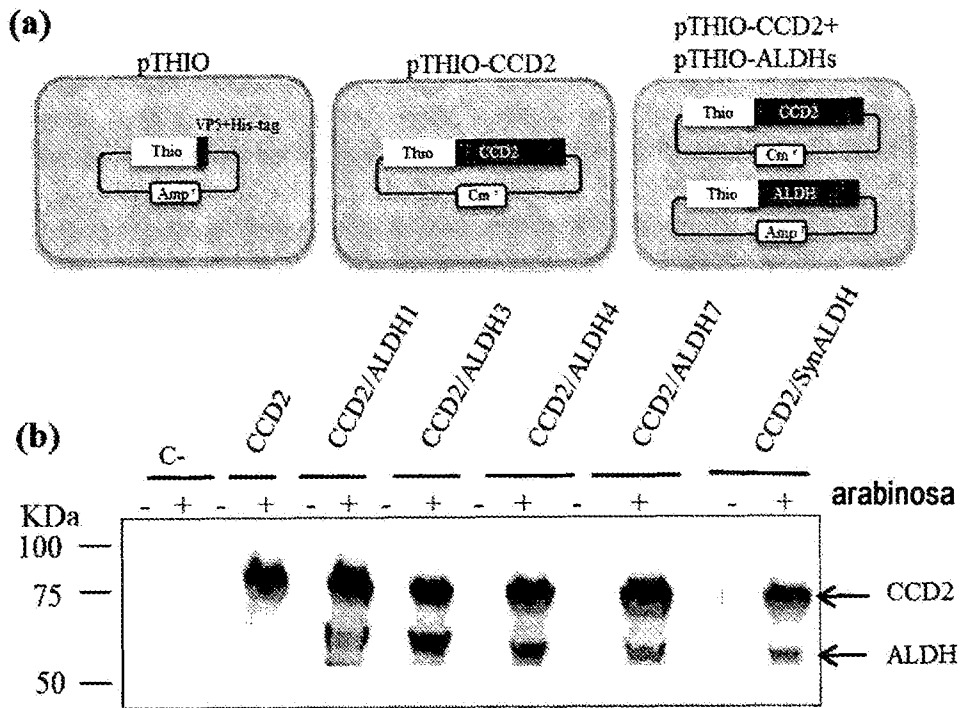


FIG. 7

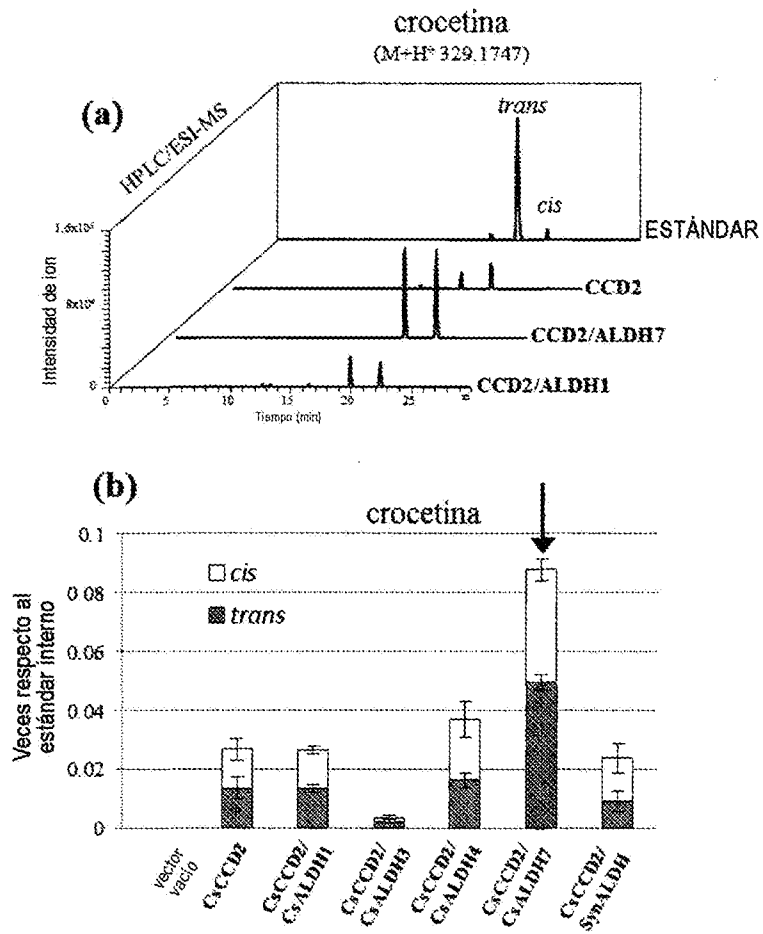


FIG. 8

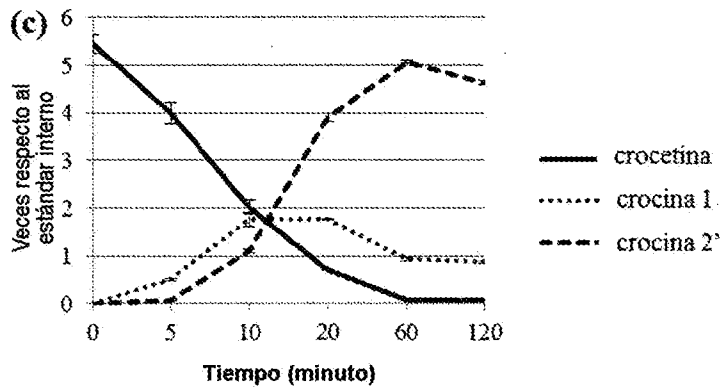
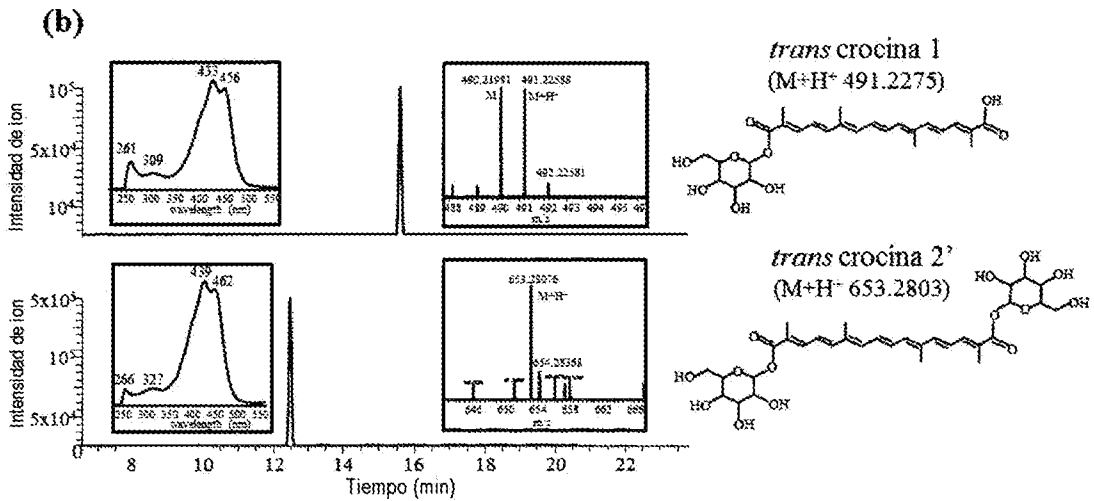
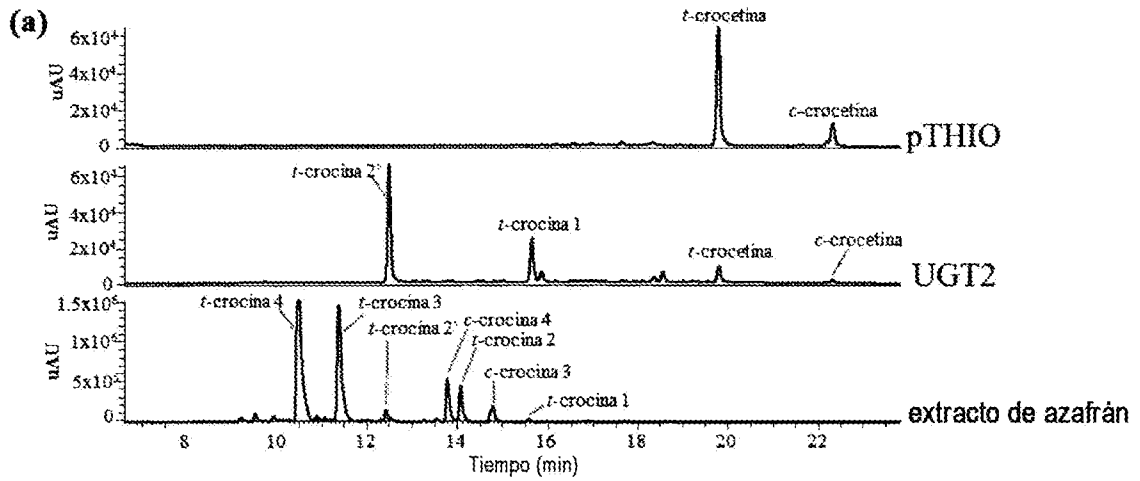


FIG. 9

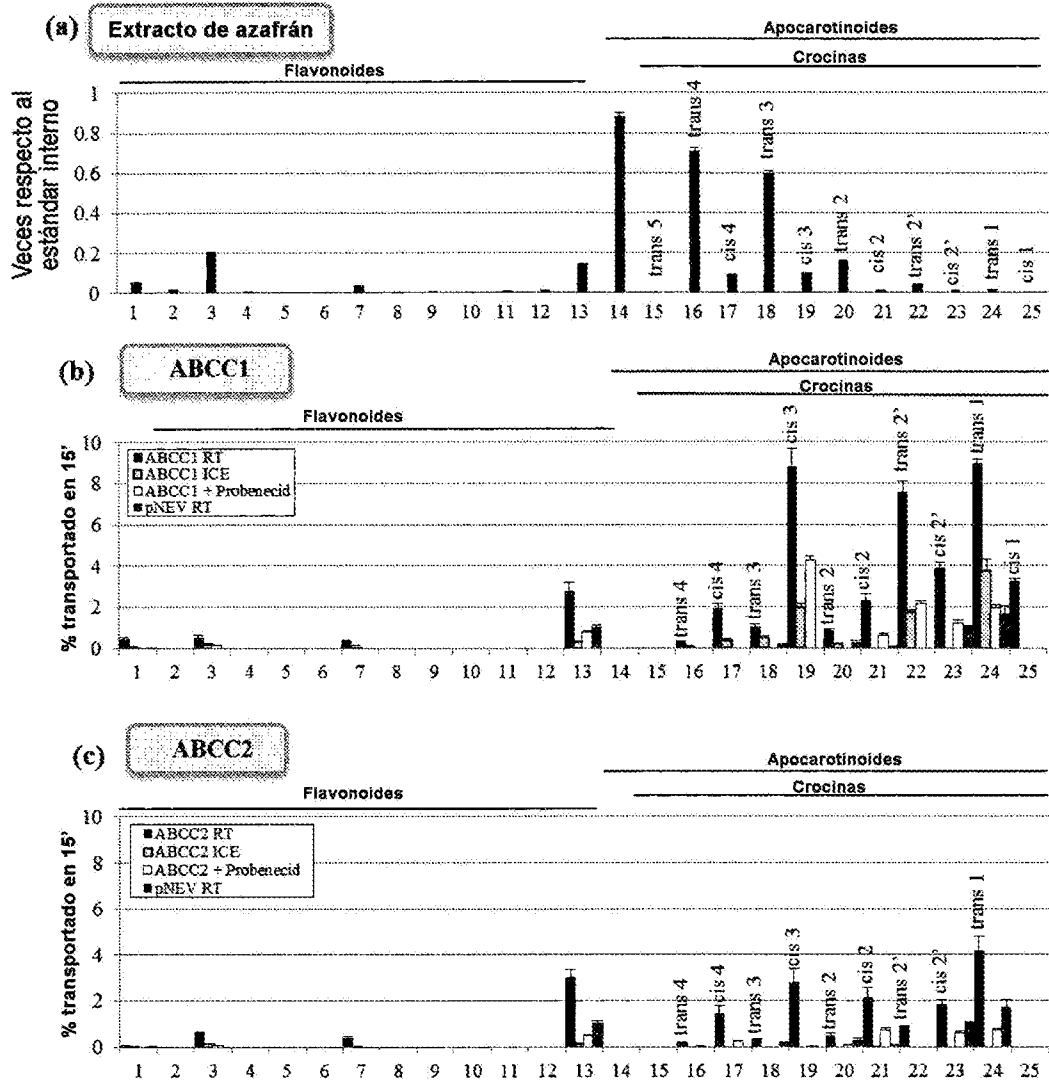


FIG. 10

REFERENCIAS CITADAS EN LA DESCRIPCIÓN

Esta lista de referencias citada por el solicitante es únicamente para mayor comodidad del lector. No forman parte del documento de la Patente Europea. Incluso teniendo en cuenta que la compilación de las referencias se ha efectuado con gran cuidado, los errores u omisiones no pueden descartarse; la EPO se exime de toda responsabilidad al respecto.

Documentos de patentes citados en la descripción

- WO 2013021261 A
- WO 2015132411 A
- WO RM2014A000415 A, Giuliano

Literatura no patente citada en la descripción

- **SHOIB AHMAD BABA et al.** *Comprehensive transcriptome analysis of Crocus sativus for discovery and expression of genes involved in apocarotenoid biosynthesis*
- **MORAGA et al.** *Planta*, 2004, vol. 219, 955-966
- **GIULIANO et al.** *Patent Application RM2014A000415*
- **FRUSCIANTE et al.** *Proc Natl Acad Sci U S A*, 2014, vol. 111, 12246-12251
- **KUMAR S et al.** *Mol Biol Evol.*
- **CHENNA et al.** *Nucleic Acids Res.*, 2003, vol. 31 (13), 3497-500
- **FRUSCIANTE et al.** *Proc Natl Acad Sci U S A*, vol. 111 (33), 12246-12251
- **DEMURTAS et al.** *Front Plant Sci.*, vol. 7, 54
- **FASANO et al.** *New Phytologist*, vol. 210, 1382-1394
- **TOMMASINI et al.** *Proc Natl Acad Sci USA.*, vol. 93 (13), 6743-6748
- **LEIER et al.** *J Biol Chem*, 1994, vol. 269 (45), 27807-10
- **TRAUTMANN et al.** *Febs j*, 2013, vol. 280 (15), 3685-96
- **FRUSCIANTE et al.** *Proc Natl Acad Sci U S A*, 2014, vol. 111 (33), 12246-12251
- **NELSON et al.** *Plant J*, 2007, vol. 51 (6), 1126-36
- **HAMILTON; BAULCOMBE.** *Science*, 1999, vol. 286 (5441), 950-2
- **SAUER; STOLZ.** *Plant J*, 1994, vol. 6 (1), 67-77
- **TOMMASINI et al.** *Proc Natl Acad Sci U S A.*, vol. 93 (13), 6743-6748
- **MANIATIS et al.** *MOLECULAR CLONING: A LABORATORY MANUAL.* Cold Spring Harbor Laboratory, 1989
- **AUSUBEL et al.** *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY.* Greene Publishing Associates and Wiley Interscience, 1989
- **INNIS et al.** *PCR Protocols: A Guide to Methods and Applications.* Academic Press, 1990