(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 555 206**

(51) Int. Cl.:

C12N 9/10 (2006.01)
C12N 15/82 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **19.12.2008** **E 08863568 (5)**

(97) Fecha y número de publicación de la concesión europea: **30.09.2015** **EP 2235051**

(54) Título: **Glicosiltransferasas, polinucleótidos que las codifican y métodos de uso**

(30) Prioridad:

21.12.2007 NZ 56469107

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
29.12.2015

(73) Titular/es:

THE NEW ZEALAND INSTITUTE FOR PLANT AND FOOD RESEARCH LIMITED (100.0%)
MT ALBERT RESEARCH CENTRE 120 MT ALBERT ROAD
MT ALBERT, AUCKLAND, NZ

(72) Inventor/es:

JUGDE, HELENE y
ATKINSON, ROSS

(74) Agente/Representante:

ISERN JARA, Jorge

ES 2 555 206 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

Glicosiltransferasas, polinucleótidos que las codifican y métodos de uso

Campo técnico

La presente divulgación se refiere a composiciones y a métodos para la producción de plantas con actividad de floretina glicosiltransferasa alterada y/o contenido de floricina alterado.

Antecedentes de la técnica

La dihidrochalcona floricina (2'-glucósido de floretina, véase la Fig. 1) es el glucósido fenólico principal que se encuentra en los manzanos. La floricina tiene un sabor amargo que contribuye al sabor característico de la sidra (Whiting y Coggins, 1975), y sus productos de oxidación dimerizados contribuyen al color de los zumos de manzana (Ridgway y Tucker, 1997). Sin embargo, desde que fue aislada de la corteza del manzano en 1835 (Petersen, 1835), la floricina ha suscitado el mayor interés científico a través de su uso como producto farmacéutico y como una herramienta para la investigación de la fisiología. Su principal acción farmacológica consiste en producir glucosuria renal y transportar glucosa en bloque mediante la inhibición de los transportadores de glucosa ligados al sodio (revisado en Ehrenkranz *et al*., 2005). La floricina y sus derivados también han mostrado ser antioxidantes sumamente eficaces *in vitro* (Ridgway *et al*., 1996) y tener una serie de funciones bioactivas tales como la inhibición de la peroxidación lipídica (Ridgway *et al*, 1997; Rezk *et al*., 2002), la prevención de la pérdida ósea (Puel *et al*. 2005), la mejora de la memoria (Boccia *et al*., 1999) y la inhibición del crecimiento de células cancerosas (Veeriah *et al*. 2006).

Hasta hace poco, se creía que la floricina solo existía en las especies del género *Malus*. Sin embargo, se han descrito glucósidos de floretina en las hojas de la zarzaparrilla nativa de Australia (*Smilax glyciphylla*, Cox *et al*., 2005), el té dulce (*Lithocarpus polystachyus*, Dong *et al*., 2007) y, a niveles muy bajos, en la fresa (Hilt *et al*., 2003). En los manzanos, la floricina se encuentra principalmente en los brotes jóvenes, en las raíces, en las hojas y en la corteza. En el fruto, la floricina es más abundante en las semillas, con niveles intermedios tanto en el núcleo como en la piel, y el nivel más bajo en la corteza. Se ha evaluado la variación en los manzanos, entre huertos, entre las diferentes variedades cultivadas y entre mutantes (Hunter y Hull, 1993, Awad *et al*., 2000). A pesar de dicha información, se sabe poco de la función *in planta* de la floricina en la fisiología del manzano, aunque se ha sugerido que podría actuar en el crecimiento y el desarrollo del manzano (Zhang *et al*., 2007), o ser un inhibidor del crecimiento bacteriano (MacDonald y Bishop, 1952) o fúngico (Gessler *et al*., 2006).

No se ha descrito la base molecular para la producción de floricina *in planta*. La floretina es un producto de la ruta de los fenilpropanoides (Watts *et al*., 2004), con la conversión a su glucósido, la floricina, susceptible de ser catalizada por la acción de una uridina difosfato (UDP) glicosiltransferasa (UGT). Las UGT median la transferencia de un residuo de azúcar desde un nucleótido de azúcar activado a moléculas aceptoras (agliconas). Las plantas contienen grandes familias de UGT, habiéndose descrito más de 100 genes en *Arabidopsis*. Estos genes tienen un motivo de firma común de 42 aminoácidos que se cree que participa en la unión del resto de UDP del azúcar activado (Li *et al*., 2001). Un análisis filogenético estableció la presencia de distintos grupos (A-N) y familias (UGT71-92) de genes de UGT en *Arabidopsis* (Ross *et al*., 2001), y esto facilitó la caracterización de muchas actividades nuevas (Jackson *et al.,* 2001; Lim *et al*., 2002; Jones *et al.,* 2003; Messner *et al.,* 2003; Lim *et al.,* 2004; Lim *et al.,* 2005). Aunque inicialmente se consideraron enzimas promiscuas, pruebas recientes sugieren que su amplia especificidad de sustrato está limitada por regio-especificidad (Hansen *et al.,* 2003; Lim *et al*., 2003) y, en algunos casos, las UGT han demostrado ser altamente específicas (Fukuchi-Mizutani *et al*., 2003). Usando una metodología de genómica funcional, los presentes inventores han identificado y caracterizado una UGT de manzana perteneciente a la familia 88 de UGT. Se ha establecido que MpUGT88A1 media la glicosilación de la floretina dihidrochalcona a floricina, lo que puede indicar que otros miembros de la familia 88 de UGT utilizan sustratos similares. JUGDE HELENE ET AL., FEBS JOURNAL, vol. 275, Nº 15, agosto de 2008, páginas 3804-3814, desvela el aislamiento y la caracterización de una glicosiltransferasa que convierte la floretina en floricina, un potente antioxidante de la manzana.

El documento JP 2006 262910 A desvela un método de producción de un glucósido floretina que comprende:

- una primera etapa de conversión de la floricina presente en el zumo exprimido o extracto de un fruto inmaduro de un manzano o un peral en floretina ligeramente hidrosoluble mediante el tratamiento del zumo exprimido o del extracto del fruto inmaduro del manzano o del peral con una enzima que tiene actividades β-flucosidasa;
- una segunda etapa de purificación y extracción de floretina; y
- una tercera etapa de conversión de floretina purificada y extraída en la segunda etapa en glucósido floretina.

OGATA JUN ET AL., PLANT BIOTECHNOLOGY, vol. 21, Nº 5, 2004, páginas 367-375, desvela el aislamiento y la caracterización de los clones de ADNc de longitud completa para 18 glicosiltransferasas (GT) de tejido de pétalo del clavel (*Dianthus caryophyllus*). Tres de las 18 se caracterizaron como 3-GT que poseían diferentes especificidades de sustrato hacia los flavonoides y la antrocianidina, y otras dos GT catalizaron la transferencia de la glucosa en el

grupo 2-hidroxilo de la chalcona. Además, estas dos enzimas glucosilaron el flavonol (3-OH y 7-OH), la flavanona (7-OH), la flavona (7-OH) y la antocianidina (3-OH y 7-OH).

El documento JP 2004 305049 A desvela un gen de la chalcona 2'-O-glicosiltransferasa (Ch 2'GT) aislado de *Dianthus caryophyllus*.

WERNER S R ET AL., JOURNAL OF BIOTECHNOLOGY, vol. 142, Nº 3-4, 15 de julio de 2009, páginas 233-241, desvela *Saccharomyces cerevisiae* que expresa una glicosiltransferasa (GT) flavonoide de *Dianthus caryophyllus* y su uso como un biocatalizador de células enteras. *S. cerevisiae* que expresó la GT constitutivamente fue significativamente más productiva que un sistema inducible de galactosa similar en condiciones *in vivo*.

Patenaude *et al.* (2002), *Nature Structural Biology*, Vol. 9, pág. 685-689 desvela las estructuras cristalinas de *N*-acetilgalactosaninil transferasa (GTA) y galactosil transferasa (GTB) humanas.

Sería beneficioso tener un medio para aumentar los niveles de floricina en las plantas.

Es un objeto de la invención proporcionar mejores composiciones y métodos de modulación de la actividad y/o del contenido de floricina en plantas o al menos proporcionar al público una elección útil.

Sumario

En un primer aspecto, la divulgación describe un método de producción de una célula vegetal o una planta con mayor contenido de floricina, método que comprende la transformación de una célula vegetal o de una planta con un polinucleótido que codifica un polipéptido con la secuencia de aminoácidos de una cualquiera de SEC ID Nº 1 a 5 o una variante del polipéptido.

Preferentemente, la variante tiene actividad de la floretina glicosiltransferasa.

En un aspecto adicional, la divulgación describe un método de producción de una célula vegetal o una planta con una mayor actividad floretina glicosiltransferasa, método que comprende la transformación de una célula vegetal o una planta con un polinucleótido que codifica un polipéptido con la secuencia de aminoácidos de una cualquiera de SEC ID Nº 1 a 5, o una variante del polipéptido.

Preferentemente, la variante tiene actividad de la floretina glicosiltransferasa.

Preferentemente, la variante comprende la secuencia de SEC ID Nº 16.

En una realización, la variante tiene al menos el 70 % de identidad de secuencia con un polipéptido con la secuencia de aminoácidos de una cualquiera de SEC ID Nº 1 a 5.

En una realización adicional, el polinucleótido codifica un polipéptido con una secuencia de aminoácidos que tiene al menos el 70 % de identidad con la secuencia de SEC ID Nº 1.

En una realización adicional, el polinucleótido codifica un polipéptido con la secuencia de aminoácidos de SEC ID Nº 1.

En una realización adicional, el polinucleótido codifica un polipéptido con una secuencia de aminoácidos que tiene al menos el 70 % de identidad con la secuencia de SEC ID Nº 2.

En una realización adicional, el polinucleótido codifica un polipéptido con la secuencia de aminoácidos de SEC ID Nº 2.

En una realización adicional, el polinucleótido codifica un polipéptido con una secuencia de aminoácidos que tiene al menos el 70 % de identidad con la secuencia de SEC ID Nº 3.

En una realización adicional, el polinucleótido codifica un polipéptido con la secuencia de aminoácidos de SEC ID Nº 3.

En una realización adicional, el polinucleótido codifica un polipéptido con una secuencia de aminoácidos que tiene al menos EL 70 % de identidad con la secuencia de SEC ID Nº 4.

En una realización adicional, el polinucleótido codifica un polipéptido con la secuencia de aminoácidos de SEC ID Nº 4.

En una realización adicional, el polinucleótido codifica un polipéptido con una secuencia de aminoácidos que tiene al menos EL 70 % de identidad con la secuencia de SEC ID Nº 5.

En una realización adicional, el polinucleótido codifica un polipéptido con la secuencia de aminoácidos de SEC ID Nº 5.

En una realización preferida del método de producción de una célula vegetal o una planta con mayor contenido de floricina, la célula vegetal o la planta también se transforma con un polinucleótido que codifica una floretina sintasa.

La transformación con la floretina glicosiltransferasa y sintasa puede ser secuencial, en cualquier orden. Como alternativa, la transformación con la floretina glicosiltransferasa y sintasa puede ser simultánea. Cuando es simultánea, las secuencias que codifican la floretina glicosiltransferasa y sintasa pueden estar en construcciones o vectores iguales o separados.

En un aspecto adicional, la divulgación describe un método de producción de una célula vegetal o una planta con mayor contenido de floricina, método que comprende la transformación de una célula vegetal o una planta con un polinucleótido que comprende una secuencia de nucleótidos seleccionada entre una cualquiera de las secuencias de SEC ID Nº 7 a 18, o una variante de las mismas.

Preferentemente, la variante codifica un polipéptido con actividad de la floretina glicosiltransferasa.

Preferentemente, la variante codifica un polipéptido que comprende la secuencia de SEC ID Nº 16.

En un aspecto adicional, la divulgación describe un método de producción de una célula vegetal o una planta con mayor actividad de la floretina glicosiltransferasa, método que comprende la transformación de una célula vegetal o una planta con un polinucleótido que comprende una secuencia de nucleótidos seleccionada entre una cualquiera de las secuencias de SEC ID Nº 6 a 15, o una variante de las mismas.

Preferentemente, la variante codifica un polipéptido que tiene actividad de la floretina glicosiltransferasa.

Preferentemente, la variante codifica un polipéptido que comprende la secuencia de SEC ID Nº 16.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con una cualquiera de las secuencias de SEC ID Nº 6 a 15.

En una realización adicional, el polinucleótido comprende una secuencia con una identidad de al menos el 70 % con una cualquiera de las secuencias de SEC ID Nº 6 a 15.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 6.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 7.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 7.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 7.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 8.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 8.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 9.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 9.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 10.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 10.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 11.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 11.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 12.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 12.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 13.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 13.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 14.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 14.

En una realización adicional, el polinucleótido comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 15.

En una realización adicional, el polinucleótido comprende la secuencia de SEC ID Nº 15.

En una realización preferida del método de producción de una célula vegetal o planta con un mayor contenido de floricina, la célula vegetal o la planta también se transforma con un polinucleótido que codifica una floretina sintasa.

La transformación con la floretina glicosiltransferasa y sintasa puede ser secuencial, en cualquier orden. Como alternativa, la transformación con la floretina glicosiltransferasa y sintasa puede ser simultánea. Cuando es simultánea, las secuencias que codifican la glicosiltransferasa y la sintasa pueden estar en construcciones o vectores iguales o separados.

En un aspecto adicional, la divulgación describe un polinucleótido aislado que codifica un polipéptido que comprende una secuencia seleccionada de una cualquiera de SEC ID Nº 2 a 5 o una variante de las mismas.

Preferentemente, la variante tiene actividad de la floretina glicosiltransferasa.

Preferentemente, la variante comprende la secuencia de SEC ID Nº 16.

En una realización, la variante comprende una secuencia con al menos el 80 % de identidad con SEC ID Nº 2.

En una realización adicional, el polipéptido comprende la secuencia de SEC ID Nº 2.

En una realización, la variante comprende una secuencia con al menos el 80 % de identidad con SEC ID Nº 3.

En una realización adicional, el polipéptido comprende la secuencia de SEC ID Nº 3.

En una realización, la variante comprende una secuencia con al menos el 80 % de identidad con SEC ID Nº 4.

En una realización adicional, el polipéptido comprende la secuencia de SEC ID Nº 4.

En una realización, la variante comprende una secuencia con al menos el 80 % de identidad con SEC ID Nº 5.

En una realización adicional, el polipéptido comprende la secuencia de SEC ID Nº 5.

En un aspecto adicional, la divulgación describe un polinucleótido aislado que comprende la secuencia de una cualquiera de SEC ID Nº 8 a 15, o una variante de las mismas.

Preferentemente, la variante codifica una floretina glicosiltransferasa.

Preferentemente, la variante codifica un polipéptido que comprende la secuencia de SEC ID Nº 16.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 8.

En una realización, el polinucleótido comprende la secuencia de una cualquiera de SEC ID Nº 8.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 9.

En una realización, el polinucleótido comprende la secuencia de una cualquiera de SEC ID Nº 9.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 10.

En una realización, el polinucleótido comprende la secuencia de una cualquiera de SEC ID Nº 10.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 11.

En una realización, el polinucleótido comprende la secuencia de una cualquiera de SEC ID Nº 11.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 12.

En una realización, el polinucleótido comprende la secuencia de una cualquiera de SEC ID Nº 12.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 13.

En una realización, el polinucleótido comprende la secuencia de una cualquiera de SEC ID Nº 13.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 14.

En una realización, el polinucleótido comprende la secuencia de una cualquiera de SEC ID Nº 14.

En una realización, la variante comprende una secuencia con al menos el 70 % de identidad de secuencia con la secuencia de SEC ID Nº 15.

En una realización, el polinucleótido comprende la secuencia de una cualquiera de SEC ID Nº 15.

En un aspecto adicional, la divulgación describe un polipéptido aislado que comprende la secuencia de aminoácidos de SEC ID Nº 2 a 5, o una variante de las mismas.

Preferentemente, la variante tiene actividad de la floretina glicosiltransferasa.

Preferentemente, la variante comprende la secuencia de SEC ID Nº 16.

En una realización, la variante polipeptídica tiene al menos el 80 % de identidad de secuencia con una secuencia de aminoácidos seleccionada de una cualquiera de SEC ID Nº 2 a 5.

En una realización adicional, el polipéptido aislado tiene al menos el 80 % de identidad de secuencia con la secuencia de aminoácidos de SEC ID Nº 2.

En una realización adicional, el polipéptido aislado comprende la secuencia de aminoácidos de SEC ID Nº 2.

En una realización adicional, el polipéptido aislado tiene al menos el 80 % de identidad de secuencia con la secuencia de aminoácidos de SEC ID Nº 3.

En una realización adicional, el polipéptido aislado comprende la secuencia de aminoácidos de SEC ID Nº 3.

En una realización adicional, el polipéptido aislado tiene al menos el 80 % de identidad de secuencia con la secuencia de aminoácidos de SEC ID Nº 4.

En una realización adicional, el polipéptido aislado comprende la secuencia de aminoácidos de SEC ID Nº 4.

En una realización adicional, el polipéptido aislado tiene al menos el 80 % de identidad de secuencia con la secuencia de aminoácidos de SEC ID Nº 5.

En un aspecto adicional, la divulgación describe un polinucleótido aislado que codifica un polipéptido descrito en el presente documento.

En un aspecto adicional, la divulgación describe un polinucleótido aislado que comprende:

a) un polinucleótido que comprende un fragmento, de al menos 15 nucleótidos de longitud, de un polinucleótido descrito en el presente documento;
b) un polinucleótido que comprende un complemento, de al menos 15 nucleótidos de longitud, del polinucleótido descrito en el presente documento; o
d) un polinucleótido que comprende una secuencia, de al menos 15 nucleótidos de longitud, capaz de hibridarse con el polinucleótido descrito en el presente documento.

En un aspecto adicional, la divulgación describe una construcción genética que comprende un polinucleótido descrito en el presente documento.

En un aspecto adicional, la divulgación describe una construcción de expresión que comprende un polinucleótido de la invención.

En un aspecto adicional, la divulgación describe una construcción de ARNi que comprende un polinucleótido de la invención.

En un aspecto adicional, la divulgación describe un vector que comprende una construcción de expresión, una construcción genética o una construcción de ARNi de la invención.

En un aspecto adicional, la divulgación describe una célula huésped modificada genéticamente para expresar un polinucleótido de la invención o un polipéptido de la invención.

Preferentemente, la célula huésped también se modifica genéticamente para expresar un polinucleótido que codifica la floretina sintasa.

En un aspecto adicional, la divulgación describe una célula huésped que comprende una construcción de expresión o construcción genética de la invención.

Preferentemente, la célula huésped también comprende una construcción de expresión o construcción genética que incluye un polinucleótido que codifica la floretina sintasa.

En un aspecto adicional, la divulgación describe un método de producción de un polipéptido floretina glicosiltransferasa, método que comprende cultivar una célula huésped que comprende una construcción de expresión de la invención o una construcción genética de la invención, capaz de expresar un polipéptido floretina glicosiltransferasa.

En un aspecto adicional, la divulgación describe un método de producción del producto enzimático de una floretina glicosiltransferasa, método que comprende cultivar una célula huésped que incluye una construcción de expresión de la invención o una construcción genética de la invención, capaz de expresar un polipéptido floretina glicosiltransferasa, en presencia de sustrato enzimático que se puede suministrar o puede estar presente de manera natural en la célula huésped.

En un aspecto adicional, la divulgación describe un método para la biosíntesis de floricina comprende las etapas de cultivar una célula huésped que comprende una construcción de expresión de la invención o una construcción genética de la invención, capaz de expresar una floretina glicosiltransferasa, en presencia de un floretina que se puede suministrar o puede estar presente de manera natural en la célula huésped.

Preferentemente, la célula huésped también se modifica genéticamente para expresar un polinucleótido que codifica la floretina sintasa.

Preferentemente, la célula huésped es una célula vegetal. Preferentemente, la célula vegetal es parte de una planta.

En un aspecto adicional, la divulgación describe una célula vegetal modificada genéticamente para expresar un polinucleótido de la invención o un polipéptido descrito en el presente documento.

En un aspecto adicional, la divulgación describe una célula vegetal que comprende una construcción de expresión de la invención o la construcción genética de la invención.

En un aspecto adicional, la divulgación describe una planta que comprende una célula vegetal de la invención.

Preferentemente, las células vegetales o las plantas también se modifican genéticamente para expresar un polinucleótido que codifica una floretina sintasa.

Preferentemente, la célula huésped también se modifica genéticamente para expresar un polinucleótido que codifica una floretina sintasa.

En algunos casos, puede ser deseable reducir la actividad de la floretina glicosiltransferasa y/o el contenido de floricina de una planta. Esto se puede lograr mediante la expresión de regulación negativa de los genes endógenos de la floretina glicosiltransferasa usando las secuencias de la floretina glicosiltransferasa, o fragmentos de las mismas, desveladas en el presente documento. Los métodos de regulación negativa o silenciamiento de genes endógenos en plantas son bien conocidos para los expertos en la materia y se describen el presente documento.

En un aspecto adicional, la divulgación describe un método de selección de una planta con la actividad de la floretina glicosiltransferasa alterada, método que comprende ensayar una planta en cuanto a la expresión alterada de un polipéptido descrito en el presente documento.

En un aspecto adicional, la divulgación describe un método de selección de una planta con la actividad de la floretina glicosiltransferasa alterada, método que comprende ensayar una planta en cuanto a la expresión alterada de un polipéptido descrito en el presente documento.

En un aspecto adicional, la divulgación describe un método de selección de una planta con contenido alterado de floricina; método que comprende ensayar una planta en cuanto a la expresión alterada de un polinucleótido o un polipéptido descrito en el presente documento.

En un aspecto adicional, la divulgación describe un método de selección de una planta con contenido alterado de floricina; método que comprende ensayar una planta en cuanto a la expresión alterada de un polipéptido o un polipéptido descrito en el presente documento.

En un aspecto adicional, la divulgación describe una célula vegetal o una planta producida mediante el método de la invención. Preferentemente, la planta se modifica genéticamente para incluir un polinucleótido o un polipéptido descrito en el presente documento.

En una realización preferida del método, la célula vegetal o la planta también se transforman con un polinucleótido que codifica una floretina sintasa.

En un aspecto adicional, la divulgación describe un grupo o una población de plantas seleccionados mediante el método descrito en el presente documento.

En un aspecto adicional, la divulgación describe un método de producción de floricina, método que comprende extraer floricina de una célula vegetal o de una planta descritas en el presente documento.

En un aspecto adicional, la invención proporciona un anticuerpo generado contra un polipéptido descrito en el presente documento.

En un aspecto adicional, la divulgación describe un método de producción de floricina, método que comprende poner en contacto floretina con el producto de expresión de una construcción de expresión que comprende un polinucleótido descrito en el presente documento o un polipéptido descrito en el presente documento para obtener floricina.

Los polinucleótidos y las variantes de polinucleótidos se pueden obtener de cualquier especie. Los polinucleótidos y las variantes también se pueden producir de forma recombinante y también pueden ser los productos de metodologías de "transposición de genes".

En una realización, el polinucleótido o la variante se obtienen de una especie vegetal.

En una realización adicional, el polinucleótido o la variante se obtienen de una especie de planta gimnosperma.

En una realización adicional, el polinucleótido o la variante se obtienen de una especie de planta angiosperma.

En una realización adicional, el polinucleótido o la variante se obtienen de especies de plantas dicotiledóneas.

Los polipéptidos y las variantes de polipéptidos se pueden obtener de cualquier especie. Los polipéptidos y las variantes también se pueden producir de forma recombinante y también se puede expresar a partir de productos de metodologías de "transposición de genes".

En una realización, los polipéptidos o las variantes de la invención se obtienen de especies vegetales.

En una realización adicional, los polipéptidos o las variantes se obtienen de especies de plantas gimnospermas.

En una realización adicional, los polipéptidos o las variantes se obtienen de especies de plantas angiospermas.

En una realización adicional, los polipéptidos o las variantes se obtienen de especies de plantas dicotiledóneas.

Las células vegetales y las plantas, incluyendo aquellas de las que se obtienen los polinucleótidos, las variantes de polinucleótidos, el polipéptido y las variantes de polipéptidos, pueden ser de cualquier especie.

En una realización, las células vegetales y las plantas son de especies gimnospermas.

En una realización adicional, las células vegetales y las plantas son de especies angiospermas.

En una realización adicional, las células vegetales y las plantas son de especies dicotiledóneas.

En una realización adicional, las células vegetales y las plantas de la invención son de especies monocotiledóneas.

En una realización preferida, las células vegetales y las plantas son de una especie seleccionada de un grupo que comprende, pero sin limitación, los siguientes géneros: *Smilax*, *Lithocarpus*, *Fragaria* y *Malus*.

Las especies vegetales preferidas son: *Smilax glyciphylla, Lithocarpus polystachyus, Fragaria, Malus domestica* y *Malus sieboldii.*

Un género particularmente preferido es *Malus*.

Las especies de *Malus* preferidas incluyen: *Malus aldenhamii Malus angustifolia, Malus asiatica, Malus baccata, Malus coronaria, Malus domestica, Malus doumeri, Malus florentina, Malus floribunda, Malus fusca, Malus halliana, Malus honanensis, Malus hupehensis, Malus ioensis, Malus kansuensis, Malus mandshurica, Malus micromalus, Malus niedzwetzkyana, Malus ombrophilia, Malus orientalis, Malus prattii, Malus prunifolia, Malus pumila, Malus sargentii, Malus sieboldii, Malus sieversii, Malus sylvestris, Malus toringoides, Malus transitoria, Malus trilobata, Malus tschonoskii, Malus x domestica, Malus x domestica x Malus sieversii, Malus sylvestris, Malus x domestica x Pyrus communis, Malus xiaojinensis, Malus yunnanensis, Malus sp., Mespilus germanica.*

La especie vegetal particularmente preferida es *Malus domestica*.

Descripción detallada

En la presente memoria descriptiva, cuando se hace referencia a especificaciones de patentes, a otros documentos externos o a otras fuentes de información, en general, se hace con el fin de proporcionar un contexto para la descripción de las características de la invención. A menos que se indique específicamente lo contrario, la referencia a dichos documentos externos no se debe interpretar como una admisión de que dichos documentos o dichas fuentes de información, en cualquier jurisdicción, sean de la técnica anterior o formen parte del conocimiento general común en la técnica.

La expresión "que comprende" como se usa en la presente memoria descriptiva significa "que consiste al menos en parte en". En la interpretación de cada afirmación de la presente memoria descriptiva que incluya la expresión "que comprende", también pueden estar presentes características distintas o aquellas introducidas por la expresión. Los términos relacionados tales como "comprenden" y "comprende" se han de interpretar de la misma manera.

Polinucleótidos y fragmentos

El término "polinucleótido/s", como se usa en el presente documento, significa un polímero desoxirribonucleótido o ribonucleótido monocatenario o bicatenario de cualquier longitud, pero preferentemente de al menos 15 nucleótidos e incluye como ejemplos no limitantes secuencias codificantes y no codificantes de una gen, complementos de secuencias sentido y antisentido, exones, intrones, ADN genómico, ADNc, pre-ARNm, ARNm, ARNr, ARNsi, ARNmi, ARNt, ribozimas, polipéptidos recombinantes, secuencias de ADN o ARN naturales aisladas y purificadas, secuencias de ARN y ADN sintéticas, sondas de ácido nucleico, cebadores y fragmentos.

Un "fragmento" de una secuencia de polinucleótidos proporcionada en el presente documento es una subsecuencia de nucleótidos contiguos que es capaz de hibridarse de manera específica a una diana de interés, por ejemplo, una secuencia que es de al menos 15 nucleótidos de longitud. Los fragmentos descritos en el presente documento comprenden 15 nucleótidos, preferentemente al menos 20 nucleótidos, más preferentemente al menos 30 nucleótidos, más preferentemente al menos 50 nucleótidos, más preferentemente al menos 50 nucleótidos y lo más preferentemente al menos 60 nucleótidos de nucleótidos contiguos de un polinucleótido descrito en el presente documento. Un fragmento de una secuencia de polinucleótidos se puede usar en tecnología antisentido, de silenciamiento génico, de triple hélice o de ribozimas, o como un cebador, una sonda, incluido en una micromatriz o usado en métodos de selección basados en los polinucleótidos descritos en el presente documento.

El término "cebador" se refiere a un polinucleótido corto que normalmente tiene un grupo 3'OH libre que se hibrida con un molde y se usa para cebar la polimerización de un polinucleótido complementario a la diana.

El término "sonda" se refiere a un polinucleótido corto que se usa para detectar una secuencia de polinucleótidos que es complementaria a la sonda en un ensayo basado en la hibridación. La sonda puede consistir en un "fragmento" de un polinucleótido como se define en el presente documento.

Polipéptidos y fragmentos

El término "polipéptido", como se usa en el presente documento, abarca cadenas de aminoácidos de cualquier longitud, pero preferentemente de al menos 5 aminoácidos, incluyendo proteínas de longitud completa, en las que los restos de aminoácidos están ligados por enlaces peptídicos covalentes. Los polipéptidos pueden ser productos naturales purificados o se pueden producir parcial o totalmente usando técnicas recombinantes o sintéticas. El término se puede referir a un polipéptido, un agregado de un polipéptido tal como un dímero u otro multímero, un polipéptido de fusión, un fragmento de polipéptido, una variante de polipéptido o un derivado del mismo.

Un "fragmento" de un polipéptido es una subsecuencia del polipéptido que realiza una función que es necesaria para la actividad biológica y/o proporciona la estructura tridimensional del polipéptido. El término se puede referir a un polipéptido, un agregado de un polipéptido tal como un dímero u otro multímero, un polipéptido de fusión, un fragmento de polipéptido, una variante de polipéptido o un derivado del mismo capaz de realizar la anterior actividad enzimática.

El término "aislado" como se aplica a las secuencias de polinucleótidos o polipéptidos desvelados en el presente documento se usa para referirse a secuencias que se retiran de su entorno celular natural. Una molécula aislada se puede obtener mediante cualquier método o combinación de métodos, incluyendo técnicas bioquímicas, recombinantes y sintéticas.

El término "recombinante" se refiere a una secuencia de polinucleótidos que se retira de las secuencias que la rodean en su contexto natural y/o se recombina con secuencias que no están presentes en su contexto natural.

Una secuencia de polipéptido "recombinante" se produce mediante la traducción de una secuencia de polinucleótidos "recombinante".

La expresión "derivado de" con respecto a los polinucleótidos o los polipéptidos de la invención que se obtienen de determinados géneros o especies significa que el polinucleótido o el polipéptido tiene la misma secuencia que un polinucleótido o un polipéptido que se encuentra de manera natural en esos géneros o especies. Por lo tanto, el polinucleótido o polipéptido, derivado de determinados géneros o especies, se puede producir de manera sintética o recombinante.

Variantes

Como se usa en el presente documento, el término "variante" se refiere a secuencias de polinucleótidos o polipéptidos diferentes de las secuencias identificadas específicamente, en las que uno o más nucleótidos o restos de aminoácidos se han eliminado, sustituido o añadido. Las variantes pueden ser variantes alélicas naturales o variantes no naturales. Las variantes pueden ser de la misma o de otras especies, y pueden englobar homólogos, parálogos y ortólogos. En ciertas realizaciones, las variantes de los polipéptidos y los polipéptidos de la invención poseen actividades biológicas que son iguales o similares a las de los polipéptidos o los polipéptidos de la invención. El término "variante" con referencia a los polipéptidos y los polipéptidos engloba todas las formas de polipéptidos y polipéptidos definidas en el presente documento.

Variantes de polinucleótidos

Las secuencias de variantes de polinucleótidos presentan preferentemente al menos el 50 %, más preferentemente al menos el 51 %, más preferentemente al menos el 52 %, más preferentemente al menos el 53 %, más preferentemente al menos el 54 %, más preferentemente al menos el 55 %, más preferentemente al menos el 56 %, más preferentemente al menos el 57 %, más preferentemente al menos el 58 %, más preferentemente al menos el 59 %, más preferentemente al menos el 60 %, más preferentemente al menos el 61 %, más preferentemente al menos el 62 %, más preferentemente al menos el 63 %, más preferentemente al menos el 64 %, más preferentemente al menos el 65 %, más preferentemente al menos el 66 %, más preferentemente al menos el 67 %, más preferentemente al menos el 68 %, más preferentemente al menos el 69 %, más preferentemente al menos el 70 %, más preferentemente al menos el 71 %, más preferentemente al menos el 72 %, más preferentemente al menos el 73 %, más preferentemente al menos el 74 %, más preferentemente al menos el 75 %, más preferentemente al menos el 76 %, más preferentemente al menos el 77 %, más preferentemente al menos el 78 %, más preferentemente al menos el 79 %, más preferentemente al menos el 80 %, más preferentemente al menos el 81 %, más preferentemente al menos el 82 %, más preferentemente al menos el 83 %, más preferentemente al menos el 84 %, más preferentemente al menos el 85 %, más preferentemente al menos el 86 %, más preferentemente al menos el 87 %, más preferentemente al menos el 88 %, más preferentemente al menos el 89 %, más preferentemente al menos el 90 %, más preferentemente al menos el 91 %, más preferentemente al menos el 92 %, más preferentemente al menos el 93 %, más preferentemente al menos el 94 %, más preferentemente al

menos el 95 %, más preferentemente al menos el 96 %, más preferentemente al menos el 97 %, más preferentemente al menos el 98 % y lo más preferentemente al menos el 99 % de identidad con una secuencia. La identidad se encuentra en una franja de comparación de al menos el 20 posiciones de nucleótidos, preferentemente de al menos 50 posiciones de nucleótidos, más preferentemente de al menos 100 posiciones de nucleótidos y lo más preferentemente a lo largo de todo un polinucleótido de la invención.

La identidad de las secuencias de polinucleótidos se puede determinar de la siguiente manera. Se compara la secuencia de polinucleótidos en cuestión con una secuencia de polinucleótidos candidata usando BLASTN (de la serie de programas BLAST, versión 2.2.5 [noviembre de 2002]) de bl2seq (Tatiana A. Tatusova, Thomas L. Madden (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", *FEMS Microbiol Lett.* 174:247-250), que se encuentra a disposición del público en NCBI (ftp://ftp.ncbi.nih.gov/blast/). Se utilizan los parámetros por defecto de bl2seq, salvo que se deba desactivar el filtrado de partes de baja complejidad.

La identidad de las secuencias de polinucleótidos se puede examinar usando los siguientes parámetros de línea de comandos de Unix:

bl2seq -i nucleotideseq1 -j nucleotideseq2 -FF -p blastn.

El parámetro -F F desactiva el filtrado de las secciones de baja complejidad. El parámetro -p selecciona el algoritmo apropiado para el par de secuencias. El programa bl2seq presenta la identidad de secuencias tanto en forma del número como del porcentaje de nucleótidos idénticos en una línea "Identidades =".

También se puede calcular la identidad de las secuencias de polinucleótidos a lo largo de toda una superposición entre una secuencia de polinucleótidos candidata y objeto usando programas de alineamiento global de secuencias (por ejemplo, Needleman, S. B. y Wunsch, C. D. (1970) *J. Mol. Biol*. 48, 443-453). En el programa Neddle del paquete EMBOSS, se encuentra una implementación completa del algoritmo de alineamiento global de Needleman-Wunsch (Rice, P. Longden, I. y Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite, Trends in Genetics, junio de 2000, vol 16, Nº 6. pág. 276-277) que se puede obtener en http://www.hgmp.mrc.ac.uk/Software/EMBOSS/. El servidor del Instituto Europeo de Bioinformática también ofrece la posibilidad de realizar alineamientos globales con Needle de EMBOSS entre dos secuencias en línea en http:/www.ebi.ac.uk/emboss/align/.

Como alternativa, se puede usar el programa GAP que calcula un alineamiento global óptimo de dos secuencias sin penalizar huecos terminales. GAP se describe en el siguiente documento: Huang, X. (1994) “On Global Sequence Alignment. Computer Applications in the Biosciences” 10, 227-235.

Un método preferido para el cálculo del % de identidad de secuencias de polinucleótidos se basa en el alineamiento de las secuencias que se van a comparar usando Clustal X (Jeanmougin *et al.*, 1998, *Trends Biochem. Sci.* 23, 403-5).

Las variantes de polinucleótidos también engloban aquellas que presentan una similitud con una o más de las secuencias identificadas específicamente que es probable que conserven la equivalencia funcional de esas secuencias y que no era razonable esperar que se produjeran al azar. Dicha similitud de secuencias con respecto a los polipéptidos se puede determinar usando el programa bl2seq a disposición del público de la serie de programas BLAST (versión 2.2.5 [noviembre de 2002]) del NCBI (ftp://ftp.ncbi.nih.gov/blast/).

La similitud de las secuencias de polinucleótidos se puede examinar usando los siguientes parámetros de línea de comandos de Unix:

bl2seq -i nucleotideseq1 -j nucleotideseq2 -F F -p tblastx.

El parámetro -F F desactiva el filtrado de las secciones de baja complejidad. El parámetro -p selecciona el algoritmo apropiado para el par de secuencias. Este programa encuentra regiones de similitud entre las secuencias, y para cada una de estas regiones presenta un "valor de E", que es el número esperado de veces que cabría esperar observar dicha coincidencia al azar en una base de datos de un tamaño fijo de referencia que contiene secuencias aleatorias. El tamaño de esta base de datos se establece por defecto en el programa bl2seq. Para los valores bajos de E, muy inferiores a uno, el valor de E es aproximadamente la probabilidad de dicha coincidencia aleatoria.

Las variantes de secuencias de polinucleótidos presentan preferentemente un valor de E inferior a 1×10^{-6}, más preferentemente inferior a 1×10^{-9}, más preferentemente inferior a 1×10^{-12}, más preferentemente inferior a 1×10^{-15}, más preferentemente inferior a 1×10^{-18}; más preferentemente inferior a 1×10^{-21}, más preferentemente inferior a 1×10^{-30}, más preferentemente inferior a 1×10^{-40}, más preferentemente inferior a 1×10^{-50}; más preferentemente inferior a 1×10^{-60}, más preferentemente inferior a 1×10^{-70}; más preferentemente inferior a 1×10^{-80}; más preferentemente inferior a 1×10^{-90} y lo más preferentemente inferior a 1×10^{-100} en comparación con una cualquiera de las secuencias identificadas específicamente.

Como alternativa, las variantes de polinucleótidos se hibridan a las secuencias de polinucleótidos especificadas o a complementos de las mismas en condiciones rigurosas.

La expresión "hibridar en condiciones rigurosas", y los equivalentes gramaticales de la misma, se refieren a la capacidad de una molécula de polinucleótido para hibridarse a una molécula de polinucleótido diana (tal como una molécula de polinucleótido diana inmovilizada en una transferencia de ADN o ARN, tal como una transferencia Southern o una transferencia Northern) en condiciones definidas de temperatura y concentración de sal. La capacidad para hibridarse en condiciones de hibridación rigurosas se puede determinar mediante la hibridación inicial en condiciones menos rigurosas, y aumentando después la rigurosidad hasta la rigurosidad deseada.

Con respecto a moléculas de polinucleótido superiores a aproximadamente 100 bases de longitud, las condiciones de hibridación rigurosas típicas no son superiores 25 a 30 ºC (por ejemplo, 10 ºC) por debajo de la temperatura de fusión (Tf) del híbrido natural (véase, en general, Sambrook *et al.*, Eds, 1987, “Molecular Cloning, a Laboratory Manual”, 2ª Ed. Cold Spring Harbor Press; Ausubel *et al.*, 1987, “Current Protocols in Molecular Biology”, Greene Publishing) . Se puede calcular la Tf para las moléculas de polinucleótido superiores a aproximadamente 100 bases mediante la fórmula Tf = 81,5 + 0,41 % (G + C-log (Na+). (Sambrook *et al.*, Eds, 1987, “Molecular Cloning, a Laboratory Manual”, 2ª Ed. Cold Spring Harbor Press; Bolton y McCarthy, 1962, PNAS 84: 1390). Las condiciones rigurosas típicas para un polinucleótido superior a 100 bases de longitud serían condiciones de hibridación tales como prelavado en una solución de SSC x 6, SDS al 0,2 %; hibridación a 65 ºC, SSC x 6, SDS al 0,2 % durante la noche; seguido de dos lavados de 30 minutos cada uno en SSC x 1, SDS al 0,1 % a 65 ºC y dos lavados de 30 minutos cada uno en SSC x 0,2, SDS al 0,1 % a 65 ºC.

Con respecto a moléculas de polinucleótido con una longitud inferior a 100 bases, las condiciones de hibridación rigurosas ilustrativas son de 5 a 10 ºC por debajo de Tf. Como media, la Tf de una molécula de polinucleótido de longitud inferior a 100 pb se reduce en aproximadamente (500/longitud del oligonucleótido) ºC.

Con respecto a los miméticos de ADN conocidos como ácidos nucleicos peptídicos (PNA) (Nielsen *et al., Science*. 6 de diciembre de 1991; 254 (5037): 1497-500), los valores de Tf son superiores a los de los híbridos de ADN-ADN o ADN-ARN, y se pueden calcular usando la fórmula descrita en Giesen *et al., Nucleic Acids Res*. 1 de noviembre de 1998; 26 (21): 5004-6. Las condiciones de hibridación rigurosas ilustrativas para un híbrido ADN-PNA con una longitud inferior a 100 bases son de 5 a 10 ºC por debajo de la Tf.

Las variantes de polinucleótidos también engloban polinucleótidos que difieren de las secuencias desveladas en el presente documento, pero que, como consecuencia de la degeneración del código genético, codifican un polipéptido que tiene actividad similar a un polipéptido codificado por un polinucleótido de la presente invención. Una modificación de la secuencia que no cambia la secuencia de aminoácidos del polipéptido es una "variación silenciosa". A excepción de ATG (metionina) y TGG (triptófano), se pueden cambiar otros condones para el mismo aminoácido mediante técnicas reconocidas en la materia, por ejemplo, para optimizar la expresión de condones en un determinado organismo hospedador.

Las modificaciones de la secuencia de polinucleótidos que generan sustituciones conservadoras de uno o varios aminoácidos en la secuencia polipeptídica codificada sin modificar de manera significativa su actividad biológica también se incluyen en la invención. El experto en la materia conocerá los métodos de preparación de sustituciones de aminoácidos fenotípicamente silenciosas (véase, por ejemplo, Bowie *et al.*, 1990, *Science* 247, 1306).

Las variantes de polinucleótidos debido a variaciones silenciosas y sustituciones conservadoras en la secuencia polipeptídica codificada se pueden determinar usando el programa bl2seq que se encuentra a disposición del público de la serie de programas BLAST (versión 2.2.5 [noviembre de 2002]) de NCBI (ftp://ftp.ncbi.nih.gov/blast/) a través del algoritmo tblastx como se ha descrito anteriormente.

La función de una variante de polinucleótido como una floretina glicosiltransferasa se puede evaluar, por ejemplo, expresando dicha secuencia en bacterias y ensayando la actividad de la proteína codificada como se describe en el apartado de ejemplos. También se puede ensayar la capacidad de una función de una variante para modificar la actividad de la floretina glicosiltransferasa o el contenido de floretina de las plantas, también como se describe en el apartado de ejemplos del presente documento.

Variantes de polipéptidos

El término "variante" con referencia a los polipéptidos engloba polipéptidos naturales, producidos de manera recombinante y sintética. Las secuencias de variantes de polipéptidos presentan preferentemente al menos el 50%, más preferentemente al menos el 51%, más preferentemente al menos el 52%, más preferentemente al menos el 53%, más preferentemente al menos el 54%, más preferentemente al menos el 55%, más preferentemente al menos el 56%, más preferentemente al menos el 57%, más preferentemente al menos el 58%, más preferentemente al menos el 59%, más preferentemente al menos el 60%, más preferentemente al menos el 61%, más preferentemente al menos el 62%, más preferentemente al menos el 63%, más preferentemente al menos el 64%, más preferentemente al menos el 65%, más preferentemente al menos el 66%, más preferentemente al menos el 67%,

más preferentemente al menos el 68%, más preferentemente al menos el 69%, más preferentemente al menos el 70%, más preferentemente al menos el 71%, más preferentemente al menos el 72%, más preferentemente al menos el 73%, más preferentemente al menos el 74%, más preferentemente al menos el 75%, más preferentemente al menos el 76%, más preferentemente al menos el 77%, más preferentemente al menos el 78%, más preferentemente al menos el 79%, más preferentemente al menos el 80%, más preferentemente al menos el 81%, más preferentemente al menos el 82%, más preferentemente al menos el 83%, más preferentemente al menos el 84%, más preferentemente al menos el 85%, más preferentemente al menos el 86%, más preferentemente al menos el 87%, más preferentemente al menos el 88%, más preferentemente al menos el 89%, más preferentemente al menos el 90%, más preferentemente al menos el 91%, más preferentemente al menos el 92%, más preferentemente al menos el 93%, más preferentemente al menos el 94%, más preferentemente al menos el 95%, más preferentemente al menos el 96%, más preferentemente al menos el 97%, más preferentemente al menos el 98%, y lo más preferentemente al menos el 99% de identidad con secuencias de la presente invención. Se encuentra identidad en una franja de comparación de al menos 20 posiciones de aminoácidos, preferentemente al menos 50 posiciones de aminoácidos, más preferentemente al menos 100 posiciones de aminoácidos y lo más preferentemente en toda la longitud de un polipéptido.

Se puede determinar la identidad de secuencias de polipéptidos de la siguiente manera. Se compara la secuencia de polipéptidos objeto con una secuencia de polipéptidos candidata usando BLASTP (de la serie de programas BLAST, versión 2.2.5 [noviembre de 2002]) en bl2seq, que se encuentra a disposición del público en NCBI (ftp://ffp.ncbi.nih.gov/blast/). Se utilizan los parámetros por defecto de bl2seq, excepto que se deba desactivar la filtración de las regiones de baja complejidad.

La identidad de secuencias de polinucleótidos también se puede calcular a lo largo de toda la superposición entre las secuencias de polinucleótidos candidata y objeto usando programas de alineamiento global de secuencias. Needle de EMBOSS (disponible en http:/www.ebi.ac.uk/emboss/align/) y GAP (Huang, X. (1994) “On Global Sequence Alignment. Computer Applications in the Biosciences”, 10, 227-235) como se ha descrito anteriormente también son programas de alineamiento global de secuencias adecuados para el cálculo de la identidad de secuencias de polipéptidos.

Un método preferido para calcular el % de identidad de secuencias de los polipéptidos se basa en el alineamiento de las secuencias que se van a comparar usando Clustal X (Jeanmougin et al., 1998, *Trends Biochem. Sci.* 23, 403-5).

Las variantes de polipéptidos también engloban aquéllas que presentan una similitud que presentan una similitud con una o más de las secuencias identificadas específicamente que es probable que conserven la equivalencia funcional de esas secuencias y que no era razonable esperar que se produjeran al azar. Dicha similitud de secuencias con respecto a los polipéptidos se puede determinar usando el programa bl2seq a disposición del público de la serie de programas BLAST (versión 2.2.5 [noviembre de 2002]) del NCBI (ftp://ftp.ncbi.nih.gov/blast/). La similitud de las secuencias de polipéptidos se puede examinar usando los siguientes parámetros de línea de comandos de Unix:

bl2seq -i peptideseq1 -j peptideseq2 -F -p blastp.

Las variantes de secuencias de polipéptidos presentan preferentemente un valor de E inferior a 1×10^{-6}, más preferentemente inferior a 1×10^{-9}, más preferentemente inferior a 1×10^{-12}, más preferentemente inferior a 1×10^{-15}, más preferentemente inferior a 1×10^{-18}; más preferentemente inferior a 1×10^{-21}, más preferentemente inferior a 1×10^{-30}, más preferentemente inferior a 1×10^{-40}, más preferentemente inferior a 1×10^{-50}; más preferentemente inferior a 1×10^{-60}, más preferentemente inferior a 1×10^{-70}; más preferentemente inferior a 1×10^{-80}; más preferentemente inferior a 1×10^{-90} y lo más preferentemente inferior a 1×10^{-100} en comparación con una cualquiera de las secuencias identificadas específicamente.

El parámetro -F F desactiva el filtrado de las secciones de baja complejidad. El parámetro -p selecciona el algoritmo apropiado para el par de secuencias. Este programa encuentra regiones de similitud entre las secuencias, y para cada una de estas regiones presenta un "valor de E", que es el número esperado de veces que cabría esperar observar dicha coincidencia al azar en una base de datos de un tamaño fijo de referencia que contiene secuencias aleatorias. Para los valores bajos de E, muy inferiores a uno, es aproximadamente la probabilidad de dicha coincidencia aleatoria.

También se incluyen las sustituciones conservadoras de uno o varios aminoácidos de una secuencia de polipéptidos descrita sin modificar de manera significativa su actividad biológica. El experto en la materia conocerá los métodos de preparación de sustituciones de aminoácidos fenotípicamente silenciosas (véase, por ejemplo, Bowie *et al.,* 1990, *Science* 247, 1306).

La función de una variante de polipéptido como una floretina glicosiltransferasa se puede evaluar mediante los métodos descritos en el apartado de ejemplos del presente documento.

Construcciones, vectores y componentes de los mismos

La expresión "construcción genética" se refiere a una molécula de polinucleótido, normalmente ADN bicatenario, que puede tener insertada otra molécula de polinucleótido (la molécula de polinucleótido insertada) tal como, pero sin limitación, una molécula de ADNc. Una construcción genética puede contener los elementos necesarios que permitan transcribir la molécula de polinucleótido insertada y, opcionalmente, traducir la transcripción a un polipéptido. La molécula de polinucleótido insertada puede proceder de la célula hospedadora o puede proceder de una célula o un organismo diferente y/o puede ser un polinucleótido recombinante. Una vez que se encuentra en el interior de la célula hospedadora, la construcción genética puede llegar a integrarse en el ADN cromosómico del hospedador. La construcción genética se puede ligar a un vector.

El término "vector" se refiere a una molécula de polinucleótido, normalmente ADN bicatenario, que se usa para transportar la construcción genética a una célula huésped. El vector puede ser capaz de replicarse en al menos un sistema hospedador adicional, tal como *E. coli*.

La expresión "construcción de expresión" se refiere a una construcción genética que incluye los elementos necesarios que permiten la transcripción de la molécula de polinucleótidos insertada y, opcionalmente, la traducción de la transcripción a un polipéptido. Por lo general, una construcción de expresión comprende en sentido de 5' a 3':

a) un promotor funcional en la célula hospedadora en la que se transformará la construcción;
b) el polinucleótido que se va a expresar y
c) un terminador funcional en la célula hospedadora en que se transformará la construcción.

La expresión "región codificante" o "marco de lectura abierto" (ORF) se refiere a cadena sentido de una secuencia de ADN genómico o una secuencia de ADNc que es capaz de producir un producto de transcripción y/o un polipéptido bajo el control de secuencias reguladoras apropiadas. La secuencia codificante se identifica por la presencia de un codón de inicio de la traducción 5’ y un codón de parada de la traducción 3’. Cuando se inserta en una construcción genética, una "secuencia codificante" es capaz de expresarse cuando se une de forma operativa a secuencias promotoras y de terminación.

"Unido de forma operativa" significa que la secuencia que se va a expresar se pone bajo el control de elementos reguladores que incluyen promotores, elementos reguladores específicos del tejido, elementos reguladores temporales, potenciadores, represores y terminadores.

La expresión "región no codificante" se refiere a secuencias no traducidas que se encuentran secuencia arriba del sitio de inicio de la traducción y secuencia abajo del sitio de parada de la traducción. Estas secuencias también se denominan respectivamente la 5' UTR y la 3' UTR. Estas regiones incluyen los elementos necesarios para el inicio y la terminación de la transcripción y para la regulación de la eficacia de la traducción.

Los terminadores son secuencias, que finalizan la transcripción, y se encuentran en los extremos no traducidos 3’ de los genes secuencia abajo de la secuencia traducida. Los terminadores son determinantes importantes de la estabilidad del ARNm y, en algunos casos, han resultado presentar funciones reguladoras espaciales.

El término "promotor" se refiere a elementos reguladores *cis* no transcritos secuencia arriba de la región codificante que regula la transcripción de genes. Los promotores comprenden elementos iniciadores *cis* que especifican el sitio de inicio de la transcripción y conservan cajas tales como la caja TATA y motivos que son unidos por factores de transcripción.

Un "transgén" es un polinucleótido que se toma de un organismo y se introduce en un organismo diferente por transformación. El transgén puede proceder de la misma especie o de una especie diferente como la especie del organismo en el que se introduce el transgén.

Una "repetición invertida" es una secuencia que se repite, donde la segunda mitad de la repetición está en la cadena complementaria, por ejemplo,

(5') GATCTA……TAGATC (3')
(3') CTAGAT……ATCTAG (5').

La transcripción por lectura producirá una transcripción que se somete a apareamiento de bases complementarias para formar una estructura en horquilla, siempre que haya un espaciador de 3-5 pb entre las regiones repetidas.

Células hospedadoras

Las células hospedadoras se pueden obtener, por ejemplo, de organismos bacterianos, micóticos, insectos, mamíferos o vegetales.

Una "planta transgénica" se refiere a una planta que contiene material genético nuevo como resultado de la manipulación o la transformación genética. El material genético nuevo se puede obtener de una planta de la misma especie que la planta transgénica resultante o de una especie diferente.

Los solicitantes han identificado nuevos polinucleótidos (SEC ID Nº 6 a 15) que codifican nuevos polipéptidos (SEC ID Nº 2 a 5) que tienen actividad de la floretina glicosiltransferasa tales como los resumidos en la Tabla 1.

Los solicitantes han demostrado que todas las secuencias de polipéptidos de floretina glicosiltransferasa desveladas (SEC ID Nº 1 a 5) muestran una importante conservación de secuencias y son variantes entre sí.

Del mismo modo, los solicitantes han demostrado que todas las secuencias de polinucleótidos de floretina glicosiltransferasa desveladas (SEC ID Nº 6 a 15) muestran una importante conservación de secuencias y son variantes entre sí.

La divulgación describe construcciones genéticas, vectores y plantas que contienen las nuevas secuencias de polinucleótidos (SEC ID Nº 6 a 15) o secuencias que codifican las nuevas secuencias de polipéptidos (SEC ID Nº 2 a 5). La divulgación también describe plantas que comprenden la construcción genética y los vectores de la invención.

La invención proporciona plantas con la actividad de la floretina glicosiltransferasa modificada con respecto a plantas de control adecuadas, y plantas con el contenido de floricina modificado con respecto a plantas de control adecuadas. La invención proporciona plantas con una mayor actividad de la floretina glicosiltransferasa y un mayor contenido de floricina.

La divulgación también describe métodos de producción de dichas plantas y métodos de selección de dichas plantas.

Las plantas de control adecuadas incluyen plantas no transformadas de la misma especie o variedad, o plantas transformadas con construcciones de control.

Métodos de aislamiento o producción de polinucleótidos

Las moléculas de polinucleótido se pueden aislar usando una variedad de técnicas conocidas para los expertos en la materia. A modo de ejemplo, dichos polipéptidos se pueden aislar mediante el uso de la reacción en cadena de la polimerasa (PCR) descrito en Mullis *et al.,* Eds. 1994, “The Polimerase Chain Reaction”, Birkhauser. Los polipéptidos se pueden amplificar usando cebadores, como se define en el presente documento, procedentes de las secuencias de polinucleótidos descritas en el presente documento.

Otros métodos adicionales de aislamiento de polinucleótidos incluyen el uso de todos, o partes de, los polipéptidos que tienen la secuencia expuesta en el presente documento como sondas de hibridación. Se puede usar la técnica de hibridación de sondas de polinucleótidos marcadas a polinucleótidos inmovilizados en soportes sólidos, tales como filtros de nitrocelulosa o membranas de nailon, para rastrear bibliotecas genómicas o de ADNc. Las condiciones de hibridación y de lavado ilustrativas son: hibridación durante 20 horas a 65 ºC en SSC x 5,0, dodecilsulfato de sodio al 0,5 %, disolución de Denhardt x 1; lavado (tres lavados de veinte minutos cada uno a 55 ºC) en SSC x 1,0, dodecilsulfato de sodio al 1 % (p/v) y, opcionalmente, un lavado (durante veinte minutos) en SSC x 0,5, dodecilsulfato de sodio al 1 % (p/v), a 60 ºC. Se puede realizar un lavado adicional opcional (durante veinte minutos) en las condiciones de SSC x 0, 1, dodecilsulfato de sodio al 1 % (p/v) a 60 ºC.

Se pueden producir fragmentos de polinucleótidos mediante técnicas conocidas en la materia tales como la digestión con endonucleasas de restricción, la síntesis de oligonucleótidos y la amplificación por PCR.

Se puede usar una secuencia de polinucleótidos parcial en métodos conocidos en la materia para identificar la correspondiente secuencia de polinucleótidos de longitud completa. Dichos métodos incluyen métodos basados en PCR, 5'RACE (Frohman MA, 1993, *Methods Enzymol.* 218: 340-56) y el método basado en hibridación, métodos basados en ordenador/bases de datos. Además, a modo de ejemplo, la PCR inversa permite la adquisición de secuencias desconocidas, que flanquean las secuencias de polinucleótidos desveladas en el presente documento, partiendo de cebadores basados en una región conocida (Triglia *et al.,* 1998, *Nucleic Acids Res* 16, 8186). El método usa varias enzimas de restricción para generar un fragmento adecuado en la región conocida de un gen. A continuación, se circulariza el fragmento mediante ligadura intramolecular y se usa como un molde de PCR. Se diseñan cebadores divergentes de la región conocida. Para ensamblar físicamente clones de longitud completa, se pueden utilizar metodologías convencionales de biología molecular (Sambrook *et al*., “Molecular Cloning: A Laboratory Manual”, 2ª Ed. Cold Spring Harbor Press, 1987).

Puede ser beneficioso, cuando se produce una planta transgénica de una determinada especie, transformar dicha planta con una secuencia o secuencias derivadas de esa especie. El beneficio puede ser reducir la preocupación de algunas personas con respecto a la transformación de especies cruzadas en la generación de organismos transgénicos. Además, cuando se desea la regulación negativa de un gen, puede ser necesario utilizar una

secuencia idéntica (o al menos muy similar) a la de la planta para la que se desea una expresión reducida. Por esta razón, entre otras, es deseable poder identificar y aislar ortólogos de un determinado gen en varias especies vegetales diferentes.

Las variantes (incluyendo los ortólogos) se pueden identificar mediante los métodos descritos.

Métodos de identificación de variantes

Métodos físicos

Se pueden identificar variantes de polipéptidos usando métodos basados en PCR (Mullis *et al.*, Eds. 1994, “The Polimerase Chain Reaction”, Birkhauser). Por lo general, la secuencia de polinucleótidos de un cebador, útil para amplificar variantes de moléculas de polinucleótidos por PCR, se puede basar en una secuencia que codifique una región conservada de la correspondiente secuencia de aminoácidos.

Como alternativa, se pueden emplear métodos de exploración de bibliotecas conocidos por los expertos en la materia (Sambrook *et al.,* “Molecular Cloning: A Laboratory Manual”, 2ª Ed. Cold Spring Harbor Press, 1987). Cuando se identifican variantes de la secuencia de la sonda, normalmente se disminuirá la rigurosidad de la hibridación y/o del lavado con respecto a cuando se buscan coincidencias de secuencias exactas.

Las variantes de polipéptidos también se pueden identificar mediante métodos físicos, por ejemplo, mediante la exploración de bibliotecas de expresión usando anticuerpos generados contra polipéptidos (Sambrook *et al*., “Molecular Cloning: A Laboratory Manual”, 2ª Ed. Cold Spring Harbor Press, 1987) o mediante la identificación de polipéptidos de fuentes naturales con la ayuda de dichos anticuerpos.

Métodos informáticos

Las secuencias de variantes, incluyendo variantes tanto de polinucleótidos como de polipéptidos, también se pueden identificar mediante métodos informáticos conocidos para los expertos en la materia, usando algoritmos de alineamiento de secuencias de dominio público y herramientas de búsqueda de similitud de secuencias para buscar bases de datos de secuencias (las bases de datos de dominio público incluyen Genbank, EMBL, Swiss-Prot, PIR y otras). Véase, por ejemplo, *Nucleic Acids Res.* 29: 1-10 y 11-16, 2001 para consultar ejemplos de recursos en línea. Las búsquedas de similitud recuperan y alinean secuencias diana para compararlas con una secuencia que se vaya a analizar (es decir, una secuencia de consulta). Los algoritmos de comparación de secuencias usan matrices de puntuación para asignar una puntuación global a cada uno de los alineamientos.

Un ejemplo de familia de programas útil para identificar variantes en bases de datos de secuencias es la serie de programas BLAST (versión 2.2.5 [noviembre de 2002]) que incluye BLASTN, BLASTP, BLASTX, tBLASTN y tBLASTX, que se encuentran a disposición del público en (ftp://ftp.ncbi.nih.gov/blast/) o en el Centro Nacional de Información de Biotecnología (NCBI), la Biblioteca Nacional de Medicina, Edificio 38A, Sala 8N805, Betesda, MD 20894, EE.UU. El servidor del NCBI también permite el uso de los programas para explorar una serie de bases de datos de secuencias a disposición del público. BLASTN compara una secuencia de consulta de nucleótidos frente a una base de datos de secuencias de nucleótidos. BLASTP compara una secuencia de consulta de aminoácidos frente a una base de datos de secuencias de proteínas. BLASTX compara una secuencia de consulta de nucleótidos traducida en todos los marcos de lectura frente a una base de datos de secuencias de proteínas. El programa tBLASTN compara una secuencia de consulta de proteínas frente a una base de datos de secuencias de nucleótidos traducida dinámicamente en todos los marcos de lectura. El programa tBLASTX compara las traducciones de seis marcos de una secuencia de consulta de nucleótidos frente a las traducciones de seis marcos de una base de datos de secuencias de nucleótidos. Los programas BLAST se pueden usar con parámetros por defecto o los parámetros se pueden modificar según sea necesario para perfeccionar la exploración.

El uso de la familia de algoritmos BLAST, que incluye BLASTN, BLASTP y BLASTX, se describe en la publicación de Altschul *et al., Nucleic Acids Res*. 25: 3389-3402, 1997.

Los "éxitos" para una o más secuencias de la base de datos mediante una secuencia consultada producida por BLASTN, BLASTP, BLASTX, tBLASTN, tBLASTX, o un algoritmo similar, alinean o identifican partes similares de las secuencias. Los éxitos se disponen por orden del grado de similitud y longitud de la secuencia superpuesta. En general, los éxitos para una secuencia de la base de datos representan una superposición solo sobre una fracción de la longitud de secuencia de la secuencia consultada.

Los algoritmos BLASTN, BLASTP, BLASTX, tBLASTN y tBLASTX también producen valores "Esperados" para los alineamientos. El Valor esperado (E) indica el número de éxitos cuya observación se puede “esperar" al azar cuando se busca en una base de datos del mismo tamaño que contiene secuencias contiguas aleatorias. El valor Esperado se usa como un umbral de significación para determinar si el éxito para una base de datos indica una verdadera similitud. Por ejemplo, un valor E de 0,1 asignado a un éxito de polinucleótido se interpreta como que, en una base de datos del tamaño de la base de datos examinada, cabría esperar la observación de 0,1 coincidencias sobre la

parte alineada de la secuencia con una puntuación similar simplemente al azar. Para las secuencias que tienen un valor E de 0,01 o inferior sobre partes alineadas y coincidentes, la probabilidad de encontrar una coincidencia al azar en esa base de datos es del 1 % o inferior usando el algoritmo BLASTN, BLASTP, BLASTX, tBLASTN o tBLASTX.

Los alineamientos de secuencias múltiples de un grupo de secuencias relacionadas se pueden llevar a cabo con CLUSTALW (Thompson, J. D., Higgins, D. G. y Gibson, T. J. (1994) "CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matriz choice". *Nucleic Acids Research*, 22: 4673-4680, http://www-igbmc.u-strasbg.fr/BioInfo/ClustalW/Top.html) o T-COFFEE (Cedric Notredame, Desmond G. Higgins, Jaap Heringa, T-Coffee: "A novel metod for fast and accurate multiple sequence alignment", *J. Mol. Biol.* (2000) 302: 205-217)) o PILEUP, que usa alineamientos por parejas, progresivos. (Feng y Dolittle, 1987, *J. Mol. Evol.* 25, 351).
Hay aplicaciones de programas informáticos de reconocimiento de patrones disponibles para buscar motivos o secuencias firma. Por ejemplo, MEME (Em múltiple para obtención de motivos) encuentra motivos y secuencias firma en una serie de secuencias y MAST (alineamiento de motivos y herramienta de búsqueda) usa estos motivos para identificar motivos similares o los mismos motivos en secuencias consultadas. Los resultados de MAST se proporcionan como una serie de alineamientos con datos estadísticos apropiados y una visión general visual de los motivos encontrados. MEME y MAST se desarrollaron en la Universidad de California, San Diego.

PROSITE (Bairoch y Bucher, 1994, *Nucleic Acids Res*. 22, 3583; Hofmann *et al.*, 1999, *Nucleic Acids Res.* 27, 215) es un método de identificación de las funciones de proteínas no caracterizadas traducidas de secuencias genómicas o de secuencias de ADNc. La base de datos PROSITE (www.expasy.org/prosite) contiene patrones y perfiles biológicamente significativos, y está diseñada de modo que se puede usar con herramientas computacionales apropiadas para asignar una nueva secuencia a una familia conocida de proteínas o para determinar qué dominio/s conocido/s está/n presente/s en la secuencia (Falquet *et al.,* 2002, *Nucleic Acids Res*. 30, 235). Prosearch es una herramienta que puede buscar en las bases de datos SWISS-PROT y EMBL con un patrón o una firma de secuencia determinada.

La función de una variante de polinucleótido en la codificación de floretina glicosiltransferasas se puede ensayar en cuanto a la actividad, o se puede ensayar en cuanto a su capacidad para modificar el contenido de floricina de las plantas mediante métodos descritos en el apartado de ejemplos del presente documento.

Métodos de aislamiento de polipéptidos

Los polipéptidos, incluyendo las variantes de polipéptidos, se pueden preparar usando métodos de síntesis de péptidos conocidos en la técnica tales como la síntesis directa de péptidos usando técnicas de fase sólida (por ejemplo, Stewart *et al*., 1969, en "Solid-Phase Peptide Synthesis", W. H. Freeman Co, San Francisco California, o síntesis automatizada, por ejemplo, usando un sintetizador de péptidos Applied Byosystems 431A (Foster City, California). También se pueden producir formas mutadas de los polipéptidos durante dichas síntesis.

Los polipéptidos y las variantes de polipéptidos también se pueden purificar de fuentes naturales usando una variedad de técnicas que son conocidas en la materia (por ejemplo, Deutscher, 1990, Ed, *Methods in Enzymology*, Vol. 182, "Guide to Protein Purification").

Como alternativa, los polipéptidos y las variantes de polipéptidos se pueden expresar de manera recombinante en células hospedadoras adecuadas y separarlos de las células como se describe a continuación.

Métodos de producción de construcciones y vectores

Las construcciones genéticas comprenden una o más secuencias de polinucleótidos de la invención y/o polinucleótidos que codifican polipéptidos de la invención, y pueden ser útiles para transformar, por ejemplo, organismos bacterianos, micóticos, de insectos, mamíferos u vegetales. Las construcciones genéticas pretenden incluir construcciones de expresión según lo definido en el presente documento.

Los métodos de producción y uso de construcciones y vectores genéticos son conocidos en la técnica y se describen, en general, en Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", 2ª Ed. Cold Spring Harbor Press, 1987; Ausubel *et al.*, "Current Protocols in Molecular Biology", Greene Publishing, 1987).

Métodos de producción de células hospedadoras que comprenden polinucleótidos, construcciones o vectores

La divulgación describe una célula hospedadora que comprende una construcción genética o un vector genético de la invención.

Las células hospedadoras que comprenden construcciones genéticas, tales como construcciones de expresión, son útiles en métodos conocidos en la técnica (por ejemplo, Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", 2ª Ed. Cold Spring Harbor Press, 1987; Ausubel *et al*., "Current Protocols in Molecular Biology", Greene Publishing, 1987) para la producción recombinante de polipéptidos de la invención. Dichos métodos pueden implicar el cultivo

de células hospedadoras en un medio apropiado en condiciones adecuadas para o que conduzcan a la expresión de un polipéptido. El polipéptido recombinante expresado, que se puede segregar opcionalmente en el cultivo, se puede separar después del medio, de las células hospedadoras o del medio de cultivo mediante métodos conocidos en la técnica (por ejemplo, Deutscher, Ed, 1990, *Methods in Enzymology,* Vol 182, "Guide to Protein Purification").

Métodos de producción de células vegetales y plantas que comprenden construcciones y vectores

La invención proporciona además células vegetales que comprenden una construcción genética y células vegetales modificadas para modificar la expresión de un polinucleótido o polipéptido. Las plantas que comprenden dichas células también forman un aspecto.

La modificación de la actividad de la floretina glicosiltransferasa se puede modificar en una planta a través de métodos desvelados en el presente documento. Dichos métodos pueden implicar la transformación de células vegetales y plantas, con una construcción diseñada para modificar la expresión de un polinucleótido o polipéptido que module la actividad de la floretina glicosiltransferasa o el contenido de floricina en dichas células vegetales y plantas. Dichos métodos también incluyen la transformación de células vegetales y plantas con una combinación de la construcción y una o más construcciones distintas diseñadas para modificar la expresión de uno o más polinucleótidos o polipéptidos que modulan la actividad de la floretina glicosiltransferasa y/o el contenido de floricina de dichas células vegetales y plantas.

Los métodos de transformación de células vegetales, plantas y partes de las mismas con polipéptidos se describen en Draper *et al.*, 1988, "Plant Genetic Transformation and Gene Expression. A Laboratory Manual". Blackwell *Sci. Pub. Oxford*, pág. 365; Potrykus y Spangenburg, 1995, "Gene Transfer to Plants". Springer-Verlag, Berlín; y Gelvin *et al*., 1993, *Plant Molecular Biol. Manual*. Kluwer Acad. Pub. Dordrecht. En Galun y Breiman, 1997, "Transgenic Plants". Imperial College Press, Londres, se proporciona una revisión de plantas transgénicas, incluyendo técnicas de transformación.

Métodos de manipulación genética de plantas

Hay una serie de estrategias de transformación de plantas disponible (por ejemplo, Birch, 1997, *Ann Rev Plant Phys Plant Mol Biol*, 48, 297, Hellens R. P., *et al* (2000) *Plant Mol Biol* 42: 819-32, Hellens R. *et al., Plant Meth* 1: 13). Por ejemplo, se pueden diseñar estrategias para aumentar la expresión de un polinucleótido/polipéptido en una célula vegetal, órgano vegetal y/o en una determinada fase de desarrollo donde/cuando se exprese de manera normal o para expresar de manera ectópica un polinucleótido/polipéptido en una célula, un tejido, un órgano y/o en una determinada fase de desarrollo que/cuando no se exprese de manera normal. El polinucleótido/polipéptido expresado puede proceder de la especie vegetal que se vaya a transformar o puede proceder de una especie vegetal diferente.

Se pueden diseñar estrategias de transformación para reducir la expresión de un polinucleótido/polipéptido en una célula, un tejido, un órgano o una determinada fase de desarrollo vegetal que/cuando se expresa de manera normal. Dichas estrategias se conocen como estrategias de silenciamiento génico.

Por lo general, las construcciones genéticas para la expresión de genes en plantas transgénicas incluyen promotores para conducir la expresión de uno o más polinucleótidos clonados, terminadores y secuencias marcadoras seleccionables para detectar la presencia de la construcción genética en la planta transformada.

Los promotores adecuados para su uso en las construcciones son funcionales en una célula, un tejido u un órgano de una planta monocotiledónea o dicotiledónea, e incluyen promotores específicos de la célula, del tejido y del órgano, promotores específicos del ciclo celular, promotores temporales, promotores inducibles, promotores constitutivos que son activos en la mayoría de los tejidos vegetales y promotores recombinantes. La elección del promotor dependerá de la expresión temporal y espacial del polinucleótido clonado, así deseado. Los promotores pueden ser aquellos que normalmente están asociados a un transgén de interés o los promotores que se obtienen de genes de otras plantas, virus, y bacterias y hongos patógenos vegetales. Los expertos en la materia podrán, sin la necesidad de realizar experimentación, seleccionar promotores que sean adecuados para su uso en la modificación y la modulación de rasgos vegetales usando construcciones genéticas que comprenden las secuencias de polinucleótidos de la invención. Los ejemplos de promotores de plantas constitutivos incluyen el promotor CaMV 35S, el promotor de nopalina sintasa y el promotor de octopina sintasa y el promotor de Ubi 1 del maíz. Los promotores de plantas que son activos en tejidos específicos, que responden a señales evolutivas internas o a estreses abióticos o bióticos externos se describen en la bibliografía científica. Por ejemplo, en el documento WO 02/00894, se describen ejemplos de promotores).

Los ejemplos de terminadores que se usan comúnmente en la construcción genética de transformación de plantas incluyen, por ejemplo, el terminador del virus del mosaico de la coliflor (CaMV) 35S, los terminadores de la nopalina sintasa u octopina sintasa de *Agrobacterium tumefaciens*, el terminador del gen zein de *Zea mays*, el terminador de la ADP-glucosa pirofosforilasa de *Oryza sativa* y el terminador de PI-II de *Solanum tuberosum*.

Los marcadores seleccionables usados comúnmente en la transformación de plantas incluyen el gen de la neomicina fosfotransferasa II (NPT II) que confiere resistencia a la kanamicina, el gen aadA, que confiere resistencia a la espectinomicina y a la estreptomicina, la fosfinotricina acetil transferasa (gen *bar*) para la resistencia a Ignite (AgrEvo) y Basta (Hoechst), y el gen de higromicina fosfotransferasa (hpt) para la resistencia a la higromicina.

También se contempla el uso de construcciones genéticas que comprende genes indicadores (secuencias de codificación que expresan una actividad que es extraña para el hospedador, normalmente una actividad enzimática y/o una señal visible (por ejemplo, luciferasa, GUS, GFP) que se puede usar para el análisis de expresión de promotores en plantas y tejidos vegetales. La bibliografía de genes indicadores se revisa en Herrera-Estrella *et al.*, 1993, *Nature* 303, 209 y Schrott, 1995. En: “Gene Transfer to Plants” (Potrykus, T., Spangenberg. Eds) Springer Verlag. Berlín, pág. 325-336.

Las estrategias de silenciamiento génico se pueden centrar en el propio gen o en elementos reguladores que efectúen la expresión del polipéptido codificado. La expresión “elementos reguladores” se usa en el presente documento en el sentido más amplio posible, e incluye otros genes que interaccionan con del gen de interés.

Las construcciones genéticas diseñadas para reducir o silenciar la expresión de un polinucleótido/polipéptido pueden incluir una copia antisentido de un polinucleótido. En dichas construcciones, el polinucleótido se sitúa en una orientación antisentido con respecto al promotor y al terminador.

Un polinucleótido “antisentido” se obtiene invirtiendo un polinucleótido o un segmento del polinucleótido de modo que la transcripción producida sea complementaria a la transcripción del ARNm del gen, por ejemplo:

5'GATCTA 3' (cadena codificante)	3'CTAGAT 5' (cadena antisentido)
3'CUAGAU5' ARNm	5'GAUCUCG3' ARN antisentido.

Las construcciones genéticas diseñadas para el silenciamiento génico también pueden incluir una repetición invertida. Una “repetición invertida” es una secuencia que se repite, donde la segunda mitad de la repetición es la cadena complementaria, por ejemplo:

5'-GATCTA.....TAGATC-3'
3'-CTAGAT.....ATCTAG-5'

La transcripción formada puede someterse al apareamiento de bases complementarias para formar una estructura de horquilla. Normalmente, se requiere un espaciador de al menos 3-5 pb entre la región repetida para permitir la formación de la horquilla.

Otra metodología de silenciamiento implica el uso de un ARN antisentido pequeño dirigido a la transcripción equivalente a un ARNmi (Llave *et al*., 2002, *Science* 297, 2053). Se contempla de manera expresa el uso de un ARN antisentido pequeño correspondiente al polinucleótido.

La expresión “construcción genética” como se usa en el presente documento también incluye los ARN antisentido pequeños y otros de dichos polipéptidos que efectúan el silenciamiento génico.

La transformación con una construcción de expresión, como se define en el presente documento, también puede generar el silenciamiento génico a través de un proceso conocido como supresión de sentido (por ejemplo, Napoli *et al.*, 1990, *Plant Cell* 2, 279; de Carvalho Niebel *et al.,* 1995, *Plant Cell,* 7, 347). En algunos casos la supresión de sentido puede implicar la sobreexpresión de toda o parte de la secuencia codificante, pero también puede implicar la expresión de la región no codificante del gen, tal como un intrón o una región no traducida 5' o 3' (UTR). Se pueden usar construcciones sentido parciales quiméricas para silenciar de manera coordinada múltiples genes, (Abbott *et al*., 2002, *Plant Physiol.* 128 (3): 844-53; Jones *et al.*, 1998, *Planta* 204: 499-505). También se contempla el uso de dichas estrategias de supresión de sentido para silenciar la expresión de un polinucleótido.

Las inserciones de polinucleótidos en las construcciones genéticas diseñadas para silenciar genes pueden corresponder a la secuencia codificante y/o secuencia no codificante, tal como promotor y/o intrón y/o secuencia 5' o 3' UTR o el gen correspondiente.

Otras estrategias de silenciamiento génico incluyen metodologías negativas dominantes y el uso de construcciones de ribozimas (Mclntyre, 1996, *Transgenic Res*, 5, 257).

El silenciamiento previo a la transcripción se puede realizar mediante la mutación del propio gen o sus elementos reguladores. Dichas mutaciones pueden incluir mutaciones puntuales, desplazamientos del marco de lectura, inserciones, supresiones y sustituciones.

A continuación, se presentan publicaciones representativas que desvelan protocolos de transformación genética que se pueden usar para transformar genéticamente las siguientes especies vegetales: arroz (Alam *et al.,* 1999, *Plant*

Cell Rep. 18, 572); manzana (Yao *et al.*, 1995, *Plant Cell Reports* 14, 407-412); maíz (patentes estadounidenses Nº de serie 5.177.010 y 5.981.840); trigo (Ortiz *et al.,* 1996, *Plant Cell Rep*. 15, 1996, 877); tomate (patente estadounidense Nº de serie 5.159.135); patata (Kumar *et al.*, 1996 *Plant J.* 9,: 821); yuca (Li *et al.*, 1996 *Nat. Biotechnology* 14, 736); lechuga (Michelmore *et al.,* 1987, *Plant Cell Rep.* 6, 439); tabaco (Horsch *et al.,* 1985, *Science* 227, 1229); algodón (patentes estadounidenses Nº de serie 5.846.797 y 5.004.863); pastos (patentes estadounidenses Nº de serie 5.187.073 y 6.020.539); menta (Niu *et al*., 1998, *Plant Cell Rep*. 17, 165); plantas cítricas (Pena *et al.*, 1995, *Plant Sci.*104, 183); alcaravea (Krens *et al.,* 1997, *Plant Cell Rep*, 17, 39); plátano (patente estadounidense Nº de serie 5.792.935); soja (patentes estadounidenses Nº de serie 5.416.011; 5.569.834; 5.824.877; 5.563.044 y 5.968.830); piña (patente estadounidense Nº de serie 5.952.543); álamo (patente estadounidense Nº de serie 4.795.855); monocotiledóneas en general (patentes estadounidenses Nº de serie 5.591.616 y 6.037.522); *Brassica* (patentes estadounidenses Nº de serie 5.188.958; 5.463.174 y 5.750.871); cereales (patente estadounidense Nº de serie 6.074.877); pera (Matsuda *et al.,* 2005, *Plant Cell Rep*. 24(1):45-51); Pruno (Ramesh *et al*., 2006 *Plant Cell Rep.* 25(8):821-8; Song y Sink 2005 *Plant Cell Rep.* 2006 ;25(2):117-23; Gonzalez Padilla *et al.,* 2003 *Plant Cell Rep*.22(1):38-45); fresa (Oosumi *et al.,* 2006 *Planta*. 223(6):1219-30; Folta *et al.,* 2006 *Planta*, 14 de abril; PMID: 16614818), rosa (Li *et al.*, 2003), Rubus (Graham *et al.,* 1995 *Methods Mol Biol*. 1995; 44:129-33), tomate (Dan *et al.,* 2006, *Plant Cell Reports* V25:432-441), manzana (Yao *et al.*, 1995, *Plant Cell Rep.* 14, 407-412) y *Actinidia eriantha* (Wang *et al.,* 2006, *Plant Cell Rep.* 25,5: 425-31). La invención también contempla la transformación de otras especies. En la literatura científica, se pueden encontrar métodos y protocolos adecuados.

Se pueden emplear varios métodos adicionales conocidos en la técnica para modificar la expresión de un nucleótido y/o polipéptido de la invención. Dichos métodos incluyen, pero sin limitación, Tilling (Till *et al*., 2003, *Methods Mol Biol*, 2 %, 205), denominado tecnología "Deletagene" (Li *et al.,* 2001, *Plant Journal* 27 (3), 235) y el uso de factores de transcripción artificiales tales como factores de transcripción de dedo de cinc sintéticos (por ejemplo, Jouvenot *et al.,* 2003, *Gene Therapy* 10, 513). Además, también se pueden expresar anticuerpos o fragmentos de los mismos, dirigidos a un determinado polipéptido en plantas para modular la actividad de ese polipéptido (Jobling *et al.,* 2003, *Nat. Biotechnol.*, 21(1), 35). También se pueden aplicar metodologías de marcaje con transposón. Además, los péptidos que interaccionan con un polipéptido de la invención se pueden identificar mediante tecnologías tales como presentación de fases (Dyax Corporation). Dichos péptidos que interaccionan se pueden expresar en o aplicar a una planta para afectar a la actividad de un polipéptido de la invención. Se contempla específicamente el uso de cada una de las metodologías anteriores en la modificación de la expresión de un nucleótido y/o un polipéptido.

Las expresiones "modificación de la expresión de" y “expresión modificada" de un polinucleótido o de un polipéptido de la invención pretenden englobar la situación en la que se modifica el ADN genómico correspondiente a un polinucleótido de la invención, conduciendo así a la expresión alterada de un polinucleótido o de un polipéptido de la invención. La modificación del ADN genómico puede ser a través de la transformación genética u otros métodos conocidos en la técnica para la inducción de mutaciones. La "expresión alterada" puede estar relacionada con un aumento o una reducción de la cantidad de ARN mensajero y/o polipéptido producido, y también puede modificar la actividad de un polipéptido debido a alteraciones en la secuencia de un polinucleótido y un polipéptido producido.

Métodos de selección de plantas

También se proporcionan métodos para seleccionar plantas con actividad de la floretina glicosiltransferasa o contenido de floricina modificados. Dichos métodos implican ensayar las modificaciones de la expresión de un polinucleótido o un polipéptido en las plantas. Dichos métodos se pueden aplicar a una edad joven o fase de desarrollo temprana cuando la actividad de la floretina glicosiltransferasa o el contenido de floricina modificados pueden no medirse fácilmente necesariamente.

La expresión de un polinucleótido, tal como un ARN mensajero, se suele usar como un indicador de la expresión de un polipéptido correspondiente. Los ejemplos de métodos de medición de la expresión de un polinucleótido incluyen, pero sin limitación análisis Northern, RT-PCR y análisis de transferencia de puntos (Sambrook *et al.*, “Molecular Cloning: A Laboratory Manual”, 2ª Ed. Cold Spring Harbor Press, 1987). Por lo tanto, los polinucleótidos o las partes de los polinucleótidos son útiles como sondas o cebadores, como se define en el presente documento, en métodos para la identificación de plantas con niveles modificados de floretina glicosiltransferasa o floricina. Los polinucleótidos se pueden usar como sondas en experimentos de hibridación o como cebadores en experimentos basados en PCR, diseñados para identificar dichas plantas.

Como alternativa, se pueden generar anticuerpos contra polipéptidos. Los métodos de generación y uso de anticuerpos son convencionales en la técnica (véase, por ejemplo: “Antibodies, A Laboratory Manual”, Harlow A. Lane, Eds, Cold Spring Harbour Laboratory, 1998). Dichos anticuerpos se pueden usar en métodos de detección de la expresión modificada de polipéptidos que modulan el tamaño de las flores de las plantas. Dichos métodos pueden incluir ELISA (Kemeny, 1991, “A Practical Guide to ELISA”, NY Pergamon Press) y análisis Western (Towbin y Gordon, 1994, *J Immunol Methods*, 72, 313).

Estas metodologías de análisis de la expresión de polinucleótidos o polipéptidos y la selección de plantas con floretina glicosiltransferasa modificada o contenido de floricina modificado son útiles en los programas de

reproducción convencionales diseñados para producir variedades con la actividad de la floretina glicosiltransferasa o el contenido de floricina modificados.

Plantas

El término "planta" pretende incluir una planta completa, cualquier parte de una planta, propágulos y progenie de una planta.

El término “propágulo” significa cualquier parte de una planta que se puede usar en la reproducción o propagación, bien sexual o asexual, incluyendo las semillas y los esquejes.

Las plantas se pueden cultivar y bien autosembrar o cruzar con una cepa de planta diferente, pudiéndose identificar los híbridos resultantes, con las características de fenotipo deseadas. Se pueden cultivar dos o más generaciones para garantizar que las características fenotípicas objeto se mantengan y se hereden de manera estable. Las plantas procedentes de dichas metodologías de reproducción convencionales también forman un aspecto.

La función de una variante de polinucleótido en la codificación de una floretina glicosiltransferasa se puede evaluar, por ejemplo, expresando dicha secuencia en bacterias y ensayando la actividad de la proteína codificada según lo descrito en el apartado de ejemplos del presente documento.

También se puede alterar la modificación de la actividad de la floretina glicosiltransferasa y/o el contenido de floricina en una planta mediante métodos. Dichos métodos pueden implicar la transformación de células vegetales y plantas, con una construcción de la invención diseñada para modificar la expresión de un polinucleótido o un polipéptido que module la actividad de la floretina glicosiltransferasa y/o el contenido de floricina de dichas células vegetales y plantas. Preferentemente, dichos métodos también incluyen la transformación de células vegetales y plantas con una combinación de la construcción y una o más construcciones distintas diseñadas para modificar la expresión de uno o más polinucleótidos o polipéptidos distintos que modulen el contenido de ácido ascórbico en dichas células vegetales y plantas. Preferentemente, se expresa una combinación de floretina glicosiltransferasa y una floretina sintasa en las células vegetales o plantas.

Métodos de extracción de floricina de plantas

También se proporcionan métodos de producción de floricina mediante la extracción de floricina de una planta. Se puede extraer floricina de plantas mediante muchos métodos diferentes conocidos por los expertos en la materia.

Ridgway y Tucker, 1997, extrajeron floricina de hojas y ramitas de manzano joven (que contenía hasta el 10 % en peso seco de floricina). Demostraron que el determinante más importante del contenido de floricina de un tejido dado es su período de desarrollo, y que la variedad cultivada, incluyendo el tipo de portainjerto, tenía un efecto mucho menos destacado. Recomiendan que la producción de floricina mejor se lleve a cabo mediante el rebrote de los portainjertos de crecimiento rápido, tales como el M25 o MM106, siendo similar a la producción de sauce o álamo para biomasa; para lo que ya se ha desarrollado la cosecha mecánica. Entonces, se deben obtener fácilmente rendimientos de 250 kg de floricina por hectárea. Los rendimientos se podrían mejorar mediante ingeniería genética. (Ridgway T. y Tucker G. (1997), “Apple: a new agrochemical crop”. *Biochemical Society Transactions* 25, 110S).

Hunter y Hull (1993) también proporcionan un método para el manzano. Dong, H. *et al.* (2007) proporcionan una separación preparativa de floricina del extracto en bruto de *Lithocarpus polystachyus*. Cox *et al.* proporcionan un método de extracción de *Smilax*.

Estos métodos se pueden aumentar a escala para la extracción de floricina a mayor escala utilizando metodologías bien conocidas por los expertos en la materia.

Breve descripción de las figuras

La presente divulgación se entenderá mejor con referencia a las figuras adjuntas en las que:

La Figura 1 muestra la estructura de la floricina y de su aglicona floretina. La conversión de la floretina en floricina (2'-glucósido de floretina) está mediada por la uridina difosfato glicosiltransferasa de *Malus domestica* (MpUGT88A1) en presencia de uridina difosfato glucosa (UDP-glc).

La Figura 2 muestra la expresión relativa de MpUGT88A1 *in silico*. Los datos se basan en la distribución de las EST de MpUGT88A1 mediante la comparación con la biblioteca de manzanas de GenBank. Se dividió el número de EST de MpUGT88A1 de cada tejido entre el número total de EST secuenciadas para ese tejido y expresadas en relación con la muestra de floema. Tipos de tejidos: Fruto [1] = fruto pequeño cosechado 10 DDFP (días después de la floración plena); Fruto [2] = fruto en desarrollo cosechado 24-87 DDFP; Fruto [3] = fruto maduro cosechado 126-150 DDFP; Fruto [3-sk] = piel de fruto maduro; Fruto [3-cr] = corteza de fruto maduro; Fruto [3-cr] = núcleo de fruto maduro; Fruto [1-3] = fruto maduro que combina piel, corteza y núcleo; Hoja [inf] = tejido foliar

desafiado con patógeno; Hoja [1] = hoja joven y en expansión; Hoja [2] = hoja en senescencia, floema; xilema; capullo; raíz, flor, brotes, semillas y otros tejidos.

La Figura 3 muestra un alineamiento del polipéptido MpUGT88A1 (SEC ID Nº 1) y otras UGT glicosiltransferasas vegetales. MpUGT88A1 (*Malus domestica*; manzana, EU246349); OsBAC10743 (*Oryza sativa*, arroz, BAC10743), SrUGT88B1 (*Stevia rebaudiana*, Stevia, AAR06919), AtUGT88A1 (*Arabidopsis thaliana*, At3g16520) y VmUFGT4 (*Vigna mungo*, gramo negro, BAA36412). Los recuadros negros y grises contienen residuos que son idénticos y similares, respectivamente. La región subrayada indica el motivo “PSPG” conservado en todas las glicosiltransferasas vegetales.

La Figura 4 muestra una comparación filogenética de MpUGT88A1 y otras secuencias de glicosiltransferasas vegetales. Se dan valores de arranque de > 80 % (1.000 repeticiones de arranque). Los identificadores de genes son como en la Fig. 3 más *Arabidopsis thaliana* AtUGT72B1 (At4g01070); AtUGT72C1 (At4g36770); AtUGT72D1 (At2g18570); AtUGT72E1 (At3g50740); AtUGT71B1 (At3g21750); AtUGT71C1 (At2g29750) y AtUGT71D1 (At2g29730). Los miembros de la familia 88 de UGT están en un recuadro.

La Figura 5 muestra la expresión relativa del gen de MpUGT88A1 en diferentes tejidos vegetales. Se usó RT-PCR cuantitativa usando cebadores específicos de MpUGT88A1 para medir los niveles de transcripción en cuatro tejidos de manzana: L = hoja en expansión; F = fruto maduro (150 DDFP), P = flor abierta y R = punta de la raíz. La expresión se da con respecto a la muestra de la raíz.

La Figura 6 muestra la purificación de MpUGT88A1 recombinante producido en *Escherichia coli.* Separación SDS-PAGE de las proteínas purificadas mediante cromatografía de afinidad Ni^{2+} (His-Trap) y filtración en gel (G200 Superdex). 1) escalera Benchmark (Invitrogen, Auckland, NZ); 2 y 3) fracciones de MpUGT88A1 purificado; 4 y 5) fracciones equivalentes de control de vector pET30a(+) purificado; 6, inmunodetección de MpUGT88A1 usando un anticuerpo monoclonal His_6.

La Figura 7 muestra el análisis de LC-MS de los productos de la reacción entre MpUGT88A1, floretina y azúcares UDP. Cromatogramas de HPLC: a) patrón mixto de floricina y floretina; b) floretina + UDP-galactosa; c) floretina + UDP-glucosa; d) floretina + UDP-xilosa; e) floretina + UDP-galactosa salpicada con floricina; f) floretina + UDP-glucosa salpicada con floricina; espectros MS: g) FullScan, datos de MS^2 y MS^3 para la floricina; h) FullScan, datos de MS^2 y MS^3 para la floretina; i) FullScan, datos de MS^2 y MS^3 para el pico 1; j) FullScan, datos de MS^2 y MS^3 para el pico 2; k) FullScan, datos de MS^2 y MS^3 para el pico 3; y l) FullScan, datos de MS^2 y MS^3 para el pico 4.

La Figura 8 muestra (A) la actividad de MdPGT1 hacia la floretina (2 μM), ensayada en mezclas de reacción que contienen Tris-HCl 50 mM al pH indicado; y (B) la actividad de MdPGT1 hacia la floretina (2 μM), ensayada en condiciones convencionales a las temperaturas indicadas.

La Figura 9 muestra un alineamiento CLUSTALX del polipéptido MpUGT88A1 y variantes del mismo (SEC ID Nº 1 a 5), y muestra la posición de un motivo completamente conservado (SEC ID Nº 16) presente en las cinco secuencias.

La Figura 10 muestra el % de identidad entre cada una de las secuencias alineadas en la Figura 9.

Ejemplos

Ejemplo 1: Identificación de una secuencia de floretina glicosiltransferasa de manzana

Identificación de la secuencia y filogenética

Se sometieron secuencias EST de manzana del GenBank a búsqueda con BLAST (a excepción del valor de exp^{-05}) usando los genes de UGT previamente publicados del GenBank. Se analizaron las secuencias EST de forma automática a través de dos series de la construcción de cóntigos, y se obtuvo un conjunto de secuencias "de cóntigo no redundante". Se construyeron alineamientos de aminoácidos de las proteínas predichas usando ClustalX (versión 1.8). Se comprobó en todas las proteínas la presencia del motivo de firma común de 42 aminoácidos encontrado en las UGT vegetales (Li *et al*., 2001). Se seleccionó una secuencia de longitud completa de cada cóntigo para la secuenciación completa.

Para el análisis filogenético, primero se alinearon todas las secuencias de aminoácidos usando ClustalX, y después se editaron manualmente. Las secuencias de UGT de *Arabidopsis* se obtuvieron de la siguiente página web http://www.p450.kvl.dk/UGT.shtml. Los valores de confianza para las agrupaciones de árboles filogenéticos se obtuvieron usando el árbol BOOTSTRAP N-J TREE, usando 1.000 ensayos de arranque. Los árboles se visualizaron en TREEVIEW (v.1.6.6).

Gen MpUGT88A y su proteína predicha

La extracción de genes identificó más de 60 genes codificantes de UDP-glicosiltransferasas entre las ~270.000 secuencias EST de manzana del GenBank. Los perfiles de expresión *in silico* de las EST de UGT de manzana identificaron el gen “MpUGT88A1” que fue altamente expresado (representado por 255 EST) en todos los tejidos conocidos por producir floricina incluyendo las hojas, las semillas y los frutos (Fig. 2). El clon de ADNc para MpUGT88A1 es de 1.745 nucleótidos de longitud y contiene un marco de lectura abierto de 483 aminoácidos con 64 nucleótidos de 5' UTR y 232 nucleótidos de 3' UTR. La proteína MpUGT88A1 predicha tiene una masa de 53,5 kDa y un pI de 5,8, y contiene la secuencia consenso "PSPG” de 42 aminoácidos que se encuentra en todas las UGT y que se cree que participa en la unión del resto UDP del azúcar activado (Li *et al*., 2001). En esta secuencia consenso, MpUGT88A1 contiene dos motivos altamente conservados "WXPQ" y "HCGWNS", que se encuentran en el 95 % de las secuencias de UGT, y una glutamina absolutamente conservada en el aminoácido 42 de la caja PSPG (Vogt, 2000).

Se construyó un árbol filogenético marco usando MpUGT88A1 y miembros representativos del árbol de UGT de *Arabidopsis* publicado por Ross *et al*. (2001). Este árbol marco indicó que MpUGT88A1 se agrupó con el único AtUGT88A1 de secuencia de la familia 88 de UGT de *Arabidopsis* (datos no mostrados). Por lo tanto, el clon de manzana fue designado MpUGT88A1 de acuerdo con la nomenclatura descrita en Ross *et al*. (2001). MpUGT88A1 mostraba el 42,5 % de identidad de aminoácidos con AtUGT88A1, y una homología significativa con dos secuencias de UGT de longitud completa del arroz (*Oryza sativa*, BAC10743) y de la Stevia (*Stevia rebaudiana*, AAR06919) y una secuencia parcial del gramo negro (*Vigna mungo*, BAA36412, 52 % de identidad de aminoácidos en la región de solapamiento). MpUGT88A1 y estas cuatro secuencias se muestran alineadas en la Fig. 3. A continuación, se construyó un árbol filogenético más detallado usando estas secuencias alineadas con miembros seleccionados de las familias 71 y 72 de UGT de *Arabidopsis* (las familias de UGT más similares a la familia 88 de UGT). Este árbol indica que las secuencias de MpUGT88A1 de manzana y del arroz relacionado, *S. rebaudiana*, y del gramo negro dadas en la Fig. 3 se separan claramente con AtUGT88A1 como parte de la familia 88 de UGT y se diferencian de otras secuencias de UGT de *Arabidopsis* (Fig. 4).

Ejemplo 2: Análisis del perfil de expresión de la floretina glicosiltransferasa en plantas

Material vegetal, extracción de ARN y qPCR

Se aisló ARN total de tejidos de manzana mediante un método de extracción de ARN en sílice modificada (Nieuwenhuizen *et al*., 2007). Se midió la concentración de ARN de cada muestra usando un espectrofotómetro ND-1000 NanoDrop® (NanoDrop Technologies, Inc., Wilmington, DE, EE.UU.). Se trataron las muestras de ARN (~3 µg en reacciones de 30 µl) durante 30 minutos a 37 ºC con DNasa I (Ambion, Inc., Austin, TX, EE.UU.) para eliminar cualquier contaminación de ADN genómico menor. Se inactivó por calor la DNasa durante 30 min a 50 ºC. La transcripción inversa se realizó en reacciones de 20 µl según las instrucciones del fabricante usando RNasa H-transcriptasa inversa SuperScript III™ (Invitrogen, Auckland, NZ), 500 ng de ARN y el cebador NotI-PA 5'-GACTAGTTCT AGATCGCGAG CGGCCGCCCT$_{(15)}$-3' (SEC ID Nº 17).

Se repitieron las reacciones de PCR cuantitativa (20 µl) cuatro veces en una ABI Prism 7900HT (Applied Biosystems, Foster City, CA, EE.UU.) con 3 µl de ADNc diluido, cebadores específicos de MpUGT88A1 0,5 µM (directo 5'-GAAGGGTGTG TTGCCAGAAG GGT-3' (SEC ID Nº 18); inverso 5'-GTCACGAACC CACCAACCGA CT-3' [SEC ID Nº 19]), y 5 µl de LightCycler 480 SYBR Green Master (Roche Diagnostics) siguiendo las instrucciones del fabricante. Las condiciones de ciclación incluían un arranque en caliente inicial a 95 ºC durante 5 min, seguido de 45 ciclos de 95 ºC durante 10 s, 60 ºC durante 10 s y 72 ºC durante 12 s. Se midió la fluorescencia al final de cada etapa de hibridación. Cada reacción de PCR fue seguida por un programa de curva de fusión para comprobar que solo se amplificaron los productos individuales, partiendo de la desnaturalización a 95 ºC durante 5 s antes de enfriar hasta 65 ºC durante 1 min y después aumentando a 0,1 ºC•s^{-1} con la medición de fluorescencia continua hasta alcanzarse 97 ºC. Los controles negativos consistieron en agua en lugar de ADNc y se ejecutaron con todas las reacciones. Los datos se analizaron usando el programa LightCycler 480 versión 1.2.0.169. Para cada gen, se generó una curva patrón usando una dilución en serie de ADNc, y se usaron las eficacias resultantes para calcular la expresión relativa a la actina de manzana (CN938023) con el fin de minimizar la variación en los niveles de molde de ADNc.

Expresión de MpUGT88A1

Los perfiles de expresión *in silico* de MpUGT88A1 indicaron que las EST se representaron en las bibliotecas realizadas a partir de todos los tejidos del manzano, a excepción de la hoja en senescencia y del fruto en desarrollo cosechado 24-87 DDFP (Fig. 2). Las EST fueron más abundantes en las bibliotecas construidas a partir de tejido de flor, de floema y de brotes. En el fruto maduro, la expresión fue mayor en el tejido del núcleo, aunque también se detectó la expresión en la corteza y en la piel. Como los perfiles de expresión *in silico* pueden verse influidos por la profundidad del muestreo de la biblioteca y el sesgo de la clonación, también se midió la expresión génica de MpUGT88A1 mediante qPCR en cuatro tejidos del manzano: hoja en expansión, fruto maduro (150 DDFP), flor abierta y punta de la raíz (Fig. 5). Los datos de expresión de qPCR confirmaron los resultados de los perfiles de

expresión *in silico*. MpUGT88A1 se expresó en los cuatro tejidos, detectándose la expresión más elevada en la flor abierta y en el fruto maduro. En conjunto, los datos de expresión indican que, en el manzano, MpUGT88A1 es una transcripción muy abundante (0,1 % de las transcripciones totales), se expresa en una amplia selección de tejidos y que la transcripción se encuentra en todos los tejidos en los que se ha informado de la presencia de floricina.

Ejemplo 3: Expresión de la floretina glicosiltransferasa en *E. coli* y caracterización de la actividad enzimática

Expresión de MpUGT88A1 en *E. coli*

Se amplificó el marco de lectura abierto de EST111441 (MpUGT88A1) usando los cebadores RA335 5'-ACGGGATCCA TGGGAGACGT CATTGTACTG-3' (SEC ID Nº 20) y RA336 5'-CCCAAGCTTT TATGTAATGC TACTAACAAA GTTGAC-3' (SEC ID Nº 21). Se purificaron las bandas amplificadas usando columnas de limpieza de PCR Qiaquick® (Qiagen GmbH, Hilden, Alemania), se digirieron con *BamH*I y *Hin*dIII (subrayadas en los cebadores anteriores) y se ligaron en los sitios correspondientes del vector pET30a(+) (Novagen, Madison, WI, EE.UU). La secuencia del clon se verificó contra de la secuencia EST original. Se expresó la proteína marcada con His_6 N-terminal recombinante de plásmidos pET-30a(+) en células BL21-Codon-Plus™-RIL de *E. coli*. Los cultivos se desarrollaron en medios autoinducibles ZYM-5052 (Studier, 2005) a 37 ºC durante 4 horas a 300 rpm. A continuación, se redujo la temperatura hasta 16 ºC y se prosiguió con la incubación durante otras 60 h. Se purificaron las proteínas recombinantes en columnas HP quelantes His-Trap de 5 ml (Amersham Biosciences, Buckinghamshire, RU) y se eluyeron usando un gradiente de imidazol de 0-250 mM continuo como se describe en Green *et al.,* (2007). Después, se aplicó el concentrado a una columna de filtración en gel G200 Superdex de 40 cm y 1,6 veces (Pharmacia Biotech, Auckland, NZ) previamente equilibrada con Tris-HCl 50 mM, pH 7,5, NaCl 500 mM, DTT 5 mM a un caudal de 1 $ml \bullet min^{-1}$. Se combinaron las fracciones de pureza más alta, se ajustaron a glicerol al 15 % y se almacenaron a -80 ºC. Se expresó un control de solo vector pET-30(+) y se purificó como se ha explicado anteriormente.

Se analizó la proteína recombinante sobre geles de SDS-Tris-Tricina al 12 % (p/v), se electrotransfirieron a una membrana de polivinildifluoruro y se bloquearon como se describe en (Nieuwenhuizen *et al*., 2007). Se inmunolocalizaron las proteínas con un anticuerpo monoclonal His_6 (Roche, Mannheim, Alemania), 1:1000, (p/v), se diluyeron en tampón TBS que contenía leche en polvo. Se incubaron las membranas con anticuerpo secundario conjugado con fosfatasa alcalina anti-conejo (Sigma-Aldrich, St. Louis, MO, EE.UU.) y la unión se visualizó usando el sistema de sustrato líquido BCIP®NBT (nitroazul de tetrazolio) (Sigma-Aldrich).

Ensayos de actividad de UGT

Los ensayos de actividad de UGT se realizaron por triplicado en reacciones de 50 µl usando ~1-2 µg de proteína recombinante purificada en columnas de His-Trap y G200 Superdex. Las reacciones se realizaron en tampón de ensayo de glicosiltransferasa (Tris HCl 50 mM, pH 7,5, DTT 2 mM) con sustrato 2 mM y 1 µl de [^{3}H]-UDP-glucosa (uridina difosfo-D-[6-^{3}H] glucosa, 13,6 Ci/mmol, GE Heathcare Buckinghamshire, RU). Las reacciones se realizaron a 30 ºC durante 30 min y se finalizaron mediante la adición de 10 µl de HCl 2 M. Se extrajeron las mezclas de reacción dos veces con 100 µl de acetato de etilo, y se combinaron 20 µl de la fase orgánica con 1 ml de fluido de centelleo no acuoso y se analizaron por recuento de centelleo líquido (Tri-Carb 2900TR, PerkinElmer, Boston, MA, EE.UU.). La enzima hervida y los controles de vector pET30a(+) se ejecutaron en paralelo con todas las reacciones enzimáticas.

Se determinó el efecto sobre la actividad de UGT de diferente pH (5-10), temperatura (15-50 ºC) y fuerza iónica de iones monovalentes (Na^+, 0-100 mM; K^+, de 0-10 mM) y divalentes (Mg^+ o Mn^{2+}, 0-25 mM) como se ha descrito anteriormente usando floretina 2 µM. Las reacciones mostraron ser lineales durante 1 h en condiciones convencionales. El valor K_m aparente para la floretina se determinó variando la concentración de la floretina de 4 µM a 0,01 µM con una concentración de [^{3}H]-UDP-glucosa fija de 27 uM. El valor de K_m para la UDP-glucosa se determinó mediante la variación de la concentración de [^{3}H]-UDP-glucosa de 27 uM a 0,135 Um y mezclando [^{3}H]-UDP-glucosa (13,5 uM) con UDP-glucosa fría (10-100 uM) a una concentración de floretina fija de 2 µM.

Para el análisis de LC-MS, las reacciones contenían ~10 µg de enzima, floretina 10 µM y UDP-glucosa (Sigma Aldrich), UDP-galactosa o UDP-xilosa (CarboSource Servicios, Universidad de Georgia, GA, EE.UU.) a una concentración final de 250 µM. Las reacciones se realizaron durante 16 h, y se detuvieron mediante la adición de 10 µl de ácido acético glacial al 10 %. Se combinaron los productos de 3 reacciones de 100 µl para el análisis de LC-MS.

Preferencia de sustrato de MpUGT88A1

Se clonó toda la región de codificación del ADNc de MpUGT88A1 en marco en el vector de expresión pET30a(+) para la producción de proteína recombinante marcada con His_6 en *Escherichia coli*. Se aisló la proteína recombinante mediante cromatografía de afinidad de Ni^{2+} y de filtración en gel. La expresión con éxito de la proteína de fusión se indicó en SDS-PAGE por la aparición de una proteína del tamaño esperado de 58 kDa (53,5 kDa para

MpUGT88A1 + 4,5 kDa para el marcador His_6 y restos asociados) en los extractos de MpUGT88A1, pero no en los extractos de control de vectores equivalentes (Fig. 6, carriles 2-5). El análisis Western usando anticuerpos monoclonales His_6 confirmó que la 58 kDa era una proteína recombinante marcada con His (Fig. 6, carril 6).

Para la caracterización funcional de MpUGT88A1, se ensayaron alícuotas de la proteína recombinante purificada en cuanto a la actividad glicosiltransferasa usando [^{3}H]-UDP-glucosa como donante de azúcar. Se exploró la actividad de quince sustratos (ácido cafeico, catequina, ácido clorogénico, ácido 2-cumárico, ácido 3-cumárico, ácido 4-cumárico, cianidina, ácido 3-(3,4-dihidroxifenil)propiónico, epicatequina, ácido 3-hidroxibenzoico, naringenina, floretina, ácido protocatéquico, quercetina y rutina) que son componentes naturales de la manzana o compuestos fenólicos comúnmente disponibles. De estos sustratos, solo la floretina fue utilizada por MpUGT88A1 como un aceptor. El control de vector no mostró ninguna actividad hacia la floretina, lo que indica que la reacción de glicosilación era específica de la enzima MpUGT88A1.

Se determinaron los parámetros cinéticos para MpUGT88A1 con respecto a la floretina y la UDP-glucosa. La enzima mostró una K_m de 0,62 uM ± 0,1 μM para la floretina con una tasa de renovación de 9,72 x 10^{-4} mol/s/mol. La K_m observada para UDP-glucosa fue de ~13 uM, coincidiendo con la observada para otras enzimas UGT.

Se ensayó la proteína MpUGT88A1 en un intervalo de pH de 5-10. La enzima mostró una actividad significativa de pH 7-8 con actividad máxima a pH 7,7 (Figura 8A). La actividad se redujo al 20 % a pH 6 y 9, y a menos del 5 % a pH 5 y 10. La enzima mostró un amplio intervalo de temperaturas de 15-50 ºC con actividad máxima a 25 ºC (Figura 8B). La actividad enzimática no se inhibió mediante la adición de cationes divalentes (Mg^{2+} o Mn^{2+}, 25 mM) ni cationes monovalentes (Na^+, 100 mM o K^+, 10 mM).

Análisis LC-MS de mezclas de reacción *in vitro*

La LC-MS empleó un espectrómetro de masas con trampa de iones lineal LTQ dotado de una interfaz ESI (Thermoquest, Finnigan, San Jose, CA, EE.UU.) acoplado a un MDLC Ettan™ (GE Healthcare Bio-Sciences, Uppsala, Suecia).

La separación de los compuestos fenólicos se realizó usando una columna analítica de 150 x 2 mm de 100 Å (Phenomenex, Torrance, CA, EE.UU.) ODS (3) de 5 μm Prodogy, mantenida a 35 ºC. Se instaló un filtro en línea de 0,2 μm (Alltech, Deerfield, Illinois, EE.UU.) antes de la columna. Los disolventes fueron (A) acetonitrilo + ácido fórmico al 0,1 % y (B) agua + ácido fórmico al 0,1 %, y el caudal fue de 200 μl•min^{-1}. La fase móvil inicial, A al 5 %/B al 95 %, se mantuvo durante 5 min y luego en rampa linealmente hasta A al 10 % durante 10 min, A al 17 % durante 25 min, A al 23 % durante 30 min, A al 30 % durante 40 min, A al 97 % durante de 48 a 53 min antes de restablecer las condiciones originales. El volumen de inyección de las muestras fue de 50 μl.

Los datos de MS se adquirieron en el modo negativo usando un método de LC-MS^3 dependiente de los datos. Este método aísla y fragmenta el ion primario más intenso, dando los datos de MS^2, y a continuación, aísla y fragmenta el ion secundario más intenso (datos de MS^3). Para maximizar la sensibilidad, se fijó el intervalo de exploración completo en *m/z* 420-490 de 0 a 42,9 min para la detección de glucósidos de floretina, y luego en *m/z* 270-280 a partir de 43 min para la detección de floretina sin reaccionar. El voltaje de ESI, la temperatura capilar, la presión del gas de la cubierta y el gas de barrido se ajustaron a -10 V, 275 ºC, 275,79 kPa (40 psi) y 34,47 kPa (5 psi), respectivamente.

Análisis LC-MS

Se analizaron los productos de la reacción entre MpUGT88A1, floretina y UDP-glucosa mediante HPLC (Fig. 7c), y se compararon con los cromatogramas de los patrones de floretina y floricina (Fig. 7a). El producto de la reacción enzimática de MpUGT88A1 se ejecutó en 38 min - el mismo tiempo de retención que el patrón de floricina. No se observó floretina a los 47 min, lo que indicó que la reacción se había realizado hasta el final (Fig. 7c). Una serie de HPLC en la que se añadió el patrón de floricina a la mezcla de reacción de MpUGT88A1 confirmó además que los productos de la reacción tenían tiempos de retención idénticos (Fig. 7f).

Se ensayó la enzima MpUGT88A1 con floretina en presencia de dos donantes de azúcar activados adicionales - UDP-xilosa y UDP-galactosa. Los cromatogramas de HPLC indicaron que se formó un solo producto glicosilado con UDP-xilosa con un tiempo de retención de 40,5 min (Fig. 7d). Se observaron dos picos en presencia de UDP-galactosa - un pico pequeño a los 36,5 min y un pico mayor a los 38 min (Fig. 7b). El pico mayor se eluyó junto con el patrón de floricina a los 38 min, cuando los productos de la reacción de UDP-galactosa fueron salpicados con floricina auténtica (Fig. 7e).

Se usó la espectrometría de masas para caracterizar mejor los productos de las reacciones de MpUGT88A1. La floretina (Fig. 7h) se detectó como su ion pseudomolecular *m/z* 273 $[M-1]^-$, mientras que la floricina (Fig. 7g) y los productos de reacción del glicósido de floretina se detectaron predominantemente como los aductos de formiato correspondientes, *m/z* 481 $[M + formiato]^{-1}$ para el glucósido (Pico 3; Fig. 7c) y los galactósidos (picos 1 y 2; Fig. 7b)

y *m/z* 451 [M+formiato]$^{-1}$ para el xilósido (pico 4; Fig. 7d). MS^2 en los aductos de formiato identificó el ión pseudomolecular esperado a *m/z* 435 [M-1]$^-$ para el glucósido (Fig. 7k) y los galactósidos (Fig. 7i y 7j), y *m/z* 405 [M-1]$^-$ para el xilósido (Fig. 7l). MS^3 en los iones de glucósido y galactósido *m/z* 435 [M-1]$^-$ y el ion de xilósido *m/z* 405 [M-1]$^-$, todos identificaron el *m/z* 273 [M-1]$^-$ de la floretina aglicona.

Expresión de Md_138221 en *E. coli*

También se amplificó la secuencia de Md_138221 (SEC ID Nº 14) y se clonó en pET30a(+) mediante técnicas convencionales como se ha descrito anteriormente para MpUGT88A1.

Se confirmó la secuencia clonada mediante secuenciación antes de la transferencia a células RIL como se ha descrito anteriormente. A continuación, se finalizaron los ensayos para la actividad de la floretina glicosiltransferasa como se ha descrito anteriormente. Md_138221 también mostró actividad floretina glicosiltransferasa en aproximadamente el 3 % de la mostrada para MpUGT88A1.

Discusión

Se usó una metodología de genómica para identificar el gen MpUGT88A1 del manzano (cv. Royal Gala) con homología con las UGT de la familia 88 de función desconocida de *Arabidopsis*, el arroz y estevia (*S. rebaudiana*). Los ensayos de glicosiltransferasa y LC-MS revelaron que la enzima MpUGT88A1 recombinante pudo glicosilar específicamente la floretina en presencia de UDP-glucosa, UDP-xilosa y UDP-galactosa. Este es el primer informe de caracterización funcional de una UGT que utiliza una dihidrochalcona como su sustrato primario.

El producto glicosilado de floretina y UDP-glucosa migra junto con un patrón de floricina conocido, lo que indica que es probable que MpUGT88A1 sea la enzima que glicosila la floretina en floricina *in planta*. Esta afirmación se apoya en otras líneas de evidencia. Los datos de la expresión indicaron que el gen MpUGT88A1 se expresó a un alto nivel en la hoja, la raíz, la flor y el fruto, lo que coincide con la presencia de floricina en estos tejidos del manzano. La proteína MpUGT88A1 recombinante también produjo glucósidos con UDP-xilosa y UDP-galactosa (teniendo uno de los productos glicosilados con UDP-galactosa el mismo tiempo de retención que la floricina). Se ha informado que los manzanos producen dos glicósidos de floretina además de 2'-glucósido de floretina (floricina), en concreto, 2'-galactósido de floretina (Barth *et al*., 2005) y 2'-xiloglucósido de floretina (Lommen *et al*., 2000). Aunque no se ha determinado la naturaleza exacta de la glicosilación en estas reacciones (pues no había auténticos compuestos de manzano disponibles para la comparación), los espectros de MS indicaron claramente que las moléculas de azúcar se unieron a la floretina. Parece probable que la regio-especificidad de MpUGT88A1 es con respecto a la posición 2'-hidroxilo, pues no se utilizaron como sustratos aquellos (por ejemplo, la naringenina y la quercetina) con grupos hidroxilo en posiciones alternativas.

Hasta la fecha, a ninguna secuencia de la familia de UGT 88 se le ha atribuido una función definitiva. Se ha informado que SvUGT88B1 de *S. rebaudiana* muestra actividad menor con el kaempferol (Richman *et al*., 2005), mientras que VmUFGT4, un clon parcial aislado de tejido de plántulas de *V. mungo* (Mato *et al*., 1998), y OsBAC10743 del arroz no se han caracterizado en absoluto. Aunque las secuencias de UGT de *Arabidopsis* se han caracterizado extensamente *in vitro* (por ejemplo, Lim *et al*., 2002), no se ha atribuido ninguna función a AtUGT88A1. AtUGT88A1 se expresa moderadamente en una serie de tejidos de *Arabidopsis*, pero especialmente en las hojas caulinares (navegador eFP de *Arabidopsis*; http://www.bar.utoronto.ca).

Las UGT de la familia 88 se clasifican como pertenecientes a las UGT del Grupo E mayor, que también contiene las familias de UGT 71 y 72. Algunas enzimas de estas familias se han caracterizado funcionalmente, y muestran actividad hacia una amplia selección de sustratos, incluyendo: fenilpropanoides, benzoatos, fenoles clorados y anilinas (AtUGT71C1, Lim y Bowles 2004; AtUGT72E2, Lim y Bowles, 2004; AtUGT72B, Loutre *et al.,* 2003); cumarinas, flavonoides y naftoles (NtGT1a, NtGT3, *Nicotiana tabacum*, Taguchi *et al*., 2003); quercetina, isoflavonas y triterpeno saponinas (UGT71G1, *Medicago truncatula*, Achnine *et al*., 2005) e hidroxiquinona (arbutina sintasa, *Rauvolfia serpentina*, Hefner y Stockigt, 2003). Es posible que algunas de estas enzimas también acepten las dihidrochalconas como sustratos, pero esto no se ha publicado. Otras UGT de la familia 88 pueden ser específicas de las dihidrochalconas, pero esto es poco probable, dado que los glicósidos de dihidrochalcona se encuentran casi exclusivamente en *Malus* (y se han usado como marcadores para la detección de mezclas de manzana en otros zumos y papillas de frutas, Dragovic-Uzelac *et al*., 2005). Las UGT de la familia 88 tienen más probabilidades de mostrar, aunque no se ha determinado, estéreo-especificidad representada por la posición 2'-hidroxilo de la floretina.

En conclusión, los presentes resultados han identificado una enzima clave en la ruta bioquímica de producción floricina en los manzanos. La capacidad de manipular los niveles de floricina en el fruto es un objetivo importante en los programas de reproducción de los manzanos, pues "el uso de floricina puede proporcionar la base molecular de la observación clínica de que *una manzana al día del médico te libraría*" (Ehrenkranz *et al*., 2005).

Ejemplo 4: Aumento de la actividad de la floretina glicosiltransferasa en las plantas mediante la expresión de un polinucleótido de floretina glicosiltransferasa de la invención en las plantas

Transformación transitoria de hojas de tabaco

Se transforma transitoriamente tabaco (*Nicotiana benthamiana*) con cultivos de *Agrobacterium* que contienen un gen/polinucleótido (por ejemplo, SEC ID Nº 6-15) que codifica la floretina glicosiltransferasa (por ejemplo, SEC ID Nº: 1-5) clonado en pGreen (Hellens *et al.*, 2000) mezclado con *Agrobacterium* que contiene el gen para el supresor de el silenciamiento P19 como se ha descrito previamente (Hellens *et al.*, 2005). Los controles se realizaron usando *Agrobacterium* que contenían P19 en pGreen solo. Las hojas de tabaco se recogieron 9 días después de la transformación y se congelaron en nitrógeno líquido.

Transformación estable de plantas de manzano

Se produjeron dos construcciones que contenían MdPGT1 (EST111441), en pHEX2 (Hellens *et al.*, 2005) mediante técnicas convencionales, para caracterizar el papel del gen MdPGT1 en plantas de manzano transgénicas.

Se produjo una construcción con ARNi desactivado. Dicha construcción contenía 455 nucleótidos del gen EST111441 (nucleótidos 473-928 de SEC ID Nº 6) en orientación sentido, y los mismos 455 nucleótidos orientados en la dirección antisentido para formar una repetición invertida. La expresión constitutiva de la horquilla de ARNbc fue dirigida por un promotor 35S del virus del mosaico de la coliflor (CaMV).

Se produjo una construcción de sobreexpresión. Esta construcción contenía EST111441 de longitud completa (SEC ID Nº 6) dirigida por un promotor 35S de CaMV.

Se regeneraron plantas de manzano transgénicas como se ha descrito en Yao *et al.*, (1995) *Plant Cell Reports* 14: 407-412.

Se regeneraron catorce líneas de manzano que contenían la construcción desactivada EST111441 en un medio que contenía kanamicina. Se confirmó por PCR que once de las líneas contenían la construcción desactivada EST111441. Se transfirieron cinco líneas que mostraron la reducción más significativa en la expresión del ARN de EST111441 (en comparación con la expresión de EST111441 en hojas de tipo silvestre) al invernadero de contención para su posterior análisis (incluyendo el análisis de los niveles de floretina y floricina).

Se regeneraron ocho líneas de manzano que contenían la construcción de sobreexpresión EST111441 en un medio que contenía kanamicina. Estas líneas se caracterizarán del mismo modo que las líneas desactivadas para elegir las mejores plantas para la transferencia al invernadero de contención.

Medición de floricina

Las hojas recogidas en N_2 líquido, en el campo, se congelan instantáneamente, se almacenan a -40 ºC, se liofilizan y se muelen hasta obtener un polvo fino. Se preparan los extractos lavando previamente 200 mg de polvo en éter dietílico y realizando la extracción a 40 ºC durante 3 horas en acetona al 70 %-ácido ascórbico 8,6 mM. Se retira la acetona por evaporación bajo presión y se añade H_2O destilada a un volumen constante de 5 ml. Se estiman las concentraciones de floridizina y floretina mediante HPLC de fase inversa frente a patrones comerciales, con un disolvente de MeOH:H_2O (45:55), tamponado con reactivo PIC-A (Supelco, columna de 25 cm x 4,6 mm que contiene una fase ligada C_{18} 5 pm, precedida de una columna protectora de 2 cm empaquetada con Porasil B 37-75 um, caudal de 0,9 ml.min^{-1}) y se detecta a 254 nm con un detector de UV. Hunter y Hull, 1993.

Dichos experimentos pueden mostrar que la sobreexpresión de la secuencia de floretina glicosiltransferasa de la invención puede aumentar la actividad de la floretina glicosiltransferasa o la producción de floricina en las plantas.

Ejemplo 5: Identificación de secuencias de variantes de floretina glicosiltransferasa

Se usó la secuencia Md_111441 para identificar los genes ortólogos de la floretina glicosiltransferasa de las bases de datos de secuencias propiedad de HortResearch.

Se identificaron cuatro variantes de secuencias como se resume en la siguiente Tabla 1.

Tabla 1

Referencia de la floretina glicosiltransferasa	Especie de *Malus*	SEC ID Nº del polinucleótido		SEC ID Nº del polipéptido
		ADNc	Marco de lectura abierto (ORF)	
Md_111441	*x domestica*	6	7	1
Ms_297292	*sieboldii*	8	9	2
Md_88077	*x domestica*	10	11	3
Md_87623	*x domestica*	12	13	4
Md_138221	*x domestica*	14	15	5

En la Figura 9, se muestra el alineamiento realizado con ClustalX (versión 1.8) (Hall, 1999) de las secuencias polipeptídicas de Md_111441 y las cuatro variantes. También se ilustra la región completamente conservada (SEC ID Nº 16) presente en Md_111441 y el resto de variantes. La función de estas variantes como floretina glicosiltransferasas se puede confirmar usando los métodos descritos en los ejemplos anteriores.

Referencias

Achnine, L., Huhman, D. V., Farag, M. A., Sumner, L. W., Blount, J. W. y Dixon, R. A. (2005). “Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula”. *Plant Journal* 41: 875-887.

Awad, M. A., de Jager, A. y van Westing, L. M. (2000). “Flavonoid and chlorogenic acid levels in apple fruit: characterisation of variation”. *Scientia Horticulturae* 83: 249-263.

Barth, S. W., Fahndrich, C., Bub, A., Dietrich, H., Watzl, B., Will, F., Briviba, K. y Rechkemmer, G. (2005). “Cloudy apple juice decreases DNA damage, hyperproliferation and aberrant crypt foci development in the distal colon of DMH-initiated rats”. *Carcinogenesis* 26: 1414-1421.

Boccia, M. M., Kopf, S. R. y Baratti, C. M. (1999). “Phlorizin, a competitive inhibitor of glucose transport, facilitates memory storage in mice”. *Neurobiology of Learning and Memory* 71: 104-112.

Cox, S. D., Jayasinghe, K. C. y Markham, J. L. (2005). “Antioxidant activity in Australian native sarsaparilla (Smilax glyciphylla)”. *Journal of Ethnopharmacology* 101: 162-168.

Dong, H., Ning, Z., Yu, L., Li, L., Lin, L. y Huang, J. (2007). “Preparative separation and identification of the flavonoid phlorhizin from the crude extract of Lithocarpus polystachyus Rehd”. *Molecules* 12: 552-62.

Dragovic-Uzelac, V., Pospisil, J., Levaj, B. y Delonga, K. (2005). “The study of phenolic profiles of raw apricots and apples and their purees by HPLC for the evaluation of apricot nectars and jams authenticity”. *Food Chemistry* 91: 373-383.

Ehrenkranz, J. R., Lewis, N. G., Kahn, C. R. y Roth, J. (2005). “Phlorizin: a review”. *Diabetes/Metabolism Research and Reviews* 21: 31-38.

Fukuchi-Mizutani, M., Okuhara, H., Fukui, Y., Nakao, M., Katsumoto, Y., Yonekura-Sakakibara, K., Kusumi, T., Hase, T. y Tanaka, Y. (2003). “Biochemical and molecular characterization of a novel UDP-glucose:anthocyanin 3'-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian”. *Plant Physiology* 132: 1652-1663.

Gessler, C., Patocchi, A., Sansavini, S., Tartarini, S. y Gianfranceschi, L. (2006). “Venturia inaequalis resistance in apple”. *Critical Reviews in Plant Sciences* 25: 473-503.

Green, S., Friel, E. N., Matich, A., Beuning, L. L., Cooney, J. M., Rowan, D. D. y MacRae, E. (2007). “Unusual features of a recombinant apple alpha-farnesene synthase”. *Phytochemistry* 68: 176-188.

Hall, T. A. 1999. “BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT”. *Nucl. Acids. Symp. Ser*. 41:95-98.

Hansen, K. S., Kristensen, C., Tattersall, D. B., Jones, P. R., Olsen, C. E., Bak, S. y Moller, B. L. (2003). “The in vitro substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from Sorghum bicolor”. *Phytochemistry* 64: 143-151.

Hefner, T. y Stockigt, J. (2003). “Probing suggested catalytic domains of glycosyltransferases by site-directed mutagenesis”. *European Journal of Biochemistry* 270: 533-538.

Hellens, R. P., Allan, A. C., Friel, E. N., Bolitho, K., Grafton, K., Templeton, M. D., Karunairetnam, S. y Laing, W. A. (2005), “Transient plant expression vectors for functional genomics, quantification of promoter activity and RNA silencing”. *Plant Methods*, 1:13

Hilt, P., Schieber, A., Yildirim, C., Arnold, G., Klaiber, I., Conrad, J., Beifuss, U. y Carle, R. (2003). “Detection of phloridzin in strawberries (Fragaria x ananassa Duch.) by HPLC-PDA-MS/MS and NMR spectroscopy”. *Journal of Agricultural and Food Chemistry* 51: 2896-2899.

Hunter, M. D. y Hull, L. A. (1993). “Variation in concentrations of phlorhizin and phloretin in apple foliage”. *Phytochemistry* 34: 1251-1254.

Jackson, R. G., Lim, E. K., Li, Y., Kowalczyk, M., Sandberg, G., Hoggett, J., Ashford, D. A. y Bowles, D. J. (2001). “Identification and biochemical characterization of an Arabidopsis indole-3-acetic acid glucosyltransferase”. *Journal of Biological Chemistry* 276: 4350-4356.

Jones, P., Messner, B., Nakajima, J., Schaffner, A. R. y Saito, K. (2003). “UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana”. *Journal of Biological Chemistry* 278: 43910-43918.

Li, Y., Baldauf, S., Lim, E. K. y Bowles, D. J. (2001). “Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana”. *Journal of Biological Chemistry* 276: 4338-4343.

Lim, E. K., Ashford, D. A., Hou, B., Jackson, R. G. y Bowles, D. J. (2004). “Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides”. *Biotechnology and Bioengineering* 87: 623-631.

Lim, E. K. y Bowles, D. J. (2004). “A class of plant glycosyltransferases involved in cellular homeostasis”. *EMBO Journal* 23: 2915-2922.

Lim, E. K., Doucet, C. J., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Ross, J. y Bowles, D. J. (2002). “The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates”. J*ournal of Biological Chemistry* 277: 586-592.

Lim, E. K., Higgins, G. S., Li, Y. y Bowles, D. J. (2003). “Regioselectivity of glucosylation of caffeic acid by a UDP-glucose:glucosyltransferase is maintained in planta”. *Biochemical Journal* 373: 987-992.

Lim, E. K., Jackson, R. G. y Bowles, D. J. (2005). “Identification and characterisation of Arabidopsis glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde”. *FEBS Letters* 579: 2802-2806.

Lommen, A., Godejohann, M., Venema, D. P., Hollman, P. C. y Spraul, M. (2000). “Application of directly coupled HPLC-NMR-MS to the identification and confirmation of quercetin glycosides and phloretin glycosides in apple peel”. *Analytical Chemistry* 72: 1793-1797.

Loutre, C., Dixon, D. P., Brazier, M., Slater, M., Cole, D. J. y Edwards, R. (2003). “Isolation of a glucosyltransferase from Arabidopsis thaliana active in the metabolism of the persistent pollutant 3,4-dichloroaniline”. *Plant Journal* 34: 485-493.

MacDonald, R. E. y Bishop, C. J. (1952). “Phloretin: An antibacterial substance obtained from apple leaves”. *Canadian Journal of Botany* 30: 486-489.

Mato, M., Ozeki, Y., Itoh, Y., Higeta, D., Yoshitama, K., Teramoto, S., Aida, R., Ishikura, N. y Shibata, M. (1998). “Isolation and characterization of a cDNA clone of UDP-galactose: flavonoid 3-O-galactosyltransferase (UF3GaT) expressed in Vigna mungo seedlings”. *Plant and Cell Physiology* 39: 1145-1155.

Messner, B., Thulke, O. y Schaffner, A. R. (2003). “Arabidopsis glucosyltransferases with activities toward both endogenous and xenobiotic substrates”. *Planta* 217: 138-146.

Newcomb, R. D., Crowhurst, R. N., Gleave, A. P., Rikkerink, E. H. A., Allan, A. C., Beuning, L. L., Bowen, J. H., Gera, E., Jamieson, K. R., Janssen, B. J., Laing, W. A., McArtney, S., Nain, B., Ross, G. S., Snowden, K. C., Souleyre, E. J. F., Walton, E. F. y Yauk, Y.-K. (2006). “Analyses of expressed sequence tags from apple”. *Plant Physiology* 141: 147-166.

Nieuwenhuizen, N. J., Beuning, L. L., Sutherland, P. W., Sharma, N. N., Cooney, J. M., Bieleski, L. R. F., Schröder, R., MacRae, E. A. y Atkinson, R. G. (2007). “Identification and characterisation of acidic and novel basic forms of actinidin, the highly abundant cysteine protease from kiwifruit”. *Functional Plant Biology* 34: 946-961.

Petersen, C. (1835). “Analyse des phloridzins”. *Annales Academie Science Francaise* 15: 178.

Puel, C., Quintin, A., Mathey, J., Obled, C., Davicco, M. J., Lebecque, P., Kati-Coulibaly, S., Horcajada, M. N. y Coxam, V. (2005). “Prevention of bone loss by phloridzin, an apple polyphenol, in ovariectomized rats under inflammation conditions”. *Calcified Tissue International* 77: 311-318.

Rezk, B. M., Haenen, G. R., van der Vijgh, W. J. y Bast, A. (2002). “The antioxidant activity of phloretin: the disclosure of a new antioxidant pharmacophore in flavonoids”. *Biochemical and Biophysical Research Communications* 295: 9-13.

Richman, A., Swanson, A., Humphrey, T., Chapman, R., McGarvey, B., Pocs, R. y Brandle, J. (2005). “Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of Stevia rebaudiana”. *Plant Journal* 41: 56-67.

Ridgway, T., O'Reilly, J., West, G., Tucker, G. y Wiseman, H. (1996). “Potent antioxidant properties of novel apple-derived flavonoids with commercial potential as food additives”. *Biochemical Society Transactions* 24: 391S.

Ridgway, T., O'Reilly, J., West, G., Tucker, G. y Wiseman, H. (1997). “Antioxidant action of novel derivatives of the apple-derived flavonoid phloridzin compared to oestrogen: relevance to potential cardioprotective action”. *Biochemical Society Transactions* 25: 106S.

Ridgway, T. y Tucker, G. (1997). “Phloridzin derivatives; food additives/chemopreventative drugs of the future”. *Biochemical Society Transactions* 25: 109S.

Ross, J., Li, Y., Lim, E. y Bowles, D. J. (2001). “Higher plant glycosyltransferases”. *Genome Biology* 2: 3004.1-3004.6.

Schaffer, R. J., Friel, E. N., Souleyre, E. J., Bolitho, K., Thodey, K., Ledger, S., Bowen, J. H., Ma, J. H., Nain, B., Cohen, D., Gleave, A. P., Crowhurst, R. N., Janssen, B. J., Yao, J. L. y Newcomb, R. D. (2007). “A genomics approach reveals that aroma production in apple is controlled by ethylene predominantly at the final step in each biosynthetic pathway”. *Plant Physiology* 144: 1899-912.

Studier, F. W. (2005). “Protein production by auto-induction in high-density shaking cultures”. *Protein Expression and Purification* 41: 207-234.

Taguchi, G., Nakamura, M., Hayashida, N. y Okazaki, M. (2003). “Exogenously added naphthols induce three glucosyltransferases, and are accumulated as glucosides in tobacco cells”. *Plant Science* 164: 231-240.
Veeriah, S., Kautenburger, T., Habermann, N., Sauer, J., Dietrich, H., Will, F. y Pool-Zobel, B. L. (2006). “Apple flavonoids inhibit growth of HT29 human colon cancer cells and modulate expression of genes involved in the biotransformation of xenobiotics”. *Molecular Carcinogenesis* 45: 164-174.
Vogt, T. (2000). “Glycosyltransferases involved in plant secondary metabolism. Evolution of Metabolic Pathways”. J. T. Romeo, R. Ibrahim, L. Varin y V. De Luca. Oxford, Elsevier Science Ltd.: 317-347.
Watts, K. T., Lee, P. C. y Schmidt-Dannert, C. (2004). “Exploring recombinant flavonoid biosynthesis in metabolically engineered Escherichia coli”. *Chembiochem* 5: 500-507.
Whiting, G. C. y Coggins, R. A. (1975). “Estimation of the monomeric phenolics from ciders”. *Journal of the Science of Food and Agriculture* 26: 1833-1838.
Zhang, X. Z., Zhao, Y. B., Li, C. M., Chen, D. M., Wang, G. P., Chang, R. F. y Shu, H. R. (2007). “Potential polyphenol markers of phase change in apple (Malus domestica)”. *Journal of Plant Physiology* 164: 574-580.

LISTADO DE SECUENCIAS

<110> THE HORTICULTURE AND FOOD RESEARCH INSTITUTE OF NEW ZEALAND

<120> GLICOSILTRANSFERASAS, POLINUCLEÓTIDOS QUE LAS CODIFICAN Y MÉTODOS DE USO

<130> 586955 HCF

<150> NZ 564691
<151> 21-12-2007

<160> 21

<170> PatentIn versión 3.3

<210> 1
<211> 483
<212> PRT
<213> *Malus domestica*

<400> 1

Met Gly Asp Val Ile Val Leu Tyr Ala Ser Pro Gly Met Gly His Ile
1 5 10 15

Val Ser Met Val Glu Leu Gly Lys Phe Ile Val His Arg Tyr Gly Pro
20 25 30

His Lys Phe Ser Ile Thr Ile Leu Tyr Thr Cys Gly Ser Ile Val Asp
35 40 45

Thr Ala Ser Ile Pro Val Tyr Ile Arg Arg Ile Ser His Ser His Pro
50 55 60

Phe Ile Ser Phe Arg Gln Phe Pro Arg Val Thr Asn Asn Ile Thr Arg
65 70 75 80

Asn Ile Ser Val Pro Ala Ile Thr Phe Asp Phe Ile Arg Gln Asn Asp
85 90 95

Pro His Val Arg Ser Ala Leu Gln Glu Ile Ser Lys Ser Ala Thr Val
100 105 110

Arg Ala Phe Ile Ile Asp Leu Phe Cys Thr Ser Ala Leu Pro Ile Gly
115 120 125

Lys Glu Phe Asn Ile Pro Thr Tyr Tyr Phe Cys Thr Ser Gly Ala Ala
130 135 140

Ile Leu Ala Ala Phe Leu Tyr Leu Pro Lys Ile Asp Glu Gln Thr Lys
145 150 155 160

Thr Thr Glu Ser Phe Lys Asp Leu Arg Asp Thr Val Phe Glu Phe Pro
165 170 175

Gly Trp Lys Ser Pro Leu Lys Ala Thr His Met Val Gln Leu Val Leu

180 185 190

Asp Arg Asn Asp Pro Ala Tyr Ser Asp Met Ile Tyr Phe Cys Ser His
195 200 205

Leu Pro Lys Ser Asn Gly Ile Ile Val Asn Thr Phe Glu Glu Leu Glu
210 215 220

Pro Pro Ser Val Leu Gln Ala Ile Ala Gly Gly Leu Cys Val Pro Asp
225 230 235 240

Gly Pro Thr Pro Pro Val Tyr Tyr Val Gly Pro Leu Ile Glu Glu Glu
245 250 255

Lys Glu Leu Ser Lys Asp Ala Asp Ala Ala Glu Lys Glu Asp Cys Leu
260 265 270

Ser Trp Leu Asp Lys Gln Pro Ser Arg Ser Val Leu Phe Leu Cys Phe
275 280 285

Gly Ser Met Gly Ser Phe Pro Ala Ala Gln Leu Lys Glu Ile Ala Asn
290 295 300

Gly Leu Glu Ala Ser Gly Gln Arg Phe Leu Trp Val Val Lys Lys Pro
305 310 315 320

Pro Val Glu Glu Lys Ser Lys Gln Val His Gly Val Asp Asp Phe Asp
325 330 335

Leu Lys Gly Val Leu Pro Glu Gly Phe Leu Glu Arg Thr Ala Asp Arg
340 345 350

Gly Met Val Val Lys Ser Trp Ala Pro Gln Val Val Val Leu Lys Lys
355 360 365

Glu Ser Val Gly Gly Phe Val Thr His Cys Gly Trp Asn Ser Val Leu
370 375 380

Glu Ala Val Val Ala Gly Val Pro Met Ile Ala Trp Pro Leu Tyr Ala
385 390 395 400

Glu Gln His Met Asn Arg Asn Val Leu Val Thr Asp Met Glu Ile Ala
405 410 415

Ile Gly Val Glu Gln Arg Asp Glu Glu Gly Gly Phe Val Ser Gly Glu
420 425 430

Glu Val Glu Arg Arg Val Arg Glu Leu Met Glu Ser Glu Gly Gly Arg
435 440 445

Val Leu Arg Glu Arg Cys Lys Lys Leu Gly Glu Met Ala Ser Ala Ala

450 455 460

Leu Gly Glu Thr Gly Ser Ser Thr Arg Asn Leu Val Asn Phe Val Ser
465 470 475 480

Ser Ile Thr

<210> 2
<211> 483
<212> PRT
<213> *Malus sieboldii*

<400> 2

Met Gly Asp Val Ile Val Leu Tyr Ala Ser Pro Gly Met Gly His Ile
1 5 10 15

Val Ala Met Val Glu Leu Gly Lys Phe Ile Val His Arg Tyr Gly Pro
20 25 30

His Lys Phe Ser Ile Thr Ile Leu Tyr Thr Cys Gly Ser Ile Val Asp
35 40 45

Thr Ala Ser Ile Pro Val Tyr Ile Arg Arg Ile Ser His Ser His Pro
50 55 60

Phe Ile Ser Phe Arg Gln Phe Pro Arg Val Thr Asn Asn Ile Thr Arg
65 70 75 80

Asn Ile Ser Val Pro Ala Ile Thr Phe Asp Phe Ile Arg Gln Asn Asp
85 90 95

Pro His Val Arg Ser Ala Leu Gln Glu Ile Ser Lys Ser Ala Thr Val
100 105 110

Arg Ala Phe Ile Ile Asp Leu Phe Cys Thr Ser Ala Leu Pro Ile Gly
115 120 125

Lys Glu Phe Asn Ile Pro Thr Tyr Tyr Phe Cys Thr Ser Gly Ala Ala
130 135 140

Ile Leu Ala Ala Phe Leu Tyr Leu Pro Lys Ile Asp Glu Gln Thr Lys
145 150 155 160

Thr Thr Glu Ser Phe Lys Asp Leu Arg Asp Thr Val Phe Glu Phe Pro
165 170 175

Gly Trp Lys Ser Pro Leu Lys Ala Thr His Met Val Gln Leu Val Leu
180 185 190

Asp Arg Asn Asp Pro Ala Tyr Ser Asp Met Ile Tyr Phe Cys Ser His
195 200 205

Leu Pro Lys Ser Asn Gly Ile Ile Val Asn Thr Phe Glu Glu Leu Glu
210 215 220

Pro Pro Ser Val Leu Gln Ala Ile Ala Gly Gly Leu Cys Val Pro Asp
225 230 235 240

Gly Pro Thr Pro Pro Val Tyr Tyr Val Gly Pro Leu Ile Glu Glu Glu
245 250 255

Lys Glu Leu Ser Lys Asp Ala Asp Ala Ala Glu Lys Glu Asp Cys Leu
260 265 270

Ser Trp Leu Asp Lys Gln Pro Ser Arg Ser Val Leu Phe Leu Cys Phe
275 280 285

Gly Ser Met Gly Ser Phe Pro Ala Ala Gln Leu Lys Glu Ile Ala Asn
290 295 300

Gly Leu Glu Ala Ser Gly Gln Arg Phe Leu Trp Val Val Lys Lys Pro
305 310 315 320

Pro Val Glu Glu Lys Ser Lys Gln Val His Gly Val Asp Asp Phe Asp
325 330 335

Leu Lys Gly Val Leu Pro Glu Gly Phe Leu Glu Arg Thr Ala Asp Arg
340 345 350

Gly Met Val Val Lys Ser Trp Ala Pro Gln Val Val Val Leu Lys Lys
355 360 365

Glu Ser Val Gly Gly Phe Val Thr His Cys Gly Trp Asn Ser Val Leu
370 375 380

Glu Ala Val Val Ala Gly Val Pro Met Ile Ala Trp Pro Leu Tyr Ala
385 390 395 400

Glu Gln His Met Asn Arg Asn Val Leu Val Thr Asp Met Glu Ile Ala
405 410 415

Ile Gly Val Glu Gln Arg Asp Glu Glu Gly Gly Phe Val Ser Gly Glu
420 425 430

Glu Val Glu Arg Arg Val Arg Glu Leu Met Glu Ser Glu Gly Gly Arg
435 440 445

Ala Leu Arg Glu Arg Cys Lys Lys Leu Gly Glu Met Ala Ser Ala Ala
450 455 460

Leu Gly Glu Thr Gly Ser Ser Thr Arg Asn Leu Val Asn Phe Val Ser
465 470 475 480

Ser Ile Thr

<210> 3
<211> 483
<212> PRT
<213> *Malus domestica*

<400> 3

```
Met Gly Asp Val Ile Val Leu Tyr Ala Ser Pro Gly Met Gly His Ile
1               5                   10                  15

Val Ser Met Val Glu Leu Gly Lys Phe Ile Val His Arg Tyr Gly Pro
            20                  25                  30

His Lys Phe Ser Ile Thr Ile Leu Tyr Thr Cys Gly Ser Ile Val Asp
        35                  40                  45

Thr Ala Ser Ile Pro Val Tyr Ile Arg Arg Ile Ser His Ser His Pro
    50                  55                  60

Phe Ile Ser Phe Arg Gln Phe Pro Arg Val Thr Asn Asn Ile Thr Arg
65                  70                  75                  80

Asn Ile Ser Val Pro Ala Ile Thr Phe Asp Phe Ile Arg Gln Asn Asp
                85                  90                  95

Pro His Val Arg Ser Ala Leu Gln Glu Ile Ser Lys Ser Ala Thr Val
            100                 105                 110

Arg Ala Phe Ile Ile Asp Leu Phe Cys Thr Ser Ala Leu Pro Ile Gly
        115                 120                 125

Lys Glu Phe Asn Ile Pro Thr Tyr Tyr Phe Cys Thr Ser Gly Ala Ala
    130                 135                 140

Ile Leu Ala Ala Phe Leu Tyr Leu Pro Lys Ile Asp Glu Gln Thr Lys
145                 150                 155                 160

Thr Thr Glu Ser Phe Lys Asp Leu Arg Asp Thr Val Phe Glu Phe Pro
                165                 170                 175

Gly Trp Lys Ser Pro Leu Lys Ala Thr His Met Val Gln Leu Val Leu
            180                 185                 190

Asp Arg Asn Asp Pro Ala Tyr Ser Asp Met Ile Tyr Phe Cys Ser His
        195                 200                 205

Leu Pro Lys Ser Asn Gly Ile Ile Val Asn Thr Phe Glu Glu Leu Glu
    210                 215                 220
```

Pro Pro Ser Val Leu Gln Ala Ile Ala Gly Gly Leu Cys Val Pro Asp
225 230 235 240

Gly Pro Thr Pro Pro Val Tyr Tyr Val Gly Pro Leu Ile Glu Glu Glu
245 250 255

Lys Glu Leu Ser Lys Asp Ala Asp Ala Ala Glu Lys Glu Asp Cys Leu
260 265 270

Ser Trp Leu Asp Lys Gln Pro Ser Arg Ser Val Leu Phe Leu Cys Phe
275 280 285

Gly Ser Met Gly Ser Phe Pro Ala Ala Gln Leu Lys Glu Ile Ala Asn
290 295 300

Gly Leu Glu Ala Ser Gly Gln Arg Phe Leu Trp Val Val Lys Lys Pro
305 310 315 320

Pro Val Glu Glu Lys Ser Lys Gln Val His Gly Val Asp Asp Phe Asp
325 330 335

Leu Lys Gly Val Leu Pro Glu Gly Phe Leu Glu Arg Thr Ala Asp Arg
340 345 350

Gly Met Val Val Lys Ser Trp Ala Pro Gln Val Val Val Leu Lys Lys
355 360 365

Glu Ser Val Gly Gly Phe Val Thr His Cys Gly Trp Asn Ser Val Leu
370 375 380

Glu Ala Val Val Ala Gly Val Pro Met Ile Ala Trp Pro Leu Tyr Ala
385 390 395 400

Glu Gln His Met Asn Arg Asn Val Leu Val Thr Asp Met Glu Leu Ala
405 410 415

Ile Gly Val Glu Gln Arg Asp Glu Glu Gly Gly Phe Val Ser Gly Glu
420 425 430

Glu Val Glu Arg Arg Val Arg Glu Leu Met Glu Ser Glu Gly Gly Arg
435 440 445

Val Leu Arg Glu Arg Cys Lys Lys Leu Gly Glu Met Ala Ser Ala Ala
450 455 460

Leu Gly Glu Thr Gly Ser Ser Thr Arg Asn Leu Val Asn Phe Val Ser
465 470 475 480

Ser Ile Thr

<210> 4
<211> 483
<212> PRT
<213> *Malus domestica*

<400> 4

Met Gly Asp Val Ile Val Leu Tyr Ala Ser Pro Gly Met Gly His Ile
1 5 10 15

Val Ser Met Val Glu Leu Gly Lys Phe Ile Val His Arg Tyr Gly Pro
20 25 30

His Lys Phe Ser Ile Thr Ile Leu Tyr Thr Cys Gly Ser Ile Val Asp
35 40 45

Thr Ala Ser Ile Pro Val Tyr Ile Arg Arg Ile Ser His Ser His Pro
50 55 60

Phe Ile Ser Phe Arg Gln Phe Pro Arg Val Thr Asn Asn Ile Thr Arg
65 70 75 80

Asn Ile Ser Val Pro Ala Ile Thr Phe Asp Phe Ile Arg Gln Asn Asp
85 90 95

Pro His Val Arg Ser Ala Leu Gln Glu Ile Ser Lys Ser Ala Thr Val
100 105 110

Arg Ala Phe Ile Ile Asp Leu Phe Cys Thr Ser Ala Leu Pro Ile Gly
115 120 125

Lys Glu Phe Asn Ile Pro Thr Tyr Tyr Phe Cys Thr Ser Gly Ala Ala
130 135 140

Ile Leu Ala Ala Phe Leu Tyr Leu Pro Lys Ile Asp Glu Gln Thr Lys
145 150 155 160

Thr Thr Glu Ser Phe Lys Asp Leu Arg Asp Thr Val Phe Glu Phe Pro
165 170 175

Gly Trp Lys Ser Pro Leu Lys Ala Pro His Met Val Gln Leu Val Leu
180 185 190

Asp Arg Asn Asp Pro Ala Tyr Ser Asp Met Ile Tyr Phe Cys Ser His
195 200 205

Leu Pro Lys Ser Asn Gly Ile Ile Val Asn Thr Phe Glu Glu Leu Glu
210 215 220

Pro Pro Ala Val Leu Gln Ala Ile Ala Gly Gly Leu Cys Val Pro Asp
225 230 235 240

Gly Pro Thr Pro Pro Val Tyr Tyr Val Gly Pro Leu Ile Glu Glu Glu
245 250 255

Lys Glu Leu Ser Lys Asp Ala Asp Ala Ala Gly Glu Glu Asp Cys Leu
260 265 270

Ser Trp Leu Asp Lys Gln Pro Ser Arg Ser Val Leu Phe Leu Cys Phe
275 280 285

Gly Ser Met Gly Ser Phe Pro Ala Ala Gln Leu Lys Glu Ile Ala Asn
290 295 300

Gly Leu Glu Ala Ser Gly Gln Arg Phe Leu Trp Val Val Lys Lys Pro
305 310 315 320

Pro Val Glu Glu Lys Ser Lys Gln Val His Gly Val Asp Asp Phe Asp
325 330 335

Leu Lys Gly Val Leu Pro Glu Gly Phe Leu Glu Arg Thr Ala Asp Arg
340 345 350

Gly Met Val Val Lys Ser Trp Ala Pro Gln Val Val Val Leu Lys Lys
355 360 365

Glu Ser Val Gly Gly Phe Val Thr His Cys Gly Trp Asn Ser Val Leu
370 375 380

Glu Ala Val Val Ala Gly Val Pro Met Ile Ala Trp Pro Leu Tyr Ala
385 390 395 400

Glu Gln His Met Asn Arg Asn Val Leu Val Thr Asp Met Glu Ile Ala
405 410 415

Ile Gly Val Glu Gln Arg Asp Glu Glu Gly Gly Phe Val Ser Gly Glu
420 425 430

Glu Val Glu Arg Arg Val Arg Glu Leu Met Glu Ser Glu Gly Gly Arg
435 440 445

Ala Leu Arg Glu Arg Cys Lys Lys Leu Gly Glu Met Ala Ser Ala Ala
450 455 460

Leu Gly Glu Thr Gly Ser Ser Thr Arg Asn Leu Val Asn Phe Val Ser
465 470 475 480

Ser Ile Thr

<210> 5
<211> 480
<212> PRT
<213> *Malus domestica*

<400> 5

Met Gly Asp Val Ile Val Leu Tyr Ala Ala Pro Gly Met Gly His Val
1 5 10 15

Ile Ser Met Val Glu Leu Gly Lys Leu Ile Leu His Arg Tyr Gly Pro
20 25 30

His Lys Phe Ser Ile Thr Ile Leu Tyr Thr Cys Gly Ser Phe Phe Asp
35 40 45

Thr Pro Ser Ile Pro Ala Tyr Ile Arg Arg Ile Ser His Ser His Pro
50 55 60

Ser Ile Ser Phe Arg Gln Phe Pro Arg Val Thr Asn Lys Ile Thr Gln
65 70 75 80

Asn Ile Ser Gly Thr Ala Ile Val Val Asp Phe Val Arg Gln Asn Asp
85 90 95

Pro His Val Arg Arg Ala Leu Gln Asp Ile Ser Lys Ser Ala Val Val
100 105 110

Arg Ala Phe Ile Ile Asp Leu Val Cys Thr Ser Ala Met Thr Ile Ser
115 120 125

Lys Glu Phe Asp Ile Pro Thr Tyr Tyr Phe Tyr Thr Ser Gly Ala Ala
130 135 140

Ala Leu Gly Ala Phe Leu Tyr Phe Pro Lys Ile His Glu Gln Thr Thr
145 150 155 160

Gln Ser Phe Lys Asp Leu Thr Asp Thr Val Ile Glu Phe Pro Gly Arg
165 170 175

Lys Ser Pro Leu Lys Ala Ile His Met Ile Glu Pro Leu Leu Asp Arg
180 185 190

Asp Asp Pro Ala Tyr Trp Asp Phe Leu Ser Phe Cys Ser Asp Leu Pro
195 200 205

Lys Ser Lys Gly Ile Ile Val Asn Thr Phe Glu Glu Leu Glu Pro Pro
210 215 220

Ala Val Leu His Ala Ile Ala Glu Gly Leu Cys Val Pro Asp Gly Pro
225 230 235 240

Thr Ser Pro Val Tyr Tyr Val Gly Pro Leu Ile Asp Glu Glu Lys Val
245 250 255

Ser Gly Asn Asp Ala Ala Ala Ala Glu Glu Asp Cys Leu Ser Trp Leu

260 265 270

Asp Lys Gln Pro Ser Arg Ser Val Val Phe Leu Cys Phe Gly Ser Arg
275 280 285

Gly Ser Leu Pro Ala Ile Gln Leu Lys Glu Ile Ala Lys Ala Leu Glu
290 295 300

Ala Ser Gly Gln Arg Phe Leu Trp Val Val Lys Lys Pro Pro Val Asp
305 310 315 320

Glu Lys Thr Lys Gln Val Leu Gly Val Asp Asp Phe Asp Leu Glu Gly
325 330 335

Val Leu Pro Glu Gly Phe Leu Glu Arg Thr Lys Asp Arg Gly Met Val
340 345 350

Val Lys Ser Trp Ala Pro Gln Ala Glu Val Leu Lys Lys Glu Ser Val
355 360 365

Gly Gly Phe Val Thr His Cys Gly Trp Asn Ser Val Leu Glu Ala Val
370 375 380

Val Ala Gly Val Pro Met Ile Ala Trp Pro Leu Tyr Ala Glu Gln His
385 390 395 400

Leu Asn Arg Asn Val Met Ala Thr Asp Met Glu Ile Ala Ile Ala Val
405 410 415

Glu Gln Arg Asp Glu Glu Asp Gly Phe Val Ser Gly Glu Glu Leu Glu
420 425 430

Arg Arg Val Arg Glu Leu Met Glu Ser Glu Glu Gly Arg Val Leu Arg
435 440 445

Glu Arg Ser Lys Lys Ile Gly Glu Met Ala Met Ala Ala Leu Gly Glu
450 455 460

Asn Gly Ser Ser Thr Arg Asn Leu Val Asn Phe Val Asn Ser Leu Thr
465 470 475 480

<210> 6
<211> 1696
<212> ADN
<213> *Malus domestica*

<400> 6

gtcgcagctg ccgccatggg agacgtcatt gtactgtacg catctccggg gatggggcat 60
atcgtctcca tggtggagct gggcaagttc attgtccacc gctacggccc ccacaaattc 120
tccatcacca ttctctacac ctgcggcagc attgtcgaca ccgctagcat ccccgtctac 180
atccgccgca tctcccactc ccacccttc atttccttcc gccaattccc tcgcgtcacc 240

aataatatta cccgaaacat aagcgtcccc gcaatcacgt tcgacttcat ccgccagaac 300
gatcctcatg tccgcagtgc cctccaagaa atctctaaat ccgccaccgt tcgcgccttc 360
atcatcgacc tcttctgcac ctccgctctt cccataggga aggaattcaa catcccaaca 420
tactacttct gcacttctgg tgccgcaatt cttgctgctt ttttgtattt gcccaagatc 480
gatgagcaaa ccaaaaccac cgagagtttc aaagacctcc gcgacaccgt tttcgaattc 540
cccggatgga agtctcctct gaaggctaca cacatggtcc aactggtgct cgaccggaac 600
gaccctgctt attcggacat gatctatttc tgctcacatc ttcccaaatc caacggaatc 660
atcgtcaaca cgttcgaaga gctcgagcca cctagcgtcc tccaggccat tgctggaggc 720
ctgtgtgttc ctgatgggcc aactccgccc gtgtactacg ttggtccatt gattgaggaa 780
gagaaagaat tgagtaagga tgcagatgcg gccgagaagg aggactgctt gtcatggctc 840
gataagcagc caagtcgaag cgtgctgttt ctctgtttcg gaagcatggg atcatttccg 900
gctgctcaac tgaaggagat agcgaacggg ttggaggcga gcgggcagag gttcctgtgg 960
gtggtgaaga agccgccggt tgaagagaaa tcaaagcagg tccatggagt tgacgacttt 1020
gatttgaagg gtgtgttgcc agaagggttt ttggagagga cggccgacag ggggatggta 1080
gtgaagtcat gggcgccgca ggtggtggtg ttgaagaagg agtcggttgg tgggttcgtg 1140
acacattgcg gatggaactc ggtactggaa gcagtggttg cgggggtgcc gatgattgct 1200
tggccgcttt acgcggagca gcatatgaac aggaatgttc tagtgacgga catggaaatc 1260
gcgatcgggg tggagcagag agacgaggaa ggtgggttcg tgagcgggga agaagtggag 1320
aggagagtga gggagttgat ggagtcggaa ggaggaagag tgcttagaga gaggtgcaag 1380
aaacttgggg agatggcttc ggctgctttg ggagagaccg gttcgtccac cagaaacttg 1440
gtcaactttg ttagtagcat tacataaccg tgcaagtttt gggttcgcat ttcgcatttg 1500
gatttcgaga ggagtcgtgt gcaagaataa accctcttaa ccaactccgc gatatggttt 1560
tctgttactt agtttggcac ctcgtgggat tttggtgctc tcttttgaga gttttaagaa 1620
agcatcaaac tattcggatg taatgttgat gtaacgtcga ccatcagctt aattccaacc 1680
aataagatta tgtaaa 1696

<210> 7
<211> 1452
<212> ADN
<213> *Malus domestica*

<400> 7

atgggagacg tcattgtact gtacgcatct ccggggatgg ggcatatcgt ctccatggtg 60
gagctgggca agttcattgt ccaccgctac ggcccccaca aattctccat caccattctc 120
tacacctgcg gcagcattgt cgacaccgct agcatccccg tctacatccg ccgcatctcc 180
cactcccacc ctttcatttc cttccgccaa ttccctcgcg tcaccaataa tattacccga 240
aacataagcg tccccgcaat cacgttcgac ttcatccgcc agaacgatcc tcatgtccgc 300

agtgccctcc aagaaatctc taaatccgcc accgttcgcg ccttcatcat cgacctcttc 360
tgcacctccg ctcttcccat agggaaggaa ttcaacatcc caacatacta cttctgcact 420
tctggtgccg caattcttgc tgctttttg tatttgccca agatcgatga gcaaaccaaa 480
accaccgaga gtttcaaaga cctccgcgac accgttttcg aattccccgg atggaagtct 540
cctctgaagg ctacacacat ggtccaactg gtgctcgacc ggaacgaccc tgcttattcg 600
gacatgatct atttctgctc acatcttccc aaatccaacg gaatcatcgt caacacgttc 660
gaagagctcg agccacctag cgtcctccag gccattgctg gaggcctgtg tgttcctgat 720
gggccaactc cgcccgtgta ctacgttggt ccattgattg aggaagagaa agaattgagt 780
aaggatgcag atgcggccga gaaggaggac tgcttgtcat ggctcgataa gcagccaagt 840
cgaagcgtgc tgtttctctg tttcggaagc atgggatcat ttccggctgc tcaactgaag 900
gagatagcga acgggttgga ggcgagcggg cagaggttcc tgtgggtggt gaagaagccg 960
ccggttgaag agaaatcaaa gcaggtccat ggagttgacg actttgattt gaagggtgtg 1020
ttgccagaag ggtttttgga gaggacggcc gacaggggga tggtagtgaa gtcatgggcg 1080
ccgcaggtgg tggtgttgaa gaaggagtcg gttggtgggt tcgtgacaca ttgcggatgg 1140
aactcggtac tggaagcagt ggttgcgggg gtgccgatga ttgcttggcc gctttacgcg 1200
gagcagcata tgaacaggaa tgttctagtg acggacatgg aaatcgcgat cggggtggag 1260
cagagagacg aggaaggtgg gttcgtgagc ggggaagaag tggagaggag agtgagggag 1320
ttgatggagt cggaaggagg aagagtgctt agagagaggt gcaagaaact tggggagatg 1380
gcttcggctg ctttgggaga gaccggttcg tccaccagaa acttggtcaa ctttgttagt 1440
agcattacat aa 1452

<210> 8
<211> 1714
<212> ADN
<213> *Malus sieboldii*

<400> 8

cgcagctgcc gccatgggag acgtcattgt actgtacgca tctccaggga tggggcacat 60
cgtcgccatg gtggagctgg gcaagttcat tgtccaccgc tacggccccc acaaattctc 120
catcaccatt ctctacacct gcggcagcat tgtcgacacc gctagcatcc ccgtctacat 180
ccgccgcatc tcccactccc acccttttcat ttccttccgc caattccctc gcgtcaccaa 240
taatattacc cgaaacataa gcgtccccgc aatcacgttc gacttcatcc gccagaacga 300
tcctcatgtc cgcagtgccc tccaagaaat ctctaaatcc gccaccgttc gcgccttcat 360
catcgacctc ttctgcacct ccgctcttcc catagggaag gaattcaaca tcccaacata 420
ctacttctgc acttctggtg ccgcaattct tgctgctttt ttgtatttgc ccaagatcga 480
tgagcaaacc aaaaccaccg agagtttcaa agacctccgc gacaccgttt tcgaattccc 540
cggatggaag tctcctctga aggctacaca catggtccaa ctggtgctcg accggaacga 600
ccctgcttat tcggacatga tctatttctg ctcacatctt cccaaatcca acggaatcat 660

```
cgtcaacacg ttcgaagagc tcgagccacc tagcgtcctc caggccattg ctggaggcct    720
gtgtgttcct gatgggccaa ctccgcccgt gtactacgtt ggtccattga ttgaggaaga    780
gaaagaattg agtaaggatg cagatgcggc cgagaaggag gactgcttgt catggctcga    840
taagcagcca agtcgaagcg tgctgtttct ctgtttcgga agcatgggat catttccggc    900
tgctcaactg aaggagatag cgaacgggtt ggaggcgagc gggcagaggt tcctgtgggt    960
ggtgaagaag ccgccggttg aagagaaatc aaagcaggtc catggagttg acgactttga   1020
tttgaagggt gtgttgccag aagggttttt ggagaggacg gcagacaggg ggatggtagt   1080
gaagtcatgg gcgccgcagg tggtggtgtt gaagaaggag tcggttggtg ggttcgtgac   1140
acattgcgga tggaactcgg tactggaagc agtggttgcg ggggtgccga tgattgcttg   1200
gccgctttac gcggagcagc atatgaacag gaatgttcta gtgacggaca tggaaatcgc   1260
gatcggggtg gagcagagag acgaggaagg tgggttcgtg agcggggaag aagtggagag   1320
gagagtgagg gagttgatgg agtcggaagg aggaagagcg cttagagaga ggtgcaagaa   1380
acttggggag atggcttcgg ctgctttggg agagaccggt tcgtccacca gaaacttggt   1440
caactttgtt agtagcatta cataaccgtg caagttttgg gttcgcattt cgcatttgga   1500
tttcgagagg agtcgtgtgc aagaataaac cctcttaacc aactccgcga tatggttttc   1560
tgttacttag tttggcacct cgtgggattt tggtgctctc ttttgagagt tttaagaaag   1620
catcaaacta ttcggatgta atgttgatgt aacgtcgacc atcagcttaa ttccaaccaa   1680
taagattatg tacttctttc ttaaaaaaaa aaaa                               1714
```

<210> 9
<211> 1452
<212> ADN
<213> *Malus sieboldii*

<400> 9

```
atgggagacg tcattgtact gtacgcatct ccagggatgg ggcacatcgt cgccatggtg     60
gagctgggca agttcattgt ccaccgctac ggcccccaca aattctccat caccattctc    120
tacacctgcg gcagcattgt cgacaccgct agcatccccg tctacatccg ccgcatctcc    180
cactcccacc ctttcatttc cttccgccaa ttccctcgcg tcaccaataa tattacccga    240
aacataagcg tccccgcaat cacgttcgac ttcatccgcc agaacgatcc tcatgtccgc    300
agtgccctcc aagaaatctc taaatccgcc accgttcgcg ccttcatcat cgacctcttc    360
tgcacctccg ctcttcccat agggaaggaa ttcaacatcc caacatacta cttctgcact    420
tctggtgccg caattcttgc tgctttttg tatttgccca agatcgatga gcaaaccaaa    480
accaccgaga gtttcaaaga cctccgcgac accgttttcg aattccccgg atggaagtct    540
cctctgaagg ctacacacat ggtccaactg gtgctcgacc ggaacgaccc tgcttattcg    600
gacatgatct atttctgctc acatcttccc aaatccaacg gaatcatcgt caacacgttc    660
gaagagctcg agccacctag cgtcctccag gccattgctg gaggcctgtg tgttcctgat    720
```

gggccaactc cgcccgtgta ctacgttggt ccattgattg aggaagagaa agaattgagt 780
aaggatgcag atgcggccga gaaggaggac tgcttgtcat ggctcgataa gcagccaagt 840
cgaagcgtgc tgtttctctg tttcggaagc atgggatcat ttccggctgc tcaactgaag 900
gagatagcga acgggttgga ggcgagcggg cagaggttcc tgtgggtggt gaagaagccg 960
ccggttgaag agaaatcaaa gcaggtccat ggagttgacg actttgattt gaagggtgtg 1020
ttgccagaag ggtttttgga gaggacggca gacaggggga tggtagtgaa gtcatgggcg 1080
ccgcaggtgg tggtgttgaa gaaggagtcg gttggtgggt tcgtgacaca ttgcggatgg 1140
aactcggtac tggaagcagt ggttgcgggg gtgccgatga ttgcttggcc gctttacgcg 1200
gagcagcata tgaacaggaa tgttctagtg acggacatgg aaatcgcgat cggggtggag 1260
cagagagacg aggaaggtgg gttcgtgagc ggggaagaag tggagaggag agtgagggag 1320
ttgatggagt cggaaggagg aagagcgctt agagagaggt gcaagaaact tggggagatg 1380
gcttcggctg ctttgggaga gaccggttcg tccaccagaa acttggtcaa ctttgttagt 1440
agcattacat aa 1452

<210> 10
<211> 1722
<212> ADN
<213> *Malus domestica*

<400> 10

ccaccagtcg cagctgccgc catgggagac gtcattgtac tgtacgcatc tccggggatg 60
gggcatatcg tctccatggt ggagctgggc aagttcattg tccaccgcta cggcccccac 120
aaattctcca tcaccattct ctacacctgc ggcagcattg tcgacaccgc tagcatcccc 180
gtctacatcc gccgcatctc ccactcccac cctttcattt ccttccgcca attccctcgc 240
gtcaccaata atattacccg aaacataagc gtccccgcaa tcacgttcga cttcatccgc 300
cagaacgatc ctcatgtccg cagtgccctc caagaaatct ctaaatccgc caccgttcgc 360
gccttcatca tcgacctctt ctgcacctcc gctcttccca tagggaagga attcaacatc 420
ccaacatact acttctgcac ttctggtgcc gcaattcttg ctgctttttt gtatttgccc 480
aagatcgatg agcaaaccaa aaccaccgag agtttcaaag acctccgcga caccgttttc 540
gaattccccg gatggaagtc tcctctgaag gctacacaca tggtccaact ggtgctcgac 600
cggaacgacc ctgcttattc ggacatgatc tatttctgct cacatcttcc caaatccaac 660
ggaatcatcg tcaacacgtt cgaagagctc gagccaccta gcgtcctcca ggccattgct 720
ggaggcctgt gtgttcctga tgggccaact ccgcccgtgt actacgttgg tccattgatt 780
gaggaagaga aagaattgag taaggatgca gatgcggccg agaaggagga ctgcttgtca 840
tggctcgata agcagccaag tcgaagcgtg ctgtttctct gtttcggaag catgggatca 900
tttccggctg ctcaactgaa ggagatagcg aacgggttgg aggcgagcgg gcagaggttc 960
ctgtgggtgg tgaagaagcc gccggttgaa gagaaatcaa agcaggtcca tggagttgac 1020
gactttgatt tgaagggtgt gttgccagaa gggtttttgg agaggacggc cgacaggggg 1080

atggtagtga agtcatgggc gccgcaggtg gtggtgttga agaaggagtc ggttggtggg 1140
ttcgtgacac attgcggatg gaactcggta ctggaagcag tggttgcggg ggtgccgatg 1200
attgcttggc cgctttacgc ggagcagcat atgaacagga atgttctagt gacggacatg 1260
gaactcgcga tcggggtgga gcagagagac gaggaaggtg ggttcgtgag cggggaagaa 1320
gtggagagga gagtgaggga gttgatggag tcggaaggag gaagagtgct tagagagagg 1380
tgcaagaaac ttggggagat ggcttcggct gctttgggag agaccggttc gtccaccaga 1440
aacttggtca actttgttag tagcattaca taaccgtgca agttttgggt tcgcatttcg 1500
catttggatt tcgagaggag tcgtgtgcaa gaataaaccc tcttaaccaa ctccgcgata 1560
tggttttctg ttacttagtt tggcacctcg tgggattttg gtgctctctt ttgagagttt 1620
taagaaagca tcaaactatt cggatgtaat gttgatgtaa cgtcgaccat cagcttaatt 1680
ccaaccaata agattatgta cttctttctt aaaaaaaaaa aa 1722

<210> 11
<211> 1452
<212> ADN
<213> *Malus domestica*

<400> 11

atgggagacg tcattgtact gtacgcatct ccggggatgg ggcatatcgt ctccatggtg 60
gagctgggca agttcattgt ccaccgctac ggcccccaca aattctccat caccattctc 120
tacacctgcg gcagcattgt cgacaccgct agcatccccg tctacatccg ccgcatctcc 180
cactcccacc ctttcatttc cttccgccaa ttccctcgcg tcaccaataa tattacccga 240
aacataagcg tccccgcaat cacgttcgac ttcatccgcc agaacgatcc tcatgtccgc 300
agtgccctcc aagaaatctc taaatccgcc accgttcgcg ccttcatcat cgacctcttc 360
tgcacctccg ctcttcccat agggaaggaa ttcaacatcc caacatacta cttctgcact 420
tctggtgccg caattcttgc tgctttttg tatttgccca agatcgatga gcaaaccaaa 480
accaccgaga gtttcaaaga cctccgcgac accgttttcg aattccccgg atggaagtct 540
cctctgaagg ctacacacat ggtccaactg gtgctcgacc ggaacgaccc tgcttattcg 600
gacatgatct atttctgctc acatcttccc aaatccaacg gaatcatcgt caacacgttc 660
gaagagctcg agccacctag cgtcctccag gccattgctg gaggcctgtg tgttcctgat 720
gggccaactc cgcccgtgta ctacgttggt ccattgattg aggaagagaa agaattgagt 780
aaggatgcag atgcggccga gaaggaggac tgcttgtcat ggctcgataa gcagccaagt 840
cgaagcgtgc tgtttctctg tttcggaagc atgggatcat ttccggctgc tcaactgaag 900
gagatagcga acgggttgga ggcgagcggg cagaggttcc tgtgggtggt gaagaagccg 960
ccggttgaag agaaatcaaa gcaggtccat ggagttgacg actttgattt gaagggtgtg 1020
ttgccagaag ggtttttgga gaggacggcc gacaggggga tggtagtgaa gtcatgggcg 1080
ccgcaggtgg tggtgttgaa gaaggagtcg gttggtgggt tcgtgacaca ttgcggatgg 1140

```
aactcggtac tggaagcagt ggttgcgggg gtgccgatga ttgcttggcc gctttacgcg   1200
gagcagcata tgaacaggaa tgttctagtg acggacatgg aactcgcgat cggggtggag   1260
cagagagacg aggaaggtgg gttcgtgagc ggggaagaag tggagaggag agtgagggag   1320
ttgatggagt cggaaggagg aagagtgctt agagagaggt gcaagaaact tggggagatg   1380
gcttcggctg ctttgggaga gaccggttcg tccaccagaa acttggtcaa ctttgttagt   1440
agcattacat aa                                                       1452
```

<210> 12
<211> 1719
<212> ADN
<213> *Malus domestica*

<400> 12

gtcgcagctg ccgccatggg agacgtcatt gtactgtacg catctccggg gatggggcac 60
atcgtctcca tggtggagct gggcaagttc attgtccacc gctacggccc ccacaaattc 120
tccatcacca ttctctacac ctgcggcagc attgtcgaca ccgctagcat ccccgtctac 180
atccgccgca tctcccactc ccacccttc atttccttcc gccaattccc tcgcgtcacc 240
aataatatta cccgaaacat aagcgtcccc gcaatcacgt tcgacttcat ccgccagaac 300
gatcctcatg tccgcagtgc cctccaagaa atctctaaat ccgccaccgt tcgcgccttc 360
atcatcgacc tcttctgcac ctccgctctt cccataggga aggaattcaa catcccaaca 420
tactacttct gcacttctgg agccgcaatt cttgctgctt ttttgtattt gcccaagatc 480
gatgagcaaa ccaaaaccac cgagagtttc aaagacctcc gcgacaccgt tttcgaattc 540
cccggatgga agtctcctct gaaggctcca cacatggtcc aactggtgct cgaccggaac 600
gaccctgctt attcggacat gatctatttc tgctcacatc ttcccaaatc caacggaatc 660
atcgtcaaca cgttcgaaga gctcgagcca cctgccgtcc tccaggccat tgctggaggc 720
ctgtgtgttc ctgatgggcc aactccgccc gtgtactacg ttggtccatt gattgaggaa 780
gagaaagaat tgagtaagga tgcagatgcg gccggggagg aggactgctt gtcatggctc 840
gataagcagc caagtcgaag cgtgctgttt ctctgtttcg gaagcatggg atcatttccg 900
gctgctcaac tgaaggagat agcgaacggg ttggaggcga gcgggcagag gttcctgtgg 960
gtggtgaaga agccgccggt tgaagagaaa tcaaagcagg tccatggagt tgacgacttt 1020
gatttgaagg gtgtgttgcc agaagggttt ttggagagga cggccgacag ggggatggta 1080
gtgaagtcat gggcgccgca ggtggtggtg ttgaagaagg agtcggttgg tgggttcgtg 1140
acacattgcg gatggaactc ggtactggaa gcagtggttg cgggggtgcc gatgattgct 1200
tggccgcttt acgcggagca gcatatgaac aggaatgttc tagtgacgga catggaaatc 1260
gcgatcgggg tggagcagag agacgaggaa ggtgggttcg tgagcgggga agaagtggag 1320
aggagagtga gggagttgat ggagtcggaa ggaggaagag cgcttagaga gaggtgcaag 1380
aaacttgggg agatggcttc ggctgctttg ggagagaccg gttcgtccac cagaaacttg 1440
gtcaactttg ttagtagcat tacataaccg tgcaagtttt gggttcgcat ttcgcatttg 1500

gatttcgaga ggagtcgtgt gcaagaataa accctcttaa ccaactccgc gatatggttt 1560
tctgttactt agtttggcac ctcgtgggat tttggtgctc tcttttgaga gttttaagaa 1620
agcatcaaac tattcggatg taatgttgat gtaacgtcga ccatcagctt aattccaacc 1680
aataagatta tctacttctt tcttaccaaa aaaaaaaaa 1719

<210> 13
<211> 1452
<212> ADN
<213> *Malus domestica*

<400> 13

```
atgggagacg tcattgtact gtacgcatct ccggggatgg ggcacatcgt ctccatggtg     60
gagctgggca agttcattgt ccaccgctac ggcccccaca aattctccat caccattctc    120
tacacctgcg gcagcattgt cgacaccgct agcatccccg tctacatccg ccgcatctcc    180
cactcccacc ctttcatttc cttccgccaa ttccctcgcg tcaccaataa tattacccga    240
aacataagcg tccccgcaat cacgttcgac ttcatccgcc agaacgatcc tcatgtccgc    300
agtgccctcc aagaaatctc taaatccgcc accgttcgcg ccttcatcat cgacctcttc    360
tgcacctccg ctcttcccat agggaaggaa ttcaacatcc caacatacta cttctgcact    420
tctggagccg caattcttgc tgctttttg tatttgccca agatcgatga gcaaaccaaa    480
accaccgaga gtttcaaaga cctccgcgac accgttttcg aattccccgg atggaagtct    540
cctctgaagg ctccacacat ggtccaactg gtgctcgacc ggaacgaccc tgcttattcg    600
gacatgatct atttctgctc acatcttccc aaatccaacg gaatcatcgt caacacgttc    660
gaagagctcg agccacctgc cgtcctccag gccattgctg gaggcctgtg tgttcctgat    720
gggccaactc cgcccgtgta ctacgttggt ccattgattg aggaagagaa agaattgagt    780
aaggatgcag atgcggccgg ggaggaggac tgcttgtcat ggctcgataa gcagccaagt    840
cgaagcgtgc tgtttctctg tttcggaagc atgggatcat ttccggctgc tcaactgaag    900
gagatagcga acgggttgga ggcgagcggg cagaggttcc tgtgggtggt gaagaagccg    960
ccggttgaag agaaatcaaa gcaggtccat ggagttgacg actttgattt gaagggtgtg   1020
ttgccagaag ggtttttgga gaggacggcc gacaggggga tggtagtgaa gtcatgggcg   1080
ccgcaggtgg tggtgttgaa gaaggagtcg gttggtgggt tcgtgacaca ttgcggatgg   1140
aactcggtac tggaagcagt ggttgcgggg gtgccgatga ttgcttggcc gctttacgcg   1200
gagcagcata tgaacaggaa tgttctagtg acggacatgg aaatcgcgat cggggtggag   1260
cagagagacg aggaaggtgg gttcgtgagc ggggaagaag tggagaggag agtgagggag   1320
ttgatggagt cggaaggagg aagagcgctt agagagaggt gcaagaaact tggggagatg   1380
gcttcggctg ctttgggaga gaccggttcg tccaccagaa acttggtcaa ctttgttagt   1440
agcattacat aa                                                       1452
```

<210> 14
<211> 1659
<212> ADN
<213> *Malus domestica*

<400> 14

```
ctcaaatcaa atcagataat ttcagcagaa aacacaacac acaaatcagc agcagcagct     60
gccatgggag acgtaatagt gctgtacgca gctccaggaa tggggcatgt catctccatg    120
gtggagctgg gcaagctcat cctccaccgc tacggccccc acaagttttc catcaccatt    180
ctctacacct gcggcagctt cttcgacacc cctagcatcc ccgcctacat ccgccgcatc    240
tcccactccc acccttccat ttccttccgc caattccctc gcgtcaccaa taaaattacc    300
caaaacatca gcggcaccgc aatcgtggtt gacttcgttc gccagaacga tccccacgtc    360
cgccgtgccc tccaagacat ctccaaatcc gccgtcgtcc gcgccttcat catcgacctc    420
gtctgcacct ccgcgatgac catcagcaag gaattcgaca ttcccacata ctacttctac    480
acttctggtg ccgcagctct gggtgctttt ttgtatttcc ctaagatcca tgaacaaacc    540
acccagagtt tcaaggacct caccgacacc gttatcgaat tccccggacg gaaatctcct    600
ctgaaggcta tacacatgat cgaaccgctg ctcgaccgag acgaccctgc ttattgggac    660
ttcctctcct tttgctccga tcttcccaaa tccaaaggaa tcatcgtcaa cacgttcgaa    720
gagctcgagc cgccagccgt cctccatgcc attgctgaag gcctgtgtgt tcctgatggg    780
ccaacttcgc ctgtgtacta cgttggacca ttgattgacg aagaaaaagt atcgggtaat    840
gatgcagctg cggccgagga ggactgcttg tcatggctcg ataagcagcc aagtcgaagc    900
gtggtttttc tctgtttcgg aagcagggga tcactccctg caattcaact gaaggagata    960
gcgaaagcgt tggaggcgag cgggcagagg ttcctgtggg tggtgaagaa gccgccggtt   1020
gatgagaaaa caaagcaggt ccttggagtt gacgactttg atttggaggg tgtgttgcca   1080
gaagggttct tggagaggac caaagacagg gggatggtag tgaagtcatg ggcaccgcag   1140
gcagaggtgt tgaagaagga atcggttggt gggttcgtga cacattgcgg atggaactcg   1200
gtgctggaag cagtggttgc gggggtgccc atgattgctt ggccgttgta cgcggagcag   1260
catttgaaca ggaatgttat ggcgacggac atggaaatag cgattgcggt tgagcagaga   1320
gatgaggaag atgggttcgt gagtggggag gaattggaga ggagagtgag ggagttgatg   1380
gagtcggaag aagggagagt gcttagagag aggagcaaga aaattgggga gatggctatg   1440
gctgctttgg gagagaatgg ttcgtccacc agaaacttgg ttaactttgt taatagctta   1500
acataattat gcaagttctg gattttcatt tgcatttgga ttttggtgaa aaataaaata   1560
agaaaaattg tatctgtttg atgacaattg tattgtcttc ccatcaccat tttttttttt   1620
tataataaac tcctcgttct ttatattaaa aaaaaaaaa                          1659
```

<210> 15
<211> 1443
<212> ADN
<213> *Malus domestica*

<400> 15

```
atgggagacg taatagtgct gtacgcagct ccaggaatgg ggcatgtcat ctccatggtg     60
gagctgggca agctcatcct ccaccgctac ggcccccaca agttttccat caccattctc    120
tacacctgcg gcagcttctt cgacacccct agcatccccg cctacatccg ccgcatctcc    180
cactcccacc cttccatttc cttccgccaa ttccctcgcg tcaccaataa aattacccaa    240
aacatcagcg gcaccgcaat cgtggttgac ttcgttcgcc agaacgatcc ccacgtccgc    300
cgtgccctcc aagacatctc caaatccgcc gtcgtccgcg ccttcatcat cgacctcgtc    360
tgcacctccg cgatgaccat cagcaaggaa ttcgacattc ccacatacta cttctacact    420
tctggtgccg cagctctggg tgctttttg tatttcccta agatccatga acaaaccacc     480
cagagtttca aggacctcac cgacaccgtt atcgaattcc ccggacggaa atctcctctg    540
aaggctatac acatgatcga accgctgctc gaccgagacg accctgctta ttgggacttc    600
ctctcctttt gctccgatct tcccaaatcc aaaggaatca tcgtcaacac gttcgaagag    660
ctcgagccgc cagccgtcct ccatgccatt gctgaaggcc tgtgtgttcc tgatgggcca    720
acttcgcctg tgtactacgt tggaccattg attgacgaag aaaaagtatc gggtaatgat    780
gcagctgcgg ccgaggagga ctgcttgtca tggctcgata agcagccaag tcgaagcgtg    840
gtttttctct gtttcggaag caggggatca ctccctgcaa ttcaactgaa ggagatagcg    900
aaagcgttgg aggcgagcgg gcagaggttc ctgtgggtgg tgaagaagcc gccggttgat    960
gagaaaacaa agcaggtcct tggagttgac gactttgatt tggagggtgt gttgccagaa   1020
gggttcttgg agaggaccaa agacaggggg atggtagtga agtcatgggc accgcaggca   1080
gaggtgttga agaaggaatc ggttggtggg ttcgtgacac attgcggatg gaactcggtg   1140
ctggaagcag tggttgcggg ggtgcccatg attgcttggc cgttgtacgc ggagcagcat   1200
ttgaacagga atgttatggc gacggacatg gaaatagcga ttgcggttga gcagagagat   1260
gaggaagatg ggttcgtgag tggggaggaa ttggagagga gagtgaggga gttgatggag   1320
tcggaagaag ggagagtgct tagagagagg agcaagaaaa ttggggagat ggctatggct   1380
gctttgggag agaatggttc gtccaccaga aacttggtta actttgttaa tagcttaaca   1440
taa                                                                 1443
```

<210> 16
<211> 39
<212> PRT
<213> Artificial

<220>
<223> motivo consenso

<400> 16

```
Val Leu Lys Lys Glu Ser Val Gly Gly Phe Val Thr His Cys Gly Trp
1               5                   10                  15

Asn Ser Val Leu Glu Ala Val Val Ala Gly Val Pro Met Ile Ala Trp
            20                  25                  30
```

Pro Leu Tyr Ala Glu Gln His 35

<210> 17

<211> 30
<212> ADN
<213> Artificial

<220>
<223> Cebador

<400> 17
gactagttct agatcgcgag_cggccgccct 30

<210> 18
<211> 25
<212> ADN
<213> Artificial

<220>
<223> Cebador

<400> 18
gaagggtgtg ttgccagaag ggtct 25

<210> 19
<211> 20
<212> ADN
<213> Artificial

<220>
<223> Cebador

<400> 19
gtcacgaacc caccaaccga 20

<210> 20
<211> 30
<212> ADN
<213> Artificial

<220>
<223> Cebador

<400> 20
acgggatcca tgggagacgt cattgtactg 30

<210> 21
<211> 36
<212> ADN
<213> Artificial

<220>
<223> Cebador

<400> 21
cccaagcttt tatgtaatgc tactaacaaa gttgac 36

REIVINDICACIONES

1. Un método de producción de una célula vegetal o una planta con al menos uno de entre:

a) mayor contenido de floricina y
b) mayor actividad de la floretina glicosiltransferasa;
método que comprende la transformación de una célula vegetal o de una planta con un polinucleótido que codifica un polipéptido con la secuencia de aminoácidos de SEC ID Nº 1 o un polipéptido con al menos el 85 % de identidad de secuencia con la secuencia de aminoácidos de SEC ID Nº 1 y actividad de la floretina glicosiltransferasa.

2. Un método de la reivindicación 1, en el que el polinucleótido comprende la secuencia de nucleótidos de una cualquiera de SEC ID Nº 6 y 7, o una secuencia con al menos el 70 % de identidad con una cualquiera de SEC ID Nº 6 y 7.

3. El método de la reivindicación 1 o 2, en el que la célula vegetal o la planta también se transforma con un polinucleótido que codifica una floretina sintasa.

4. Una célula hospedadora, una célula vegetal, una planta o una parte de planta producida mediante el método de una cualquiera de las reivindicaciones 1 a 3, en la que la célula hospedadora, la célula vegetal o la planta está modificada genéticamente para expresar el polinucleótido definido en una cualquiera de las reivindicaciones a 1 a 3.

5. Un método de biosíntesis o de producción de floricina que comprende:

a) cultivar una célula hospedadora, una célula vegetal o una planta de la reivindicación 4, capaz de expresar una floretina glicosiltransferasa, en presencia de floretina que se puede suministrar a o puede estar presente de manera natural en la célula hospedadora, la célula vegetal o la planta;
b) extraer floricina de una célula vegetal de o de una planta de la reivindicación 4 o
c) poner en contacto floretina con el producto de expresión de una construcción de expresión que comprende un polinucleótido o un polipéptido según lo definido en la reivindicación 1.

6. Un método de selección de una planta modificada en cuanto a la actividad de la floretina glicosiltransferasa o el contenido de floricina, método que comprende ensayar en una planta la expresión modificada de un polinucleótido definido en la reivindicación 1 o un polipéptido definido en la reivindicación 1.

FIGURA 1

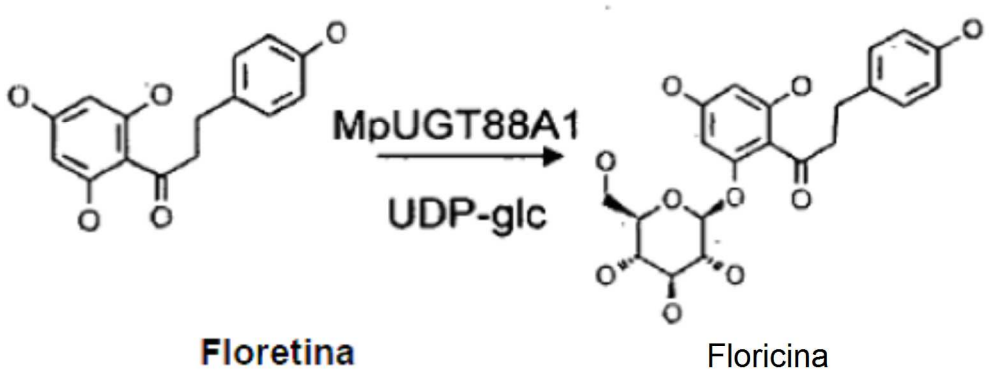


FIGURA 2

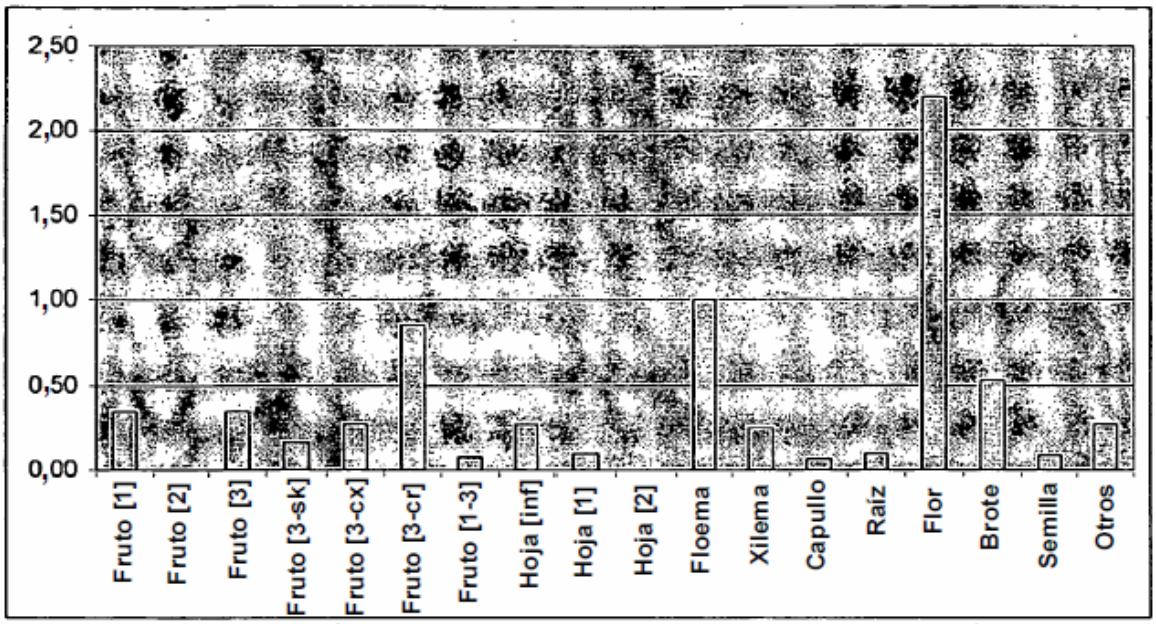

2,50
2,00
1,50
1,00
0,50
0,00
Fruto [1]
Fruto [2]
Fruto [3]
Fruto [3-sk]
Fruto [3-cx]
Fruto [3-cr]
Fruto [1-3]
Hoja [inf]
Hoja [1]
Hoja [2]
Floema
Xilema
Capullo
Raíz
Flor
Brote
Semilla
Otros

FIGURA 3

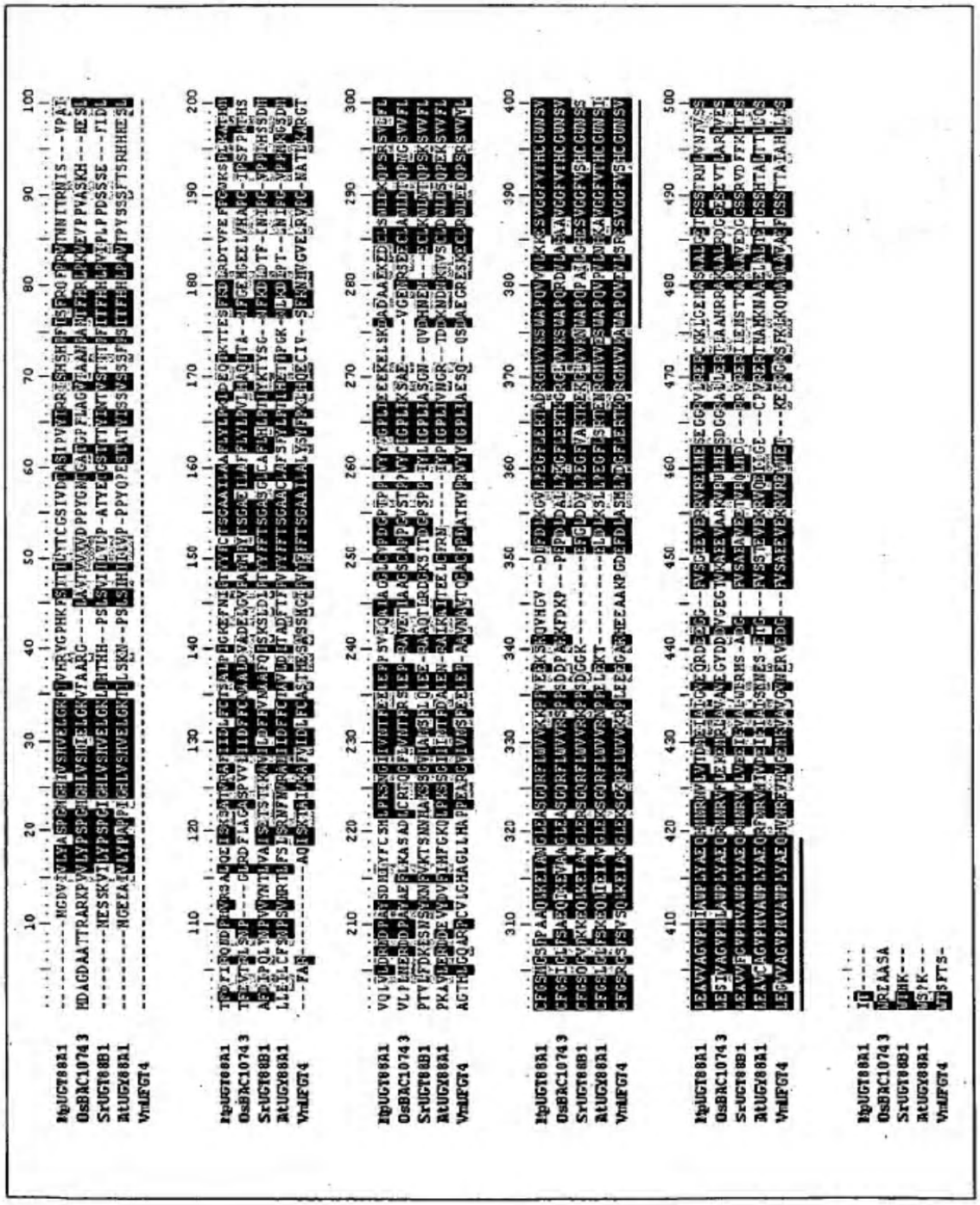


FIGURA 4

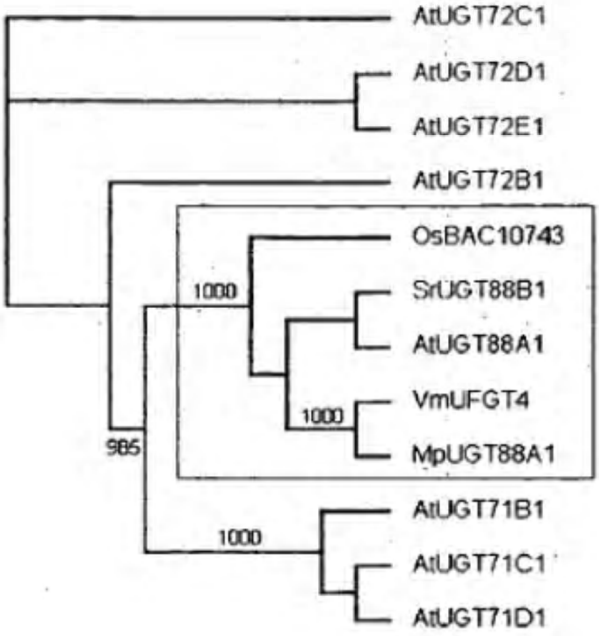

AtUGT72C1
AtUGT72D1
AtUGT72E1
AtUGT72B1
OsBAC10743
SrUGT88B1
AtUGT88A1
VmUFGT4
MpUGT88A1
AtUGT71B1
AtUGT71C1
AtUGT71D1
1000
1000
985
1000

FIGURA 5

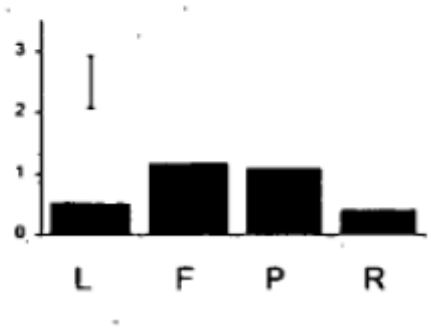


FIGURA 6

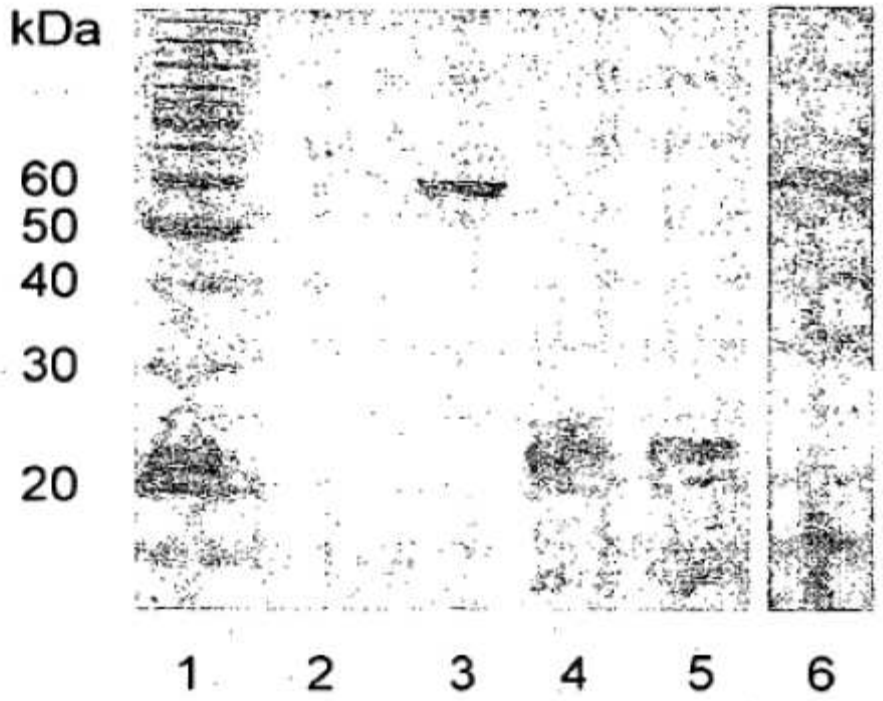


FIGURA 7

Fig. 7

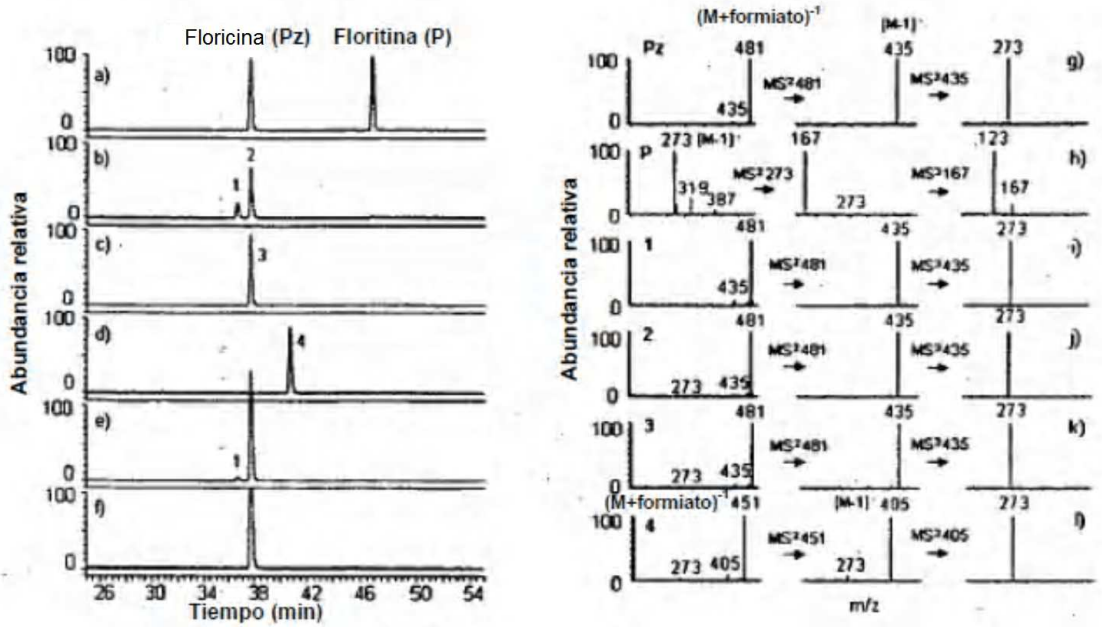


FIGURA 8

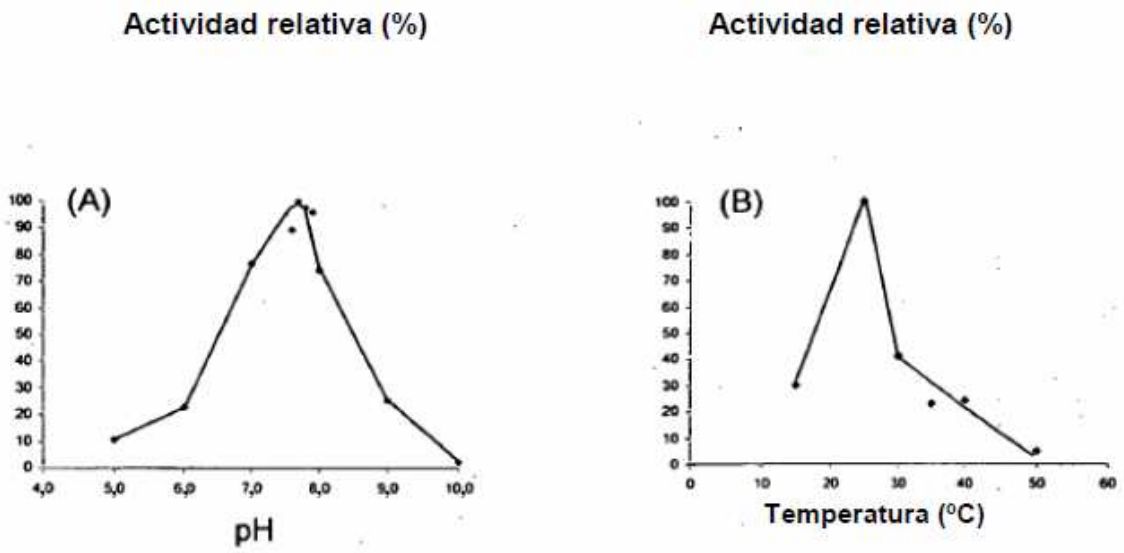
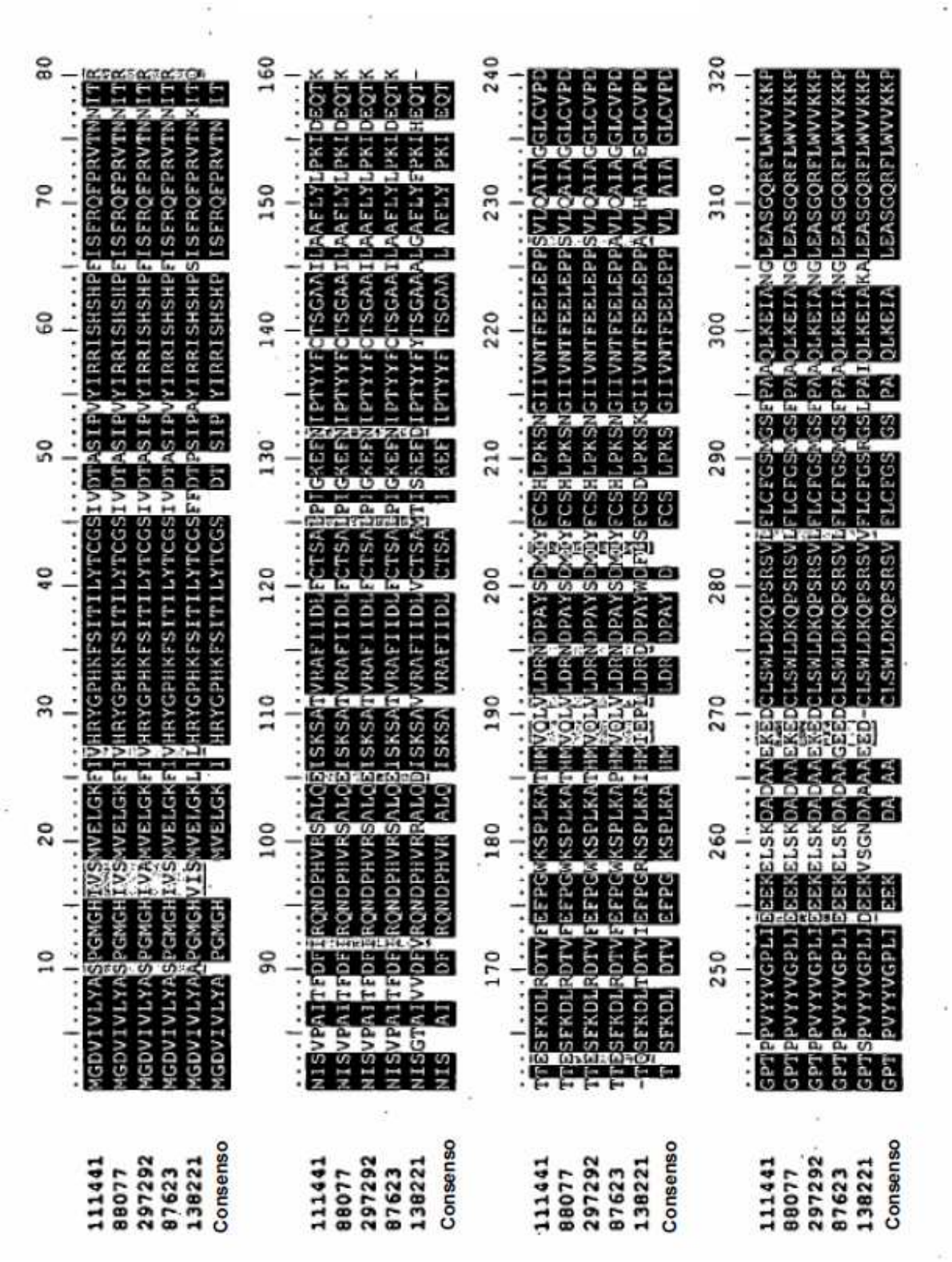

Actividad relativa (%)
(A)
100
90
80
70
60
50
40
30
20
10
0
4,0
5,0
6,0
7,0
8,0
9,0
10,0
pH
Actividad relativa (%)
(B)
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
Temperatura (ºC)

FIGURA 9



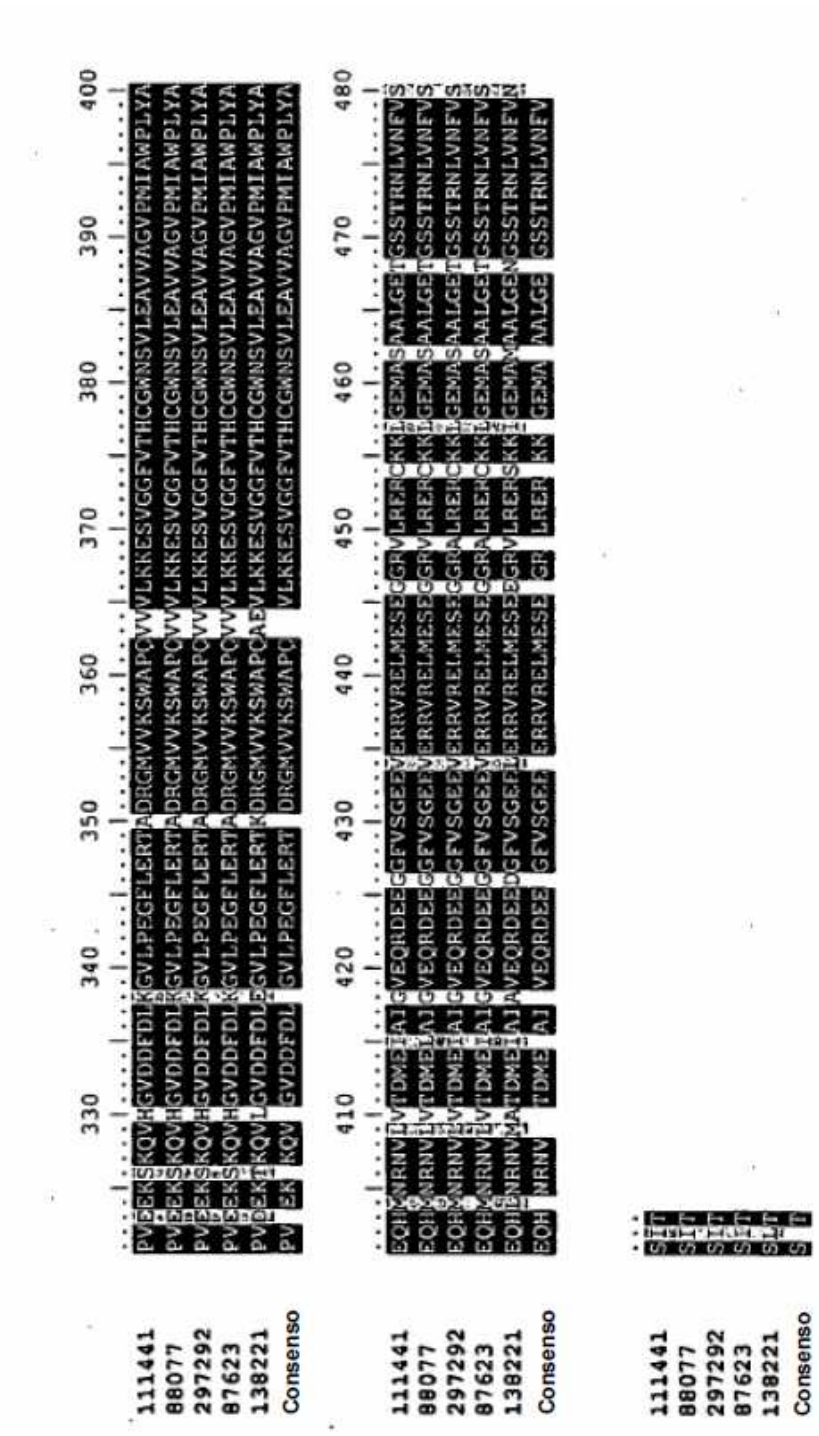

111441
88077
297292
87623
138221
Consenso
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480

FIGURA 10

Polipéptido	Md_88077	Ms_297292	Md_87623	Md_138221
Md_111441	99,79	99,59	98,96	82,4
Md_88077		99,38	98,76	82,19
Ms_297292			98,96	81,99
Md_87623				82,4