

VECTORES Y MÉTODOS PARA MEJORAR LA FRECUENCIA DE TRANSFORMACIÓN DE PLANTAS DICOTILEDÓNEAS

REFERENCIA CRUZADA A SOLICITUD RELACIONADA

La presente solicitud reivindica la prioridad de la solicitud provisional de EE. UU. n.º 63/271 654, presentada el 25 de octubre de 2021, cuyo contenido se incorpora al presente documento mediante referencia.

ANTECEDENTES

Los genes morfogénicos BabyBoom (BBM) y Wuschel (WSU), juntos o individualmente, se han utilizado para aumentar la eficiencia de la transformación de monocotiledóneas. El uso de estos genes para la transformación de dicotiledóneas fue difícil debido a un fenotipo anormal y un problema de regeneración, especialmente BBM no mostró éxito. En consecuencia, se necesitan métodos para aumentar la eficiencia de la transformación de dicotiledóneas mediante el uso de BBM con un fenotipo vegetal mínimo o nulo. También se necesitan métodos para aumentar la eficiencia de la transformación de gimnospermas mediante el uso de BBM con un fenotipo vegetal mínimo o nulo.

COMPENDIO

En determinadas modalidades de la invención se proporciona una célula vegetal transformada que comprende una construcción de ácido nucleico heteróloga que comprende un promotor y un factor de transcripción relacionado con la embriogénesis y/u organogénesis (tal como un factor de transcripción de genes morfogénicos o un factor de transcripción de genes de desarrollo), donde el promotor es un promotor específico de embriones de plantas o específico de semillas.

En determinadas modalidades de la invención se proporciona una construcción de ácido nucleico que comprende un promotor específico de embriones de plantas y un factor de transcripción relacionado con la embriogénesis y/u organogénesis, donde el factor de transcripción es BabyBoom (BBM) o Wuschel (WSU).

En determinadas modalidades de la invención se proporciona una construcción de ácido nucleico que comprende pGmLe1AtWUS o pGmLe1AtBBM.

En determinadas modalidades de la invención se proporciona un método de transformación de tejidos vegetales, que comprende poner en contacto a los tejidos vegetales con una construcción de ácido nucleico heteróloga que comprende un promotor y un factor de transcripción relacionado con la embriogénesis y/u organogénesis (tal como un factor de transcripción de genes morfogénicos o un factor de transcripción de genes de desarrollo), donde el promotor es un promotor específico de embriones de plantas o específico de semillas.

En determinadas modalidades de la invención se proporciona un método de aumento de la producción de semillas en una planta en comparación con una planta no transformada, que comprende poner en contacto a los tejidos vegetales con una construcción de ácido nucleico heteróloga que comprende un promotor y un factor de transcripción relacionado con un factor de embriogénesis y organogénesis (tal como un factor de transcripción de genes morfogénicos o un factor de transcripción de genes de desarrollo), donde el promotor es un promotor específico de embriones de plantas o específico de semillas.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Figuras 1A-1B. Diagrama esquemático de dos vectores de transformación utilizados para la transformación de *Theobroma cacao* L. **(a)** pDDNPTYFP-1 es un fragmento de ADN-T de 3637 pb que contiene *neomicina fosfotransferasa II (nptII)* y una fusión de traducción de proteína fluorescente amarilla mejorada (*eyfp*, por su sigla en inglés) impulsada por un promotor CaMV35S. **(b)** pDDNPTYFP-2 es un fragmento de ADN-T de 4416 pb que contiene dos casetes de genes donde la *eyfp* es impulsada por un promotor de Nos/potenciador de TMV Ω y la *nptII* es impulsada por un promotor CaMV35S mejorado.

Figura 2. Formación de embriones somáticos a partir de materiales florales para 9 cultivares diferentes de *Theobroma cacao* L. Se utilizaron explantes de pétalos y estaminodios para inducir embriones somáticos primarios. Cada histograma representa el nivel medio ($\pm$ error estándar).

Figuras 3A-3B. Embriogénesis somática secundaria y conversión de embriones en plántulas para cultivares seleccionados de *Theobroma cacao* L. **(a)** Se observaron explantes de cotiledón en cuatro categorías según la cantidad de embriones que produjeron: ningún embrión producido, de 1 a 4 embriones producidos, de 5 a 9 embriones producidos y más de 10 embriones producidos. **(b)** Los embriones maduros se expusieron a niveles crecientes de

luz mientras estaban en un medio de desarrollo embrionario que contenía sacarosa al 6 % y luego en un medio ligero de desarrollo embrionario y se contaron como convertidos si producían un brote de 1 cm con meristemo y una raíz de 1 cm. Cada histograma representa el nivel medio ($\pm$ error estándar).

Figura 4. Efecto de la concentración de sal basal del medio de plantas leñosas (WPM, por su sigla en inglés) duplicada y WPM con ajuste de $5,0 \mu$ de CuSO_4 en medio de generación de callo secundario (SCG, por su sigla en inglés) sobre la embriogénesis somática secundaria en *Theobroma cacao* L. cv. INIAPG-038. Se observaron explantes de cotiledón en cuatro categorías según la cantidad de embriones que produjeron: ningún embrión producido, de 1 a 4 embriones producidos, de 5 a 9 embriones producidos y más de 10 embriones producidos. Cada histograma representa el nivel medio ($\pm$ error estándar).

Figura 5. Expresión transitoria de proteína fluorescente amarilla (YFP, por su sigla en inglés) para cultivares seleccionados de *Theobroma cacao* L. Se observaron explantes de cotiledón en tres categorías según el alcance de expresión de proteína fluorescente (FTE, por su sigla en inglés): de 0 a 5 % de alcance de FPE, de 6 a 20 % de alcance de FPE, >20 % de alcance de FPE. Cada histograma representa el nivel medio ($\pm$ error estándar).

Figura 6. Efecto de la cepa de *Agrobacterium* sobre la expresión transitoria de la proteína fluorescente amarilla (YFP) en *Theobroma cacao* L. cv. INIAPG-038. Se observaron explantes de cotiledón en tres categorías según el alcance de expresión de proteína fluorescente (FTE, por su sigla en inglés): de 0 a 5 % de alcance de FPE, de 6 a 20 % de alcance de FPE, >20 % de alcance de FPE. Cada histograma representa el nivel medio ($\pm$ error estándar).

Figura 7. Efecto del tamaño de embriones somáticos sobre la expresión transitoria de la proteína fluorescente amarilla (YFP) en *Theobroma cacao* L. cv. INIAPG-038. Para los experimentos se utilizó tejido de cotiledón derivado de embriones somáticos. Se observaron explantes de cotiledón en tres categorías según el alcance de expresión de proteína fluorescente (FTP): de 0 a 5 % de alcance de FPE, de 6 a 20 % de alcance de FPE, >20 % de alcance de FPE. Cada histograma representa el nivel medio ($\pm$ error estándar).

Figuras 8A-8G. Expresión transitoria de proteína fluorescente amarilla (YFP) y desarrollo de eventos transgénicos estables en *Theobroma cacao* L. cv. INIAPG-038.

(a) Cotiledón con expresión de YFP transitoria. (b) Embrión globular con expresión de YFP. (c) Embrión maduro con expresión de YFP. (d) Embriones somáticos cuaternarios con expresión de YFP. (e) Plántula con expresión de YFP. (f) Una planta no transgénica (la más a la izquierda) y 3 plantas transgénicas INIAPG-038 en el invernadero. (g) Una plántula transgénica INIAPG-038 injertada en un pélet de turba Jiffy; se injertó un vástago transgénico INIAPG-038 con raíces pobres en un portainjerto no transgénico.

Figuras 9A-9D. Análisis por PCR del ADN genómico de tres eventos transgénicos diferentes de *Theobroma cacao* L. cv. INIAPG-038. El ADN genómico se extrajo de tejidos foliares de un control no transgénico y de tres plantas de cada evento transgénico independiente. M escalera de ADN GeneRuler de 1 kb, C Control no transgénico INIAPG-038, Carriles 1-3 evento transgénico INIAPG-038 #1-1, #1-2 y #1-3, Carriles 4-6 evento transgénico INIAPG-038 #2-1, #2-2 y #2-3, Carriles 7-9 evento transgénico INIAPG-038 #3-1, #3-2 y #3-3, P pDDNPTYFP-2 como control positivo. (a) Productos de fragmentos de 595 pb amplificados mediante el uso de cebadores específicos de genes de *proteína fluorescente amarilla mejorada (eyfp)*, EYFP 3F y EYFP 4R. (b) Productos de fragmentos de 761 pb amplificados mediante el uso de cebadores específicos de genes de *neomicina fosfotransferasa II (nptII)*, NPTII 3F y NPTII 4R (c) Productos de fragmentos de 562 pb amplificados mediante el uso de cebadores específicos de replicones de pVS1, PVS1 1F y PVS1 2R. (d) Productos de fragmentos de 672 pb amplificados mediante el uso de cebadores específicos de genes de *neomicina fosfotransferasa I (npt1)*, NPT1 1F y NPT1 2R.

Figuras 10A-10E. Diagrama esquemático de cinco vectores de transformación utilizados para la transformación del cacao. (A) vector de control, (B) pGmLe1AtBBM, (C) pGmLe1AtWUS, (D) pGmLe1VvGRF/GIF y (E) pZmUbi1VvGRF/GIF.

Figura 11. Generación de embriones somáticos de cacao transgénico mediante el uso de pGmLe1AtBBM.

Figura 12. Brotes putativos de tabaco transgénico 30 días después de la infección por *Agrobacterium*.

Figura 13. Brotes putativos de *Nicotiana benthamiana* transgénica 30 días después de la infección por *Agrobacterium*.

Figura 14. Brotes putativos de tabaco transgénico 66 días después de la infección por

Agrobacterium.

Figura 15. Brotes putativos de *Nicotiana benthamiana* transgénica 66 días después de la infección por *Agrobacterium*.

Figura 16. Las plantas de tabaco transgénicas transformadas con pGmLe1AtBBM y pGmLe1AtWUS muestran un fenotipo normal. Las fotos fueron tomadas 66 días después de la infección por *Agrobacterium*.

Figuras 17A-17B. Expresión de la proteína fluorescente amarilla (YFP) en (a) discos foliares de eventos transgénicos de *Theobroma cacao* L. cv. INIAPG-038 y (b) cepas de *Agrobacterium* que contienen vectores que incluyen el casete del gen de la proteína fluorescente amarilla mejorada utilizado para la transformación de INIAPG-038. Arriba, Brightfield; abajo, filtro de YFP. (a) Discos foliares en orden desde la columna más a la izquierda hasta la más a la derecha: INIAPG-038 no transgénico, evento #1, evento #2, evento #3. La muestra de cada fila provino de una hoja o planta diferente. (b) Cepas de *Agrobacterium* en orden de izquierda a derecha: AGL1 sin un vector binario de ADN-T, AGL1 con pDDNPTYFP-1, AGL1 con pDDNPTYFP-2.

Figura 18. Construcciones pGmLe1pAtBBM y pGmLe1pAtWUS.

Figuras 19A-19C. Embriones somáticos de cacao en germinación transformados con GmLe1AtBBM de fenotipo normal. Fig. 19A Matina 1-6. Fig. 19B INIAPG-038. Fig. 19C. CCN-51.

Figura 20. Plántulas Matina 1-6 transformadas con pGmLe1AtBBM que tienen un fenotipo vegetal normal.

Figura 21. Formación de callos de cacao no regenerables transformados con pGmLe1AtWUS.

Figura 22. Formación de múltiples embriones somáticos a partir de un único embrión INIAPG-038 transformado con pGmLe1AtBBM.

Figura 23. Embriones somáticos de Matina 1-6 transformados con pGmLe1AtBBM de fenotipo anormal.

Figura 24. Plantas de tabaco transgénicas transformadas con control, pGmLe1AtBBM y pGmLe1AtWUS de fenotipo normal.

Figura 25. Formación de tejido de callo verde M82 no regenerable transformado con pGmLe1AtWUS.

Figura 26. Plantas de tomate M82 transgénicas transformadas con control, pGmLe1AtBBM y pGmLe1AtWUS de fenotipo normal. La planta de control se transfirió al suelo en una fecha posterior que las otras plantas.

Figuras 27A-27D. Plantas de yuca transgénicas transformadas con control, pGmLe1AtBBM, pGmLe1AtWUS y pZmUbi1VvGRF/GIF de fenotipo normal. Fig. 27A. Control. Fig. 27B. pGmLe1AtBBM. Fig. 27C. pGmLe1AtWUS. Fig. 27D. pZmUbiVvGRF/GIF.

DESCRIPCIÓN DETALLADA

La presente descripción está relacionada con el campo de la transformación de plantas. Se proporcionan métodos para aumentar la eficiencia de la transformación de dicotiledóneas mediante el uso de un factor de transcripción relacionado con genes de embriogénesis y/u organogénesis con un fenotipo vegetal mínimo o nulo. También se proporcionan diseños vectoriales para lograr estos resultados, así como métodos para su uso.

Un protocolo de transformación altamente eficiente es un requisito previo para desarrollar cultivos modificados genéticamente y editados con genoma. La presente invención es útil como herramienta de transformación y edición del genoma, especialmente para cultivos y cultivares de dicotiledóneas recalcitrantes.

En determinadas modalidades de la invención se proporciona una célula vegetal transformada que comprende una construcción de ácido nucleico heteróloga que comprende un promotor y un factor de transcripción relacionado con la embriogénesis y/u organogénesis, donde el promotor es un promotor específico de embriones de plantas o específico de semillas.

En determinadas modalidades, la célula vegetal es una célula dicotiledónea.

En determinadas modalidades, la célula dicotiledónea es una célula de cacao, yuca, soja, tomate, tabaco, café, girasol, alfalfa, pimienta, caupí, girasol, uva, álamo, café, girasol, nuez, almendra, melocotón, nectarina, ciruela, albaricoque, manzana, pera, caqui u otro árbol

frutal.

En determinadas modalidades, la célula vegetal es una célula de gimnosperma.

5 En determinadas modalidades, la célula de gimnosperma se selecciona de entre células de pino, ginkgo, abeto o abeto de Douglas.

En determinadas modalidades, el promotor es un promotor inducible.

10 En determinadas modalidades, el promotor es un promotor específico de embriones de plantas, y el promotor específico de embriones de plantas es un promotor específico de embriones de dicotiledóneas.

En determinadas modalidades, el promotor específico de embriones de dicotiledóneas es un promotor específico de embriones de soja.

15 En determinadas modalidades, el promotor específico de embriones de soja es un promotor de lectina de soja.

En determinadas modalidades, el promotor específico de embriones de soja es un promotor de glicinina de soja.

20 En determinadas modalidades, el promotor específico de embriones de soja es un promotor de conglicinina de soja.

25 En determinadas modalidades, el promotor es un promotor específico de semillas de plantas, y el promotor específico de semillas de plantas es un promotor específico de semillas de dicotiledóneas.

En determinadas modalidades, el promotor específico de semillas de dicotiledóneas es un promotor específico de semillas de soja.

30 En determinadas modalidades, el factor de transcripción relacionado con la embriogénesis y/u organogénesis es un factor de transcripción de genes morfogénicos o un factor de transcripción de genes de desarrollo.

En determinadas modalidades, el factor de transcripción es BabyBoom (BBM).

35

En determinadas modalidades, la construcción comprende además un vector.

En determinadas modalidades, el vector es un plásmido.

5 En determinadas modalidades, la construcción comprende además una recombinasa con especificidad de sitio.

10 En determinadas modalidades, la construcción comprende además un gen de rasgo. Tal como se usa en el presente documento, un «gen de rasgo» puede ser un producto génico que confiera mejora nutricional, producción de proteínas farmacéuticas, mayor rendimiento de semillas o biomasa, tolerancia al estrés abiótico, tolerancia a la sequía, tolerancia al frío, tolerancia a herbicidas, resistencia a plagas, resistencia a insectos, resistencia a enfermedades, eficiencia de uso de nitrógeno (NUE, por su sigla en inglés) o capacidad de alterar una vía metabólica.

15 En determinadas modalidades, la construcción comprende además un gen marcador visual/seleccionable.

En determinadas modalidades, la construcción comprende además un casete de CRISPR.

20 En determinadas modalidades de la invención se proporciona una construcción de ácido nucleico que comprende un promotor específico de embriones de plantas y un factor de transcripción, donde el factor de transcripción es BabyBoom (BBM) o Wuschel (WSU).

En determinadas modalidades, el factor de transcripción es BabyBoom (BBM).

25 En determinadas modalidades, el factor de transcripción es Wuschel (WSU).

En determinadas modalidades de la invención se proporciona una construcción de ácido nucleico que comprende pGmLe1AtWUS o pGmLe1AtBBM.

30 En determinadas modalidades de la invención se proporciona un método de tejido vegetal transformante, que comprende poner en contacto al tejido vegetal con una construcción de ácido nucleico heteróloga que comprende un promotor y un factor de transcripción, donde el promotor es un promotor específico de embriones de plantas o específico de semillas.

35

En determinadas modalidades, el tejido vegetal es tejido de dicotiledónea.

En determinadas modalidades, el tejido de dicotiledónea es una célula de cacao, yuca, soja, tomate, tabaco, café, girasol, alfalfa, pimienta, caupí, girasol, uva, álamo, café, girasol, nuez, almendra, melocotón, nectarina, ciruela, albaricoque, manzana, pera, caqui u otro árbol frutal.

En determinadas modalidades, el tejido vegetal es tejido de gimnosperma.

En determinadas modalidades, el tejido de gimnosperma se selecciona de entre células de pino, ginkgo, abeto o abeto de Douglas.

En determinadas modalidades, el tejido vegetal es tejido de embrión cigótico.

En determinadas modalidades, el tejido vegetal es tejido de embrión somático.

En determinadas modalidades, el tejido vegetal es tejido de callo embriogénico somático.

En determinadas modalidades, el tejido vegetal es tejido organogénico.

En determinadas modalidades, el tejido vegetal es tejido meristemático de brotes.

En determinadas modalidades, el tejido vegetal es tejido de plántulas germinantes.

En determinadas modalidades, el tejido vegetal es tejido foliar.

En determinadas modalidades, el tejido vegetal es tejido derivado de protoplastos.

Aspectos de la invención

Un aspecto de la invención es una célula vegetal de dicotiledónea o gimnosperma transformada que comprende una construcción de ácido nucleico heterólogo que comprende un promotor específico de semillas de plantas y un factor de transcripción de genes morfogénicos BabyBoom (BBM).

En un aspecto, la célula vegetal es una dicotiledónea.

En un aspecto, la célula dicotiledónea es una célula de cacao, yuca, soja, tomate,

tabaco, alfalfa, pimienta, caupí, girasol, uva, álamo, café, nuez, almendra, melocotón, nectarina, ciruela, albaricoque, manzana, pera, caqui u otro árbol frutal.

En un aspecto, la célula vegetal es una célula de gimnosperma.

5

En un aspecto, la gimnosperma se selecciona de entre una célula de pino, ginkgo, abeto o abeto de Douglas.

En un aspecto, el promotor es un promotor específico de semillas de plantas.

10

En un aspecto, el promotor específico de semillas de plantas es un promotor específico de semillas de cotiledones.

En un aspecto, el promotor específico de semillas de cotiledones es un promotor del gen de la proteína de almacenamiento de la semilla.

15

En un aspecto, el promotor del gen de la proteína de almacenamiento de semillas es un promotor del gen de la proteína conglucina, glicina, lectina, faseolina, legumina, vicilina, convicilina, napina o cruciferina.

En un aspecto, el promotor específico de semillas de cotiledones es un promotor específico de embriones.

20

En un aspecto, el promotor específico de embriones es un promotor específico de semillas de soja.

En un aspecto, el promotor específico de semillas de soja es un promotor específico de cotiledones de soja.

25

En un aspecto, el promotor específico de cotiledones de soja es un promotor del gen de la proteína de almacenamiento de la semilla de soja.

En un aspecto, el promotor específico de cotiledones de soja es un promotor específico de embriones de soja.

30

En un aspecto, el promotor específico de embriones de soja es un promotor de lectina de soja.

35

En un aspecto, el promotor específico de embriones de soja es un promotor de glicinina de soja.

5 En un aspecto, el promotor específico de embriones de soja es un promotor de conglucina de soja.

En un aspecto, la construcción comprende además un vector.

En un aspecto, el vector es un plásmido.

10 En un aspecto, la construcción comprende además una recombinasa con especificidad de sitio.

En un aspecto, la construcción comprende además un gen de rasgo.

15 En un aspecto, la construcción comprende además un gen marcador seleccionable/visual.

En un aspecto, la construcción comprende además un casete de CRISPR.

20 Un aspecto de la invención es una construcción de ácido nucleico que comprende un promotor específico de semillas de plantas y un factor de transcripción de genes morfogénicos BabyBoom (BBM).

Un aspecto de la invención es una construcción de ácido nucleico que comprende pGmLe1AtBBM.

25 Un aspecto de la invención es un método de transformación de tejido vegetal de dicotiledónea o gimnosperma, que comprende poner en contacto al tejido vegetal con una construcción de ácido nucleico heterólogo que comprende un promotor específico de semillas de plantas y un factor de transcripción de genes morfogénicos BabyBoom (BBM).

30 En un aspecto, el tejido vegetal es tejido de dicotiledónea.

En un aspecto, el tejido de dicotiledónea es un tejido de cacao, yuca, soja, tomate, tabaco, alfalfa, pimiento, caupí, girasol, uva, álamo, café, nuez, almendra, melocotón, nectarina, ciruela, albaricoque, manzana, pera, caqui u otro árbol frutal.

En un aspecto, el tejido es tejido de gimnosperma.

En un aspecto, el tejido de gimnosperma se selecciona de entre tejido de pino, ginkgo, abeto o abeto de Douglas.

5

En un aspecto, el tejido vegetal es un embrión cigótico.

En un aspecto, el tejido vegetal es un embrión somático.

En un aspecto, el tejido vegetal es tejido de callo embriogénico somático.

10

En un aspecto, el tejido vegetal es tejido organogénico.

En un aspecto, el tejido vegetal es tejido meristemático de brotes.

En un aspecto, el tejido vegetal es una plántula germinante.

15

En un aspecto, el tejido vegetal es tejido foliar.

En un aspecto, el tejido vegetal es tejido derivado de protoplastos.

Ácidos nucleicos, casetes de expresión, vectores y células

20

En determinadas modalidades de la invención se proporciona un ácido nucleico que codifica un promotor y un factor de transcripción relacionado con la embriogénesis y/u organogénesis tal como se describe en el presente documento.

25

En determinadas modalidades de la invención se proporciona un casete de expresión que comprende un ácido nucleico tal como se describe en el presente documento y un promotor.

30

En determinadas modalidades de la invención se proporciona un vector (p. ej., un plásmido) que comprende un casete de expresión tal como se describe en el presente documento.

En determinadas modalidades de la invención se proporciona una célula que comprende un vector tal como se describe en el presente documento.

La expresión «ácido nucleico» hace referencia a desoxirribonucleótidos o

35

ribonucleótidos y polímeros de estos en forma mono o bicatenaria, compuestos por monómeros (nucleótidos) que contienen un azúcar, fosfato y una base que es una purina o pirimidina. A menos que se limite específicamente, el término abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales con propiedades de unión similares a las del ácido nucleico de referencia y se metabolizan de manera similar a los nucleótidos de origen natural. A menos que se indique lo contrario, una secuencia de ácido nucleico particular también abarca implícitamente variantes con modificaciones conservadoras de esta (p. ej., sustituciones de codones degenerados) y secuencias complementarias, así como la secuencia indicada explícitamente. Específicamente, las sustituciones de codones degenerados pueden lograrse mediante la generación de secuencias en las que la tercera posición de uno o más codones seleccionados (o todos ellos) se sustituye con residuos de base mixta y/o desoxiinosina (Batzner et al., *Nucleic Acid Res.*, 19:508 (1991); Ohtsuka et al., *JBC*, 260:2605 (1985); Rossolini et al., *Mol. Cell. Probes*, 8:91 (1994)). Un «fragmento de ácido nucleico» es una fracción de una molécula de ácido nucleico determinada. El ácido desoxirribonucleico (ADN) en la mayoría de los organismos es el material genético, mientras que el ácido ribonucleico (ARN) participa en la transferencia de la información contenida en el ADN a las proteínas. La expresión «secuencia de nucleótidos» hace referencia a un polímero de ADN o ARN que puede ser mono o bicatenario, que contiene opcionalmente bases de nucleótidos de carácter sintético, no natural o alterado con capacidad de incorporación a polímeros de ADN o ARN. Las expresiones «ácido nucleico», «molécula de ácido nucleico», «fragmento de ácido nucleico», «secuencia o segmento de ácido nucleico» o «polinucleótido» también pueden usarse indistintamente con un gen, ADNc, ADN y ARN codificados por un gen.

«Parte» o «fragmento», en lo que respecta a una molécula, una secuencia o un segmento de ácido nucleico de la invención, cuando se enlaza a otras secuencias para su expresión, hace referencia a una secuencia que tiene al menos 80 nucleótidos, más específicamente al menos 150 nucleótidos, e incluso más específicamente al menos 400 nucleótidos. Si no se emplea para la expresión, una «parte» o un «fragmento» hace referencia a al menos 9, específicamente 12, más específicamente 15, incluso más específicamente al menos 20 nucleótidos consecutivos, p. ej., sondas y cebadores (oligonucleótidos), correspondientes a la secuencia de nucleótidos de las moléculas de ácido nucleico de la invención.

La invención abarca composiciones de proteínas o ácidos nucleicos aislados o

considerablemente purificados. En el contexto de la presente invención, una molécula de ADN «aislada» o «purificada» o un polipéptido «aislado» o «purificado» es una molécula de ADN o un polipéptido que existe aparte de su entorno natural y, por lo tanto, no es un producto de la naturaleza. Una molécula de ADN o un polipéptido aislado puede existir en una forma purificada o puede existir en un entorno no natural tal como, por ejemplo, una célula hospedadora transgénica. Por ejemplo, una proteína o molécula de ácido nucleico «aislada» o «purificada», o una parte con actividad biológica de esta, está considerablemente libre de otro material celular, o medio de cultivo cuando se produce mediante técnicas recombinantes, o considerablemente libre de precursores químicos u otras sustancias químicas cuando se sintetiza químicamente. En una modalidad, un ácido nucleico «aislado» está libre de secuencias que flanquean naturalmente el ácido nucleico (es decir, secuencias ubicadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo del cual deriva el ácido nucleico. Por ejemplo, en diversas modalidades, la molécula de ácido nucleico aislada puede contener menos de alrededor de 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb o 0,1 kb de secuencias de nucleótidos que flanquean naturalmente la molécula de ácido nucleico en el ADN genómico de la célula de la que deriva el ácido nucleico. Una proteína considerablemente libre de material celular incluye preparaciones de proteína o polipéptido que tienen menos de alrededor de 30 %, 20 %, 10 %, 5 % (en peso seco) de proteína contaminante. Cuando la proteína de la invención, o una parte con actividad biológica de esta, se produce de forma recombinante, el medio de cultivo puede representar menos de alrededor de 30 %, 20 %, 10 % o 5 % (en peso seco) de precursores químicos o sustancias químicas que no sean proteínas de interés. La presente invención también abarca fragmentos y variantes de las secuencias de nucleótidos descritas y proteínas o proteínas de longitud parcial codificadas por estas. «Fragmento» o «parte» hace referencia a una longitud completa o menos que la longitud completa de la secuencia de nucleótidos que codifica, o la secuencia de aminoácidos de, un polipéptido o una proteína.

«De origen natural» se utiliza para describir un objeto que se puede encontrar en la naturaleza en contraposición a ser producido artificialmente. Por ejemplo, una proteína o una secuencia de nucleótidos presente en un organismo (incluido un virus), que puede aislarse de una fuente en la naturaleza y que no ha sido modificada intencionalmente por el ser humano en el laboratorio, es de origen natural.

Una «variante» de una molécula es una secuencia considerablemente similar a la

secuencia de la molécula natural. Para las secuencias de nucleótidos, las variantes incluyen aquellas secuencias que, debido a la degeneración del código genético, codifican la secuencia de aminoácidos idéntica de la proteína natural. Variantes alélicas de origen natural tales como estas pueden identificarse mediante el uso de técnicas de biología molecular muy conocidas, como, por ejemplo, la reacción en cadena de la polimerasa (PCR, por su sigla en inglés) y las técnicas de hibridación. Las variantes de secuencias de nucleótidos también incluyen secuencias de nucleótidos obtenidas sintéticamente, tales como las generadas, por ejemplo, mediante el uso de mutagénesis dirigida al sitio, que codifican la proteína natural, así como aquellas que codifican un polipéptido que tiene sustituciones de aminoácidos. Generalmente, las variantes de secuencias de nucleótidos de la invención tendrán al menos 40, 50, 60 o 70 %, p. ej., 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, generalmente al menos 80 %, p. ej., 81 %-84 %, al menos 85 %, p. ej., 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, de identidad de secuencia con respecto a la secuencia de nucleótidos natural (endógena).

«Variaciones con modificaciones conservadoras» de una secuencia de ácido nucleico particular hace referencia a aquellas secuencias de ácido nucleico que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o donde la secuencia de ácido nucleico no codifica una secuencia de aminoácidos, a secuencias esencialmente idénticas. Debido a la degeneración del código genético, una gran cantidad de ácidos nucleicos funcionalmente idénticos codifican cualquier polipéptido determinado. Por ejemplo, los codones CGT, CGC, CGA, CGG, AGA y AGG codifican en todos los casos el aminoácido arginina. Por lo tanto, en cada posición donde una arginina está especificada por un codón, el codón puede alterarse a cualquiera de los codones correspondientes descritos sin alterar la proteína codificada. Esas variaciones de ácido nucleico son «variaciones silenciosas» que son una especie de «variaciones con modificaciones conservadoras». Cada secuencia de ácido nucleico descrita en el presente documento que codifica un polipéptido también describe cada posible variación silenciosa, excepto que se indique lo contrario. Una persona experta reconocerá que cada codón en un ácido nucleico (excepto ATG, que normalmente es el único codón para metionina) puede modificarse para proporcionar una molécula funcionalmente idéntica mediante técnicas convencionales. En consecuencia, cada «variación silenciosa» de un ácido nucleico que codifica un polipéptido está implícita en cada secuencia descrita.

«Molécula de ADN recombinante» es una combinación de secuencias de ADN que se

unen mediante el uso de tecnología de ADN recombinante y procedimientos usados para unir secuencias de ADN tal como se describe, por ejemplo, en Sambrook y Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (3.^a edición, 2001).

5

Las expresiones «secuencia de ADN heteróloga», «segmento de ADN exógeno» o «ácido nucleico heterólogo» hacen referencia a una secuencia que se origina de una fuente extraña a la célula hospedadora particular o, si proviene de la misma fuente, se modifica a partir de su forma original. Por lo tanto, un gen heterólogo en una célula hospedadora incluye un gen que es endógeno con respecto a la célula hospedadora particular, pero que ha sido modificado. Esas expresiones también incluyen copias múltiples de origen no natural de una secuencia de ADN de origen natural. Por lo tanto, esas expresiones hacen referencia a un segmento de ADN que es extraño o heterólogo con respecto a la célula, u homólogo con respecto a la célula, pero que se encuentra en una posición dentro del ácido nucleico de la célula hospedadora en la que el elemento normalmente no se encuentra. Los segmentos de ADN exógenos se expresan para proporcionar polipéptidos exógenos.

10

15

Una secuencia de ADN «homóloga» es una secuencia de ADN que se asocia naturalmente a una célula hospedadora en la que se introduce.

20

«En estado natural» hace referencia al gen normal u organismo que se encuentra en la naturaleza sin ninguna mutación conocida.

«Genoma» hace referencia al material genético completo de un organismo.

25

Se define que un «vector» incluye, entre otras cosas, cualquier vector plasmídico, cosmídico, fágico o binario en forma circular o linear bi o monocatenaria que puede o no ser autotransmisible o movilizable y que puede transformar al hospedador procariota o eucariota, ya sea mediante integración en el genoma celular o existir extracromosómicamente (p. ej., un plásmido replicante autónomo con un origen de replicación).

30

Los «vectores de clonación» contienen normalmente uno o una pequeña cantidad de sitios de reconocimiento de endonucleasas de restricción en los que se pueden insertar secuencias de ADN extrañas de manera determinable sin pérdida de la función biológica esencial del vector, así como un gen marcador adecuado para el uso en la identificación y selección de células transformadas con el vector de clonación. Los genes marcadores incluyen

35

normalmente genes que proporcionan resistencia a la tetraciclina, resistencia a la higromicina o resistencia a la ampicilina.

«Casete de expresión», tal como se usa en el presente documento, significa una
5 secuencia de ADN capaz de dirigir la expresión de una secuencia de nucleótidos particular en una célula hospedadora adecuada, que comprende un promotor enlazado de manera operativa a la secuencia de nucleótidos de interés que está enlazada de manera operativa a las señales de terminación. También comprende normalmente secuencias necesarias para la traducción adecuada de la secuencia de nucleótidos. La región codificante habitualmente
10 codifica una proteína de interés, pero también puede codificar un ARN funcional de interés, por ejemplo, un ARN no codificante o un ARN no traducido, en la dirección codificante o no codificante. El casete de expresión que comprende la secuencia de nucleótidos de interés puede ser quimérico, lo que significa que al menos uno de sus componentes es heterólogo con respecto al menos a uno de sus otros componentes. El casete de expresión también puede
15 ser uno de origen, natural pero que se haya obtenido en una forma recombinante útil para la expresión heteróloga. La expresión de la secuencia de nucleótidos en el casete de expresión puede estar bajo el control de un promotor constitutivo o de un promotor inducible que inicie la transcripción solamente cuando la célula hospedadora se exponga a algún estímulo externo particular. En el caso de un organismo multicelular, el promotor también puede ser específico
20 de un tejido, un órgano o una etapa de desarrollo particular.

Esos casetes de expresión comprenderán la región de inicio de la transcripción de la invención enlazada a una secuencia de nucleótidos de interés. Ese casete de expresión está provisto de múltiples sitios de restricción para que la inserción del gen de interés esté bajo la
25 regulación de transcripción de las regiones reguladoras. El casete de expresión puede contener además genes marcadores seleccionables.

Por lo general, el casete de expresión comprenderá un gen marcador seleccionable para la selección de células transformadas. Los genes marcadores seleccionables se utilizan para la selección de células o tejidos transformados. Los genes marcadores incluyen genes
30 que codifican la resistencia a los antibióticos. Tales como los que codifican la neomicina fosfotransferasa II (nptII) y la higromicina fosfotransferasa (HPT), así como genes que confieren resistencia a compuestos herbicidas, tales como el glufosinato de amonio, el bromoxinilo, las imidazolinonas, el glifosato y el 2,4-diclorofenoxiacetato (2,4-D). En la

presente invención se puede utilizar cualquier gen marcador seleccionable.

La expresión «transcripción de ARN» hace referencia al producto resultante de la transcripción catalizada por ARN polimerasa de una secuencia de ADN. Cuando la transcripción de ARN es una copia complementaria perfecta de la secuencia de ADN, se la denomina «transcripción primaria», o puede ser una secuencia de ARN derivada del procesamiento posterior a la transcripción de la transcripción primaria y se la denomina «ARN maduro». El «ARN mensajero» (ARNm) hace referencia al ARN que no tiene intrones y que la célula puede traducir en proteína. El «ADNc» hace referencia a un ADN mono o bicatenario que es complementario con respecto al ARNm y deriva de este.

«Secuencias reguladoras» y «secuencias reguladoras adecuadas» hacen referencia en cada caso a secuencias de nucleótidos en direcciones retrógradas (secuencias 5' no codificantes), dentro o en direcciones anterógradas (secuencias 3' no codificantes) con respecto a una secuencia codificante, y que influyen en la transcripción, el procesamiento o la estabilidad del ARN, o la traducción de la secuencia codificante asociada. Las secuencias reguladoras incluyen potenciadores, promotores, secuencias líderes de traducción, intrones y secuencias señal de poliadenilación. Incluyen secuencias naturales y sintéticas, así como secuencias que pueden ser una combinación de secuencias sintéticas y naturales. Como se ha señalado anteriormente, la expresión «secuencias reguladoras adecuadas» no se limita a los promotores. Sin embargo, algunas secuencias reguladoras adecuadas útiles en la presente invención incluirán, de modo no taxativo, promotores constitutivos, promotores específicos de tejidos, promotores específicos de desarrollo, promotores inducibles y promotores virales.

«Secuencia 5' no codificante» hace referencia a una secuencia de nucleótidos ubicada en dirección 5' (retrógrada) con respecto a la secuencia codificante. Está presente en el ARNm completamente procesado en dirección retrógrada con respecto al codón de inicio y puede afectar el procesamiento del transcrito primario para obtener ARNm, la estabilidad del ARNm o la eficiencia de la traducción (Turner et al., Mol. Biotech., 3:225 (1995)).

«Secuencia 3' no codificante» hace referencia a secuencias de nucleótidos ubicadas en dirección 3' (en dirección anterógrada) a una secuencia codificante e incluye secuencias de señales de poliadenilación y otras secuencias que codifican señales reguladoras capaces de afectar el procesamiento de ARNm o la expresión génica. La señal de poliadenilación habitualmente se caracteriza por afectar la adición de trectos de ácido poliadenílico al

extremo 3' del precursor de ARNm.

La expresión «secuencia líder de traducción» hace referencia a la parte de la secuencia de ADN de un gen entre el promotor y la secuencia codificante que se transcribe en ARN y está presente en el ARNm completamente procesado en dirección retrógrada (5') con respecto al codón de inicio de la traducción. La secuencia líder de traducción puede afectar el procesamiento del transcrito primario para obtener ARNm, la estabilidad del ARNm o la eficiencia de la traducción.

El término proteína «madura» hace referencia a un polipéptido procesado después de la traducción sin su péptido de señal. La proteína «precursora» hace referencia al producto primario de la traducción de un ARNm. «Péptido señal» hace referencia a la extensión hacia el extremo amino de un polipéptido, que se traduce junto con el polipéptido que forma un péptido precursor y que se requiere para su entrada en la vía secretora. La expresión «secuencia [de] señal» hace referencia a una secuencia de nucleótidos que codifica el péptido señal.

«Promotor» hace referencia a una secuencia de nucleótidos, habitualmente en dirección retrógrada (5') con respecto a su secuencia codificante, que controla la expresión de la secuencia codificante al proporcionar el reconocimiento de la ARN polimerasa y otros factores necesarios para una transcripción adecuada. «Promotor» incluye un promotor mínimo que es una secuencia de ADN corta compuesta por una caja TATA y otras secuencias que sirven para especificar el sitio de inicio de la transcripción, que recibe la adición de elementos reguladores para el control de la expresión. «Promotor» también hace referencia a una secuencia de nucleótidos que incluye un promotor mínimo más elementos reguladores que son capaces de controlar la expresión de una secuencia codificante o ARN funcional. Este tipo de secuencia promotora consiste en elementos en dirección retrógrada próximos y más distantes; estos últimos elementos a menudo se denominan «potenciadores». En consecuencia, un «potenciador» es una secuencia de ADN que puede estimular la actividad promotora y puede ser un elemento innato del promotor o un elemento heterólogo insertado para mejorar el nivel o la especificidad tisular de un promotor. Los promotores pueden derivar en su totalidad de un gen natural, o estar compuestos por diferentes elementos derivados de diferentes promotores que se encuentren en la naturaleza, o incluso estar compuestos por segmentos de ADN sintético. Un promotor también puede contener secuencias de ADN que

participen de la unión de factores proteicos que controlen la eficacia del inicio de la transcripción en respuesta a condiciones fisiológicas o de desarrollo.

El «sitio de inicio» es la posición que rodea el primer nucleótido que es parte de la secuencia transcrita, que también se define como posición +1. Con respecto a este sitio, todas las demás secuencias del gen y sus regiones controladoras están numeradas. Las secuencias en dirección anterógrada (es decir, otras secuencias que codifican proteínas en la dirección 3') se denominan «positivas», mientras que las secuencias en dirección retrógrada (en su mayoría, de las regiones de control en la dirección 5') se denominan «negativas».

Los elementos promotores, particularmente un elemento TATA, que están inactivos o que han reducido en gran medida la actividad promotora en ausencia de activación en dirección retrógrada se conocen como «promotores mínimos o centrales». En presencia de un factor de transcripción adecuado, el promotor mínimo funciona para permitir la transcripción. Un «promotor mínimo o central», por lo tanto, consiste solo en todos los elementos basales necesarios para el inicio de la transcripción, p. ej., una caja TATA y/o un iniciador.

«Expresión constitutiva» hace referencia a la expresión que usa un promotor constitutivo o regulado. «Expresión regulada» y «condicional» hacen referencia a la expresión controlada por un promotor regulado.

«Enlazado de manera operativa» hace referencia a la asociación de secuencias de ácido nucleico en un solo fragmento de ácido nucleico de forma tal que la función de una de las secuencias se vea afectada por la otra. Por ejemplo, se dice que una secuencia de ADN reguladora está «enlazada de manera operativa a» o «asociada a» una secuencia de ADN que codifica un ARN o un polipéptido si las dos secuencias están situadas de forma tal que la secuencia de ADN reguladora afecte la expresión de la secuencia de ADN codificante (es decir, que la secuencia codificante o el ARN funcional esté bajo el control de transcripción del promotor). Las secuencias codificantes pueden enlazarse de manera operativa a secuencias reguladoras en orientación codificante o no codificante.

«Expresión» hace referencia a la transcripción y/o traducción en una célula de un gen endógeno, transgén, así como la transcripción y acumulación estable de ARN codificante (ARNm) o funcional. En el caso de las construcciones no codificantes, la expresión puede hacer referencia únicamente a la transcripción del ADN no codificante. La expresión también

puede hacer referencia a la producción de proteínas.

«Fragmento de terminación de la transcripción» hace referencia a secuencias de nucleótidos que contienen una o más señales reguladoras, tales como secuencias de señales de poliadenilación, capaces de terminar la transcripción. En la técnica se conocen ejemplos de fragmentos de terminación de la transcripción.

«Fragmento de terminación de la traducción» hace referencia a secuencias de nucleótidos que contienen una o más señales reguladoras, tales como uno o más codones de terminación en los tres marcos, capaces de terminar la traducción. La inserción de un fragmento de terminación de la traducción adyacente o cerca del codón de inicio en el extremo 5' de la secuencia codificante no dará como resultado ninguna traducción o traducción incorrecta. La escisión del fragmento de terminación de la traducción mediante recombinación con especificidad de sitio dejará una secuencia con especificidad de sitio en la secuencia codificante que no interfiera con la traducción adecuada mediante el uso del codón de inicio.

Las expresiones «secuencia de acción en *cis*» y «elemento de acción en *cis*» hacen referencia a secuencias de ADN o ARN cuyas funciones requieren que estén en la misma molécula.

Las expresiones «secuencia de acción en *trans*» y «elemento de acción en *trans*» hacen referencia a secuencias de ADN o ARN cuyas funciones no requieren que estén en la misma molécula.

Las siguientes expresiones se utilizan para describir las relaciones de secuencias entre dos o más secuencias (p. ej., ácidos nucleicos, polinucleótidos o polipéptidos): (a) «secuencia de referencia», (b) «ventana de comparación», (c) «identidad de secuencia», (d) «porcentaje de identidad de secuencia» y (e) «identidad considerable».

(a) Tal como se usa en el presente documento, la «secuencia de referencia» es una secuencia definida utilizada como base para la comparación de secuencias. Una secuencia de referencia puede ser un subconjunto o la totalidad de una secuencia especificada; por ejemplo, como un segmento de un ADNc, secuencia genética o secuencia peptídica de longitud completa, o el ADNc, la secuencia genética o la secuencia peptídica en su totalidad.

(b) Tal como se usa en el presente documento, la «ventana de comparación» hace

referencia a un segmento contiguo y especificado de una secuencia, donde la secuencia en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para una alineación óptima de las dos secuencias. Generalmente, la ventana de comparación tiene al menos 20 nucleótidos contiguos de longitud y, opcionalmente, puede tener 30, 40, 50, 100 o más. Las personas expertas en la técnica entienden que, para evitar una gran similitud con una secuencia de referencia debido a la inclusión de huecos en la secuencia, normalmente se introduce una penalización por hueco que se resta de la cantidad de coincidencias.

Los métodos de alineación de secuencias para la comparación son muy conocidos en la técnica. Por lo tanto, la determinación del porcentaje de identidad entre dos secuencias cualesquiera puede lograrse mediante el uso de un algoritmo matemático. Los ejemplos no taxativos de esos algoritmos matemáticos son el algoritmo de Myers y Miller, CABIOS, 4:11 (1988); el algoritmo de homología local de Smith et al., Adv. Appl. Math., 2:482 (1981); el algoritmo de alineación por homología de Needleman y Wunsch, JMB, 48:443 (1970); el método de búsqueda de similitudes de Pearson y Lipman, Proc. Natl. Acad. Sci. USA, 85:2444 (1988); el algoritmo de Karlin y Altschul, Proc. Natl. Acad. Sci. USA, 87:2264 (1990), con las modificaciones de Karlin y Altschul, Proc. Natl. Acad. Sci. USA, 90:5873 (1993).

Se pueden utilizar implementaciones informáticas de estos algoritmos matemáticos para comparar secuencias y determinar la identidad de la secuencia. Esas implementaciones incluyen, de modo no taxativo, CLUSTAL en el programa PC/Gene (disponible a través de Intelligenetics, Mountain View, California); el programa ALIGN (versión 2.0) y GAP, BESTFIT, BLAST, FASTA y TFASTA en el paquete de software de Wisconsin Genetics, versión 8 (disponible a través de Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, EUA). Las alineaciones que utilizan estos programas pueden realizarse mediante el uso de los parámetros predeterminados. El programa CLUSTAL está bien descrito por Higgins et al., Gene, 73:237 (1988); Higgins et al., CABIOS, 5:151 (1989); Corpet et al., Nucl. Acids Res., 16:10881 (1988); Huang et al., CABIOS, 8:155 (1992); y Pearson et al., Meth. Mol. Biol., 24:307 (1994). El programa ALIGN se basa en el algoritmo de Myers y Miller, *supra*. Los programas BLAST de Altschul et al., JMB, 215:403 (1990); Nucl. Acids Res., 25:3389 (1990), se basan en el algoritmo de Karlin y Altschul *supra*.

El software para realizar análisis BLAST está disponible públicamente a través del

Centro Nacional de Información Biotecnológica (disponible en la red mundial en ncbi.nlm.nih.gov/). Este algoritmo implica primero identificar pares de secuencias de puntuación alta (HSP) mediante la identificación de palabras cortas de longitud W en la secuencia de consulta, que coinciden o cumplen con alguna puntuación umbral de valor positivo T cuando se alinean con una palabra de la misma longitud en una secuencia de base de datos. Se hace referencia a T como el umbral de puntuación de palabras vecinas. Estos aciertos iniciales de palabras vecinas actúan como semillas para iniciar búsquedas para encontrar HSP más largas que las contengan. Luego, los aciertos de palabras se extienden en ambas direcciones a lo largo de cada secuencia hasta donde se pueda aumentar la puntuación de alineación acumulada. Las puntuaciones acumuladas se calculan mediante el uso, para secuencias de nucleótidos, de los parámetros M (puntuación de recompensa para un par de residuos coincidentes; siempre > 0) y N (puntuación de penalización para residuos no coincidentes; siempre < 0). Para las secuencias de aminoácidos, se utiliza una matriz de puntuación para calcular la puntuación acumulada. La extensión de los aciertos de palabras en cada dirección se detiene cuando la puntuación de alineación acumulada cae en la cantidad X con respecto a su valor máximo alcanzado, la puntuación acumulada llega a cero o menos debido a la acumulación de una o más alineaciones de residuos de puntuación negativa, o se llega al final de cualquiera de las secuencias.

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias. Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña ($P(N)$), que proporciona una indicación de la probabilidad por la cual se produciría una coincidencia entre dos secuencias de nucleótidos o aminoácidos por casualidad. Por ejemplo, una secuencia de ácido nucleico de prueba se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación de la secuencia de ácido nucleico de prueba con la secuencia de ácido nucleico de referencia es menor que alrededor de 0,1, más específicamente menor que alrededor de 0,01, y más específicamente menor que alrededor de 0,001.

Para obtener alineaciones con espacios con propósitos de comparación, se puede utilizar Gapped BLAST (en BLAST 2.0) tal como se describe en Altschul et al., *Nucleic Acids Res.* 25:3389 (1997). De manera alternativa, se puede utilizar PSI-BLAST (en BLAST 2.0) para realizar una búsqueda iterada que detecte relaciones distantes entre moléculas. Véase

Altschul et al., *supra*. Cuando se utiliza BLAST, Gapped BLAST, PSI-BLAST, se pueden usar los parámetros predeterminados de los respectivos programas (p. ej., BLASTN para secuencias de nucleótidos, BLASTX para proteínas). El programa BLASTN (para secuencias de nucleótidos) utiliza por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, un límite de 100, M=5, N=-4 y una comparación de ambas cadenas. Para secuencias de aminoácidos, el programa BLASTP utiliza por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10 y la matriz de puntuación BLOSUM62. Consulte la red mundial en ncbi.nlm.nih.gov. La alineación también puede realizarse manualmente mediante inspección visual.

A los efectos de la presente invención, la comparación de secuencias para la determinación del porcentaje de identidad de secuencia con otra secuencia puede realizarse mediante el uso del programa BlastN (versión 1.4.7 o posterior) con sus parámetros predeterminados o cualquier programa equivalente. Por «programa equivalente» se entiende cualquier programa de comparación de secuencias que, para dos secuencias cualesquiera en cuestión, genera una alineación que tiene coincidencias idénticas de nucleótidos o residuos de aminoácidos y una identidad de porcentaje de secuencia total en comparación con la alineación correspondiente generada por el programa preferido.

(c) Tal como se usa en el presente documento, «identidad de secuencia» o «identidad» en el contexto de dos secuencias de ácido nucleico o polipéptido hace referencia a un porcentaje especificado de residuos en las dos secuencias que son iguales cuando se alinean para la correspondencia máxima a través de una ventana de comparación especificada, tal como se mide mediante algoritmos de comparación de secuencias o mediante inspección visual. Cuando se utiliza el porcentaje de identidad de secuencia en referencia a proteínas, se reconoce que las posiciones de los residuos que no son idénticas a menudo difieren mediante sustituciones conservadoras de aminoácidos, donde los residuos de aminoácidos se sustituyen por otros residuos de aminoácidos con propiedades químicas similares (p. ej., carga o hidrofobicidad) y, por lo tanto, no cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en sustituciones conservadoras, el porcentaje de identidad de secuencia puede ajustarse al alza para corregir la naturaleza conservadora de la sustitución. Se dice que las secuencias que difieren por esas sustituciones conservadoras tienen «similitud de secuencia» o «similitud». Los medios para realizar este ajuste son muy conocidos por las personas expertas en la técnica. Normalmente, esto implica calificar una sustitución

conservadora como una discrepancia parcial en lugar de completa, aumentando así el porcentaje de identidad de secuencia. Así, por ejemplo, cuando a un aminoácido idéntico se le asigna una puntuación de 1 y a una sustitución no conservadora se le asigna una puntuación de cero, a una sustitución conservadora se le asigna una puntuación entre cero y 1. La puntuación de las sustituciones conservadoras se calcula, p. ej., tal como se implementa en el programa PC/GENE (Intelligenetics, Mountain View, California).

(d) Tal como se usa en el presente documento, «porcentaje de identidad de secuencia» significa el valor determinado al comparar dos secuencias alineadas óptimamente a través de una ventana de comparación, donde la parte de la secuencia en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para la alineación óptima de las dos secuencias. El porcentaje se calcula mediante la determinación de la cantidad de posiciones en las que aparece la base de ácido nucleico o el residuo de aminoácido idéntico en ambas secuencias para proporcionar la cantidad de posiciones coincidentes, dividiendo la cantidad de posiciones coincidentes por la cantidad total de posiciones en la ventana de comparación y multiplicar el resultado por 100 para proporcionar el porcentaje de identidad de secuencia.

(e)(i) La expresión «identidad considerable» de secuencias significa que un polinucleótido comprende una secuencia que tiene al menos 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 % o 79 %, al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 % u 89 %, al menos 90 %, 91 %, 92 %, 93 % o 94 %, y al menos 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia, en comparación con una secuencia de referencia mediante el uso de uno de los programas de alineación descritos con parámetros convencionales. Una persona experta en la técnica reconocerá que estos valores pueden ajustarse de manera adecuada para determinar la identidad correspondiente de las proteínas codificadas por dos secuencias de nucleótidos teniendo en cuenta la degeneración de codones, la similitud de aminoácidos, el posicionamiento del marco de lectura y similares. Identidad considerable de secuencias de aminoácidos para estos fines normalmente significa una identidad de secuencia de al menos 70 %, al menos 80 %, 90 % o al menos 95 %.

Otra indicación de que las secuencias de nucleótidos son considerablemente idénticas es si dos moléculas se hibridan entre sí en condiciones rigurosas (véase más adelante). Generalmente, las condiciones rigurosas se seleccionan para que sean alrededor de 5 °C más

bajas que el punto de fusión térmica (T_m) para la secuencia específica con una fuerza iónica y un pH definidos. Sin embargo, las condiciones rigurosas abarcan temperaturas en el intervalo de alrededor de 1 °C a alrededor de 20 °C, dependiendo del grado de rigurosidad deseado, tal como se califica de otro modo en el presente documento. Los ácidos nucleicos que no se hibridan entre sí en condiciones rigurosas siguen siendo considerablemente idénticos si los polipéptidos que codifican son considerablemente idénticos. Esto puede ocurrir, p. ej., cuando se crea una copia de un ácido nucleico mediante el uso de la máxima degeneración de codones permitida por el código genético. Una indicación de que dos secuencias de ácidos nucleicos son considerablemente idénticas es cuando el polipéptido codificado por el primer ácido nucleico tiene reactividad cruzada inmunológica con el polipéptido codificado por el segundo ácido nucleico.

(e)(ii) La expresión «identidad considerable» en el contexto de un péptido indica que un péptido comprende una secuencia con al menos 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, o 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, u 89 %, al menos 90 %, 91 %, 92 %, 93 % o 94 %, o 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia con respecto a la secuencia de referencia a través de una ventana de comparación especificada. La alineación óptima se realiza mediante el uso del algoritmo de alineación por homología de Needleman y Wunsch, J. Mol. Biol. 48:443 (1970). Una indicación de que dos secuencias peptídicas son considerablemente idénticas es que un péptido es inmunológicamente reactivo con anticuerpos generados contra el segundo péptido. Por lo tanto, un péptido es considerablemente idéntico a un segundo péptido, por ejemplo, cuando los dos péptidos difieren solamente por una sustitución conservadora.

Para la comparación de secuencias, normalmente una secuencia actúa como secuencia de referencia con la que se comparan las secuencias de prueba. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se ingresan en una computadora, se designan las coordenadas de la subsecuencia si es necesario y se designan los parámetros del programa del algoritmo de secuencia. Luego, el algoritmo de comparación de secuencias calcula el porcentaje de identidad de secuencia para la o las secuencias de prueba en relación con la secuencia de referencia, en función de los parámetros designados del programa.

Tal como se señaló anteriormente, otra indicación de que dos secuencias de ácidos

nucleicos son considerablemente idénticas es que las dos moléculas se hibridan entre sí en condiciones rigurosas. La expresión «hibridación específicamente con» hace referencia a la unión, duplexación o hibridación de una molécula solamente con una secuencia de nucleótidos particular en condiciones rigurosas cuando esa secuencia está presente en un ADN o ARN de
 5 mezcla compleja (p. ej., celular total). «Unir[se] considerablemente» hace referencia a la hibridación complementaria entre un ácido nucleico sonda y un ácido nucleico objetivo y abarca desajustes menores que pueden acomodarse mediante la reducción de la rigurosidad de los medios de hibridación para lograr la detección deseada de la secuencia de ácido
 10 nucleico objetivo.

Las «condiciones rigurosas de hibridación» y las «condiciones rigurosas de hibridación y lavado», en el contexto de los experimentos de hibridación de ácidos nucleicos, tales como las hibridaciones Southern y Northern, dependen de la secuencia y son diferentes en función de diferentes parámetros ambientales. Las secuencias más largas se hibridan específicamente
 15 a temperaturas más altas. El punto de fusión térmica (T_m) es la temperatura (con fuerza iónica y pH definidos) a la que el 50 % de la secuencia objetivo se hibrida con una sonda con coincidencia perfecta. La especificidad suele ser la función de los lavados posteriores a la hibridación, siendo los factores críticos la fuerza iónica y la temperatura de la solución de lavado final. Para híbridos ADN-ADN, la T_m puede aproximarse a partir de la ecuación de
 20 Meinkoth y Wahl, Anal. Biochem., 138:267 (1984); $T_m = 81,5\text{ °C} + 16,6 (\log M) + 0,41 (\%GC) - 0,61 (\% \text{ forma}) - 500/L$; donde M es la molaridad de cationes monovalentes, %GC es el porcentaje de nucleótidos de guanosina y citosina en el ADN, % forma es el porcentaje de formamida en la solución de hibridación y L es la longitud del híbrido en pares de bases. La T_m se reduce en alrededor de 1 °C para cada 1 % de falta de coincidencia; por lo tanto, la T_m , las
 25 condiciones de hibridación y/o lavado pueden ajustarse para hibridarse con secuencias de la identidad deseada. Por ejemplo, si se buscan secuencias con una identidad >90 %, la T_m puede disminuir 10 °C. Generalmente, las condiciones rigurosas se seleccionan para que sean alrededor de 5 °C más bajas que la T_m para la secuencia específica y su complemento con una fuerza iónica y un pH definidos. Sin embargo, las condiciones severamente rigurosas
 30 pueden utilizar una hibridación y/o un lavado a 1, 2, 3, o 4 °C menos que la T_m ; las condiciones moderadamente rigurosas pueden utilizar una hibridación y/o un lavado a 6, 7, 8, 9 o 10 °C menos que la T_m ; las condiciones de baja rigurosidad pueden utilizar una hibridación y/o un lavado a 11, 12, 13, 14, 15, o 20 °C menos que la T_m . Mediante el uso de la ecuación, las composiciones de hibridación y lavado, y la temperatura deseada, las personas expertas

entenderán que las variaciones en la rigurosidad de las soluciones de hibridación y/o lavado se describen inherentemente. Si el grado deseado de falta de coincidencia da como resultado una temperatura de menos de 45 °C (solución acuosa) o 32 °C (solución de formamida), se prefiere aumentar la concentración de SSC para poder utilizar una temperatura más alta. Una extensa guía para la hibridación de ácidos nucleicos se encuentra en Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology Hybridization with Nucleic Acid Probes, parte I, capítulo 2 «Overview of principles of hybridization and the strategy of nucleic acid probe assays» Elsevier, Nueva York (1993). Generalmente, las condiciones de hibridación y lavado altamente rigurosas se seleccionan para que sean alrededor de 5 °C inferiores a la T_m para la secuencia específica con una fuerza iónica y un pH definidos.

Un ejemplo de condiciones de lavado muy rigurosas es NaCl a 0,15 M a 72 °C durante alrededor de 15 minutos. Un ejemplo de condiciones de lavado rigurosas es un lavado con SSC 0,2X a 65 °C durante 15 minutos (véase Sambrook, *infra*, para obtener una descripción del agente de amortiguación SSC). A menudo, un lavado de alta rigurosidad va precedido por un lavado de baja rigurosidad para eliminar la señal de fondo de la sonda. Un ejemplo de lavado de rigurosidad media para una estructura doble de, p. ej., más de 100 nucleótidos, es SSC 1X a 45 °C durante 15 minutos. Un ejemplo de lavado de baja rigurosidad para una estructura doble de, p. ej., más de 100 nucleótidos, es SSC 4-6X a 40 °C durante 15 minutos. Para sondas cortas (p. ej., alrededor de 10 a 50 nucleótidos), las condiciones rigurosas normalmente involucran concentraciones de sal de menos de alrededor de 1,5 M, más específicamente alrededor de 0,01 a 1,0 M, concentración de iones de Na (u otras sales) con un pH de 7,0 a 8,3, y la temperatura es normalmente de al menos alrededor de 30 °C y al menos alrededor de 60 °C para sondas largas (p. ej., >50 nucleótidos). También se pueden lograr condiciones rigurosas con la adición de agentes de desestabilización tales como formamida. En general, una relación señal-ruido de 2X (o superior) con respecto a la observada para una sonda no relacionada en el ensayo de hibridación particular indica la detección de una hibridación específica. Los ácidos nucleicos que no se hibridan entre sí en condiciones rigurosas siguen siendo considerablemente idénticos si las proteínas que codifican son considerablemente idénticas. Esto ocurre, p. ej., cuando se crea una copia de un ácido nucleico mediante el uso de la máxima degeneración de codones permitida por el código genético.

Se seleccionan condiciones muy rigurosas para que sean iguales a la T_m para una

sonda particular. Un ejemplo de condiciones rigurosas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 residuos complementarios en un filtro en una transferencia Southern o Northern es formamida al 50 %, p. ej., hibridación en formamida al 50 %, NaCl a 1 M, SDS al 1 % a 37 °C y un lavado en SSC 0,1X a entre 60 y 65 °C. Los ejemplos de condiciones de baja rigurosidad incluyen la hibridación con una solución amortiguadora de formamida a entre 30 y 35 %, NaCl a 1 M, SDS (dodecil sulfato de sodio) al 1 % a 37 °C, y un lavado en SSC a entre 1X y 2X (SSC 20X = NaCl a 3,0 M/citrato trisódico a 0,3 M) a entre 50 y 55 °C. Los ejemplos de condiciones de rigurosidad moderada incluyen la hibridación en formamida a entre 40 y 45 %, NaCl a 1,0 M, SDS al 1 % a 37 °C, y un lavado en SSC a entre 0,5X y 1X a entre 55 y 60 °C.

Por «variante» de polipéptido se entiende un polipéptido derivado de la proteína natural por delección (que se denomina «truncamiento») o adición de uno o más aminoácidos en el extremo N y/o el extremo C de la proteína natural; delección o adición de uno o más aminoácidos en uno o más sitios en la proteína natural; o sustitución de uno o más aminoácidos en uno o más sitios en la proteína natural. Esas variantes pueden ser el resultado de, por ejemplo, polimorfismo genético o de manipulación humana. Los métodos para esas manipulaciones son generalmente conocidos en la técnica.

Por lo tanto, los polipéptidos de la invención pueden alterarse de diversas maneras, lo cual incluye sustituciones, delecciones, truncamientos e inserciones de aminoácidos. Los métodos para esas manipulaciones son generalmente conocidos en la técnica. Por ejemplo, se pueden preparar variantes de secuencias de aminoácidos de los polipéptidos mediante mutaciones en el ADN. Los métodos para mutagénesis y alteraciones de secuencias de nucleótidos son muy conocidos en la técnica. Véase, por ejemplo, Kunkel, Proc. Natl. Acad. Sci. USA, 82:488 (1985); Kunkel et al., Meth. Enzymol., 154:367 (1987); la patente de EE. UU. n.º 4 873 192; Walker y Gaastra, Techniques in Mol. Biol. (MacMillan Publishing Co. (1983), y las referencias allí citadas. Se puede encontrar orientación sobre las sustituciones de aminoácidos adecuadas que no afectan la actividad biológica de la proteína de interés en el modelo de Dayhoff et al., Atlas of Protein Sequence and Structure (Natl. Biomed. Res. Found. 1978). Se prefieren las sustituciones conservadoras, tales como el intercambio de un aminoácido por otro que tenga propiedades similares.

Por lo tanto, los genes y las secuencias de nucleótidos de la invención incluyen tanto

las secuencias de origen natural como las formas mutantes. Asimismo, los polipéptidos de la invención abarcan proteínas de origen natural, así como variaciones y formas modificadas de estas. Esas variantes seguirán poseyendo la actividad deseada. No se espera que las delecciones, inserciones y sustituciones de la secuencia polipeptídica abarcada en el presente documento produzcan cambios radicales en las características del polipéptido. Sin embargo, cuando es difícil predecir el efecto exacto de la sustitución, delección o inserción antes de realizarla, una persona experta en la técnica se dará cuenta de que el efecto se evaluará mediante ensayos de detección de rutina.

Las sustituciones, delecciones o adiciones individuales que alteran, agregan o eliminan un solo aminoácido o un pequeño porcentaje de aminoácidos (normalmente menos del 5 %, más normalmente menos del 1 %) en una secuencia codificada son «variaciones con modificaciones conservadoras», donde las alteraciones dan como resultado la sustitución de un aminoácido con un aminoácido químicamente similar. Las tablas de sustituciones conservadoras que proporcionan aminoácidos funcionalmente similares son muy conocidas en la técnica. Cada uno de los cinco grupos siguientes contiene aminoácidos que son sustituciones conservadoras entre sí: alifáticos: glicina (G), alanina (A), valina (V), leucina (L), isoleucina (I); aromáticos: fenilalanina (F), tirosina (Y), triptófano (W); que contienen azufre: metionina (M), cisteína (C); básicos: arginina (R), lisina (K), histidina (H); ácidos: ácido aspártico (D), ácido glutámico (E), asparagina (N), glutamina (Q). Además, las sustituciones, delecciones o adiciones individuales que alteran, agregan o eliminan un solo aminoácido o un pequeño porcentaje de aminoácidos en una secuencia codificada también son «variaciones con modificaciones conservadoras».

El término «transformación» hace referencia a la transferencia de un fragmento de ácido nucleico al genoma de una célula hospedadora, lo que da como resultado una herencia genéticamente estable. Las células hospedadoras que contienen los fragmentos de ácido nucleico transformados se conocen como células «transgénicas», y los organismos que comprenden células transgénicas se conocen como «organismos transgénicos».

«Transformado/a», «transgénico/a» y «recombinante» hacen referencia a una célula o un organismo hospedador en el que se ha introducido una molécula de ácido nucleico heteróloga. La molécula de ácido nucleico puede integrarse de manera estable en el genoma generalmente conocido en la técnica y se describe en Sambrook y Russell, *supra*. Véase

también Innis et al., PCR Protocols, Academic Press (1995); y Gelfand, PCR Strategies, Academic Press (1995); e Innis y Gelfand, PCR Methods Manual, Academic Press (1999). Los métodos conocidos de PCR incluyen, de modo no taxativo, métodos que utilizan cebadores emparejados, cebadores anidados, cebadores específicos individuales, cebadores degenerados, cebadores específicos de genes, cebadores específicos de vectores, cebadores parcialmente no coincidentes y similares. Por ejemplo, las células «transformadas», «transformantes» y «transgénicas» han pasado por el proceso de transformación y contienen un gen extraño integrado en su cromosoma. La expresión «sin transformar» hace referencia a las células normales que no han pasado por el proceso de transformación.

Ácidos nucleicos

En la presente invención se proporcionan ácidos nucleicos que codifican partes o la totalidad de una proteína. La expresión «ácido nucleico» hace referencia a ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN) y polímeros de estos en forma mono o bicatenaria, compuestos por monómeros (nucleótidos) que contienen un azúcar, fosfato y una base que es una purina o pirimidina. A menos que se limite específicamente, el término abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales con propiedades de unión similares a las del ácido nucleico de referencia y se metabolizan de manera similar a los nucleótidos de origen natural. A menos que se indique lo contrario, una secuencia de ácido nucleico particular también incluye implícitamente variantes con modificaciones conservadoras de esta (p. ej., sustituciones de codones degenerados) y secuencias complementarias, así como la secuencia indicada explícitamente. Específicamente, las sustituciones de codones degenerados pueden lograrse mediante la generación de secuencias en las que la tercera posición de uno o más codones seleccionados (o todos ellos) se sustituye con residuos de base mixta y/o desoxiinosina. Un «fragmento de ácido nucleico» es una parte de una molécula de ácido nucleico determinada.

Las expresiones «polinucleótido», «ácido nucleico» y «fragmento de ácido nucleico» se utilizan indistintamente en el presente documento. Estos términos abarcan nucleótidos conectados por enlaces fosfodiéster. Un «polinucleótido» puede ser un polímero de ácido ribonucleico (ARN) o ácido desoxirribonucleico (ADN) mono o bicatenario, que opcionalmente contiene bases de nucleótidos sintéticas, no naturales o alteradas. Un polinucleótido en forma de polímero de ADN puede comprender uno o más segmentos de ADNc, ADN genómico, ADN

sintético o mezclas de estos. Las bases de nucleótidos se indican en el presente documento mediante un código de una sola letra: adenina (A), guanina (G), timina (T), citosina (C), inosina (I) y uracilo (U). A menos que se limite específicamente, el término abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades de unión similares a las del ácido nucleico de referencia y se metabolizan de manera similar a los nucleótidos naturales. A menos que se indique lo contrario, una secuencia de ácido nucleico particular también abarca implícitamente variantes con modificaciones conservadoras de esta (p. ej., sustituciones de codones degenerados) y secuencias complementarias, así como la secuencia indicada explícitamente. Específicamente, las sustituciones de codones degenerados pueden lograrse mediante la generación de secuencias en las que la tercera posición de uno o más codones seleccionados (o todos ellos) se sustituye con residuos de base mixta y/o desoxiinosina.

Una «secuencia de nucleótidos» es un polímero de ADN o ARN que puede ser monocatenario o bicatenario, que contiene opcionalmente bases de nucleótidos de carácter sintético, no natural o alterado con capacidad de incorporación a polímeros de ADN o ARN. Las expresiones «ácido nucleico», «molécula de ácido nucleico», «fragmento de ácido nucleico», «secuencia o segmento de ácido nucleico» o «polinucleótido» se usan indistintamente y también pueden usarse indistintamente con un gen, ADNc, ADN y ARN codificados por un gen.

La invención abarca composiciones de ácido nucleico considerablemente purificado. En el contexto de la presente invención, una molécula de ADN o molécula de ARN «aislada» o «purificada» es una molécula de ADN o molécula de ARN que existe aparte de su entorno natural y, por lo tanto, no es un producto de la naturaleza. Una molécula de ADN o molécula de ARN aislada puede existir en una forma purificada o puede existir en un entorno no natural tal como, por ejemplo, una célula hospedadora transgénica. Por ejemplo, una molécula de ácido nucleico «aislada» o «purificada», o una parte con actividad biológica de esta, está considerablemente libre de otro material celular, o medio de cultivo cuando se produce mediante técnicas recombinantes, o considerablemente libre de precursores químicos u otras sustancias químicas cuando se sintetiza químicamente. En una modalidad, un ácido nucleico «aislado» está libre de secuencias que flanquean naturalmente el ácido nucleico (es decir, secuencias ubicadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo del cual deriva el ácido nucleico. La presente invención también abarca fragmentos

y variantes de las secuencias de nucleótidos descritas. Por «fragmento» o «parte» se entiende una longitud completa o menos que la longitud completa de la secuencia de nucleótidos.

«De origen natural», «natural» o «en estado natural» se utiliza para describir un objeto que se puede encontrar en la naturaleza en contraposición a ser producido artificialmente. Por ejemplo, una proteína o una secuencia de nucleótidos presente en un organismo (incluido un virus), que puede aislarse de una fuente en la naturaleza y que no ha sido modificada intencionalmente por una persona en el laboratorio, es de origen natural.

«Genoma» hace referencia al material genético completo de un organismo.

La «identidad», tal como se conoce en la técnica, es una relación entre dos o más secuencias de polipéptidos o dos o más secuencias de polinucleótidos, tal como se determina al comparar las secuencias. En la técnica, «identidad» también hace referencia al grado de relación de secuencias entre secuencias de polipéptidos o polinucleótidos, según sea el caso, tal como se determina según la coincidencia entre cadenas de esas secuencias. La «identidad» y «similitud» pueden calcularse fácilmente mediante métodos conocidos, lo cual incluye, de modo no taxativo, los descritos en (Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, Nueva York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, Nueva York, 1993; Computer Analysis of Sequence Data, Parte I, Griffin, A. M., y Griffin, H. G., eds., Humana Press, Nueva Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; y Sequence Analysis Primer, Gribskov, M. y Devereux, J., eds., M Stockton Press, Nueva York, 1991; y Carillo, H., y Lipman, D., SIAM J. Applied Math., 48: 1073 (1988). Los métodos preferidos para determinar la identidad están diseñados para proporcionar la mayor coincidencia entre las secuencias analizadas. Los métodos para determinar la identidad y la similitud están codificados en programas informáticos disponibles públicamente. Los métodos de programas informáticos preferidos para determinar la identidad y similitud entre dos secuencias incluyen, de modo no taxativo, el paquete de programas GCG (Devereux, J., et al. 1984), BLASTP, BLASTN y FASTA (Altschul, S. F., et al, 1990. El programa BLASTX está disponible públicamente en NCBI y otras fuentes (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, Md. 20894; Altschul, S., et al., 1990). El conocido algoritmo de Smith Waterman también puede utilizarse para determinar la identidad.

Las moléculas de ácido nucleico que codifican variantes de secuencias de aminoácidos

de una proteína se preparan mediante una variedad de métodos conocidos en la técnica. Estos métodos incluyen, de modo no taxativo, el aislamiento de una fuente natural (en el caso de variantes de secuencias de aminoácidos de origen natural) o la preparación mediante mutagénesis mediada por oligonucleótidos (o dirigida al sitio), mutagénesis por PCR y mutagénesis en casete de una variante preparada anteriormente o una versión no variante de la proteína.

Promotores

«Promotor» hace referencia a una secuencia de nucleótidos, habitualmente en dirección retrógrada (5') con respecto a su secuencia codificante, que controla la expresión de la secuencia codificante al proporcionar el reconocimiento de la ARN polimerasa y otros factores necesarios para una transcripción adecuada. «Promotor» incluye un promotor mínimo que es una secuencia de ADN corta compuesta por una caja TATA y otras secuencias que sirven para especificar el sitio de inicio de la transcripción, que recibe la adición de elementos reguladores para el control de la expresión. «Promotor» también hace referencia a una secuencia de nucleótidos que incluye un promotor mínimo más elementos reguladores que son capaces de controlar la expresión de una secuencia codificante o ARN funcional. Este tipo de secuencia promotora consiste en elementos en dirección retrógrada próximos y más distantes; estos últimos elementos a menudo se denominan «potenciadores». En consecuencia, un «potenciador» es una secuencia de ADN que puede estimular la actividad promotora y puede ser un elemento innato del promotor o un elemento heterólogo insertado para mejorar el nivel o la especificidad tisular de un promotor. Es capaz de funcionar en ambas orientaciones (normal o invertida) y es capaz de funcionar incluso cuando se mueve en dirección retrógrada o anterógrada con respecto al promotor. Tanto los potenciadores como otros elementos promotores en dirección retrógrada se unen a proteínas de unión al ADN específicas de secuencia que median sus efectos. Los promotores pueden derivar en su totalidad de un gen natural, o estar compuestos por diferentes elementos derivados de diferentes promotores que se encuentren en la naturaleza, o incluso estar compuestos por segmentos de ADN sintético. Un promotor también puede contener secuencias de ADN que participen de la unión de factores proteicos que controlen la eficacia del inicio de la transcripción en respuesta a condiciones fisiológicas o de desarrollo. En la patente de EE. UU. n.º 7 501 129, que se incorpora al presente documento mediante referencia, se proporciona una descripción general sobre los promotores.

La invención abarca composiciones de ácido nucleico considerablemente purificado. En el contexto de la presente invención, una molécula de ADN o molécula de ARN «aislada» o «purificada» es una molécula de ADN o molécula de ARN que existe aparte de su entorno natural y, por lo tanto, no es un producto de la naturaleza. Una molécula de ADN o molécula de ARN aislada puede existir en una forma purificada o puede existir en un entorno no natural tal como, por ejemplo, una célula hospedadora transgénica. Por ejemplo, una molécula de ácido nucleico «aislada» o «purificada», o una parte con actividad biológica de esta, está considerablemente libre de otro material celular, o medio de cultivo cuando se produce mediante técnicas recombinantes, o considerablemente libre de precursores químicos u otras sustancias químicas cuando se sintetiza químicamente. En una modalidad, un ácido nucleico «aislado» está libre de secuencias que flanquean naturalmente el ácido nucleico (es decir, secuencias ubicadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo del cual deriva el ácido nucleico. Por ejemplo, en diversas modalidades, la molécula de ácido nucleico aislada puede contener menos de alrededor de 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb o 0,1 kb de secuencias de nucleótidos que flanquean naturalmente la molécula de ácido nucleico en el ADN genómico de la célula de la que deriva el ácido nucleico. La presente invención también abarca fragmentos y variantes de las secuencias de nucleótidos descritas. Por «fragmento» o «parte» se entiende una longitud completa o menos que la longitud completa de la secuencia de nucleótidos.

«De origen natural», «natural» o «en estado natural» se utiliza para describir un objeto que se puede encontrar en la naturaleza en contraposición a ser producido artificialmente. Por ejemplo, una proteína o una secuencia de nucleótidos presente en un organismo (incluido un virus), que puede aislarse de una fuente en la naturaleza y que no ha sido modificada intencionalmente por una persona en el laboratorio, es de origen natural.

En determinadas modalidades, en la presente invención se proporcionan vectores y casetes de expresión que contienen los promotores descritos anteriormente. Se define que un «vector» incluye, entre otras cosas, cualquier vector viral, así como cualquier vector plasmídico, cosmídico, fágico o binario en forma circular o lineal bi o monocatenaria que puede o no ser autotransmisible o movilizable y que puede transformar al hospedador procariota o eucariota, ya sea mediante integración en el genoma celular o existir extracromosómicamente (p. ej., un plásmido replicante autónomo con un origen de replicación).

Los ácidos nucleicos que codifican composiciones terapéuticas pueden diseñarse en un vector mediante el uso de técnicas de ligadura convencionales, tales como las descritas en Sambrook y Russell, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY (2001). Por ejemplo, las ligaduras pueden lograrse en Tris-Cl a 20 mM, pH 7,5, MgCl₂ a 10 mM, DTT a 10 mM, BSA a 33 µg/mL, NaCl a 10 mM-50 mM y ATP a 40 µM, ADN ligasa T4 a 0,01-0,02 unidades (Weiss) a 0 °C (para la ligadura de «extremos pegajosos») o ATP a 1 mM, ADN ligasa T4 a 0,3-0,6 unidades a 14 °C (para la ligadura de «extremos romos»). Las ligaduras intermoleculares de «extremos pegajosos» habitualmente se realizan en concentraciones totales de ADN de 30-100 µg/mL (concentración total de extremo de 5-100 nM).

En determinadas modalidades, en la presente invención se proporciona un vector que contiene un casete de expresión que comprende un promotor enlazado de manera operativa a una secuencia objetivo (p. ej., una proteína de interés) para la producción de una vacuna. «Casete de expresión», tal como se usa en el presente documento, significa una secuencia de ácido nucleico capaz de dirigir la expresión de una secuencia de nucleótidos particular en una célula hospedadora adecuada, que incluye un promotor enlazado de manera operativa a la secuencia de nucleótidos de interés que puede estar enlazada de manera operativa a las señales de terminación. La región codificante habitualmente codifica un ARN funcional de interés, por ejemplo una molécula de ARNi. El casete de expresión que incluye la secuencia de nucleótidos de interés puede ser quimérico.

«Enlazado/a de manera operativa» hace referencia a la asociación de secuencias de ácido nucleico o aminoácidos en un solo fragmento molecular de forma tal que la función de una de las secuencias se ve afectada por otra. Por ejemplo, se dice que una secuencia de ADN reguladora está «enlazada de manera operativa a» o «asociada a» una secuencia de ADN que codifica un ARN o un polipéptido si las dos secuencias están situadas de forma tal que la secuencia de ADN reguladora afecte la expresión de la secuencia de ADN codificante (es decir, que la secuencia codificante o el ARN funcional está bajo el control de transcripción del promotor). Las secuencias codificantes pueden enlazarse de manera operativa a secuencias reguladoras en orientación codificante o no codificante. El ácido nucleico está «enlazado de manera operativa» cuando se coloca en una relación funcional con otra secuencia de ácido nucleico. Generalmente, «enlazado de manera operativa» significa que las secuencias de ADN que se enlazan son contiguas. Sin embargo, los potenciadores no tienen

que ser contiguos. La unión se logra mediante ligadura en sitios de restricción convenientes. Si esos sitios no existen, los adaptadores o enlazadores de oligonucleótidos sintéticos se utilizan de acuerdo con la práctica convencional. Además, se pueden enlazar entre sí múltiples copias de las enzimas codificadoras de ácido nucleico en el vector de expresión. Esos múltiples ácidos nucleicos pueden estar separados por enlazadores.

«Expresión» hace referencia a la transcripción y/o traducción de un gen endógeno o un transgén en las células. Por ejemplo, en el caso de las construcciones no codificantes, la expresión puede hacer referencia únicamente a la transcripción del ADN no codificante. Además, la expresión hace referencia a la transcripción y acumulación estable de ARN codificante (ARNm) o funcional. La expresión también puede hacer referencia a la producción de proteínas.

«Casete de expresión», tal como se usa en el presente documento, significa una secuencia de ADN capaz de dirigir la expresión de una secuencia de nucleótidos particular en una célula hospedadora adecuada, que comprende un promotor enlazado de manera operativa a la secuencia de nucleótidos de interés que se enlaza de manera operativa a las señales de terminación. También comprende normalmente secuencias necesarias para la traducción adecuada de la secuencia de nucleótidos. La región codificante habitualmente codifica una proteína de interés, pero también puede codificar un ARN funcional de interés, por ejemplo, un ARN no codificante o un ARN no traducido, en la dirección codificante o no codificante. El casete de expresión que comprende la secuencia de nucleótidos de interés puede ser quimérico, lo que significa que al menos uno de sus componentes es heterólogo con respecto al menos a uno de sus otros componentes. El casete de expresión también puede ser uno de origen, natural pero que se haya obtenido en una forma recombinante útil para la expresión heteróloga. Esos casetes de expresión comprenderán la región de inicio de la transcripción enlazada a una secuencia de nucleótidos de interés. Ese casete de expresión puede estar provisto de múltiples sitios de restricción para que la inserción del gen de interés esté bajo la regulación de transcripción de las regiones reguladoras. El casete de expresión puede contener además genes marcadores seleccionables.

En la presente descripción también se proporciona una célula de mamífero que contiene un vector descrito en el presente documento.

Proteínas

Los términos «proteína», «péptido» y «polipéptido» se usan indistintamente en el presente documento.

5

Una proteína de interés puede conjugarse con o enlazarse a otro péptido o a un polisacárido.

10

El término «aminoácido» incluye los residuos de los aminoácidos naturales (p. ej., Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Hyl, Hyp, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr y Val) en formas estereoisoméricas dextrógiras o levógiras, así como aminoácidos no naturales (p. ej., fosfoserina, fosfotreonina, fosfotirosina, hidroxiprolina y gamma-carboxiglutamato; ácido hipúrico, ácido octahidroindol-2-carboxílico, estatina, ácido 1,2,3,4-tetrahidroisoquinolin-3-carboxílico, penicilamina, ornitina, citrulina, alfa-metil-alanina, para-benzoilfenilalanina, fenilglicina, propargilglicina, sarcosina y tert-butilglicina). El término también comprende aminoácidos naturales y no naturales (estereoisómeros dextrógiros y levógiros) que llevan un grupo protector amino convencional (p. ej., acetilo o benciloxycarbonilo), así como aminoácidos naturales y no naturales protegidos en el extremo carboxi (p. ej., como éster o amida de bencilo, fenilo o alquilo(C₁-C₆); o como una α -metilbencilamida). Las personas expertas en la técnica conocen otros grupos protectores de amino y carboxi adecuados (véase, por ejemplo, Greene, T. W.; Wutz, P. G. M., Protecting Groups In Organic Synthesis; segunda edición, 1991, Nueva York, John Wiley & sons, Inc. , y los documentos allí citados). Un aminoácido puede enlazarse al resto de un compuesto a través del extremo carboxi, extremo amino o mediante cualquier otro punto de unión conveniente, tal como, por ejemplo, a través del azufre de la cisteína.

15

20

25

30

La invención abarca composiciones de proteínas aisladas o considerablemente purificadas. En el contexto de la presente invención, un polipéptido «aislado» o «purificado» es un polipéptido que existe separado de su entorno natural y, por lo tanto, no es un producto de la naturaleza. Un polipéptido puede existir en forma purificada o puede existir en un entorno no natural tal como, por ejemplo, una célula hospedadora transgénica. Por ejemplo, una proteína «aislada» o «purificada», o una parte con actividad biológica de esta, está considerablemente libre de otro material celular, o medio de cultivo cuando se produce mediante técnicas recombinantes, o considerablemente libre de precursores químicos u otras sustancias químicas cuando se sintetiza químicamente. Una proteína considerablemente libre

35

de material celular incluye preparaciones de proteína o polipéptido que tienen menos de alrededor de 30 %, 20 %, 10 %, 5 % (en peso seco) de proteína contaminante. Cuando la proteína de la invención, o una parte con actividad biológica de esta, se produce de forma recombinante, preferiblemente el medio de cultivo representa menos de alrededor de 30 %, 20 %, 10 % o 5 % (en peso seco) de precursores químicos o sustancias químicas que no sean proteínas de interés. La presente invención también abarca fragmentos y variantes de las proteínas descritas o proteínas de longitud parcial codificadas por estas. «Fragmento» o «parte» hace referencia a una longitud completa o menos que la longitud completa de la secuencia de aminoácidos de, un polipéptido o una proteína.

Una «variante» de una molécula es una secuencia considerablemente similar a la secuencia de la molécula natural.

«Aislado/a» significa alterado/a o extraído/a del estado natural. Por ejemplo, un ácido nucleico o un péptido presente naturalmente en un animal vivo no está «aislado», sino que el mismo ácido nucleico o péptido parcial o completamente separado de los materiales coexistentes de su estado natural está «aislado». Un ácido nucleico o una proteína aislados pueden existir en forma considerablemente purificada o pueden existir en un entorno no natural tal como, por ejemplo, una célula hospedadora. A menos que se especifique particularmente lo contrario en el presente documento, las proteínas, los complejos de viriones, los anticuerpos y otras moléculas biológicas que forman la materia de la presente invención se aíslan o se pueden aislar.

La expresión «identidad considerable» en el contexto de un péptido indica que un péptido comprende una secuencia que tiene al menos 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, o 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 % u 89 %, al menos 90 %, 91 %, 92 %, 93 % o 94 %, o 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia con respecto a una secuencia de referencia a través de una ventana de comparación especificada. La alineación óptima se realiza mediante el uso del algoritmo de alineación por homología de Needleman y Wunsch, J. Mol. Biol. 48:443 (1970). Una indicación de que dos secuencias peptídicas son considerablemente idénticas es que un péptido es inmunológicamente reactivo con anticuerpos generados contra el segundo péptido. Por lo tanto, un péptido es considerablemente idéntico a un segundo péptido, por ejemplo, cuando los dos péptidos difieren solamente por una sustitución conservadora.

Todas las publicaciones, patentes y solicitudes de patentes citadas en el presente documento se incorporan al presente documento mediante referencia. Si bien en la memoria descriptiva anterior se ha descrito la presente invención en relación con determinadas modalidades de esta, y se han establecido muchos detalles con fines ilustrativos, será evidente para las personas expertas en la técnica que la invención es pasible de modalidades adicionales y que algunos de los detalles descritos en el presente documento pueden variar considerablemente sin apartarse de los principios básicos de la invención.

El uso de las palabras «un», «una», «el», «la» y «o» y referencias similares en el contexto de la descripción de la invención se deben entender como abarcativas tanto del singular como del plural, salvo que se indique lo contrario en el presente documento o el contexto indique claramente lo contrario. Por lo tanto, por ejemplo, la referencia a «un polipéptido en cuestión» incluye múltiples polipéptidos de ese tipo y la referencia al «agente» incluye la referencia a uno o más agentes y equivalentes de estos conocidos por las personas expertas en la técnica, etc.

Las expresiones «que comprende/n», «que tiene/n», «que incluye/n» y «que contiene/n» deben interpretarse como expresiones abiertas (es decir, con el significado de «que incluye/n, de modo no taxativo») a menos que se indique lo contrario. La enumeración de los intervalos de valores en la presente tiene la intención meramente de servir como un método abreviado para referirse individualmente a cada valor separado que se encuentra dentro del intervalo, a menos que se indique lo contrario en la presente, y cada valor separado se incorpora a la memoria descriptiva como si se mencionara individualmente en la presente. Todos los métodos descritos en la presente se pueden realizar en cualquier orden adecuado a menos que se indique lo contrario en la presente o que el contexto lo contradiga claramente. El uso de cualquiera y todos los ejemplos, o lenguaje ilustrativo (*p. ej.*, "tal como") proporcionado en la presente, pretende simplemente ilustrar mejor la invención y no representa una limitación en el alcance de la invención a menos que se reivindique lo contrario. Ninguna expresión en la memoria descriptiva debe interpretarse como una indicación de que cualquier elemento no reivindicado sea esencial para la puesta en práctica de la invención.

En el presente documento se describen modalidades de la presente invención, incluido el mejor modo conocido por los inventores para llevar a cabo la invención. Las variaciones de esas modalidades pueden resultar evidentes para las personas expertas en la técnica al leer

la descripción anterior. Los inventores esperan que las personas expertas empleen esas variaciones según sea adecuado, y los inventores pretenden que la invención se ponga en práctica de forma diferente a como se describe específicamente en el presente documento. En consecuencia, la presente invención incluye todas las modificaciones y los equivalentes de la materia mencionada en las reivindicaciones anexadas según lo permitido por la ley aplicable. Además, cualquier combinación de los elementos descritos anteriormente en todas las variaciones posibles de estos está abarcada por la invención a menos que se indique lo contrario en el presente documento o que el contexto indique claramente lo contrario.

Con respecto a intervalos de valores, la invención abarca cada valor intermedio entre los límites superior e inferior del intervalo hasta al menos una décima parte de la unidad del límite inferior, a menos que el contexto indique claramente lo contrario. Además, la invención abarca cualquier otro valor intermedio indicado. Además, la invención también abarca intervalos que excluyan uno o ambos de los límites superior e inferior del intervalo, a menos que se excluyan específicamente del intervalo indicado.

Además, todos los números que expresan cantidades de ingredientes, condiciones de reacción, % de pureza, longitudes de polipéptidos y polinucleótidos, etc, utilizados en la memoria descriptiva y las reivindicaciones, están modificados por la expresión «alrededor de», a menos que se indique lo contrario. En consecuencia, los parámetros numéricos establecidos en la memoria descriptiva y las reivindicaciones son aproximaciones que pueden variar dependiendo de las propiedades deseadas de la presente invención. Como mínimo, y no como un intento de limitar la aplicación de la doctrina de equivalentes al alcance de las reivindicaciones, cada parámetro numérico debería al menos interpretarse a la luz de la cantidad de dígitos significativos informados, aplicando técnicas habituales de redondeo. No obstante, los valores numéricos establecidos en los ejemplos específicos se informan con la mayor precisión posible. Cualquier valor numérico, sin embargo, contiene inherentemente ciertos errores con respecto a la desviación estándar de su medición experimental.

A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que le otorgaría comúnmente una persona experta en la técnica a la que pertenece la presente invención. Una persona experta en la técnica también podrá reconocer muchos métodos y materiales similares o equivalentes a los descritos en el presente documento que se podrían utilizar para poner en práctica la

presente invención. Además, todas las publicaciones mencionadas en el presente documento se incorporan en su totalidad mediante referencia.

A continuación, la invención se ilustrará mediante los siguientes ejemplos no taxativos.

5

EJEMPLO 1

Cribado de cultivares para determinar la respuesta del cultivo de tejidos y el establecimiento de transformación genética en un cultivar de *Theobroma cacao* de alto rendimiento y resistente a enfermedades.

10

Resumen

Un protocolo de transformación altamente eficiente es un requisito previo para desarrollar cultivos modificados genéticamente y con edición de genoma. En *Theobroma cacao* se ha evaluado y mejorado un sistema de cultivo de tejidos que abarca el inicio del cultivo desde el material floral hasta la conversión de embriones en plantas. Se cribaron nueve cultivares para determinar su respuesta en cultivo de tejido y la susceptibilidad al suministro de ADN de transferencia de *Agrobacterium* tal como se mide a través de expresión transitoria. Estos factores clave se utilizaron para determinar la transformabilidad genética de diversos cultivares. Se seleccionó el cultivar INIAPG-038 de alto rendimiento y resistente a enfermedades para una transformación estable y el método se optimizó de manera adicional. Se produjeron múltiples eventos transgénicos mediante el uso de dos vectores que contenían genes de proteína fluorescente amarilla y de neomicina fosfotransferasa II. Se llevó a cabo una estrategia doble para mejorar los índices de la embriogénesis somática secundaria y el suministro de ADN-T para mejorar la frecuencia general de transformación. El uso de la cepa AGL1 de *Agrobacterium* y tejido de cotiledón derivado de embriones somáticos secundarios con un tamaño de entre 4 y 10 mm dio como resultado la mayor eficiencia de suministro de ADN-T. Además, el uso de concentraciones más altas de sales basales y sulfato cúprico en el medio aumentó cinco veces y cuatro veces la frecuencia de explantes que produjeron más de diez embriones durante la embriogénesis somática secundaria, respectivamente. En consecuencia, una combinación óptima de todos estos componentes dio como resultado una transformación exitosa de INIAPG-038 con una frecuencia del 3,7 % en el nivel de planta T₀. El injerto de vástagos transgénicos con raíces no desarrolladas en plántulas no transgénicas con raíces sanas ayudó a que las plántulas sobrevivieran y facilitó un rápido trasplante al suelo.

35

La estrategia presentada puede aplicarse para mejorar la respuesta del cultivo de tejidos y la frecuencia de transformación en otros cultivares de *Theobroma cacao*.

Introducción

Theobroma cacao L. es un cultivo de importancia económica que se cultiva predominantemente en África, aunque la especie fue domesticada originalmente en América del Sur, donde aún se mantiene una producción significativa. A pesar de que los granos de *Theobroma cacao* son la materia prima de una industria multimillonaria, las enfermedades han causado graves pérdidas tanto en el rendimiento de los cultivos como en la supervivencia de los árboles. Obstáculos tales como el largo período juvenil, los problemas con las poblaciones heterocigotas y heterogéneas, el largo ciclo de vida y las limitaciones de financiamiento de los programas convencionales de fitomejoramiento de *Theobroma cacao* hacen que rasgos deseados tales como el alto rendimiento, la resistencia a enfermedades y la calidad del grano sean difíciles de combinar en un solo cultivar. Sin embargo, tanto la edición del genoma como las estrategias transgénicas pueden utilizarse para mejorar y acelerar en gran medida el ritmo de desarrollo de rasgos.

La primera transformación estable para regenerar plantas de *Theobroma cacao* se informó en 2003 y desde entonces ha habido solamente unos pocos informes exitosos de la transformación de *TcChi1*, *TcLEC2* y *D5C* y *pD4e1* en *Theobroma cacao*. También se utilizaron expresiones transitorias de CRISPR/Cas9 para inactivar parcialmente *TcNPR3* y mejorar la respuesta de defensa en *Theobroma cacao*. La introducción del factor de transcripción *BABY BOOM* en *Theobroma cacao* aumentó significativamente la frecuencia de transformación, pero provocó un problema de regeneración con un fenotipo anormal. Estos experimentos mencionados anteriormente se realizaron en el cultivar adecuado PSU SCA-6. Antes de este estudio, todos los informes publicados en la literatura sobre plantas transgénicas recuperadas se referían únicamente al cultivar PSU SCA-6. Aunque PSU SCA-6 no es idéntico a SCA-6, el SCA-6 original fue recolectado de la naturaleza en la Amazonía peruana y tiene cierta resistencia natural a las enfermedades. Sin embargo, este cultivar no es muy productivo. Para que la transformación genética se utilice de manera más eficaz, se debe establecer la capacidad de transformar muchos cultivares de *Theobroma cacao*, incluidos aquellos menos susceptibles a la transformación. Lo ideal sería desarrollar un único método de transformación independiente del genotipo, pero al igual que la embriogénesis somática, que ha tenido un

gran éxito en todo el reino vegetal, todavía existe un alto grado de variación de un genotipo a otro, por lo que es necesaria la personalización del protocolo. La optimización de las sales basales, la duración de los tratamientos hormonales, la calidad y el tipo de explante son formas eficaces de mejorar la respuesta de cultivo de cultivares recalcitrantes.

Se utilizan varios tipos de explantes en el cultivo de tejidos de *Theobroma cacao* y el proceso de transformación, que incluyen pétalos y estaminodios de flores inmaduras sin abrir, tejido de cotiledones de embriones somáticos y embriones somáticos intactos. Los pétalos y estaminodios son capaces de realizar una embriogénesis somática primaria y son el punto de partida para todos los procesos posteriores. El tejido de cotiledón derivado de embriones somáticos es un tipo de tejido alternativo y eficiente en respuestas embriogénicas somáticas secundarias, terciarias y cuaternarias. Los embriones somáticos maduros se convierten en plántulas y pueden trasplantarse al suelo. Hasta ahora, todos los métodos de transformación anteriores utilizaban una estrategia embriogénica mediante el uso de tejido de cotiledones de embriones somáticos. Las ventajas de esta estrategia incluyen las metodologías descritas anteriormente, el origen unicelular predominante de los embriones y la gran cantidad de embriones que el tejido de cotiledón es capaz de producir. De manera alternativa, no se exploró a fondo una estrategia organogénica debido a la falta de métodos previamente descritos en *Theobroma cacao*, la aparente dificultad de establecer un nuevo sistema de cultivo de tejidos y la naturaleza no clonal de la semilla que daría como resultado la pérdida del cultivar original si se utiliza como material donante organogénico.

En este estudio, informamos un protocolo exitoso de transformación y regeneración mediada por *Agrobacterium* para INIAPG-038, mediante el uso de embriogénesis somática mejorada e infección por *Agrobacterium*. INIAPG-038 es una línea de fitomejoramiento desarrollada por Mars Wrigley, Incorporated en colaboración con el Departamento de Agricultura de los Estados Unidos (USDA, por su sigla en inglés) y el Instituto Nacional de Investigaciones Agropecuarias (INIAP). Es un cultivar altamente productivo que tiene tolerancia a enfermedades fúngicas tales como la enfermedad de la vaina helada y la enfermedad de la escoba de bruja. También presentamos un método para evaluar qué tan pasible de cultivo y transformación es un cultivar, además de proporcionar datos sobre los cultivares cribados.

Materiales y métodos

Material vegetal

Se utilizaron nueve cultivares de *Theobroma cacao*, SCA-6, CCN-51, INIAPG-038, INIAPT-374, INIAPG-152, INIAPM-053, INIAPG-029, INIAPT-484 y Matina 1-6, para experimentos de respuesta al cultivo de tejidos. Las flores inmaduras y sin abrir se originaron en la Estación de Investigación de Horticultura Subtropical del Servicio de Investigación Agrícola del USDA en Miami, FL, donde se cosecharon en el invernadero por la mañana y luego se enviaron durante la noche a las Instalaciones de Transformación y Genómica de Plantas del Innovative Genomics Institute (IGI), Berkeley, CA. Los cultivares se eligieron principalmente en función de su constante disponibilidad de flores. Las flores se sellaron dentro de un tubo de centrífuga cónico Falcon de 50 mL que contenía una solución de ditiotreitól a 5 mM y se enviaron en una caja aislada de Styrofoam junto con bolsas de hielo para mantener una temperatura de aproximadamente 4 °C.

Inicio de la embriogénesis primaria mediante el uso de material floral y mejoras

Se usaron flores inmaduras para iniciar embriones somáticos primarios de acuerdo con los protocolos anteriores (Li Z, Traore A, Maximova SN, Gultinan MJ (1998) Somatic embryogenesis and plant regeneration from floral explants of cacao (*Theobroma cacao* L.) using thidiazuron. In Vitro Cell Dev Biol – Plant 34:293–299; Garcia C, Corrêa F, Findley S, Almeida AA, Costa M, Motamayor JC, Ray S, Marelli JP (2016) Optimization of somatic embryogenesis procedure for commercial clones of *Theobroma cacao* L. Afr J Biotechnol 15:1936–1951 («García et al. 2016»)) con algunas modificaciones. La composición del medio seguida fue la de García et al. 2016. Se observaron explantes de 9 cultivares después de 3 a 4 meses de cultivo para determinar los índices con los que tanto los explantes de estaminodios como los de pétalos formaron embriones; a continuación, sigue un breve resumen del método. Para iniciar el material floral, se esteriliza la superficie mediante el uso de una solución de lejía al 20 % (v/v) (concentración final de hipoclorito de sodio al 1,05 %) que contiene Tween 20 al 0,002 % y luego se diseca, los pétalos y estaminodios se colocan en el medio de crecimiento de callos primarios (PCG, por su sigla en inglés) durante 2 sem., luego se transfiere a medio de crecimiento de callos secundarios (SCG, por su sigla en inglés) (García et al. 2016) durante dos sem., tras lo cual se transfiere a ED4 (medio de desarrollo de embriones que contiene sacarosa al 4 %, García et al. 2016) durante dos sem. y luego a ED3 (medio de crecimiento

de embriones que contiene sacarosa al 3 %, García et al. 2016) donde permanece hasta que se hayan formado los embriones y se hayan realizado las observaciones. La cantidad total de estaminodios y pétalos utilizados para cada cultivar, así como la cantidad de réplicas para cada cultivar variaron, tal como se indicará a continuación: SCA-6 con 1879 explantes totales en 2 réplicas, CCN-51 con 985 explantes totales en 6 réplicas, INIAPG-038 con 716 explantes totales en 7 réplicas, INIAPT-374 con 623 explantes totales en 4 réplicas, INIAPG-152 con 395 explantes totales en 4 réplicas, INIAPM-053 con 295 explantes totales en 4 réplicas, INIAPG-029 con 260 explantes totales en 4 réplicas, INIAPT-484 con 281 explantes totales en 4 réplicas y Matina 1-6 con 645 explantes totales en 4 réplicas.

Inicio de la embriogénesis somática secundaria mediante el uso de embriones primarios

Para inducir la embriogénesis somática secundaria, se extirparon cotiledones de los embriones somáticos primarios y se disecaron en segmentos de 4 a 10 mm. Estos explantes fueron manipulados según el método descrito por García et al. 2016, aunque el medio SCG utilizado fue de este último. Se recolectaron embriones somáticos secundarios de estos cultivos mientras estaban en ED3 y se usaron para la conversión en plántulas y la transformación con *Agrobacterium*. Inicialmente, se cribaron cuatro cultivares, INIAPG-038, CCN-51, SCA-6 e INIAPT-374, para determinar los índices con los que los explantes de embriones somáticos primarios formaron embriones somáticos secundarios. Los embriones producidos a partir de estos explantes se contaron aproximadamente a las 10 sem. y cada explante se clasificó en una de cuatro categorías: no se produjeron embriones, se produjeron de 1 a 4 embriones, se produjeron de 5 a 9 embriones y se produjeron ≥ 10 embriones. La cantidad total de explantes utilizados para cada cultivar, así como la cantidad de réplicas para cada cultivar, variaron tal como se indicará a continuación: INIAG-038 con 2698 explantes en 9 réplicas, CCN-51 con 791 explantes en 7 réplicas, SCA-6 con 2514 explantes en 8 réplicas e INIAT-374 con 752 explantes en 5 réplicas.

Conversión de embriones y regeneración de plántulas

Los embriones somáticos permanecieron en medio ED3 hasta que alcanzaron aproximadamente entre 15 y 20 mm de longitud, momento en el cual estaban listos para su conversión en plántulas. El proceso de conversión comenzó cuando estos embriones somáticos se colocaron en medio ED6 (desarrollo de embriones que contiene sacarosa al 6 %,

García et al. 2016) y se expusieron a bajos niveles de luz de diodos emisores de luz (LED, por su sigla en inglés), de 10 a 30 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Estos embriones se incubaron a 27 °C y luego las placas de Petri se envolvieron con cinta 3M Micropore (3M, St. Paul, MN) para permitir el intercambio de gases. Después de dos sem. en ED6, los embriones se transfirieron a un medio ligero de desarrollo embrionario (EDL, por su sigla en inglés) y luego se mantuvieron en EDL con transferencias cada dos semanas. A medida que los cultivos se desarrollaron, fueron expuestos lentamente a intensidades de luz LED más altas, de 30 a 60 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Se midieron y compararon los índices de conversión de diversos cultivares. Un embrión se consideró convertido si producía un brote con un meristemo y una raíz de al menos 1 cm de longitud. Cuando las plántulas alcanzaron 2,5 cm de altura, se transfirieron a un Phytatray™ II (Sigma-Aldrich, San Luis, MO) que contenía 100 mL de medio y se sellaron con cinta 3M Micropore. Los cultivares evaluados incluyen INIAPT-374, CCN-51 e INIAPG-038. La cantidad total de embriones utilizados para cada cultivar, así como la cantidad de réplicas para cada cultivar, variaron tal como se indicará a continuación: T-374 con 256 embriones en 4 réplicas, CCN-51 con 227 embriones en 8 réplicas e INIAPG-038 con 868 embriones en 9 réplicas.

Prueba de eficiencia del suministro de ADN de transferencia (ADN-T)

Se analizaron cuatro cultivares, INIAPG-038, SCA-6, CCN-51 e INIAPT-374 para determinar la expresión de proteínas fluorescentes (FPE, por su sigla en inglés) de carácter transitorio después de la transformación. Se utilizaron dos vectores binarios diferentes, pDDNPTYFP-1 y pDDNPTYFP-2, en experimentos de transformación de *Theobroma cacao* (Fig. 1A-1B); pDDNPTYFP-1 contiene una fusión de traducción de neomicina fosfotransferasa II (*nptII*)-proteína fluorescente amarilla (YFP) GlyLink impulsada por el promotor 35S (CaMV35S) del virus del mosaico de la coliflor (Fig. 1A), al tiempo que pDDNPTYFP-2 contiene dos casetes de genes separados, *nptII* impulsado por el promotor CaMV35S potenciado e YFP que el promotor Nos/potenciador TMV Ω impulsa (Fig. 1B). Ambos vectores se construyeron en la estructura principal de pCAMBIA2300. Se seleccionaron embriones secundarios de tamaño, color y morfología similares de cada cultivar y se infectaron mediante el uso de los métodos descritos más adelante utilizando la cepa AGL1 de *Agrobacterium*. Los explantes se evaluaron en función de la FPE 7 d después de la infección por *Agrobacterium*. Cada explante fue examinado y clasificado en función del porcentaje de su superficie que expresa YFP. Hubo tres categorías en total: de 0 a 5 % de alcance de FPE, de 6 a 20 % de alcance de FPE y >20 % de alcance de FPE. El tratamiento

con mayor porcentaje de explantes con >20 % de alcance de FPE fue seleccionado como el mejor, sin embargo, a veces fue útil considerar la distribución en todas las clasificaciones. La cantidad total de explantes utilizados para cada cultivar, así como la cantidad de réplicas para cada cultivar, variaron tal como se indicará a continuación: INIAPG-038 con 2025 explantes en 16 réplicas, CCN-51 con 472 explantes en 6 réplicas, SCA-6 con 2905 explantes en diez réplicas e INIAPT-374 con 449 explantes en 6 réplicas.

Transformación y mejoras de cultivos de tejidos en INIAPG-038

Se analizó la eficacia del suministro de ADN-T en cuatro cepas de *Agrobacterium*, LBA4404, AGL1, EHA105 y GV3101, con pDDNPTYFP-1. Se realizaron tres réplicas separadas y se compararon las 4 cepas, el total de explantes utilizados para cada tratamiento de cada réplica osciló entre 130 y 200 y todo el material provino de INIAPG-038. Todo el tejido de cotiledón extirpado de los embriones somáticos secundarios se recolectó en una única placa de Petri y se distribuyó de manera aleatoria antes de separarlo y tratarlo con cuatro cepas diferentes de *Agrobacterium*. Antes de la infección, cada cepa de *Agrobacterium* se suspendió en el medio de infección y luego se diluyó para que cada tratamiento con *Agrobacterium* tuviera la misma densidad óptica con una longitud de onda de 600 nm ($DO_{600\text{ nm}}$). Estos experimentos se evaluaron en función del alcance de FPE 1 sem. después de la infección mediante el uso del método descrito anteriormente.

En INIAPG-038 también se examinó el efecto del tamaño de los embriones somáticos secundarios sobre el suministro de ADN-T. Se compararon embriones somáticos con un tamaño de entre 4 y 10 mm con embriones somáticos de 10 a 20 mm. La cantidad total de explantes utilizados y la cantidad de réplicas para cada tratamiento son las siguientes: para el tamaño de embriones de 4 a 10 mm se usaron 1007 explantes en 9 réplicas, y para el tamaño de embriones de 10 a 20 mm se usaron 714 explantes en 5 réplicas. Los explantes se evaluaron con el método descrito anteriormente.

Se examinó el efecto de la concentración de sal basal del medio de plantas leñosas (WPM) duplicada, así como del WPM con ajuste a 5,0 μM de CuSO_4 en SCG, para determinar su efecto sobre la producción secundaria de embriones en INIAPG-038. Los tratamientos, la cantidad total de explantes y réplicas son los siguientes: medio SCG 1x estándar evaluado con 198 explantes en 3 réplicas, SCG 1x con CuSO_4 adicional evaluado con 327 explantes en 5 réplicas y SCG 2x con doble concentración de sales basales de WPM evaluado con

219 explantes en 5 réplicas. Los explantes se evaluaron con el método descrito anteriormente.

Transformación estable de INIAPG-038

Infección, cocultivo, embriogénesis y conversión de Agrobacterium

Las transformaciones se realizaron mediante el uso de tejido de cotiledón derivado de embriones somáticos de 4 a 10 mm de longitud, de color blanco o tostado y con morfología normal. El material donante utilizado para el proceso de transformación fueron embriones somáticos secundarios, convirtiendo la transformación en un proceso embriogénico somático terciario. Para preparar los explantes, se escindió tejido de cotiledón de los embriones secundarios previamente seleccionados y luego se disecó en segmentos de 4 a 10 mm. La etapa de infección se realizó mediante la suspensión de los explantes en entre 10 y 20 mL de medio líquido que albergaba *Agrobacterium* contenido dentro de un tubo de ensayo sellado de 50 mL. El medio de infección consistió en sal basal de MS y vitaminas con acetosiringona a 200 μ M, 1 mg de L^{-1} ácido indolacético y 2 mg de L^{-1} zeatina. Se suspendió exhaustivamente *Agrobacterium* en el medio de infección líquido y se midió la $DO_{600\text{ nm}}$ antes de su uso y se ajustó a 0,400. Después de 1 a 2 h de agitación, los explantes de cotiledones infectados con *Agrobacterium* se transfirieron a un medio de cocultivo basado en SCG y se incubaron a 21 °C durante 92 ± 8 h antes de la remoción mediante lavado de *Agrobacterium*. Para remover *Agrobacterium*, los explantes se transfirieron del medio de cocultivo a un tubo de centrifuga cónico Falcon de 50 mL que contenía 5 mL de medio SCG líquido y se agitaron mediante el uso de una pipeta de 1 mL durante 1 min, luego se decantó el medio SCG líquido y se añadieron 20 mL adicionales de medio SCG líquido y el tubo se agitó durante 1 h mediante el uso de un agitador orbital a 75 rpm. Luego, el tejido del cotiledón se transfirió a un medio SCG modificado que contenía 200 mg de L^{-1} cefotaxima y 200 mg de L^{-1} timentina. El tejido permaneció en este antibiótico que contenía SCG durante entre 2 y 4 sem. antes de transferirlo al medio ED4 que contenía los mismos antibióticos durante otras 2 sem., después el tejido se mantuvo en ED3 con antibióticos. Los antibióticos permanecieron en el medio siempre que hubiera explantes que habían estado expuestos directamente a *Agrobacterium*. Sin embargo, más adelante en el proceso, durante la conversión de embriones somáticos cuaternarios, se extrajeron los antibióticos del medio. La primera, segunda y tercera transformación se llevaron a cabo sin el uso de selección química y los embriones transgénicos se identificaron mediante selección con marcadores visuales. La cuarta transformación incluyó el uso de 100 mg de L^{-1}

de kanamicina para la selección química. Después de dos sem. de reposo en SCG que contenía timentina y cefotaxima, pero sin kanamicina, los explantes se trasladaron a ED4 que contenía antibióticos y kanamicina. Esta concentración de kanamicina se mantuvo durante 1 mes y luego se aumentó hasta 150 mg de L⁻¹ mientras los explantes estaban en ED3.

El tejido infectado con *Agrobacterium* se sometió a embriogénesis somática indirecta y se observó mediante el uso de un microscopio fluorescente varias veces entre uno y cuatro meses durante la formación de los embriones. Después de la identificación de un evento transgénico mediante el uso del microscopio fluorescente, los embriones somáticos terciarios (TSE, por su sigla en inglés) transgénicos se extrajeron del resto del tejido no transgénico y se aislaron en ED3. Para generar múltiples embriones somáticos cuaternarios (QSE, por su sigla en inglés) a partir de un único TSE transgénico, se utilizó una etapa adicional de embriogénesis somática. Esta etapa adicional de embriogénesis somática sigue el mismo método para generar embriones somáticos secundarios, los cotiledones se disecan y se colocan en SCG durante 2 sem., luego en ED4 durante 2 sem. y luego se mantienen en ED3. La conversión de QSE y la regeneración de plántulas se llevaron a cabo tal como se describió anteriormente.

Análisis de PCR

Se utilizó un protocolo de extracción de ADN de CTAB para extraer ADN genómico de materiales foliares de plantas INIAPG-038 no transgénicas e INIAPG-038 transgénicas. Para evaluar la presencia de *nptII* en plantas transgénicas, se usó el conjunto de cebadores, NPTII 3F (5'-CAAGATGGATTGCACGCAGGTT-3') y NPTII 4R (5'-TAGAAGGCGATGCGCTGCGAAT-3'), mientras que la presencia de *eyfp* se determinó mediante el uso del conjunto de cebadores EYFP 3F (5'-TAAACGGCCACAAGTTCAGCG-3') y EYFP 4R (5'-AGGACCATGTGATCGCGCTTC-3'). Para evaluar la posible presencia de contaminación por *Agrobacterium* en las muestras de ADN extraídas o un evento de lectura, también se analizaron tres genes en la estructura principal de *Agrobacterium*. La presencia de pVS1 se evaluó mediante el uso del conjunto de cebadores PVS1 1F (5'-ATGAAGGTTATCGCTGTACT-3') y PVS1 2R (5'-CTGATTCAAGAACGGTTGTG-3'), mientras que la presencia de *npt1* se evaluó mediante el uso del conjunto de cebadores NPT1 1F (5'-CTCCTGCTAAGGTATATAAGC-3') y NPT1 2R (5'-AATCAGGCTTGATCCCCAGT-3'). Para cada mezcla de reacción de PCR, se usaron los siguientes reactivos: 25 µL de DreamTaq PCR Master Mix (2X) (Thermo Fisher Scientific, Grand Island, NY), 1,0 µL de cebador directo a

10 μM , 1,0 μL de cebador inverso a 10 μM , 21 μL de H_2O , y 2,0 μL de ADN genómico a aproximadamente 100 ng de μL^{-1} , para un volumen total de 50 μL . La PCR se realizó con la siguiente programación del termociclador: mantener a 95 °C durante 3 min, 16 ciclos de 94 °C durante 30 seg, 58 °C durante 30 seg (-0,5 °C/ciclo), 68 °C durante 1 min, 50 ciclos de 94 °C durante 30 seg, 54 °C durante 30 seg, 69 °C durante 1 min; y luego elongación a 68 °C durante 7 min, y permaneció a 4 °C por infinito. Para cada reacción de PCR, se cargaron 20 μL en un gel de agarosa al 0,8 % para electroforesis.

Visualización fluorescente

Se visualizaron imágenes fluorescentes de embriones y plántulas de eventos transgénicos de INIAPG-038 con un estereomicroscopio fluorescente Leica M165 FC, equipado con Leica DFC7000 T (JH Technologies, Fremont, CA); mediante el uso de dos filtros microscópicos, campo claro y ET YFP con excitación de 514 nm y emisión de 527 nm. El microscopio está vinculado a un software de imágenes de cámara, Leica Application Suite versión 4.9, que se utilizó para capturar las imágenes fluorescentes. La detección de la actividad fluorescente se midió con un aumento de 30x.

Trasplante al suelo y aclimatación

Después de la conversión de los embriones, mientras las plántulas miden de 10 a 13 cm de altura y tienen de seis a diez hojas, se transfirieron al suelo. Las plántulas se transfirieron a un suelo con buen drenaje, mezclando partes iguales de mezcla para macetas Sun Gro™ (Sun Gro Horticulture Distribution Inc., Agawam, MA) y se descubrió que la perlita funciona bien. Al trasplantar, el agar se eliminó manualmente y sumergiendo las raíces en agua. Luego, las plántulas se plantaron en macetas cuadradas de 4 pulgadas con una mezcla de tierra previamente humedecida y se colocaron inmediatamente en una bandeja de agua 1020 (Greenhouse Megastore, West Sacramento, CA) y luego se cubrieron con un domo de humedad ventilado de 7 pulgadas (Greenhouse Megastore, West Sacramento, CA) con todos los respiraderos cerrados. Encima del domo de humedad se colocaron cortinas para reducir la intensidad de la luz a 60 $\mu\text{mol m}^{-2} \text{s}^{-1}$. La bandeja, el domo de humedad, las sombras y las plantas se guardaron dentro de una cámara de crecimiento de plantas Conviron con los parámetros que se describirán más adelante. En el transcurso de una semana, se abrieron las rejillas de ventilación del domo de humedad y se quitaron las cortinas. Después de este proceso de aclimatación, las plantas pudieron sobrevivir en la cámara de crecimiento en un

entorno de 27 °C, de 60 a 70 % de HR y de 120 a 180 $\mu\text{mol m}^{-2} \text{s}^{-1}$ de intensidad de luz.

Injerto de vástagos transgénicos sobre portainjertos germinados in vitro

El injerto en forma de cuña se realizó de la siguiente manera. Se esterilizó la superficie de vainas de *Theobroma cacao* de un grupo Amelonado no especificado (n.º de registro en GRIN PI 668412) mediante el uso de una solución de lejía al 10 % (v/v) (concentración final de hipoclorito de sodio al 0,525 %) durante 15 min, luego la vaina se cortó y se quitaron las semillas. Luego se quitó la pulpa espesa de las semillas. Las semillas se esterilizaron de manera adicional en la superficie mediante el uso de una solución de lejía al 10 % (v/v) (concentración final de hipoclorito de sodio al 0,525 %) durante 15 min. Luego, las semillas esterilizadas se remojaron en agua esterilizada durante 72 h. Posteriormente, estas semillas previamente germinadas se plantaron en pelets de turba esterilizados en autoclave (Jiffy Products of America Inc. Lorain, OH) remojados en un medio EDL líquido. Las plántulas se mantuvieron en Phytatrays[™] y se cultivaron de manera similar a las plántulas derivadas de cultivos de tejidos.

Cuando estaban disponibles tanto un vástago transgénico con raíces pobres como un portainjerto *in vitro* no transgénico con raíces fuertes y sanas, el injerto se realizó mediante el uso de un corte con forma de «V» similar a un injerto con forma de silla de montar o en hendidura en la plántula y dos cortes diagonales que exponían la capa de cámbium del vástago transgénico. Luego, el vástago y el portainjerto se combinaron y aseguraron mediante el uso de un clip de papel y un segmento de 1 cm de una pajita de polipropileno cortada lateralmente para parecerse a una «C». Todos los cortes se realizaron y aseguraron mediante el uso de herramientas y materiales esterilizados. Luego, los injertos se sellaron en Phytatrays[™] con cinta microporosa y periódicamente se añadió medio EDL líquido para mantener la humedad adecuada. Una vez fusionados el vástago y el portainjerto, se trasplantaron al suelo mediante el uso del método descrito anteriormente.

Resultados y descripción

Para establecer un protocolo de transformación exitoso y altamente eficiente, se ha evaluado y mejorado en *Theobroma cacao* un sistema de cultivo de tejidos que abarca desde el inicio del material floral hasta la regeneración de las plántulas en el suelo. Se criaron múltiples cultivares para determinar su respuesta a (1) la inducción de embriones primarios a

partir de material floral, (2) la producción de embriones secundarios a partir de tejido de cotiledón derivado de embriones somáticos primarios, (3) la conversión de embriones en plántulas y (4) la eficiencia del suministro de ADN-T mediante *Agrobacterium*. Estos cuatro factores se utilizaron para determinar qué tan pasibles serían diversos cultivares a la transformación genética.

Respuesta al cultivo de tejidos y mejoras de INIAPG-038

Se utilizaron pétalos y estaminodios para evaluar inicialmente la embriogénesis somática en nueve cultivares de *Theobroma cacao*. Los explantes de estaminodios generalmente formaron embriones con mayor frecuencia que los explantes de pétalos en todos los cultivares excepto SCA-6, donde fueron iguales. (Fig. 2). El cultivar con mayor porcentaje de explantes florales que formaron embriones fue SCA-6 con 44 % de pétalos y 44 % de estaminodios que formaron embriones (Fig. 2). Los siguientes cultivares con mejor desempeño fueron CCN-51 con 6 % de pétalos y 29 % de estaminodios, e INIAPT-374 con 4 % de pétalos y 21 % de estaminodios e INIAPG-038 con 7 % de pétalos y 15 % de estaminodios que produjeron embriones con sus respectivos índices. El resto de los cultivares tuvo menos de un 10 % de explantes florales que produjeron embriones; INIAPG-152 e INIAPM-053 tuvieron un desempeño ligeramente mejor que los INIAPT-484, INIAPG-029 y Matina 1-6 más recalcitrantes.

Los cuatro cultivares, SCA-6, CCN-51, INIAPT-374 e INIAPG-038, que mostraron la mejor formación de embriones somáticos primarios (Fig. 2), se evaluaron de manera adicional para determinar su producción de embriones somáticos secundarios (Fig. 3A). Se observó que INIAPG-038 tuvo el mayor porcentaje de explantes al producir más de 10 embriones con un 6,2 %. También tuvo la mayor cantidad de explantes que produjeron entre 5 y 9 embriones con un 12,6 %. SCA-6 y CCN-51 tuvieron una producción secundaria de embriones similar a la de INIAPG-038, mientras que INIAPT-374 tuvo la producción más baja de embriones secundarios con un 1,0 % y un 2,8 % de explantes que formaron más de 10 embriones y explantes que formaron entre 5 y 9 embriones, respectivamente. En total, se evaluaron tres cultivares para determinar el índice con el que los embriones se regenerarían en plántulas (Fig. 3B). INIAPT-374 tuvo índices de regeneración de embriones del 33 %. Los embriones de CCN-51 e INIAPG-038 se regeneraron en plántulas con un índice del 28 % y 22 %, respectivamente. Sin embargo, no hubo diferencias significativas entre estos 3 cultivares.

Con respecto a los experimentos realizados para mejorar la embriogénesis somática secundaria en INIAPG-038, se observó que, mediante el uso de una concentración de 2 veces de sales basales de WPM o WPM con ajuste a 5,0 μM de CuSO_4 en SCG tuvo un mejor desempeño que el medio SCG estándar que contenía sales basales de WPM de 1 vez (Fig. 4). El SCG estándar tuvo un 2,5 % de explantes que formaron más de 10 embriones y un 12,0 % que formaron entre 5 y 9 embriones, mientras que el medio basal WPM de 2 veces dio como resultado un 12,3 % y un 16,5 %, respectivamente. El medio basal WPM de 2 veces tuvo un porcentaje significativamente mayor de explantes que formaron más de 10 embriones producidos, en comparación con el medio SCG estándar. El medio SCG estándar, medio basado en WPM, contiene CuSO_4 a 1,0 μM , 10 veces más que el medio MS. Un nivel de cobre cinco veces mayor (5,0 μM) en medio SCG dio como resultado que el 10,9 % de los explantes produjeran más de 10 embriones y el 23,4 % de los explantes formaran de 5 a 9 embriones en INIAPG-038 (Fig. 4); tuvo un porcentaje significativamente mayor de explantes que formaron más de 10 embriones producidos y un porcentaje significativamente menor de explantes que no formaron embriones producidos, en comparación con el medio SCG estándar. Esto es coherente con la conclusión anterior de que la calidad de los callos, el crecimiento de los callos y la regenerabilidad puede mejorarse al aumentar el nivel del micronutriente cobre en la cebada y la avena.

Mejoras y optimizaciones en la eficiencia del suministro de ADN-T

Los experimentos para la prueba de eficiencia de suministro de ADN-T se realizaron mediante el uso de pDDNPTYFP-1 que contenía *nptII::eyfp* en los 4 cultivares con buena respuesta al cultivo de tejidos: INIAPG-038, CCN-51, SCA-6 y INIAPT-374 (Fig. 5). De ellos, INIAPG-038 y CCN-51 tuvieron eficiencia de suministro de ADN-T con 16 % y 14 % de los explantes con >20 % de alcance de FPE, respectivamente. Los otros 2 cultivares, SCA-6 con 9 %, e INAPT-374 con 6 % de explantes entraron en la categoría de >20 % de alcance de FPE.

De las cuatro cepas de *Agrobacterium* evaluadas, AGL1 y EHA105 produjeron el alcance de FPE de carácter transitorio más alto en INIAPG-038. Tanto LBA4404 como GV3101 tuvieron un desempeño deficiente en cuanto a la eficiencia de suministro de ADN-T (Fig. 6). La cepa AGL1 de *Agrobacterium* produjo un promedio de 13 % de explantes clasificados en la categoría de >20 % de FPE, mientras que EHA105 produjo un promedio de 12 %. GV3101

produjo un promedio de 4 % de los explantes clasificados en la categoría de >20 % FPE, mientras que LBA4404 produjo un promedio de 3 %. La cepa AGL1 de *Agrobacterium* que produjo una mayor frecuencia de transformación se observó previamente en la transformación estable de un maíz de élite endogámico (Cho et al. 2014). Sin embargo, otras especies de plantas o explantes favorecieron diferentes cepas de *Agrobacterium* tales como LBA4404 que se usó en la transformación del tabaco y GV3101 que se usó en la transformación del tomate.

Con respecto al tipo de explante INIAPG-038 utilizado para la transformación de *Theobroma cacao*, observamos que el tejido de cotiledón derivado de embriones somáticos primarios o secundarios tuvo la FPE transitoria más alta, mientras que los explantes de hipocótilo y radícula mostraron muy poca FPE transitoria y producción de embriones después de la infección por *Agrobacterium* (datos no mostrados). Anteriormente, se utilizaban tejidos de cotiledones derivados de embriones maduros de PSU SCA-6 para el objetivo de transformación. Nuestro estudio demostró que el tamaño de los embriones de INIAPG-038 también afectó la FPE transitoria después de la transformación (Fig. 7). Los tejidos de cotiledones derivados de embriones somáticos inmaduros con un tamaño de entre 4 y 10 mm superaron a los derivados de embriones maduros con un tamaño de entre 10 y 20 mm. Se observó que los embriones en el intervalo de 4 a 10 mm produjeron explantes con un alcance de FPE >20 % con un índice del 19 %, mientras que los embriones de 10 a 20 mm solo produjeron explantes con un alcance de FPE >20 % con un índice del 6 %.

Transformación estable de INIAPG-038

Se seleccionó INIAPG-038 para mejorar de manera adicional la respuesta del cultivo de tejidos, la eficiencia del suministro de ADN-T y, en última instancia, para una transformación estable porque tuvo el mejor desempeño tanto en cuanto a la embriogénesis somática secundaria como al suministro de ADN-T. Aunque de los 4 mejores cultivares del cribado de embriogénesis somática primaria tuvo un desempeño menos eficiente en la embriogénesis somática primaria (Fig. 2), este paso no es tan crucial para la transformación y no se considera un posible cuello de botella. Además, INIAPG-038 es un cultivar altamente productivo con tolerancia a la enfermedad de las vainas heladas y a la enfermedad de la escoba de bruja. Se realizaron transformaciones previas en PSU SCA-6 porque es un cultivar susceptible con potencial de transformación genética y embriogénesis somática en niveles altos. Este cultivar Forastero también muestra una alta producción de embriones somáticos primarios en

comparación con los cultivares Trinitario. Nuestra decisión de centrarnos en INIAPG-038 dependió en gran medida de la importancia del cultivar, así como del potencial embriogénico somático (Fig. 2 y 3A) y la eficiencia de suministro de ADN-T (Fig. 5).

Se realizaron experimentos de transformación estable en INIAPG-038 mediante la infección de tejidos de cotiledones derivados de los embriones somáticos secundarios con AGL1 que contenía pDDNPTYFP-1 o pDDNPTYFP-2 (Fig. 1A-1B). Se podrían disecar de seis a diez explantes de cotiledones de cada embrión utilizado en este proceso. La expresión transitoria de YFP se observó claramente de 5 a 7 d después de la infección por *Agrobacterium* (Fig. 8A). Se realizaron cinco experimentos exitosos separados de transformación estable en INIAPG-038 mediante selección de marcadores visuales de YFP. La selección de kanamicina no fue lo suficientemente rigurosa e igual se formó tejido embrionario no transgénico en los explantes de cotiledones, incluso con un alto nivel de kanamicina. En un método descrito anteriormente, se utilizaron 50 mg de L⁻¹ de genetina (G418) *paranptII*, lo que reduce en gran medida la formación de tejidos no transgénicos. Se generaron ocho eventos de transformación independientes en forma de embriones globulares con expresión de YFP de INIAPG-038 a partir de los primeros 4 conjuntos de experimentos mediante el uso de tejidos de cotiledones derivados de 82 embriones somáticos; la frecuencia de transformación en el nivel de tejido T₀ fue del 9,8 % (8 de 82) (Tabla 1, Fig. 8B). Se generaron once eventos de transformación independientes más mediante el uso de 82 embriones somáticos adicionales en el 5.º conjunto de experimentos; la frecuencia de transformación en el nivel de tejido T₀ fue del 13,4 % (11 de 82) (Tabla 1).

Tabla 1. Frecuencias de transformación estable de *Theobroma cacao* cv. INIAPG-038

Experimentos de transformación	Vector usado	Cantidad de embriones disecados ^a	Cantidad de eventos de expresión de YFP	Frecuencia de transformación a nivel de tejido T ₀ ^b	Cantidad de eventos regenerables	Frecuencia de transformación a nivel de planta T ₀ ^c	Cantidad de embriones somáticos cuaternarios producidos	Cantidad de plántulas regeneradas	Cantidad de plántulas en el suelo
Conjunto 1	pDDNPTYFP-1	32	2	6,3 %	1	3,1 %	68	15	11
Conjunto 2	pDDNPTYFP-2	10	1	10,0 %	0	0,0 %	0	0	0
Conjunto 3	pDDNPTYFP-2	24	3	12,5 %	1	4,2 %	16	4	3
Conjunto 4	pDDNPTYFP-2	16	2	12,5 %	1	6,3 %	95	20	3 (1) ^d
	Subtotal	82	8	9,8 %	3	3,7 %	179	39	17
Conjunto 5	pDDNPTYFP-2	82	11	13,4 %	n.d.	n.d.	n.d.	n.d.	n.d.
	Total	164	19	11,6 %					

^a Después de la disección, un embrión somático produciría de 6 a 10 explantes de cotiledones

^b Frecuencia de transformación a nivel de tejido T₀ = (cantidad de eventos de expresión de YFP/cantidad de embriones disecados) x 100 %

^c Frecuencia de transformación a nivel de planta T₀ = (cantidad de eventos regenerables/cantidad de embriones disecados) x 100 %

^d Uno de cada 3 es una planta injertada y desplazada al suelo para crecer en el invernadero

n.d.: no determinada

El proceso de transformación descrito en el presente documento proporciona solamente un único TSE transgénico (Fig. 8B) para cada evento de transformación y el índice promedio de conversión de embriones para INIAPG-038 fue solo del 22,4 % (Fig. 3B), por lo que si se intentara regenerar TSE transgénicos directamente, la mayoría de los eventos de transformación probablemente no se recuperarían como plantas. Para aumentar la probabilidad de regenerar eventos de transformación en plantas, se llevó a cabo un ciclo de embriogénesis adicional para generar múltiples QSE transgénicos a partir de cada evento de transformación antes de la etapa de conversión embrionaria. De los 8 eventos de transformación generados a partir de los primeros 4 conjuntos de experimentos de transformación, 3 TSE transgénicos se desarrollaron en embriones maduros (Fig. 8C) que proliferaron en múltiples QSE (Fig. 8D) y fueron capaces de regenerarse en plántulas (Fig. 8E), lo que dio como resultado una frecuencia de transformación en el nivel de planta T₀ del 3,7 % (3 de 82) (Tabla 1). La frecuencia de transformación T₀ se calculó como (cant. de eventos regenerables/cant. de embriones disecados) x 100 %. Con embriogénesis somática cuaternaria, los tres eventos transgénicos, #1, 2 y 3, proliferaron en 68, 16 y 95 embriones en total, respectivamente (Tabla 1). Uno (evento #3) de los tres embriones transgénicos fue quimérico y los QSE no transgénicos se descartaron en función de la expresión de YFP. De los QSE generados para cada evento, 15, 4 y 20 de esos embriones se convirtieron en

plántulas, respectivamente (Tabla 1), lo que arroja un índice promedio del 21,8 % de conversión de embriones transgénicos en plántulas. De los embriones convertidos en plantas para cada evento, 11, 3 y 3 plantas de cada evento se aclimataron exitosamente en el suelo, respectivamente (Tabla 1 y Fig. 8F).

Algunas plántulas transgénicas regeneradas de INIAPG-038 tuvieron un desarrollo de raíces lento o pobre y no pudieron sobrevivir en el suelo; por lo tanto, se intentó el injerto para resolver este problema. Las plántulas con sistemas de raíces poco desarrollados mueren habitualmente durante el trasplante al suelo; también existe el riesgo de dañar los sistemas de raíces bien establecidos durante la extracción del agar. Sin embargo, el injerto de vástagos transgénicos *in vitro* evita estos dos problemas ya que se utiliza una plántula no transgénica robusta y bien enraizada y todo el sistema de raíces intacto dentro de un cubo de turba Jiffy esterilizado puede colocarse en el suelo sin alterar las raíces (Fig. 8G). En total, se injertaron seis vástagos transgénicos con raíces pobres en portainjertos, y cuatro de estos injertos sobrevivieron. De ellas, 1 planta del evento #3 se transfirió al suelo para crecer en el invernadero (Tabla 1). Además, los injertos podrían facilitar el uso de brotes transgénicos que de otro modo no podrían formar raíces. El uso de injertos para superar un enraizamiento deficiente *in vitro* y mejorar la recuperación transformante también se utilizó anteriormente en cártamo y algodón.

La presencia de los transgenes, *eyfp* y *nptII*, se confirmó mediante análisis de PCR. El producto amplificado de 595 pb correspondiente al fragmento interno del gen *eyfp* se observó a partir del ADN genómico de las 9 plantas transgénicas analizadas derivadas de tres eventos independientes mediante el uso de cebadores específicos del gen *eyfp*, EYFP 3F y EYFP 4R (Fig. 9A). También se observó un fragmento amplificado de 761 pb de todas las plantas transgénicas analizadas derivadas de tres eventos independientes mediante el uso de cebadores específicos del gen *nptII*, NPTII 3F y NPTII 4R (Fig. 9B). No se detectaron fragmentos de *eyfp* y *nptII* amplificados por PCR en la planta de control no transgénica (Fig. 9A, 9B) tal como se esperaba. Se realizaron controles de PCR adicionales mediante el uso de cebadores para el origen de replicación de pVS1 bacteriano y el gen *nptI* ubicado fuera de los bordes del ADN-T para verificar la posibilidad de que la contaminación por *Agrobacterium* diera resultados falsos positivos. La ausencia de los transgenes, *pvs1* y *npt1* se confirmó mediante análisis de PCR tanto en el primer como en el segundo evento; sin embargo, el tercer evento produjo productos amplificados correspondientes a dos secuencias

principales del vector de *Agrobacterium* (Fig. 9C, 9D). Se compararon las dos cepas de *Agrobacterium* utilizadas para la transformación, así como muestras de hojas intactas del tercer evento, para determinar si la contaminación por *Agrobacterium* fue la causa de la detección de los cuatro genes. Las cepas de *Agrobacterium* que contenían el casete del gen *eyfp* no mostraron expresión de YFP y la expresión de YFP de las muestras de hojas del tercer evento fue uniforme y coherente con la de los otros dos eventos (Fig. 17A-17B). Se observó una expresión de YFP más fuerte impulsada por el promotor CaMV35S en los tejidos de las hojas en comparación con la impulsada por el promotor de Nos/potenciador de TMV Ω . La presencia de las secuencias de la estructura principal del vector de *Agrobacterium* indicaría entonces que el tercer evento fue un evento de lectura completa y que esos genes están presentes adicionalmente con los otros transgenes. La lectura completa es la integración en el genoma de la planta de secuencias completas de la estructura principal del vector que podrían ser el resultado de una transferencia conjugativa iniciada en el borde derecho y la posterior copia continua en los bordes izquierdo y derecho. La relación de plantas leídas en eventos transgénicos oscila normalmente entre el 20 % y el 50 % y, a veces, llega hasta el 75 % o más.

En conclusión, hemos establecido un exitoso sistema de transformación mediado por *Agrobacterium* para la producción de plantas transgénicas de *Theobroma cacao* mediante el uso de INIAPG-038, un cultivar de alto rendimiento y resistente a enfermedades. Anteriormente, solo se había informado de transformación estable de PSU SCA-6. Los resultados en este estudio pueden aplicarse para mejorar la respuesta del cultivo de tejidos y la frecuencia de transformación en otros cultivares de *Theobroma cacao*.

EJEMPLO 2

Uso de factores de transcripción para mejorar la frecuencia de transformación y reducir la dependencia del genotipo en *Theobroma cacao*

Construcción de vectores y cepas de Agrobacterium

Se clonaron casetes de expresión de genes morfogénicos y marcadores visuales en el vector binario pCambia2300 CR3 (Fig. 10A-10E). Inicialmente, el promotor de *ubiquitina-10* de *A. thaliana* (AtUbi10), el gen de la proteína fluorescente amarilla mejorada (*eyfp*) y dos elementos de terminación de nopalina sintasa (Nos) de *A. tumefaciens* se amplificaron por

PCR mediante el uso de Phusion® High-Fidelity Polymerase [New England Biolabs (NEB)]. Los amplicones purificados se ensamblaron en el pCAMBIA2300 CR3 digerido con KpnI-HF mediante el uso del kit de clonación de ensamblaje de ADN NEBuilder® HiFi (NEB). A continuación, se extrajo el casete CR3 mediante digestión con PmeI y KpnI-HF durante la noche. Los fragmentos de ADN se separaron en un gel TAE al 0,75 % y el fragmento pCAMBIA2300 AtUbi10p:YFP deseado se extrajo mediante el uso del kit de extracción en gel de ADN Monarch® (NEB). Para el ensamblaje de construcciones con Le1p:AtBBM, Le1p:AtWUS y Le1p:GRF/GIF, se amplificaron y ensamblaron los siguientes elementos de ADN con el fragmento lineal pCAMBIA2300 AtUbi10p:eYFP mediante el uso del kit de clonación de ensamblaje de ADN NEBuilder® HiFi. El elemento de terminación y promotor del gen *Le1* y la lectina de soja con especificidad de semilla (Le1p) se amplificaron a partir del ADN genómico de la soja. Se amplificaron los genes *Baby boom* (*bbm*) y *Wuschel2* (*wus*) a partir de ADN genómico de *A. thaliana*. El gen quimérico del FACTOR REGULADOR DEL CRECIMIENTO (*grf*, por su sigla en inglés) y del cofactor FACTOR DE INTERACCIÓN CON EL GRF (*gif*, por su sigla en inglés) derivó de *Vitis vinifera* (que proporcionó amablemente el laboratorio Jorge Dubcovsky). Para el ensamblaje de la construcción con ZmUbi1p:GRF/GIF, se amplificó y ensambló el promotor del gen de *ubiquitina-1* de *Zea mays* (Ubi1) tal como se describió anteriormente. Para el ensamblaje de AtUBQ10p:eYFP+enh35Sp:nptII, los extremos del ADN del fragmento pCAMBIA2300 AtUbi10p:eYFP digerido con PmeI/KpnI se hicieron romos mediante el uso de ADN polimerasa T4 (NEB). Los extremos romos del fragmento se ligaron mediante ADN ligasa T4 (NEB) en alta concentración para producir un plásmido circular. Cada construcción se transformó en células electrocompetentes de AGL1 mediante electroporación.

Transformación de *Theobroma cacao*

Las transformaciones del cacao se realizaron mediante el uso de tejido de cotiledón derivado de embriones somáticos de 4 a 10 mm de longitud, de color blanco o tostado y con morfología normal, tal como se describió en el Ejemplo 1. Cinco transformaciones realizadas solo con pGmLe1AtBBM (Fig. 10B). Se pudieron detectar embriones con expresión de YFP transformados con este vector 1-2 meses después de la infección por *Agrobacterium* en comparación con 3_4 meses después de la infección por *Agrobacterium* con el vector de control. Se generaron treinta eventos en los 3 cultivares evaluados, Matina, INIAPG-038 y CCN_51 (Tabla 2). La mayoría de estos eventos se generaron en Matina y 18 de ellos llegaron

a la etapa QSE. Según las primeras observaciones, la mayoría de estos eventos parecen tener una morfología normal (Fig. 11). El fenotipo anormal del embrión se observó en la literatura anterior y produjo embriones incapaces de regenerarse cuando se utilizó el promotor constitutivo CaMV35S. Para cada uno de los primeros eventos de Matina hay al menos unos pocos embriones que tienen una morfología normal y carecen del color marrón o verde del tejido que muestra parte del material, por lo que esperamos que la mayoría de los eventos tengan capacidad de regeneración.

Se realizaron dos conjuntos más de experimentos de transformación para comparar las siguientes 5 construcciones tanto en Matina como en INIAPG-038; control, pGmLe1AtBBM, pGmLe1AtWUS, pGmLe1VvGRF/GIF y pZmUbi1VvGRF/GIF (Tabla 2, Fig. 10A-10E, Figura 18). La primera transformación proporcionó 1 evento en Matina de pGmLe1AtBBM. Un solo evento de ambos genotipos progresó a la etapa QSE en este experimento y el segundo tuvo 2 eventos en INIAPG-038 (Tabla 2). Hasta el momento no se generaron otros eventos a partir de ninguna de las otras construcciones. El material donante total utilizado para cada construcción fue pequeño, para Matina promedió 8,4 embriones por construcción y para INIAPG-038 fue 14,8. La tercera y cuarta transformación que compararon las 5 construcciones utilizaron 40 embriones de Matina por construcción y 32 embriones de INIAPG-038 por construcción (Tabla 2).

Tabla 2. Estado de la transformación del cacao mediante el uso de vectores de factores de transcripción

Cultivar	Material donante	Construcción	Cant. de embriones infectados	Cant. de eventos con expresión de YFP	Eventos con QSE
Matina	Embrión secundario	pGmLe1AtBBM	40	4	4
G_038	Embrión secundario	pGmLe1AtBBM	75	6	1
Matina	Embrión secundario	pGmLe1AtBBM	75	13	11
G_038	Embrión	pGmLe1AtBBM	75	5	1

Cultivar	Material donante	Construcción	Cant. de embriones infectados	Cant. de eventos con expresión de YFP	Eventos con QSE
	secundario				
CCN_51	Embrión secundario	pGmLe1AtBBM	75	2	1
Matina	Embrión secundario, cultivo SSE	control	8,4	0	0
Matina	Embrión secundario, cultivo SSE	pGmLe1AtBBM	8,4	1	1
Matina	Embrión secundario, cultivo SSE	pGmLe1AtWUS	8,4	0	0
Matina	Embrión secundario, cultivo SSE	pGmLe1VvGRF/GIF	8,4	0	0
Matina	Embrión secundario, cultivo SSE	pZmUbi1VvGRF/GIF	8,4	0	0
G_038	Embrión secundario, cultivo SSE	control	14,8	0	0
G_038	Embrión secundario, cultivo SSE	pGmLe1AtBBM	14,8	2	1
G_038	Embrión secundario, cultivo SSE	pGmLe1AtWUS	14,8	0	0
G_038	Embrión	pGmLe1VvGRF/GIF	14,8	0	0

	Cultivar	Material donante	Construcción	Cant. de embriones infectados	Cant. de eventos con expresión de YFP	Eventos con QSE
5		secundario, cultivo SSE				
10	G_038	Embrión secundario, cultivo SSE	pZmUbi1VvGRF/GIF	14,8	0	0
	Matina	Embrión secundario	control	40	nd	nd
15	Matina	Embrión secundario	pGmLe1AtBBM	40	nd	nd
	Matina	Embrión secundario	pGmLe1AtWUS	40	nd	nd
	Matina	Embrión secundario	pGmLe1VvGRF/GIF	40	nd	nd
20	Matina	Embrión secundario	pZmUbi1VvGRF/GIF	40	nd	nd
	G_038	Embrión secundario	control	32	nd	nd
25	G_038	Embrión secundario	pGmLe1AtBBM	32	nd	nd
	G_038	Embrión secundario	pGmLe1AtWUS	32	nd	nd
	G_038	Embrión secundario	pGmLe1VvGRF/GIF	32	nd	nd
30	G_038	Embrión secundario	pZmUbi1VvGRF/GIF	32	nd	nd

EJEMPLO 3**Uso de genes de factores de transcripción para pruebas de transformación en
Nicotiana tabacum y *Nicotiana benthamiana****Protocolo de transformación para el tabaco y Nicotiana benthamiana*

De cuatro a seis semanas antes de la transformación, las semillas se esterilizaron con lejía al 20 % durante 15 minutos, se lavaron con H₂O destilada estéril de cuatro a cinco veces y se germinaron en Phytatray™ II (Sigma-Aldrich, San Luis, MO) que contenía medios de germinación de tomate (SIG). Un día antes de la transformación, se colocó en estrías una cepa de *Agrobacterium* seleccionada que contenía un vector binario sobre placas de inducción de *Agrobacterium* con acetosiringona (AS) a 100 µM y el antibiótico adecuado para la selección. El agente de amortiguación de infección que contenía *Agrobacterium* se diluyó en un tubo Falcon de 15 mL suplementado con AS a 200 µM, 6-bencilaminopurina (BAP) a 1 mg/L y ácido 1-naftalenacético (NAA) a 0,1 mg/L. Para la transformación se utilizó una lectura de DO₆₀₀ de 0,400 (+/- 0,015) del agente de amortiguación de infección. Se disecaron explantes de *Nicotiana tabacum* (tabaco) o *Nicotiana benthamiana* mediante el corte de tejido foliar (5-10 mm²) y su colocación encima de un papel de filtro Whatman estéril de 70 mm (cat. n.º 1001-070) en una placa de Petri esterilizada de 100 mm x 25 mm, humedecida con 800 µL de la mezcla bacteriana. Los tejidos se infectaron al pipetear al menos 5 mL de mezcla de bacterias diluidas en placas de Petri. La placa de Petri que contenía los tejidos foliares se selló con parafilm y se colocó en un agitador orbital a temperatura ambiente durante dos horas. Luego se eliminó el agente de amortiguación de infección y los tejidos se transfirieron a un medio de cocultivo durante tres o cuatro días de incubación a 21 °C. Los cultivos se comprobaron tres días después del cocultivo para supervisar el crecimiento excesivo de *Agrobacterium* que rodea los tejidos. Los tejidos se lavaron durante la noche en un tubo Falcon de 50 mL que contenía 15 mL de medio de infección con cefotaxima a 200 mg/L, timentina a 200 mg/L, BAP a 1 mg/L y NAA a 0,1 mg/L si se observa crecimiento excesivo de *Agrobacterium*. Los tejidos foliares que parecían estar libres de crecimiento bacteriano excesivo se lavaron durante un día más y se secaron en un papel de filtro esterilizado para su posterior transferencia a la primera ronda de selección. Los tejidos se transfirieron a medios de selección nuevos cada dos semanas manteniendo contacto con los medios para una selección eficiente. Los brotes transgénicos putativos que alcanzaron la altura de al menos 1,5 cm se transfirieron a un medio

de enraizamiento en Phytatray™ II para generar las plantas transgénicas.

Efectos de BABYBOOM (BBM) y WUSCHEL (WUS) en la frecuencia de transformación y el fenotipo vegetal en tabaco transgénico y N. benthamiana

Se tenía conocimiento de que genes morfogénicos de plantas tales como *BABY BOOM* (*bbm*) y *WUSCHEL* (*wus*) inducían estructuras similares a embriones o embriones somáticos, además de mejorar la regeneración, lo que condujo al aumento en la frecuencia de transformación de genotipos recalcitrantes en el maíz. Evaluamos los efectos de estos dos genes en sistemas modelo de plantas dicotiledóneas de *tabaco* y *Nicotiana benthamiana*.

Según los resultados preliminares que obtuvimos del primer conjunto de transformación, parece que se observaron diferencias entre las plantas de tabaco y *N. benthamiana* que fueron transformadas con genes *bbm* y *wus* impulsados por el promotor de la lectina con especificidad con respecto a embriones de soja (*Le1*) (Tablas 3 y 4, Figuras 12-15). Después de seleccionar los tejidos de las hojas transformadas en un medio que contenía kanamicina durante la selección 30 días después de la infección, observamos el desarrollo de abundantes sectores verdes/brotes pequeños en tejidos transgénicos putativos pGmLe1AtBBM, seguidos de pGmLe1AtWUS, y los sectores menos verdes fueron de la construcción de control NPTII-eYFP para tabaco y *N. benthamiana*. Aquí, encontramos que la mayor eficiencia de transformación puede observarse en pGmLe1AtBBM>pGmLe1AtWUS>control. Después de 66 días de selección, se generaron supuestas plantas de tabaco transgénico en pGmLe1AtBBM y episodios transgénicos menores en plantas en pGmLe1WUS. Sin embargo, las plantas de *N. benthamiana* parecen tener una menor velocidad de crecimiento después de la selección donde no se generaron plántulas (con formación de raíces) al momento en que se registraron estos datos. Los datos incluyen una evaluación cualitativa tanto de la eficiencia de la transformación como de la regeneración de las plántulas. La eficiencia de la transformación mediante el uso de estos genes de factores de transcripción se generará una vez que se completen los recuentos finales de plántulas transgénicas putativas. Actualmente, no existen efectos fenotípicos significativos en el crecimiento de brotes y desarrollo de raíces de plantas de tabaco transgénicas.

Tabla 3. Eficiencia de transformación y efectos fenotípicos en plantas de tabaco transgénica 66 días después de la infección por *Agrobacterium*.

Construcción	Eficiencia de transformación	Desarrollo de brotes	Crecimiento de raíces
Control	++	-	-
pGmLe1AtBBM	++++	++++	++++
pGmLe1AtWUS	+++	++	+

- Ninguna planta con raíces completas y brotes regenerados al momento del registro de datos

+ Leve

++ Moderado/a

+++ Alto/a

++++ Máximo/a

Tabla 4. Eficiencia de transformación y efectos fenotípicos en plantas transgénicas de *Nicotiana benthamiana* 66 días después de la infección por *Agrobacterium*.

Construcción	Eficiencia de transformación	Desarrollo de brotes
Control	++	+
pGmLe1AtBBM	++++	++++
pGmLe1AtWUS	+++	++

+ Leve

++ Moderado/a

+++ Alto/a

++++ Máximo/a

EJEMPLO 4**Uso de genes de factores de transcripción para la transformación en el tomate***Transformación del tomate*

Las semillas de *Solanum lycopersicum* M82 se esterilizaron superficialmente en una solución de lejía al 20 % (v/v) que contenía una gota de Tween-20 durante 15 min antes de enjuagarse tres veces con agua esterilizada en autoclave. Las semillas se germinaron en placas de Petri que contenían 30 mL de medio basado en Murashige y Skoog (MS) (Phytotech M524) que contenía sales MS a 2,15 g/L, *mio*-inositol (Phytotech I703) a 100 mg/L, sacarosa a 10 g/L, solución vitamínica de Gamborg 1000x (Phytotech G219) a 1 mL/L, MES a 195,2 g/L y Phytigel a 3,5 g/L (Sigma-Aldrich P8169), pH 5,8. Los cultivos se mantuvieron a 26 °C bajo un fotoperíodo de 16 h de luz/8 h de oscuridad a 57-65 $\mu\text{E m}^{-2} \text{s}^{-1}$ durante seis o siete días.

Un día antes de la infección, se colocaron en estrías cepas de *Agrobacterium* en placas de inducción de *Agrobacterium* de 20 mL que contenían extracto de levadura Bacto (Gibco 288620) a 5 g/L, peptona de Bacto (Gibco 211684) a 10 g/L, cloruro de sodio a 5 g/L, agar de Bacto (BD Difco 214050) a 15 g/L, acetosiringona (AS) a 100 μM y kanamicina a 50 mg/L y se mantuvieron durante 16 h a 28 °C. Las *Agrobacterium* recién formadas se suspendieron en 5 mL de medio de infección líquido que contenía sales MS (Sigma-Aldrich M5519) a 4,33 g/L, *mio*-inositol a 100 mg/L, solución vitamínica de Gamborg 1000x (Phytotech G219) a 1 mL/L, MES a 195,2 g/L, glucosa (Sigma-Aldrich G7021) a 30 g/L, AS a 200 μM , ribósido de *trans*-zeatina a 2 mg/L y ácido indol-3-acético (IAA) a 1 mg/L. Las soluciones de *Agrobacterium* se diluyeron para lograr una DO_{600} de 0,400 +/- 0,015. Se extirparon tejidos de cotiledones M82 de plántulas en germinación y se colocaron en papel de filtro esterilizado (Whatman 1001-070) que contenía 1 mL de mezcla de *Agrobacterium*. Los tejidos se colocaron con el lado abaxial hacia arriba y se cortaron para que no alcanzaran más de 5 mm de longitud. La solución bacteriana restante se transfirió a la placa de Petri y los tejidos se colocaron en un agitador de incubación de 80 rpm a temperatura ambiente durante dos horas. Luego, los tejidos se transfirieron a un medio de cocultivo con el lado adaxial hacia arriba que contenía sales MS a 4,33 g/L, *mio*-inositol a 100 mg/L, solución vitamínica de Gamborg 1000x a 1 mL/L, MES a 195,2 g/L, glucosa a 20 g/L, agar a 6 g/L (Sigma-Aldrich A1296), AS a 200 μM , ribósido de *trans*-zeatina a 2 mg/L e IAA a 1 mg/L. Los tejidos se

mantuvieron a 21 °C durante 96 h en la oscuridad.

Después de 4 d, los tejidos se transfirieron a un papel de filtro limpio con la adición de 800 uL de medio de infección líquido durante 16-24 h a 21 °C. Luego, los tejidos se transfirieron a la primera ronda de medio de selección que contenía sales MS a 4,33 g/L, *mio*-inositol a 100 mg/L, solución vitamínica de Gamborg 1000x, a 1 mL/L, MES a 195,2 g/L, sacarosa a 20 g/L, agar vegetal (Phytotech A296) a 8 g/L, ribósido de *trans*-zeatina a 2 mg/L y 0,5 mg/L, cefotaxima (Phytotech C380) a 200 mg/L, timentina (Phytotech T869) a 200 mg/L y kanamicina (Phytotech K378) a 90 mg/L. Los tejidos se mantuvieron a 26 °C bajo 16 h de luz/8 h de oscuridad durante dos semanas antes de transferirlos a la segunda ronda de medio de selección que contenía IAA a 0,1 mg/L. Cuando los brotes alcanzaron al menos 3 mm de altura, se transfirieron a un medio de elongación de brotes que contenía sales MS a 4,33 g/L, *mio*-inositol a 100 mg/L, sacarosa a 20 g/L, agar vegetal a 8 g/L, ribósido de *trans*-zeatina a 1 mg/L, cefotaxima a 200 mg/L, timentina a 200 mg/L, vitaminas de Nitsch & Nitsch (1000x) a 1 mL/L y kanamicina a 90 mg/L. Después de que los brotes alcanzaron una altura de 25 mm, se transfirieron a un medio de enraizamiento que contenía sales MS a 4,33 g/L, sacarosa a 30 g/L, agar vegetal a 8 g/L, IAA a 0,2 mg/L, cefotaxima a 200 mg/L, timentina a 200 mg/L, vitaminas de Nitsch & Nitsch (1000x) a 1 mL/L y kanamicina a 60 mg/L en Phytatrays (Sigma-Aldrich P5929).

Todos los medios se ajustaron hasta un pH de 5,8 antes del tratamiento en autoclave. Ingredientes tales como ribósido de *trans*-zeatina, IAA, kanamicina, cefotaxima, timentina y vitaminas de Nitsch y Nitsch se esterilizaron mediante filtración y se agregaron a un medio tratado en autoclave que se dejó enfriar hasta 55 °C.

Efectos de BABYBOOM (BBM) y WUSCHEL (WUS) en la frecuencia de transformación y el fenotipo vegetal en tomate transgénico

Después de seleccionar los tejidos foliares transformados en medio que contenía kanamicina durante la selección 2,5 meses después de la infección, se observó la frecuencia de transformación más alta y el desarrollo de brotes más rápido de pGmLe1AtWUS. No hubo diferencias en la frecuencia de transformación y el desarrollo de brotes entre el control y pGmLe1AtBBM (Tabla 5).

Tabla 5. Eficiencia de transformación y fenotipo en plantas de tomate transgénicas 75 días

después de la infección por *Agrobacterium*.

5	Construcción	Frecuencia de transformación	Desarrollo de brotes	Crecimiento de raíces
	control	+++	++	-
	pGmLe1AtBBM	+++	++	-
	pGmLe1AtWUS	++++	++++	-

10 - Sin plantas con raíz completa y brote regenerado al momento del registro de datos

+ Leve

++ Moderado/a

15 +++ Alto/a

++++ Máximo/a

El fenotipo vegetal de las plantas M82 transformadas con pGmLe1AtWUS de experimentos anteriores fue normal.

20

EJEMPLO 5

Uso de genes de factores de transcripción para la transformación en la yuca

Transformación de yuca

25

La transformación mediada por Agrobacterium se realizó mediante el uso del callo embriogénico friable (FEC, por su sigla en inglés) de los registros de yuca 60444, TME 419 y 91/02324, con la posterior regeneración de la planta. Se indujeron embriones somáticos a partir de lóbulos de hojas de plantas micropropagadas *in vitro* mediante cultivo en medio basal de Murashige y Skoog (MS) suplementado con sacarosa (MS2) a 20 g/L más picloram a 50 µM. Los embriones en etapa precotiledónea se subcultivaron en medio basal de Gresshoff y Doy (GD) suplementado con sacarosa (MS2) a 20 g/L y picloram a 50 µM (GD2 50P) para inducir la producción de FEC.

30

Se seleccionaron FEC homogéneos que se usaron como tejido objetivo para la

35

transformación con la cepa AGL1 de *A. tumefaciens* portadora de cada una de las 5 construcciones de TF descritas en el ejemplo 2. Un día antes de la transformación, las cepas de AGL1 se colocaron en estrías sobre medio LB sólido que contenía AS a 200 μ M y esto se colocó en un incubador durante la noche a 26–28 °C. El día de la transformación, las *Agrobacterium* se diluyeron en tubos cónicos de 50 mL que contenían medio líquido GD2 50P suplementado con AS a 200 μ M para lograr una lectura de DO a 600 nm de hasta 0,75. Para la infección, los FEC se colocaron en tubos cónicos de 50 mL que contenían suspensión de *Agrobacterium* y se sometieron a sonicación con un limpiador ultrasónico Branson 3510-DTH durante 3 segundos. Posteriormente, los tubos se colocaron en un agitador durante entre 3 y 4 h a temperatura ambiente. Luego, se extrajo la suspensión de *Agrobacterium* de la placa de Petri y los FEC infectados se extendieron sobre 2 capas de papel de filtro cualitativo Whatman Grado 1 (70 mm de diámetro) en una placa de Petri nueva. La capa superior de papel de filtro que contenía FEC infectado se colocó en una placa de Petri que contenía medio GD2 50P semisólido suplementado con AS a 200 μ M. Después de 3 a 5 días de cocultivo a 21 °C, se controlaron las *Agrobacterium* mediante la adición de cefotaxima a 200 mg/L al medio basal GD2 50P y al medio MS2. Las células transgénicas se seleccionaron y proliferaron en GD2 50P que contenía paromomicina, antes de la regeneración de los embriones en medio MS2 suplementado con ácido naftalenoacético (NAA, por su sigla en inglés). Los embriones somáticos se germinaron en medio MS2 que contenía BAP y las plantas regeneradas se mantuvieron en MS2.

De diez a catorce días después de la distribución en el medio de selección, se identificaron grupos de FEC en crecimiento activo de más de 2-3 mm de diámetro bajo el microscopio de disección y se subcultivaron mediante el uso de pinzas finas en el medio de regeneración de la Etapa 1 que consistía en medio MS2 suplementado con NAA a 5 μ M, cefotaxima a 200 mg/L y paromomicina a 45 μ M, con entre 6 y 10 callos colocados en cada placa de Petri. Este proceso se repitió 10 y 20 días después, retirando las unidades de FEC en crecimiento de las placas de selección de GD2 50P en cada momento y transfiriéndolas al medio de regeneración de la Etapa 1. Después de 21 días de cultivo en medio de la Etapa 1, se subcultivaron colonias de FEC en crecimiento activo en un medio de regeneración de la Etapa 2 que consistía en medio MS2 suplementado con NAA a 0,5 μ M y paromomicina a 45 μ M. Después de 21 días de cultivo en medio de regeneración de la Etapa 2, se seleccionaron embriones en etapa de cotiledón de color verde de cada línea transgénica putativa y se transfirieron de 5 a 6 embriones individualmente a una placa de Petri que contenía

medio MS2 suplementado con BAP a 2 μ M (MS2 2BAP). La germinación de embriones somáticos para producir un tallo distinto y hojas trilobuladas se produjo de forma asincrónica durante las siguientes 3 a 4 semanas. Una vez que se formaron brotes distintos, se extirparon de los tejidos de los cotiledones juveniles y se transfirieron a medio MS2 para obtener el
 5 alargamiento del tallo y el establecimiento de las plántulas. Después de tres a 4 semanas de cultivo en medio MS2, se retuvo y micropropagó la planta más robusta.

Efectos de BABYBOOM (BBM) y WUSCHEL (WUS) en la frecuencia de transformación y el fenotipo vegetal en la yuca transgénica

10 Las plantas de yuca transformadas con pGmLe1AtWUS y pGmLe1AtBBM mostraron un fenotipo normal.

EJEMPLO 6

15 **Uso de genes de factores de transcripción para la transformación en la soja**

Transformación de la soja

La transformación de la soja se realizó mediante el uso de cotiledones inmaduros y cultivos embriogénicos en suspensión de Williams 82 mediante bombardeo con
 20 micropartículas. Las vainas inmaduras se retiraron cuando las semillas tenían de 3 a 7 mm de largo y se esterilizaron en superficie durante 15 min en Clorox al 20 % más 2 gotas de Tween 20. Después de enjuagar 3 veces con agua esterilizada, se extirparon los cotiledones inmaduros de las vainas. Después de retirar el extremo de los cotiledones que contenían el
 25 eje embrionario, la superficie abaxial de cada mitad de cotiledón se colocó en un medio sólido que contenía sales MS, vitaminas B5, sacarosa al 3 %, 2,4-D a 20 mg/L y agar de Difco Bacto al 0,8 % (pH 5,8). Los cultivos se incubaron a 26 °C con fotoperíodos de 16 h a 50-60 μ mol m⁻² s⁻¹. Después de tres a seis semanas, se transfirieron embriones somáticos globulares a
 30 30 mL de medio líquido 10A40N suplementado con glutamina a 15 mM en un matraz Erlenmeyer de 125 mL. Los cultivos en suspensión de carácter embriogénico se mantuvieron en un agitador rotatorio a 130 rpm y se subcultivaron de forma rutinaria cada mes.

Se utilizó una pistola de helio PDS-1000 (Bio-Rad, Hercules, CA) para el bombardeo de partículas de ADN. Se colocaron alrededor de 0,6-0,8 g de cultivos en suspensión de carácter embriogénico en el centro de una placa de Petri de 100 mm de diámetro y se eliminó

el exceso de líquido. Se precipitaron de 0,1 a 1 µg de cada construcción de ADN y después los precipitados se volvieron a suspender en 85 µ de etanol. Para cada bombardeo, se cargaron 10 µL de esta mezcla en un microportador y se dejó secar. Se utilizó un disco de presión de ruptura de 650 psi y el vacío de la cámara de muestra fue de 28 pulgadas de mercurio durante el bombardeo. Los tejidos bombardeados se incubaron a 28 °C durante la noche y luego se volvieron a suspender en el medio líquido 10A40N. Dos semanas después del bombardeo, los cultivos se colocaron en el mismo medio pero que contenía G418 a 50 mg/L. El medio se cambió en intervalos de 2 semanas. De seis a nueve semanas después de la selección de G418, se separaron las excrecencias de color amarillo verdoso de los grupos muertos blanquecinos y se transfirieron individualmente para una mayor proliferación a 30 mL de medio 10A40N nuevo que contenía G418. Se recolectaron clones transgénicos putativos para determinar la expresión de YFP y se analizaron mediante análisis de amplificación por reacción en cadena de la polimerasa (PCR).

Se transfirieron grumos embriogénicos, de 2-3 mm de diámetro, a un medio líquido que contenía sales MS, vitaminas B5, maltosa al 6 % (pH 5,7) o a un medio sólido de la misma composición pero solidificado con Phytigel al 0,2 %. Después de dos y cuatro semanas de crecimiento en el medio de desarrollo de embriones líquido o sólido, respectivamente, los embriones en desarrollo se separaron manualmente y se cultivaron como embriones individuales en el medio sólido fresco. Los embriones individuales se mantuvieron en placas de Petri durante 4 semanas y una porción de los embriones maduros se disecó durante entre 2 y 10 días en placas de Petri vacías. Los embriones no disecados se mantuvieron en el mismo medio. La germinación de los embriones se inició a partir de los embriones disecados transfiriéndolos a un medio sólido que contenía sales MS, vitaminas B5, sacarosa A 3 y Phytigel al 0,2 % (pH 5,8). Después del desarrollo de los brotes y las raíces, las plántulas se transfirieron a Phytatrays que contenían el mismo medio para su posterior crecimiento. Las plantas regeneradas con sistemas de raíces bien desarrollados se transfirieron luego a macetas con turba (A. H. Hummert Seed Co., San Luis, MO) que contenían una mezcla esterilizada 3:1 de tierra y vermiculita. Las plantas en macetas se colocaron dentro de una bandeja con tapa de plástico transparente. Después de una aclimatación completa abriendo gradualmente la tapa de la bandeja durante 2 semanas, las plantas se transfirieron a las macetas y estas se trasladaron al invernadero.

La transformación mediada por *Agrobacterium* se realizó mediante el uso de

cotiledones inmaduros. La cepa AGL1 de *A. tumefaciens* que portaba cada una de las 5 construcciones de TF se diluyó en tubos cónicos de 50 mL que contenían medio de infección líquido suplementado con AS a 200 μ M para lograr una lectura de DO a 600 nm de hasta 0,5. Para la infección, se colocaron cotiledones inmaduros en tubos cónicos de 50 mL que contenían suspensión de *Agrobacterium* y se sometieron a sonicación con un limpiador ultrasónico Branson 3510-DTH durante 3 segundos. Posteriormente, los tubos se colocaron en un agitador durante entre 3 y 4 h a temperatura ambiente. Luego, se extrajo la suspensión de *Agrobacterium* de la placa de Petri y los cotiledones infectados se extendieron sobre 2 capas de papel de filtro cualitativo Whatman Grado 1 (70 mm de diámetro) en una placa de Petri nueva. La capa superior de papel de filtro que contenía cotiledones infectados se colocó sobre un medio de inducción de embriones somáticos sólido suplementado con AS a 200 μ m. Después de 3 a 5 días de cocultivo a 21 °C, se controlaron las *Agrobacterium* mediante la adición de cefotaxima a 200 mg/L al mismo medio. Después de una incubación de dos a tres semanas a 26 °C, los tejidos se trasladaron a un medio líquido 10A40N que contenía G418 a 50 mg/L. El resto del procedimiento es igual al protocolo descrito anteriormente para la transformación mediada por bombardeo.

EJEMPLO 7

Uso de Baby Boom (BBM) impulsado por un promotor específico de cotiledones para mejorar la frecuencia de transformación y reducir el fenotipo vegetal anormal en *Theobroma cacao*

Se completaron dos conjuntos de experimentos de transformación para comparar las siguientes cinco construcciones tanto en Matina 1-6 como en INIAPG-038: control, pGmLe1AtBBM, pGmLe11AtWUS, pGmLe1VvGRF/GIF y pZmUbi1VvGRF/GIF (Tabla 7). Se produjo un único evento de embrión somático transgénico en Matina 1-6 solo a partir de pGmLe1AtBBM, mientras que no se generó ningún evento de embrión somático transgénico a partir de ninguna de las otras 4 construcciones (Tabla 7). Se generaron dos eventos de embriones somáticos transgénicos en INIAPG-038: uno a partir de la construcción de control y otro a partir de pGmLe1AtBBM. Sin embargo, el evento de la construcción de control no logró producir QSE y, finalmente, plantas. Además, no se generaron eventos de embriones somáticos transgénicos a partir de ninguna de las otras tres construcciones, pGmLe11AtWUS, pGmLe1VvGRF/GIF y pZmUbi1VvGRF/GIF. Los eventos de embriones somáticos

transgénicos únicos de ambos cultivares transformados con pGmLe1AtBBM se desarrollaron en embriones maduros normales (**Fig. 19A y 19B**), y luego regeneraron plantas con un fenotipo normal (**Fig. 20**). Las frecuencias de transformación estable mediante el uso de pGmLe1AtBBM a nivel de planta T₀ de Matina 1-6 e INIAPG-038 de fueron 11,9 % (1/8,4) y 6,8 % (1/14,8), respectivamente (Tabla 7). Se generaron múltiples eventos de expresión de YFP tanto en Matina 1-6 como en INIAPG-038 a partir de pGmLe11AtWUS, pero todos los eventos produjeron tejidos de callos no regenerables (**Fig. 21**). Tampoco se observaron embriones somáticos transgénicos regenerables a partir de ninguna de las construcciones GRF/GIF (Tabla 7).

Se realizó otro conjunto de experimentos de transformación a gran escala para comparar las construcciones de control frente a pGmLe1AtBBM en Matina 1-6 (Tabla 7). No se generaron eventos de embriones somáticos transgénicos a partir de tejidos de cotiledones derivados de 680 embriones somáticos secundarios mediante el uso de la construcción de control. Por el contrario, se generaron 17 eventos independientes de expresión de YFP a partir de pGmLe1AtBBM. Estos embriones somáticos transgénicos podrían detectarse 1-2 meses después de la infección por *Agrobacterium* en comparación con 3-4 meses después de la infección por *Agrobacterium* con la construcción de control. Todos los eventos excepto uno podrían producir embriones somáticos cuaternarios (QSE) (Tabla 7). Siete de 16 eventos generadores de QSE regeneraron plantas con un fenotipo normal, lo que dio como resultado una frecuencia de transformación a nivel de planta T₀ del 6,1 % (7 de 115). Los nueve eventos restantes que generaron QSE no fueron regenerables. Estos eventos produjeron múltiples embriones somáticos en los mismos tejidos con un fenotipo de leve a moderadamente anormal (**Fig. 22 y Fig. 23**) y una germinación fallida (**Fig. 22**) posiblemente debido a una expresión demasiado alta de BBM en los tejidos impulsados por el promotor Le1 (GmLe1) de la soja. De manera similar, en otros casos se observó que la introducción del factor de transcripción BBM del cacao impulsado por el fuerte promotor constitutivo CaMV35S en el cacao causó un problema de regeneración con un fenotipo anormal a pesar de que aumentó significativamente la frecuencia de transformación. Se informó que los embriones que sobreexpresan BBM impulsados por este promotor constitutivo nunca alcanzaron la etapa normal de desarrollo del embrión somático maduro ni la conversión en una nueva plántula. Se realizaron experimentos de transformación adicionales mediante el uso de la construcción pGmLe1AtBBM en INIAPG-038 y CCN-51 (Tabla 7). Se generaron dos y tres eventos de embriones somáticos transgénicos regenerables (**Fig. 19B y Fig. 19C**) de 150 explantes de cada uno de INIAPG-

038 y CCN-5, lo que dio como resultado frecuencias de transformación a nivel de planta T₀ de 1,3 % y 2,0 %, respectivamente (Tabla 7).

En conclusión, se ha establecido un protocolo exitoso de transformación mediada por *Agrobacterium* para producir plantas transgénicas de *Theobroma cacao* mediante el uso de tres cultivares de élite, Matina 1-6, INIAPG-038 y CCN-51. El uso de un promotor moderadamente fuerte específico de semillas/cotiledones/embriones dio como resultado una regeneración exitosa de plantas transgénicas que expresaban BBM, así como en un aumento significativo de la frecuencia de transformación en el cacao. Se sabe que el promotor Le1 de la soja es un promotor específico de semillas moderadamente fuerte, pero puede ser importante reducir el nivel de expresión de BBM para mejorar el protocolo actual. Inesperadamente, la sobreexpresión de WUS mediante el uso del mismo promotor GmLe1 causó un problema de regeneración al producir tejidos de callos no regenerables en el cacao.

Tabla 7. Frecuencia de transformación y fenotipo vegetal del cacao estable mediante el uso de diferentes construcciones de factores de transcripción

Cultivar	Construcción	Cant. de embriones infectados	Cant. de eventos de embriones somáticos con expresión de YFP	Cant. de eventos con QSE	Cant. de eventos con fenotipo de planta normal	Frecuencia de transformación al nivel de embrión somático T ₀	Transformación al nivel de fenotipo de planta normal T ₀ Frecuencia
Matina 1-6	Control	8,4	0	0	0	0,0 %	0,0 %
	pGmLe1AtBBM	8,4	1	1	1	11,9 %	11,9 %
	pGmLe11AtWUS	8,4	0	0	0	0,0 %	0,0 %
	pGmLe1VvGRF/GIF	8,4	0	0	0	0,0 %	0,0 %

		pZmUbiVvGRF/GIF	8,4	0	0	0	0,0 %	0,0 %
5	INIAPG-038	Control	14,8	1	0	0	6,8 %	0,0 %
		pGmLe1AtBBM	14,8	2	1	1	13,5 %	6,8 %
		pGmLe11AtWUS	14,8	0	0	0	0,0 %	0,0 %
		pGmLe1VvGRF/GIF	14,8	0	0	0	0,0 %	0,0 %
		pZmUbiVvGRF/GIF	14,8	0	0	0	0,0 %	0,0 %
10	Matina 1-6	Control	680	0	0	0	0,0 %	0,0 %
	Matina 1-6	pGmLe1AtBBM	115	17	16	7	14,8 %	6,1 %
	INIAPG-038	pGmLe1AtBBM	150	11	2	2	7,3 %	1,3 %
15	CCN_51	pGmLe1AtBBM	150	6	6	3	4,0 %	2,0 %

EJEMPLO 8.

20 Efectos de BABYBOOM (BBM) y WUSCHEL (WUS) en el fenotipo vegetal en tabaco transgénico

Además de los datos presentados en el Ejemplo 3 anterior, no se observó ninguna diferencia significativa en el fenotipo vegetal entre las plantas de tabaco transformadas con control, pGmLe1AtBBM o pGmLe1AtWUS (Fig. 24).

EJEMPLO 9.

Efectos de BABYBOOM (BBM) y WUSCHEL (WUS) en la frecuencia de transformación y el fenotipo vegetal en tomate transgénico

30 Además de los datos presentados en el Ejemplo 4 anterior, se observó un alto porcentaje de tejido tipo callo verde a partir del tratamiento de infección con pGmLe1AtWUS (Fig. 25). No eran regenerables o inicialmente su regeneración fue lenta. Sin embargo, este tipo de tejido no se produjo en gran medida a partir de pGmLe1AtBBM y construcciones de

control. Las frecuencias de transformación estable a nivel de planta T₀ de M82 mediante el uso de pGmLe1AtBBM, pGmLe1AtWUS y construcciones de control fueron 43,3 % (13/30), 36,7 % (11/30) y 13,3 % (4/30), respectivamente (Tabla 8). pGmLe1AtBBM y pGmLe1AtWUS fueron 3,3 y 2,8 veces más altos que el control, respectivamente. Las plantas de tomate M82 regeneradas transformadas con pGmLe1AtWUS y pGmLe1AtBBM tuvieron un fenotipo normal (Fig. 26).

Tabla 8. Frecuencia de transformación y fenotipo vegetal del tomate estable mediante el uso de diferentes construcciones de factores de transcripción

Cultivar	Construcción	Cant. de explantes (mitades de cotiledones) infectados	Cant. de eventos con regeneración/enraizamiento de plantas	Frecuencia de transformación a nivel de planta T ₀
M82	control	30	4	13,3 %
M82	pGmLe1AtBBM	30	13	43,3 %
M82	pGmLe1AtWUS	30	11	36,7 %

EJEMPLO 10

Además de los datos presentados en el Ejemplo 5 anterior, se seleccionó aproximadamente un gramo (3-4 partes de FEC) de FEC de yuca homogénea (cv. 60444), lo cual se usó como tejido diana para la transformación con la cepa AGL1 de *A. tumefaciens* que portaba cada una de las cinco construcciones de TF diferentes descritas en el Ejemplo 2. La construcción con la frecuencia de transformación más alta al nivel de planta T₀ fue pGmLe1AtBBM (Tabla 9). pGmLe11AtBBM produjo 14 eventos regenerables, un aumento de 7 veces con respecto a la construcción de control que produjo 2 eventos. Solo se produjo un evento a partir de pGmLe1AtWUS o pZmUbi1VvGRF/GIF (Tabla 9). No se observó ningún evento transgénico regenerable a partir de pGmLe1VvGRF/GIF. Las plantas de yuca regeneradas transformadas pGmLe1AtBBM, pGmLe1AtWUS y pZmUbi1VvGRF/GIF parecieron tener un fenotipo normal (Fig. 27A-27D). Sin embargo, el fenotipo de los transformantes pGmLe1VvGRF/GIF no se ha evaluado porque todavía no hemos producido

plantas completamente regeneradas a partir de estas construcciones (Tabla 9).

Tabla 9. Frecuencia de transformación y fenotipo vegetal del cacao estable mediante el uso de diferentes construcciones de factores de transcripción

Cultivar	Construcción	Cantidad de FEC infectados	Cant. total de eventos recuperados en el medio de selección de la etapa 2	Cant. total de eventos recuperados en medio de germinación	Cant. de eventos con resultado positivo de PCR con regeneración vegetal	Fenotipo vegetal
60444	Control	1 g	50	10	2	Normal
	pGmLe1AtBBM	1 g	130	45	14	Normal
	pGmLe1AtWUS	1 g	28	20	1	Normal
	pGmLe1VvGRF/GIF	1 g	15	7	0	N/A
	pZmUbiVvGRF/GIF	1 g	21	13	1	Normal

Aunque en la memoria descriptiva y los ejemplos anteriores se describió y habilitó por completo la presente invención, no se pretende que estos limiten el alcance de la invención, que está definido por las reivindicaciones anexadas.

Todas las publicaciones, patentes y solicitudes de patente se incorporan a la presente mediante referencia. Si bien en la memoria descriptiva anterior se ha descrito la presente invención en relación con determinadas modalidades de esta, y se han establecido muchos detalles con fines ilustrativos, será evidente para las personas expertas en la técnica que la invención es pasible de modalidades adicionales y que algunos de los detalles descritos en el presente documento pueden variar considerablemente sin apartarse de los principios básicos de la invención.

El uso de los términos “un”, “una”, “el/la” y referencias similares en el contexto de la descripción de la invención se deben entender como que abarcan el singular y el plural, salvo

que se indique lo contrario en la presente o se contradiga claramente por contexto. Las expresiones «que comprende/n», «que tiene/n», «que incluye/n» y «que contiene/n» deben interpretarse como expresiones abiertas (es decir, con el significado de «que incluye/n, de modo no taxativo») a menos que se indique lo contrario. La enumeración de los intervalos de valores en la presente tiene la intención meramente de servir como un método abreviado para referirse individualmente a cada valor separado que se encuentra dentro del intervalo, a menos que se indique lo contrario en la presente, y cada valor separado se incorpora a la memoria descriptiva como si se mencionara individualmente en la presente. Todos los métodos descritos en la presente se pueden realizar en cualquier orden adecuado a menos que se indique lo contrario en la presente o que el contexto lo contradiga claramente. El uso de cualquiera y todos los ejemplos, o lenguaje ilustrativo (*p. ej.*, "tal como") proporcionado en la presente, pretende simplemente ilustrar mejor la invención y no representa una limitación en el alcance de la invención a menos que se reivindique lo contrario. Ninguna expresión en la memoria descriptiva debe interpretarse como una indicación de que cualquier elemento no reivindicado sea esencial para la puesta en práctica de la invención.

En la presente se describen modalidades de la presente invención, incluido el mejor modo conocido por los inventores para llevar a cabo la invención. Las variaciones de esas modalidades pueden resultar evidentes para las personas expertas en la técnica al leer la descripción anterior. Los inventores esperan que las personas expertas en la técnica empleen dichas variaciones según sea apropiado, y los inventores pretenden que la invención se practique de forma diferente a como se describe específicamente en la presente. En consecuencia, la presente invención incluye todas las modificaciones y los equivalentes de la materia mencionada en las reivindicaciones anexadas según lo permitido por la ley aplicable. Además, cualquier combinación de los elementos descritos anteriormente en todas las variaciones posibles de estos está abarcada por la invención a menos que se indique lo contrario en el presente documento o que el contexto indique claramente lo contrario.