

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 632 007**

51 Int. Cl.:

C12N 15/82 (2006.01)

A01H 5/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.04.2012 PCT/EP2012/056324**

87 Fecha y número de publicación internacional: **11.10.2012 WO12136788**

96 Fecha de presentación y número de la solicitud europea: **05.04.2012 E 12712280 (2)**

97 Fecha y número de publicación de la concesión europea: **17.05.2017 EP 2694658**

54 Título: **Promotor específico para semillas en algodón**

30 Prioridad:

07.04.2011 EP 11075060
07.04.2011 US 201161472816 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
07.09.2017

73 Titular/es:

BAYER CROPSCIENCE NV (100.0%)
J.E. Mommaertsiaan 14
1831 Diegem, BE

72 Inventor/es:

SCHEIRLINCK, MARIE-THERESE;
MEULEWAETER, FRANK;
JACOBS, JOHN;
VANDEWIELE, MARTINE y
BOUDONCK, KURT

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 632 007 T3

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

DESCRIPCIÓN

Promotor específico para semillas en algodón

La presente solicitud da a conocer una secuencia de ácido nucleico (aislado) que comprende una secuencia de nucleótidos seleccionada de (a) SEQ ID NO: 1 o un fragmento del mismo, en donde dicho fragmento comprende al menos 600 nucleótidos consecutivos de SEQ ID NO: 1, comprendiendo el fragmento, además, un motivo T/G que tiene la secuencia de nucleótidos desde la posición de nucleótidos 298 a 303 de SEQ ID NO: 1 y un motivo de unión MYB que tiene la secuencia de nucleótidos desde la posición 846 a 851 de SEQ ID NO: 1 y en donde el fragmento, cuando se introduce en algodón, tiene actividad del promotor específica para la semilla con altos niveles de expresión en fibras desde el día 5 post-antesis hasta el día 30 post-antesis; (b) una secuencia de nucleótidos con al menos 95% de identidad de secuencia con la secuencia de nucleótidos de (a), comprendiendo el fragmento, además, un motivo T/G que tiene la secuencia de nucleótidos desde la posición de nucleótidos 298 a 303 de SEQ ID NO: 1 y un motivo de unión MYB que tiene la secuencia de nucleótidos desde la posición 846 a 851 de SEQ ID NO: 1 y que tiene actividad de promotor específica para la semilla con altos niveles de expresión en fibras desde el día 5 post-antesis hasta el día 30 post-antesis cuando se introduce en algodón; y (c) una secuencia de nucleótidos complementaria a la secuencia de nucleótidos de una cualquiera de (a) o (b). Además, en esta memoria se describe un gen quimérico que comprende el ácido nucleico (aislado) descrita en esta memoria enlazado operativamente a un ácido nucleico que codifica un producto de expresión de interés, y opcionalmente una secuencia de terminación de la transcripción y de poliadenilación. También se describen en esta memoria un vector, una célula vegetal de algodón transgénica, una planta de algodón transgénica y una semilla tal como se caracteriza en las reivindicaciones. Métodos descritos en esta memoria se refieren a efectuar la expresión específica para la semilla de un producto en algodón y a alterar propiedades de la fibra en una planta de algodón según se caracteriza en las reivindicaciones.

Tricomas son apéndices epidérmicos especializados que se encuentran en la superficie de órganos aéreos de la mayoría de las plantas terrestres. Hay varios tipos de tricomas: unicelulares o multicelulares, ramificados o no ramificados, y glandulares o no glandulares. Los tricomas contribuyen a muchos aspectos de la adaptación de las plantas a tensiones bióticas y abióticas tales como para cercar insectos herbívoros, regular la temperatura de la superficie, disminuir la pérdida de agua por transpiración, aumentar la tolerancia a la congelación, ayudar a la dispersión de semillas, y proteger los tejidos vegetales frente a la luz UV (Eisner et al, 1998; Werker, 2000; Wagner et al, 2004). Tricomas secretores glandulares (GSTs) secretan a menudo metabolitos secundarios de plantas para constituir la resistencia natural basada en el producto de herbívoros y patógenos (Werker, 2000; Ranger y Hower, 2001; Wagner et al., 2004; Medeiros y Tingey, 2006).

Las diferentes especies vegetales pueden tener diferentes tipos de tricomas, y una planta puede formar más de un tipo de tricomas. La maleza anual *Arabidopsis thaliana* produce tricomas no glandulares unicelulares, que pueden ser ramificados o no ramificados (Szymanski et al., 2000). Las plantas de tabaco contienen habitualmente tricomas multicelulares, incluyendo tricomas altos secretores glandulares (GSTs) y tricomas aglandulares simples (Wagner et al., 2004). Las fibras de algodón son tricomas unicelulares y de semillas ampliamente alargados (Kim y Triplett, 2001).

La fibra de algodón es el textil sencillo más importante en todo el mundo. Alrededor de 80 millones de acres de algodón se cosechan anualmente en todo el mundo. El algodón es el quinto más grande de los cultivos en los EE.UU. en términos de superficie de producción, con una media de 10,3 millones de acres plantados en los años 2006 a 2008. Alrededor del 90% del algodón cultivado en todo el mundo es *Gossypium hirsutum*, mientras que *Gossypium barbadense* supone alrededor del 8%. En consecuencia, la modificación de las características de la fibra de algodón para adaptarse mejor a los requisitos de la industria y el consumidor es un esfuerzo importante en el cultivo por cualquiera de los métodos clásicos o alterando genéticamente el genoma de las plantas de algodón. Objetivos a alcanzar incluyen el aumento de longitud de la fibra de pelusa, resistencia, capacidad de tinción, disminución de la producción de fibra rizada, la relación de madurez de la fibra, el contenido de fibra inmadura, la uniformidad de la fibra y el micronaire.

El desarrollo de la fibra de algodón es un proceso de múltiples etapas bajo la regulación de un gran número de genes, muchos de los cuales están supra-regulados o muy expresados en el desarrollo de células de fibra (Li, C.H. et al., 2002; Ruan et al., 2003; Wang, S. et al., 2004; Li et al., 2005; Luo et al., 2007).

Se han descrito diversos promotores que impulsan la expresión de genes en semillas de algodón. Mientras que actualmente se conocen promotores específicos para semillas o específicos para tricomas de algodón, también se utilizan promotores heterólogos para controlar la expresión en el algodón específica para la semilla, específica para la envuelta de la semilla o específica para tricomas.

E6 fue el primer gen identificado de fibra de algodón, y el promotor E6 se ha utilizado para diseñar la calidad de la fibra de algodón (John y Keller, 1996). GhRDL1, un gen altamente expresado en células de fibras de algodón en la fase de alargamiento, codifica una proteína que contiene el dominio BURP (Li, C.H. et al., 2002), y el promotor GaRDL1 exhibía actividad específica para tricomas en plantas transgénicas de *Arabidopsis* (Wang, S. et al., 2004).
 5 Transcripciones GhTUB1 se acumulan preferentemente en altos niveles en fibra, por consiguiente, el gen de fusión pGhTUB1::GUS se expresó en un nivel alto en las fibras, pero a niveles mucho menores en otros tejidos (Li, X.B. et al., 2002). Los promotores de tres genes de la proteína de transferencia de lípidos de algodón, LTP3, LTP6 y FSIt4, fueron capaces de dirigir la expresión del gen GUS en GSTs de las hojas y los tallos en plantas de tabaco transgénicas (Hsu et al., 1999; Liu et al., 2000; Delaney et al., 2007), sin embargo, no exhibían una clara
 10 especificidad tisular. P. ej., en plantas de tabaco transgénicas pFSIt4::GUS, se pudo detectar una fuerte actividad de GUS en todos los tipos de tricomas; además, la expresión de GUS también era visible en el margen de la hoja, el tejido vascular, los óvulos y las puntas de las raíces (Delaney et al., 2007).

El factor de transcripción R2R3 MYB del algodón GaMYB2 ha demostrado ser un homólogo funcional de *Arabidopsis* GLABRA1 (GL1), un regulador clave de la formación de tricomas de *Arabidopsis*. GaMYB2 se expresa en células de
 15 fibras de algodón en las fases tempranas del desarrollo (Wang, S. et al., 2004). Su promotor impulsa la expresión específica para tricomas también en *Arabidopsis* y la expresión específica para la cabeza de GST en el tabaco (Shangguan et al., 2008).

La patente de EE.UU. 7.626.081 describe un promotor específico para semillas de algodón encontrado en el gen de alfa globulina. El promotor *Gh-sp* se deriva de un gen de proteínas de las semillas y es activo sólo en las semillas de
 20 algodón en maduración (Song et al., 2000).

El promotor FBP7 de *Petunia* controla un factor de transcripción MADS-box y se sabe que es específico para las semillas. Se ha demostrado que plantas de algodón transformadas con una construcción de informador impulsada por el promotor FBP7 específicamente expresa dicho informador en la envuelta de las semillas (Pei et al., 2008).

A pesar del hecho de que actualmente existen muchos promotores que se sabe impulsan la expresión específica para semillas, específica para la envuelta de semillas o específica para tricomas en plantas de algodón, sería deseable tener más promotores específicos para semillas, específicos para la envuelta de semillas o específicos para tricomas disponibles para la expresión específica para las semilla en el algodón.
 25

Por consiguiente, en un aspecto, la presente solicitud da a conocer una secuencia de ácido nucleico (aislado) que comprende una secuencia de nucleótidos seleccionada de (a) SEQ ID NO: 1 o un fragmento del mismo, en donde dicho fragmento comprende al menos 600 nucleótidos consecutivos de SEQ ID NO: 1, comprendiendo el fragmento, además, un motivo de la caja T/G que tiene la secuencia de nucleótidos desde la posición de nucleótidos 298 a 303 de SEQ ID NO: 1 y un motivo de unión MYB que tiene la secuencia de nucleótidos desde la posición 846 a 851 de SEQ ID NO: 1 y en donde el fragmento, cuando se introduce en algodón, tiene actividad del promotor específico para la semilla con altos niveles de expresión en fibras desde el día 5 post-antesis hasta el día 30 post-antesis; (b)
 30 una secuencia de nucleótidos con al menos 95% de identidad de secuencia con la secuencia de nucleótidos de (a), comprendiendo el fragmento, además, un motivo de la caja T/G que tiene la secuencia de nucleótidos desde la posición de nucleótidos 298 a 303 de SEQ ID NO: 1 y un motivo de unión MYB que tiene la secuencia de nucleótidos desde la posición 846 a 851 de SEQ ID NO: 1 y que tiene actividad de promotor específico para la semilla con altos niveles de expresión en fibras desde el día 5 post-antesis hasta el día 30 post-antesis cuando se introduce en algodón; y (c) una secuencia de nucleótidos complementaria a la secuencia de nucleótidos de una cualquiera de (a) o (b).
 35
 40

A la secuencia de ácido nucleico (aislado) de este aspecto se la designa en adelante también la "secuencia del promotor".

A menos que se indique lo contrario, las realizaciones descritas más adelante para la secuencia del promotor descrita en esta memoria son aplicables a realizaciones respectivas de otros aspectos descritos en esta memoria.
 45

Tal como se utiliza en esta memoria, la expresión "que comprende" debe interpretarse como que especifica la presencia de las características, números enteros, etapas o componentes establecidas, pero no excluye la presencia o adición de una o más características, números enteros, etapas o componentes, o grupos de los mismos. Así, p. ej., un ácido nucleico que comprende una secuencia de nucleótidos puede comprender más nucleótidos que los realmente citados, es decir, embebidos en un ácido nucleico más grande. Un gen quimérico, como se describirá más adelante, que comprende un ácido nucleico que se define funcional o estructuralmente, puede comprender ácidos nucleicos adicionales, etc. Sin embargo, en el contexto de la presente divulgación, la expresión "que comprende" también incluye "consiste en".
 50

En otras palabras, la terminología relativa a un ácido nucleico "que comprende" una determinada secuencia de nucleótidos o una proteína que comprende una determinada secuencia de aminoácidos, tal como se utiliza a lo largo
 55

del texto, se refiere a un ácido nucleico o una proteína que incluye o contiene al menos la secuencia descrita, de modo que otras secuencias de nucleótidos o aminoácidos pueden incluirse en el extremo 5' (o N-terminal) y/o 3' (o C-terminal), p. ej., (la secuencia de nucleótidos de) una proteína marcadora seleccionable, de) un péptido de tránsito, y/o una secuencia 5' conductora o secuencia 3' conductora.

5 Los ácidos nucleicos pueden ser ADN o ARN, monocatenario o bicatenario. Los ácidos nucleicos pueden sintetizarse químicamente o producirse mediante expresión biológica in vitro o in vivo. Los ácidos nucleicos se pueden sintetizar químicamente utilizando fosforamiditas de ribonucleósidos apropiadamente protegidas y un sintetizador de ADN/ARN convencional. Proveedores de reactivos de síntesis de ARN son Prologo (Hamburg, Alemania), Dharmacon Research (Lafayette, CO, EE.UU.), Pierce Chemical (parte de Perbio Science, Rockford, IL, EE.UU.), Glen Research (Sterling, VA, EE.UU.), ChemGenes (Ashland, MA, EE.UU.) y Cruachem (Glasgow, Reino Unido).
10 En relación con el gen quimérico de la presente divulgación, el ADN incluye ADNc y ADN genómico.

15 Un "ácido nucleico aislado" o "secuencia de ácido nucleico aislado", tal como se utiliza en la presente solicitud, se refiere a un ácido nucleico tal como se define arriba que no se produce de forma natural (tal como un ácido nucleico artificial o sintético con una secuencia de nucleótidos diferente que el ácido nucleico que se produce de forma natural o un ácido nucleico que es más corto que uno que se produce de forma natural) o que ya no está en el entorno natural en el que estaba originalmente presente, p. ej., una secuencia codificante de ácidos nucleicos asociada con un elemento regulador heterólogo (tal como una secuencia codificante bacteriana enlazada operativamente a un promotor expresable en plantas) en un gen quimérico o un ácido nucleico transferido a otra
20 célula huésped tal como una célula vegetal transgénica.

La longitud de un fragmento de SEQ ID NO: 1 tal como se describe en esta memoria y su posición dentro de SEQ ID NO: 1 se ha de elegir de manera que sea suficientemente larga, p. ej., que comprenda todos los elementos necesarios y suficientes, y este situado de manera que sea capaz de inducir la expresión específica para la semilla, específica para la envuelta de la semilla o específica para tricomas.

25 Métodos de evaluar si una secuencia de ácido nucleico tal como se describe arriba, que en la presente solicitud representa una secuencia de promotor, es capaz de inducir la expresión de la secuencia codificante o un gen quimérico que está comprendido en o, en particular, de una secuencia de ácido nucleico enlazada operativamente a la misma, de una manera específica para la semilla, específica para la envuelta de la semilla o específica para tricomas son conocidos por la persona experta.

30 Por ejemplo, se pueden realizar estudios de gen informador con el fin de evaluar la función inductora de una secuencia de ácido nucleico. Un ejemplo incluye enlazar operativamente dicha primera secuencia de ácido nucleico a un gen informador tal como GUS, introducir la construcción de ácido nucleico resultante en una planta o célula vegetal tal como en una planta de algodón, y evaluar la inducción de la expresión de dicho gen informador en diferentes tejidos de dicha planta tal como también se describe con más detalle más adelante.

35 Dicho fragmento de la secuencia de ácido nucleico descrita en esta memoria y que tiene actividad del promotor específica para la semilla, específica para la envuelta de la semilla o específica para fibras de tricomas en algunos ejemplos puede comprender, por consiguiente, al menos 600, al menos 650, al menos 700, al menos 800, al menos 900, al menos 1000, al menos 1100, al menos 1200, al menos 1300 o al menos 1400 nucleótidos consecutivos de SEQ ID NO: 1. En otro ejemplo, dicho fragmento comprende la secuencia de nucleótidos desde la posición 1 hasta la posición 748 de SEQ ID NO: 1, en donde se encuentra la posición -1. En otro ejemplo, dicho fragmento comprende la secuencia de nucleótidos de SEQ ID NO: 1. Aún en otro ejemplo, dicha secuencia de ácidos nucleicos
40 consiste en SEQ ID NO: 1.

45 Sin embargo, resultará evidente que también se pueden utilizar para el mismo efecto variantes de la presente secuencia de nucleótidos, incluyendo inserciones, deleciones y sustituciones de los mismos. Generalmente, dichas variantes tienen una identidad de la secuencia de al menos 95% o incluso al menos 98% con la SEQ ID NO: 1 y conservan su actividad del promotor específica para la semilla, específica para la envuelta de la semilla o específica para fibras de tricomas.

50 Tal como se utiliza en esta memoria, el término "promotor" designa cualquier secuencia de ácido nucleico tal como la secuencia de ADN, que está reconocida y unida (directa o indirectamente) por una ARN polimerasa dependiente de ADN durante la iniciación de la transcripción, lo que resulta en la generación de una molécula de ARN que es complementaria al ADN transcrito; a esta región también se la puede aludir como una "región reguladora 5'". Los promotores están habitualmente ubicados aguas arriba de la región 5' no traducida (UTR) que precede a la secuencia codificante a transcribir y tienen regiones que actúan como sitios de unión para la ARN polimerasa II y otras proteínas tales como factores de transcripción para iniciar la transcripción de un gen enlazado operativamente.
55 Los promotores pueden contener por sí mismos sub-elementos (es decir, motivos de promotor) tales como

elementos cis o dominios potenciadores que regulan la transcripción de genes enlazados operativamente. El promotor y una 5' UTR conectada también se designan como 'región del promotor'.

Un promotor "específico para la semilla" en el contexto de la presente invención significa que la transcripción de una secuencia de ácido nucleico controlada por un promotor es al menos 5 veces mayor, al menos 10 veces mayor, al menos 20 veces mayor o al menos 50 veces mayor en una célula de semillas que en las células de cualquier otro tejido vegetal.

En un ejemplo, específico para la semilla significa específico para la envuelta de la semilla, es decir, no tiene lugar transcripción alguna en los tejidos derivados de los gametofitos de la semilla. En otro ejemplo, específico para la semilla o específico para la envuelta de la semilla significa específico para tricomas, es decir, no tiene lugar transcripción alguna en partes de la semilla o de la envuelta de la semilla distintas de tricomas. Tricomas incluyen fibras, p. ej., de una planta de algodón. Por consiguiente, la expresión "específico para la envuelta de la semilla" o "específico para tricomas" significa que la transcripción de una secuencia de ácido nucleico controlada por un promotor se efectúa de manera que la transcripción de dicho ácido nucleico en la semilla, la envuelta de la semilla, el tricoma o la fibra, respectivamente, es al menos 5 veces mayor, al menos 10 veces mayor, al menos 20 veces mayor o al menos 50 veces mayor que en células de cualquier otro tejido vegetal, preferiblemente tejido vegetal presente durante el desarrollo de semillas tal como durante el desarrollo de tricomas de la semilla.

Para la presente invención, el promotor también puede ser preferencial para la semilla. La expresión "preferencial para la semilla" (o "transcripción", que es equivalente) en el contexto de esta invención significa la transcripción de una secuencia de ácido nucleico mediante un elemento regulador de la transcripción tal como un promotor de manera que la transcripción de dicha secuencia de ácido nucleico en las semillas contribuye en más de 50%, preferiblemente en más de 60%, más preferiblemente en más de 70%, incluso más preferiblemente en más de 80% de toda la cantidad del ARN transcrito a partir de dicha secuencia de ácido nucleico en la planta entera durante cualquiera de sus fases de desarrollo.

La confirmación de la actividad del promotor para una secuencia del promotor o un fragmento de promotor funcional puede determinarse por los expertos en la técnica, por ejemplo utilizando una construcción de promotor-informador que comprende la secuencia de promotor unida operativamente a un marcador fácilmente puntuable explicado adicionalmente en esta memoria. La capacidad de la expresión específica para la semilla, específica para la envuelta de la semilla o específica para tricomas de los fragmentos o variantes del promotor descritos en esta memoria puede ser convenientemente testada enlazando operativamente estas secuencias de ácido nucleico a una secuencia de nucleótidos que codifica un marcador fácilmente evaluable, p. ej., un gen de beta-glucuronidasa, introducir un gen quimérico en una planta y analizar el patrón de expresión del marcador en semillas, la envuelta de la semilla o tricomas, en comparación con el patrón de expresión del marcador en otras partes de la planta. Candidatos para un marcador (o un gen informador) que no sea el GUS arriba mencionado son cloranfenicol-acetiltransferasa (CAT), beta-galactosidasa (beta-GAL), y proteínas con propiedades fluorescentes o fosforescentes tales como la proteína fluorescente verde (GFP) de *Aequora Victoria* o luciferasa. Para definir un promotor mínimo, una secuencia de ácido nucleico que representa el promotor está enlazada operativamente a la secuencia codificante de un gen marcador (informador) por técnicas de ADN recombinante bien conocidas en la técnica. El gen informador está enlazado operativamente aguas abajo del promotor, de modo que las transcripciones que se inician en el promotor prosiguen a través del gen informador. La casete de expresión que contiene el gen informador bajo el control del promotor puede introducirse en un tipo celular apropiado por técnicas de transformación bien conocidas en la técnica y descritas en otra parte en esta solicitud. Para el ensayo de la proteína informadora, se preparan lisados de células y se realizan ensayos adecuados, que son bien conocidos en la técnica, por la proteína informadora. Por ejemplo, si CAT era el gen informador de elección, los lisados de células transfectadas con construcciones que contienen CAT bajo el control de un promotor en estudio se mezclan con cloranfenicol marcado isotópicamente y acetil-coenzima A (acetil-CoA). La enzima CAT transfiere el grupo acetilo de la acetil-CoA a la posición 2 o 3 de cloranfenicol. La reacción se controla por cromatografía en capa fina, que separa cloranfenicol acetilado a partir de material que no ha reaccionado. Los productos de reacción se visualizan después por autorradiografía. El nivel de actividad enzimática corresponde a la cantidad de enzima que se hizo, que a su vez revela el nivel de expresión y la funcionalidad específica para la semilla, específica para la envuelta de la semilla o específica para tricomas del promotor o fragmento o variante del mismo. Este nivel de expresión se puede comparar también con otros promotores para determinar la fuerza relativa del promotor sometido a estudio. Una vez confirmada la actividad y la funcionalidad, análisis mutacionales y/o de inserción y/o de delección adicionales se pueden emplear para determinar p. ej., una región mínima y/o secuencias requeridas para iniciar la transcripción. Así, las secuencias se pueden eliminar en el extremo 5' de la región del promotor y/o en el extremo 3' de la región del promotor, o dentro de la secuencia del promotor y/o se pueden introducir sustituciones de nucleótidos. Estas construcciones se introducen de nuevo en las células y se determinan su actividad y/o funcionalidad.

En lugar de medir la actividad de una enzima informadora, la actividad (y funcionalidad) del promotor de la transcripción también se pueden determinar midiendo el nivel de ARN que se produce a partir de la secuencia codificante enlazada operativamente a un promotor o fragmento del mismo. Este nivel de ARN, tal como ARNm, se puede medir ya sea en un solo instante o en múltiples instantes y como tal, el aumento múltiple puede ser un aumento múltiple medio o un valor extrapolado derivado de los valores medidos experimentalmente. Como es una comparación de los niveles, se puede utilizar cualquier método que mida los niveles de ARNm. En un ejemplo, el

tejido u órganos comparados son una semilla o tejido de semilla con una hoja o tejido de hoja. En otro ejemplo, se comparan múltiples tejidos u órganos. Un ejemplo para comparaciones múltiples es una semilla o un tejido de la semilla o de las semillas comparado con 2, 3, 4 o más tejidos u órganos seleccionados del grupo que consiste en tejido floral, ápice floral, polen, hoja, embrión, brotes, primordios de hojas, ápice de los brotes, raíz, punta de la raíz, tejido vascular y cotiledón. Tal como se utiliza en esta memoria, ejemplos de órganos de la planta son semilla, hoja, raíz, etc. y ejemplo de tejidos son los primordios de las hojas, ápice de los brotes, tejido vascular, etc. La actividad o resistencia de un promotor puede ser medida en términos de la cantidad de ARNm o acumulación de proteína que produce específicamente, respecto a la cantidad total de ARNm o proteína. El promotor expresa una secuencia de ácido nucleico operativamente unida, por ejemplo, a un nivel mayor que aproximadamente 0,1%, aproximadamente 0,2%, mayor que aproximadamente 0,5, 0,6, 0,7, 0,8 o aproximadamente 0,9%, mayor que aproximadamente 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8% o aproximadamente 9%, o mayor que aproximadamente 10% del ARNm total de la célula en la que está contenida. Alternativamente, la actividad o la resistencia de un promotor puede expresarse con relación a un promotor bien caracterizado (para el que fue evaluada previamente la actividad transcripcional) o la resistencia en un tejido específico puede ser expresada con respecto a la de otro tejido.

Se describen promotores específicos para la semilla, específicos para la envuelta de la semilla o específicos para tricomas que comprenden una secuencia de nucleótidos que tiene al menos 95% o al menos 98% de identidad de secuencia con SEQ ID NO: 1 o un fragmento de la misma tal como se define anteriormente. Variantes que se producen de forma natural del promotor descrito en esta memoria se pueden identificar con el uso de técnicas de biología molecular bien conocidas tales como, por ejemplo, con la reacción en cadena de la polimerasa (PCR) y técnicas de hibridación como se ha esbozado antes en esta memoria. Tales secuencias de ácidos nucleicos también incluyen secuencias de ácidos nucleicos derivados sintéticamente tales como las generadas, por ejemplo, mediante el uso de mutagénesis dirigida al sitio de SEQ ID NO: 1 o un fragmento del mismo. En general, variantes de la secuencia de nucleótidos tendrán al menos 95%, 96%, 97% a 98% y 99% de identidad de secuencia con la secuencia de ácido nucleico de SEQ ID NO: 1. Derivados de las secuencias de ácido nucleico descritas en esta memoria y que tienen la identidad de secuencia requerida pueden incluir, pero no se limitan a delecciones de secuencias, mutaciones puntuales simples o múltiples, alteraciones en un sitio de reconocimiento de enzima de restricción particular, adición de elementos funcionales, u otros medios de modificación molecular que pueden mejorar o alterar de otro modo la expresión del promotor. Técnicas para obtener tales derivados son bien conocidas en la técnica (véase, por ejemplo, J.F. Sambrook, D.W. Russell y N. Irwin (2000) *Molecular Cloning: A Laboratory Manual*). Por ejemplo, un experto ordinario en la técnica puede delimitar los elementos funcionales dentro de los promotores descritos en esta memoria y eliminar cualesquiera elementos no esenciales. Los elementos funcionales pueden ser modificados o combinados para aumentar la utilidad o la expresión de las secuencias de la invención para cualquier aplicación particular. Los expertos en la técnica están familiarizados con los materiales de recurso estándares que describen condiciones y procedimientos específicos para la construcción, manipulación y aislamiento de macromoléculas (p. ej., moléculas de ADN, plásmidos, etc.), así como la generación de organismos recombinantes y el rastreo y aislamiento de moléculas de ADN.

La secuencia del promotor de SEQ ID NO: 1 y sus fragmentos funcionales y variantes puede ser alterada, p. ej., para que contenga, p. ej., "ADN de potenciador" para ayudar a elevar la expresión génica. Como es bien conocido en la técnica, se pueden utilizar determinados elementos de ADN para potenciar la transcripción del ADN. Estos potenciadores se encuentran a menudo en 5' con respecto al inicio de la transcripción en un promotor que funciona en células eucariotas, pero a menudo pueden ser insertados aguas arriba (5') o aguas abajo (3') a la secuencia codificante. En algunos casos, estos elementos de ADN potenciador son intrones. Entre los intrones que son útiles como ADN potenciador se encuentran los intrones en posición 5' con respecto al gen actina de arroz 1 (véase el documento US5641876), el gen actina de arroz 2, el gen de alcohol deshidrogenasa de maíz, el gen proteína choque térmico 70 de maíz (véase el documento US5593874), el gen shrunken 1 de maíz, el gen 1 sensible a la luz de *Solanum tuberosum* y el gen proteína choque térmico 70 de maíz de *Petunia hybrida* (véase el documento US 5659122). Por lo tanto, tal como se contempla en esta memoria, un promotor o una región promotora incluye variaciones de promotores derivados al insertar o eliminar regiones reguladoras, someter el promotor a mutagénesis al azar o dirigida al sitio, etc. La actividad o la resistencia de un promotor se pueden medir en términos de las cantidades de ARN que produce, o la cantidad de acumulación de proteínas en una célula o tejido, con relación a un promotor cuya actividad transcripcional ha sido evaluada previamente tal como se describió anteriormente.

Tal como se utiliza en esta memoria, la expresión "porcentaje de identidad de secuencia" se refiere al porcentaje de nucleótidos idénticos entre dos segmentos de una ventana de ADN alineados de manera óptima. La alineación óptima de secuencias para alinear una ventana de comparación es bien conocida para los expertos en la técnica y puede realizarse mediante herramientas tal como el algoritmo de homología local de Smith y Waterman (Waterman, M.S., Chapman & Hall. Londres, 1995), el algoritmo de alineación de homología de Needleman y Wunsch (1970), el método de búsqueda de similitud de Pearson y Lipman (1988) y, preferiblemente, mediante implementaciones computarizadas de estos algoritmos tales como GAP, BESTFIT, FASTA y TFASTA, disponibles como parte de la GCG (marca registrada), Wisconsin Package (marca registrada de Accelrys Inc., San Diego, Calif.). Una "fracción de identidad" para segmentos alineados de una secuencia de ensayo y una secuencia de referencia es el número de componentes idénticos que son compartidos por las dos secuencias alineadas, dividido por el número total de componentes en el segmento de la secuencia de referencia, es decir, toda la secuencia de referencia o una parte definida menor de la secuencia de referencia. El porcentaje de identidad de secuencia se representa como los

tiempos de la fracción de identidad 100. La comparación de una o más secuencias de ADN puede ser a una secuencia de ADN de longitud completa o a una porción de la misma, o a una secuencia de ADN más larga.

5 El término "hibridación" se refiere a la capacidad de una primera cadena de ácido nucleico de unirse con una segunda cadena a través del apareamiento de bases del enlace hidrógeno cuando las dos cadenas de ácido nucleico tienen suficiente identidad de secuencia. La hibridación se produce cuando las dos moléculas de ácido nucleico se hibridan entre sí en condiciones apropiadas. La hibridación de ácidos nucleicos es una técnica bien conocida por los expertos en la técnica de la manipulación de ADN. La propiedad de hibridación de un par de ácidos nucleicos dada es una indicación de su similitud o identidad. Otra indicación de que dos secuencias de ácido nucleico son ampliamente idénticas es que las dos moléculas se hibridan entre sí en condiciones rigurosas. 10 "Condiciones de hibridación rigurosas" y "condiciones de lavado de hibridación rigurosas" en el contexto de experimentos de hibridación de ácidos nucleicos tales como la hibridación Southern y Northern son dependientes de la secuencia y son diferentes bajo diferentes parámetros ambientales. Un ejemplo de condiciones de lavado altamente rigurosas es 0,1 X SSC, solución 5X de Denhardt, SDS al 0,5% a 65°C durante, p. ej., aproximadamente 15 minutos. Un ejemplo de condiciones de lavado apropiadas para la presente invención es 2 X SSC, SDS al 0,1% lavado a 65°C durante, p. ej., aproximadamente 15 minutos. A menudo, un lavado de alta rigurosidad es precedido por un lavado de baja rigurosidad para eliminar la señal de fondo de la sonda. Las condiciones rigurosas también se pueden lograr con la adición de agentes desestabilizantes tales como formamida. En general, una relación de señal a ruido de 2 X (o superior) que la observada para una sonda no relacionada en el ensayo de hibridación particular indica detección de una hibridación específica. Condiciones muy rigurosas se seleccionan para que sean iguales a la T_m para una sonda particular. 20

En el curso de la presente invención, un promotor específico para la semilla, junto con una 5' UTR subsiguiente (juntos también designados como 'región del promotor') ha sido identificado en *Gossypium hirsutum*. Dicho promotor, también denominado pMADS6 o GhpMADS6, en su contexto natural, controla el gen MADS-box, MADS6, en algodón. 25

G. hirsutum es un alotetraploide que resulta de la fusión de dos especies diploides ancestrales de hace 1-1,2 millones de años y el genoma puede contener hasta cuatro alelos de cada uno de los genes. Considerando que al menos 100 genes MADS-box están presentes en el genoma de *Arabidopsis* (Parenicová et al. 2003), con un número similar en álamo (Leseberg et al. 2006), petunia (Immink et al. 2003) y arroz (Nam et al. 2004), es probable que la familia de genes MADS-box en algodón sea grande y compleja, con un alto nivel de homología y/o redundancia funcional. 30

El pMADS6 promotor tiene una identidad de secuencia muy baja de 46% con el promotor FBP7 de petunia y se expresa altamente en el óvulo, la fibra y en el tejido de la flor. Se ha demostrado que la expresión es particularmente fuerte durante el desarrollo temprano de la fibra; la expresión del producto génico se detectó hasta un máximo de 24 DPA (Lightfoot et al., 2008). 35

Se espera que el promotor actual sea adecuado para la expresión específica para la semilla, específica para la envuelta de la semilla o específica para tricomas o la expresión al menos durante las primeras fases del desarrollo de fibras de transgenes en algodón. El promotor actual se puede utilizar para expresar genes que modifican propiedades de la fibra de algodón o de otra manera están implicados en la formación de fibras de algodón tales como los genes implicados en la síntesis de auxinas. El presente promotor también podría ser más fácil controlable, ya que se deriva de la planta en la que está destinado a ser reintroducido. Alternativamente, o además, se pueden evitar potenciales efectos secundarios desconocidos o no deseados de promotores heterólogos mediante el uso del presente promotor en algodón. 40

A menos que se indique lo contrario, definiciones específicas o características específicas de determinados ejemplos descritos en la presente solicitud en relación con un aspecto pueden ser introducidas en cualquier otro aspecto descrito en esta memoria. 45

Se identificó un cierto número de supuestos elementos de respuesta en la secuencia del promotor descrita en esta memoria. La búsqueda se limitó a elementos específicos para tricomas y a un motivo que corresponde a una caja L1 y a motivos de unión a MYB. Los dos últimos se han descrito como motivos que confieren en potencia una expresión específica para tricomas (Wang y Chen, 2004). La búsqueda reveló dos motivos que confieren en potencia una expresión específica para semilla, específica para envuelta de semilla o específica para tricomas. El primero es una caja T/G que corresponde al motivo de tricomas RPSP01178 situado comenzando en la posición 298 en SEQ ID NO: 1 (correspondiente a la posición -451). La secuencia exacta del motivo de la caja T/G es AACGTG. Dicho motivo de unión ha sido identificado en el promotor de un gen MYB de la fibra de algodón (Shangguan et al., 2008), en donde la delección redujo la actividad del promotor en *Arabidopsis* y tabaco. 50

El otro motivo de unión corresponde a un motivo de unión MYB y se encuentra comenzando en la posición 846 en SEQ ID NO: 1 y tiene la secuencia cagtta. Curiosamente, dicho motivo de unión MYB también está presente dentro de la secuencia codificante del gen MADS6 regulado de forma natural por el presente promotor. Se ha identificado por Wang y Chen (2004) y se encuentra en al menos el promotor RDL1, en donde confiere especificidad para 55

tricomas en Arabidopsis, y en los promotores GL1 (que controla el gen myb en Arabidopsis) y GaMyb2 (que controla el gen MYB en algodón). Se ha demostrado anteriormente que la interrupción del motivo de unión MYB conduce a una reducción en la producción de tricomas.

5 Variantes del promotor descritas en esta memoria incluyen las que comprenden ambos elementos identificados, pero por lo demás se han modificado para eliminar tramos de nucleótidos dentro de la secuencia que no son necesarios para que el promotor sea funcional de una manera específica para la semilla, específica para la envuelta de la semilla o incluso específica para tricomas o fibras. Por ejemplo, cualquier tramo de nucleótidos situado entre los dos motivos identificados y/o entre el inicio de la transcripción y el primer motivo puede ser al menos parcialmente eliminado para dar lugar a una secuencia de nucleótidos más corta que la secuencia de
10 aproximadamente 1,5 Kb representada en SEQ ID NO: 1.

Se espera que la secuencia de nucleótidos del presente promotor, así como fragmentos y variantes del mismo tal como se define anteriormente, ejerza una actividad específica para la semilla, específica para la envuelta de la semilla o incluso específica para tricomas o fibras.

15 En un ejemplo, la actividad del promotor específica para la semilla está en algodón. Otros ejemplos para los que se pueden tratar el promotor, fragmentos y variantes de los mismos incluyen otras plantas productoras de tricomas o de fibras tales como cáñamo, yute, lino y plantas leñosas, incluyendo, pero no limitadas a *Pinus spp.*, *Populus spp.*, *Picea spp.*, *Eucalyptus spp.* etc.

En un ejemplo de la secuencia de ácido nucleico descrita en esta memoria, la actividad del promotor específica para la semilla es específica para tricomas o específica para fibras.

20 En otro aspecto, la presente solicitud describe un gen quimérico que comprende el ácido nucleico (aislado) descrito en esta memoria, enlazado operativamente a un ácido nucleico que codifica un producto de expresión de interés y, opcionalmente, una secuencia de terminación de la transcripción y de poliadenilación funcional en células vegetales.

Un gen quimérico es un gen artificial construido enlazando operativamente fragmentos de genes no relacionados u otras secuencias de ácidos nucleicos. En otras palabras "gen quimérico" designa un gen que no se encuentra normalmente en una especie vegetal o se refiere a cualquier gen en el que el promotor, partes adjuntas del promotor o una o más de otras regiones reguladoras del gen no están asociados en la naturaleza con una parte o la totalidad del ácido nucleico transcrito enlazado operativamente con el mismo, es decir, son heterólogos con respecto al ácido nucleico transcrito. Más particularmente, un gen quimérico que es un gen artificial, es decir, que no se produce de forma natural, el gen producido por un enlace operativo de la secuencia de ácido nucleico de la invención tal como, p. ej. el ácido nucleico de SEQ ID NO: 1, un fragmento de 600 nt del mismo o una secuencia de ácido nucleico que tiene al menos 95% de identidad de secuencia con la misma, todos ellos capaces de dirigir una expresión específica para la semilla, específica para la envuelta de la semilla o específica para tricomas-fibras de un producto de expresión de interés como se describe anteriormente, con una segunda secuencia de ácido nucleico que codifica dicho producto de expresión de interés que no está enlazado operativamente de forma natural a dicha secuencia de ácido nucleico. Una secuencia de ácido nucleico enlazada operativamente de forma natural a dicha secuencia de ácido nucleico es la secuencia codificante del gen MADS6 de algodón.

El término "heterólogo" se refiere a la relación entre dos o más secuencias de ácido nucleico o proteína que se derivan de diferentes fuentes. Por ejemplo, un promotor es heterólogo con respecto a una secuencia de ácido nucleico enlazada operativamente, tal como una secuencia codificante, si dicha combinación no se encuentra normalmente en la naturaleza. Además, una secuencia particular puede ser "heteróloga" con respecto a una célula u organismo en el que se inserta (es decir, no se produce de forma natural en esa célula u organismo particular). Por ejemplo, el gen quimérico descrito en esta memoria es un ácido nucleico heterólogo.

La expresión "enlazada operativamente" se refiere a la disposición espacial funcional de dos o más regiones de ácido nucleico o secuencias de ácido nucleico. Por ejemplo, una región del promotor puede estar situada con relación a una secuencia de ácido nucleico que codifica un producto de expresión de interés, de manera que la transcripción de dicha secuencia de ácido nucleico está dirigida por la región del promotor. Por lo tanto, una región del promotor está "enlazada operativamente" a la secuencia de ácido nucleico.

El promotor, fragmento o variante del mismo tal como se describe arriba pueden ser enlazados operativamente a una secuencia de ácido nucleico que codifica un producto de expresión de interés que es heterólogo con respecto al promotor. Generalmente, la secuencia de ácido nucleico puede ser cualquier secuencia de ácido nucleico para la que se desea un nivel alterado tal como un nivel de transcripción incrementado. La secuencia de ácido nucleico puede codificar, por ejemplo, un polipéptido que es capaz de modificar propiedades de la fibra en algodón o que está implicado en la biosíntesis de auxinas.

Las auxinas son una clase de hormonas vegetales que juegan un papel esencial en la coordinación de muchos procesos de crecimiento y de comportamiento en el ciclo vital de las plantas. En el nivel molecular, las auxinas tienen un anillo aromático y un grupo carboxílico (Taiz y Zeiger, 1998). El miembro más importante de la familia de auxinas es el ácido indol-3-acético (IAA). Genera la mayor parte de los efectos de auxinas en plantas intactas, y es la auxina nativa más potente.

Secuencias de ácido nucleico heterólogas adecuadas adicionales para modificar las propiedades de las fibras de algodón incluyen, sin limitación, las descritas en el documento WO02/45485, con lo que la calidad de la fibra en plantas productoras de fibras tales como algodón, se modifica modulando la actividad de sacarosa sintasa y/o la expresión en tales plantas, los ácidos nucleicos que median en una alteración de una fase de alargamiento de las células de las fibras modulando la deposición de callosa tal como se describe en el documento WO2005/017157, en particular un gen que codifica una proteína β -1,3 glucano sintasa, o en el documento WO2006/136351.

Un "producto de expresión" designa un producto intermedio o final que surge de la transcripción y, opcionalmente, la traducción del ácido nucleico tal como ADN o ARN, que codifica dicho producto. Durante el proceso de transcripción, una secuencia de ADN bajo el control de regiones reguladoras, en particular la secuencia de promotor descrita en esta memoria, se transcribe en una molécula de ARN. Una molécula de ARN puede formar por sí misma un producto de expresión y entonces es capaz, por ejemplo, de interactuar con otro ácido nucleico o proteína. Alternativamente, una molécula de ARN puede ser un producto intermedio cuando es capaz de ser traducida en un péptido o una proteína. Se dice que un gen codifica una molécula de ARN como producto de expresión cuando el ARN como el producto final de la expresión del gen es capaz de interactuar con otro ácido nucleico o proteína. Ejemplos de productos de expresión de ARN incluyen ARNs inhibidores tales como, p. ej., ARN sentido, ARN antisentido, ARN de horquilla, ribozimas, ARNm_i o ARNs_i, ARNm, ARNr y ARNt. Se dice que un gen codifica una proteína o péptido como producto de expresión cuando el producto final de la expresión del gen es una proteína o un péptido.

Productos de expresión ilustrativos adicionales de interés incluyen proteínas implicadas en la síntesis de la pared celular y la formación de fibras tal como se describe en el documento WO2005/017157, en particular un gen que codifica una proteína β -1,3 glucano sintasa, o en el documento WO2006/136351, PCT/EP2011/ 004929 o WO2011/089021, en particular una N-acetilglucosamina transferasa que puede ser dirigida a las membranas del aparato de Golgi tal como una N-acetilglucosamina transferasa del tipo NODC, o una quitina sintasa.

Dentro del alcance de la presente divulgación, puede hacerse uso también, en combinación con el gen quimérico descrito anteriormente, de otras secuencias reguladoras, que están situadas entre dicha secuencia de ácido nucleico que comprende un promotor y dicha secuencia de ácido nucleico que comprende la secuencia codificante del producto de expresión. Este es especialmente el caso si la secuencia de nucleótidos utilizada como promotor es desde la posición 1 hasta la posición 748 de SEQ ID NO: 1 que corresponde al promotor sin 5' UTR. Ejemplos no limitantes de tales secuencias reguladoras incluyen activadores de traducción ("potenciadores"), por ejemplo el activador de traducción del virus del mosaico del tabaco (TMV) descrito en la solicitud WO 87/07644, o del virus del grabado del tabaco (TEV) descrito por Carrington y Freed 1990, J. Virol. 64: 1590-1597, o intrones tales como el intrón de la histona 3 de Arabidopsis (Chaubet et al., 1992).

Otras secuencias reguladoras adecuadas incluyen 5' UTRs. Tal como se utiliza en esta memoria, una 5' UTR, a la que también se alude como secuencia conductora, es una región particular de un ARN mensajero (ARNm) situada entre el sitio de inicio de la transcripción y el codón de inicio de la región codificante. Está implicado en la estabilidad del ARNm y la eficiencia de la traducción. Por ejemplo, el conductor 5' no traducido de un gen de proteína de unión a/b de clorofila de petunia (cab22L) aguas abajo del sitio de inicio de la transcripción 35S se puede utilizar para aumentar los niveles de estado estacionario de la expresión del gen informador (Harpster et al., 1988, Mol Gen Genet. 212(1): 182-90). El documento WO95/006742 describe el uso de secuencias conductoras no traducidas 5' derivadas de genes que codifican proteínas de choque térmico para aumentar la expresión del transgen.

El gen quimérico puede comprender también una secuencia de terminación de la transcripción o una secuencia de poliadenilación operativa en una célula vegetal, en particular una célula de la planta de algodón. Como una secuencia de terminación de la transcripción o de poliadenilación puede hacerse uso de cualquier secuencia correspondiente de origen bacteriano tal como, por ejemplo, el terminador nos de *Agrobacterium tumefaciens* de origen viral tales como, por ejemplo, el terminador CaMV 35S, o de origen vegetal tal como, por ejemplo, un terminador de histona tal como se describe en la solicitud de patente publicada EP 0 633 317 A1.

En un ejemplo del gen quimérico descrito en esta memoria, dicho producto de expresión de interés es una proteína, un péptido o una molécula de ARN, dicha molécula de RNA es capaz de modular la expresión de un gen endógeno a dicha planta. En un ejemplo, dicha proteína, péptido o molécula de ARN es capaz de modular una propiedad de la fibra. En otro ejemplo, dicha proteína, péptido o molécula de ARN está implicado en la biosíntesis de la auxina.

El término "proteína", tal como se utiliza en esta memoria, describe un grupo de moléculas que consiste en más de 30 aminoácidos, mientras que el término "péptido" describe moléculas que consisten en hasta 30 aminoácidos. Las proteínas y los péptidos pueden formar, además, dímeros, trímeros y oligómeros superiores, es decir, consisten en más de una molécula (poli)peptídica. Moléculas de proteínas o péptidos que forman tales dímeros, trímeros, etc., pueden ser idénticos o no idénticos. Las correspondientes estructuras de orden superior se denominan, por consiguiente, homodímeros o heterodímeros, homotrímeros o heterotrímeros, etc. Los términos "proteína" y "péptido" también se refieren a proteínas o péptidos modificados de forma natural en los que la modificación se efectúa, p. ej., mediante glicosilación, acetilación, fosforilación y similares. Tales modificaciones son bien conocidas en la técnica.

Proteínas a modo de ejemplo, adecuadas como productos de expresión, incluyen proteínas implicadas en la modificación de propiedades de la fibra tales como las que median en un aumento en la longitud de la fibra, una alteración en la resistencia de la fibra o una alteración en las propiedades de la pared celular resultante, p. ej., en una carga alterada de dichas paredes celulares, una alteración en la capacidad de tinción, una producción de fibra de pelusa disminuida, una alteración en la relación de madurez de la fibra, una disminución en el contenido de fibra inmadura o un aumento en la uniformidad de la fibra y micronaire.

Dicho producto de expresión de interés también puede ser una molécula de ARN capaz de modular la expresión de un gen endógeno a dicha planta de algodón.

Ejemplos de genes diana adecuados para los productos de expresión de ARN a este respecto incluyen los implicados en la modificación de propiedades de la fibra, tales como los que median en un incremento en la longitud de la fibra, una alteración en la resistencia de la fibra o una alteración en las propiedades de la pared celular resultante, p. ej., en una carga alterada de dichas paredes celulares, una alteración en la capacidad de tinción, producción de fibra de pelusa disminuida, una alteración en la relación de madurez de la fibra, una disminución en el contenido de fibra inmadura o un incremento en la uniformidad de la fibra y micronaire.

Para el caso de moléculas de ARN, resultará claro que siempre que las secuencias de nucleótidos de las moléculas de ARN se definan con referencia a la secuencia de nucleótidos de moléculas de ADN correspondiente, la timina (T) en la secuencia de nucleótidos debe ser sustituida por uracilo (U). El que se haga referencia a moléculas de ARN o ADN resultará claro por el contexto de la solicitud.

La expresión "capaz de modular la expresión de un gen" se refiere a la acción de una molécula de ARN, tal como una molécula de ARN inhibidora tal como se describe en esta memoria, para influir sobre el nivel de expresión de genes diana de diferentes maneras. Esto puede efectuarse, p. ej., mediante la inhibición de la expresión de un gen diana mediante la interacción directa con los componentes que impulsan dicha expresión tal como el gen propiamente dicho o el ARNm transcrito que resulta en una disminución de la expresión, o mediante la inhibición de otro gen implicado en la activación de la expresión de un gen diana, aboliendo de este modo dicha activación, o inhibiendo otro gen implicado en la inhibición de la expresión de un gen diana que resulta en un aumento de la expresión. La inhibición de un gen implicado en la inhibición de la expresión de un gen diana utilizando ARN inhibidor puede resultar, por el contrario, en una activación de la expresión de dicho gen diana.

Moléculas de ARN inhibidoras disminuyen los niveles de ARNms de sus proteínas diana disponibles para la traducción a dicha proteína diana. De esta manera, se puede inhibir la expresión de proteínas implicadas en las respuestas no deseadas a condiciones de estrés. Esto se puede lograr a través de técnicas bien establecidas, incluyendo la co-supresión (supresión de ARN sentido), ARN antisentido, ARN de doble cadena (ARNds), ARNsi o microARN (miARN).

Una molécula de ARN como producto de expresión tal como se describe en esta memoria comprende una parte de una secuencia de nucleótidos que codifica una proteína diana o una secuencia homóloga para regular a la baja la expresión de dicha proteína diana. Otro ejemplo de una molécula de ARN como producto de expresión para uso en la expresión de regulación a la baja son moléculas de ARN antisentido que comprenden una secuencia de nucleótidos complementaria a al menos una parte de un nucleótido que codifica una proteína de interés o una secuencia homóloga. Aquí, la regulación a la baja puede efectuarse, p. ej., introduciendo este ARN antisentido o un ADN quimérico que codifica dicha molécula de ARN. Todavía en otro ejemplo, la expresión de una proteína de interés se regula a la baja introduciendo una molécula de ARN de doble cadena que comprende una región de ARN sentido y una región de ARN antisentido correspondiente a y respectivamente complementaria a al menos parte de una secuencia génica que codifica dicha proteína de interés, región de ARN sentido y antisentido que son capaces de formar una región de ARN de doble cadena entre sí. Una molécula de ARN de doble cadena de este tipo puede ser codificada tanto por moléculas sentido como antisentido tal como se describe anteriormente y por una molécula de una sola cadena que es procesada para formar ARNsi o ARNmi.

En un ejemplo, la expresión de una proteína diana puede ser regulada a la baja introduciendo una construcción de ADN quimérico que proporciona una molécula de ARN sentido capaz de regular a la baja la expresión mediante la co-supresión. La región de ADN transcrita proporcionará, después de la transcripción, una denominada molécula de ARN sentido capaz de reducir la expresión de un gen que codifica una proteína diana en la planta o célula vegetal diana de una manera transcripcional o post-transcripcional. La región de ADN transcrito (y la molécula de ARN resultante) comprende al menos 20 nucleótidos consecutivos que tienen al menos 95% de identidad de secuencia con la porción correspondiente de la secuencia de nucleótidos que codifica la proteína diana presente en la planta o célula vegetal.

Alternativamente, un producto de expresión para regular a la baja la expresión de una proteína diana es una molécula de ARN antisentido. Regular a la baja o reducir la expresión de una proteína de interés en la planta de algodón o célula vegetal diana se efectúa de nuevo de una manera transcripcional o post-transcripcional. La región de ADN transcrita (y la molécula de ARN resultante) comprende al menos 20 nucleótidos consecutivos que tienen al menos 95% de identidad de la secuencia con el complemento de la parte correspondiente de la secuencia de ácido nucleico que codifica dicha proteína diana presente en la planta o célula vegetal.

Sin embargo, la secuencia de nucleótidos mínima de la región de ARN antisentido o sentido de aproximadamente 20 nt de la secuencia de ácido nucleico que codifica una proteína diana puede estar comprendida dentro de una molécula de ARN mayor, variando en tamaño de 20 nt a una longitud igual al tamaño del gen diana. Las regiones de nucleótidos antisentido o sentido mencionados pueden ser, por lo tanto, de aproximadamente 21 nt a aproximadamente 5000 nt de longitud tal como 21 nt, 40 nt, 50 nt, 100 nt, 200 nt, 300 nt, 500 nt, 1000 nt, 2000 nt o incluso aproximadamente 5000 nt o mayores en longitud. Además de ello, no se requiere para el propósito de la invención que la secuencia de nucleótidos de la molécula de ARN inhibidora utilizada o la región codificante del transgen, sea completamente idéntica o complementaria al gen endógeno que codifica la proteína diana, cuya expresión está dirigida a ser reducida en la célula vegetal. Cuanto más larga sea la secuencia, menos riguroso será el requisito para la identidad de secuencia global. Por lo tanto, las regiones sentido o antisentido pueden tener una identidad de secuencia global de aproximadamente 40% o 50% o 60% o 70% u 80% o 90% o 100% a la secuencia de nucleótidos de un gen endógeno o el complemento de la misma. Sin embargo, tal como se menciona, las regiones antisentido o sentido deben comprender una secuencia de nucleótidos de 20 nucleótidos consecutivos que tiene de aproximadamente 95 a aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos del gen endógeno que codifica el gen diana. El tramo de aproximadamente 95 a aproximadamente 100% de identidad de secuencia puede ser de aproximadamente 50, 75 o 100 nt.

La eficiencia de los genes quiméricos mencionados anteriormente para la regulación a la baja del nivel de expresión génica mediada por ARN antisentido o ARN sentido puede potenciarse adicionalmente por la inclusión de elementos de ADN que resulta en la expresión de moléculas de ARN inhibitoras aberrantes, no poliadeniladas. Un elemento de ADN de este tipo, adecuado para ese propósito, es una región de ADN que codifica una ribozima de auto-corte y empalme. La eficiencia puede potenciarse también proporcionando las moléculas de ARN generadas con señales de localización o retención nuclear.

Además, un producto de expresión tal como se describe en esta memoria puede ser una secuencia de ácido nucleico que proporciona una molécula de ARN de doble cadena, capaz de regular a la baja la expresión de un gen que codifica una proteína diana. Tras la transcripción de la región de ADN, el ARN es capaz de formar una molécula de ARNs a través de apareamiento convencional de bases entre una región sentido y antisentido, en que la región sentido y antisentido son secuencias de nucleótidos tal como se describe anteriormente en esta memoria. Los productos de expresión, que son ARNs de acuerdo con la invención pueden comprender, además, un intrón tal como un intrón heterólogo, situado, p. ej., en la secuencia de espaciador entre las regiones de ARN sentido y antisentido de acuerdo con la divulgación del documento WO 99/53050. Para lograr la construcción de un transgen de este tipo, se puede hacer uso de los vectores descritos en el documento WO 02/059294 A1.

En un ejemplo, dicha molécula de ARN comprende una primera y una segunda región de ARN, en donde 1. dicha primera región de ARN comprende una secuencia de nucleótidos de al menos 19 nucleótidos consecutivos que tiene al menos aproximadamente 94% de identidad de secuencia con la secuencia de nucleótidos de dicho gen endógeno; 2. dicha segunda región de ARN comprende una secuencia de nucleótidos complementaria a dichos 19 nucleótidos consecutivos de dicha primera región de ARN; 3. dicha primera y segunda región de ARN son capaces de apareamiento de bases para formar una molécula de ARN de doble cadena entre al menos dichos 19 nucleótidos consecutivos de dichas primera y segunda región. En otros ejemplos, se aplican las mismas consideraciones que se ha descrito anteriormente para el ARN sentido y antisentido.

Otro producto de expresión a modo de ejemplo es una molécula de microARN (mirARN, que puede ser procesado a partir de una molécula de pre-microARN) capaz de guiar la escisión del ARNm transcrito a partir del ADN que codifica la proteína diana que ha de ser traducido a dicha proteína diana. Moléculas de miARN se pueden introducir convenientemente en células vegetales a través de la expresión de un gen quimérico tal como se describe en esta

memoria que comprende una (segunda) secuencia de ácido nucleico que codifica, como producto de expresión de interés dicha transcripción de miARN, pre-miARN o miARN primaria.

Las miARNs son pequeños ARNs endógenos que regulan la expresión génica en plantas, pero también en otros eucariotas. Tal como se utiliza en esta memoria, un "miARN" es una molécula de ARN de aproximadamente 20 a 30 nucleótidos (Siomi y Siomi, 2009) de longitud, que se pueden cargar en un complejo RISC y dirigir la escisión de una molécula de ARN diana, en donde la molécula de ARN diana comprende una secuencia de nucleótidos esencialmente complementaria a la secuencia de nucleótidos de la molécula de miARN. En el ejemplo de miARNs, se pueden producir uno o más de los siguientes apareamientos erróneos en el miARN esencialmente complementarios con el ARN diana:

- 10 - Un apareamiento erróneo entre el nucleótido en el extremo 5' de dicho miARN y la secuencia de nucleótidos correspondiente en la molécula de ARN diana;
- Un apareamiento erróneo entre uno cualquiera de los nucleótidos en la posición 1 a la posición 9 de dicho miARN y la secuencia de nucleótidos correspondiente en la molécula de ARN diana;
- 15 - Tres apareamientos erróneos entre uno cualquiera de los nucleótidos en la posición 12 a la posición 21 de dicho miARN y la secuencia de nucleótidos correspondiente en la molécula de ARN diana siempre que no haya más de dos apareamientos erróneos consecutivos;
- No se permite un apareamiento erróneo en las posiciones 10 y 11 del miARN (todas las posiciones del miARN se indican a partir del extremo 5' de la molécula de miARN).

Un ejemplo adicional de un producto de expresión capaz de regular a la baja la expresión de una proteína diana es codificado por una secuencia de ácido nucleico que produce una molécula de ARN pre-miARN que se procesa en un miARN capaz de guiar la escisión del ARNm que codifica dicha proteína diana. En las plantas, miARNs son procesados a partir de las regiones de tallo-bucle de largos pre-miARNs endógenos por la actividad de escisión de DICERLIKE1 (DCL1). miARNs de plantas son altamente complementarios a ARNm diana conservados, y guían la escisión de sus dianas. Los miRNAs parecen ser componentes clave en la regulación de la expresión génica de redes complejas de las vías implicadas, entre otras cosas, en el desarrollo. Tal como se utiliza en esta memoria, una molécula de "pre-miARN" es una molécula de ARN de aproximadamente 100 a aproximadamente 200 nucleótidos, de preferencia de aproximadamente 100 a aproximadamente 130 nucleótidos, que puede adoptar una estructura secundaria que comprende un tallo de ARNs y un bucle de ARN de cadena sencilla y que comprende, además, la secuencia de nucleótidos del miARN y su secuencia de complemento del miARN en el tallo de ARN de doble cadena. Preferiblemente, el miARN y su complemento se encuentran a una distancia de aproximadamente 10 a aproximadamente 20 nucleótidos de los extremos libres del tallo de miARN ARNs. La longitud y la secuencia de la región de bucle de cadena sencilla no son críticas y pueden variar considerablemente, p. ej., entre 30 y 50 nt de longitud. Preferiblemente, la diferencia en la energía libre entre la estructura del ARN no apareado y apareado es de entre -20 y -60 kcal/mol, p. ej., alrededor de -40 kcal/mol. La complementariedad entre el miARN y el miARN no tiene por qué ser perfecta y se pueden tolerar alrededor de 1 a 3 engrosamientos de nucleótidos no apareados. La estructura secundaria adoptada por una molécula de ARN puede predecirse mediante algoritmos informáticos convencionales en la técnica tales como mFold, UNAFold y RNAfold. La cadena particular del tallo de ARNs desde el pre-miARN que se libera por la actividad DCL y se carga en el complejo RISC se determina por el grado de complementariedad en el extremo 5', con lo que la cadena que en su extremo 5' es la menos implicada en el enlace de hidrógeno entre los nucleótidos de las diferentes cadenas del tallo de ARNs escindido se carga en el complejo RISC y determinará la especificidad de secuencia de la degradación de la molécula de ARN diana. Sin embargo, si empíricamente la molécula de miARN de una molécula de pre-miARN sintética particular no es funcional debido a que se carga la cadena "equivocada" en el complejo RISC, resultará inmediatamente evidente que este problema puede ser resuelto mediante el intercambio de la posición de la molécula de miARN y su complemento en las respectivas cadenas del tallo de ARNs de la molécula de pre-miARN. Tal como se sabe en la técnica, la unión entre A y U que implican dos uniones hidrógeno, o G y U que implican dos uniones hidrógeno es menos fuerte que entre G y C que implica tres uniones hidrógeno.

Moléculas de miARN pueden estar comprendidos dentro de sus moléculas de pre-miARN que se producen de forma natural, pero también pueden ser introducidas en armazones de moléculas pre-miARN existentes mediante el intercambio de la secuencia de nucleótidos de la molécula de miARN normalmente procesada a partir de dicha molécula de pre-miARN existente para la secuencia de nucleótidos de otro miARN de interés. El armazón del pre-miARN también puede ser completamente sintético. Del mismo modo, moléculas de miARN sintéticas pueden estar comprendidas dentro y procesadas de armazones de moléculas de pre-miARN existentes o armazones de pre-miARN sintéticos.

55 Productos de expresión a modo de ejemplo también pueden ser ribozimas que catalizan su propio escisión o la escisión de otros ARN.

- En un ejemplo del gen quimérico descrito en esta memoria, la modulación de la expresión es incrementar la expresión y dicha secuencia de ácido nucleico que codifica un producto de expresión de interés codifica un ARN que, cuando se transcribe, 1. proporciona una molécula de ARN capaz de aumentar la expresión de un gen endógeno a dicha planta de algodón. Genes de este tipo podrían correlacionarse positivamente con la longitud de la fibra, la resistencia de la fibra o una alteración deseada en las propiedades de la pared celular, que resultan, p. ej., en una carga alterada de dichas paredes celulares, una alteración en la capacidad de tinción, una producción disminuida de fibra de pelusa, una alteración en la relación de madurez de la fibra y micronaire, o 2. proporciona una molécula de ARN capaz de disminuir la expresión de un gen endógeno a dicha planta de algodón, en donde dicho gen puede ser correlacionado negativamente con la longitud de la fibra, la resistencia de la fibra o una alteración deseada en las propiedades de la pared celular que resultan, p. ej., en una carga alterada de dichas paredes celulares, una alteración en la capacidad de tinción, una producción disminuida de fibra de pelusa, una alteración en la relación de madurez de la fibra, una disminución en el contenido de fibra inmadura, o un incremento en la uniformidad de la fibra y micronaire.
- 15 Productos de expresión basados en ARN a modo de ejemplo incluyen ARNs inhibidores tales como miARNs, siARNs, ARNs antisentido, ARNs sentido, ARNs de horquilla o ribozimas dirigidos a glucanasa, factor despolimerizante de actina que codifica ADF, CPC que codifica caprice o TRY que codifica triptychon, entre otros.

En otro ejemplo del gen quimérico descrito en esta memoria, dicho producto de expresión es un gen informador o un gen específico para fibras tal como se describe en otra parte en esta solicitud.

- 20 En un ejemplo del gen quimérico descrito en esta memoria, dicha molécula de ARN comprende una primera y una segunda región de ARN, en donde 1. dicha primera región de ARN comprende una secuencia de nucleótidos de al menos 19 nucleótidos consecutivos que tienen al menos aproximadamente 94% de identidad de secuencia con la secuencia de nucleótidos de dicho gen endógeno; 2. dicha segunda región de ARN comprende una secuencia de nucleótidos complementaria a dichos 19 nucleótidos consecutivos de dicha primera región de ARN; y 3. dichas primera y segunda regiones de ARN son capaces de un apareamiento de las bases para formar una molécula de ARN de doble cadena entre al menos dicho 19 nucleótidos consecutivos de dichas primera y segunda regiones.

La presente solicitud también describe un vector que comprende el gen quimérico descrito en esta memoria.

- Un "vector" se refiere a cualquier agente basado en ácido nucleico capaz de transportar y transferir información genética tal como un plásmido, cósmido, virus, secuencia de replicación autónoma, fago, o secuencia de nucleótidos de ADN o ARN lineal de cadena sencilla, circular de cadena sencilla, lineal de doble cadena o circular de doble cadena. El vector recombinante se puede derivar de cualquier fuente y es capaz de integración genómica o replicación autónoma. Por lo tanto, el gen quimérico arriba descrito puede proporcionarse en un vector recombinante. Un vector recombinante comprende típicamente, en una orientación 5' a 3': un promotor para dirigir la transcripción de una secuencia de ácido nucleico y una secuencia de ácido nucleico a transcribir. Estos elementos se corresponden con el gen quimérico descrito en esta memoria a ser introducido. El vector recombinante puede comprender, además, un terminador de la transcripción 3', una señal de poliadenilación 3', otras secuencias de ácido nucleico no traducidas, secuencias de ácido nucleico de tránsito y de fijación de objetivo, marcadores seleccionables, potenciadores y operadores, según se desee. La expresión "5' UTR" se refiere a la región no traducida de ADN aguas arriba, o 5' de la región codificante de un gen, y "3' UTR" se refiere a la región no traducida de ADN aguas abajo, o 3' de la región codificante de un gen. Medios para preparar vectores recombinantes son bien conocidos en la técnica. Métodos para preparar vectores recombinantes particularmente adecuados para la transformación de plantas se describen en los documentos US4971908, US4940835, US4769061 y US4757011. El vector descrito en esta memoria puede ser un vector de expresión. Vectores típicos, útiles para la expresión de ácidos nucleicos en plantas superiores, son bien conocidos en la técnica e incluyen vectores derivados del plásmido inductor de tumores (Ti) de *Agrobacterium tumefaciens*.

- En otro aspecto, la presente solicitud describe una célula de plantal de algodón transgénica que comprende el gen quimérico descrito en esta memoria o el vector descrito en esta memoria. La presente invención también se dirige a células de plantas de algodón transgénicas y a plantas de algodón transgénicas que comprenden una secuencia de ácido nucleico tal como se describe arriba, es decir, la secuencia del promotor descrita en esta memoria, enlazada operativamente a una secuencia de ácido nucleico heterólogo que codifica un producto de expresión de interés. Alternativamente, dichas células de plantas transgénicas o plantas comprenden el gen quimérico descrito en esta memoria. Secuencias de promotor preferidas y productos de expresión de interés y otros elementos reguladores se han descrito arriba.

- Una planta de algodón transgénica puede ser producida introduciendo la o las secuencias de ácido nucleico tal como se describe arriba en plantas o células vegetales. "Introducción" en relación con la presente solicitud se refiere

a la disposición de información genética en una planta o célula vegetal por medios artificiales. Esto puede efectuarse por cualquier método conocido en la técnica para la introducción de ARN o ADN en células vegetales, protoplastos, callos, raíces, tubérculos, semillas, tallos, hojas, plántulas, embriones, polen y microesporas, otros tejidos vegetales o plantas enteras. Más particularmente, "introducción" significa la integración de forma estable en el genoma de la planta.

A las plantas que contienen la secuencia de ácido nucleico transformadas se las alude como "plantas transgénicas". Transgénico y recombinante se refieren a un organismo huésped tal como una planta en la que se ha introducido una molécula de ácido nucleico heterólogo (p. ej., la secuencia de ácido nucleico, el gen quimérico o el vector tal como se describe en esta memoria). El ácido nucleico puede ser integrado de forma estable en el genoma de la planta. Métodos específicos para la introducción se describen en relación con los métodos descritos en esta memoria.

"Algodón" o "planta de algodón", tal como se utiliza en esta memoria, puede ser cualquier variedad útil para el cultivo de algodón. Las variedades de algodón más comúnmente utilizadas son *Gossypium barbadense*, *G. hirsutum*, *G. arboreum* y *G. herbaceum*. Variedades adicionales incluyen *G. africanum* y *G. raimondii*. También se incluye la progenie de los cruces de cualquiera de las especies anteriores con otras especies o cruces entre tales especies.

Una célula de planta de algodón puede ser cualquier célula que comprende esencialmente la información genética necesaria para definir una planta de algodón, que puede, aparte del gen quimérico descrito en esta memoria, ser complementada con uno o más transgenes. Las células pueden derivar de los distintos órganos y/o tejidos que forman una planta de algodón, incluyendo pero no limitados a frutos, semillas, embriones, tejido reproductor, regiones meristémicas, tejido de callo, hojas, raíces, brotes, flores, tejido vascular, gametofitos, esporofitos, polen y microesporas.

La presente solicitud también describe una planta transgénica que consiste en la célula de la planta de algodón transgénica descrita anteriormente en esta memoria, o que comprende el gen quimérico o el vector descrito en esta memoria integrado de manera estable en el genoma de la planta. Esto puede efectuarse por protocolos de transformación descritos en otra parte en esta solicitud.

En otra realización, la presente invención se refiere a una semilla generada a partir de una planta de algodón transgénica descrita en esta memoria, en donde dicha semilla comprende el gen quimérico descrito en esta memoria.

La semilla se forma por una planta embrionaria encerrada junto con nutrientes almacenados por una envuelta de la semilla. Es el producto del óvulo maduro de las plantas gimnospermas y angiospermas, a la última de las cuales pertenece el algodón, que se produce después de la fertilización y en cierta medida crece dentro de la planta madre.

En esta memoria se describen, además, fibras de algodón y aceite de semilla de algodón obtenibles u obtenidos de las plantas descritas en esta memoria. Las fibras de algodón descritas en esta memoria pueden distinguirse de otras fibras mediante la aplicación del método de detección descrito en el documento WO2010/015423 y la comprobación de la presencia del ácido nucleico de (a) o gen quimérico de (b) en las fibras. Por consiguiente, el ácido nucleico de (a) también se puede utilizar para el seguimiento de paredes celulares, en particular fibras de algodón de acuerdo con la invención.

También se describen en esta memoria hilos y materiales textiles hechos a partir de las fibras descritas en esta memoria, así como los productos alimenticios y piensos que comprenden o están hechos del aceite de semilla de algodón descrito en esta memoria. También se describe un método para obtener el aceite de semilla de algodón, que comprende recolectar las semillas de algodón de la planta de algodón descrita en esta memoria y extraer dicho aceite de dichas semillas. Además, también se describe un método para producir fibras de algodón, que comprende cultivar la planta de algodón descrita en esta memoria y recolectar el algodón a partir de dichas plantas de algodón.

Está disponible un cierto número de métodos para introducir ADN en células vegetales o plantas, ya sea por transformación o introgresión. La transformación mediada por *Agrobacterium* de algodón se ha descrito, p. ej., en la patente de EE.UU. 5.004.863, en la patente de EE.UU. 6.483.013 y en el documento WO2000/71733.

Las plantas también pueden ser transformadas por bombardeo de partículas: partículas de oro o wolframio se recubren con ADN y luego se disparan en células de plantas jóvenes o embriones de plantas. Este método también permite la transformación de plastidios de plantas. La transformación del algodón mediante bombardeo de partículas se reseña, p. ej., en el documento WO 92/15675.

La transformación (transducción) viral se puede utilizar para la expresión transitoria o estable de un gen, dependiendo de la naturaleza del genoma del virus. El material genético deseado se empaqueta en un virus de

planta adecuado y se permite que el virus modificado infecte la planta. La progenie de las plantas infectadas está libre de virus y también libre del gen insertado. Se describen o se detallan adicionalmente métodos adecuados para la transformación viral en el documento WO 90/12107, WO 03/052108 o WO 2005/098004.

5 "Introgresión" significa la integración de un gen en el genoma de una planta por medios naturales, es decir, mediante el cruce de una planta que comprende el gen quimérico descrito en esta memoria con una planta que no comprende dicho gen quimérico. La descendencia se puede seleccionar para las que comprenden el gen quimérico.

Otros protocolos de transformación y de introgresión también se pueden encontrar en la patente de EE.UU. 7.172.881.

10 En un aspecto adicional, la presente solicitud describe un método de cultivo de algodón, que comprende (a1) proporcionar la planta transgénica descrita en esta memoria o producida por el método descrito en esta memoria ; o (a2) introducir un gen quimérico según se describe en esta memoria o un vector descrito en esta memoria en una planta; (b) cultivar la planta de (a1) o (a2); y (c) recolectar el algodón producido por dicha planta.

"Cultivo" se refiere a la creación del entorno para que la planta crezca, se multiplique y/o envejezca. Condiciones de cultivo adecuadas para plantas específicas son bien conocidas en la técnica.

15 La presente solicitud describe también un método de producir una semilla que comprende el gen quimérico descrito en esta memoria, que comprende (a) cultivar una planta transgénica que comprende el gen quimérico descrito en esta memoria o el vector descrito en esta memoria, una planta transgénica descrita en esta memoria o una planta transgénica obtenido por el método descrito en esta memoria, en que dicha planta transgénica produce dicha semilla y dicho gen quimérico está comprendido en dicha semilla, y (b) aislar dicha semilla de dicha planta transgénica.

20 En otro aspecto, la presente solicitud describe un método de efectuar la expresión específica para la semilla de un producto en algodón, que comprende introducir el gen quimérico descrito en esta memoria o el vector descrito en esta memoria en el genoma de una planta de algodón; o proporcionar la planta transgénica descrita en esta memoria. En un ejemplo, la expresión específica para la semilla puede ser la expresión específica para la envuelta de la semilla, la expresión específica para tricomas o la expresión específica para fibras.

25 La presente solicitud describe, además, un método para alterar propiedades de la fibra en una planta de algodón, que comprende introducir el gen quimérico descrito en esta memoria o el vector descrito en esta memoria en el genoma de una planta de algodón; o proporcionar la planta transgénica descrita en esta memoria.

En un ejemplo, el método comprende, además, cultivar dicha planta hasta que se generen semillas.

30 En otro ejemplo basado en la etapa adicional de arriba, el método es para aumentar la producción de algodón de una planta de algodón y comprende, además, recolectar el algodón producido por dicha planta de algodón. En otras palabras, en esta memoria se describe un método para aumentar la producción de algodón a partir de una planta de algodón, que comprende introducir el gen quimérico descrito en esta memoria o el vector descrito en esta memoria en el genoma de una planta de algodón; o proporcionar la planta transgénica descrita en esta memoria; cultivar dicha planta hasta que se generen semillas; y recolectar el algodón producido por dicha planta de algodón.

35 La expresión "aumentar la producción" en relación con la presente solicitud se refiere a un aumento en la producción de fibras de algodón que se puede conseguir, p. ej., aumentando el número de fibras producidas en una semilla de algodón, la longitud de las fibras o la resistencia de la fibras. Genes y productos de expresión de los mismos implicados en conferir estas propiedades se han descrito anteriormente.

40 En otro aspecto, la presente solicitud describe el uso del gen quimérico descrito en esta memoria, el vector descrito en esta memoria o la planta transgénica o célula vegetal descrita en esta memoria para la expresión específica para la semilla de un producto en algodón, para alterar las propiedades de la fibra en el algodón o para aumentar la producción de algodón. Las definiciones y ejemplos adicionales arriba descritos para otros aspectos descritos en esta memoria se aplican igualmente al presente aspecto.

45 Las células de plantas de algodón transformadas y plantas de algodón descritas en esta memoria o los obtenidos por los métodos descritos en esta memoria puede contener, además del gen quimérico arriba descrito, al menos otro gen quimérico que comprende un ácido nucleico que codifica un producto de expresión de interés. Ejemplos de un producto de expresión de este tipo incluyen moléculas de ARN o proteínas tales como, p. ej., una enzima para la resistencia a un herbicida.

Productos de expresión de interés adicionales confieren resistencia a los insectos a una planta de algodón, es decir, resistencia al ataque por ciertos insectos diana, o la tolerancia a estreses abióticos.

5 Las células vegetales y plantas transformadas descritas en esta memoria tales como las obtenidas por los métodos descritos en esta memoria se pueden utilizar adicionalmente en los procedimientos de reproducción bien conocidos en la técnica tales como cruce, autofecundación y retrocruzamiento. Los programas de reproducción pueden implicar el cruce para generar una generación F1 (primera filial), seguido de varias generaciones de autopolinización (generando F2, F3, etc.). El programa de cultivo también puede implicar etapas de retrocruzamiento (BC), con lo que la descendencia se retro-cruza a una de las líneas parentales, denominada el parental recurrente.

10 Por consiguiente, también se describe en esta memoria un método para producir plantas que comprenden el gen quimérico descrito en esta memoria, que comprende la etapa de cruzar la planta de algodón descrita en esta memoria con otra planta o consigo misma y seleccionar la descendencia que comprende dicho gen quimérico.

Las células de plantas transgénicas y plantas obtenidas por los métodos descritos en esta memoria también se pueden utilizar adicionalmente en procedimientos de transformación posteriores, p. ej., para introducir un gen quimérico adicional.

15 Las figuras muestran:

Figura 1: Gel de agarosa que presenta los resultados de la reacción PCR para amplificar la secuencia de ADN de cuatro genes MADS.

Figura 2: Esquema del procedimiento de PCR inversa.

20 Figura 3: Etapas de clonación para recuperar el promotor MADS6 (PMADS6) tal como se describe en el Ejemplo 1. Esquema 3a/3b de la recuperación de la secuencia genómica y enfoque de la PCR inversa; 3c. Fragmentos recuperados de la PCR inversa; 3d a f: creación de vectores que comprenden el promotor MADS6 completo. Abreviaturas: bla: gen de resistencia a ampicilina; ORI ColE1: origen de replicación de plásmido procedente de pMB1; lacZ: secuencia codificante que codifica el péptido beta-galactosidasa alfa de Escherichia coli; MCS: sitio de multi-clonación; 5' UTR: región 5' no traducida; Plac: Promotor del operón lac de Escherichia coli; Pmads6: promotor MADS6; Phis: secuencia que incluye la región del promotor del gen histona H4 de Arabidopsis thaliana y el primer intrón del gen II de la variante de histona H3.III de Arabidopsis thaliana; 2mepsps: secuencia codificante del gen doble mutante 5-enol piruvilshikimate-3-fosfato sintasa de Zea mays (maíz); (Lebrun et al., 1997) TPotp C: secuencia codificante del péptido de tránsito optimizado, que contiene la secuencia de los genes de la subunidad pequeña RuBisCO de Zea mays (maíz) y Helianthus annuus (girasol); aadA: estreptomomicina y resistencia a espectinomomicina; la secuencia codificante del gen aminoglicósido adeniltransferasa (aadA) del transposón Tn7 de Escherichia coli (Fling et al., 1985); bar: secuencia codificante del gen fosfinotricina acetiltransferasa (= gen de resistencia bialafos) de Streptomyces hygroscopicus (Thompson et al, 1987); 3'nos: fragmento del extremo 3' no traducido del gen nopalina sintasa del ADN-T de pTIT37 y que contiene señales de poliadenilación de plantas; ori pVS1: origen de replicación de plásmido procedente de pVS1 para el mantenimiento estable en Agrobacterium; P35S3: fragmento de la región del promotor de la transcripción de virus del mosaico de la coliflor 35S; GUS: secuencia codificante del gen beta-glucuronidasa de Escherichia coli, incluyendo el segundo intrón del gen ST-LS1 de Solanum tuberosum (patata).

Figura 4: Producción del vector de la transformación de algodón pTTS108 que comprende el promotor MADS6 putativo, la secuencia codificante de GUS y el marcador de selección bar.

40 Figura 5: Óvulos transformadas con el gen informador GUS controlado por PMADS6 (Figura 5a) y un gen informador GUS controlado por PFBP7 (Figura 5b) 6 dap. Los círculos indican manchas azules que se encuentran en los óvulos que corresponde a la expresión de GUS.

Los ejemplos ilustran la invención.

Materiales

45 A menos que se indique lo contrario, los productos químicos y reactivos en los ejemplos se obtuvieron de Sigma Chemical Company, las endonucleasas de restricción eran de Fermentas o Roche-Boehringer, y otras enzimas modificadoras o kits con respecto a productos bioquímicos y ensayos biológicos moleculares eran de Qiagen, Invitrogen y Q-BIOgene. Las cepas bacterianas eran de Invitrogen. Las etapas de clonación llevadas a cabo tales como, p. ej., escisiones por restricción, electroforesis en gel de agarosa, purificación de fragmentos de ADN, enlace de fragmentos de ADN, transformación de células de *E. coli*, bacterias en crecimiento, fagos que se multiplican y

análisis de la secuencia de ADN recombinante se llevan a cabo tal como se describe por Sambrook (1989). La secuenciación de moléculas de ADN recombinante se lleva a cabo usando el secuenciador de ADN por fluorescencia láser ABI según el método de Sanger.

Ejemplo 1: Seguimiento del promotor MADS6 en algodón

- 5 El promotor FBP7 de petunia es conocido por expresar tricomas específicamente. En el curso de la presente invención, se identificó en algodón un promotor con propiedades similares.

Se efectuó una búsqueda BLAST con la secuencia de la proteína FBP7 controlada por el promotor FBP7 en la base de datos TrEMBL para identificar homólogos potenciales en el algodón.

- 10 La búsqueda recuperó dos aciertos en *Gossypium hirsutum*, ambos de los cuales son proteínas MADS-box denominadas GhMADS6 y GhMADS7.

En plantas, los genes MADS-box codifican una gran familia de factores de transcripción de al menos 100 miembros (revisado por De Bodt et al. 2003; Kofuji et al 2003; Parenicová et al. 2003; Nam et al., 2004).

- 15 Se ha encontrado por Lightfoot et al. (2007) que GhMADS6 y GhMADS7, además de dos más, GhMADS4 y GhMADS5, son homólogos de FBP7. Este grupo demostró que MADS5 y MADS6 se expresan altamente en el óvulo, la flor y tejido de las fibras. Además de ello, ambas proteínas se expresan en el desarrollo temprano de las fibras (0 a 6 DPA). No se detectó expresión alguna en las hojas y el tallo. MADS5 muestra una baja expresión en la raíz.

Con el fin de encontrar un promotor específico para las fibras, el gen que codifica MADS6 fue escogido para la investigación adicional.

- 20 Se utilizó una búsqueda blastn para identificar el ADNc que codifica la proteína MADS6. La información recuperada reveló un fragmento de ADNc de 1040 pares de bases. Sin embargo, una búsqueda de esta secuencia en bases de datos genómicas no produjo información alguna. Por consiguiente, no se pudo determinar la secuencia de nucleótidos 5' con respecto a la secuencia codificante MADS6.

- 25 En el presente caso, los genes MADS identificados muestran un nivel muy alto de identidad de secuencia. Esto hace una tarea difícil el seguimiento del promotor de un gen MADS específico, en este caso el gen MADS6. Para un enfoque basado en la PCR inversa, se necesita encontrar cebadores específicos para el gen MADS seleccionado, los cuales, considerando la alta identidad de secuencia entre los miembros de la familia, es bastante elaborada.

La secuencia genómica fue trazada por PCR utilizando los siguientes cebadores.

MADS4

- 30 TSOL454_directo: 5' - tcgaggccatacattctcag - 3' (SEQ ID NO: 2)
TSOL455_inverso: 5' - gtcttacacactctacacatc - 3' (SEQ ID NO: 3)

MADS5

TSOL456_directo: 5' - agaggaactcccactccctac - 3' (SEQ ID NO: 4)
TSOL457_inverso: 5' - atgtagagtacatattggtga - 3' (SEQ ID NO: 5)

- 35 MADS6

TSOL458_directo: 5' - catccatctgcttactcccat - 3' (SEQ ID NO: 6)
TSOL459_inverso: 5' - tacatcatacgaactcaca - 3' (SEQ ID NO: 7)

MADS7

- 40 TSOL460_directo: 5' - caaaccagctgatgcaagcagc - 3' (SEQ ID NO: 8)
TSOL461_inverso: 5' - caacaactaggcttcaactgt - 3' (SEQ ID NO: 9)

Las reacciones de PCR se realizaron en ADN genómico de tipo salvaje Cocker.

Para todos los cuatro genes, se amplificó un único fragmento mayor que el fragmento esperado, en base a secuencias de ADNc (véase la Figura 1). Esto indica que los cuatro genes contienen intrones dentro de la región codificante amplificada.

5 Los fragmentos de PCR obtenidos para MADS 4, MADS5 y MADS7 se almacenaron a -20° C. El fragmento de PCR obtenido para MADS6 se clonó en un vector de clonación.

10 Para comparación, se realizó una reacción PCR adicional sobre el ADN genómico para MADS6 utilizando los cebadores reseñados en Lightfoot et al. (2007). Una vez más, el fragmento obtenido (1040 pb) era mayor que el fragmento de ADNc correspondiente (394 pb). El fragmento obtenido se clonó y secuenció y se alineó con un fragmento genómico obtenido con cebadores específicos para el extremo 3' de MADS6, con la secuencia de ADNc de MADS6 y con los fragmentos obtenidos a partir de ADNc utilizando cebadores específicos para el extremo 3' de MADS6. Se pudieron confirmar intrones dentro de la secuencia codificante.

El fragmento de 1040 pares de bases obtenido, correspondiente a la parte 3' del gen MADS6, se tomó entonces como base para la amplificación de la secuencia 5' genómica que comprende el promotor.

15 Se juzgó el enfoque de la PCR inversa. Un esquema de las etapas a realizar se representa en la figura 2. Un buen candidato para la enzima A se podría definir como NheI. Ensayos con esta enzima resultaron en fragmentos que se hibridan con fragmentos genómicos de más de 5 kb.

Después se diseñaron cebadores anidados para la PCT inversa que fijan como objetivo el extremo 3' de la secuencia codificante MADS6 que se encontró que era al menos idéntica de los cuatro genes MADS identificados en *G. hirsutum*.

20 Las reacciones PCR se establecieron con estos cebadores y algunos de los candidatos para la enzima A.

Sólo podían obtenerse múltiples bandas no específicas utilizando los cebadores anteriores dirigidos específicamente a MADS6 en la región 3' del gen, y no pudo determinarse la secuencia 5' aguas arriba del gen MADS6.

25 Se realizó un intento adicional con la secuencia 5' menos específica del gen MADS6. En primer lugar, secuencias 5'-genómicas de MADS6 se amplificaron utilizando cebadores directos TSOL472 (5' - GGACAAGTGATCAAAGAG - 3'; SEQ ID NO: 10) resp. TSOL473 (5' - ATTGGCCGGAAGCTTACCA - 3'; SEQ ID NO: 11), la unión a la región 5' no traducida (UTR) y cebador inverso TSOL465 que tiene la secuencia 5' - GGACCTGATCCTAGTAATTCC - 3' (SEQ ID NO: 12) y la unión al ADNc de MADS6. Aunque la distancia entre TSOL472 y TSOL473 en la 5'UTR de la secuencia de ADNc es solamente 30 pb, los fragmentos amplificados (2500 pb, resp. 3250 pb) difieren en 750 pb de longitud; lo que indica una secuencia de intrón de 750 pb en la 5'UTR del gen MADS6. La clonación y secuenciación de ambos fragmentos confirmaron esta hipótesis.
30 El fragmento más largo de 3251 pares de bases se tomó como base para la amplificación de la secuencia 5' que comprende el promotor (véase la figura 3a).

35 En un enfoque de PCR inversa adicional, un buen candidato para la enzima B se definió por hibridación de ADN Coker genómico digerido a una sonda 5'UTR de MADS6 (fragmento de 540 pb obtenido mediante amplificación por PCR con los cebadores TSOL473 (SEQ ID NO: 11) y TSOL512 (5' - AGCCATTCCTATCCCATAC - 3'; SEQ ID NO: 13).

Para todas las enzimas de restricción testadas se obtuvieron al menos 2 fragmentos de hibridación, lo que indica una hibridación cruzada con al menos otro miembro de la familia MADS.

40 De todas las enzimas, BclI, NdeI y HindIII fueron consideradas como las más prometedoras. Todas estas enzimas proporcionaron 2 bandas de hibridación de entre 3,5 y 0,8 kb.

A continuación, el ADN Coker genómico se digirió en tres viales de reacción con BclI, NdeI y HindIII. 10 µg de ADN genómico digerido, limpiado por precipitación y resuspendido en 85 µl de tampón TE se autoligó (en un volumen de reacción de 100 µl) para obtener fragmentos circulares (fig.3 b), adecuados para el análisis por PCR anidada.

45 Se realizó una PCR anidada utilizando cebadores TSOL502 (5'-CTGTTCTATCTTTCCCTTCTTG - 3', SEQ ID NO: 14) y TSOL503 (5' - AGAAAGAAAGCATGCATTTAGG - 3'; SEQ ID NO: 15) para la primera reacción PCR y los cebadores TSOL500 (5' - ATACATGATGGGTTCTCTTC - 3'; SEQ ID NO: 16) y TSOL501 (5'-AAGCATGCATTTAGGTAAG - 3'; SEQ ID NO: 17) para la segunda reacción PCR. La PCR anidada dio como resultado la amplificación de una banda que tiene un tamaño correspondiente a la banda de hibridación para dos de las tres enzimas ensayadas: amplificación de un fragmento de 850 pb en HindIII resp. un fragmento de 1,7 kb en

Ndel digerido y ADN genómico re-ligado. La banda amplificada de 1,7 kb se clonó en un vector de clonación y se secuenció. El vector recombinante se denominó pTS478 (Figura 3b).

Los dos fragmentos obtenidos y clonados que corresponden al 5'UTR y la región del promotor putativo se unieron como sigue (Figura 3c).

5 Como primer paso, ambos fragmentos fueron amplificados por PCR añadiendo sitios de clonación adecuados para la clonación adicional.

10 A / Amplificación de la mayor parte de aguas arriba: amplificación por PCR en el ADN molde de pTS478, utilizando el cebador directo TSOL558 (5'-GCGCGGTACCGAATTCCATATGTATATTATATATT-3' que contiene sitios de restricción KpnI y EcoRI; SEQ ID NO: 18) y el cebador inverso TSOL559 (5'-TATAACTAGTAGTGTGCTGGAATTCGC-3' que contiene un sitio de restricción SpeI; SEQ ID NO: 19). El fragmento se clonó como vector intermedio (pTS412) y se secuenció (Figura 3d).

15 B / amplificación de la mayor parte 3' de 5'UTR: amplificación por PCR en el ADN molde de pTS469, utilizando el cebador directo TSOL473 (5'-ATTGGCCGGAAGCTTACCA-3') y el inverso TSOL560 (5'-GCATCCATGGTCTCTTTGATCACTTGTA-3' que contiene un sitio de restricción NcoI en el inicio de la traducción; SEQ ID NO: 20). Después de la digestión de la restricción SphI, el fragmento se ligó en pTS412 y se linearizó por digestión de la restricción SphI. Como resultado, los dos fragmentos podían ser unidos en el vector pTS413; con ello, se pudo reconstruir la secuencia 5' aguas arriba completa (1,5 kb) (promotor + 5'UTR incluyendo el intrón de 750 pb) (Figura 3e). La secuencia de la inserción de pTS413 completo (1,5 kb) fue confirmada por secuenciación.

20 Para una fácil clonación e intercambio entre diferentes vectores de transformación de plantas, el fragmento aguas arriba MADS6 de 1,5 kb se clonó en un plásmido como fragmento KpnI/NcoI.

El vector pTS414 resultante contiene un fragmento de 1,5 kb que comprende la secuencia 5' aguas arriba del gen MADS6 de *G. hirsutum*, que comprende la secuencia de promotor putativo.

Ejemplo 2: Construcción de un casete de expresión que comprende el promotor pMADS6 y una secuencia que codifica un producto de expresión de interés

25 El vector pTS414 que comprende el fragmento de 1,5 kb obtenido en el ejemplo 1 y un vector que comprende la secuencia codificante para un producto de expresión de interés se digirió con las enzimas de restricción apropiadas. El fragmento de la secuencia que codifica el producto de expresión de interés se ligó en el vector que comprende dicho fragmento de 1,5 kb.

30 El vector resultante se digiere entonces con enzimas de restricción apropiadas, el casete de expresión que comprende la secuencia que codifica el producto de expresión de interés unido al promotor MADS6 putativo se purifica y clona en dos vectores, uno que comprende el marcador de selección bar y uno que comprende la selección de epsps.

Ejemplo 3: Construcción de un casete de expresión que comprende el promotor pMADS6 y el gen informador GUS

35 El vector pTS414 que comprende el fragmento de 1,5 kb obtenido en el ejemplo 1 y un vector que comprende la secuencia codificante de GUS (SEQ ID NO: 21) se digirió con enzimas de restricción EcoRI y NcoI. El fragmento de la secuencia de GUS se ligó en pTS414 para obtener pTS415 (véase la figura 4). Se realizó una construcción de control por clonación del promotor FBP7 unido al informador de GUS.

40 pTS415 se digirió entonces con EcoRI y PstI, el casete de expresión que comprende la secuencia codificante de GUS unido al promotor MADS6 putativo se purificó y se clonó en un vector que comprende el marcador de selección bar, resultando el vector pTTS108 (Figura 4). Este vector se utilizó para la transformación estable de algodón.

Ejemplo 4: Análisis de la expresión de construcciones informadoras que comprenden pMADS6 y pFBP7, ambos unidos a un gen informador GUS

45 Óvulos de algodón se transformaron por bombardeo de partículas utilizando una Pistola de Genes Bio-RAD Modelo PDS 1000/He Biolistic. 60 mg de partículas de oro de un diámetro medio de 0,3 a 3 µm se lavaron una vez con etanol al 70% y posteriormente 3 veces con agua destilada. Las partículas se resuspendieron en 1 ml de agua destilada. A 50 µl de suspensión, se añadieron 5 µg de ADN, 50 µl de CaCl₂ y 20 µl de espermidina, la mezcla

resultante se agitó suavemente durante 10 min a TA y después se dejó a TA durante 5 min más. El sedimento resultante se lavó 3 veces con etanol al 100% y luego se re-suspendió en 55 µl de etanol al 100%.

6 µl de la suspensión se sembraron en estrías en macrotransportadores (Bio-Rad) y se dejaron secar. El bombardeo de partículas se llevó a cabo utilizando discos Rupture de 900, 1100, 1350 psi (Bio-RAD).

5 La expresión del gen informador GUS controlado por PMADS6 (construcción pTS417) o PFBP7 se analizó 6 días después de la polinización (dap).

La Figura 5 muestra óvulos transformados con el gen informador GUS controlado por PMADS6 (Figura 5a) y un gen informador GUS controlado por PFBP7 (Figura 5b) 6 dap. Los círculos indican manchas azules que se encuentran en los óvulos que corresponde a la expresión de GUS. A partir de estos resultados, es evidente que PMADS6 conduce la expresión en el óvulo.

Ejemplo 5: Análisis de pMADS6 por motivos de unión específica para semillas y tricomas

Se llevó a cabo un análisis del promotor utilizando bases de datos disponibles públicamente tales como PLACE (<http://www.dna.affrc.go.jp/PLACE/>), RegSite (<http://linux1.softberry.com/berry.phtml?topic=regsitelist>), PlantCare (Lescot et al., 2002; disponible en <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) y AtcisDB (Davuluri et al., 2003). La búsqueda se limitó a elementos específicos para tricomas, así como a motivos correspondientes a una caja L1 y motivos de unión a MYB. Los dos últimos se han descrito como motivos que confieren potencialmente una expresión específica de tricomas (Wang y Chen, 2004).

La búsqueda reveló dos motivos que confieren en potencia una expresión específica para la semilla, específica para la envuelta de la semilla o específica para tricomas. El primero es una caja T/G que corresponde al motivo RPSP01178 de tricomas situado partiendo de la posición -1224 (correspondiente a la posición 298 de SEQ ID NO:1 La secuencia exacta del motivo caja T/G es AACGTG. Dicho motivo de unión ha sido identificado en el promotor de un gen MYB de fibra de algodón (Shangguan et al., 2008), en donde su delección reduce la actividad del promotor en Arabidopsis y tabaco.

El otro motivo de unión corresponde a un motivo de unión MYB y se encuentra en la posición -676 (correspondiente a la posición 846 de SEQ ID NO: 1. Curiosamente, dicho motivo de unión MYB también está presente en la secuencia codificante del gen MADS6 regulado de forma natural por el presente promotor. Ha sido identificado por Wang y Chen (2004) y se encuentra en al menos el promotor RDL1, en donde confiere especificidad para tricomas en Arabidopsis, y en los promotores GL1 (que controla el gen myb en Arabidopsis) y GaMyb2 (que controla el gen MYB en algodón). Se ha demostrado anteriormente que la interrupción del motivo de unión MYB conduce a una reducción en la producción de tricomas.

Ejemplo 6: Análisis de la expresión de pMADS6 en fibras de algodón y otros tejidos vegetales de algodón

Algodón fue transformado por un método de transformación mediado por Agrobacterium bien conocido en la técnica con la construcción de acuerdo con el ejemplo 3 de acuerdo con protocolos bien conocidos en la técnica y plantas transformadas se cultivaron en el invernadero. Las plantas se analizaron para la expresión de GUS.

La expresión de GUS en la fibra en desarrollo podría ser detectada 1 DPA, una alta expresión 5 DPA, 8 DPA, 10 DPA, 13 DPA y 30 DPA. Una baja expresión de GUS se pudo detectar en las anteras. No se detectó expresión en los sépalos, el tallo, el eje floral y las raíces.

Referencias

Braasch y Corey (2001). Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. Chem Biol 8(1), p. 1-7.

Carrington y Freed 1990. Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. J. Virol. 64, p. 1590-1597.

Chaubet et al. (1992). Genes encoding a histone H3.3-like variant in Arabidopsis contain intervening sequences. J Mol Biol. 225(2):569-74.

Davuluri et al. (2003). AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics 4(1):25.

De Bodt, S. et al. (2003). Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J Mol Evol 56, p. 573-586.

Delaney et al. (2007). The fiber specificity of the cotton FSItp4 gene promoter is regulated by an AT-rich promoter region and the AT-hook transcription factor GhAT1. Plant and Cell Physiology 48, 1426-1437.

Eisner, T. et al. (1998). When defence backfires: detrimental effect of a plant's protective trichomes on an insect beneficial to the plant. Proc Natl Acad Sci USA 95, p. 4410-4414.

- Hsu et al. (1999). Analysis of promoter activity of cotton lipid transfer protein gene LTP6 in transgenic tobacco plants. *Plant Science* 143, p. 63-70.
- Immink, R. et al. (2003). Analysis of the petunia MADS-box transcription factor family. *Mol Gen Genomics* 268, p. 598-606.
- 5 John y Keller (1996). Metabolic pathway engineering in cotton: biosynthesis of polyhydroxybutyrate in fiber cell. *Proceeding of the National Academy of Sciences, USA* 93, 12678-12773.
- Kim y Triplett, 82001). Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiology* 127, p. 1361-1366.
- 10 Kofuji, R. et al. (2003). Evolution and divergence of the MADS-box gene family based on genome-wide expression analysis. *Mol Biol Evol* 20, p. 1963-77.
- Lescot et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002 30(1):325-327.
- Leseberg, C. et al. (2006). Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene* 378, p. 84-94.
- 15 Li et al. (2005). The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. *The Plant Cell* 17, p. 859-875.
- Li, C.H. et al. (2002). Isolation of genes preferentially expressed in cotton fibers by cDNA filter array and RT-PCR. *Plant Science* 163, p. 1113-1120.
- Li, X.B. et al. (2002). Molecular characterization of the cotton GhTUB1 gene that is preferentially expressed in fiber. *Plant Physiology* 130, p. 666-674.
- 20 Lightfoot et al. (2008). Evidence for alternative splicing of MADS-box transcripts in developing cotton fiber cells. *Mol Genet Genomics* 279:75-85.
- Liu et al. (2000). Cloning and promoter analysis of the cotton lipid transfer protein gene Ltp3. *Biochimica et Biophysica Acta* 1487, p. 106-111.
- 25 Luo et al. (2007). GhDET2, a steroid 5 α -reductase, plays an important role in cotton fiber cell initiation and elongation. *The Plant Journal* 51, p. 419-430.
- Medeiros and Tingey, (2006). Glandular trichomes of *Solanum berthaultii* and its hybrids with *Solanum tuberosum* affect nymphal emergence, development, and survival of *Empoasca fabae* (Homoptera: Cicadellidae). *Journal of Economic Entomology* 99, p. 1483-1489.
- 30 Nam, J. et al. (2004). Type I MADS-box genes have experienced faster birth- and death-evolution than type II MADS-box genes in angiosperms. *PNAS* 101, p. 1919-15.
- Needleman y Wunsch (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 48, p.443-453.
- 35 Parenicová, L. et al. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15, p. 1538-51.
- Pearson y Lipman (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci* 85, p.2444-48.
- Pei et al. (2008). Improvements of Fiber Yield and Fiber Fineness by Expressing the *iaaM* Gene in Cotton Seed Coat. *Cotton Science* 20, Suplemento, p. 44.
- 40 Ranger y Hower, 2001. Role of the glandular trichomes in resistance of perennial alfalfa to the potato leafhopper (Homoptera: Cicadellidae). *Journal of Economic Entomology* 94, p. 950-957.
- Ruan et al. (2003). Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *The Plant Cell* 15, p. 952-964.
- Sambrook, J.F., Russell, D.W. e Irwin, N. (2000). *Molecular Cloning: A Laboratory Manual*, 3rd edition Volumes 1, 2, and 3. Cold Spring Harbor Laboratory Press.
- 45 Shanguan et al. (2008). Promoter of a cotton fiber MYB gene functional in trichomes of *Arabidopsis* and glandular trichomes of tobacco. *Journal of Experimental Botany* 59(13), p. 3533-3542.
- Siomi, H. y Siomi, M. (2009). On the road to reading the RNA-interference code. *Nature* 457, 396-404.
- Song et al. (2000). Expression of Two Tissue-Specific Promoters in Transgenic Cotton Plants. *Journal of Cotton Science* 4, p. 217-223.
- 50 Szymanski et al. (2000). Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*. *Trends in Plant Science* 5, p. 214-219.
- Wagner et al. (2004). New approaches for studying and exploiting an old protuberance, the plant trichome. *Annals of Botany* 93, p. 3-11.
- 55 Wang, S. et al. (2004). Control of plant trichome development by a cotton fiber MYB gene. *The Plant Cell* 16, p. 2323-2334.
- Waterman, M. S. (1995). *Introduction to Computational Biology: Maps, sequences and genomes*. Chapman & Hall. London.
- Werker, E. (2000). Trichome diversity and development. *Advances in Botanical Research* 31, p. 1-35.

LISTADO DE SECUENCIAS

<110> Bayer CropScience N.V.

<120> Nuevo promotor específico para la envuelta de la semilla en algodón

<130> bcs11_2007

5 <160> 21

<170> PatentIn version 3.5

<210> 1

<211> 1522

<212> ADN

10 <213> Gossypium hirsutum

<400> 1

tatgtatatt	atatattatt	aataaaattg	gaaataaact	atatgtatga	aatTTTTTaa	60
aaatattgag	gccattaaag	ttcatcaaaa	cacatcattt	ggcatcaaat	taaacaaaac	120
cttaaaatat	atttagaaat	atttaaaaaa	atattgataa	actttgatag	attaatacca	180
agctaaaatc	aatagttccc	ctaatgtagt	agtactgaaa	ttgtagaata	ttatcttgat	240
atggggtag	actaatctta	gtgggggtat	tattacacag	tgtacgtagt	gcaggacaac	300
gtgtatggca	ccaaaaccct	aactccatag	aaatggaag	caaagataga	gagtgagaac	360
gttgagtctg	agtctaaaca	ccgccatta	aatggagaca	gtagaaaaga	ttcgagaagg	420
gcacgttagt	gagttcagg	actttaaata	aggattacaa	aatatggaaa	cctcattgct	480
aaaaatggtg	gaaatttgca	cctgcaaagc	tcttttcatt	ttgctaattc	tacaccctct	540
tctctaacc	tttcggcaaa	ttacaattcc	aaggtctttt	tccaactctg	aaaacttgag	600
tgagtctccc	accccatctc	tctaagctcc	ctccctataa	tatcacctat	aagcttaggc	660
ctaccctcct	cggaggaaa	ttattaataa	tatactctca	aaacactttg	ctttcttttg	720
gattcgtgat	tttccgaaa	ctgaattggc	cggaactctt	accaaatac	ctcaggtttg	780
agggttttcg	tttcttttcc	ttcatcatc	ttcttttaat	gaagaaaac	aaaaacaatt	840
gctctcagtt	agatcttgtt	aaatatgtat	atacttactt	tttttaactc	ttatttcttt	900
acctaaatgc	atgcttttct	tcttttttcc	tcatcttctt	agatctgtgt	tgtcagaacc	960
ttctacttct	ttgcatgatg	aagatctgtt	ctgttctatc	tttcccttct	tgcctttttt	1020
cttcaccata	atcatacaaa	aatacatgat	gggttctctt	catatttcat	gaatattcat	1080
ccctacttgt	gtttcagatc	tttccaaaaa	toccttcaac	ttgagctctt	tttttttttg	1140
tggagtaata	atatgcaacg	ataaaagtaa	tagtacatcc	atgaaataaa	ttaaaaacc	1200
aaaagtcttt	cttttctatt	ttactagccc	tttagatttg	taagtttctt	ggaattgaag	1260
ctagggtatg	ggaataggaa	tggctaaact	ttcctccata	cccaaaagtc	ttttcttttc	1320
tttttttttc	agtctttatt	atttctattt	acgtctcaaa	acaaatggat	ctacttctta	1380
tactctcaat	ttccttcatc	taaacaccat	tttttttctc	agattttcac	tgattcatac	1440
ttatcatcta	tttcccaaaa	ggtaaataat	attctaaaca	ggcagatcgg	gtagtgtttg	1500
15 caggtacaag	tgatcaaaga	ga				1522

<210> 2
 <211> 20
 <212> ADN
 <213> Secuencia Artificial

5 <220>
 <223> cebador TSOL454_directo

<400> 2

tcgaggccat acattctcag 20

<210> 3
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

10 <220>
 <223> cebador TSOL455_inverso

<400> 3

gtcttacaca ctctacacat c 21

<210> 4
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

20 <220>
 <223> cebador TSOL456_directo

<400> 4

agaggaactc ccactcccta c 21

<210> 5
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

25 <220>
 <223> cebador TSOL457_inverso

<400> 5

atgtagagta catatggttg a 21

<210> 6
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

35 <220>
 <223> cebador TSOL458_directo

<400> 6

40 catccatctg cttactccca t 21

<210> 7
 <211> 20
 <212> ADN
 <213> Secuencia Artificial
 5 <220>
 <223> cebador TSOL459_inverso
 <400> 7
 tacatcatac gaactcaca 20
 10 <210> 8
 <211> 22
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> cebador TSOL460_directo
 15 <400> 8
 caaaccagct gatgcaagca gc 22
 <210> 9
 <211> 22
 <212> ADN
 20 <213> Secuencia Artificial
 <220>
 <223> cebador TSOL461_inverso
 <400> 9
 caacaactag gctttcaact gt 22
 25 <210> 10
 <211> 19
 <212> ADN
 <213> Secuencia Artificial
 <220>
 30 <223> cebador TSOL472_directo
 <400> 10
 ggtacaagt atcaaagag 19
 <210> 11
 <211> 20
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> cebador TSOL473_directo
 <400> 11
 40 attggccgga actcttacca 20

<210> 12
 <211> 21
 <212> ADN
 <213> Secuencia Artificial
 5 <220>
 <223> cebador TSOL465_inverso
 <400> 12
 ggacctgatc ctagtaattc c 21
 <210> 13
 10 <211> 20
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> cebador TSOL512_inverso
 15 <400> 13
 agccattcct attccatac 20
 <210> 14
 <211> 22
 <212> ADN
 20 <213> Secuencia Artificial
 <220>
 <223> cebador TSOL502
 <400> 14
 ctgttctatc tttccttct tg 22
 25 <210> 15
 <211> 22
 <212> ADN
 <213> Secuencia Artificial
 <220>
 30 <223> cebador TSOL503
 <400> 15
 agaaagaaag catgcattta gg 22
 <210> 16
 <211> 20
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> cebador TSOL500
 <400> 16
 40 atacatgatg ggttctcttc 20

<210> 17
 <211> 20
 <212> ADN
 <213> Secuencia Artificial

5 <220>
 <223> cebador TSOL501

<400> 17

aagcatgcat ttaggtaaag 20

10 <210> 18
 <211> 35
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> cebador TSOL558_directo

15 <400> 18

gcgcggtacc gaattccata tgtatattat atatt 35

<210> 19
 <211> 27
 <212> ADN
 <213> Secuencia Artificial

20 <220>
 <223> cebador TSOL559_inverso

<400> 19

tataactagt agtgtgctgg aattcgc 27

25 <210> 20
 <211> 28
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> cebador TSOL560_inverso

30 <400> 20

gcatccatgg tctcttgat cacttgta 28

<210> 21
 <211> 1815
 <212> ADN
 <213> Secuencia Artificial

35 <220>
 <223> secuencia de ácido nucleico que codifica beta-glucuronidasa

<400> 21

ES 2 632 007 T3

atggtaccgc gtactgtaga aacccaacc cgtgaaatca aaaaactcga cggcctgtgg 60
 gcattcagtc tggatcgca aaactgtgga attgatcagc gttggtggga aagcgcgtta 120
 caagaaagcc gggcaattgc tgtgccaggc agttttaacg atcagttcgc cgatgcagat 180
 attcgttaatt atgcgggcaa cgtctggtat cagcgcgaag tctttatacc gaaaggttgg 240
 gcaggccagc gtatcgtgct gcgtttcgat gcggtcactc attacggcaa agtgtgggtc 300
 aataatcagg aagtgatgga gcatcagggc ggctatacgc catttgaagc cgatgtcacg 360
 ccgtatgta ttgccgggaa aagtgtacgt atcaccgttt gtgtgaacaa cgaactgaac 420
 tggcagacta tcccgccggg aatggtgatt accgacgaaa acggcaagaa aaagcagctt 480
 tacttccatg atttctttaa ctatgccgga atccatcgca gcgtaatgct ctacaccacg 540
 ccgaacacct gggtagcga tatcaccgtg gtgacgcatg tcgcgcaaga ctgtaaccac 600
 gcgtctgttg actggcaggt ggtggccaat ggtgatgtca gcgttgaact gcgtgatgcg 660
 gatcaacagg tggttgcaac tggacaaggc actagcggga ctttgcaagt ggtgaatccg 720
 cacctctggc aaccgggtga aggttatctc tatgaactgt gcgtcacagc caaaagccag 780
 acagagtgtg atatctaccc gcttcgctc ggcatccggt cagtggcagt gaagggcgaa 840
 cagttcctga ttaaccacaa accgttctac tttactggct ttggtcgtca tgaagatgcg 900
 gacttgctg gcaaaggatt cgataacgtg ctgatggtgc acgaccacgc attaatggac 960
 tggattgggg ccaactccta ccgtacctcg cattaccctt acgctgaaga gatgctcgac 1020
 tgggcagatg aacatggcat cgtggtgatt gatgaaactg ctgctgtcgg cttaacctc 1080
 tctttaggca ttggtttcga agcgggcaac aagccgaaag aactgtacag cgaagaggca 1140
 gtcaacgggg aaactcagca agcgcactta caggcgatta aagagctgat agcgcgtgac 1200
 aaaaaccacc caagcgtggt gatgtggagt attgccaacg aaccggatac ccgtccgcaa 1260
 ggtgcacggg aatatttcgc gccactggcg gaagcaacgc gtaaactcga cccgacgct 1320
 ccgatcacct gcgtcaatgt aatgttctgc gacgctcaca ccgataccat cagcgtctc 1380
 tttgatgtgc tgtgcctgaa ccgttattac ggatggtatg tccaagcgg cgatttgaa 1440
 acggcagaga aggtactgga aaaagaactt ctggcctggc aggagaaact gcatcagccg 1500
 attatcatca ccgaatacgg cgtggatac ttagccgggc tgcactcaat gtacaccgac 1560
 atgtggagtg aagagtatca gtgtgcatgg ctggatatgt atcaccgct ctttgatcgc 1620
 gtcagcgcgg tcgtcgtgta acaggtatgg aatttcgccc attttgcgac ctcgcaaggc 1680
 atattgcgcg ttggcggtaa caagaaagg atcttcactc gcgaccgcaa accgaagtcg 1740
 gcggcttttc tgtgcaaaa acgctggact ggcatgaaact tcggtgaaaa accgcagcag 1800
 ggaggcaaac aatga 1815

REIVINDICACIONES

1. Una secuencia de ácido nucleico que comprende una secuencia de nucleótidos seleccionada de
 - 5 (a) SEQ ID NO: 1 o un fragmento del mismo, en donde dicho fragmento comprende al menos 600 nucleótidos consecutivos de SEQ ID NO: 1, comprendiendo el fragmento, además, un motivo caja T/G que tiene la secuencia de nucleótidos desde la posición de nucleótidos 298 a 303 de SEQ ID NO: 1 y un motivo de unión MYB que tiene la secuencia de nucleótidos desde la posición de nucleótidos 846 a 851 de SEQ ID NO: 1 y en donde dicho fragmento, cuando se introduce en algodón, tiene actividad del promotor específica para la semilla con altos niveles de expresión en fibras desde el día 5 post-antesis hasta el día 30 post-antesis;
 - 10 (b) una secuencia de nucleótidos con al menos 95% de identidad de secuencia con la secuencia de nucleótidos de (a), comprendiendo dicho fragmento, además, un motivo caja T/G que tiene la secuencia de nucleótidos desde la posición de nucleótidos 298 a 303 de SEQ ID NO: 1 y un motivo de unión MYB que tiene la secuencia de nucleótidos desde la posición 846 a 851 de SEQ ID NO: 1 y que tiene actividad del promotor específica para la semilla con altos niveles de expresión en fibras desde el día 5 post-antesis hasta el día 30 post-antesis cuando se introduce en algodón;
 - 15 (c) una secuencia de nucleótidos complementaria a la secuencia de nucleótidos de una cualquiera de (a) o (b).
2. El ácido nucleico de la reivindicación 1, en donde dicha actividad del promotor específico para la semilla es específica para tricomas.
- 20 3. Un gen quimérico que comprende el ácido nucleico de la reivindicación 1 o la reivindicación 2, enlazado operativamente a una secuencia de ácido nucleico que codifica un producto de expresión de interés, y opcionalmente una secuencia de terminación de la transcripción y de poliadenilación.
4. El gen quimérico de la reivindicación 3, en donde dicho producto de expresión de interés es una proteína o una molécula de ARN capaz de modular la expresión de un gen endógeno a dicha planta.
- 25 5. El gen quimérico de la reivindicación 3 o 4, en donde dicho producto de expresión es un gen informador o un gen específico para fibras.
6. El gen quimérico de la reivindicación 4, en donde dicha molécula de ARN comprende una primera y una segunda región de ARN, en donde
 - 30 1. dicha primera región de ARN comprende una secuencia de nucleótidos de al menos 19 nucleótidos consecutivos que tienen al menos aproximadamente 94% de identidad de secuencia con la secuencia de nucleótidos de dicho gen endógeno;
 2. dicha segunda región de ARN comprende una secuencia de nucleótidos complementaria a dichos 19 nucleótidos consecutivos de dicha primera región de ARN; y
 - 35 3. dicha primera y segunda región de ARN son capaces de apareamientos de bases para formar una molécula de ARN de doble cadena entre al menos dichos 19 nucleótidos consecutivos de dicha primera y segunda región.
7. Un vector, que comprende el gen quimérico de una cualquiera de las reivindicaciones 3 a 6.
8. Una célula de planta de algodón transgénica que comprende el gen quimérico de una cualquiera de las reivindicaciones 3 a 6 o el vector de la reivindicación 7.
- 40 9. Una planta de algodón transgénica que comprende el gen quimérico de una cualquiera de las reivindicaciones 3 a 6 o el vector de la reivindicación 7, establemente integrado en su genoma o que consiste en la célula de la planta de algodón transgénica de la reivindicación 8.
10. La planta transgénica de la reivindicación 9, que es *G. hirsutum*, *G. barbadense*, *G. arboreum* o *G. herbaceum*.
11. Una semilla generada a partir de una planta transgénica de acuerdo con la reivindicación 9 o 10, en donde la semilla comprende el gen quimérico de acuerdo con una cualquiera de las reivindicaciones 3 a 6.

12. Fibras de algodón obtenibles a partir de la planta transgénica de la reivindicación 9 o 10.

13. Un método de efectuar la expresión específica para la semilla de un producto en algodón, que comprende introducir el gen quimérico de una cualquiera de las reivindicaciones 3 a 6 o el vector de la reivindicación 7 en el genoma de una planta de algodón; o

5 proporcionar una planta transgénica de la reivindicación 9 o 10.

14. El método de la reivindicación 13 que comprende, además, cultivar dicha planta hasta generar semillas.

15. Uso del gen quimérico de una cualquiera de las reivindicaciones 3 a 6, del vector de la reivindicación 7 o de la planta transgénica de la reivindicación 9 o 10 para la expresión específica para la semilla de un producto en algodón, o para alterar las propiedades de las fibras en algodón.

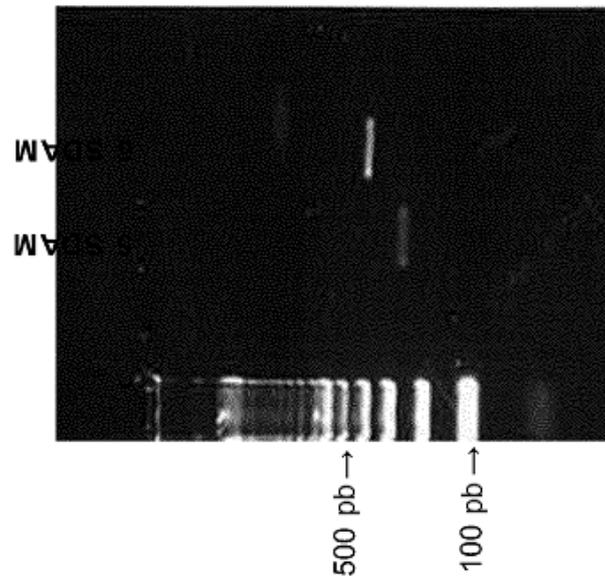
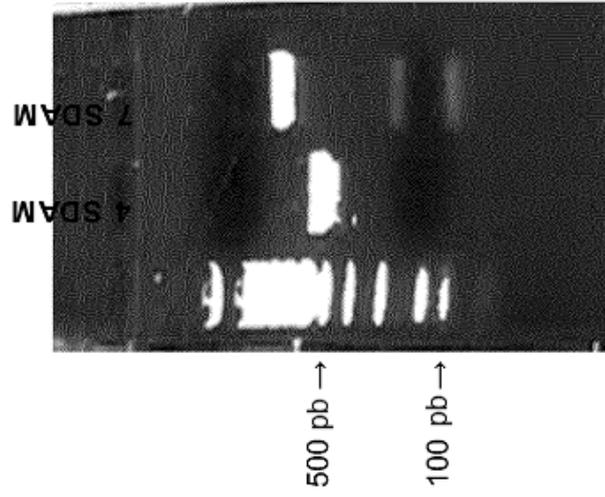


Figura 1

Figura 2

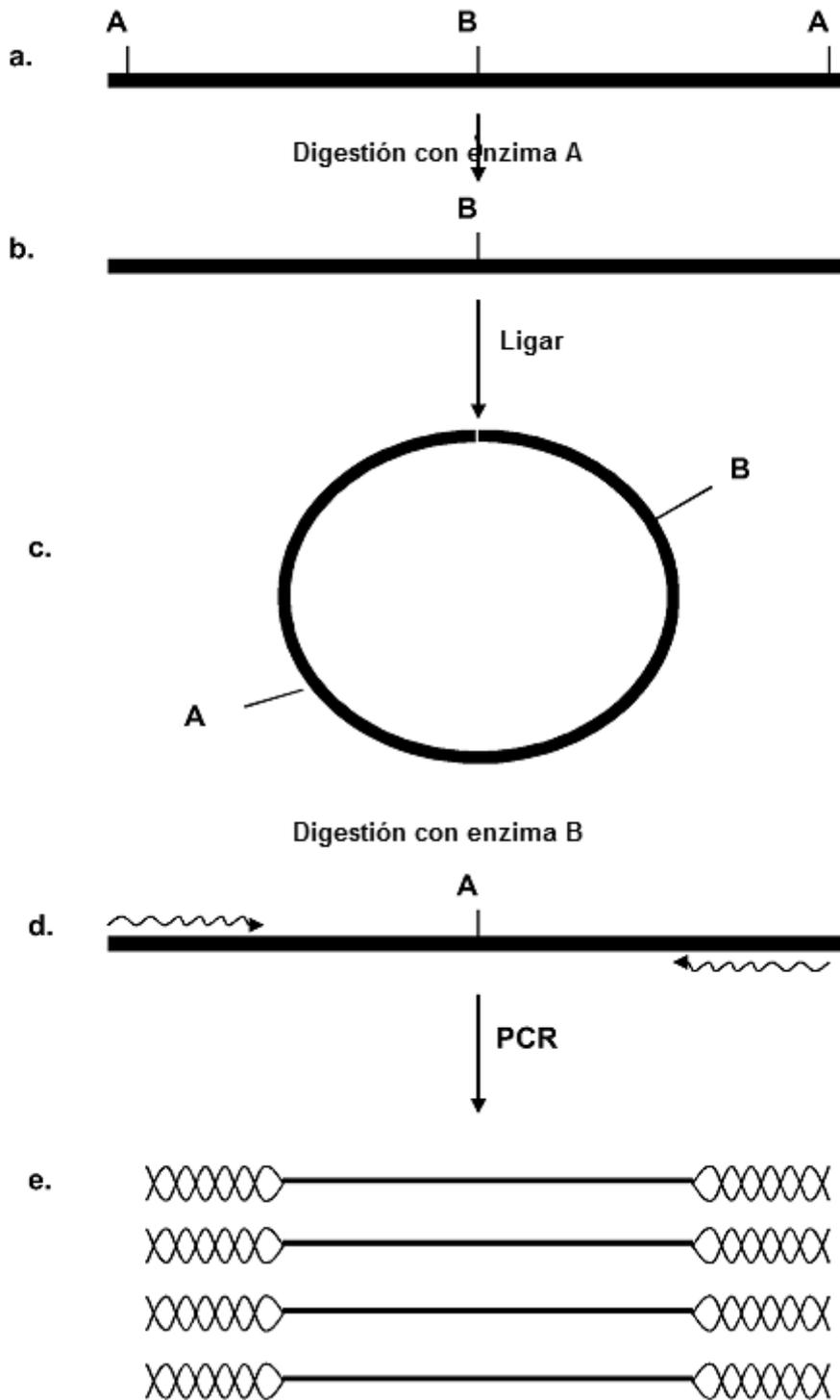


Figura 3a

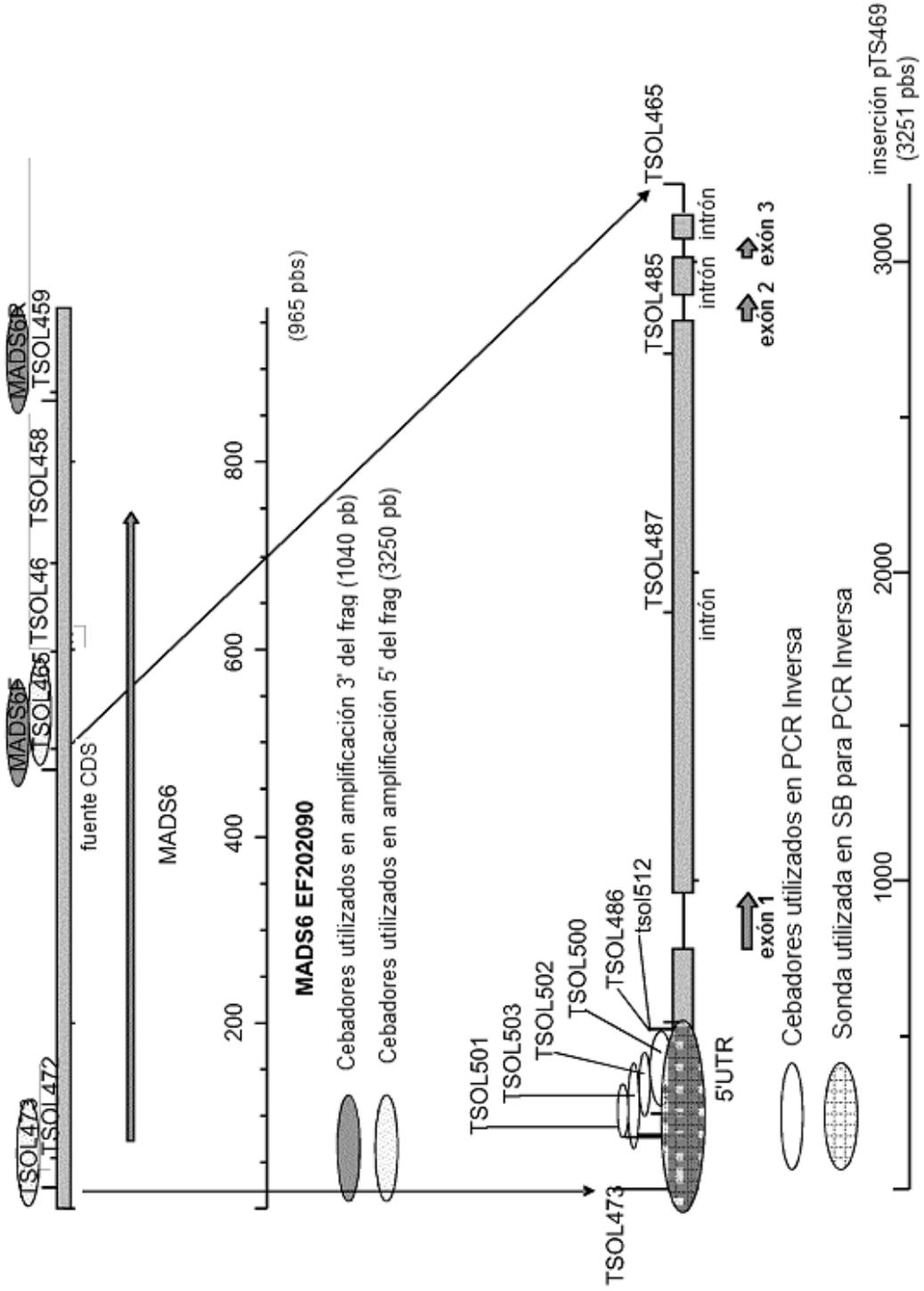


Figura 3b

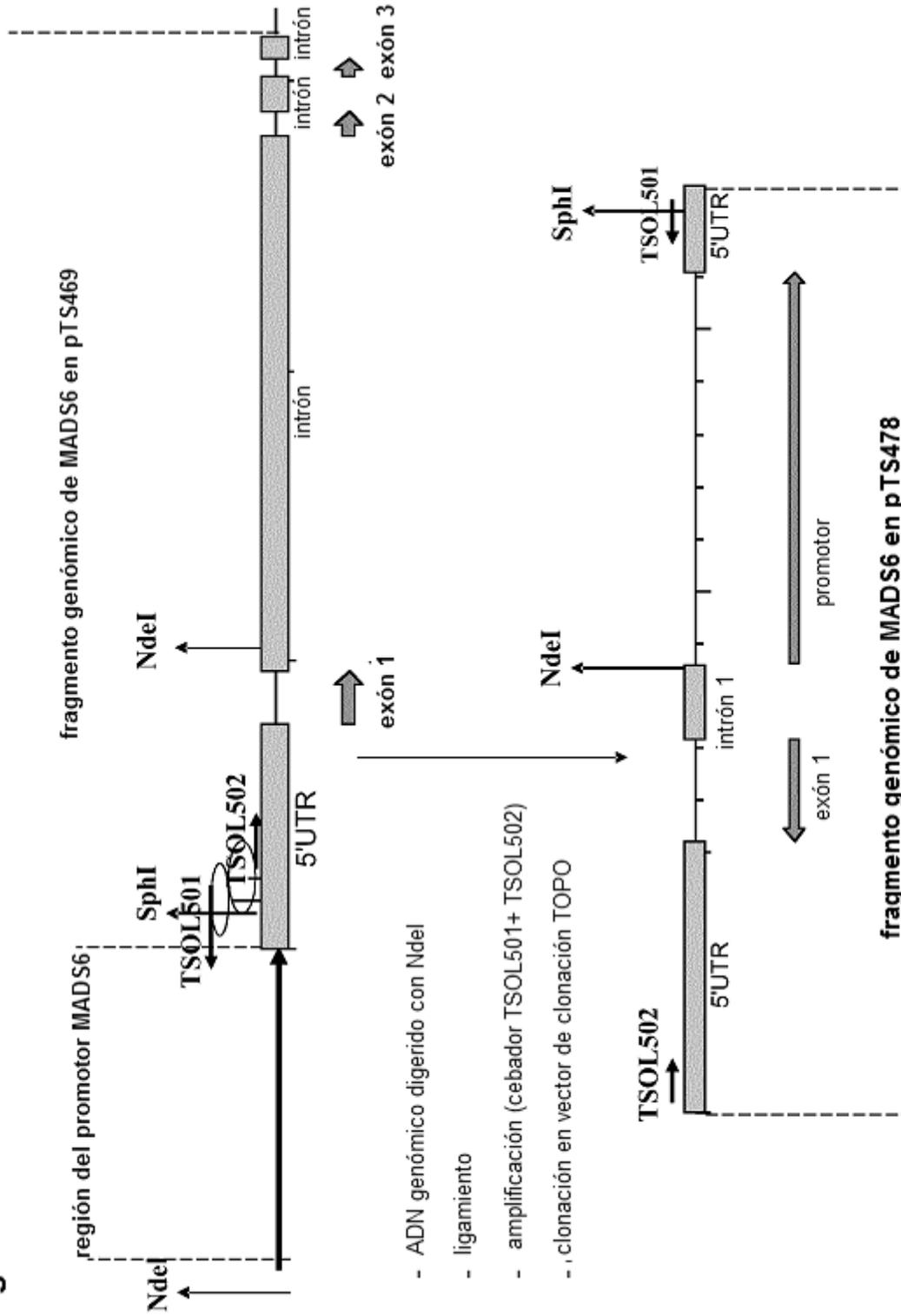
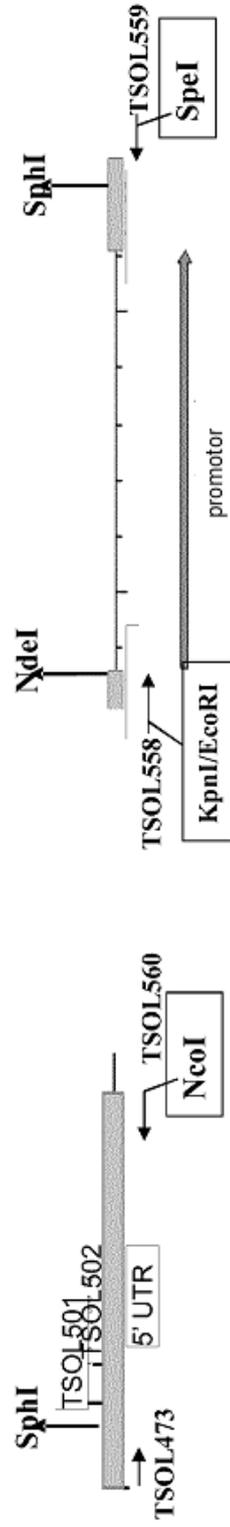


Figura 3c



parte de la inserción pTS478

parte de la inserción pTS469

Figura 3d

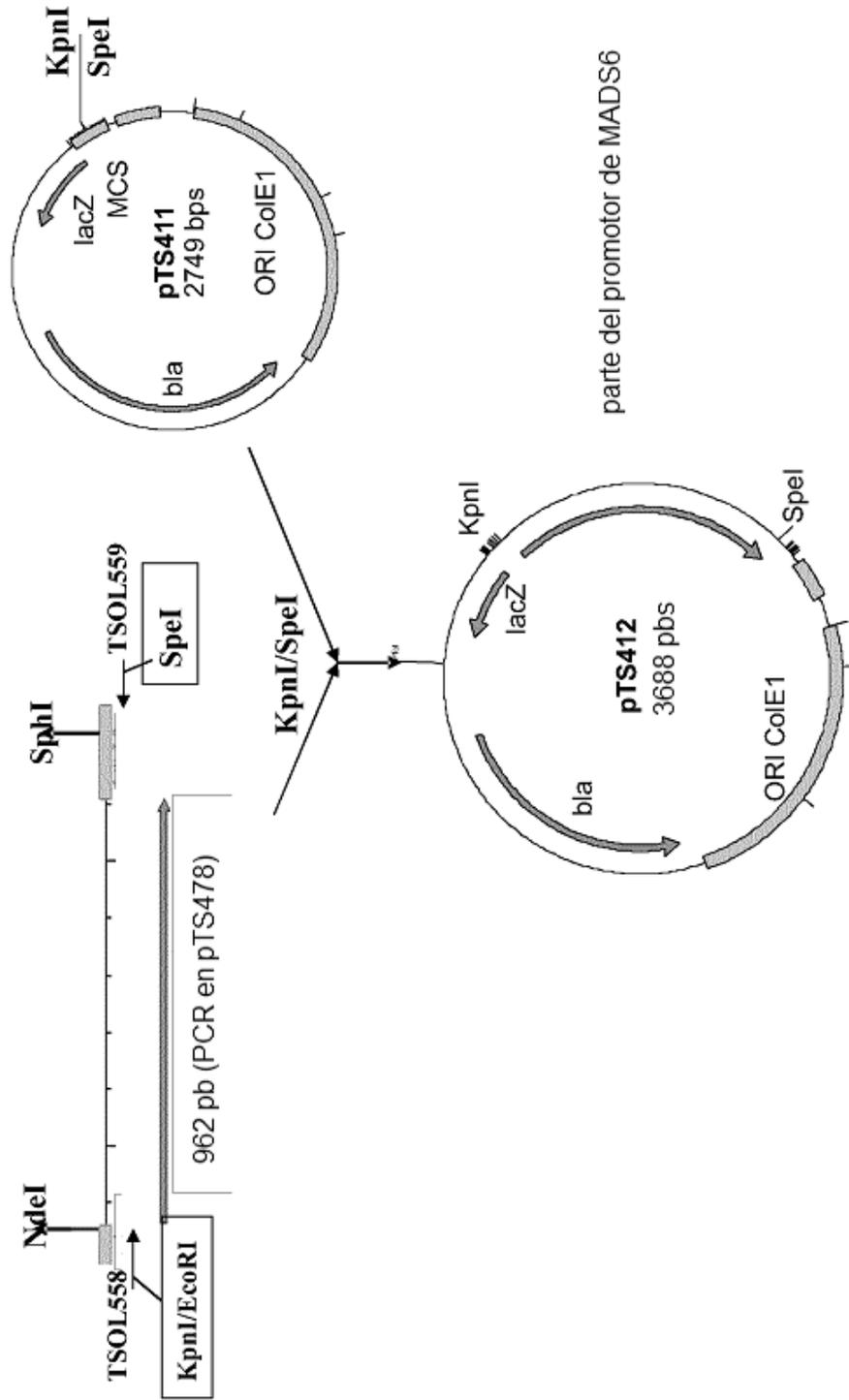


Figura 3e

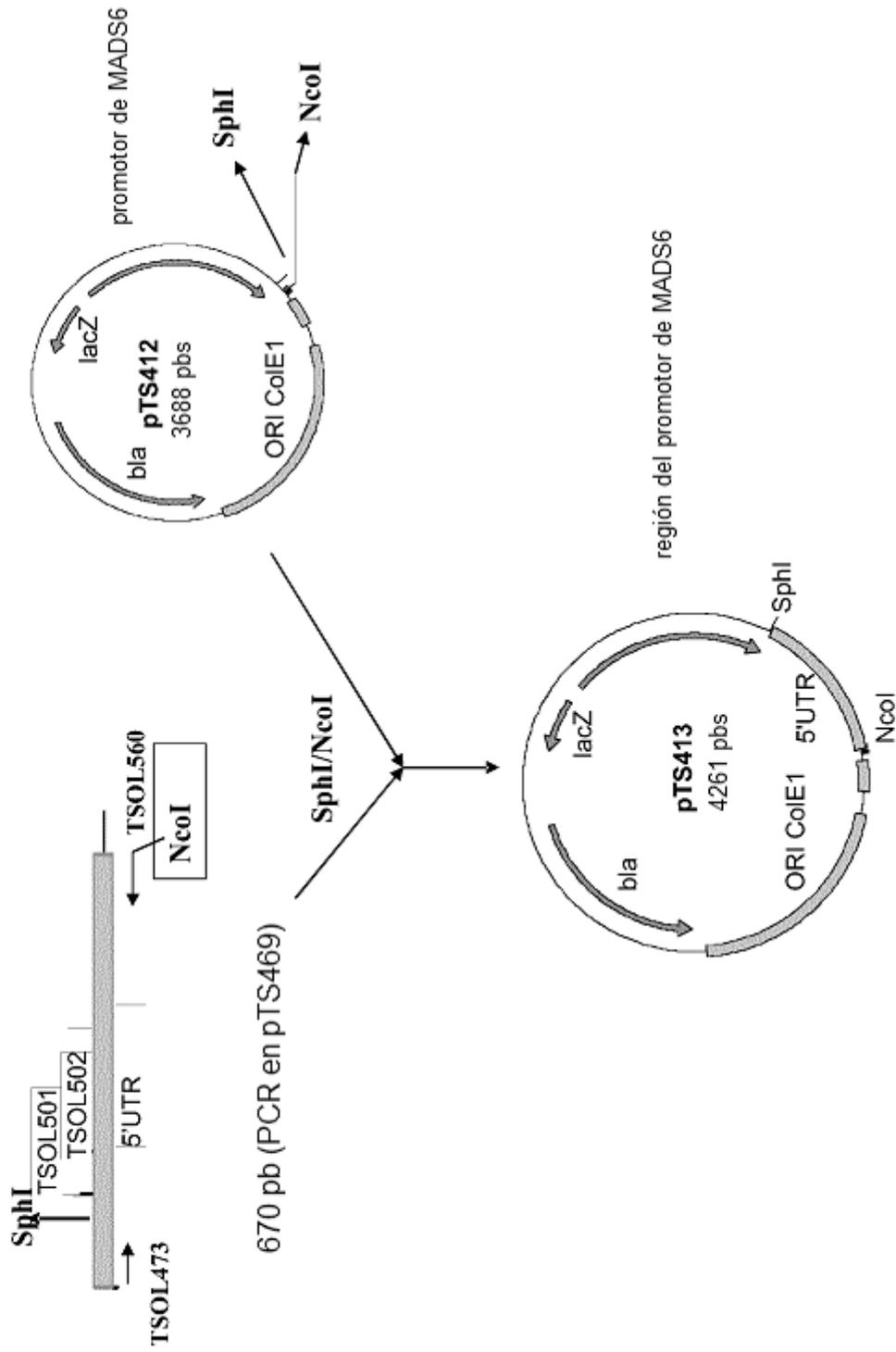


Figura 4

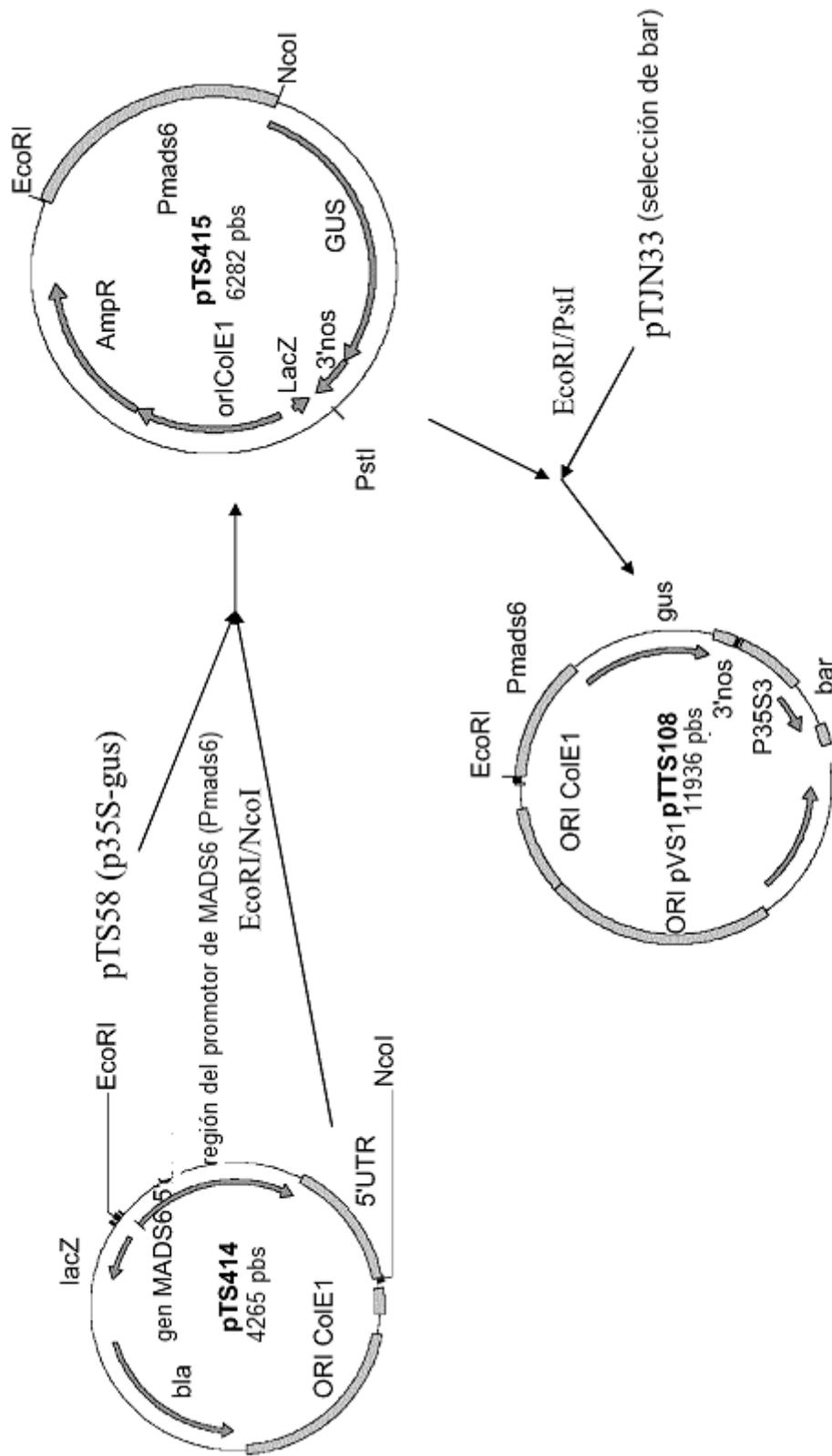


Figura 5a

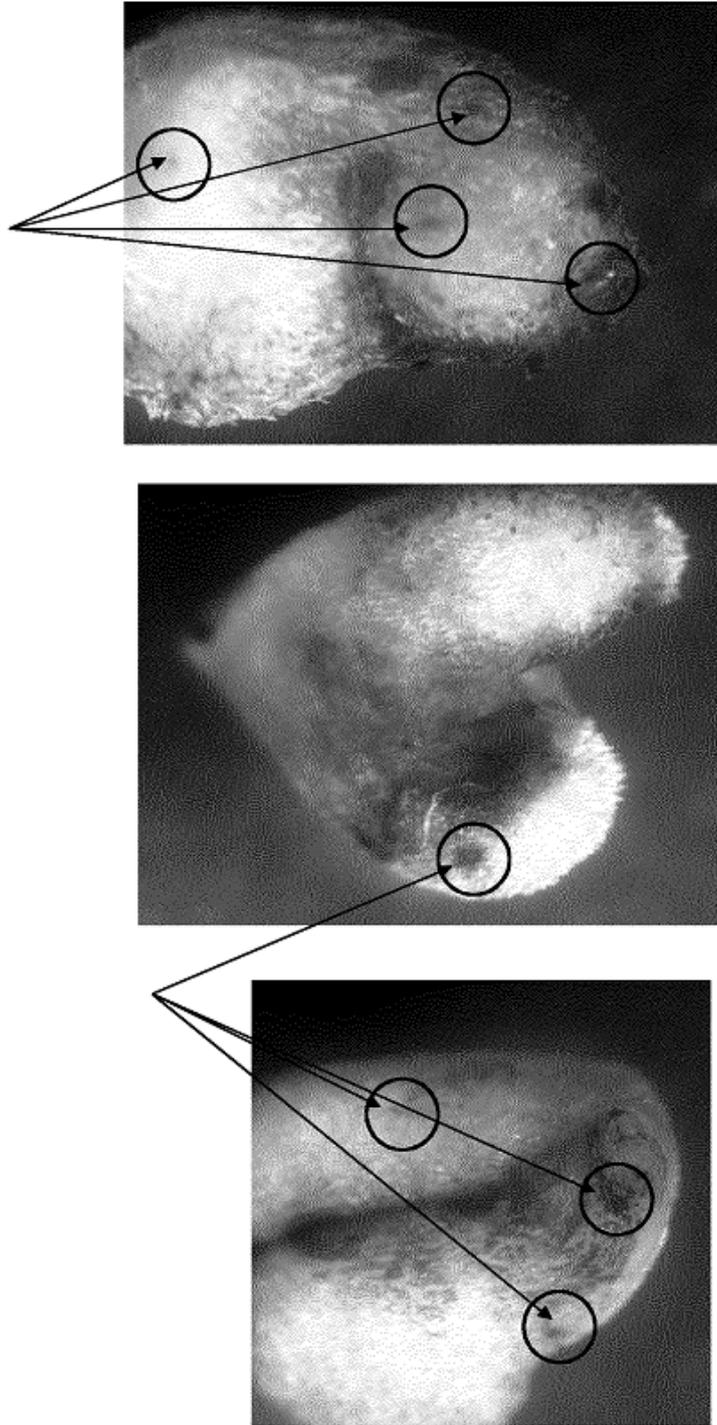


Figura 5b

