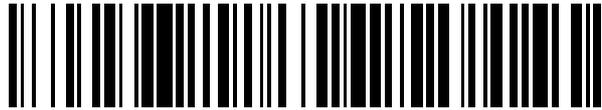


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 463 474**

51 Int. Cl.:

**C12N 15/82** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.05.2005 E 05104022 (8)**

97 Fecha y número de publicación de la concesión europea: **02.04.2014 EP 1605051**

54 Título: **Procedimiento para aumentar la expresión de transgén**

30 Prioridad:

**13.05.2004 EP 04102108**  
**18.05.2004 US 572141 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**28.05.2014**

73 Titular/es:

**CROPDESIGN N.V. (100.0%)**  
**Technologiepark 21C**  
**9052 Zwijnaarde , BE**

72 Inventor/es:

**HATZFELD, YVES;**  
**BROEKAERT, WILLEM;**  
**DE WILDE, CHRIS y**  
**ZHOU, ZHONGYI**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 463 474 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

Procedimiento para aumentar la expresión de transgén

La presente invención se refiere a procedimientos para aumentar la expresión de transgén en plantas. En particular, la presente invención se refiere al uso de la región no traducida 5' (5'UTR) de un gen GOS2, particularmente al uso del primer intrón de un gen GOS2 para potenciar la transcripción. La presente invención también se refiere a procedimientos para modificar la composición de semillas y en particular para aumentar la transcripción o expresión de proteínas en plantas.

Los niveles de proteínas en una célula se determinan por un lado por su velocidad de degradación y por otro lado por su velocidad de síntesis. La velocidad de síntesis proteica depende de diversos procesos, tales como transcripción, procesamiento de ARN postranscripcional, traducción y modificaciones postraduccionales. La primera etapa en la síntesis proteica es la transcripción, el proceso de copiar ADN a una molécula de ARN. Además del promotor, que es la parte más importante de un gen que controla el nivel y patrón de transcripción, también se conoce bien que la región no traducida 5' (5'UTR) de un transcrito de ARN primario tiene influencia en el nivel y patrón de expresión de genes vegetales. Se sabe que los genes vegetales contienen un intrón en su 5'UTR; los estudios de regulación y expresión génica han demostrado que estos intrones pueden tener un impacto positivo en la transcripción. Los ejemplos incluyen intrones de *Arabidopsis* (Rose y Last, Plant J. 11, 455-464, 1997; Chaubet-Gigot y col., Plant Mol. Biol. 45, 17-30, 2001), semilla de ricino (*Ricinus communis*; Tanaka y col., Nucleic Acids Res. 18, 6767-6770, 1990), avena (*Avena sativa*; Bruce y Quail, Plant Cell 2, 1081-1089, 1990), petunia (*Petunia hybrida*; Dean y col., Plant Cell 1, 201-208, 1989; Vain y col., Plant Cell Rep. 15, 489-494, 1996), arroz (*Oryza sativa*; McElroy y col., Plant Cell 2, 163-171, 1990; Snowden y col., Plant Mol. Biol. 31, 689-692, 1996; Rethmeier y col., Plant J. 12, 895-899, 1997), soja (*Glycine max*; Kato y col., Biosci. Biotechnol. Biochem. 62, 151-153, 1998), patata (*Solanum tuberosum*; Leon y col., Plant Physiol. 95, 968-972, 1991; Fu y col., Plant Cell 7, 1387-1394, 1995; Fu y col., Plant Cell 7, 1395-1403, 1995) y tabaco (*Nicotiana tabacum*; Plesse y col., Plant Mol. Biol. 45, 655-667, 2001).

El mecanismo que subyace en esta potenciación mediada por intrón de la transcripción aún no está claro, pero los intrones deben estar en su orientación normal con respecto a la región transcrita del gen (Callis y col., Genes Dev. 1, 1183-1200, 1987; Clancy y col., Plant Sci. 98, 151-161, 1994; Mascarenhas y col., Plant Mol. Biol. 15, 913-920, 1990). Un estudio del intrón 1 de *PAT1* en *Arabidopsis* (Rose, RNA 8, 1444-1453, 2002) demostró que se requería la maquinaria de corte y empalme, pero que el corte y empalme del intrón no era suficiente para potenciar la acumulación de ARNm. De acuerdo con el modelo presentado, los intrones podrían estimular la acumulación de ARNm aumentando la capacidad de procesamiento de la ARN polimerasa II mediante una modificación de su dominio carboxilo terminal, promoviendo de este modo la elongación del transcrito sin afectar significativamente al inicio de la transcripción. Sería por lo tanto más probable que la transcripción de genes que contenían un intrón extendiera la longitud del gen en casos en los que el procesamiento del extremo 3' produjera transcritos estables. Los estudios sugieren que la presencia de un intrón aumenta la probabilidad de que se realicen transcritos estables de longitud completa, conduciendo de este modo a acumulación de ARNm aumentada (Rose, 2002). Clancy y Hannah (Plant Physiol. 130, 918-929, 2002) usaron fusiones de genes indicadores para identificar elementos del primer intrón de *Sh1* requerido para la potenciación en células de maíz cultivado. Descubrieron que un derivado de 145 pb del intrón *Sh1* confería aproximadamente la misma estimulación de 20 a 50 veces típica para el intrón de longitud completa en su sistema de expresión transitoria. Además, se descubrió que se requería un motivo de 35 pb contenido dentro del intrón para niveles máximos de potenciación pero no para corte y empalme de transcrito eficaz. El elemento importante de este motivo de pb redundante es la riqueza en T en lugar de la secuencia específica. Cuando se anuló el corte y empalme del transcrito por mutaciones en los límites de los intrones, la potenciación se redujo a aproximadamente 2 veces. Se descubrió que el requisito de corte y empalme para potenciación no se debía a codones de inicio de la traducción cadena arriba contenidos en transcritos no cortados. Los autores concluyeron que el corte y empalme del intrón *Sh1* era integral para la potenciación y sugirieron que las modificaciones de transcrito desencadenadas por el motivo rico en T y corte y empalme pueden ligar el ARNm con el sistema de tráfico de la célula. Otros informes sugieren que la longitud y composición de las secuencias flanqueantes pueden contribuir a la expresión génica potenciada mediada por intrones (Sinibaldi y Mettler, In WE Cohn, K Moldave, eds, Progress in Nucleic Acid Research and Molecular Biology, Vol. 42. Academic Press, Nueva York, pp 229-257, 1992; Luehrsen y Walbot, Prog. Nucleic Acid Res. Mol. Biol. 47, 149-193, 1991; Maas y col., Plant Mol. Biol. 16, 199-207, 1991; Clancy y col., 1994), o que la magnitud de la estimulación también depende de las secuencias codificantes (Sinibaldi y Mettler, 1992; Rethmeier y col., 1997; Rethmeier y col., Plant J. 13, 831-835, 1998), el tejido de expresión y las condiciones fisiológicas (Tanaka y col., 1990; Sinibaldi y Mettler, 1992; Gallie y Young, Plant Physiol. 106, 929-939, 1994; Fu y Park, 1995a, 1995b; Chaubet-Gigot y col., 2001; Plesse y col., 2001).

*GOS2* del arroz es un gen de copia única que codifica el factor de inicio de la traducción en IF1, que está implicado en el ensamblaje del complejo 43S con ARNm y en la exploración correcta de la 5'UTR hacia el ATG durante el inicio de la traducción. El gen *GOS2* también se encuentra en otros organismos tales como levadura y mamíferos, y para cada uno de estos, el primer intrón se localiza en la 5'UTR. En su ambiente natural, la proteína *GOS2* del arroz se expresa bajo el control de su promotor fuerte y constitutivo. Este promotor se usa ampliamente como una alternativa al promotor CaMV35S en la ingeniería genética de plantas monocotiledóneas. Hensgens y col. (Plant Mol. Biol. 23, 643-669, 1993) estudiaron la expresión del gen indicador *gusA* conducido por el promotor *GOS2*. Describen un primer tipo de construcción que comprende, de 5' a 3', un promotor transcripcional de *GOS2*, el sitio de

inicio de la traducción, los dos primeros exones e intrones y parte del tercer exón, una región de poliligador, la secuencia codificante de *gusA* y una región de poliadenilación. Hensgens y col. también describen un segundo tipo de construcción, en la que *gusA* estaba fusionado con los primeros 36 nucleótidos del primer exón de *GOS2* y por lo tanto carecía de los primeros dos intrones. Se observó la mayor expresión de GUS para el primer tipo de construcción, sin embargo los autores no pudieron determinar si la diferencia en el nivel de expresión entre los dos tipos de construcciones era atribuible a la presencia de los dos intrones. Hasta la fecha, no se ha realizado ningún análisis del papel de los intrones en el gen *GOS2* en el proceso de transcripción.

Se determinan propiedades fenotípicas de plantas por la expresión de uno o más genes en el genoma vegetal. El nivel y patrón de expresión apropiados (patrones de expresiones temporales y espaciales) de los genes pueden depender de muchos elementos reguladores, incluyendo el tipo de promotor y las regiones no traducidas del ARNm. Estos elementos reguladores pueden usarse ventajosamente para adaptar la expresión de los genes a necesidades particulares. Por ejemplo, un promotor puede tener la expresión espacial deseada (tal como expresión específica de raíz o específica de flor), pero puede carecer de la fuerza para asegurar niveles de expresión altos. En dichos casos, un experto en la materia puede considerar usar elementos reguladores adicionales para potenciar los niveles de expresión. Se conocen en la técnica ejemplos de elementos reguladores para modular los niveles o patrones de expresión e incluyen secuencias activadoras corriente arriba, elementos próximos al promotor, potenciadores y silenciadores. Se han descrito varios genes para los que la presencia de intrones contribuyó significativamente al nivel de expresión. Sin embargo, el mecanismo de este aumento apenas se entiende.

Para obtener los niveles o patrones de expresión génica deseados, que pueden ajustarse dependiendo de las necesidades específicas, es deseable tener a disposición una amplia serie de intrones que sean capaces de estimular la expresión génica. Existe la necesidad en diversos campos de la ingeniería genética de elementos de control de la transcripción capaces de aumentar la expresión génica en plantas. Los inventores han mostrado ahora por primera vez que la región no traducida 5' (5'UTR), y particularmente el primer intrón de un gen *GOS2* de planta, tiene un efecto estimulador en la expresión de transgenes en plantas cuando se inserta entre un promotor y un ácido nucleico para expresar. Se descubrió que este era el caso no solamente cuando el primer intrón de un gen *GOS2* de planta se combinaba con el promotor de *GOS2* de arroz, sino que también se observó sorprendentemente cuando una 5'UTR de *GOS2*, y en particular su primer intrón, estaba unida operativamente con un promotor no *GOS2* expresable en plantas.

En una realización, la invención se refiere a un procedimiento para aumentar la expresión génica en una planta transgénica, comprendiendo dicho procedimiento

- a. integrar el primer intrón de la 5' UTR de un gen *GOS2* de planta en el extremo 5' de un ácido nucleico de interés, creando de este modo una unidad transcripcional quimérica, en la que dicho primer intrón se representa en SEC ID N°: 5 o una variante funcional de dicho primer intrón, cuya variante funcional es de al menos 65 pares de base de longitud, comprende sitios de corte y empalme y una adenosina de punto de ramificación funcional, dicha adenosina de punto de ramificación es parte del motivo conservado con la siguiente secuencia consenso  $Y_{100}U_{100}R_{64}A_{100}U_{50}$  o  $Y_{100}U_{100}R_{64}A_{100}Y_{70}$ , en la que Y representa los nucleótidos C o T, y R representa A y G, y en la que los números en subíndice indican el porcentaje de conservación del nucleótido respectivo, cuya variante funcional es capaz de hibridar en condiciones rigurosas con el primer intrón como se representa en SEC ID N°: 5, en el que dichas condiciones de hibridación son hibridación en SCC 4 x a 65 °C, seguido de una etapa de lavado en SCC 0,1 x a 65 °C durante una hora;
- b. fusionar de forma operativa dicha unidad transcripcional quimérica con un promotor expresable en plantas para obtener un casete de expresión, a condición de no haya combinación de la 5' UTR como se representan en SEC ID N°: 2 del gen *GOS2* del arroz con un promotor del gen *GOS2* del arroz;
- c. introducir en y expresar en una célula vegetal dicho casete de expresión de (b), para crear una célula vegetal transgénica; y
- d. regenerar y/o cultivar una planta a partir de dicha célula vegetal transgénica de (c) para que dicha célula transgénica transcriba el transgén y en la que la 5'UTR del ácido nucleico expresado de interés comprende al menos el primer intrón de la 5'UTR como se ha definido en (a).

Se describe un procedimiento para aumentar la expresión transgénica en una planta transgénica, comprendiendo dicho procedimiento

- a. integrar toda o parte de una 5' UTR de un gen *GOS2* de planta en el extremo 5' de un ácido nucleico de interés, creando de este modo una unidad transcripcional quimérica, en la que dicha parte de una 5'UTR comprende al menos el primer intrón de dicho gen *GOS2* o una variante funcional de dicho primer intrón;
- b. fusionar de forma operativa la unidad transcripcional quimérica con un promotor expresable en plantas para obtener un casete de expresión, a condición de no haya combinación de una 5' UTR completa del gen *GOS2* del arroz con un promotor del gen *GOS2* del arroz;
- c. introducir en y expresar en una célula vegetal el casete de expresión de (b), para crear una célula vegetal transgénica; y
- d. regenerar y/o cultivar una planta a partir de la célula vegetal transgénica de (c) de modo que la célula transgénica transcriba el transgén y en la que la 5'UTR del ácido nucleico expresado de interés comprende al menos la 5'UTR (a).

El objeto de la presente invención se refiere al primer intrón de un gen *GOS2*. El gen *GOS2* (secuencia codificada proporcionada en SEC ID N°: 3), también conocido como *SUI1*, codifica un factor de inicio de la traducción eIF1 (SEC ID N°: 4). En una realización particular, se usa el primer intrón del gen *GOS2* de arroz, pero un experto en la materia apreciará que pueden usarse también los primeros intrones de otros genes *GOS2* en los procedimientos de la invención. El gen *GOS2* del arroz contiene cuatro intrones de respectivamente 998 pb, 94 pb, 90 pb y 105 pb (de Pater y col., Plant J. 2, 837-844, 1992). El primer intrón se localiza dentro de la 5'UTR (Figura 1), no solamente en arroz (X51910), sino también en *Coffea arabica* (AJ519840), y *Arabidopsis* (At5g54940, At5g54760, At1g54290, At4g27130). Puede esperarse que este sea el caso para todas las plantas, dada la alta conservación de secuencia entre genes *GOS2* de plantas. Por lo tanto, en una realización preferida de la presente invención, se derivan secuencias intrónicas alternativas de genes *GOS2* de genes *GOS2* de plantas, porque en estos genes los primeros intrones se localizan en la 5'UTR. Más preferentemente, el primer intrón de un gen *GOS2* deriva de una planta monocotiledónea, más preferentemente de arroz o maíz. No obstante, un experto en la materia sería capaz de aislar las primeras secuencias intrónicas de *GOS2* de otras plantas sin secuencias codificantes flanqueantes no deseadas usando técnicas convencionales. Por lo tanto, los primeros intrones derivados de cualquier gen *GOS2/SUI1* de planta pueden ser igualmente útiles en los procedimientos de la presente invención.

La expresión "intrón" o "secuencia intermedia" como se define en el presente documento es una secuencia de ácido nucleico no codificante que interrumpe frecuentemente secuencias codificantes de genes eucariotas pero que también puede estar presente fuera de la secuencia codificante y que se transcribe en ARN y posteriormente se retiran por corte y empalme de ARN para dejar un ARNm funcional u otro ARN. El intrón comprende sitios de corte y empalme en ambos extremos y una adenosina de punto de ramificación. Se entiende que la expresión "sitios de corte y empalme" como se define en el presente documento significa los extremos exteriores de la secuencia que se separan por corte del transcrito primario y los nucleótidos inmediatos en el transcrito primario de la secuencia que se separa por corte. La adenosina del punto de ramificación es el nucleótido que forma el lazo durante el acontecimiento de corte y empalme. La adenosina de punto de ramificación es parte de un motivo conservado con la siguiente secuencia consenso (Brown, Plant Journal 10, 771-780, 1996): Y<sub>100</sub>U<sub>100</sub>R<sub>64</sub>A<sub>100</sub>U<sub>50</sub> o Y<sub>100</sub>U<sub>100</sub>R<sub>64</sub>A<sub>100</sub>Y<sub>70</sub>, en la que Y representa los nucleótidos C o T, y R representa A o G y en la que los números en subíndice indican el porcentaje de conservación del nucleótido respectivo. Este motivo determina la funcionalidad de la adenosina del punto de ramificación.

Además, también pueden ser útiles variantes funcionales del primer intrón de un gen *GOS2* en el aumento de la expresión de un transgén. Las variantes funcionales abarcan variantes de sustitución, inserción y delección y combinaciones de las mismas. Las "variantes de sustitución" son en las que se ha retirado al menos una base en la secuencia de nucleótidos y se ha insertado una base diferente en su lugar. Las "variantes de inserción" de un ácido nucleico son en las que se introducen uno o más nucleótidos en un sitio predeterminado en la secuencia. Las "variantes de delección" de un ácido nucleico se caracterizan por la retirada de uno o más nucleótidos del ácido nucleico. Puede producirse cualquier combinación de sustitución o sustituciones, delección o delecciones o inserción o inserciones a condición de la funcionalidad del intrón permanezca esencialmente igual, es decir, que la variante intrónica, cuando se usa en los procedimientos de acuerdo con la presente invención, provoque expresión aumentada de un transgén. La expresión aumentada debe entenderse como aumentada en comparación con la expresión del mismo transgén sin la presencia del intrón. Además, una variante funcional es de al menos 65 pares de bases de longitud, comprende sitios de corte y empalme y una adenosina de punto de ramificación funcional, y es capaz de hibridar en condiciones rigurosas con un primer intrón de origen natural de un gen *GOS2* de planta.

El término "hibridación" como se define en el presente documento es un proceso en el que secuencias de nucleótidos complementarias sustancialmente homólogas hibridan entre sí. Para permitir que se produzca hibridación, las moléculas de ácido nucleico en general se desnaturalizan de forma térmica o química para fundir una doble cadena en dos cadenas individuales y/o para retirar horquillas u otras estructuras secundarias de ácidos nucleicos monocatenarios. La rigurosidad de la hibridación se ve influida por condiciones tales como la temperatura, la concentración salina y la composición del tampón de hibridación. Las "condiciones de hibridación rigurosas" y "condiciones de lavado de hibridación rigurosas" en el contexto de los experimentos de hibridación de ácido nucleico tales como hibridaciones de Southern y Northern dependen de la secuencia y son diferentes en diferentes parámetros ambientales. Por ejemplo, secuencias más largas hibridan específicamente a temperaturas mayores. La T<sub>m</sub> es la temperatura con fuerza iónica y pH definidos, a la que el 50 % de la secuencia diana hibrida con una sonda perfectamente coincidente. La especificidad es típicamente la función de lavados posthibridación. Los factores críticos de dichos lavados incluyen la fuerza iónica y temperatura de la solución de lavado final. Generalmente, se seleccionan condiciones rigurosas que sean aproximadamente 50 °C inferiores al punto de fusión térmica (T<sub>m</sub>) para la secuencia específica a una fuerza iónica y pH definidos. La T<sub>m</sub> es la temperatura con fuerza iónica y pH definidos, a la que el 50 % de la secuencia diana hibrida con una sonda perfectamente coincidente. La T<sub>m</sub> depende de las condiciones de la solución y la composición de base de la sonda, y puede calcularse usando la siguiente ecuación:

$$T_m = 79,8 \text{ } ^\circ\text{C} + (18,5 \times \log[\text{Na}^+]) + (58,4 \text{ } ^\circ\text{C} \times \%[\text{G} + \text{C}]) - (820 \times (\text{N}^\circ \text{ de pb en doble cadena})^{-1}) - (0,5 \times \% \text{ de formamida})$$

Son condiciones rigurosas preferidas cuando la temperatura es de 20 °C inferior a la T<sub>m</sub>, y son condiciones rigurosas más preferidas cuando la temperatura es de 10 °C inferior a la T<sub>m</sub>. La unión no específica también puede controlarse usando una cualquiera de varias técnicas conocidas tales como bloqueo de la membrana con soluciones que contienen proteína, adiciones de ARN heterólogo, ADN y SDS al tampón de hibridación, y tratamiento con RNasa.

Las condiciones de lavado se realizan típicamente a o por debajo de la rigurosidad de hibridación. Generalmente, las condiciones rigurosas adecuadas para ensayos de hibridación de ácido nucleico o procedimientos de detección de amplificación génica son como se ha expuesto anteriormente. También pueden seleccionarse condiciones de rigurosidad aumentada o reducida en comparación con lo anterior.

5 Con objeto de definir el nivel de rigurosidad, puede hacerse referencia convenientemente a Sambrook y col. (2001) Molecular Cloning: a laboratory manual, 3ª Edición Cold Spring Harbor Laboratory Press, CSH, Nueva York o a Current Protocols in Molecular Biology, John Wiley & Sons, N. Y. (1989). El experto en la materia conoce diversos parámetros que pueden alterarse durante la hibridación y el lavado y que mantendrán o cambiarán las condiciones de rigurosidad. Un ejemplo de las condiciones de alta rigurosidad incluye hibridación en SSC 0,1-1x / SDS 0,1 % p/v  
10 a 60 °C durante 1-3 horas. Otro ejemplo de condiciones de hibridación rigurosas es la hibridación a SSC 4x a 65 °C, seguido de un lavado en SSC 0,1x a 65 °C durante aproximadamente una hora. Como alternativa, una condición de hibridación rigurosa ejemplar es en formamida 50 %, SSC 4x a 42 °C. Otro ejemplo más de las condiciones rigurosas incluyen hibridación a 62 °C en SSC 6x, BLOTTO 0,05x y lavado en SSC 2x, SDS 0,1 % p/v a 62 °C.

15 Los genes *GOS2* de planta tienen su primer intrón localizado dentro de la 5'UTR, por lo tanto se prevé que no es solamente este primer intrón sino también el primer intrón con sus secuencias flanqueantes hasta el total de la 5'UTR lo que puede tener un efecto estimulador en la expresión génica. Por lo tanto, la presente invención proporciona en otra realización un procedimiento para aumentar la expresión de transgenes en una planta transgénica en el que se usa sustancialmente la 5'UTR completa de un gen *GOS2* de planta, o cualquier parte del mismo que comprende el primer intrón o una variante funcional del mismo, para aumentar la expresión de un transgén, a condición de que el promotor de *GOS2* de arroz no se use en combinación con la 5'UTR completa del gen *GOS2* del arroz.  
20

La 5'UTR de un gen debe entenderse como la parte de un gen que se transcribe a un transcrito de ARN primario (pre ARNm) y estando dicha parte localizada cadena arriba de la secuencia codificante. El transcrito primario es el producto de ARN inicial, que contiene intrones y exones, producido por la transcripción de ADN. Muchos transcritos primarios deben experimentar procesamiento de ARN para formar la especie de ARN fisiológicamente activa. El procesamiento en un ARNm maduro puede comprender recortar los extremos, retirar intrones, protección de los extremos y/o recortar moléculas de ARNr individuales de sus ARN precursores. La 5'UTR de un ARNm es por lo tanto la parte del ARNm que no se traduce en proteína y que se localiza cadena arriba de la secuencia codificante.  
25

De acuerdo con la presente invención, la 5'UTR de un gen *GOS2* o una parte del mismo que comprende el primer intrón, se integra en el extremo 5' de un ácido nucleico de interés. La fusión da como resultado la formación de una unidad transcripcional, que da lugar a un transcrito de ARN primario funcional individual que puede procesarse a un ARNm maduro en una célula huésped dada. La 5'UTR de un gen *GOS2*, o una parte del mismo que comprende el primer intrón, está presente en la unidad transcripcional en orientación con sentido y cadena arriba de cualquier secuencia codificante de proteína. El intrón útil en los procedimientos de la presente invención habitualmente está flanqueado por las secuencias límite que comprenden el sitio de corte y empalme, pero también puede ser funcional sin estas secuencias límite.  
30  
35

Preferentemente el ácido nucleico de interés es un transgén. La expresión "secuencia o secuencias de ácido nucleico", "ácido nucleico" o "gen o genes", cuando se usa en el presente documento se refiere a nucleótidos, bien ribonucleótidos (ARN) o bien desoxirribonucleótidos (ADN) o una combinación de ambos, en una forma polimérica de cualquier longitud. Estas expresiones incluyen además ADN y ARN bicatenario y monocatenario. Estas expresiones también incluyen modificaciones de nucleótidos conocidas en la técnica, tales como metilación, ciclación, "extremos recubiertos", sustitución de uno o más de los nucleótidos de origen natural con un análogo tal como inosina, y modificaciones de cadena principal polinucleotídica. El término "transgén" se refiere a un ácido nucleico aislado que se introduce en una planta. En una realización preferida de la presente invención, el transgén codifica un polipéptido.  
40  
45

El término "polipéptido" o "proteína" se define como una cadena molecular que comprende dos o más aminoácidos unidos covalentemente por enlaces peptídicos.

El término "aislado" en la presente solicitud significa retirado de su ambiente original. Por ejemplo, un ácido nucleico presente en su estado natural en un organismo no está aislado, mientras que el mismo ácido nucleico separado de los ácidos nucleicos adyacentes en los que está presente de forma natural, se considera "aislado". El transgén puede originarse de la misma especie vegetal o de otra, puede aislarse de cualquier fuente, tal como bacterias, levadura u hongos, plantas (incluyendo algas) o animales (incluyendo seres humanos). El transgén puede modificarse sustancialmente de su forma nativa en composición y/o ambiente genómico mediante manipulación humana deliberada. En una realización particular de la invención, el transgén comprende una secuencia que codifica una proteína. Sin embargo, no es necesario que el transgén codifique una proteína, los procedimientos pueden aplicarse para cualquier ácido nucleico para expresar, tales como ARNt, ARNr, ARN antisentido, ARNip. Una "secuencia codificante" o "fase abierta de lectura" (ORF) se define como una secuencia de nucleótidos que puede transcribirse en ARNm y puede después traducirse a un polipéptido cuando se sitúa bajo el control de secuencias reguladoras apropiadas, es decir, cuando la secuencia codificante u ORF esté presente en una forma expresable. La secuencia codificante u ORF se une con un codón de inicio de la traducción 5' y un codón de parada de la  
50  
55  
60

traducción 3'. Una secuencia codificante u ORF puede incluir, pero sin limitación, ARN, ARNm, ADNc, secuencias de nucleótidos recombinantes, secuencias de nucleótidos fabricadas de forma sintética o ADN genómico. La secuencia codificante u ORF puede estar interrumpida por intrones.

5 Ventajosamente los procedimientos de la presente invención dan como resultado expresión aumentada del transgén. Se entiende que el término "expresión" como se define en el presente documento significa la producción de una proteína o secuencia de nucleótidos en una célula en sí misma o en un sistema sin células. Incluye la transcripción en un producto de ARN, modificación post transcripcional y opcionalmente también traducción a un producto proteico o polipéptido de un ADN que codifica ese producto, así como posibles modificaciones post traduccionales. Una secuencia codificante puede transcribirse en dirección con sentido o antisentido.

10 Como se ha mencionado anteriormente, cualquiera gen *GOS2/SUI1* puede ser una fuente adecuada de una 5'UTR y/o un primer intrón para su uso en un procedimiento para aumentar la expresión de transgenes. Los procedimientos para la búsqueda e identificación de homólogos de eIF1 estarían dentro del alcance de los expertos en la materia. Dichos procedimientos comprenden la comparación de las secuencias representadas por SEC ID N° 3 o 4, en un formato legible por ordenador, con secuencias que están disponibles en bases de datos públicas tales como MIPS (http://mips.gsf.de/), GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html) o EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl/index.html), usando algoritmos bien conocidos en la técnica para el alineamiento o la comparación de secuencias, tales como GAP (Needleman y Wunsch, J. Mol. Biol. 48; 443-453 (1970)), BESTFIT (usando el algoritmo de homología local de Smith y Waterman (Advances in Applied Mathematics 2; 482-489 (1981))), BLAST (Altschul, S. F., Gish, W., Miller, W., Myers, E. W. y Lipman, D. J., J. Mol. Biol. 215: 403-410 (1990)), FASTA y TFASTA (W. R. Pearson y D. J. Lipman Proc. Natl. Acad. Sci. USA 85: 2444- 2448 (1988)). El software para realizar análisis de BLAST está disponible públicamente a través del Centro Nacional para la Información Biotecnológica. Los homólogos adecuados pueden identificarse usando parámetros por defecto de BLAST (matriz BLOSUM62, penalización de apertura de hueco 11 y penalización de extensión de hueco 1). Preferentemente, se usan secuencias de longitud completa (desde el codón de inicio al de parada en el caso del ADN, o de la proteína madura completa) para su análisis.

También es posible el análisis de secuencias genómicas para la identificación de homólogos de *GOS2/SUI1*. Están disponibles varios algoritmos y herramientas informáticas para la identificación de genes en secuencia de ADN sin tratamiento. Habitualmente estas herramientas combinan el análisis de parámetros estadísticos en la secuencia de ADN con procedimientos basados en homología para identificar secuencias homólogas en bases de datos. Aunque ninguno de estos procedimientos por sí solo es suficientemente fiable para una buena predicción, la combinación de diversos programas habitualmente proporciona buenos resultados. Los ejemplos bien conocidos de dichas herramientas incluyen GeneMark (http://opal.biology.gatech.edu/genemark, http://www.ebi.ac.uk/genemark, Borodovsky, M. y McIninch J. (1993) GeneMark: Parallel Gene Recognition for both DNA Strands. Computers & Chemistry, 17, 123-133), Gene Locator and Interpolated Markov Modeler (GLIMMER, http://www.tigr.org/softlab, A.L. Delcher y col. Improved microbial gene identification with GLIMMER. Nucleic Acids Research, 27, 4636-4641.(1999)), Gene Recognition and Assembly Internet Link (GRAIL, http://compbio.ornl.gov), GenScan (http://genes.mit.edu/GENSCAN.html, Burge, C. y Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78-94.), GeneBuilder (http://www.it-ba.mi.cnr.it/webgene Milanese L. y col. GeneBuilder: interactive in silico prediction of genes structure. Bioinformatics, 15 (7): 612-621, 1999). Puede realizarse una análisis combinado con el programa TIGR Combiner (http://www.tigr.org/softlab, J. E. Allen y col. Computational gene prediction using multiple sources of evidence. Genome Research, 14(1), 142-148, 2004) que predice modelos génicos usando el resultado de otro software de anotación tal como GeneMark, GlimmerM, GRAIL, GenScan y Fgenes. Usa un algoritmo estadístico para identificar patrones de pruebas correspondientes a modelos génicos.

45 Se conocen en la técnica procedimientos para caracterizar la 5'UTR de un gen. Por ejemplo, muchos puntos de inicio de la transcripción en eucariotas están precedidos por una caja TATA o consisten en un elemento iniciador reconocido por ARN polimerasa II. Los elementos iniciadores de origen natural en muchos genes tienen un resto de citosina (C) en la posición -1 y uno de adenina (A) en el sitio de inicio de la transcripción (+1). La secuencia consenso para un elemento iniciador consiste en 5'-Y Y A<sup>+</sup> N T/A Y Y Y-3', en la que A<sup>+</sup> es el inicio de la transcripción, Y es una pirimidina (C o T), N es cualquiera de las cuatro bases, y T/A es T o A en la posición +3. Otros genes contienen un tramo rico en CG típico de 20 a 50 nucleótidos dentro de una región de aproximadamente 100 pares de bases cadena arriba de la región del sitio de inicio, conocido como isla CpG. También se conocen en la técnica procedimientos para predecir la presencia de intrones. Los ejemplos incluyen SPL (http://www.softberry.com/berry.phtml, Softberry), NNSplice (http://www.fruit-fly.org/seq\_tools/splice.html, Reese MG y col. 1997. Improved Splice Site Detection in Genie. J Comp Biol 4(3), 311-23), SpliceView (http://www.itba.mi.cnr.it/webgene/, Rogozin I. B. y L. Milanese. Analysis of donor splice signals in different organisms. J. Mol. Evol., 1997, V.45, 50-59.), NetGene2 (http://www.cbs.dtu.dk/services, S. M. Hebsgaard, y col. Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information, Nucleic Acids Research, 1996, Vol. 24, 3439-3452.), GeneSplicer (http://www.tigr.org/softlab, Pertea M. y col. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res. 1 Mar 2001; 29(5): 1185-90).

En una realización preferida, la presente invención proporciona un procedimiento para aumentar la expresión de un transgén como se ha perfilado anteriormente, en el que dicha parte de la 5'UTR de GOS2 es como se representa

por la SEC ID N° 1. En una realización alternativa, se usa la 5'UTR de GOS2 completa como se representa por la SEC ID N° 2. En cualquier caso, los procedimientos de acuerdo con la presente invención siempre hacen uso de una unidad transcripcional quimérica que comprende el ácido nucleico de interés fusionado en su extremo 5' con el extremo 3' de la 5'UTR de un gen GOS2, o parte del mismo, en el que esa parte de la 5'UTR de GOS2 comprende al menos el primer intrón del gen GOS2 o una variante funcional de este primer intrón. Preferentemente, la parte de la 5'UTR de GOS2 es como se presenta por la SEC ID N° 1.

Como es generalmente aceptado, una cadena de ácido nucleico es direccional, teniendo el "extremo 5'" un hidroxilo (o fosfato) libre en un carbono 5' y teniendo el "extremo 3'" un hidroxilo (o fosfato) en un carbono 3'. En el caso de un ácido nucleico bicatenario, la orientación de la cadena con sentido establece la dirección. La cadena con sentido tiene, por definición, el mismo "sentido" que el ARNm; es decir, puede traducirse exactamente como puede la secuencia de ARNm.

En una siguiente etapa, la unidad transcripcional está unida operativamente con un promotor, formando de este modo un casete de expresión. Como se usa en el presente documento, un "promotor" significa una región de ADN cadena arriba del inicio de la transcripción y que está implicado en la unión de ARN polimerasa y otras proteínas para iniciar la transcripción. La referencia en el presente documento a un "promotor" debe interpretarse en su contexto más amplio e incluye las secuencias reguladoras de la transcripción derivadas de un gen genómico eucariota clásico, que incluye la caja TATA que se requiere para el inicio de la transcripción preciso, con o sin una secuencia de caja CCAAT y elementos reguladores adicionales (es decir secuencias activadoras cadena arriba, potenciadores y silenciadores) que alteran la expresión génica en respuesta a estímulos de desarrollo y/o externos, o de una manera específica de tejido. En consecuencia, la velocidad de transcripción de un promotor reprimible se reduce en respuesta a un agente represor. La velocidad de transcripción de un promotor inducible aumenta en respuesta a un agente inductor. La velocidad de promoción de un promotor constitutivo no está específicamente regulada, aunque puede variar según la influencia de las condiciones metabólicas generales. El término "promotor" también incluye las secuencias reguladoras de la transcripción de un gen procariota clásico, en cuyo caso puede incluir una secuencia de caja -35 y/o una secuencia reguladora de la transcripción de caja -10. El término "promotor" también se usa para describir una molécula sintética o de fusión, o derivado que confiere, activa o potencia la expresión de una molécula de ácido nucleico en una célula, tejido u órgano.

Ventajosamente, puede combinarse cualquier tipo de promotor con la unidad transcripcional y usarse para conducir la expresión del transgén, dependiendo del resultado deseado. Ventajosamente, el promotor es un promotor expresable en plantas. Por "promotor expresable en plantas" se entiende un promotor que es capaz de actuar en una célula vegetal. Por ejemplo, un promotor específico de meristemo, tal como el promotor rnr (ribonucleótido reductasa), cdc2a y el promotor cyc07, pueden usarse para efectuar la expresión en todas las partes crecientes de la planta, aumentando de este modo la proliferación celular, lo que a su vez aumentaría la producción o la biomasa. Si el resultado deseado fuera influir en las características de las semillas, tales como la capacidad de almacenamiento, tamaño de la semilla, número de semillas, biomasa, etc., entonces puede seleccionarse un promotor específico de semilla, tal como p2S2, pPROLAMIN, pOLEOSIN. Puede seleccionarse un promotor específico de aleurona para aumentar el crecimiento en el momento de la germinación, aumentando de este modo el transporte de azúcares al embrión. Puede utilizarse un promotor específico de inflorescencia, tal como pLEAFY, si el resultado deseado fuera modificar el número de órganos de flores. Para producir plantas estériles masculinas podría desearse usar un promotor específico de anteras. Para influir en la arquitectura floral, por ejemplo tamaño del pétalo, se puede elegir un promotor específico de pétalos. Si el resultado deseado fuera modificar el crecimiento y/o las características del desarrollo en órganos particulares, la elección del promotor para combinar dependería del órgano para modificar. Por ejemplo, el uso de un promotor específico de raíz conduciría a aumento del crecimiento y/o aumento de la biomasa o de la producción de la raíz y/o alteración fenotípica de la raíz. Esto sería particularmente importante cuando sea la raíz en sí misma lo que sea el producto final deseado; dichos cultivos incluyen remolacha azucarera, nabo, zanahoria y patata. Puede usarse un promotor específico de fruto para modificar, por ejemplo, la fuerza de la piel externa del fruto o para aumentar el tamaño del fruto. Puede usarse un promotor específico de tejido verde para aumentar el tamaño de la hoja. Puede usarse un promotor específico de pared celular para aumentar la rigidez de la pared celular, aumentando de este modo la resistencia a patógenos. Puede usarse un promotor específico de anteras para producir plantas estériles masculinas. Puede usarse un promotor específico vascular para aumentar el transporte de hojas a semillas. Puede usarse un promotor específico de nódulo para aumentar las capacidades fijadoras de nitrógeno de una planta, aumentando de este modo los niveles de nutrientes en una planta. Puede usarse también un promotor inducible por tensión para conducir la expresión de un ácido nucleico para aumentar la integridad de membrana durante condiciones de tensión. Puede usarse un promotor inducible por tensión tal como el promotor inducible por tensión de agua WSI18, el promotor Trg-31 inducible por tensión de sequía, el promotor relacionado con ABA rab21 o cualquier otro promotor que sea inducible en una condición de tensión particular tal como tensión por temperatura (frío, congelación, calor) o tensión osmótica, o puede usarse tensión de sequía o tensión oxidativa o tensión biótica para conducir la expresión de un transgén.

En una realización particular de la invención, el promotor usado para aumentar la expresión de un transgén es un promotor constitutivo. El término "constitutivo" como se define en el presente documento se refiere a un promotor expresado predominantemente en al menos un tejido u órgano y predominantemente en cualquier estadio de vida de la planta. Preferentemente, el promotor se expresa predominantemente por toda la planta. Preferentemente, el

promotor constitutivo es un promotor de *GOS2*, o un promotor de fuerza similar y/o un promotor con un patrón de expresión similar, más preferentemente el promotor es el promotor de *GOS2* del arroz, a condición de que el promotor de *GOS2* del arroz no se use en combinación con la 5'UTR completa del gen *GOS2* del arroz. En otra realización, se usa un promotor inducible. La velocidad de transcripción de un promotor inducible aumenta en respuesta a un agente o condición inductor, tal como estímulos químicos, ambientales o físicos. En otra realización más, se usa un promotor regulado por el desarrollo. La expresión "promotor regulado por el desarrollo" se refiere a un promotor cuya actividad se determina por acontecimientos del desarrollo. En una realización preferida, se usa un promotor con preferencia de órgano. Un promotor con preferencia de órgano es un promotor que se expresa predominantemente en un órgano específico de la planta. En una realización más preferida, se usa un promotor con preferencia de semillas. La expresión "promotor con preferencia de semillas" se refiere a un promotor que se expresa predominantemente en semillas de una planta. En otra realización preferida, se usa un promotor con preferencia de tejido. La expresión "preferencia de tejido" como se define en el presente documento se refiere a un promotor que se expresa predominantemente pero no necesariamente de forma exclusiva en un tejido u órgano, pero que también puede expresarse en una célula específica. Preferentemente, el promotor con preferencia de tejido es un promotor con preferencia de endospermo. Más preferentemente, el promotor con preferencia de endospermo es un promotor de RP6 de prolamina, más preferentemente el promotor de RP6 de prolamina del arroz.

Los inventores han mostrado también por primera vez que una 5'UTR que se origina de un gen *GOS2* de planta o una parte del mismo, y comprendiendo dicha parte el primer intrón, puede usarse para aumentar la expresión de un transgén en semillas de plantas.

Por lo tanto, se proporciona un procedimiento para aumentar el contenido proteico de semillas de plantas, comprendiendo dicho procedimiento

- a. integrar el primer intrón de la 5' UTR de un gen *GOS2* de planta en el extremo 5' de un ácido nucleico que codifica una proteína de interés, creando de este modo una unidad transcripcional quimérica, en la que dicho primer intrón se representa en la SEC ID N°: 5 o una variante funcional de dicho primer intrón, cuya variante funcional es de al menos 65 pares de base de longitud, comprende sitios de corte y empalme y una adenosina de punto de ramificación funcional, dicha adenosina de punto de ramificación es parte del motivo conservado con la siguiente secuencia consenso  $Y_{100}U_{100}R_{64}A_{100}U_{50}$  o  $Y_{100}U_{100}R_{64}A_{100}Y_{70}$ , en la que Y representa los nucleótidos C o T, y R representa A y G, y en la que los números en subíndice indican el porcentaje de conservación del nucleótido respectivo, cuya variante funcional es capaz de hibridar en condiciones rigurosas con el primer intrón como se representa en SEC ID N°: 5, en el que dichas condiciones de hibridación son hibridación en SCC 4 x a 65 °C, seguido de una etapa de lavado en SCC 0,1 x a 65 °C durante una hora;
- b. fusionar operativamente dicha unidad transcripcional quimérica con un promotor expresable en plantas para obtener un casete de expresión, a condición de no haya combinación de la 5' UTR como se representa en la SEC ID N°: 2 del gen *GOS2* del arroz con un promotor del gen *GOS2* del arroz;
- c. introducir y expresar en una célula vegetal el casete de expresión de (b), para crear una célula vegetal transgénica; y
- d. regenerar y/o cultivar una planta a partir de la célula vegetal transgénica de (c) de modo que dicha célula vegetal transgénica transcriba dicho ácido nucleico que codifica la proteína de interés y en la que la 5'UTR del ácido nucleico expresado que codifica dicha proteína comprende al menos el primer intrón de la 5'UTR de (a).

Se describe un procedimiento para aumentar el contenido proteico de semillas vegetales, comprendiendo dicho procedimiento

- a. integrar toda o parte de la 5' UTR de un gen *GOS2* de planta en el extremo 5' de un ácido nucleico que codifica la proteína de interés, creando de este modo una unidad transcripcional quimérica, en la que esta parte de una 5'UTR de *GOS2* comprende al menos el primer intrón del gen *GOS2* o una variante funcional de este primer intrón;
- b. fusionar de forma operativa la unidad transcripcional quimérica con un promotor expresable en plantas para obtener un casete de expresión, a condición de no haya combinación de una 5' UTR completa del gen *GOS2* del arroz con un promotor del gen *GOS2* del arroz;
- c. introducir y expresar en una célula vegetal el casete de expresión de (b), para crear una célula vegetal transgénica; y
- d. regenerar y/o cultivar una planta a partir de la célula vegetal transgénica de (c) de modo que la célula transgénica transcriba el ácido nucleico que codifica la proteína de interés y en la que la 5'UTR del ácido nucleico expresado que codifica la proteína de interés comprende al menos la 5'UTR de (a).

La expresión "unido operativamente" se refiere a una yuxtaposición entre una secuencia reguladora y una codificante, en la que los componentes descritos de este modo están en una relación que les permite actuar de su manera pretendida. Una secuencia reguladora "unida operativamente" con una secuencia codificante se liga de tal manera que se consiga expresión de la secuencia codificante en condiciones compatibles con las secuencias de control. En el caso de que la secuencia reguladora sea un promotor, se conocerá por un experto en la materia que es preferible un ácido nucleico bicatenario.

De acuerdo con una realización adicional, se proporciona un vector en el que el casete de expresión puede estar ligado. Con "vector" o "construcción genética" se entiende una secuencia de ADN, que puede introducirse en un organismo por transformación y puede mantenerse de forma transitoria o estable en este organismo. Opcionalmente, también pueden usarse una o más secuencias terminadoras en la construcción introducida en una planta. El término "terminador" abarca una secuencia de control que es una secuencia de ADN en el extremo de una unidad transcripcional que señala el procesamiento 3' y la poliadenilación de un transcrito primario y terminación de la transcripción. Los elementos reguladores adicionales pueden incluir potenciadores de la transcripción así como de la traducción. Los expertos en la materia serán conscientes de secuencias terminadoras y potenciadoras distintas del primer intrón de un gen GOS2, que pueden ser adecuadas para su uso en la realización de la invención. Dichas secuencias se conocerían o pueden obtenerse fácilmente por un experto habitual en la materia. El vector puede incluir además una secuencia de origen de replicación que se requiere para el mantenimiento y/o replicación en un tipo celular específico. Un ejemplo es cuando se requiere mantener un vector en una célula bacteriana como un elemento genético episómico (por ejemplo, molécula plasmídica o cosmídica). Los orígenes preferidos de replicación incluyen, pero sin limitación, el ori f1 y colE1. La construcción genética puede comprender opcionalmente un gen marcador seleccionable. Como se usa en el presente documento, la expresión "gen marcador seleccionable" incluye cualquier gen que confiera un fenotipo a una célula en la que se exprese para facilitar la identificación y/o selección de células que se transfectan o transforman con una construcción de ácido nucleico de la invención. Los marcadores adecuados pueden seleccionarse de marcadores que confieran resistencia a antibióticos o herbicidas, que introduzcan un nuevo rasgo metabólico o que permitan la selección visual. Los ejemplos de genes marcadores seleccionables incluyen genes que confieren resistencia a antibióticos (tales como *nptII* que fosforila la neomicina y kanamicina, o *hpt*, que fosforila la higromicina); genes que confieren resistencia a herbicidas (por ejemplo *bar* que proporciona resistencia a Basta; *aroA* o *gox* que proporciona resistencia contra glifosato); o genes que proporcionan un rasgo metabólico (tal como *manA* que permite a las plantas usar manosa como única fuente de carbono). Los genes marcadores visuales dan como resultado la formación de color (por ejemplo  $\beta$ -glucuronidasa, *GUS*), luminiscencia (tal como luciferasa) o fluorescencia (Proteína Verde Fluorescente, *GFP*, y derivados de la misma).

En una siguiente etapa, el casete de expresión, insertado o no en un vector, se introduce en una célula vegetal o en la planta en sí misma (incluyendo introducción en un tejido, órgano o cualquier otra parte de la planta). De acuerdo con una característica preferida de la presente invención, el ácido nucleico se introduce preferentemente en una planta por transformación. El término "transformación" como se indica en el presente documento abarca la transferencia de un polinucleótido exógeno en una célula huésped, independientemente del procedimiento usado para la transferencia. El tejido vegetal capaz de propagación clonal posterior, bien por organogénesis o bien por embriogénesis, puede transformarse con una construcción genética de la presente invención y regenerar una planta completa a partir del mismo.

El tejido particular seleccionado variará dependiendo de los sistemas de propagación clonales disponibles, y mejor adaptados, para la especie particular que se transforme. Las dianas tisulares ejemplares incluyen discos foliares, polen, embriones, cotiledones, hipocótilos, megagametofitos, tejido caloso, tejido meristemático existente (por ejemplo, meristemo apical, yemas axilares, y meristemas radiculares), y tejido del meristemo inducido (por ejemplo, meristemo del cotiledón y meristemo del hipocótilo). El polinucleótido puede introducirse de forma transitoria o estable en una célula huésped y puede mantenerse no integrado, por ejemplo, como un plásmido. Como alternativa, puede integrarse en el genoma huésped o en el genoma del cloroplasto. La célula vegetal transformada resultante puede usarse después para regenerar una planta transformada de una manera conocida por los expertos en la materia.

La transformación de una especie vegetal es ahora una técnica bastante rutinaria. Ventajosamente, puede usarse cualquiera de varios procedimientos de transformación para introducir el gen de interés en una célula ancestro adecuada. Los procedimientos de transformación incluyen el uso de liposomas, electroporación, productos químicos que aumentan la captación de ADN libre, inyección del ADN directamente en la planta, bombardeo con pistola de partículas, transformación usando virus o polen y microproyección. Los procedimientos pueden seleccionarse del procedimiento de calcio/poli-etilenglicol para protoplastos (Krens, F. A. y col., 1982, *Nature* 296, 72-74; Negrutiu I. y col., 1987, *Plant Mol. Biol.* 8, 363-373); electroporación de protoplastos (Shillito R. D. y col., 1985, *Bio/Technol.* 3, 1099-1102); microinyección en materia vegetal (Crossway A. y col., 1986, *Mol. Gen. Genet.* 202, 179-185); bombardeo de partículas recubiertas con ADN o ARN (Klein T. M. y col., 1987, *Nature* 327, 70) infección con virus (no integrantes) y similares. Un procedimiento preferido para la transformación del arroz de acuerdo con la presente invención es mediante transformación mediada por *Agrobacterium* usando cualquiera de los procedimientos bien conocidos para la transformación del arroz, tales como los descritos en cualquiera de los siguientes: solicitud de patente Europea publicada EP 1198985 A1, Aldemita y Hodges (*Planta*, 199, 612-617, 1996); Chan y col. (*Plant Mol. Biol.* 22 (3) 491-506, 1993), Hiei y col. (*Plant J.* 6 (2) 271-282, 1994). En el caso de la transformación del maíz, el procedimiento preferido es como se describe en Ishida y col. (*Nat. Biotechnol.* Jun 1996; 14(6): 745-50) o Frame y col. (*Plant Physiol.* Mayo 2002; 129(1): 13-22).

Generalmente después de la transformación, las células o agrupaciones de células vegetales se seleccionan con respecto a la presencia de uno o más marcadores que se codifican por genes expresables en plantas cotransferidos con el gen de interés, después de lo cual el material transformado se regenera en una planta completa.

Después de la transferencia y regeneración de ADN, pueden evaluarse plantas transformadas potencialmente, por ejemplo usando análisis de Southern o por PCR, con respecto a la presencia del gen de interés, número de copias y/u organización genómica. Como alternativa, o adicionalmente, los niveles de expresión del ADN recién introducido pueden controlarse usando análisis de Northern y/o Western, siendo ambas técnicas bien conocidas por personas que tienen experiencia habitual en la materia.

En una siguiente etapa de selección, las plantas transformadas se evalúan con respecto a la expresión deseada del transgén o con respecto a los fenotipos deseados. Se conocen por los expertos en la materia procedimientos para medir la expresión modulada (expresión aumentada o reducida). Los niveles de ARN alterados pueden medirse con técnicas como RT-PCR mientras que los cambios en los niveles proteicos pueden determinarse por ensayos enzimáticos, tinción de geles de proteína SDS-PAGE, transferencia de Western o ensayos inmunoabsorbentes ligados a enzima. Además, en el caso de las plantas transgénicas, pueden controlarse los cambios en el fenotipo de las plantas transgénicas. Por ejemplo, pueden medirse cambios en la producción o biomasa, cambios en la arquitectura vegetal o respuesta a tensión, o cambios en el contenido proteico o de aceite de las semillas. Se conoce por expertos en la materia que la expresión de transgenes en plantas, y por lo tanto también el efecto fenotípico debido a la expresión de dicho transgén, puede diferir entre diferentes líneas transgénicas obtenidas de forma independiente y la descendencia de la misma. Los transgenes presentes en diferentes plantas transgénicas obtenidas de forma independiente pueden diferir entre sí por el locus de inserción cromosómica así como por el número de copias del transgén insertadas, y por la configuración de las múltiples copias del transgén en un locus. Por lo tanto pueden esperarse diferencias en los niveles de expresión, que pueden deberse a la influencia del contexto cromosómico del transgén (el llamado efecto de posición) o de mecanismos silenciadores desencadenados por ciertas configuraciones del transgén (por ejemplo las inserciones en tándem orientadas hacia dentro de los transgenes tienden a silenciar al nivel transcripcional o postranscripcional).

Las plantas transformadas generadas pueden prepararse por una diversidad de medios, tales como por propagación clonal o técnicas de cultivo clásicas. Por ejemplo, una planta transformada de primera generación (o T1) puede auto sembrarse para proporcionar transformantes de segunda generación (o T2) homocigotos, y las plantas T2 pueden propagarse adicionalmente mediante técnicas de cultivo clásicas. Los organismos transformados generados pueden tomar una diversidad de formas. Por ejemplo, puede haber quimeras de células transformadas y células no transformadas; transformantes clonales (por ejemplo, todas las células transformadas para contener el casete de expresión); injertos de tejidos transformados y no transformados (por ejemplo, en plantas, un portainjerto transformado injertado en un injerto no transformado).

La presente invención claramente se extiende a cualquiera célula vegetal o planta producida por cualquiera de los procedimientos descritos en el presente documento, y a todas las partes de plantas y propágulos de las mismas. La presente invención se extiende adicionalmente para abarcar la descendencia de una célula primaria transformada o transfectada, tejido, órgano o planta completa que se haya producido por cualquiera de los procedimientos anteriormente mencionados, siendo el único requisito que la descendencia muestre la misma característica o las mismas características genotípicas y/o fenotípicas que las producidas en el parental por los procedimientos de acuerdo con la presente invención. La invención también se extiende a partes recolectables de una planta, tales como (pero sin limitación), semillas, hojas, frutos, flores, tallos o cultivos de tallos, rizomas, raíces, tubérculos y bulbos. La invención se refiere además a productos derivados directamente de una parte cosechable de dicha planta, incluyendo dichos productos microgránulos secos o polvos, aceite, grasa y ácidos grasos, almidón o proteínas.

El término "planta" como se usa en el presente documento abarca plantas completas, ancestros y descendencia de las plantas y partes de plantas, incluyendo semillas, frutos, flores, hojas, brotes, tallos, raíces (incluyendo tubérculos) y células, tejidos y órganos vegetales. El término "planta" abarca por lo tanto también cultivos en suspensión, embriones, regiones meristemáticas, tejidos callosos, gametofitos, esporofitos, polen y microesporas. Las plantas que son particularmente útiles en los procedimientos de la invención incluyen algas, helechos y todas las plantas que pertenecen a la superfamilia Viridiplantae, en particular plantas monocotiledóneas y dicotiledóneas, incluyendo legumbres de pienso o forraje, plantas ornamentales, cultivos alimentarios, árboles o arbustos seleccionados de la lista que comprende *Abelmoschus* spp., *Acer* spp., *Actinidia* spp., *Agropyron* spp., *Allium* spp., *Amaranthus* spp., *Ananas comosus*, *Annona* spp., *Apium graveolens*, *Arabidopsis thaliana*, *Arachis* spp., *Artocarpus* spp., *Asparagus officinalis*, *Avena sativa*, *Averrhoa carambola*, *Benincasa hispida*, *Bertholletia excelsea*, *Beta vulgaris*, *Brassica* spp., *Cadaba farinosa*, *Camellia sinensis*, *Canna indica*, *Capsicum* spp., *Carica papaya*, *Carissa macrocarpa*, *Carthamus tinctorius*, *Carya* spp., *Castanea* spp., *Cichorium endivia*, *Cinnamomum* spp., *Citrullus lanatus*, *Citrus* spp., *Cocos* spp., *Coffea* spp., *Cola* spp., *Colocasia esculenta*, *Corylus* spp., *Crataegus* spp., *Cucumis* spp., *Cucurbita* spp., *Cynara* spp., *Daucus carota*, *Desmodium* spp., *Dimocarpus longan*, *Dioscorea* spp., *Diospyros* spp., *Echinochloa* spp., *Eleusine coracana*, *Eriobotrya japonica*, *Eugenia uniflora*, *Fagopyrum* spp., *Fagus* spp., *Ficus carica*, *Fortunella* spp., *Fragaria* spp., *Ginkgo biloba*, *Glycine* spp., *Gossypium hirsutum*, *Helianthus* spp., *Hibiscus* spp., *Hordeum* spp., *Ipomoea batatas*, *Juglans* spp., *Lactuca sativa*, *Lathyrus* spp., *Lemna* spp., *Lens culinaris*, *Linum usitatissimum*, *Litchi chinensis*, *Lotus* spp., *Luffa acutangula*, *Lupinus* spp., *Macrotyloma* spp., *Malpighia emarginata*, *Malus* spp., *Mammea americana*, *Mangifera indica*, *Manihot* spp., *Manilkara zapota*, *Medicago sativa*, *Melilotus* spp., *Mentha* spp., *Momordica* spp., *Morus nigra*, *Musa* spp., *Nicotiana* spp., *Olea* spp., *Opuntia* spp., *Omithopus* spp., *Oryza* spp., *Panicum miliaceum*, *Passiflora edulis*, *Pastinaca sativa*, *Persea* spp., *Petroselinum crispum*, *Phaseolus* spp., *Phoenix* spp., *Physalis* spp., *Pinus* spp., *Pistacia vera*, *Pisum* spp., *Poa* spp., *Populus* spp., *Prosopis* spp., *Prunus*

spp., *Psidium* spp., *Punica granatum*, *Pyrus communis*, *Quercus* spp., *Raphanus sativus*, *Rheum rhabarbarum*, *Ribes* spp., *Rubus* spp., *Saccharum* spp., *Sambucus* spp., *Secale cereale*, *Sesamum* spp., *Solanum* spp., *Sorghum bicolor*, *Spinacia* spp., *Syzygium* spp., *Tamarindus indica*, *Theobroma cacao*, *Trifolium* spp., *Triticosecale rimpaii*, *Triticum* spp., *Vaccinium* spp., *Vicia* spp., *Vigna* spp., *Vitis* spp., *Zea mays*, *Zizania palustris*, *Ziziphus* spp., entre otros.

De acuerdo con una característica preferida de la presente invención, la planta es una planta de cultivo que comprende soja, girasol, canola, alfalfa, colza o algodón. Más preferentemente, la planta de acuerdo con la presente invención es una planta monocotiledónea, incluyendo miembros de las Poáceas, tales como caña de azúcar; más preferentemente un cereal, tal como arroz, maíz, trigo, mijo, cebada, avena, sorgo.

- 10 La presente invención también describe el uso de la 5'UTR de un gen *GOS2* de planta o parte del mismo que comprende el primer intrón. Se prevé cualquier uso de la 5'UTR de un gen *GOS2* de planta o parte del mismo que comprende el primer intrón, para aumentar la expresión del transgén, a condición de no haya combinación de una 5'UTR completa del gen *GOS2* del arroz con un promotor del gen *GOS2* del arroz. En particular, la 5'UTR de un gen *GOS2* de planta o parte del mismo que comprende el primer intrón puede usarse para modificar las características de crecimiento de una planta, tales como producción, cantidad de biomasa producida, arquitectura, tolerancia a la tensión. Otros ejemplos de uso de la 5'UTR de un gen *GOS2* de planta o parte del mismo que comprende el primer intrón incluyen la modificación de la composición de una planta o partes cosechables de la misma, tales como semillas. La invención se refiere al uso del primer intrón que se representa en la SEC ID N°: 5 o una variante funcional de dicho primer intrón, cuya variante funcional es de al menos 65 pares de bases de longitud, comprende sitios de corte y empalme y una adenosina de punto de ramificación funcional, siendo dicha adenosina de punto de ramificación parte del motivo conservado con la siguiente secuencia consenso  $Y_{100}U_{100}R_{64}A_{100}U_{50}$  o  $Y_{100}U_{100}R_{64}A_{100}Y_{70}$ , en la que Y representa los nucleótidos C o T, y R representa A y G, y en la que los números en subíndice indican el porcentaje de conservación del nucleótido respectivo, cuya variante funcional es capaz de hibridar en condiciones rigurosas con el primer intrón como se representa en la SEC ID N°: 5, en la que dichas condiciones de hibridación son hibridación en SCC 4 x a 65 °C, seguido de una etapa de lavado en SCC 0,1 x a 65 °C durante una hora en un procedimiento para modificar la composición de semillas de una planta, a condición de no haya combinación de las 5'UTR como se representa en la SEC ID N°: 2 del gen *GOS2* del arroz con un promotor del gen *GOS2* del arroz. Dicho uso puede dar como resultado por ejemplo semillas con contenido proteico aumentado. La invención se refiere al uso del primer intrón que se representa en la SEC ID N°: 5 o una variante funcional de dicho primer intrón, cuya variante funcional es de al menos 65 pares de bases de longitud, comprende sitios de corte y empalme y una adenosina de punto de ramificación funcional, siendo dicha adenosina de punto de ramificación parte del motivo conservado con la siguiente secuencia consenso  $Y_{100}U_{100}R_{64}A_{100}U_{50}$  o  $Y_{100}U_{100}R_{64}A_{100}Y_{70}$ , en la que Y representa los nucleótidos C o T, y R representa A y G, y en la que los números en subíndice indican el porcentaje de conservación del nucleótido respectivo, cuya variante funcional es capaz de hibridar en condiciones rigurosas con el primer intrón como se representa en la SEC ID N°: 5, en la que dichas condiciones de hibridación son hibridación en SCC 4 x a 65 °C, seguido de una etapa de lavado en SCC 0,1 x a 65 °C durante una hora en un procedimiento para aumentar el contenido proteico de semillas de plantas, a condición de no haya combinación de las 5'UTR como se representa en la SEC ID N°: 2 del gen *GOS2* del arroz con un promotor del gen *GOS2* del arroz. El contenido proteico aumentado de las semillas es ventajoso para aumentar el valor nutricional de estas semillas o es ventajoso además en el campo de la agricultura molecular. Las plantas con características de crecimiento mejoradas también pueden usarse para producir proteínas industriales y/u otros compuestos. La invención se refiere al uso del primer intrón que se representa en la SEC ID N°: 5 o una variante funcional de dicho primer intrón, cuya variante funcional es de al menos 65 pares de bases de longitud, comprende sitios de corte y empalme y una adenosina de punto de ramificación funcional, siendo dicha adenosina de punto de ramificación parte del motivo conservado con la siguiente secuencia consenso  $Y_{100}U_{100}R_{64}A_{100}U_{50}$  o  $Y_{100}U_{100}R_{64}A_{100}Y_{70}$ , en la que Y representa los nucleótidos C o T, y R representa A y G, y en la que los números en subíndice indican el porcentaje de conservación del nucleótido respectivo, cuya variante funcional es capaz de hibridar en condiciones rigurosas con el primer intrón como se representa en la SEC ID N°: 5, en la que dichas condiciones de hibridación son hibridación en SCC 4 x a 65 °C, seguido de una etapa de lavado en SCC 0,1 x a 65 °C durante una hora en un procedimiento para
- 50 (i) aumentar la expresión de un ácido nucleico de interés fusionado con el mismo;  
(ii) modificar las características de crecimiento de una planta;  
(iii) aumentar la producción de una planta;  
(iv) aumentar la tolerancia a tensión de una planta; o  
(v) alterar la arquitectura de una planta,
- 55 a condición de no haya combinación de la 5'UTR como se representa en SEC ID N° 2 del gen *GOS2* del arroz con el promotor del gen *GOS2* del arroz. En este caso, el objetivo sería asegurar la alta acumulación de los productos deseados en particular y tejidos vegetales fáciles de cosechar.

La presente invención también describe una composición que comprende la 5'UTR de un gen *GOS2* de planta, o parte del mismo, para aumentar la expresión de un transgén en plantas transgénicas a condición de (i) dicha parte de la 5'UTR de *GOS2* comprenda al menos el primer intrón de dicho gen *GOS2* o una variante funcional de dicho primer intrón y (ii) el intrón de *GOS2* del arroz no se usa en combinación con el promotor de *GOS2* del arroz.

La presente invención se describirá ahora con referencia a las siguientes figuras en las que:

**Figura 1:** representación esquemática del gen *GOS2* del arroz. La flecha inferior muestra la estructura del ARNm, las partes blancas representan los intrones, mientras que la flecha superior representa la secuencia codificante correspondiente (partes rellenas).

5 **Figura 2:** representación esquemática del ADN T del plásmido p5024, que comprende comenzando desde el extremo izquierdo (LB) un casete de expresión *nptII* para resistencia a kanamicina (SM k7) dentro de 2 sitios de recombinación (RS), el promotor del gen *RP6* de prolamina del arroz, el líder épsilon del virus del moteado del dátilo (CfMV) (potenciador, GenBank Z48630), el casete Gateway A (GW), que contiene un gen de resistencia a cloranfenicol y un gen suicida *ccdB* flanqueado por el sitio de recombinación de Gateway *attR1* y *attR2*, una secuencia terminadora doble (*T-zein* y *T-rbcS*) flanqueada por la secuencia del extremo derecho (RB) del ADN T.

10 **Figura 3:** representación esquemática del ADN T del plásmido p5942, derivado de p5024 insertando el primer intrón de *GOS2* del arroz en el sitio de restricción *SpeI*.

**Figura 4:** representación esquemática del ADN T de los plásmidos p06192 y p06193, usado para transformar plantas de arroz. El casete GW de p5942 y p5024 se reemplaza por un gen indicador.

15 **Figura 5:** secuencias de ADN útiles en la presente invención. SEC ID N° 1 representa la parte de la 5'UTR de *GOS2* del arroz que se usa en los ejemplos. La secuencia intrónica en sí misma (SEC ID N°: 5) se indica en negrita. SEC ID N° 2 representa la 5'UTR completa del gen *GOS2* del arroz definido por de Pater y col. (1992), extraído de la referencia de GenBank X51910. La transcripción puede comenzar en cualquiera de los tres primeros nucleótidos. SEC ID N°: 3 representa la secuencia codificante del factor de inicio de la traducción *GOS2* y SEC ID N°: 4 es la secuencia de proteína *GOS2* codificada por SEC ID N°: 3.

20

#### Ejemplos:

La presente invención se describirá ahora con referencia a los siguientes ejemplos, que son solamente para ilustración.

25 Manipulación de ADN: a no ser que se indique de otro modo, se realizan técnicas de ADN recombinante de acuerdo con protocolos convencionales descritos en (Sambrook (2001) Molecular Cloning: a laboratory manual, 3ª Edición Cold Spring Harbor Laboratory Press, CSH, Nueva York) o en los Volúmenes 1 y 2 de Ausubel y col. (1994), Current Protocols in Molecular Biology, Current Protocols. Se describen materiales y procedimientos convencionales para el trabajo molecular de plantas en Plant Molecular Biology Labfax (1993) por R.D.D. Croy, publicado por BIOS Scientific Publications Ltd (Reino Unido) y Blackwell Scientific Publications (Reino Unido).

#### 30 **Ejemplo 1: construcciones genéticas**

Se creó un ADN-T modificado (p5024) comprendiendo dentro de sus límites los elementos como se muestran en la Figura 2. A continuación se insertó el primer intrón del gen *GOS2* del arroz en el sitio de restricción *SpeI* (Figura 3). Finalmente, el casete GW se reemplazó por el gen *GUS* usando una reacción LR GW, dando como resultado los plásmidos p06192 y p06193 (Figura 4).

#### 35 **Ejemplo 2: transformación del arroz:**

Se descascarillaron semillas secas maduras de *Oryza sativa* japónica cultivar Nipponbare. Se realizó esterilización incubando las semillas durante un minuto en etanol al 70 %, seguido de 30 minutos en HgCl<sub>2</sub> 0,2 % y mediante 6 lavados de 15 minutos con agua destilada estéril. Las semillas estériles se germinaron después en un medio que contenía 2,4-D (medio de inducción de callo). Después de una incubación de 4 semanas en oscuridad, se escindieron callos embriogénicos, derivados de escutelo y se propagaron en el mismo medio. Dos semanas después, los callos se multiplicaron o se propagaron por subcultivo en el mismo medio durante otras 2 semanas. Tres días antes del cocultivo, se subcultivaron trozos de callos embriogénicos en medio nuevo para potenciar la actividad de división celular. La cepa de *Agrobacterium* LBA4404 que alberga el vector binario p06192 o p06193 se usó para cocultivo. La cepa de *Agrobacterium* se cultivó durante 3 días a 28 °C en medio AB con los antibióticos apropiados. Las bacterias se recogieron después y se suspendieron en medio de cocultivo líquido a una DO<sub>600</sub> de aproximadamente 1. La suspensión se transfirió a una placa de petri y los callos se sumergieron en la suspensión durante 15 minutos. Los tejidos de los callos se secaron después por transferencia en papel de filtro, se transfirieron a medio de cocultivo solidificado y se incubaron durante 3 días en oscuridad a 25 °C. A continuación, el callo cocultivado se cultivó en medio que contenía 2,4-D durante 4 semanas en oscuridad a 28 °C en presencia de un agente selectivo a una concentración adecuada. Durante este periodo, se desarrollaron islas de callos resistentes de crecimiento rápido. Tras la transferencia de este material a un medio de regeneración, e incubación a la luz, se liberó el potencial embriogénico y se desarrollaron brotes en las siguientes cuatro a cinco semanas. Se escindieron brotes del callo y se incubaron durante 2 a 3 semanas en un medio que contenía auxina del que se transfirieron al suelo. Se cultivaron brotes endurecidos con alta humedad y días cortos en un invernadero. Finalmente se cosecharon semillas de tres a cinco semanas después del trasplante. El procedimiento produjo transformantes de un único locus a una velocidad de más del 50 % (Aldemita y Hodges, Planta 199; 612-617, 1996; Chan y col., Plant Mol. Biol. 22(3),

55

491-506, 1993; Hiei y col., Plant J. 6(2), 271-282, 1994).

**Ejemplo 3: evaluación de las plantas transformadas:**

5 Se generaron aproximadamente de 15 a 20 transformantes T0 independientes. Los transformantes primarios se transfirieron de cámaras de cultivo tisular a un invernadero para cultivar y cosechar semilla T1. Se conservaron 10 y 15 acontecimientos (respectivamente sin y con intrón). Para cada acontecimiento, se molieron 10 semillas T1 (una mezcla 1:2:1 de nulicigotos, hemიცigotos y homocigotos) hasta un polvo fino y se ensayaron con respecto a actividad GUS (modificada de Breyne P, y col. (1993), Plant Mol. Biol. Reporter 11, 21-31) en mediciones por triplicado. Los resultados se proporcionan en la Tabla 1 a continuación:

**Tabla 1:** medición de las actividades de GUS en plantas transformadas con construcción con o sin intrón de GOS2:

Promotor + Potenciador + GUS					Promotor + intrón de GOS2 + Potenciador + GUS				
Línea	1	2	3	Media	Línea	1	2	3	Media
OS1861-001A	14,3	28,7	17,3	20	OS1866-001C	67,1	109,5	80,4	86
OS1861-001B	8,8	11,1	11,4	10	OS1866-002C	40,6	153,2	259,8	151
OS1861-002A	46,8	54,1	14,8	39	OS1866-003C	158,5	129,6	200,1	163
OS1861-004A	28,4	14,7	24,7	23	OS1866-005A	173,8	302,8	61,9	180
OS1861-005C	7,5	17,3	29,8	18	OS1866-005B	162,1	217,1	65,6	148
OS1861-006A	13,4	53,2	14,5	27	OS1866-007A	27,4	221,2	281,7	177
OS1861-006B	4,1	15,5	26,3	15	OS1866-007C	72,4	130,2	183,8	129
OS1861-007B	2,5	106,4	ND	54	OS1866-008A	196,6	186,2	117	167
OS1861-009A	18,3	118,7	92,4	76	OS1866-008C	161,5	336,6	61,2	186
OS1861-009C	13,5	11,7	1,9	9	OS1866-011A	162,0	238,1	403,4	268
			<b>media</b>	<b>29</b>	OS1866-012A	118,7	115,4	449,8	228
			desv est	22	OS1866-013A	33,6	55,4	ND	45
					OS1866-014A	17,7	99,8	79	66
					OS1866-014B	93,7	81,2	105,2	93
					OS1866-015A	159,4	90,0	10,8	87
								<b>media</b>	<b>149</b>
								desv est	62

10

A partir de los datos se puede deducir que la presencia del intrón de GOS2 da como resultado un aumento mayor de 5 veces en la expresión del gen indicador.

LISTADO DE SECUENCIAS

15 <110> CropDesign N.V.

<120> Procedimiento para aumentar la expresión de transgén

20 <130> CD-114-EP

<150> EP 04102108.0

<151> 13-05-2004

<150> US 60/572.141

<151> 18-05-2004

<160> 5

5 <170> PatentIn versión 3.3

<210> 1

<211> 1112

<212> ADN

10 <213> *Oryza sativa*

<400> 1

ctagtagaag	ccgagcgacc	gccttctctcg	atccatatct	tccggctcgag	ttcttggctcg	60
atctcttccc	tcctccacct	cctcctcaca	gggtatgtgc	ccttcgggtg	ttcttggatt	120
tattgttcta	ggttgtgtag	tacgggctgt	gatgttagga	aaggggatct	gtatctgtga	180
tgattcctgt	tcttggattt	gggatagagg	ggttcttgat	gttgcatggt	atcggttcgg	240
tttgattagt	agtatggttt	tcaatcgtct	ggagagctct	atggaaatga	aatggtttag	300
ggtacggaat	cttgcgattt	tgtgagtacc	ttttgtttga	ggtaaaatca	gagcaccggg	360
gattttgctt	gggtgaataa	aagtacgggt	gtttggtcct	cgattctggt	agtgatgctt	420
ctcgatttga	cgaagctatc	ctttgtttat	tcctattga	acaaaaataa	tccaactttg	480
aagacgggcc	cgttgatgag	attgaatgat	tgattcttaa	gcctgtccaa	aatttcgcag	540
ctggcttggt	tagatacagt	agtcctccatc	acgaaattca	tggaaacagt	tataatcctc	600
aggaacaggg	gattccctgt	tcttccgatt	tgctttagtc	ccagaatttt	tttcccaaa	660
tatcttaaaa	agtcactttc	tggttcagtt	caatgaattg	attgctacaa	ataatgcttt	720
tatagcgta	tcctagctgt	agttcagtta	ataggtaata	cccctatagt	ttagttagga	780
gaagaactta	tccgatttct	gatctccatt	tttaattata	tgaatgaac	tgtagcataa	840
gcagtattca	tttggattat	tttttttatt	agctctcacc	ccttcattat	tctgagctga	900
aagtctggca	tgaactgtcc	tcaattttgt	tttcaaattc	acatcgatta	tctatgcatt	960
atcctcttgt	atctacctgt	agaagtttct	ttttggttat	tccttgactg	cttgattaca	1020
gaaagaaatt	tatgaagctg	taatcgggat	agttatactg	cttgttctta	tgattcattt	1080
cctttgtgca	gttcttgggtg	tagcttgeca	ct			1112

15

<210> 2

<211> 1176

<212> ADN

20 <213> *Oryza sativa*

<400> 2

taggaggcat	ccaagccaag	aagagggaga	gcaccaagga	cacgcgacta	gcagaagccg	60
agcgaccgcc	ttcttcgatc	catatcttcc	ggctcgagttc	ttggctcgatc	tcttccctcc	120
tccacctect	cctcacaggg	tatgtgccct	tcggttggtc	ttggatttat	tgttctaggt	180
tgtgtagtag	gggcttgat	gtaggaaag	gggatctgta	tctgtgatga	ttcctgttct	240
tggatttggg	atagaggggt	tcttgatggt	gcattgtatc	ggttcgggtt	gattagtagt	300
atggttttca	atcgtctgga	gagctctatg	gaaatgaaat	ggtttagggg	acggaatcct	360
gcgattttgt	gagtaccttt	tgtttgaggt	aaaatcagag	caccgggtgat	tttgcctggg	420
gtaataaaaag	tacgggtggt	tggtcctcga	ttctggtagt	gatgcttctc	gatttgacga	480
agctatecct	tgtttattcc	ctattgaaca	aaaataatcc	aactttgaag	acgggtcccgt	540

tgatgagatt	gaatgattga	ttcttaagcc	tgtccaaaat	ttcgcagctg	gcttgttttag	600
atacagtagt	ccccatcacg	aaattcatgg	aaacagttat	aatcctcagg	aacaggggat	660
tccctgttct	tccgatttgc	tttagtccca	gaattttttt	tcccaaatat	cttaaaaagt	720
cactttctgg	ttcagttcaa	tgaattgatt	gctacaaata	atgcttttat	agcgttatcc	780
tagctgtagt	tcagttaata	ggtaataccc	ctatagttta	gtcaggagaa	gaacttatcc	840
gatttctgat	ctccattttt	aattatatga	aatgaactgt	agcataagca	gtattcattt	900
ggattatttt	ttttatttagc	tctcaccctc	tcattattct	gagctgaaaag	tctggcatga	960
actgtcctca	attttgtttt	caaattcaca	togattatct	atcgattatc	ctcttgtatc	1020
tacctgtaga	agtttctttt	tggttattcc	ttgactgctt	gattacagaa	agaaatttat	1080
gaagctgtaa	tccggatagt	tatactgctt	gttcttatga	ttcatttctc	ttgtgcagtt	1140
cttgggtgtag	cttgcacact	tcaccagcaa	agtttc			1176

<210> 3

<211> 348

25

ES 2 463 474 T3

<212> ADN  
<213> *Oryza sativa*

<400> 3

5

```

atgtctgatac tcgacattca gatcccaact gccttcgatc cctttgctga ggccaatgct      60
ggagactctg gtgcggctgc aggatcaaag gactacgttc atgtacgcat ccagcagcgt      120
aatggccgta agagcctgac cactgtccag ggattgaaga aggaattcag ctacaacaag      180
atccttaaag atctcaagaa agagttttgc tgcaatggta ctgttgtcca ggaccagag      240
cttggccagg tcattcaact tcaggggatg cagaggaaga acgtatcaa ttttctgtt      300
caggccggca ttgtgaagaa ggaacacatc aagattcatg gtttctga      348
    
```

<210> 4  
<211> 115  
<212> PRT  
<213> *Oryza sativa*

10

<400> 4

```

Met Ser Asp Leu Asp Ile Gln Ile Pro Thr Ala Phe Asp Pro Phe Ala
1           5           10           15

Glu Ala Asn Ala Gly Asp Ser Gly Ala Ala Ala Gly Ser Lys Asp Tyr
          20           25           30

Val His Val Arg Ile Gln Gln Arg Asn Gly Arg Lys Ser Leu Thr Thr
          35           40           45

Val Gln Gly Leu Lys Lys Glu Phe Ser Tyr Asn Lys Ile Leu Lys Asp
          50           55           60

Leu Lys Lys Glu Phe Cys Cys Asn Gly Thr Val Val Gln Asp Pro Glu
65           70           75           80

Leu Gly Gln Val Ile Gln Leu Gln Gly Asp Gln Arg Lys Asn Val Ser
          85           90           95

Asn Phe Leu Val Gln Ala Gly Ile Val Lys Lys Glu His Ile Lys Ile
          100          105          110

His Gly Phe
          115
    
```

15

<210> 5  
<211> 999  
<212> ADN  
<213> *Oryza sativa*

20

<400> 5

ES 2 463 474 T3

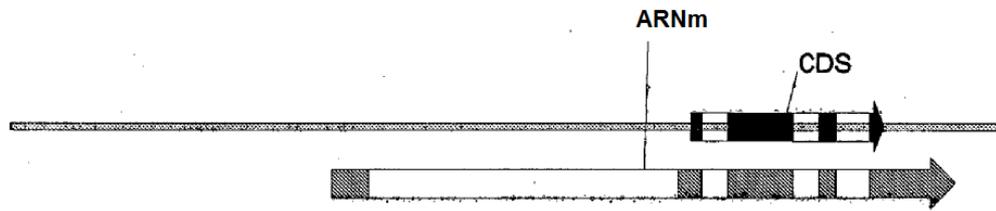
gtatgtgcc	ttcggttggt	cttggattta	ttgttctagg	ttgtgtagta	cgggcggtga	60
tgtaggaaa	gggatctgt	atctgtgatg	attcctgttc	ttggatttgg	gatagagggg	120
ttcttgatgt	tgcatgttat	cggttcgggt	tgattagtag	tatggttttc	aatcgtctgg	180
agagctctat	ggaaatgaaa	tggtttaggg	tacggaatct	tcgattttg	tgagtacctt	240
ttgtttgagg	taaaatcaga	gcaccgggtga	ttttgcttgg	tgtaataaaa	gtaccggttg	300
ttggtcctcg	attctggtag	tgatgcttct	cgatttgacg	aagctatcct	ttgtttattc	360
cctattgaac	aaaaataatc	caactttgaa	gacggtcccg	ttgatgagat	tgaatgattg	420
attcttaagc	ctgtccaaaa	tttcgcagct	ggcttgttta	gatacagtag	tccccatcac	480
gaaatccatg	gaaacagtta	taatcctcag	gaacagggga	ttccctgttc	ttccgatttg	540
ctttagtccc	agaatTTTTT	ttcccaaata	tcttaaaaag	tcactttctg	gttcagttca	600
atgaattgat	tgctacaaat	aatgctttta	tagcgttatc	ctagctgtag	ttcagttaat	660
aggtaatacc	cctatagttt	agtcaggaga	agaacttatc	cgatttctga	totocatttt	720
taattatatg	aaatgaactg	tagcataagc	agtattcatt	tggattattt	tttttattag	780
ctctcacccc	ttcattattc	tgagctgaaa	gtctggcatg	aactgtcctc	aattttgttt	840
tcaaattcac	atcgattatc	tatgcattat	cctcttgat	ctacctgtag	aagtttcttt	900
ttggttattc	cttgactgct	tgattacaga	aagaaattta	tgaagctgta	atcgggatag	960
ttatactgct	tgttcttatg	attcatttcc	tttgtgcag			999

## REIVINDICACIONES

1. Procedimiento para aumentar la expresión de transgenes en una planta transgénica, comprendiendo dicho procedimiento
  - 5 a) integrar el primer intrón de la 5' UTR de un gen *GOS2* de planta en el extremo 5' de un ácido nucleico de interés, creando de este modo una unidad transcripcional quimérica, en la que dicho primer intrón se representa en la SEC ID N°: 5 o una variante funcional de dicho primer intrón, cuya variante funcional es de al menos 65 pares de bases de longitud, comprende sitios de corte y empalme y una adenosina de punto de ramificación funcional, siendo dicha adenosina de punto de ramificación parte del motivo conservado con la siguiente secuencia consenso  $Y_{100}U_{100}R_{64}A_{100}U_{50}$  o  $Y_{100}U_{100}R_{64}A_{100}Y_{70}$ , en la que Y representa los nucleótidos C o T, y R representa A y G, y en la que los números en subíndice indican el porcentaje de conservación del nucleótido respectivo, cuya variante funcional es capaz de hibridar en condiciones rigurosas con el primer intrón como se representa en SEC ID N°: 5, en la que dichas condiciones de hibridación son hibridación en SCC 4 x a 65 °C, seguido de una etapa de lavado en SCC 0,1 x a 65 °C durante una hora;
  - 10 b) fusionar operativamente dicha unidad transcripcional quimérica con un promotor expresable en plantas para obtener un casete de expresión, a condición de no haya combinación de la 5' UTR como se representa en SEC ID N°: 2 del gen *GOS2* del arroz con un promotor del gen *GOS2* del arroz;
  - 15 c) introducir y expresar en una célula vegetal dicho casete de expresión de (b), para crear una célula vegetal transgénica; y
  - 20 d) regenerar y/o cultivar una planta a partir de dicha célula vegetal transgénica de (c) de modo que dicha célula transgénica transcriba el transgén y en la que la 5'UTR del ácido nucleico expresado de interés comprende al menos el primer intrón de la 5'UTR como se define en (a).
2. Procedimiento de la reivindicación 1, en el que dicho transgén codifica un polipéptido.
3. Procedimiento de acuerdo con las reivindicaciones 1 y 2, en el que dicha planta transgénica es una planta monocotiledónea.
- 25 4. Procedimiento de la reivindicación 3, en el que dicha planta monocotiledónea es un cereal, tal como arroz o maíz.
5. Procedimiento de una cualquiera de las reivindicaciones 1 a 4, en el que dicho gen *GOS2* de planta se origina de una planta monocotiledónea, preferentemente de arroz o maíz.
6. Procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que dicho promotor expresable en plantas se selecciona del grupo que consiste en:
  - 30 (i) un promotor constitutivo;
  - (ii) un promotor inducible;
  - (iii) un promotor con preferencia de tejido;
  - (iv) un promotor con preferencia de órgano;
  - (v) un promotor regulado por el desarrollo.
- 35 7. Procedimiento de la reivindicación 6, en el que dicho promotor con preferencia de órgano es un promotor con preferencia de semilla.
8. Procedimiento de la reivindicación 6, en el que dicho promotor con preferencia de tejido es un promotor con preferencia de endospermo.
9. Procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 8, en el que dicha unidad transcripcional quimérica comprende un ácido nucleico de interés fusionado en su extremo 5' con el extremo 3' del primer intrón de la 5'UTR de un gen *GOS2* de planta.
- 40 10. Uso del primer intrón como se representa en la SEC ID N°: 5 o una variante funcional de dicho primer intrón, cuya variante funcional es de al menos 65 pares de bases de longitud, comprende sitios de corte y empalme y una adenosina de punto de ramificación funcional, siendo dicha adenosina de punto de ramificación parte del motivo conservado con la siguiente secuencia consenso  $Y_{100}U_{100}R_{64}A_{100}U_{50}$  o  $Y_{100}U_{100}R_{64}A_{100}Y_{70}$ , en la que Y representa los nucleótidos C o T, y R representa A y G, y en la que los números en subíndice indican el porcentaje de conservación del nucleótido respectivo, cuya variante funcional es capaz de hibridar en condiciones rigurosas con el primer intrón como se representa en la SEC ID N°: 5, en la que dichas condiciones de hibridación son hibridación en SCC 4 x a 65 °C, seguido de una etapa de lavado en SCC 0,1 x a 65 °C durante una hora en un procedimiento para
  - 50 (i) aumentar la expresión de un ácido nucleico de interés fusionado con el mismo;
  - (ii) modificar las características de crecimiento de una planta;
  - (iii) aumentar la producción de una planta;
  - (iv) aumentar la tolerancia a tensión de una planta;
  - (v) alterar la arquitectura de una planta;
  - 55 (vi) modificar la composición de semillas de una planta; o

(vii) aumentar el contenido proteico de las semillas vegetales.

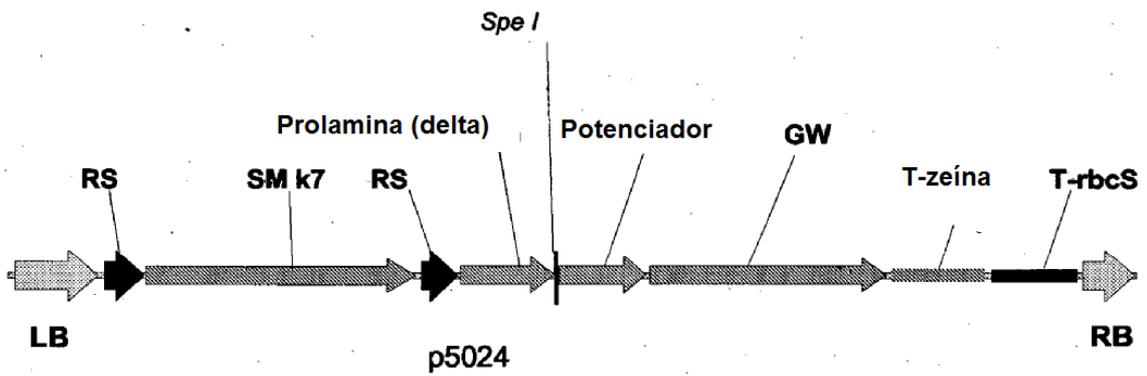
a condición de no haya combinación de la 5'UTR como se representa en SEC ID N° 2 del gen *GOS2* del arroz con el promotor del gen *GOS2* del arroz.



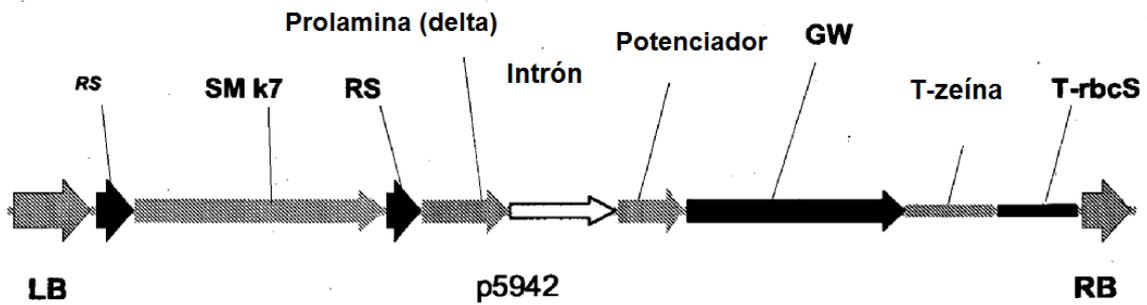
**gen GOS2**

3192 pb

**FIGURA 1**

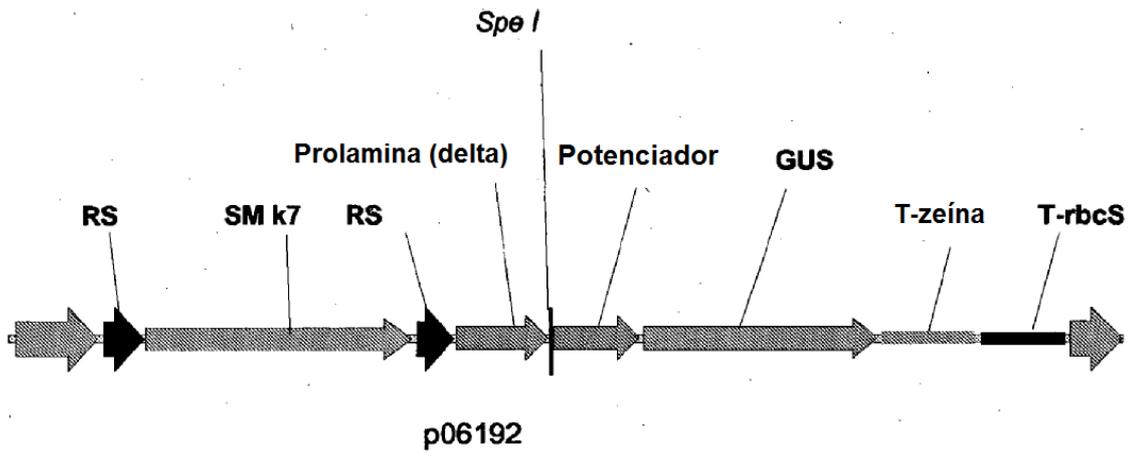


**FIGURA 2**

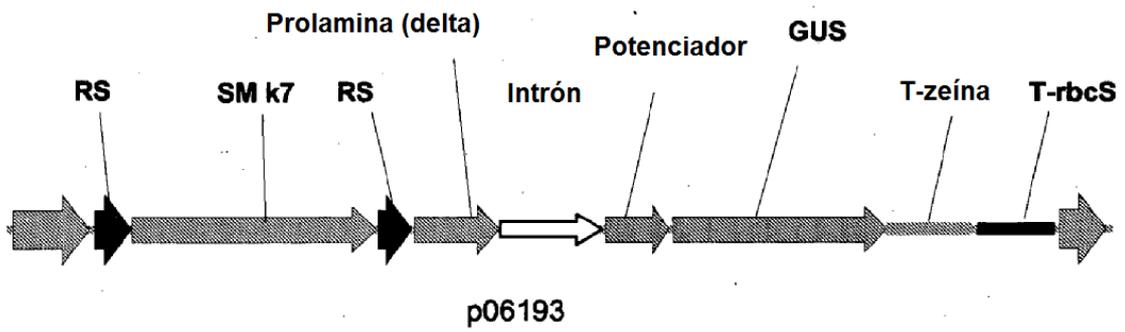


**FIGURA 3**

**A**



**B**



**FIGURA 4**

SEC ID N° 1:

ctagtagaagccgagcgcaccgccttcttcgatccatatacttcgggtcgagttcttggtegat  
ctcttccctcctccacctcctcctcacagggatgtgccccttcgggttcttggatttatt  
**gttctaggttggttagtacggggcgttgatggttaggaaaggggatctgtatctgtgatgatc**  
**ctgttcttggatttgggatagaggggttcttgatggtgcatggtatcgggtcggttgatta**  
**gtagtatggttttcaatcgtctggagagctctatggaaaatgaaatggtttagggtagcgaat**  
cttgcgattttgtgagtaccttttgtttgaggtaaaaatcagagcaccgggtgattttgcttgg  
tgtaataaaagtagcgggtggtttggctcctcgattctggtagtgatgcttctcgatttgacgaa  
gctatcctttgtttattccctattgaacaaaaataatccaactttgaagacgggtcccgttga  
tgagattgaatgattgattcttaagcctgtccaaaatttcgcagctggcttgtttagataca  
gtagtccccatcacgaaattcatggaaacagttataatcctcaggaacaggggattccctgt  
tcttccgatttgctttagtcccagaattttttttcccaaatatcttaaaaagtcactttctg  
gttcagttcaatgaattgattgctacaaaataatgctttttatagcgttatcctagctgtagtt  
cagttaataggtaataccctatagtttagtcaggagaagaacttatccgatttctgatctc  
catttttaattatatagaaatgaactgtagcataagcagttatcatttggattatttttttta  
ttagctctcacccttcattattctgagctgaaagtctggcatgaactgtcctcaattttgt  
tttcaaattcacatcgattatctatgcattatcctcttgtatctacctgtagaagttcttt  
ttgggtattccttgactgcttgattacagaaagaaatttatgaagctgtaatcgggatagtt  
atactgcttgttcttatgattcatttcctttgtgcagttcttgggtgtagcttgccact

SEC ID N° 2:

taggagcatccaagccaagaagaggggagagcaccaaggacacgcgactagcagaagccgag  
cgaccgccttcttcgatccatatacttcgggtcgagttcttggtegatctcttccctcctcca  
cctcctcctcacagggatgtgccccttcgggttcttggatttattgttctaggttggtga  
gtacggggcgttgatggttaggaaaggggatctgtatctgtgatgattcctgttcttggatttg  
ggatagaggggttcttgatggtgcatggtatcgggtcggttgattagtagtatggttttca  
atcgtctggagagctctatggaaaatgaaatggtttagggtagcgaatccttgcgattttgtga  
gtaccttttgtttgaggtaaaaatcagagcaccgggtgattttgcttgggtgtaataaaagtagc  
gttgtttggctcctcgattctggtagtgatgcttctcgatttgacgaagctatcctttgttta  
ttccctattgaacaaaaataatccaactttgaagacgggtcccgttgatgagattgaatgatt  
gattcttaagcctgtccaaaatttcgcagctggcttgttttagatacagtagtccccatcacg  
aaattcatggaaacagttataatcctcaggaacaggggattccctgttcttccgatttgctt  
tagtcccagaattttttttcccaaatatcttaaaaagtcactttctggttcagttcaatgaa  
ttgattgctacaaaataatgctttttatagcgttatcctagctgtagttcagttaataggtaat  
accctatagtttagtcaggagaagaacttatccgatttctgatctccatttttaattatata  
gaaatgaactgtagcataagcagttatcatttggattattttttttattagctctcaccct  
tcattattctgagctgaaagtctggcatgaactgtcctcaattttgttttcaaattcacatc  
gattatctatcgattatcctcttgtatctacctgtagaagtttctttttggttattccttga  
ctgcttgattacagaaagaaatttatgaagctgtaatcgggatagttatactgcttgttctt  
atgattcatttcctttgtgcagttcttgggtgtagcttgccactttcaccagcaaagtttc

FIGURA 5

**SEC ID N° 3:**

atgtctgatctcgacattcagatocccaactgccttcgatccctttgctgaggccaatgctgg  
agactctggtgocggtgocaggatcaaaggactacgttcatgtaecgatccagcagcgtaatg  
gccgtaagagcctgaccactgtccagggttgagaaggaattcagctacaacaagatcctt  
aaagatctcaagaaagagttttgctgcaatggtactggtgtccaggaccagagcttggcca  
ggtcattcaacttcagggtgatcagaggaagaacgtatcaaattttctgttcaggccggca  
ttgtgaagaaggaacacatcaagattcatggtttctga

**SEC ID N° 4:**

MSDLDIQIPTAFDPFAEANAGDSGAAAGSKDYVHVRIQQRNGRKSLLTVQGLKKEFSYNKIL  
KDLKKEFCNGTVVQDPELGQVIQLQGDQRKNVSNFLVQAGIVKKEHIKIHF

**SEC ID N° 5:**

gtatgtgccttcgggttgttcttgatttattgttctaggttggtgtagtacgggcttgatg  
ttaggaaaggggatctgtatctgtgatgattcctgttcttgatttgggatagaggggttct  
tgatggtgcatggttatcgggttcgggttgattagtagtatggttttcaatcgtctggagagct  
ctatggaaatgaaatgggttagggtagcgaatcttgcgattttgtgagtagcttttgttga  
ggtaaaatcagagcaccgggtgattttgcttggtgtaataaaagtagcgggtggttggctctcg  
attctggtagtgatgcttctcgatttgacgaagctatcctttgtttatccctattgaacia  
aaataatccaactttgaagacggtcccgttgatgagattgaaatgattgattcttaagcctgt  
ccaaaatttcgcagctggcttgttagatacagtagtccccatcacgaaattcatggaaaca  
gttataatcctcaggaacaggggattccctgttcttccgatttgcttagtccagaatttt  
tttcccaaatatcttaaaaagtcactttctgggttcagttcaatgaattgattgctacaat  
aatgcttttatagcgttatcctagctgtagttcagttaataggtaatacccctatagtttag  
tcaggagaagaacttatccgatttctgatctccatttttaattatgaaatgaactgtagc  
ataagcagattcatttggattatTTTTTTTattagctctcacccttcattattctgagct  
gaaagtctggcatgaactgtcctcaattttgTTTTTcaaattcacatcgattatctatgcatt  
atcctcttgatctacctgtagaagtttctTTTTTgggttattccttgactgcttgattacaga  
aagaaatttatgaagctgtaatcgggatagttatactgcttggttcttatgattcatttctt  
tgtgcag

**FIGURA 5 (continuación)**