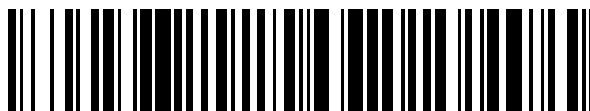


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 420 929**

51 Int. Cl.:

**C12N 15/09** (2006.01)

**C12N 15/63** (2006.01)

**C12N 15/82** (2006.01)

**C12N 15/87** (2006.01)

**C12N 15/90** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.07.2003 E 03765524 (8)**

97 Fecha y número de publicación de la concesión europea: **05.06.2013 EP 1551966**

54 Título: **Procedimientos de utilización de polinucleótidos artificiales y sus composiciones para reducir el silenciamiento de transgenes**

30 Prioridad:

**18.07.2002 US 396665 P**

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

**27.08.2013**

73 Titular/es:

**MONSANTO TECHNOLOGY LLC (100.0%)  
800 NORTH LINDBERGH BOULEVARD  
ST. LOUIS, MISSOURI 63167, US**

72 Inventor/es:

**FLASINSKI, STANISLAW**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 420 929 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

Procedimientos de utilización de polinucleótidos artificiales y sus composiciones para reducir el silenciamiento de transgenes

**Antecedentes de la invención**5 **Campo de la invención**

La presente invención se refiere a la ingeniería genética de plantas. Más en concreto, la invención también se refiere a células vegetales, plantas o su progenie tolerantes al glifosato, que contienen polinucleótidos artificiales específicos que codifican un polipéptido de EPSPS.

**Descripción de la técnica relacionada**

10 Pueden aislarse genes heterólogos a partir de una fuente diferente de la planta en la que se van a transformar, o pueden modificarse o diseñarse para que tengan cualidades diferentes o mejoradas. Los rasgos o cualidades de interés particularmente deseables para la ingeniería genética de plantas incluyen, pero no se limitan a la resistencia a insectos, a enfermedades fúngicas, y otras plagas y agentes causantes de enfermedad, la tolerancia a herbicidas, una estabilidad o caducidad potenciada, rendimiento, tolerancias a estreses ambientales, y  
15 potenciaciones nutricionales.

Los procedimientos de biología molecular tradicionales para generar nuevos genes y proteínas en general implican la mutagénesis aleatoria o dirigida. Un ejemplo de mutagénesis aleatoria es una técnica recombinante conocida como "reordenamiento del ADN", según se describe en las patentes de EEUU 5.605.793; 5.811.238; 5.830.721; 5.837.458, y los documentos WO 98/31837, WO 99/65927. Un procedimiento alternativo para la evolución  
20 molecular implica un procedimiento de extensión en etapas (StEP) para la mutagénesis y recombinación *in vitro* de secuencias de moléculas de ácidos nucleicos, según se describe en la patente de EEUU 5.965.408. Un ejemplo de mutagénesis dirigida es la introducción de una mutación puntual en un sitio específico en un polipéptido. Una estrategia alternativa, que es útil cuando el gen heterólogo procede de una fuente no vegetal, es diseñar un gen insecticida artificial que usa el codón que se emplea de modo más habitual en la tabla de utilización de codones de la planta del maíz (Kozziel *et al.*, 1993, *Biotechnology*, 11, 194-200). Fischhoff y Perlak (patente de EEUU n.º  
25 5.500.365) divulgan una mayor expresión de la proteína insecticida de *Bacillus thuringiensis* (Bt) comparada con plantas cultivadas, cuando la secuencia del polinucleótido se modifica para reducir la aparición de secuencias desestabilizantes. Es necesario modificar la secuencia del polinucleótido de Bt de tipo salvaje, porque el polinucleótido de Bt de longitud completa de tipo salvaje no expresa niveles suficientes de proteína insecticida en plantas para que sea útil desde un punto de vista agronómico.  
30

Los genes heterólogos se clonan en vectores adecuados para la transformación de plantas. Las técnicas de transformación y regeneración útiles para incorporar genes heterólogos en el genoma de una planta son muy conocidas en la técnica. El gen después puede expresarse en la célula vegetal para que muestre la característica o el rasgo añadido. Sin embargo, los genes heterólogos que normalmente se expresan bien como transgenes  
35 pueden sufrir el silenciamiento de genes cuando más de una copia del mismo gen se expresa en la misma planta. Esto puede suceder cuando un primer gen heterólogo es demasiado similar a una secuencia de ADN de un gen endógeno en la planta. Otros ejemplos incluyen cuando una planta transgénica posteriormente se cruza con otras plantas transgénicas que tienen el mismo transgén o transgenes similares, o cuando la planta transgénica se vuelve a transformar con un módulo de expresión de plantas que contiene el mismo gen o un gen similar. De forma parecida, puede producirse el silenciamiento de genes si el apilamiento de rasgos emplea los mismos elementos genéticos utilizados para dirigir la expresión del gen transgénico de interés. Para apilar rasgos, las líneas transgénicas estables deben prepararse con diferentes combinaciones de genes y elementos genéticos para evitar el silenciamiento de genes. La N-fosfonometilglicina, también conocida como glifosato, es un herbicida muy conocido que tiene actividad en un amplio espectro de especies vegetales. El glifosato es el ingrediente activo de Roundup® (Monsanto Co.), un herbicida seguro que tienen una semivida deseablemente corta en el entorno.  
40 Cuando se aplica a una superficie vegetal, el glifosato se mueve de modo sistémico a través de la planta. El glifosato es fitotóxico debido a que inhibe la vía del ácido shikímico, que proporciona un precursor para la síntesis de aminoácidos aromáticos. El glifosato inhibe la enzima 5-enolpiruvil-3-fosfoshikimato sintasa (EPSPS).  
45

También puede lograrse la tolerancia al glifosato mediante la expresión de variantes de EPSPS que tienen menor afinidad por el glifosato y, así, mantener su actividad catalítica en presencia del glifosato (patente de EEUU n.º 5.633.435). Las enzimas que degradan el glifosato en tejidos vegetales (patente de EEUU n.º 5.463.175) también son capaces de conferir tolerancia celular al glifosato. Estos genes se emplean para la producción de cultivos transgénicos que son tolerantes al glifosato, permitiendo con ello utilizar el glifosato para el control eficaz de especies adventicias con problemas mínimos de daños para el cultivo. Por ejemplo, la tolerancia al glifosato se ha  
50 introducido genéticamente en el maíz (patente de EEUU n.º 5.554.798), trigo (solicitud de patente de EEUU n.º  
55

20020062503), soja (solicitud de patente de EEUU n.º 20020157139) y canola (documento WO 9204449). Los transgenes para la tolerancia al glifosato y los transgenes para la tolerancia a otros herbicidas, por ejemplo, el gen *bar* (Toki *et al.*, Plant Physiol., 100:1503-1507, 1992; Thompson *et al.*, EMBO J., 6:2519-2523, 1987, fosfinotricina acetiltransferasa, gen BAR aislado de *Streptomyces*; DeBlock *et al.*, EMBO J., 6:2513-2522, 1987, herbicida glufosinato) también son útiles como marcadores seleccionables o marcadores puntuables, y pueden proporcionar un fenotipo útil para la selección de plantas unido a otros rasgos útiles desde el punto de vista agronómico.

En la técnica son necesarios procedimientos para diseñar genes para la expresión en plantas para mejorar rasgos útiles desde el punto de vista agronómico que eviten el silenciamiento de genes cuando se insertan múltiples copias y se produce la recombinación con genes vegetales endógenos.

10 **Breve descripción de los dibujos**

Figura 1. Comparación en apilamiento de los cambios en las secuencias polinucleotídicas de dos versiones de EPSPS de arroz artificiales (OsEPSPS\_AT, OsEPSPS\_ZM) y una EPSPS de arroz nativa (OsEPSPS\_Nat), cuyos polipéptidos están modificados para ser resistentes al glifosato.

15 Figura 2. Comparación en apilamiento de las secuencias polinucleotídicas de una EPSPS de maíz nativa (ZmEPSPS\_Nat) y artificial (ZmEPSPS\_ZM), cuyos polipéptidos están modificados para ser resistentes al glifosato.

Figura 3. Comparación en apilamiento de las secuencias polinucleotídicas de una EPSPS de soja nativa (GmEPSPS\_Nat) y una versión artificial (GmEPSPS\_GM), cuyos polipéptidos están modificados para ser resistentes al glifosato.

20 Figura 4. Comparación en apilamiento de las secuencias polinucleotídicas de un gen BAR nativo (BAR1\_Nat) y dos versiones artificiales con sesgo de codones de *Zea mays* (BAR1\_ZM) y *Arabidopsis thaliana* (BAR1\_AT).

Figura 5. Comparación en apilamiento de las secuencias polinucleotídicas de CTP2 y CP4EPSPS nativas (CTP2CP4\_Nat) y versiones artificiales (CTP2CP4\_AT, CTP2CP4\_ZM, y CTP2CP4\_GM).

25 Figures 6 a 25. Mapas de plásmidos de pMON54949, pMON54950, pMON30151, pMON59302, pMON59307, pMON42411, pMON58400, pMON58401, pMON54964, pMON25455, pMON30152, pMON54992, pMON54985, pMON20999, pMON45313, pMON59308, pMON59309, pMON59313, pMON59396 y pMON25496, respectivamente.

**Breve descripción del listado de secuencias**

SEQ ID NO:1	OsEPSPS_TIPS	Una secuencia de una proteína de EPSPS de arroz modificada para que sea resistente al glifosato, con el péptido de tránsito del cloroplasto.
SEQ ID NO:2	OsEPSPS_Nat	Una secuencia polinucleotídica de un polinucleótido de EPSPS nativo de arroz modificado para que sea codifique una proteína resistente al glifosato.
SEQ ID NO:3	OsEPSPS_AT	Una secuencia polinucleotídica de un polinucleótido de EPSPS de arroz artificial utilizando la tabla de utilización de codones de <i>Arabidopsis</i> , y posteriormente modificada para que codifique una proteína resistente al glifosato.
SEQ ID NO:4	OsEPSPS_ZM	Una secuencia polinucleotídica de un polinucleótido de EPSPS de arroz artificial utilizando la tabla de utilización de codones de <i>Zea mays</i> , y posteriormente modificada para que codifique una proteína resistente al glifosato.
SEQ ID NO:5	GmEPSPS_IKS	Una secuencia de una proteína de EPSPS de soja modificada para que sea resistente al glifosato, con el péptido de tránsito del cloroplasto.
SEQ ID NO:6	GmEPSPS_Nat	Una secuencia polinucleotídica de un polinucleótido de EPSPS nativo de soja modificado para que sea codifique una proteína resistente al glifosato.
SEQ ID NO:7	GmEPSPS_GM	Una secuencia polinucleotídica de un polinucleótido de EPSPS de soja artificial utilizando la tabla de utilización de codones de <i>Glycine max</i> , y

ES 2 420 929 T3

		posteriormente modificada para que codifique una proteína resistente al glifosato.
SEQ ID NO:8	ZmEPSPS_TIPS	Una secuencia de una proteína de EPSPS de maíz modificada para que sea resistente al glifosato, con el péptido de tránsito del cloroplasto.
SEQ ID NO:9	ZmEPSPS_Nat	Una secuencia polinucleotídica de un polinucleótido de EPSPS nativo de maíz modificado para que sea codifique una proteína resistente al glifosato.
SEQ ID NO:10	ZmEPSPS_ZM	Una secuencia polinucleotídica de un polinucleótido de EPSPS de maíz artificial utilizando la tabla de utilización de codones de <i>Zea mays</i> , y posteriormente modificada para que codifique una proteína resistente al glifosato.
SEQ ID NO:11	CTP2	Una secuencia proteica de un péptido de tránsito del cloroplasto 2 del gen EPSPS de <i>Arabidopsis</i> .
SEQ ID NO:12	CTP2_Nat	Una secuencia polinucleotídica del péptido de tránsito del cloroplasto de EPSPS de <i>Arabidopsis</i> .
SEQ ID NO:13	CTP2_AT	Una secuencia polinucleotídica de un polinucleótido artificial que codifica CTP2 utilizando la tabla de utilización de codones de <i>Arabidopsis</i> .
SEQ ID NO:14	CTP2_ZM	Una secuencia polinucleotídica de un polinucleótido artificial que codifica CTP2 utilizando la tabla de utilización de codones de <i>Zea mays</i> .
SEQ ID NO:15	CP4EPSPS	Una secuencia proteica de una proteína de EPSPS resistente al glifosato de la cepa CP4 de <i>Agrobacterium</i> .
SEQ ID NO:16	CP4EPSPS_Nat	Una secuencia polinucleotídica de un polinucleótido nativo que codifica la proteína CP4EPSPS (patente de EEUU 5.693.435).
SEQ ID NO:17	CP4EPSPS_AT	Una secuencia polinucleotídica de un polinucleótido artificial que codifica la proteína CP4EPSPS utilizando la tabla de utilización de codones de <i>Arabidopsis</i> .
SEQ ID NO:18	CP4EPSPS_ZM	Una secuencia polinucleotídica de un polinucleótido artificial que codifica la proteína CP4EPSPS utilizando la tabla de utilización de codones de <i>Zea mays</i> .
SEQ ID NO:19	BAR1	La secuencia proteica de una fosfinotricina acetiltransferasa.
SEQ ID NO:20	BAR1_Nat	Una secuencia polinucleotídica de un polinucleótido nativo aislado a partir de <i>Streptomyces</i> que codifica la fosfinotricina acetiltransferasa.
SEQ ID NO:21	BAR1_AT	Una secuencia polinucleotídica de un polinucleótido artificial que codifica la fosfinotricina acetiltransferasa utilizando la tabla de utilización de codones de <i>Arabidopsis</i> .
SEQ ID NO:22	BAR1_ZM	Una secuencia polinucleotídica de un polinucleótido artificial que codifica la fosfinotricina acetiltransferasa utilizando la tabla de utilización de codones de <i>Zea mays</i> .
SEQ ID NO:23	CP4EPSPS_Syn	Una secuencia polinucleotídica de un polinucleótido artificial con sesgo de codones de dicotiledóneas.
SEQ ID NO:24	CP4EPSPS_AT_p1	Una molécula de cebador de ADN de diagnóstico para el polinucleótido CP4EPSPS_AT.
SEQ ID NO:25	CP4EPSPS_AT_p2	Una molécula de cebador de ADN de diagnóstico para el polinucleótido

		CP4EPSPS_AT.
SEQ ID NO:26	CP4EPSPS_ZM_p1	Una molécula de cebador de ADN de diagnóstico para el polinucleótido CP4EPSPS_ZM.
SEQ ID NO:27	CP4EPSPS_ZM_p2	Una molécula de cebador de ADN de diagnóstico para el polinucleótido CP4EPSPS_ZM.
SEQ ID NO:28	CP4EPSPS_Nat_p1	Una molécula de cebador de ADN de diagnóstico para el polinucleótido CP4EPSPS_Nat.
SEQ ID NO:29	CP4EPSPS_Nat_p2	Una molécula de cebador de ADN de diagnóstico para el polinucleótido CP4EPSPS_Nat.
SEQ ID NO:30	CP4EPSPS_Syn_p1	Una molécula de cebador de ADN de diagnóstico para el polinucleótido CP4EPSPS_Syn.
SEQ ID NO:31	CP4EPSPS_Syn_p2	Una molécula de cebador de ADN de diagnóstico para el polinucleótido CP4EPSPS_Syn.
SEQ ID NO:32	ZmAdh1 cebador 1	Cebador control 1 de diagnóstico para el gen Adh1 de maíz endógeno.
SEQ ID NO:33	ZmAdh1 cebador 2	Cebador control 2 de diagnóstico para el gen Adh1 de maíz endógeno.
SEQ ID NO:34	GNAGIMKS	Motivo que proporciona resistencia al glifosato a una planta EPSPS.
SEQ ID NO:35	CTPEPSPSCP4_GM	Una secuencia polinucleotídica de un polinucleótido artificial que codifica la proteína CP4EPSPS utilizando la tabla de utilización de codones de <i>Glicina max</i> .

### **Sumario de la invención**

La presente invención proporciona una célula vegetal, una planta o su progenie tolerante al glifosato, que comprende al menos dos polinucleótidos, en la que dichos al menos dos polinucleótidos codifican polipéptidos de EPSPS que son al menos 98% idénticos, siendo al menos uno de dichos al menos dos polinucleótidos un transgén que comprende SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:35, y dichos al menos dos polinucleótidos son sustancialmente divergentes y son menos del 85% similares en sus secuencias polinucleotídicas. De particular interés son las plantas o su progenie seleccionadas del grupo que consiste en trigo, maíz, arroz, soja, algodón, patata, canola, césped, árboles forestales, grano de sorgo, cultivos vegetales, plantas ornamentales, cultivos de forraje y cultivos frutales.

La presente invención proporciona también un procedimiento para seleccionar células vegetales, plantas o su progenie tolerantes al glifosato, según se definió anteriormente en la presente, que comprende las etapas de:

(a) transformar una célula vegetal con una construcción de ADN que comprende: una molécula de promotor que opera en plantas, unida operablemente a una molécula polinucleotídica artificial seleccionada de SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:17, SEQ ID NO:18 y SEQ ID NO:35; y

(b) cultivar dicha célula vegetal en un medio selectivo que contiene glifosato para destruir selectivamente las células que no han sido transformadas con dicha construcción de ADN; y

(c) regenerar dicha célula vegetal en una planta fértil.

### **Descripción detallada de la invención**

Se proporcionan las siguientes definiciones para definir con más precisión la presente invención y para guiar a los expertos en la técnica en la práctica de la presente invención. A menos que se indique lo contrario, los términos y las expresiones deben entenderse según la utilización convencional por parte de los expertos en la técnica pertinente. Las definiciones de los términos y las expresiones habituales en la biología molecular también pueden encontrarse en Rieger *et al.*, Glossary of Genetics: Classical and Molecular, 5ª edición, Springer-Verlag, Nueva York, (1991); y Lewin, Genes V, Oxford University Press, Nueva York, (1994). Se emplea la nomenclatura para las bases de ADN como se indica en 37 CFR § 1.822. Se utiliza la nomenclatura convencional de una letra y de tres letras para los restos aminoácidos.

"Sustituciones de aminoácidos" y "variantes de aminoácidos" son preferiblemente sustituciones de un único resto aminoácido por otro resto aminoácido en cualquier posición dentro de la proteína. Las sustituciones, deleciones, inserciones o cualquiera de sus combinaciones pueden combinarse para lograr una construcción final.

5 Un "polinucleótido artificial", tal como se emplea en la presente invención, es una secuencia de ADN diseñada según los procedimientos de la presente invención y creada como una molécula de ADN aislada para su uso en una construcción de ADN que proporciona la expresión de una proteína en células hospedantes, y para el objetivo de la clonación en construcciones apropiadas u otros usos conocidos por los expertos en la técnica. Para estos fines están disponibles programas informáticos que incluyen, pero no se limitan a los programas "BestFit" o "Gap" de Sequence Analysis Software Package, Genetics Computer Group (GCG), Inc., University of Wisconsin Biotechnology Center, Madison, WI 53711. El polinucleótido artificial puede crearse mediante uno o más procedimientos conocidos en la técnica que incluyen, pero no se limitan a PCR de solapamiento. Un polinucleótido artificial es sustancialmente divergente de otros polinucleótidos que codifican la proteína idéntica o casi idéntica.

15 El término "quimérico" se refiere a una secuencia de ácido nucleico o proteína de fusión. Una secuencia codificadora de ácido nucleico quimérica está formada por dos o más secuencias unidas dentro de marco que codifican una proteína quimérica. Un gen quimérico se refiere a los múltiples elementos genéticos derivados de fuentes heterólogas que comprenden un gen.

Las expresiones "secuencia codificadora", "marco de lectura abierto" y "secuencia estructural" se refieren a la región de tripletes de ácidos nucleicos secuenciales continuos que codifican una proteína, un polipéptido, o una secuencia peptídica.

20 Un "codón" se refiere a una secuencia de tres nucleótidos que especifica un aminoácido concreto.

La "utilización de codones" o "sesgo de codones" se refieren a la frecuencia de uso de codones que codifican aminoácidos en las secuencias codificadoras de organismos.

"Complementariedad" y "complemento", cuando se refieren a secuencias de ácidos nucleicos, se refieren a la unión específica de adenina a timina (uracilo en el ARN), y de citosina a guanina en las hebras opuestas del ADN o ARN.

25 Una "construcción" se refiere a los elementos genéticos heterólogos unidos operablemente entre sí que forman una molécula de ADN recombinante, y pueden comprender elementos que proporcionan la expresión de una molécula polinucleotídica de ADN en una célula hospedante, y elementos que proporcionan el mantenimiento de la construcción. La "región C-terminal" se refiere a la región de un péptido, un polipéptido o una cadena proteica desde la mitad hasta el extremo que porta el aminoácido que tiene un grupo carboxilo libre.

30 El término "divergente", tal como se emplea en la presente, se refiere a la comparación de moléculas polinucleotídicas que codifican la misma proteína o polipéptido o casi la misma proteína o polipéptido. Las cuatro letras del código genético (A, G, C y T/U) comprenden codones de tres letras que dirigen a las moléculas de ARN para que ensamblen aminoácidos para formar un polipéptido a partir de un molde de ARNm. Cuando existe más de un codón que puede codificar el mismo aminoácido, esto se denomina degeneración. Los codones degenerados se emplean para construir moléculas polinucleotídicas sustancialmente divergentes que codifican el mismo polipéptido, en las que estas moléculas tienen una secuencia de nucleótidos en su longitud completa en que son menos del 85% idénticas, y no existen tramos de secuencia polinucleotídica mayores que 23 nucleótidos que sean idénticos.

40 La expresión "ADN codificante" se refiere a polinucleótidos de ADN cromosómico, ADN plasmídico, ADNc, o ADN artificial que codifican cualquiera de las proteínas analizadas en el presente.

45 El término "endógeno" se refiere a materiales originados desde el interior de un organismo o una célula. Una "endonucleasa" se refiere a una enzima que hidroliza el ADN bicatenario en emplazamientos internos. "Exógeno" se refiere a materiales originados desde el exterior de un organismo o una célula. Esto generalmente se aplica a moléculas de ácidos nucleicos utilizadas para producir plantas y células hospedantes transformadas o transgénicas.

Un "exón" se refiere a la porción de un gen que realmente es traducido a una proteína, es decir, una secuencia codificadora.

El término "expresión" se refiere a la transcripción o la traducción de un polinucleótido para producir un correspondiente producto génico, una ARN o una proteína.

50 Un "fragmento" de un gen es una porción de una molécula de ácido polinucleotídico de longitud completa que tiene al menos una longitud mínima capaz de transcribirse en un ARN, de traducirse en un péptido, o que es útil como sonda o cebador en un procedimiento de detección de ADN.

El término "gen" se refiere a un polinucleótido de ADN cromosómico, ADN plasmídico, ADNc, ADN artificial u otro ADN que codifica un péptido, un polipéptido, una proteína o una molécula de ARN, y los elementos genéticos que flanquean la secuencia codificadora que están implicados en la regulación de la expresión.

5 El término "genoma", según se aplica a virus, incluye todas las secuencias de ácidos nucleicos contenidas dentro de la cápsida del virus. El término "genoma", según se aplica a bacterias, incluye el cromosoma y los plásmidos dentro de una célula hospedante bacteriana. Por tanto, los ácidos nucleicos codificantes de la presente invención introducidos en células hospedantes bacterianas puede estar cromosómicamente integrados o localizados en plásmidos. El término "genoma", según se aplica a células vegetales, incluye no solo el ADN cromosómico que se encuentra dentro del núcleo, sino también el ADN de orgánulos que se encuentra dentro de componentes subcelulares de la célula. Por tanto, los ácidos nucleicos de la presente invención introducidos en células vegetales puede estar cromosómicamente integrados o localizados en orgánulos.

15 El "glifosato" se refiere a la N-fosfonometilglicina y sus sales. El glifosato es el ingrediente activo del herbicida Roundup® (Monsanto Co.). El tratamiento de plantas con "glifosato" se refiere a tratamientos con la formulación herbicida Roundup® o Roundup Ultra®, a menos que se indique lo contrario. El glifosato, N-fosfonometilglicina y sus sales (no como el herbicida Roundup® formulado), es el componente de medios de cultivo sintéticos utilizados para la selección de tolerancia de plantas y bacterias al glifosato, o se emplea para determinar la resistencia a enzimas en ensayos bioquímicos *in vitro*.

Una secuencia de "ADN heterólogo" se refiere a una secuencia polinucleotídica que se origina a partir de una fuente o una especie extraña, o si procede de la misma fuente, está modificada con respecto a su forma original.

20 Un "ADN homólogo" se refiere a ADN de la misma fuente que la de la planta receptora.

La "hibridación" se refiere a la capacidad de una hebra de un ácido nucleico para unirse con una hebra complementaria a través del apareamiento de bases. La hibridación se produce cuando las secuencias complementarias en dos hebras de ácidos nucleicos se unen entre sí. Las sondas y los cebadores de ácidos nucleicos de la presente invención se hibridan bajo condiciones rigurosas a una secuencia de ADN diana. Puede utilizarse cualquier procedimiento de hibridación o de amplificación de ácidos nucleicos convencional para identificar la presencia de ADN que surge de un acontecimiento transgénico en una muestra. Las moléculas de ácidos nucleicos o sus fragmentos son capaces de hibridarse específicamente con otras moléculas de ácidos nucleicos bajo ciertas circunstancias. Tal como se emplea en la presente, se dice que dos moléculas de ácidos nucleicos son capaces de hibridarse específicamente entre sí si las dos moléculas son capaces de formar una estructura de ácido nucleico bicatenaria antiparalela. Se dice que una molécula de ácido nucleico es el "complemento" de otra molécula de ácido nucleico si muestran complementariedad completa. Tal como se emplea en la presente, se dice que las moléculas muestran " complementariedad completa" cuando cada nucleótido de una de las moléculas es complementario con un nucleótido de la otra. Se dice que dos moléculas son "mínimamente complementarias" si puede hibridarse entre sí con la suficiente estabilidad para permitir que permanezcan asociadas entre sí al menos bajo condiciones convencionales de "baja rigurosidad". De forma similar, se dice que las moléculas son "complementarias" si pueden hibridarse entre sí con la suficiente estabilidad como para permitir que permanezcan asociadas entre sí bajo condiciones convencionales de "alta rigurosidad". Las condiciones de rigurosidad convencionales se describen en Sambrook *et al.*, 1989, y en Haymes *et al.*, en: Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, DC (1985). Por tanto, son permisibles desviaciones de la complementariedad completa, con la condición de que dichas desviaciones no impidan completamente la capacidad de la molécula para formar una estructura bicatenaria. Para que una molécula de ácido nucleico pueda actuar como sonda o cebador, sólo debe ser suficientemente complementaria en su secuencia como para ser capaz de formar una estructura bicatenaria estable con el disolvente y las concentraciones salinas concretos utilizados.

45 Tal como se emplea en la presente, una secuencia sustancialmente homóloga es una secuencia de ácido nucleico que se hibrida específicamente con el complemento de la secuencia de ácido nucleico con el que se está comparando bajo condiciones de alta rigurosidad. La expresión "condiciones rigurosas" se define desde el punto de vista funcional con respecto a la hibridación de una sonda de ácido nucleico con un ácido nucleico diana (es decir, con una secuencia de ácido nucleico concreta de interés) mediante el procedimiento de hibridación específica analizado en Sambrook *et al.*, 1989, en 9.52-9.55. Véase también, Sambrook *et al.*, 1989, en 9.47-9.52, 9.56-9.58; Kanehisa (Nucl. Acids Res., 12:203-213, 1984); y Wetmur y Davidson (J. Mol. Biol., 31:349-370, 1988). Por consiguiente, las secuencias de nucleótidos puede utilizarse por su capacidad para formar selectivamente moléculas de dúplex con tramos complementarios de fragmentos de ADN. Dependiendo de la aplicación prevista, se puede desear emplear condiciones variables de hibridación para lograr grados variables de selectividad de la sonda por la secuencia diana. Para aplicaciones que requieren una alta selectividad, generalmente se desean emplar condiciones relativamente rigurosas para formar los híbridos, por ejemplo, se seleccionarán condiciones de contenido en sal relativamente bajo y/o de temperatura alta, tales como las proporcionadas por NaCl

aproximadamente 0,02 M a aproximadamente 0,15 M a unas temperaturas de aproximadamente 50 °C a aproximadamente 70 °C. Unas condiciones rigurosas, por ejemplo, son lavar el filtro de hibridación al menos dos veces con tampón de lavado de alta rigurosidad (0,2x SSC, SDS al 0,1%, 65 °C). Las condiciones de rigurosidad apropiadas que estimulan la hibridación del ADN, por ejemplo, 6,0x cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45 °C, seguido de un lavado de 2,0x SSC a 50 °C, son conocidas por los expertos en la técnica, o pueden encontrarse en *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Por ejemplo, la concentración salina en la etapa de lavado puede seleccionarse de una baja rigurosidad de aproximadamente 2,0x SSC a 50 °C, hasta una alta rigurosidad de aproximadamente 0,2x SSC a 50 °C. Además, la temperatura en la etapa de lavado puede aumentar desde una condiciones de baja rigurosidad a temperatura ambiente, aproximadamente 22 °C, a una condiciones de alta rigurosidad a aproximadamente 65 °C. Tanto la temperatura como las sales pueden variar, o la temperatura o la concentración salina puede mantenerse constante mientras que se cambia la otra variable. Estas condiciones selectivas toleran poca, o ninguna, divergencia entre la sonda y el molde o hebra diana. La detección de secuencias de ADN mediante hibridación es muy conocida por los expertos en la técnica, y las indicaciones de las patentes de EEUU n.º 4.965.188 y 5.176.995 son ejemplares de los procedimientos de análisis de hibridación.

"Coincidencia" se refiere al grano de similitud entre dos secuencias de proteínas o de ácido polinucleotídicos. Se realiza un alineamiento de las dos secuencias mediante un programa informático adecuado. Un programa informático aceptado muy utilizado para realizar alineamientos de secuencias es CLUSTALW v1.6 (Thompson, *et al.*, *Nucl. Acids Res.*, 22: 4673-4680, 1994). El número de bases o aminoácidos apareados se divide entre el número total de bases o aminoácidos, y se multiplica por 100 para obtener un porcentaje de coincidencia. Por ejemplo, si dos secuencias de 580 pares de bases tienen 145 bases desapareadas, serán 25% idénticas. Si las dos secuencias comparadas tienen diferente longitud, el número de apareamientos se divide entre la longitud más corta de las dos. Por ejemplo, si hay 100 aminoácidos apareados entre una proteína de 200 y de 400 aminoácidos, estas son 50% idénticas con respecto a la secuencia más corta. Si la secuencia más corta tiene una longitud menor que 150 bases o 50 aminoácidos, el número de apareamientos se divide entre 150 (para las bases del ácido nucleico) o 50 (para aminoácidos) y se multiplica por 100 para obtener un porcentaje de coincidencia.

Tal como se describe en la presente, una proteína puede ser "sustancialmente idéntica" a proteínas relacionadas. Estas proteínas con coincidencia sustancial en general comprenden al menos una secuencia polipeptídica que tiene al menos 98% de coincidencia de secuencia, comparada con otras secuencias polipeptídicas relacionadas. El programa Gap del paquete WISCONSIN PACKAGE, versión 10.0-UNIX de Genetics Computer Group, Inc. se basa en el procedimiento de Needleman y Wunsch (*J. Mol. Biol.*, 48:443-453, 1970) utilizando el conjunto de parámetros por defecto para la comparación apareada (para la comparación de secuencias de aminoácidos: penalización de creación de huecos = 8, penalización de extensión de huecos = 20); o utilizando el programa TBLASTN en el paquete informático BLAST 2.2.1 (Altschul *et al.*, *Nucleic Acids Res.*, 25:3389-3402), utilizando la matriz BLOSUM62 (Henikoff y Henikoff, *Proc. Natl. Acad. Sci. U.S.A.*, 89:10915-10919, 1992) y el conjunto de parámetros por defecto para la comparación apareada (coste de creación de huecos = 11, coste de extensión de huecos = 1). En BLAST, el valor de E, o valor de esperanza, representa el número de alineamientos diferentes con unas puntuaciones equivalentes o mejores que la puntuación de alineamiento bruta, S, que se espera que aparezca por azar en la búsqueda en las bases de datos. cuanto menor sea el valor de E, más significativo será el apareamiento. Debido a que el tamaño de la base de datos es un elemento en el cálculo del valor de E, los valores de E obtenidos mediante la utilización de BLAST ("BLASTing") frente a bases de datos públicas, tales como GenBank, en general han sufrido un aumento a lo largo del tiempo para cualquier apareamiento problema/entrada. El porcentaje de coincidencia se refiere al porcentaje de restos aminoácidos idénticamente apareados que existen a lo largo de la longitud de la porción de las secuencias que es alineada por el algoritmo BLAST.

Un "intrón" se refiere a una porción de un gen que no es traducido a una proteína, aunque se transcriba a ARN.

Una secuencia de ácido nucleico "aislada" está sustancialmente separada o purificada de otras secuencias de ácidos nucleicos con las que el ácido nucleico normalmente está asociado en la célula del organismo en el que el aparece el ácido nucleico de forma natural, es decir, otro ADN cromosómico o extracromosómico. El término incluye los ácidos nucleicos que se purifican de modo bioquímico para eliminar sustancialmente los ácidos nucleicos contaminantes y otros componentes celulares. El término también incluye ácidos nucleicos recombinantes y ácidos nucleicos sintetizados de modo químico.

Polipéptidos "aislados", "purificados" y "homogéneos": Un polipéptido está "aislado" si se ha separado de los componentes celulares (ácidos nucleicos, lípidos, carbohidratos y otros polipéptidos) que le acompañan en la naturaleza, o se ha sintetizado de modo químico o es recombinante. Un polipéptido monomérico está aislado cuando al menos 60% en peso de una muestra está compuesta del polipéptido, preferiblemente 90% o más, más preferiblemente 95% o más, y lo más preferiblemente más del 99%. La pureza o homogeneidad de la proteína se determina, por ejemplo, mediante una electroforesis en gel de poliacrilamida de una muestra de proteína, seguido de la visualización de una única banda de polipéptido tras teñir el gel de poliacrilamida; una cromatografía líquida



de alta presión; u otros procedimientos convencionales. Las proteínas pueden purificarse mediante cualquier medio conocido en la técnica, por ejemplo, según se describe en *Guide to Protein Purification*, ed. Deutscher, Meth. Enzymol., 185, Academic Press, San Diego, 1990; y *Scopes, Protein Purification: Principles and Practice*, Springer Verlag, Nueva York, 1982.

5 "Marcaje" o "marcado": Existe una diversidad de procedimientos y reactivos convencionales para marcar polinucleótidos y polipéptidos y sus fragmentos. Los marcadores típicos incluyen isótopos radiactivos, ligandos y receptores de ligandos, fluróforos, agente quimioluminiscentes, y enzimas. Los procedimientos para marcar y las indicaciones para la elección de marcadores apropiados para diversos fines se analizan, por ejemplo, en Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); y *Current Protocols in Molecular Biology*, ed. Ausubel *et al.*, Greene Publishing and Wiley-Interscience, Nueva York (1992).

10 La expresión "región codificadora de la proteína madura" se refiere a la secuencia de un producto de proteína procesado, es decir, una EPSP sintasa madura que permanece después de que el péptido de tránsito del cloroplasto se ha retirado.

15 El término "nativo" se refiere, en general, a un polipéptido o un ácido polinucleotídico natural ("de tipo salvaje"). Sin embargo, en el contexto de la presente invención, pueden haberse producido algunas modificaciones de un polipéptido o un polinucleótido aislado para proporcionar un polipéptido con un fenotipo concreto, por ejemplo, la sustitución de aminoácidos en EPSPS sensible al glifosato proporciona una EPSPS resistente al glifosato. Para fines comparativos en la presente invención, el polinucleótido aislado que contiene unos pocos nucleótidos sustituidos para proporcionar una modificación de aminoácidos para la tolerancia a herbiciads se denomina el polinucleótido "nativo" cuando se compara con el polinucleótido sustancialmente divergente.

20 Sin embargo, el polinucleótido "nativo" modificado de esta manera no es nativo con respecto a los elementos genéticos que normalmente se encuentran unidos a un polinucleótido sin modificar natural.

La "región N-terminal" se refiere a una región de una cadena de un péptido, un polipéptido, o una proteína desde el aminoácido que tiene un grupo amino libre hasta la mitad de la cadena.

25 Un "ácido nucleico" se refiere al ácido desoxirribonucleico (ADN) o al ácido ribonucleico (ARN).

Códigos de los ácidos nucleicos: A = adenosina; C = citosina; G = guanosina; T = timidina. Códigos utilizados para la síntesis de oligonucleótidos: N = A, C, G y T equimolares; I = desoxiinosina; K = G y T equimolares; R = A y G equimolares; S = C y G equimolares; W = A y T equimolares; Y = C y T equimolares.

30 Un "segmento de ácido nucleico" o un "segmento de una molécula de ácido nucleico" es una molécula de ácido nucleico que se ha aislado para que esté libre del ADN genómico total de una especie concreta, o que se ha sintetizado. Se incluyen dentro de la expresión "segmento de ácido nucleico" los segmentos de ADN, vectores recombinantes, plásmidos, cósmidos, fágmidos, fagos, virus, etc.

35 "Variantes de secuencia de nucleótidos": Utilizando procedimientos muy conocidos, los expertos en la técnica pueden producir con facilidad variantes de secuencia de nucleótidos y de aminoácidos de genes y proteínas, respectivamente. Por ejemplo, las moléculas de ADN "variantes" son moléculas de ADN que contienen cambios en la secuencia de un gen EPSPS, es decir, cambios que incluyen que uno o más nucleótidos de la secuencia del gen EPSPS están delecionados, añadidos y/o sustituidos, de modo que el gen EPSPS variante codifica una proteína que mantiene la actividad EPSPS. Las moléculas de ADN variantes pueden producirse, por ejemplo, mediante técnicas de mutagénesis de ADN convencionales, o sintetizando químicamente la molécula de ADN variante o su porción. Los procedimientos para la síntesis químicas de ácidos nucleicos se analizan, por ejemplo, en Beaucage *et al.*, *Tetra. Letts.*, 22:1859-1862 (1981); y Matteucci *et al.*, *J. Am. Chem. Soc.*, 103:3185 (1981). La síntesis química de ácidos nucleicos puede realizarse, por ejemplo, en sintetizadores de oligonucleótidos automáticos. Estos variantes preferiblemente no cambian el marco de lectura de la región que codifica la proteína del ácido nucleico, y preferiblemente codifican una proteína que no tiene cambios, o sólo una pequeña reducción.

45 Un "marco de lectura abierto (ORF)" se refiere a una región del ADN o ARN que codifica un péptido, un polipéptido, o una proteína.

50 "Unido operablemente": Una primera secuencia de ácido nucleico está unida "operablemente" con una segunda secuencia de ácido nucleico cuando la primera secuencia de ácido nucleico está colocada en una relación funcional con la segunda secuencia de ácido nucleico. Por ejemplo, un promotor está unido operablemente con una secuencia que codifica una proteína si el promotor lleva a cabo la transcripción o la expresión de la secuencia codificadora. En general, las secuencias de ADN unidas operablemente están contiguas y, cuando es necesario unir dos regiones codificadoras de proteínas, en un marco de lectura.

La "sobrexpresión" se refiere a la expresión de un ARN, un polipéptido o una proteína codificados por un ADN introducido en una célula hospedante, en la que el ARN, el polipéptido o la proteína normalmente no está presente en la célula hospedante, o en la que el ARN, el polipéptido o la proteína está presente en dicha célula hospedante a un nivel mayor que el que se expresa normalmente a partir del gen endógeno que codifica el ARN, el polipéptido o la proteína.

El término "planta" incluye cualquier planta superior y su progenie, incluyendo monocotiledóneas (por ejemplo, maíz, arroz, trigo, cebada), dicotiledóneas (por ejemplo, soja, algodón, canola, tomate, patata, *Arabidopsis*, tabaco), gimnospermas (pinos, abetos, cedros), e incluye partes de plantas, que incluyen las unidades reproductoras de una planta (por ejemplo, semillas, bulbos, tubérculos, fruto, flor) u otras partes o tejidos a partir de los cuales pueda reproducirse la planta.

Un "módulo de expresión en plantas" se refiere a segmentos de ADN quiméricos que comprenden los elementos reguladores que están unidos operablemente para proporcionar la expresión de un producto transgénico en plantas. Un "plásmido" se refiere a un trozo de ADN circular, extracromosómico y autorreplicante. Una "señal de poliadenilación" o "señal de poliA" se refiere a una secuencia de ácido nucleico localizada 3' a la región codificadora que provoca la adición de nucleótidos adenilato al extremo 3' del ARNm transcrito a partir de la región codificadora.

Una "reacción en cadena de polimerasa (PCR)" se refiere a un procedimiento de amplificación del ADN que emplea una técnica enzimática para crear múltiples copias de una secuencia de ácido nucleico (amplión). Las copias de una molécula de ADN se preparan lanzando una ADN polimerasa entre dos amplímeros. La base de este procedimiento de amplificación son múltiples ciclos de cambios de temperatura para desnaturalizar, y después volver a aparear los amplímeros (moléculas de cebador de ADN), seguido de una extensión para sintetizar nuevas hebras de ADN en la región localizada entre los amplímeros flanqueantes. La amplificación de ácidos nucleicos puede realizarse mediante cualquiera de los diversos procedimientos de amplificación de ácidos nucleicos conocidos en la técnica, que incluyen la reacción en cadena de polimerasa (PCR). En la técnica se conoce una diversidad de procedimientos de amplificación y se describen, entre otros textos, en las patentes de EEUU n.º 4.683.195 y 4.683.202, y en PCR Protocols: A Guide to Methods and Applications, ed. Innis *et al.*, Academic Press, San Diego, 1990. Se han desarrollado procedimientos de amplificación de PCR para amplificar hasta 22 kb de ADN genómico y hasta 42 kb de ADN de bacteriófago (Cheng *et al.*, Proc. Natl. Acad. Sci. USA, 91:5695-5699, 1994). Estos procedimientos, así como otros procedimientos conocidos en la técnica de la amplificación de ADN, pueden utilizarse en la práctica de la presente invención.

Un polinucleótido se refiere a una longitud de dos o más moléculas de ácido desoxirribonucleico (ADN) y ácido ribonucleico (ARN), que están conectadas para formar una molécula más grande.

Fragmentos de polipéptidos: También se describen fragmentos de una proteína que carece de al menos un resto de una proteína nativa de longitud completa, pero que mantiene sustancialmente la actividad de la proteína.

El término "promotor" y la expresión "región promotora" se refieren a una molécula de ácido polinucleotídico que actúa como un elemento regulador, que se encuentra habitualmente cadena arriba (5') a una secuencia codificadora, que controla la expresión de la secuencia codificadora mediante el control de la producción de ARN mensajero (ARNm) proporcionando el sitio de reconocimiento para la ARN polimerasa y/o otros factores necesarios para el inicio de la transcripción en el sitio correcto. Tal como se contempla en la presente, un promotor o una región promotora incluye variaciones de promotores derivadas mediante el acoplamiento a diversas secuencias reguladoras, mutagénesis aleatoria o controlada, y la adición o la duplicación de secuencias potenciadoras. La región promotora descrita en la presente, y sus equivalentes biológicamente funcionales, es la responsable de dirigir la transcripción de las secuencias codificadoras bajo su control cuando se introducen en un hospedante como parte de un vector recombinante adecuado, según se demuestra por su capacidad para producir ARNm.

Un ácido nucleico "recombinante" se prepara mediante una combinación de dos segmentos de secuencia que están separados, por ejemplo, mediante una síntesis química o mediante la manipulación de segmentos aislados de ácidos nucleicos mediante técnicas de ingeniería genética.

La expresión "construcción de ADN recombinante" o "vector recombinante" se refiere a cualquier agente, tal como un plásmido, cósmido, virus, secuencia de replicación autónoma, fago, o secuencia nucleotídica de ADN o ARN monocatenaria o bicatenaria lineal o circular, derivada de cualquier fuente, capaz de una integración genómica o una replicación autónoma, que comprende una molécula de ADN a la que una o más secuencias de ADN se han unido de una manera funcionalmente operativa. Estos vectores o construcciones de ADN recombinante son capaces de introducir una región promotora o secuencia reguladora 5' y una secuencia de ADN para un producto génico seleccionado en una célula de tal manera que la secuencia de ADN se transcribe en un ARNm funcional

que se traduce y, por tanto, se expresa. Puede construirse vectores recombinantes o construcciones de ADN recombinante para que sean capaces de expresar ARN antisentido, para inhibir la traducción de un ARN de interés específico.

5 La "regeneración" se refiere al procedimiento de cultivar una planta a partir de una célula vegetal (por ejemplo, un explante o un protoplasto vegetal).

"Indicador" se refiere a un gen y al correspondiente producto génico que, cuando se expresa en organismos transgénicos, produce un producto detectable por procedimientos químicos o moleculares, o produce un fenotipo observable.

10 "Resistencia" se refiere a una enzima que es capaz de actuar en presencia de una toxina, por ejemplo, EPSP sintasas de clase II resistentes al glifosato. Una enzima que presenta resistencia a una toxina puede tener la función de desintoxicar la toxina, por ejemplo, la fosfinotricina acetiltransferasa, glifosato oxidorreductasa, o puede ser una enzima mutante que tiene actividad catalítica que no se ve afectada por un herbicida que altera la misma actividad en la enzima de tipo salvaje, por ejemplo, acetolactato sintasa, EPSP sintasas de clase I mutantes.

15 Una "enzima de restricción" se refiere a una enzima que reconoce una secuencia de nucleótidos palindrómica específica en ADN bicatenario y rompe ambas hebras; también se denomina endonucleasa de restricción. La ruptura se produce generalmente dentro del sitio de restricción.

20 Un "marcador seleccionable" se refiere a una molécula de ácido polinucleotídico que codifica una proteína, que confiere un fenotipo que facilita la identificación de células que contienen la molécula de ácido polinucleotídico. Los marcadores seleccionables incluyen los genes que confieren resistencia a antibióticos (por ejemplo, ampicilina, kanamicina), complementan una deficiencia nutricional (por ejemplo, uracilo, histidina, leucina), o imparten una característica visualmente distinguible (por ejemplo, cambios en el color o fluorescencia). Los genes marcadores seleccionables dominantes útiles incluyen genes que codifican resistencia a antibióticos (por ejemplo, neomicina fosfotransferasa, *aad*); y resistencia a herbicidas (por ejemplo, fosfinotricina acetiltransferasa, EPSP sintasa de clase II, EPSP sintasa de clase I modificada). Una estrategia útil para la selección de transformantes para la resistencia a herbicidas se describe, por ejemplo, en Vasil, *Cell Culture and Somatic Cell Genetics of Plants*, vols. I-III, Laboratory Procedures and Their Applications Academic Press, Nueva York (1984).

La expresión "específico para (una secuencia diana)" indica que una sonda de ADN o un cebador de ADN se hibrida bajo condiciones de hibridación concretas solo con la secuencia diana en una muestra que comprende la secuencia diana.

30 La expresión "sustancialmente purificado", tal como se emplea en la presente, se refiere a una molécula separada de otras moléculas con las que normalmente está asociada en su estado nativo. Más preferiblemente, una molécula sustancialmente purificada es la especie predominante presente en una preparación. Una molécula sustancialmente purificada puede estar más del 60% exenta, preferiblemente 75% exenta, más preferiblemente 90% exenta de las otras moléculas (excluyendo el disolvente) presentes en la mezcla natural. La expresión "sustancialmente purificado" no pretende incluir moléculas presentes en su estado nativo.

"Tolerante" o "tolerancia" se refiere a un efecto reducido de un agente biótico o abiótico sobre el crecimiento y el desarrollo de organismos y plantas, por ejemplo, una plaga o un herbicida.

"Transcripción" se refiere a un procedimiento para producir una copia de ARN a partir de un molde de ADN.

40 "Transformación" se refiere a un procedimiento para introducir una molécula de ácido polinucleotídico exógena (por ejemplo, una construcción de ADN, una molécula de ácido polinucleotídico recombinante) en una célula o un protoplasto, y esta molécula de ácido polinucleotídico exógena se incorpora en un cromosoma o es capaz de realizar una replicación autónoma.

45 "Transformado" o "transgénico" se refiere a una célula, un tejido, un órgano o un organismo en el cual se introduce un ácido polinucleotídico extraño, tal como un vector de ADN o una molécula de ácido polinucleotídico recombinante. Una célula u organismo "transgénico" o "transformado" también incluye la progenie de la célula u organismo y la progenie producida a partir de un programa de cultivo que emplee dicha planta "transgénica" como progenitor en un cruzamiento y que muestre un fenotipo alterado como resultado de la presencia de la molécula de ácido polinucleotídico extraña.

50 El término "transgén" se refiere a cualquier molécula de ácido polinucleotídico no nativa a una célula u organismo transformada en la célula u organismo. Un "transgén" también incluye las partes componentes de un gen vegetal nativo modificado mediante la inserción de una molécula de ácido polinucleotídico no nativa mediante la recombinación directa o la mutación específica de sitio.

Moléculas de "péptido de tránsito" o "péptido de transporte dirigido": Estos términos generalmente se refieren a moléculas peptídicas que, cuando se unen a una proteína de interés, dirigen la proteína hacia un tejido, una célula, un emplazamiento subcelular o un orgánulo celular concretos. Los ejemplos incluyen, pero no se limitan a péptidos de tránsito del cloroplasto, señales de transpote dirigido nuclear, y señales vacuolares. El péptido de tránsito del cloroplasto tiene una utilidad particular en la presente invención para dirigir la expresión de la enzima EPSP hacia el cloroplasto.

El término "traducción" se refiere a la producción del correspondiente producto génico, es decir, un péptido, un polipéptido, o una proteína, a partir de un ARNm.

Un "vector" se refiere a un plásmido, un cósmido, un bacteriófago, o un virus que porta ADN extraño hacia un organismo hospedante.

#### *Polinucleótidos*

Los procedimientos incluyen diseñar genes que confieren un rasgo de interés a la planta en la que se introducen. Los transgenes de interés agronómico que proporcionan rasgos agronómicos beneficiosos a plantas cultivadas, por ejemplo, incluyen, pero no se limitan a elementos genéticos que comprenden resistencia a herbicidas (patente de EEUU n.º 5.633.435; patente de EEUU n.º 5.463.175), mayor rendimiento (patente de EEUU n.º 5.716.837), control de insectos (patente de EEUU n.º 6.063.597; patente de EEUU n.º 6.063.756; patente de EEUU n.º 6.093.695; patente de EEUU n.º 5.942.664; patente de EEUU n.º 6.110.464), resistencia a enfermedades fúngicas (patente de EEUU n.º 5.516.671; patente de EEUU n.º 5.773.696; patente de EEUU n.º 6.121.436; y patente de EEUU n.º 6.316.407; patente de EEUU n.º 6.506.962), resistencia a virus (patente de EEUU n.º 5.304.730; y patente de EEUU n.º 6.013.864), resistencia a nemátodos (patente de EEUU n.º 6.228.992), resistencia a enfermedades bacterianas (patente de EEUU n.º 5.516.671), producción de almidón (patente de EEUU n.º 5.750.876; y patente de EEUU n.º 6.476.295), producción de aceites modificados (patente de EEUU n.º 6.444.876), mayor producción de aceite (patente de EEUU n.º 5.608.149; y patente de EEUU n.º 6.476.295), contenido modificado en ácidos grasos (patente de EEUU n.º 6.537.750), alta producción de proteínas (patente de EEUU n.º 6.380.466), maduración de la fruta (patente de EEUU n.º 5.512.466), potenciación de la nutrición animal y humana (patente de EEUU n.º 5.985.605; y patente de EEUU n.º 6.171.640), biopolímeros (patente de EEUU n.º 5.958.745; y publicación de patente de EEUU n.º US20030028917), resistencia a estreses ambientales (patente de EEUU n.º 6.072.103), péptidos farmacéuticos (patente de EEUU n.º 6.080.560), mejores rasgos de procesamiento (patente de EEUU n.º 6.476.295), mejor digestibilidad (patente de EEUU n.º 6.531.648), bajo contenido en rafinosa (patente de EEUU n.º 6.166.292), producción de enzimas industriales (patente de EEUU n.º 5.543.576), mejor sabor (patente de EEUU n.º 6.011.199), fijación de nitrógeno (patente de EEUU n.º 5.229.114), producción de semillas híbridas (patente de EEUU n.º 5.689.041), y producción de biocombustible (patente de EEUU n.º 5.998.700).

Los herbicidas para los cuales se ha demostrado tolerancia en plantas transgénicas y el procedimiento de la presente invención pueden aplicarse incluyen, pero no se limitan a glifosato, glufosinato, sulfonilureas, imidazolinonas, bromoxinilo, delapón, ciclohezandiona, inhibidores de protoporfirionógeno oxidasa, y herbicidas de isoxasflutol. Las moléculas polinucleotídicas implicadas en la tolerancia a herbicidas son conocidas en la técnica e incluyen, pero no se limitan a una molécula polinucleotídica que codifica la 5-enolpiruvilshikimato-3-fosfato sintasa (EPSPS, descrita en las patentes de EEUU n.º 5.627.061, 5.633.435, 6.040.497; Padgett *et al.*, *Herbicide Resistant Crops*, Lewis Publishers, 53-85, 1996; y Penaloza-Vázquez, *et al.*, *Plant Cell Reports*, 14:482-487, 1995); y *aroA* (patente de EEUU n.º 5.094.945) para la tolerancia al glifosato; bromoxinilo nitrilasa (*Bxn*) para la tolerancia al bromoxinilo (patente de EEUU n.º 4.810.648); fitoeno desaturasa (*crtl*, Misawa *et al.* (1993), *Plant J.*, 4:833-840, y (1994) *Plant J.*, 6:481-489) para la tolerancia al norflurazón; acetohidroxiácido sintasa (AHAS, *aka* ALS, Sathasiivan *et al.*, *Nucl. Acids Res.*, 18:2188-2193, 1990); y el gen *bar* para la tolerancia al glufosinato y bialafós (DeBlock, *et al.*, *EMBO J.*, 6:2513-2519, 1987).

La tolerancia a herbicidas es un fenotipo deseable para plantas de cultivo. La N-fosfonometilglicina, también conocida como glifosato, es un herbicida conocidos que tiene actividad frente a un amplio espectro de especies vegetales. El glifosato es el ingrediente activo de Roundup® (Monsanto Co.), un herbicida seguro que tienen una semivida deseablemente corta en el entorno. Cuando se aplica a una superficie vegetal, el glifosato se mueve de modo sistémico a través de la planta. El glifosato es tóxico para las plantas porque inhibe la vía del ácido shikímico, que proporciona un precursor para la síntesis de aminoácidos aromáticos. De forma específica, el glifosato afecta a la conversión de fosfoenolpiruvato y ácido 3-fosfoshikímico al ácido 5-enolpiruvil-3-fosfoshikímico inhibiendo la enzima 5-enolpiruvil-3-fosfoshikimato sintasa (denominada en lo sucesivo en la presente EPSP sintasa o EPSPS). Para los fines de la presente invención, el término "glifosato" debe considerarse que incluye cualquier forma eficaz como herbicida de la N-fosfonometilglicina (incluyendo cualquiera de sus sales) y otras formas que dan como resultado la producción del anión glifosato en plantas.

A través de procedimientos de ingeniería genética es posible producir plantas tolerantes al glifosato insertando en

el genoma de la planta una molécula de ADN que provoca la producción de mayores niveles de EPSPS de tipo salvaje (Shah *et al.*, Science, 233:478-481, 1986). También puede conseguirse la tolerancia al glifosato mediante la expresión de variantes de EPSPS que tienen una menor afinidad por el glifosato y, por tanto, mantienen su actividad catalítica en presencia de glifosato (patente de EEUU n.º 5.633.435). Las enzimas que degradan el glifosato en los tejidos vegetales (patente de EEUU n.º 5.463.175) también son capaces de conferir tolerancia celular al glifosato. Por tanto, estos genes permiten la producción de cultivos transgénicos que son tolerantes al glifosato, permitiendo con ello utilizar el glifosato para un control eficaz de las hierbas adventicias con problemas mínimos de daños para el cultivo. Por ejemplo, la tolerancia al glifosato se ha introducido genéticamente en el maíz (patente de EEUU n.º 5.554.798, 6.040.497), trigo (Zhou *et al.*, Plant Cell Rep., 15:159-163,1995), soja (documento WO 9200377) y canola (documento WO 9204449).

Se han aislado variantes de la enzima EPSPS de tipo salvaje que son resistentes al glifosato como resultado de alteraciones en la secuencia codificadora de aminoácidos de EPSPS (Kishore *et al.*, Annu. Rev. Biochem., 57:627-663,1988; Schulz *et al.*, Arch. Microbiol., 137:121-123, 1984; Sost *et al.*, FEBS Lett., 173:238-241, 1984; Kishore *et al.*, en "Biotechnology for Crop Protection", ACS Symposium Series n.º 379, eds. Hedlin *et al.*, 37-48,1988). Estas variantes generalmente tiene una  $K_i$  mayor para el glifosato que la enzima EPSPS de tipo salvaje que confiere el fenotipo de tolerancia al glifosato, pero estas variantes también se caracterizan por una alta  $K_m$  para PEP que hace que la enzima sea cinéticamente menos eficaz. Por ejemplo, la  $K_m$  aparente para PEP y la  $K_i$  aparente para el glifosato para la EPSPS nativa de *E. coli* son 10  $\mu\text{M}$  y 0,5  $\mu\text{M}$ , mientras que para un aislado resistente al glifosato que tiene una sustitución de un único aminoácido de alanina a glicina en la posición 96, estos valores son de 220  $\mu\text{M}$  y 4,0 mM, respectivamente. La patente de EEUU n.º 6.040.497 divulga que la mutación conocida como la mutación TIPS (una sustitución de isoleucina por treonina en la posición del aminoácido 102, y una sustitución de serina por prolina en la posición del aminoácido 106) comprende dos mutaciones que, cuando se introducen en la secuencia polipeptídica de EPSPS de *Zea mays* confieren a la enzima resistencia al glifosato. Las plantas transgénicas que contienen esta enzima mutante son tolerantes al glifosato. Pueden realizarse mutaciones idénticas en enzimas EPSPS sensibles al glifosato procedentes de otras fuentes vegetales para crear enzimas resistencias al glifosato.

Se ha expresado una diversidad de enzimas EPSPS nativas y variantes en plantas transgénicas para conferir tolerancia al glifosato (Singh, *et al.*, en "Biosynthesis and Molecular Regulation of Amino Acids in Plants", Amer. Soc. Plant Phys. Pubs., 1992). Los ejemplos de algunas de estas EPSPS incluyen las descritas y/o aisladas según la patente de EEUU n.º 4.940.835, patente de EEUU n.º 4.971.908, patente de EEUU n.º 5.145.783, patente de EEUU n.º 5.188.642, patente de EEUU n.º 5.310.667, y patente de EEUU n.º 5.312.910. También puede derivarse de una clase estructuralmente diferenciada de genes EPSPS no homólogos, tal como los genes EPSPS de clase II aislados a partir de *Agrobacterium* sp. cepa CP4, según se describe en la patente de EEUU n.º 5.633.435, y la patente de EEUU n.º 5.627.061. Se modifican péptidos de tránsito del cloroplasto (CTP) para que se unan al N-terminal de la EPSPS bacteriana para dirigir a las enzimas resistentes al glifosato hacia el cloroplasto de la planta. En la EPSPS vegetal nativa, las regiones del péptido de tránsito del cloroplasto están dentro de la secuencia codificadora nativa (por ejemplo, CTP2, Klee *et al.*, Mol. Gen. Genet. 210:47-442, 1987). El CTP nativo puede sustituirse por un CTP heterólogo durante la construcción de un módulo de expresión en plantas de transgenes. Muchas proteínas localizadas en los cloroplastos, incluyendo EPSPS, se expresan a partir de genes nucleares en forma de precursores, y son dirigidas hacia el cloroplasto por un péptido de tránsito del cloroplasto (CTP) que se retira durante las etapas de importación. Los ejemplos de otras proteínas de cloroplastos de este tipo incluyen la subunidad pequeña (SSU) de la ribulosa-1,5-bisfosfato carboxilasa, ferredoxina, ferredoxina oxidoreductasa, la proteína I y la proteína II del complejo recolector de luz, y tiorredoxina F. Se ha demostrado *in vivo* e *in vitro* que las proteínas que no son de cloroplastos puede dirigirse a los cloroplastos mediante el uso de fusiones de proteínas con un CTP, y que una secuencia de CTP es suficiente para dirigir una proteína al cloroplasto. La incorporación de un péptido de tránsito del cloroplasto adecuado, tal como el CTP de EPSPS de *Arabidopsis thaliana* (Klee *et al.*, Mol. Gen. Genet., 210:437-442 (1987)), y el CTP de EPSPS de *Petunia hybrida* (Della-Cioppa *et al.*, Proc. Natl. Acad. Sci. USA, 83:6873-6877 (1986)) ha demostrado poder dirigir secuencias de proteínas de EPSPS heterólogas a los cloroplastos en plantas transgénicas. La producción de plantas tolerantes al glifosato mediante la expresión de una proteína de fusión que comprende un CTP amino-terminal con una enzima EPSPS resistente al glifosato es muy conocida por los expertos en la técnica (patente de EEUU n.º 5.627.061, patente de EEUU n.º 5.633.435, patente de EEUU n.º 5.312.910, documento EP 0218571, documento EP 189707, documento EP 508909, y documento EP 924299). Los expertos en la técnica reconocerán que puede fabricarse diversas construcciones quiméricas que emplean la funcionalidad de un CTP concreto para importar enzimas EPSPS resistentes al glifosato hacia el interior del cloroplasto de la célula vegetal.

Pueden realizarse modificaciones y cambios en la estructura de los polinucleótidos de la invención y seguir obteniendo una molécula que codifica una proteína o péptido funcional con características deseables. A continuación se indica un procedimiento basado en la sustitución de un codón o codones de un primer polinucleótido para crear un polinucleótido artificial de segunda generación equivalente, o incluso mejorado, en el

5 que este nuevo polinucleótido artificial es útil en los procedimientos de apilamiento de genes transgénicos y para potenciar la expresión. Se contempla que las sustituciones de codones en el polinucleótido de segunda generación, en ciertos casos, produzcan al menos un aminoácido diferente con respecto al primer polinucleótido. La sustitución del aminoácido puede proporcionar mejores características a la proteína, por ejemplo, una EPSP sintasa resistente al glifosato, o puede ser un cambio conservado que no afecta sustancialmente a las características de la proteína. El procedimiento proporciona un polinucleótido artificial creado mediante la retrotraducción de una secuencia polipeptídica en un polinucleótido empleando una tabla de utilización de codones, seguido de las etapas de potenciar características del polipéptido artificial que hacen que sea particularmente útil en plantas transgénicas.

10 En particular, se contempla que los polipéptidos modificados que codifican proteínas resistentes a herbicidas sean útiles para al menos uno de lo siguiente: para conferir tolerancia a herbicidas en una planta transformada o transgénica, para mejorar la expresión de genes de resistencia a herbicidas en plantas, para su uso como marcadores seleccionables para la introducción de otros rasgos de interés en una planta, y para evitar la recombinación con un gen de la planta endógeno similar o con transgenes existentes, lo cual permite el apilamiento de genes sin el silenciamiento de genes.

15 Se sabe que el código genético está degenerado. A continuación, en la tabla 1 se listan los aminoácidos y su codón o codones de ARN.

Tabla 1. Aminoácidos y los codones de ARN que los codifican

Aminoácido Nombre completo; código de 3 letras; código de 1 letra	Codones
Alanina; Ala; A	GCA GCC GCG GCU
Cisteína; Cys; C	UGC UGU
Ácido aspártico; Asp; D	GAC GAU
Ácido glutámico; Glu; E	GAA GAG
Fenilalanina; Phe; F	UUC UUU
Glicina; Gly; G	G GGA GGC GGG GGU
Histidina; His; H	CAC CAU
Isoleucina; Ile; I	AUA AUC AUU
Lisina; Lys; K	AAA AAG
Leucina; Leu; L	UUA UUG CUA CUC CUG CUU
Metionina; Met; M	AUG
Asparagina; Asn; N	AAC AAU
Prolina; Pro; P	CCA CCC CCG CCU
Glutamina; Gln; Q	CAA CAG
Arginina; Arg; R	AGA AGG CGA CGC CGG CGU
Serina; Ser; S	AGC AGU UCA UCC UCG UCU
Treonina; Thr; T	ACA ACC ACG ACU
Valina; Val; V	GUA GUC GUG GUU
Triptófano; Trp; W	UGG
Tirosina; Tyr; Y	UAC UAU

Los codones se describen en términos de bases de ARN, por ejemplo, adenina, uracilo, guanina y citosina. El ARNm es lo que se transcribe directamente en polipéptidos. Debe entenderse que cuando se diseña un polinucleótido de ADN para su uso en una construcción, las bases del ADN estarán sustituidas, por ejemplo, timina en lugar de uracilo.

5 Resulta deseable proporcionar plantas transgénicas que tengan múltiples fenotipos mejorados desde el punto de vista agronómico. A menudo se utiliza la tolerancia a herbicidas como marcador seleccionable para ayudar a la producción de plantas transgénicas que puedan poseer otros genes de importancia agronómica. El apilamiento de los transgenes mediante procedimientos de cultivo tradicionales o mediante retransformación de una primera planta transgénica con otro modulo de expresión en plantas puede incluir la introducción de genes o elementos genéticos que tienen una secuencia de polinucleótidos idéntica o casi idéntica. La progenie que contiene estos genes apilados puede ser susceptible a la pérdida de expresión génica debido al silenciamiento de genes. El procedimiento proporciona una molécula polinucleotídica modificada que codifica una proteína resistente a herbicidas. Las moléculas polinucleotídicas se diseñan para que sean suficientemente divergentes en la secuencia polinucleotídica con respecto a otras moléculas polinucleotídicas que codifican la misma proteína de resistencia a herbicidas. Entonces, estas moléculas pueden coexistir en la misma célula vegetal sin el problema del silenciamiento de genes.

La secuencia de polinucleótidos divergente se crea utilizando una tabla de utilización de codones construida a partir de secuencias codificadoras conocidas de diversas especies vegetales. Por ejemplo, pueden emplearse las tablas de utilización de codones para *Arabidopsis thaliana*, *Zea mays* y *Glycine max* en el procedimiento para diseñar los polinucleótidos. Los expertos en la técnica también pueden emplear otras tablas de utilización de codones de otras plantas.

La primera etapa en el procedimiento para diseñar una nueva molécula polinucleotídica artificial que codifica una proteína de tolerancia a herbicidas es el uso de una tabla de utilización de codones para determinar el porcentaje de utilización de codones en una especie vegetal para cada aminoácido de la proteína de tolerancia a herbicidas, seguido de la sustitución de al menos uno de cada ocho codones contiguos con un codón diferente seleccionado de la tabla de utilización de codones, y el ajuste del porcentaje de utilización de codones para cada aminoácido codificado por el polinucleótido hasta sustancialmente el mismo porcentaje de utilización de codones que se encuentra en la tabla de utilización de codones. Otras etapas pueden incluir introducir un codón de fin de la traducción en el segundo y tercer marco de lectura abierto de la nueva secuencia polinucleotídica; eliminar algunos codones de inicio de la traducción en el segundo y tercer marco de lectura abierto; ajustar la proporción de GC:AT local a aproximadamente 2:1 a lo largo de un intervalo de 50 nucleótidos; alterar las señales de poliadenilación potenciales o los sitios de ruptura de intrones potenciales; eliminar al menos un sitio de enzimas de restricción de seis nucleótidos contiguos más; y comparar la coincidencia de secuencia del nuevo polinucleótido artificial con un polinucleótido existente que codifica la misma proteína o una proteína similar, de forma que la coincidencia de secuencia entre los dos polinucleótidos no sea mayor que 85%.

Puede realizarse una retrotraducción de una secuencia de proteína a una secuencia de nucleótidos empleando una tabla de utilización de codones, tal como la que se encuentra en Genetics Computer Group (GCG) SeqLab u otros programas de análisis de ADN conocidos por los expertos en la técnica del análisis de ADN, o según se proporciona en las tablas 2, 3 y 4 de la presente invención. La tabla de utilización de codones para *Arabidopsis thaliana* (tabla 2), *Zea mays* (tabla 3) y *Glycine max* (tabla 4) son ejemplos de tablas que pueden construirse para especies vegetales. También pueden construirse tablas de utilización de codones que representen la utilización de codones de monocotiledóneas o dicotiledóneas.

Tabla 2. Tabla de utilización de codones de *Arabidopsis thaliana*

Aminoácido	Codón	Número	/1000	Fracción
Gly	GGG	188335,00	10,18	0,16
Gly	GGA	443469,00	23,98	0,37
Gly	GGT	409478,00	22,14	0,34
Gly	GGC	167099,00	9,03	0,14
Glu	GAG	596506,00	32,25	0,48

ES 2 420 929 T3

Glu	GAA	639579,00	34,58	0,52
Asp	GAT	683652,00	36,96	0,68
Asp	GAC	318211,00	17,20	0,32
Val	GTG	320636,00	17,34	0,26
Val	GTA	185889,00	10,05	0,15
Val	GTT	505487,00	27,33	0,41
Val	GTC	235004,00	12,71	0,19
Ala	GCG	162272,00	8,77	0,14
Ala	GCA	323871,00	17,51	0,27
Ala	GCT	521181,00	28,18	0,44
Ala	GCC	189049,00	10,22	0,16
Arg	AGG	202204,00	10,93	0,20
Arg	AGA	348508,00	18,84	0,35
Ser	AGT	260896,00	14,11	0,16
Ser	AGC	206774,00	11,18	0,13
Lys	AAG	605882,00	32,76	0,51
Lys	AAA	573121,00	30,99	0,49
Asn	AAT	418805,00	22,64	0,52
Asn	AAC	385650,00	20,85	0,48
Met	ATG	452482,00	24,46	1,00
Ile	ATA	235528,00	12,73	0,24
Ile	ATT	404070,00	21,85	0,41
Ile	ATC	341584,00	18,47	0,35
Thr	ACG	140880,00	7,62	0,15
Thr	ACA	291436,00	15,76	0,31
Thr	ACT	326366,00	17,65	0,34
Thr	ACC	190135,00	10,28	0,20



ES 2 420 929 T3

Trp	TGG	231618,00	12,52	1,00
End	TGA	19037,00	1,03	0,43
Cys	TGT	196601,00	10,63	0,60
Cys	TGC	131390,00	7,10	0,40
End	TAG	9034,00	0,49	0,20
End	TAA	16317,00	0,88	0,37
Tyr	TAT	276714,00	14,96	0,52
Tyr	TAC	254890,00	13,78	0,48
Leu	TTG	389368,00	21,05	0,22
Leu	TTA	237547,00	12,84	0,14
Phe	TTT	410976,00	22,22	0,52
Phe	TTC	380505,00	20,57	0,48
Ser	TCG	167804,00	9,07	0,10
Ser	TCA	334881,00	18,11	0,20
Ser	TCT	461774,00	24,97	0,28
Ser	TCC	203174,00	10,99	0,12
Arg	CGG	88712,00	4,80	0,09
Arg	CGA	115857,00	6,26	0,12
Arg	CGT	165276,00	8,94	0,17
Arg	CGC	69006,00	3,73	0,07
Gln	CAG	280077,00	15,14	0,44
Gln	CAA	359922,00	19,46	0,56
His	CAT	256758,00	13,88	0,62
His	CAC	160485,00	8,68	0,38
Leu	CTG	183128,00	9,90	0,11
Leu	CTA	184587,00	9,98	0,11
Leu	CTT	447606,00	24,20	0,26
Leu	CTC	294275,00	15,91	0,17

ES 2 420 929 T3

Pro	CCG	155222,00	8,39	0,17
Pro	CCA	298880,00	16,16	0,33
Pro	CCT	342406,00	18,51	0,38
Pro	CCC	97639,00	5,28	0,11

5

Tabla 3. Tabla de utilización de codones de *Zea mays*

Aminoácido	Codón	Número	/1000	Fracción
Gly	GGG	8069,00	15,19	0,21
Gly	GGA	7100,00	13,37	0,18
Gly	GGT	7871,00	14,82	0,20
Gly	GGC	15904,00	29,94	0,41
Glu	GAG	22129,00	41,67	0,68
Glu	GAA	10298,00	19,39	0,32
Asp	GAT	11996,00	22,59	0,41
Asp	GAC	17045,00	32,09	0,59
Val	GTG	13873,00	26,12	0,38
Val	GTA	3230,00	6,08	0,09
Val	GTT	8261,00	15,55	0,23
Val	GTC	11330,00	21,33	0,31
Ala	GCG	11778,00	22,18	0,24
Ala	GCA	8640,00	16,27	0,18
Ala	GCT	11940,00	22,48	0,24
Ala	GCC	16768,00	31,57	0,34
Arg	AGG	7937,00	14,94	0,27

ES 2 420 929 T3

Arg	AGA	4356,00	8,20	0,15
Ser	AGT	3877,00	7,30	0,10
Ser	AGC	8653,00	16,29	0,23
Lys	AAG	22367,00	42,11	0,74
Lys	AAA	7708,00	14,51	0,26
Asn	AAT	6997,00	13,17	0,36
Asn	AAC	12236,00	23,04	0,64
Met	ATG	12841,00	24,18	1,00
Ile	ATA	3997,00	7,53	0,16
Ile	ATT	7457,00	14,04	0,31
Ile	ATC	12925,00	24,34	0,53
Thr	ACG	5665,00	10,67	0,22
Thr	ACA	5408,00	10,18	0,21
Thr	ACT	5774,00	10,87	0,22
Thr	ACC	9256,00	17,43	0,35
Trp	TGG	6695,00	12,61	1,00
End	TGA	591,00	1,11	0,45
Cys	TGT	2762,00	5,20	0,30
Cys	TGC	6378,00	12,01	0,70
End	TAG	411,00	0,77	0,32
End	TAA	299,00	0,56	0,23
Tyr	TAT	4822,00	9,08	0,31
Tyr	TAC	10546,00	19,86	0,69
Leu	TTG	6677,00	12,57	0,14
Leu	TTA	2784,00	5,24	0,06
Phe	TTT	6316,00	11,89	0,32
Phe	TTC	13362,00	25,16	0,68

ES 2 420 929 T3

Ser	TCG	5556,00	10,46	0,14
Ser	TCA	5569,00	10,49	0,15
Ser	TCT	6149,00	11,58	0,16
Ser	TCC	8589,00	16,17	0,22
Arg	CGG	4746,00	8,94	0,16
Arg	CGA	2195,00	4,13	0,07
Arg	CGT	3113,00	5,86	0,10
Arg	CGC	7374,00	13,88	0,25
Gln	CAG	13284,00	25,01	0,64
Gln	CAA	7632,00	14,37	0,36
His	CAT	5003,00	9,42	0,39
His	CAC	7669,00	14,44	0,61
Leu	CTG	13327,00	25,09	0,28
Leu	CTA	3785,00	7,13	0,08
Leu	CTT	8238,00	15,51	0,17
Leu	CTC	12942,00	24,37	0,27
Pro	CCG	8274,00	15,58	0,27
Pro	CCA	7845,00	14,77	0,26
Pro	CCT	7129,00	13,42	0,23
Pro	CCC	7364,00	13,87	0,24

Tabla 4. Tabla de utilización de codones de *Glycine max*

Aminoácido	Codón	Número	/1000	Fracción
Gly	GGG	3097,00	12,82	0,18
Gly	GGA	5434,00	22,49	0,32
Gly	GGT	5248,00	21,72	0,31
Gly	GGC	3339,00	13,82	0,20
Glu	GAG	8296,00	34,33	0,50
Glu	GAA	8194,00	33,91	0,50

ES 2 420 929 T3

Asp	GAT	7955,00	32,92	0,62
Asp	GAC	4931,00	20,40	0,38
Val	GTG	5342,00	22,11	0,32
Val	GTA	1768,00	7,32	0,11
Val	GTT	6455,00	26,71	0,39
Val	GTC	2971,00	12,29	0,18
Ala	GCG	1470,00	6,08	0,08
Ala	GCA	5421,00	22,43	0,31
Ala	GCT	6796,00	28,12	0,38
Ala	GCC	4042,00	16,73	0,23
Arg	AGG	3218,00	13,32	0,28
Arg	AGA	3459,00	14,31	0,30
Ser	AGT	2935,00	12,15	0,17
Ser	AGC	2640,00	10,92	0,15
Lys	AAG	9052,00	37,46	0,59
Lys	AAA	6370,00	26,36	0,41
Asn	AAT	5132,00	21,24	0,48
Asn	AAC	5524,00	22,86	0,52
Met	ATG	5404,00	22,36	1,00
Ile	ATA	3086,00	12,77	0,23
Ile	ATT	6275,00	25,97	0,47
Ile	ATC	3981,00	16,47	0,30
Thr	ACG	1006,00	4,16	0,08
Thr	ACA	3601,00	14,90	0,29
Thr	ACT	4231,00	17,51	0,34
Thr	ACC	3562,00	14,74	0,29
Trp	TGG	2866,00	11,86	1,00

ES 2 420 929 T3

End	TGA	221,00	0,91	0,36
Cys	TGT	1748,00	7,23	0,49
Cys	TGC	1821,00	7,54	0,51
End	TAG	143,00	0,59	0,23
End	TAA	256,00	1,06	0,41
Tyr	TAT	3808,00	15,76	0,51
Tyr	TAC	3667,00	15,17	0,49
Leu	TTG	5343,00	22,11	0,24
Leu	TTA	2030,00	8,40	0,09
Phe	TTT	4964,00	20,54	0,49
Phe	TTC	5067,00	20,97	0,51
Ser	TCG	1107,00	4,58	0,06
Ser	TCA	3590,00	14,86	0,21
Ser	TCT	4238,00	17,54	0,24
Ser	TCC	2949,00	12,20	0,17
Arg	CGG	683,00	2,83	0,06
Arg	CGA	964,00	3,99	0,08
Arg	CGT	1697,00	7,02	0,15
Arg	CGC	1538,00	6,36	0,13
Gln	CAG	4147,00	17,16	0,46
Gln	CAA	4964,00	20,54	0,54
His	CAT	3254,00	13,47	0,55
His	CAC	2630,00	10,88	0,45
Leu	CTG	2900,00	12,00	0,13
Leu	CTA	1962,00	8,12	0,09
Leu	CTT	5676,00	23,49	0,26
Leu	CTC	4053,00	16,77	0,18

Pro	CCG	1022,00	4,23	0,08
Pro	CCA	4875,00	20,17	0,37
Pro	CCT	4794,00	19,84	0,36
Pro	CCC	2445,00	10,12	0,19

Las tablas de utilización de codones son muy conocidas en la técnica y pueden encontrarse en bases de datos de genes, por ejemplo, la base de datos de Genbank. La base de datos de utilización de codones es un versión en la red extendida de CUTG (Codon Usage Tabulated from Genbank). La frecuencia de utilización de codones en cada organismo puede buscarse en este sitio de la red (Nakamura *et al.*, Nucleic Acids Res., 28:292, 2000). Las etapas pueden realizarse en cualquier orden o de modo simultáneo. Pueden realizarse cualquiera o todas las etapas en el diseño de un polinucleótido artificial. Cada etapa se describe en detalle a continuación.

Los diferentes codones para un aminoácido deben distribuirse a través del polinucleótido basándose en el porcentaje aproximado de utilización de codones para la especie concreta a partir de una tabla de utilización de codones. Debe evitarse la acumulación local de codones idénticos. Al menos un codón es sustituido por cada ocho codones contiguos para proporcionar una divergencia suficiente de secuencias de polinucleótidos que codifican proteínas idénticas o similares. Excepto cuando se desee específicamente, por ejemplo, para proporcionar una enzima tolerante a herbicidas, la proteína codificada permanece sin cambios cuando se sustituye un codón por otro codón que se traduce en el mismo aminoácido, según se lista en la tabla 1.

Se realizan correcciones en la proporción de GC:AT local de un polinucleótido ajustando la proporción GC:AT local a aproximadamente la misma proporción que el polinucleótido de longitud completa, pero no más de 2x a lo largo de un intervalo de aproximadamente 50 nucleótidos contiguos de la molécula polinucleotídica. El intervalo de proporciones GC:AT de un polinucleótido utilizando tablas de utilización de codones de plantas dicotiledóneas debería ser de aproximadamente 0,9 a aproximadamente 1,3, y para plantas monocotiledóneas de aproximadamente 1,2 a aproximadamente 1,7. La proporción GC:AT local puede ser importante para el mantenimiento de una estructura secundaria apropiada del ARN. Se predice que las regiones que comprenden muchas bases A+T o bases G+C consecutivas tienen una mayor probabilidad para formar estructuras de horquilla debido a la autocomplementariedad. Por tanto, la sustitución por un codón diferente reduciría la probabilidad de la formación de una estructura secundaria autocomplementaria, que se sabe que reduce la transcripción y/o la traducción en algunos organismos. En la mayoría de los casos, los efectos adversos pueden minimizarse utilizando moléculas polinucleotídicas que no contienen más de cinco A+T o G+C consecutivas. La longitud máxima del tramo de GC local (sin nucleótidos AT) no debería ser mayor que 10 nucleótidos. Por tanto, los codones que codifican proteínas ricas en Gly, Ala, Arg, Ser y Pro pueden ser sustituidos para evitar acumulaciones largas de nucleótidos GC. Los codones ricos en GC listados pueden utilizarse en combinación con los codones ricos en AT para los aminoácidos Lys, Asn, Ile, Tyr, Leu, Phe y viceversa para corregir la proporción GC:AT local.

Puede realizarse una comprobación de la coincidencia de secuencia utilizando herramientas de alineamiento de secuencias de nucleótidos, tales como el programa GAP (GCG, Madison, WI) inmediatamente después de una retrotraducción para asegurarse de que la secuencia generada tiene el grado apropiado de diversidad de secuencia. Las secuencias de polinucleótidos contiguas con una longitud mayor que 23 nucleótidos que tengan 100% de coincidencia de secuencia deben eliminarse realizando sustituciones de aminoácidos en estos tramos de la secuencia.

Los codones de inicio de la traducción (ATG para ADN, AUG para ARNm) están presentes en el segundo marco de lectura (marco "b"), el tercer marco de lectura (marco "c"), y los marcos de lectura inversos (marco "d", "e", "f"). Los codones de inicio del segundo y tercer marco pueden iniciar la traducción, aunque con mucha menos eficacia que el primero. Por tanto, si en el marco "b" o "c" se encuentran uno o dos AUG cerca del extremo 5' de una molécula de ARNm, resultaría beneficioso eliminarlos de la región del polinucleótido que contiene al menos los primeros tres codones Met en el marco "a". Además, si la secuencia de proteína no tiene más de un Met en el marco "a", entonces se eliminarán las más posibles de los marcos directos "b" o "c". Para realizar esto, por ejemplo, los codones para los aminoácidos Asp, Asn, Tyr, e His en la proteína de interés, seguidos de cualquiera de los aminoácidos Gly, Glu, Asp, Val o Ala, pueden ser sustituidos para eliminar un codón de inicio en el segundo marco. La secuencia GATGGG codifica los aminoácidos Asp-Gly y forma un ATG en el marco de lectura "b". Cuando la secuencia se modifica a GACGGG, el inicio ATG se elimina y la secuencia sigue codificando Asp-Gly. Se emplea una estrategia similar para eliminar los codones de inicio en el marco de lectura "c". La combinación de un aminoácido seleccionado del grupo de Gly, Glu, Val, Ala, Arg, Lys, Ile, Thr, Cys, Tyr, Leu, Ser, His o Pro, seguido de Trp, puede dar como resultado la formación de ATG en el tercer marco de lectura del gen. En esta

situación, el primer codón puede cambiarse para que tenga un nucleótido distinto de A en la tercera posición.

La eliminación de un codón ATG en la hebra de ADN complementaria del gen en marcos alternos ("d", "e" y/o "f") sin cambiar la secuencia de aminoácidos de la proteína puede lograrse de la misma manera. Esta modificación reduce la probabilidad de traducción aunque el transgén se integre en un genoma vegetal en una orientación que permita la transcripción del ARNm complementario inverso a partir de un promotor de la planta nativo. La traducción a partir de cualquier marco de lectura inverso puede minimizarse mediante la introducción de un codón de fin en los tres marcos de lectura inversos según se describe a continuación.

La creación de codones de fin para los tres marcos del ADN complementario puede realizarse como sigue. El codón Leu (TTA y CTA) y Ser (TCA) produce tres codones de fin diferentes en la hebra complementaria inversa. Si estos aminoácidos están presentes en el C-terminal de la proteína de interés, sus codones pueden utilizarse para generar fines en la hebra complementaria en el marco de lectura "d". Para generar un codón de fin en el marco de lectura "e" de la hebra complementaria, deben estar presentes los aminoácidos Ala, Arg, Asn, Asp, Cys, Gly, His, Ile, Leu, Phe, Pro, Ser, Thr, Tyr o Val, seguidos de los aminoácidos Gln, His o Tyr en la proteína de interés. Por ejemplo, la secuencia polinucleotídica de GCCCAC que codifica los aminoácidos Ala-His puede modificarse a GCTCAC. La secuencia complementaria, GTGAGC, ahora tendrá un codón de fin TGA que se muestra en cursiva. Cuando la proteína de interés tiene Ala, Ile, Leu, Phe, Pro, Ser, Thr o Val, seguidos de Arg, Asn, Ile, Lys, Met, o Ser, el marco de lectura en la hebra complementaria puede modificarse para que tenga un codón de fin en el marco de lectura "f" de la hebra complementaria. La secuencia polinucleotídica ATATCT para Ile y Ser puede modificarse a ATCAGT para generar un codón de fin en la hebra complementaria, tal como se muestra en cursiva, ACTGAT. La combinación de codones para Phe, seguidos de cualquiera de los codones para los aminoácidos Asn, Ile, Lys, Met o Thr, siempre generará un codón de fin en el marco de la hebra complementaria "e" o "f".

Para crear un codón de fin en el marco de lectura directo "b", el marco de lectura "a" debe terminar con los nucleótidos TA o TG. Se buscan en la proteína de interés los aminoácidos Ile, Leu, Met o Val, en combinación con cualquiera de los siguientes aminoácidos: Ala, Arg, Asn, Asp, Glu, Gly, Ile, Lys, Met, Ser, Thr o Val. Por ejemplo, si la secuencia polinucleotídica que codifica los aminoácidos Met-Ser es ATGTCT, entonces puede modificarse a ATGAGT para producir un codón de fin TGA en el segundo marco de lectura.

Para crear un codón de fin para el marco de lectura "c", el marco de lectura "a" debe tener el nucleótido T en tercera posición y el siguiente codón debe comenzar con AA, AG o GA. Para encontrar los codones adecuados para ser modificados, se busca en la proteína de interés cualquiera de los aminoácidos: Ala, Asn, Asp, Arg, Cys, Gly, His, Ile, Leu, Phe, Pro, Ser, Thr, Tyr o Val, seguidos de cualquiera de los siguientes aminoácidos: Arg, Asn, Asp, Glu, Lys o Ser. Por ejemplo, si la secuencia de nucleótidos para los aminoácidos Gly-Glu es GGAGAG, la secuencia puede modificarse a GGTGAG para crear un codón de fin TGA en el tercer marco de lectura.

Otra modificación útil en los procedimientos de diseño de polinucleótidos artificiales es eliminar los sitios de restricción no deseados y otros patrones de secuencia específicos. Los sitios de restricción pueden interferir con la posterior clonación y manipulación del gen. Por ejemplo, algunos sitios de restricción que se emplean habitualmente en la clonación de genes incluyen, pero no se limitan a las enzimas de restricción de tipo II con 6 o más bases no-N listadas en la siguiente tabla 5, que es un extracto de la base de datos de endonucleasas de restricción de New England Biolabs, Inc. (Beverly, MA, EE.UU.). La búsqueda de sitios de reconocimiento de enzimas de restricción puede realizarse utilizando la aplicación funcional Map de GDG SeqLab u otra aplicación similar contenida en otros programas de análisis de ADN conocidos por los expertos en la técnica del análisis de ADN. Las enzimas de restricción también pueden añadirse a la secuencia para facilitar la clonación. Por ejemplo, el sitio de restricción ClaI está introducido en CP4EPSPS versión AT (SEQ ID NO:17) y ZM (SEQ ID NO:18) para generar secuencias recombinantes mediante intercambio de fragmentos y para facilitar la síntesis génica utilizando fragmentos de nucleótidos que pueden ensamblarse para formar el gen completo. La secuencia polinucleotídica del péptido de tránsito CTP2 (SEQ ID NO:12) está conectada con CP4EPSPS mediante el sitio de restricción SphI para facilitar la sustitución de CTP2 con diferentes versiones de nucleótidos de CTP2 (SEQ ID NO:13, SEQ ID NO:14) o polinucleótidos que codifican diferentes péptidos de tránsito del cloroplasto. Por ejemplo, en la EPSPS del arroz, el sitio de restricción NgaMIV se conserva en aproximadamente la posición del nucleótido 205 en todas las versiones artificiales para facilitar el intercambio de la región codificadora del péptido de tránsito del cloroplasto. Además, para la EPSPS de soja, la secuencia polinucleotídica para el péptido de tránsito del cloroplasto está separada del péptido maduro por el sitio de restricción para la endonucleasa SacII.

Se entiende que no es necesaria la modificación de los sitios de endonucleasas de restricción, sino que es útil para la posterior manipulación de las moléculas de ADN. La tabla 5 proporciona un listado de endonucleasas de restricción; las de interés particular están marcadas con asteriscos. Otros sitios de endonucleasas de restricción deseables para la eliminación o la adición de un polinucleótido artificial serán evidentes para los expertos en la técnica y no se limitan a los listados en la tabla 5.



ES 2 420 929 T3

Tabla 5. Secuencias de reconocimiento de enzimas de restricción

Enzimas	Secuencia de reconocimiento	Enzimas	Secuencia de reconocimiento
AatII	G_ACGT^C	*BclI	T^GATC_A
*AclI	GT^MK_AC	BetI	W^CCGG_W
Acc65I	G^GTAC_C	BfrBI	ATG^CAT
AceII	G_CTAG^C	*BglII	A^GATC_T
AclI	AA^CG_TT	BloHII	CTGCA^G
AcyI	GR^CG_YC	BlpI	GC^TNA_GC
AfeI	AGC^GCT	Bme1580I	G_KGCM^C
AfIII	C^TTAA_G	BmgI	GKGCCC
*AflIII	A^CRYG_T	Bpu10I	CC^TNA_GC
AgeI	A^CCGG_T	BsaI	GGTCTCN^NNNN_
AhaIII	TTT^AAA	BsaAI	YAC^GTR
ApaI	G_GGCC^C	BsaHI	GR^CG_YC
ApalI	G^TGCA_C	BsaWI	W^CCGG_W
ApoI	R^AATT_Y	Bsbl	CAACAC
AscI	GG^CGCG_CC	BsePI	G^CGCG_C
Asel	AT^TA_AT	BseSI	G_KGCM^C
AsiSI	GCG_AT^CGC	Bsil	C^ACGA_G
AsuII	TT^CG_AA	BsiEI	CG_RY^CG
AvaI	C^YCGR_G	BsiWI	C^GTAC_G
Avall	ATGCAT	BsmI	GAATG_CN^
AvrII	C^CTAG_G	Bsp1286I	G_DGCH^C
BalI	TGG^CCA	Bsp1407I	T^GTAC_A
*RamHI	G^GATC_C	BspEI	T^CCGG_A
BanI	G^GYRC_C	BspGI	CTGGAC
BanII	G_RGCY^C	3spHI	T^CATG_A
*BbeI	G_GCGC^C	3spLU11I	A^CATG_T
BbvCI	CC^TCA_GC	BspMII	T^CCGG_A
BsrBI	CCG^CTC	MscI	TGG^CCA
BsrDI	GCAATG_NN^	MspA1I	CMG^CKG
BsrFI	R^CCGG_Y	MstI	TGC^GCA
BsrGI	T^GTAC_A	NaeI	GCC^GGC
BssHII	G^CGCG_C	NarI	GG^CG_CC

ES 2 420 929 T3

BssSI	C^ACGA_G	*NcoI	C^CATG_G
BstAPI	GCAN_NNN^NTGC	*NdeI	CA^TA_TG
BstBI	TT^CG_AA	*NgoMIV	G^CCGG_C
BstEII	G^GTNAC_C	*NheI	G^CTAG_C
BstXI	CCAN_NNNN^NTGG	N1i38771	C_YCGR^G
BstYI	R^GATC_Y	*NotI	GC^GGCC_GC
BstZ17I	GTA^TAC	*NruI	TCG^CGA
Bsu36I	CC^TNA_GG	*NsiI	A_TGCA^T
BtgI	C^CRYG_G	NspI	R_CATG^Y
BtrI	CAC^GTC	NspBII	CMG^CKG
BtsI	GCAGTG_NN^	*PacI	TTA_AT^TAA
CfrI	Y^GGCC_R	*PciI	A^CATG_T
Cfr10I	R^CCGG_Y	Pf11108I	TCGTAG
*C1aI	AT^CG_AT	*PfiMI	CCAN_NNN^NTGG
DraI	TTT^AAA	PmaCI	CAC^GTG
Drall	RG^GNC_CY	PmeI	GTTT^AAAC
DrdII	GAACCA	PmII	CAC^GTG
Dsal	C^CRYG_G	Ppu10I	A^TGCA_T
EaeI	Y^GGCC_R	*PpuMI	RG^GWC_CY
EagI	C^GGCC_G	PshAI	GACNN^NNGTC
Ec1136II	GAG^CTC	PsiI	TTA^TAA
Eco47III	AGC^GCT	*PspOMI	G^GGCC_C
EcoNI	CCTNN^N_NNAGG	PssI	RG_GNC^CY
EcoO109I	RG^GNC_CY	*PstI	C_TGCA^G
*EcoRI	G^AATT_C	*PvuI	CG_AT^CG
*EcoRV	GAT^ATC	*PvuII	CAG^CTG
Espl	GC^TNA_GC	RsrII	CG^GWC_CC
*FseI	GG_CCGG^CC	*SacI	G_AGCT^C
*FspI	TGC^GCA	*SacII	CC_GC^GG
FspAI	RTGC^GCAY	*Sall	G^TCGA_C
GdIII	C^GGCC_R	SanDI	GG^GWC_CC
HaeI	WGG^CCW	SapI	GCTCTTCN^NNN_
HaeII	R_GCGC^Y	SauI	CC^TNA_GG
HgiAI	G_WGCW^C	SbfI	CC_TGCA^GG

ES 2 420 929 T3

HgiCI	G^GYRC_C	*Scal	AGT^ACT
HgiJII	G_RGCY^C	Scil	CTC^GAG
*HincII	GTY^RAC	Sdul	G_DGCH^C
HindII	GTY^RAC	SexAI	A^CCWGG_T
*HindIII	A^AGCT_T	Sfcl	C^TRYA_G
*HpaI	GTT^AAC	Sfel	C^TRYA_G
KasI	G^GCGC_C	Sfil	GGCCN_NNN^NGGCC
*KpnI	G_GTAC^C	Sfol	GGC^GCC
LpnI	RGC^GCY	Sgfl	GCG AT^CGC
McrI	CG_RY^CG	SgrAI	CR^CCGG_YG
MfeI	C^AATT_G	*SmaI	CCC^GGG
*MluI	A^CGCG_T	SmlI	C^TYRA_G
NotI	GTATAC		
*SnaBI	TAC^GTA		
*SpeI	A^CTAG_T		
*SphI	G_CATG^C		
SplI	C^GTAC_G		
SrfI	GCCC^GGGC		
Sse232I	CG^CCGG_CG		
Sse8387I	CC_TGCA^GG		
Sse8647I	AG^GWC_CT		
*SspI	AAT^ATT		
*StuI	AGG^CCT		
*StyI	C^CWWG_G		
*SwaI	ATTT^AAAT		
TatI	W^GTAC_W		
UbaMI	TCCNGGA		
UbaPI	CGAACG		
*VspI	AT^TA_AT		
*XbaI	T^CTAG_A		
*XhoI	C^TCGA_G		
XhoII	R^GATC_Y		
*XmaI	C^CCGG_G		
XmaIII	C^GGCC_G		

Xmnl	GAANN^NNTTC
Zral	GAC^GTC

Puede realizarse un búsqueda de patrón para encontrar secuencias desestabilizantes potenciales y sitios de polinadenilación, y después alterarlas o eliminarlas según se describe en la patente de EEUU n.º 5.500.365. Ciertos tramos largos de regiones ricas en AT, por ejemplo, el motivo de secuencia ATTTA (o AUUUA, según aparece en el ARN) se han implicado como secuencias desestabilizantes en ARNm de células de mamífero (Shaw y Kamen, Cell, 46:659-667, 1986). Muchos ARNm de vida corta tienen regiones no traducidas 3' ricas en A+T, y estas regiones a menudo presentan la secuencia ATTTA, que a veces está presente en múltiples copias o en forma de multímeros (por ejemplo, ATTTATTA...). Shaw y Kamen demuestran que la transferencia del extremo 3' de un ARNm inestable a un ARN estable (globina o VA1) disminuye la semivida del ARN estable de forma acusada. También demostraron que un pentámero de ATTTA tiene un profundo efecto desestabilizante sobre un mensaje estable, y que esta señal puede ejercer su efecto tanto si está localizada en el extremo 3' como dentro de la secuencia codificadora. Sin embargo, el número de secuencias ATTTA y/o el contexto de la secuencia en que aparecen también parecen ser importantes para determinar si actúan como secuencias desestabilizantes. También demostraron que un trímero de ATTTA tiene mucho menos efecto que un pentámero sobre la estabilidad del ARNm, y que un dímero o un monómero no tienen efecto sobre la estabilidad. Nótese que los multímeros de ATTTA, tales como un pentámero, crean automáticamente una región rica en A+T. En otros ARNm inestables, la secuencia ATTTA puede estar presente sólo en una única copia, pero a menudo está contenida en una región rica en A+T. Se ha demostrado que una repetición de 11 pentámeros AUUUA se dirige a transcripciones indicadoras para la degradación rápida en plantas (Ohme-Takagi *et al.*, Proc. Nat. Acad. Sci. USA, 90, 11811-11815, 1993). La secuencia ATTTA puede ser formada por la combinación de codones para los aminoácidos Ile (ATT) y Tyr (TAT), como en ATTTAT. Otro ejemplo pueden ser codones que terminan en AT, como en Asn, Asp, His o Tyr, seguidos de un codón TTA para Leu (por ejemplo, AATTTA). También el codón para Phe (UUU), cuando se coloca entre codones que terminan en A y comienzan en A, formará un motivo ATTTA. Para eliminar este motivo normalmente un único cambio de un nucleótido es suficiente, como en el ejemplo mostrado: GCATTTAGC cambia a GCATTCAGC o GCCTTTAGC. Los tres polinucleótidos mostrados codifican Ala-Phe-Arg.

Se han identificado más secuencias de acción en cis que se dirigen a transcripciones para un recambio rápido en plantas y en otros sistemas (Abler y Green, Plant Mol. Biol., 32:63-78, 1997). Estas incluyen el elemento DST que consiste en tres subdominios altamente conservados, separados por dos regiones variables que se encuentran cadena abajo del codón de fin de las transcripciones SAUR (Newmaan *et al.*, Plant Cell, 5: 701-714, 1993). La secuencia conservada DST consiste en GGAG(N<sub>5</sub>)CATAGATTG(N<sub>7</sub>)CATTTTGTAT, en la que los restos altamente conservados se muestran en cursiva. El segundo y tercer subdominio de los elementos DST contienen restos que no varían entre los elementos DST, y se denominan ATAGAT y GTA, respectivamente. Ambos subdominios son necesarios para la función DST. Se seleccionan nuevas secuencias polinucleotídicas artificiales para la presencia de los motivos conservados de los elementos DST GGAG, ATAGATT, CATTTT y CATTTTGTAT. Estas secuencias se eliminan mediante sustituciones de bases de codones que conservan la secuencia de la proteína codificada por el polinucleótido. Los motivos de secuencia de DST GGAG, ATAGAT, CATTTT y GTAT que aparecen en agrupaciones o patrones similares a la secuencia DST conservada también se eliminan mediante sustituciones de bases. Las secuencias de polinucleótidos que podrían actuar como sitios de poliadenilación son eliminadas en el diseño del nuevo polinucleótido (patente de EEUU n.º 5.500.365). Estas señales de poliadenilación pueden no actuar como sitios de poliA verdaderos, pero actúan como sitios aberrantes que producen ARNm inestables.

La adición de una cadena de poliadenilato al extremo 3' de un ARNm es común en la mayoría de los ARNm eucariotas. Contenidas dentro de esta transcripción de ARNm hay señales para la poliadenilación y la formación de un extremo 3' adecuado. Este procesamiento en el extremo 3' implica la ruptura del ARNm y la adición de poliA al extremo 3' maduro. La búsqueda de secuencias consenso cerca del tramo de poliA en ARNm vegetal y animal ha posibilitado identificar secuencias consenso que aparentemente están implicadas en la adición de poliA y la ruptura del extremo 3'. Las mismas secuencias consenso parecen ser importantes para ambos procedimientos. Estas señales generalmente son una variación de la secuencia AATAAA. En células animales, se han identificado algunos variantes de esta secuencia que son funcionales; en células vegetales parece haber una gama extensa de secuencias funcionales (Dean *et al.*, Nucl Acid Res., 14:2229-2240, 1986; Hunt, Annu Rev. Plant Physiol. Plant Mol. Biol., 45:47-60, 1994; Rothine, Plant Mol. Biol., 32:43-61, 1996). Todas estas secuencias consenso son variaciones de AATAA y, por tanto, todas son secuencias ricas en A+T.

Generalmente, para obtener una expresión suficiente de los transgenes modificados en plantas, se modifica la secuencia codificadora polinucleotídica estructural ("gen estructural") que codifica la proteína de interés, mediante la eliminación de las secuencias ATTTA y las señales de poliadenilación putativas mediante mutagénesis dirigida específica de sitio del ADN que comprende el gen estructural. Se eliminan sustancialmente todas las señales de

poliadenilación conocidas y secuencias ATTTA en el polinucleótido modificado, aunque a menudo se observan unos niveles de expresión potenciados solo con la eliminación de algunas de las secuencias de señal de poliadenilación identificadas anteriormente. Como alternativa, si se prepara un polinucleótido artificial que codifica la proteína concreta, los codones se seleccionan para evitar la secuencia ATTTA y las señales de poliadenilación putativas. Para los fines de la presente invención, las señales de poliadenilación putativas incluyen, pero no están necesariamente limitadas a AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAT, AATTA, AATACA y CATAAA.

La secuencia de ADN seleccionada se barre para identificar regiones con más de cuatro nucleótidos de adenina (A) o timina (T) consecutivos. Las regiones A+T se barren para detectar señales de poliadenilación de vegetales potenciales. Aunque la ausencia de cinco o más nucleótidos A o T consecutivos elimina la mayoría de las señales de poliadenilación de vegetales, si existen más de una de las señales de poliadenilación secundarias a una distancia de diez nucleótidos entre sí, entonces la secuencia de nucleótidos de esta región se altera para eliminar estas señales, manteniendo la secuencia de aminoácidos codificados original.

La siguiente etapa es considerar los aproximadamente 15 o aproximadamente 30 restos nucleótidos que rodean a la región rica en A+T. Si el contenido en A+T de la región circundante es menor que 80%, la región debe estudiarse para detectar señales de poliadenilación. La alteración de la región basada en las señales de poliadenilación depende (1) del número de señales de poliadenilación presentes, y (2) de la presencia de una señal de poliadenilación de vegetales importante. Las señales de poliadenilación se retiran mediante sustitución de bases de la secuencia del ADN en el contexto de la sustitución de codones.

Se buscan dos patrones adicionales no identificados en la patente de EEUU n.º 5.500365 y se eliminan en las realizaciones de la presente invención. La secuencia AGGTAA y GCAGGT son secuencias consenso para el intrón de los sitios de corte 5' y 3', respectivamente, en plantas monocotiledóneas y plantas dicotiledóneas. Sólo GT en el sitio de corte 5' y el AG en el sitio de corte 3' son necesarios para que se produzca un apareamiento exacto. Sin embargo, cuando se realiza una búsqueda para estas secuencias consenso, no se permiten desapareamientos en las bases.

Después de cada etapa se realiza un cartografiado de la secuencia utilizando el programa MAP de GCG para determinar la localización de los marcos de lectura abiertos y para identificar patrones de secuencia que tengan que modificarse. La etapa final sería realizar un análisis de coincidencia de secuencia utilizando, por ejemplo, el programa GAP del paquete informático GCG para determinar el grado de divergencia y porcentaje de coincidencia de secuencia.

#### Polipéptidos

En general, la proteína traducida del polinucleótido artificial tendrá la misma secuencia de aminoácidos que la proteína traducida a partir de la región codificadora no modificada. Sin embargo, la sustitución de codones que codifican aminoácidos que proporcionan un homólogo funcional de la proteína es un aspecto. Por ejemplo, ciertos aminoácidos pueden ser sustituidos por otros aminoácidos en una estructura de proteína sin una pérdida apreciable de capacidad de unión interactiva con estructuras tales como, por ejemplo, regiones de unión al antígeno de anticuerpos o sitios de unión sobre moléculas de sustrato. Puesto que es la capacidad interactiva y la naturaleza de una proteína lo que definen la actividad biológica funcional de la proteína, pueden realizarse ciertas sustituciones en la secuencia de aminoácidos en una secuencia de proteína y, por supuesto, en su secuencia codificadora de ADN subyacente, y obtener no obstante una proteína con propiedades similares. Por tanto, los inventores contemplan que puedan realizarse diversos cambios en las secuencias peptídicas de las composiciones descritas realizando cambios en la correspondiente secuencia de ADN que codifica los péptidos, en los que los péptidos no muestren una pérdida apreciable en su utilidad o actividad biológica.

Otro aspecto comprende a los homólogos funcionales, que se diferencian en uno o más aminoácidos de los de un polipéptido proporcionado en la presente como resultado de una o más sustituciones conservativas de aminoácidos. En la técnica se sabe que uno o más aminoácidos en una secuencia nativa pueden sustituirse por al menos uno aminoácido distinto cuya carga y polaridad sea similar al del aminoácido nativo, produciendo un cambio silencioso. Por ejemplo, la valina es un sustituto conservativo para la alanina, y la treonina es un sustituto conservativo para la serina. Las sustituciones conservativas para un aminoácido dentro de una secuencia polipeptídica nativa pueden seleccionarse de otros miembros de la clase a la cual pertenece el aminoácido natural. Los aminoácidos pueden dividirse en los siguientes cuatro grupos: (1) aminoácidos ácidos, (2) aminoácidos básicos, (3) aminoácidos polares neutros, y (4) aminoácidos no polares neutros. Los aminoácidos representativos dentro de estos grupos incluyen, pero no se limitan a: (1) aminoácidos ácidos (cargados negativamente), tales como ácido aspártico y ácido glutámico; (2) aminoácidos básicos (cargados positivamente), tales como arginina, histidina y lisina; (3) aminoácidos polares neutros, tales como glicina, serina, treonina, cisteína, tirosina, asparagina y glutamina; y (4) aminoácidos no polares (hidrófobos) neutros, tales como alanina, leucina, isoleucina,

valina, prolina, fenilalanina, triptófano y metionina. Los sustitutos conservados para un aminoácido con una secuencia de aminoácidos nativa pueden seleccionarse de otros miembros del grupo al cual pertenece el aminoácido natural. Por ejemplo, un grupo de aminoácidos que tiene cadenas laterales alifáticas son la glicina, la alanina, la valina, la leucina y la isoleucina; un grupo de aminoácidos que tienen cadenas laterales de hidroxilo alifáticas son la serina y la treonina; un grupo de aminoácidos que tienen cadenas laterales que contienen amida son la asparagina y la glutamina; un grupo de aminoácidos que tienen cadenas laterales aromáticas son la fenilalanina, la tirosina y el triptófano; un grupo de aminoácidos que tienen cadenas laterales básicas son la lisina, la arginina y la histidina; y un grupo de aminoácidos que tienen cadenas laterales que contienen azufre son la cisteína y la metionina. Los grupos de sustitución de aminoácidos conservativos naturales son: valina-leucina, valina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina, ácido aspártico-ácido glutámico, y asparagina-glutamina.

#### Construcciones de ADN

Puede transferirse material genético exógeno a una planta mediante el uso de una construcción de ADN diseñada para dicho objetivo mediante procedimientos que emplean *Agrobacterium*, bombardeo de partículas u otros procedimientos conocidos por los expertos en la técnica. El diseño de dicha construcción de ADN está, en general, dentro de los conocimientos de la técnica (Plant Molecular Biology: A Laboratory Manual, eds. Clark, Springer, Nueva York (1997)). Los ejemplos de plantas a las que puede transferirse material genético exógeno incluyen, sin limitación, alfalfa, *Arabidopsis*, cebada, *Brassica*, brócoli, repollo, cítricos, algodón, ajo, avena, colza, cebolla, canola, lino, maíz, plantas anuales ornamentales y perennes ornamentales, guisante, cacahuete, pimienta, patata, arroz, centeno, sorgo, soja, fresa, azúcar de caña, remolacha azucarera, tomate, trigo, álamo, pino, abeto, eucalipto, manzana, lechuga, lentejas, uvas, plátanos, té, césped, girasol, palma aceitera, *Phaseolus*, árboles, arbustos y viñas. Se sabe que las plantas importantes desde un punto de vista agronómico comprenden genotipos, variedades y cultivares, y que pueden ser ensayadas por los expertos en la técnica de la biología molecular vegetal y la selección y cultivo de plantas.

Se han descrito un gran número de moléculas de promotores de ADN aisladas que son activas como elemento genético de un transgén en células vegetales. Estas incluyen el promotor de nopalina sintasa (P-nos) (Ebert *et al.*, Proc. Natl. Acad. Sci. (U.S.A.), 84:5745-5749, 1987), el promotor de octopina sintasa (P-ocs), que son transportados en plásmidos inductores de tumores de *Agrobacterium tumefaciens*, los promotores de caulimovirus, tales como el promotor 19S del virus del mosaico de la coliflor (CaMV) (Lawton *et al.*, Plant Mol. Biol., 9:315-324, 1987) y el promotor 35S de CaMV (Odell *et al.*, Nature, 313:810-812, 1985)), el promotor 35S del virus del mosaico de *Scrophularia* (patente de EEUU n.º 6.018.100), el promotor fotoinducible de la subunidad pequeña de la ribulosa-1,5-bis-fosfato carboxilasa (ssRUBISCO), el promotor Adh (Walker *et al.*, Proc. Natl. Acad. Sci. (U.S.A.), 84:6624-6628, 1987), el promotor de la sacarosa sintasa (Yang *et al.*, Proc. Natl. Acad. Sci. (U.S.A.), 87:4144-4148, 1990), el promotor del complejo del gen R (Chandler *et al.*, Plant Cell 1:1175-1183, 1989), y el promotor del gen de la proteína de unión de clorofila *a/b*.

Puede utilizarse una diversidad de promotores específicamente activos en tejidos vegetales, tales como hojas, tallos, raíces y tubérculos, para expresar las moléculas de ácidos nucleicos. Los ejemplos de promotores específicos de tubérculos incluyen, pero no se limitan a los promotores de patatina de clase I y II (Bevan *et al.*, EMBO J., 8:1899-1906, 1986; Koster-Topfer *et al.*, Mol. Gen. Genet., 219:390-396, 1989; Mignery *et al.*, Gene, 62:27-44, 1988; Jefferson *et al.*, Plant Mol. Biol., 14: 995-1006, 1990), el promotor para los genes ADPGPP del tubérculo de la patata, ambas subunidades grande y pequeña; el promotor de sacarosa sintasa (Salanoubat y Belliard, Gene, 60:47-56, 1987; Salanoubat y Belliard, Gene, 84:181-185, 1989); y el promotor para las proteínas de tubérculo principales que incluyen los complejos de proteína de 22 kd e inhibidores de proteinasa (Hannapel, Plant Physiol., 101: 703-704, 1993). Los ejemplos de promotores específicos de hojas incluyen, pero no se limitan a los promotores de ribulosa bifosfato carboxilasa (RbcS o RuBISCO) (véase, por ejemplo, Matsuoka *et al.*, Plant J., 6:311-319, 1994); el promotor del gen de la proteína de unión de clorofila *a/b* recolectora de luz (véase, por ejemplo, Shiina *et al.*, Plant Physiol., 115:477-483, 1997; Casal *et al.*, Plant Physiol., 116:1533-1538, 1998); y el promotor del gen relacionado con *myb* de *Arabidopsis thaliana* (Atmyb5) (Li *et al.*, FEBS Lett., 379:117-121, 1996). Los ejemplos de promotores específicos de raíces incluyen, pero no se limitan al promotor para el gen de quitinasa ácida (Samac *et al.*, Plant Mol. Biol., 25:587-596, 1994); los subdominios específicos de raíces del promotor CaMV35S que se han identificado (Lam *et al.*, Proc. Natl. Acad. Sci. (U.S.A.), 86:7390-7894, 1989); el promotor ORF13 de *Agrobacterium rhizogenes* que muestra una alta actividad en raíces (Hansen *et al.*, Mol. Gen. Genet., 254:337-343, 1997); el promotor para el gen RB7 específico de raíces del tabaco (patente de EEUU 5.750.386; Yamamoto *et al.*, Plant Cell, 3:371-382, 1991); y los promotores específicos de células de la raíz indicados por Conkling *et al.* (Conkling *et al.*, Plant Physiol., 93:1203-1211, 1990), y el promotor POX1 (Pox1, *poxl*) (Hertig, *et al.*, Plant Mol. Biol., 16:171, 1991).

Otra clase de promotores específicos de tejidos vegetativos útiles son los promotores meristemáticos (ápice de la raíz y de brotes). Por ejemplo, pueden utilizarse los promotores "SHOOTMERISTEMLESS" y "SCARECROW", que

son activos en el desarrollo de los meristemos apicales de brotes o raíces (Di Laurenzio *et al.*, Cell, 86:423- 433, 1996; Long, Nature, 379:66-69, 1996). Otro ejemplo de un promotor útil es el que controla la expresión del gen HMG2 de 3-hidroxi-3-metilglutaril coenzima A reductasa, cuya expresión está limitada a tejidos meristemáticos y florales (zona secretora del estigma, granos de polen mduros, tejido vascular del gineceo, y óvulos fertilizados) (véase, por ejemplo, Enjuto *et al.*, Plant Cell, 7:517-527, 1995). Otro ejemplo de un promotor útil es el que controla la expresión de genes relacionados con knl del maíz y otras especies que muestran expresión específica de meristemos (véase, por ejemplo, Granger *et al.*, Plant Mol. Biol., 31:373-378, 1996; Kerstetter *et al.*, Plant Cell, 6:1877-1887, 1994; Hake *et al.*, Philos. Trans. R. Soc. Lond. B. Biol. Sci., 350:45-51, 1995). Otro ejemplo de un promotor meristemático es el promotor KNAT1 de *Arabidopsis thaliana*. En el ápice de los brotes, la transcripción de KNAT1 se localiza principalmente en el meristemo apical del brote; la expresión de KNAT1 en el meristemo del brote disminuye durante la transición floral y se limita a la corteza del tallo de la inflorescencia (véase, por ejemplo, Lincoln *et al.*, Plant Cell, 6:1859-1876, 1994).

Pueden obtenerse promotores potenciados en semillas y específicos de semillas adecuados de los siguientes genes: MAC1 del maíz (Sheridan *et al.*, Genetics, 142:1009-1020, 1996); Cat3 del maíz (Genbank n.º L05934, Abler *et al.*, Plant Mol. Biol., 22:10131-1038, 1993); vivparous-1 de *Arabidopsis* (Genbank n.º U93215); Atimycl de *Arabidopsis* (Urao *et al.*, Plant Mol. Biol., 32:571-57, 1996; Conceicao *et al.*, Plant, 5:493-505, 1994); napA de *Brassica napus* (Genbank n.º J02798); la familia del gen napin de *Brassica napus* (Sjodahl *et al.*, Planta, 197:264-271, 1995). El promotor específico de óvulos para el gen BEL1 (Reiser *et al.*, Cell, 83:735-742, 1995, Genbank n.º U39944; Ray *et al.*, Proc. Natl. Acad. Sci. USA, 91:5761-5765, 1994) también puede utilizarse. Los promotores MEA (FIS 1) y FIS2 específicos de células centrales y del huevo también son promotores específicos de tejidos reproductivos útiles (Luo *et al.*, Proc. Natl. Acad. Sci. USA, 97:10637-10642, 2000; Vielle-Calzada *et al.*, Genes Dev., 13:2971-2982, 1999).

Se ha identificado un promotor específico del polen del maíz en el maíz (Guerrero *et al.*, Mol. Gen. Genet., 224:161-168, 1990). Se han descrito otros genes expresados específicamente en el polen (véase, por ejemplo, Wakeley *et al.*, Plant Mol. Biol., 37:187-192, 1998; Ficker *et al.*, Mol. Gen. Genet., 257:132-142, 1998; Kulikauskas *et al.*, Plant Mol. Biol., 34:809-814, 1997; Treacy *et al.*, Plant Mol. Biol., 34:603-611, 1997).

También pueden utilizarse promotores derivados de genes que codifican proteínas de almacenaje embrionarias, que incluyen el gen que codifica la proteína de almacenaje 2S de *Brassica napus* (Dasgupta *et al.*, Gene, 133:301-302, 1993); la familia de genes de proteínas de almacenaje de la semilla 2s de *Arabidopsis*; el gen que codifica la oleosina de 20 kD de *Brassica napus* (GenBank n.º M63985); los genes que codifican la oleosina A (Genbank n.º U09118) y la oleosina B (GenBank n.º U09119) de soja; el gen que codifica la oleosina de *Arabidopsis* (GenBank n.º Z17657); el gen que codifica la oleosina 18kD del maíz (GenBank n.º J05212, Lee, Plant Mol. Biol., 26:1981-1987, 1994); y el gen que codifica la proteína rica en azufre de bajo peso molecular de soja (Choi *et al.*, Mol. Gen. Genet., 246:266-268, 1995).

También pueden utilizarse promotores derivados de genes que codifican la zeína (que incluyen los genes 15 kD, 16 kD, 19 kD, 22 kD, 27 kD, y gamma; Pedersen *et al.*, Cell, 29:1015-1026, 1982). Las zeínas son un grupo de proteínas de almacenaje que se encuentran en el endospermo del maíz.

Otro promotores que se sabe que actúan, por ejemplo, en el maíz, incluyen los promotores para los siguientes genes: *waxy*, *Brittle*, *Shrunken 2*, enzimas ramificadas I y II, almidón sintasas, enzimas desramificantes, oleosinas, glutelinas, y sacarosa sintasas. Un promotor particularmente preferido para la expresión en el endospermo de maíz es el promotor para el gen de glutelina del arroz, más en concreto el promotor Osgt-1 (Zheng *et al.*, Mol. Cell Biol., 13:5829-5842, 1993). Los ejemplos de promotores adecuados para la expresión en trigo incluyen los promotores para las subunidades de ADPglucosa pirofosforilasa (ASPGPP), las almidón sintasas unidas a gránulos y otras almidón sintasas, las enzimas ramificantes y desramificantes, las proteínas abundantes en la embriogénesis, las gliadinas y las gluteninas. Los ejemplos de estos promotores en el arroz incluyen los promotores para las subunidades de ADPGPP, las almidón sintasas unidas a gránulos y otras almidón sintasas, las enzimas ramificantes, las enzimas desramificantes, las sacarosa sintasas y las glutelinas. Un promotor particularmente preferido es el promotor para la glutelina del arroz, Osgt-1. Los ejemplos de estos promotores para la cebada incluyen los promotores para las subunidades de ADPGPP, las almidón sintasas unidas a gránulos y otras almidón sintasas, las enzimas ramificantes, las enzimas desramificantes, las sacarosa sintasas, las hordeínas, las globulinas embrionarias, y las proteínas específicas de aleurona.

Puede utilizarse un promotor del tomate activo durante la maduración del fruto, la senescencia y abscisión de las hojas y, en un menor grado, de la flores (Blume *et al.*, Plant J., 12:731-746, 1997). Otros ejemplos de promotores incluyen el promotor del gen SK2 específico del pistilo en la patata (*Solanum tuberosum* L.), que codifica una endoquitinasa básica específica del pistilo (Ficker *et al.*, Plant Mol. Biol., 35:425-431, 1997); el gen Blec4 del guisante (*Pisum sativum* cv. Alaska), activo en el tejido epidérmico de los ápices de los brotes vegetativos y florales de la alfalfa transgénica. Esto lo convierte en una herramienta útil para dirigir la expresión de genes extraños hacia

la capa epidérmica de brotes en crecimiento activo. El promotor E8 específico de tejidos para el tomate también es útil para dirigir la expresión de genes en frutos (Deikman *et al.*, *Plant Physiology*, 100:2013-2017, 1992).

También se reconoce que, puesto que en muchos casos, los límites exactos de las secuencias reguladoras aún no se han definido por completo, los fragmentos de ADN con diferente longitud pueden tener idéntica actividad promotora.

Puede utilizarse promotores que se sabe que provocan o que se descubre que provoquen la transcripción del ADN en células vegetales en la presente invención. Estos promotores pueden obtenerse a partir de una diversidad de fuentes, tales como plantas y virus vegetales. Además de los promotores que se sabe que provocan la transcripción de ADN en células vegetales, pueden identificarse otras moléculas de promotores para su uso en la presente invención mediante la selección de un banco de ADNc vegetal para descubrir genes que se expresen de forma selectiva o preferiblemente en tejidos o células diana, y el aislamiento de la región genómica 5' de los ADNc identificados.

Las construcciones o los vectores también pueden incluir, con la región codificadora de interés, un ácido polinucleotídico que actúe, por entero o en parte, para terminar la transcripción de esta región. Por ejemplo, se han aislado secuencias de este tipo, que incluyen la secuencia 3' Tr7 y la secuencia 3' *nos* (Ingelbrecht *et al.*, *The Plant Cell*, 1:671-680, 1989; Bevan *et al.*, *Nucleic Acids Res.*, 11:369-385, 1983).

Un vector o una construcción también puede incluir elementos reguladores. Los ejemplos de estos incluyen el intrón 1 Adh (Callis *et al.*, *Genes and Develop.*, 1:1183-1200, 1987), el intrón de sacarosa sintasa (Vasil *et al.*, *Plant Physiol.*, 91:1575-1579, 1989) y el elemento omega TMV (Gallie *et al.*, *Plant Cell*, 1:301-311, 1989). Pueden incluirse estos y otros elementos reguladores cuando sea apropiado.

Un vector o una construcción también puede incluir un marcador seleccionable. Los marcadores seleccionables también pueden utilizarse para seleccionar plantas o células vegetales que contengan un material genético exógeno. Los ejemplos de estos incluyen, pero no se limita al gen *neo* (Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183-188, 1985) que codifica la resistencia a kanamicina y puede seleccionarse para utilizar kanamicina, G418; un gen *bar* que proporciona la resistencia al bialaphos; un gen de EPSP sintasa mutante (Hinchee *et al.*, *Bio/Technology*, 6:915-922 (1988)) que proporciona resistencia al glifosato; un gen de nitrilasa que proporciona resistencia al bromoxinilo (Stalker *et al.*, *J. Biol. Chem.*, 263:6310-6314 (1988)); un gen de acetolactato sintasa mutante (ALS) que confiere imidazolinona o sulfonilurea; y un gen DHFR resistente al metotrexato (Thillet *et al.*, *J. Biol. Chem.*, 263:12500-12508, 1988).

Un vector o una construcción también puede incluir un marcador seleccionable. Los marcadores seleccionables pueden utilizarse para controlar la expresión. Los ejemplos de marcadores seleccionables incluyen un gen de  $\beta$ -glucuronidasa o *uidA* (GUS) que codifica una enzima para la cual son conocidos diversos sustratos cromogénicos (Jefferson, *Plant Mol. Biol. Rep.*, 5:387-405 (1987); Jefferson *et al.*, *EMBO J.*, 6:3901-3907 (1987)); un gen del locus R, que codifica un producto que regula la producción de pigmentos de antocianina (color rojo) en tejidos vegetales (Dellaporta *et al.*, *Stadler Symposium*, 11:263-282 (1988)); un gen de  $\beta$ -lactamasa (Sutcliffe *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 75:3737-3741 (1978)), un gen que codifica una enzima para la cual son conocidos diversos sustratos cromogénicos (por ejemplo, PADAC, una cefalosporina cromogénica); un gen de luciferasa (Ow *et al.*, *Science*, 234:856-859 (1986)); un gen *xyle* (Zukowsky *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 80:1101-1105 (1983)) que codifica una catecol dioxigenasa que puede convertir catecoles cromogénicos; un gen de  $\alpha$ -amilasa (Ikata *et al.*, *Bio/Technol.*, 8:241-242, 1990); un gen de tirosinasa (Katz *et al.*, *J. Gen. Microbiol.*, 129:2703-2714, 1983) que codifica una enzima capaz de oxidar la tirosina a DOPA y dopaquinona que, a su vez, se condensa en melanina; y una  $\alpha$ -galactosidasa.

#### Introducción de los polinucleótidos en plantas

Existen muchos procedimientos para introducir moléculas de ácidos nucleicos transformantes en células vegetales. Se cree que los procedimientos adecuados incluyen casi cualquier procedimiento mediante el cual puedan introducirse moléculas de ácidos nucleicos en una célula, tal como mediante infección con *Agrobacterium* o transporte directo de moléculas de ácidos nucleicos tales como, por ejemplo, mediante transformación mediada por PEG, mediante electroporación o mediante aceleración de partículas revestidas con ADN (Potrykus, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 42:205-225 (1991); Vasil, *Plant Mol. Biol.*, 25:925-937 (1994)). Por ejemplo, se ha utilizado la electroporación para transformar protoplastos de *Zea mays* (Fromm *et al.*, *Nature*, 312:791-793, 1986). Otros sistemas de vectores adecuados para introducir ADN transformante en una célula vegetal hospedante incluyen, pero no se limitan a vectores de cromosomas artificiales binarios (BIBAC) (Hamilton *et al.*, *Gene*, 200:107-116, 1997), y la transfección con vectores víricos de ARN (Della-Cioppa *et al.*, *Ann. N.Y. Acad. Sci.* (1996), 792 pp., *Engineering Plants for Commercial Products and Applications*, pp. 57-61).

La tecnología para la introducción de ADN en células es muy conocida por los expertos en la técnica. Se han



- 5 descrito cuatro procedimientos generales para transportar un gen hacia el interior de células: (1) procedimientos químicos (Graham y van der Eb, *Virology*, 54:536-539, 1973); (2) procedimientos físicos, tales como microinyección (Capecchi, *Cell*, 22:479-488 (1980)), electroporación (Wong y Neumann, *Biochem. Biophys. Res. Commun.*, 107:584-587 (1982); Fromm *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 82:5824-5828 (1985); patente de EEUU n.º 5.384.253); y la pistola de genes (Johnston y Tang, *Methods Cell Biol.*, 43:353-365 (1994)); (3) vectores víricos (Clapp, *Clin. Perinatol.*, 20:155-168, 1993; Lu *et al.*, *J. Exp. Med.*, 178:2089-2096, 1993; Eglitis y Anderson, *Biotechniques*, 6:608-614, 1988); y (4) mecanismos mediados por receptores (Curiel *et al.*, *Hum. Gen. Ther.*, 3:147-154, 1992; Wagner *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:6099-6103, 1992).
- 10 Los procedimientos de aceleración que pueden utilizarse incluyen, por ejemplo, bombardeo de microproyectiles. Un ejemplo de un procedimiento para transportar moléculas de ácidos nucleicos transformantes a células vegetales es el bombardeo de microproyectiles. Este procedimiento ha sido divulgado por Yang y Christou, eds., *Particle Bombardment Technology for Gene Transfer*, Oxford Press, Oxford, Reino Unido (1994). Las partículas no biológicas (microproyectiles) puede revestirse con ácidos nucleicos y transportarse al interior de células mediante una fuerza propulsora. Los ejemplos de partículas incluyen las formadas por wolframio, oro y platino.
- 15 La transferencia mediada por *Agrobacterium* es un sistema ampliamente aplicable para introducir genes en el interior de células vegetales porque el ADN puede introducirse en tejidos vegetales completos, obviando así la necesidad de regenerar una planta intacta a partir de un protoplasto. El uso de vectores integradores en plantas mediados por *Agrobacterium* para introducir ADN en células vegetales es muy conocido en la técnica. Véanse, por ejemplo, los procedimientos descritos por Fraley *et al.*, *Bio/Technology*, 3:629-635 (1985), y Rogers *et al.*, *Methods Enzymol.*, 153:253-277 (1987). Además, la integración del Ti-ADN es un procedimiento relativamente preciso que produce pocas redistribuciones. La región del ADN que se va a transferir está definida por las secuencias de frontera, y el ADN intermedio normalmente se inserta en el genoma vegetal como se ha descrito (Spielmann *et al.*, *Mol. Gen. Genet.*, 205:34 (1986)).
- 20 Una planta transgénica que resulta de procedimientos de transformación con *Agrobacterium* con frecuencia contiene un único gen en un cromosoma. Estas plantas transgénicas pueden denominarse hemigóticas para el gen añadido. Es más preferida una planta transgénica que sea homocigótica para el gen estructural añadido, es decir, una planta transgénica que contenga dos genes añadidos, un gen en el mismo locus sobre cada cromosoma de un par cromosómico. Una planta transgénica homocigótica puede obtenerse cruzando sexualmente (autofecundación) una planta transgénica con segregación independiente que contiene un único gen añadido, germinando algunas de las semillas producidas, y analizando las plantas resultantes producidas para detectar el gen de interés.
- 25 También se entiende que pueden cruzarse dos plantas transgénicas diferentes para producir una descendencia que contenga dos genes exógenos añadidos que se segregan independientemente. La autofecundación de la progenie apropiada puede producir plantas que sean homocigóticas para ambos genes exógenos añadidos, que codifican un polipéptido de interés. También se contempla el retrocruzamiento con una planta progenitora y el cruzamiento exogámico con una planta no transgénica, así como la propagación vegetativa.
- 30 La regeneración, el desarrollo y el cultivo de plantas a partir de un transformante de protoplastos vegetales individuales o a partir de varios explantes transformados es muy conocido en la técnica (Weissbach y Weissbach, en: *Methods for Plant Molecular Biology* (eds.), Academic Press, Inc., San Diego, CA, (1988)). Este procedimiento de regeneración y crecimiento incluye generalmente las etapas de selección de las células transformadas, y cultivo de estas células individualizadas a través de las etapas normales del desarrollo embrionario hasta la etapa de plántula enraizada. Los embriones y las semillas transgénicos se regeneran de forma similar. Los brotes enraizados transgénicos resultantes después se plantan en un medio de crecimiento vegetal apropiado, tal como suelo.
- 35 El desarrollo o la regeneración de plantas que contienen el gen extraño exógeno que codifica una proteína de interés es muy conocido en la técnica. Preferiblemente, las plantas regeneradas se autopolinizan para proporcionar plantas homocigóticas transgénicas. De otra forma, el polen obtenido de las plantas regeneradas se cruza con plantas cultivadas desde la semilla de líneas con importancia agronómica. A la inversa, el polen de plantas de estas líneas importantes se emplea para polinizar plantas regeneradas. Una planta transgénica de la presente invención que contiene un polipéptido deseado se cultiva utilizando procedimientos muy conocidos por los expertos en la técnica.
- 40 La presente invención también proporciona partes de las plantas de la presente invención. Las partes de las plantas, sin limitación, incluyen semillas, endospermo, óvulos y polen. En una realización particularmente preferida de la presente invención, la parte de la planta es una semilla.
- 45 Se han divulgado procedimientos para transformar dicotiledóneas, principalmente mediante el uso de
- 50
- 55

*Agrobacterium tumefaciens*, y para obtener plantas transgénicas, por ejemplo, algodón (patente de EEUU n.º 5.004.863, patente de EEUU n.º 5.159.135, patente de EEUU n.º 5.518.908), soja (patente de EEUU n.º 5.569.834) y *Brassica* (patente de EEUU n.º 5.463.174).

5 También se ha divulgado la transformación de monocotiledóneas utilizando electroporación, bombardeo de partículas, y *Agrobacterium*. Por ejemplo, la transformación y la regeneración de plantas se ha logrado en espárragos, cebada, *Zea mays* (Fromm *et al.*, Bio/Technology, 8:833 (1990); Armstrong *et al.*, Crop Science, 35:550-557 (1995)); avena; arroz; centeno; caña de azúcar; festuca alta; y trigo (patente de EEUU n.º 5.631.152).

10 Además de los procedimientos analizados anteriormente, los expertos en la técnica conocen los materiales didácticos que describen condiciones y procedimientos específicos para la construcción, la manipulación y el aislamiento de macromoléculas (por ejemplo, moléculas de ADN, plásmidos), la generación de organismos recombinantes y la selección y el aislamiento de clones (véase, por ejemplo, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press (1989); Mailga *et al.*, Methods in Plant Molecular Biology, Cold Spring Harbor Press (1995); Birren *et al.*, Genome Analysis: Detecting Genes, 1, Cold Spring Harbor, Nueva York (1998); Birren *et al.*, Genome Analysis: Analyzing DNA, 2, Cold Spring Harbor, Nueva York (1998); Plant Molecular Biology: A Laboratory Manual, eds. Clark, Springer, Nueva York (1997)).

Habiendo descrito en general la invención, esta se entenderá con más facilidad haciendo referencia a los siguientes ejemplos. Los ejemplos que no están cubiertos por el alcance de las reivindicaciones se ofrecen con fines ilustrativos.

## Ejemplos

### 20 Ejemplo 1

Cuando un polinucleótido de una planta nativo aislado que comprende una secuencia codificadora se reconstruye como un transgén, y luego se introduce en la planta mediante procedimientos de transformación de plantas, existe el riesgo de que la expresión del gen de la planta homólogo endógeno interaccione de forma negativa con el transgén. Para evitar estas interacciones negativas puede ser necesario proporcionar un polinucleótido del transgén con una secuencia sustancialmente divergente de la del gen de la planta nativo. Puede producirse una molécula polinucleotídica artificial mediante el procedimiento descrito en la presente y emplearse para reducir la aparición del silenciamiento de transgenes.

Este ejemplo sirve para ilustrar la producción de un polinucleótido que codifica una EPSP sintasa vegetal modificada. Se emplea la enzima EPSPS del arroz (*Oryza sativa*) nativa y el péptido de tránsito del cloroplasto para construir una molécula polinucleotídica artificial que también incluye codones que codifican aminoácidos sustituidos que no aparecen en la naturaleza en la enzima EPSPS del arroz. Estos aminoácidos sustituidos proporcionan una enzima EPSPS del arroz resistente al glifosato (OsEPSPS\_TIPS, SEQ ID NO:1).

30 Se emplean las etapas descritas en la tabla 6 para construir esta secuencia polinucleotídica artificial (OsEPSPS\_AT, SEQ ID NO:3) utilizando la tabla de utilización de codones de *Arabidopsis* y los parámetros para la construcción de una molécula polinucleotídica sustancialmente divergente, que cuando se expresa en plantas codifica una enzima EPSPS del arroz modificada resistente al herbicida glifosato. En la figura 1 se muestra la comparación de la secuencia del gen EPSPS del arroz nativa, denominada OsEPSPS\_Nat (SEQ ID NO:2) que previamente se ha modificado para que codifique una enzima resistente al glifosato, con la molécula polinucleotídica modificada para la utilización de codones de *Arabidopsis*, OsEPSPS\_AT (SEQ ID NO:3) y con la secuencia modificada para la utilización de codones de *Zea mays*, OsEPSPS\_ZM (SEQ ID NO:4) mediante este procedimiento. La figura 1 muestra las bases de nucleótidos cambiadas en los polinucleótidos modificados, comparado con OsEPSPS\_Nat (SEQ ID NO:2).

Tabla 6. Diseño de polinucleótidos para una EPSP sintasa del arroz modificada (OsEPSPS\_AT).

1. Se sustituyen los aminoácidos en las posiciones 173 y 177 para proporcionar una enzima EPSPS del arroz modificada resistente al herbicida glifosato mostrada en SEQ ID NO:1.

2. Se retrotraduce la SEQ ID NO:1 para generar una secuencia polinucleotídica artificial utilizando la tabla de utilización de codones de *Arabidopsis thaliana* (tabla 2).

3. Se realiza el alineamiento de la secuencia con la secuencia polinucleotídica de OsEPSPS nativa (SEQ ID NO:2) y la secuencia polinucleotídica artificial para determinar el grado de coincidencia de secuencia, cartografiar los marcos de lectura abierto, y seleccionar los patrones para buscar e identificar las secuencias de reconocimiento de enzimas de restricción.

4. Se realizan correcciones en los codones utilizados en la secuencia polinucleotídica artificial para lograr el

porcentaje deseado de coincidencia de secuencia y para evitar el agrupamiento de codones idénticos. Esto resulta especialmente importante para los aminoácidos que aparecen con alta frecuencia, es decir, alanina, glicina, histidina, leucina, serina y valina. En la tabla 2 aparece la distribución aproximada de la utilización de codones en la secuencia polinucleotídica según la utilización de codones de *Arabidopsis*.

- 5 5. La secuencia polinucleotídica se inspecciona para detectar regiones locales que tengan una proporción GC:AT mayor que aproximadamente 2 a lo largo de un intervalo de aproximadamente 50 nucleótidos contiguos. La secuencia polinucleotídica se ajusta según sea necesario mediante la sustitución de codones en estas regiones, de forma que la proporción GC:AT local sea menor que aproximadamente 2, y la composición polinucleotídica completa esté en el intervalo de 0,9-1,3.
- 10 6. Se introducen codones de fin para los marcos de traducción "b", "c", "d", "e" y "f". Los codones de fin de la traducción se crean en los marcos de traducción "b", "c", "d", "e" y "f" reemplazando uno o más codones dentro de aproximadamente 130 pares de bases (pb) de los extremos del polinucleótido artificial, lo cual crea un codón de fin sin cambiar la secuencia codificadora de aminoácidos del marco uno.
- 15 7. Se eliminan los codones ATG de los marcos de lectura abiertos directos (marcos "b" y "c") e inversos (marcos "d", "e", "f"). Los marcos de lectura directos e inversos se inspeccionan para determinar la presencia de codones ATG. Cualquier codón ATG en el marco "b" y "c" encontrado en la secuencia polinucleotídica antes de la tercera Met en el marco "a" del polinucleótido se elimina reemplazando uno o más codones que se solapan con ATG, cambiando uno de los nucleótidos sin cambiar la secuencia codificadora de aminoácidos del marco "a". En los marcos inversos, el reemplazamiento de ATG o la introducción del codón de fin puede realizarse para interrumpir los marcos de lectura potenciales.
- 20 8. Se eliminan los patrones de reconocimiento de enzimas de restricción no deseados y otros patrones específicos (poliadenilación, ruptura del ARN, patrones de inestabilidad de secuencia). La secuencia polinucleotídica se inspecciona para determinar la presencia de cualquier patrón polinucleotídico no deseado, y los patrones se alteran sustituyendo codones en estas regiones.
- 25 9. Se comprueba la coincidencia de secuencia entre un primer polinucleótido y el polinucleótido artificial creado mediante el procedimiento de la presente invención. Se elimina la coincidencia de secuencia en un polinucleótido contiguo que sea mayor que 23 pb. Resulta deseable eliminar la coincidencia de secuencia mayor que aproximadamente 15 pb. Resulta útil seleccionar los aminoácidos, tales como serina, arginina y leucina, que tienen 6 codones, o los aminoácidos con 4 codones, para eliminar la coincidencia de secuencia.
- 30 10. Se examina la secuencia polinucleotídica artificial que surge de cualquiera de las etapas 1 a 9 para detectar cualquiera de las características de secuencia identificadas en las etapas 4-9, y si la secuencia no cumple con las condiciones se realizan más sustituciones de codones en la secuencia hasta que se cumplan las condiciones de las etapas 4-9.
- 35 11. Se construye la molécula polinucleotídica artificial mediante procedimientos conocidos en la técnica, por ejemplo, utilizando PCR con una mezcla de cebadores solapantes. Los cebadores en los extremos del gen pueden contener sitios de restricción convenientes que permitan la clonación fácil del gen con el vector seleccionado. En el extremo 5' normalmente AlflII, BspHI, NcoI, NdeI, PciI, o SphI son las más convenientes puesto que sus secuencias contienen un codón de inicio ATG; sin embargo, también pueden utilizarse otras enzimas, si se diseña un polinucleótido modificado para crear una fusión con otro segmento polinucleotídico, por ejemplo, una secuencia codificadora de EPSPS y péptido de tránsito del cloroplasto.
- 40 12. Se realiza un análisis de la secuencia de ADN del polinucleótido artificial para confirmar que la construcción sintética ha dado como resultado la molécula polinucleotídica deseada. Si se encuentran errores, entonces estos se eliminan mediante mutagénesis dirigida específica de sitio, para la cual existen muchos procedimientos conocidos por los expertos en la técnica de la mutagénesis del ADN.
- 45 Se realiza una versión con la utilización de codones de *Zea mays* (*Zea mays*, tabla 3) de la secuencia de la EPSPS del arroz resistente al glifosato (enzima EPSPS de *Oryzae sativa* con mutaciones TIPS, SEQ ID NO:1). El polinucleótido que codifica esta enzima incluye codones que codifican aminoácidos sustituidos que no aparecen en la naturaleza en la enzima EPSPS del arroz nativa. Estos aminoácidos sustituidos proporcionan una enzima EPSPS del arroz resistente al glifosato. Se emplean las etapas descritas en la tabla 7 para construir una secuencia polinucleotídica artificial modificada (OsEPSPS\_ZM, SEQ ID NO:4) basada en la tabla de utilización de codones de *Zea mays* que codifica una enzima EPSPS del arroz modificada resistente al herbicida glifosato. La comparación de la secuencia polinucleotídica de OsEPSPS\_Nat (SEQ ID NO:2) con la secuencia polinucleotídica artificial OsEPSPS\_ZM (SEQ ID NO:4) utilizando la utilización de codones de *Zea mays* se muestra en la tabla 2.
- 50

Tabla 7. Construcción de polinucleótidos para una EPSP sintasa del arroz modificada (OsEPSPS\_ZM).

1. Se retrotraduce la SEQ ID NO:1 para generar una secuencia polinucleotídica artificial utilizando la tabla de utilización de codones de *Zea mays* (tabla 3).
2. Se realiza el alineamiento de la secuencia con la secuencia polinucleotídica de OsEPSPS nativa (SEQ ID NO:2) y la secuencia polinucleotídica artificial para determinar el grado de coincidencia de secuencia, cartografiar los marcos de lectura abiertos, y seleccionar los patrones para buscar e identificar las secuencias de reconocimiento de enzimas de restricción.
3. Se realizan correcciones en los codones utilizados en la secuencia polinucleotídica artificial para lograr el porcentaje deseado de coincidencia de secuencia y para evitar el agrupamiento de codones idénticos. Esto resulta especialmente importante para los aminoácidos que aparecen con alta frecuencia, es decir, alanina, glicina, histidina, leucina, serina y valina. En la tabla 3 aparece la distribución aproximada de la utilización de codones en la secuencia polinucleotídica según la utilización de codones de *Zea mays*.
4. La secuencia polinucleotídica se inspecciona para detectar regiones locales que tengan una proporción GC:AT mayor que aproximadamente 2 a lo largo de un intervalo de aproximadamente 50 nucleótidos contiguos. La secuencia polinucleotídica se ajusta según sea necesario mediante la sustitución de codones en estas regiones, de forma que la proporción GC:AT local sea menor que aproximadamente 2, y la composición polinucleotídica completa esté en el intervalo de 1,2-1,7.
5. Se siguen las etapas 6-12 de la tabla 6.

Tabla 8. Porcentaje de coincidencia de secuencia entre los polinucleótidos OsEPSPS.

	OsEPSPS_ZM	OsEPSPS_AT	OsEPSPS_Nat
OsEPSPS_ZM	100,00	73,51	71,58
OsEPSPS_AT		100,00	74,03
OsEPSPS_Nat			100,00

Tabla 9. Composición de nucleótidos y proporción GC:AT de las secuencias polinucleotídicas modificadas para la secuencia del gen EPSPS del arroz.

	A	C	G	T	GC:AT
OsEPSPS_AT	377	336	444	391	1,02
OsEPSPS_ZM	365	381	470	332	1,22

Las dos secuencias polinucleotídicas artificiales de EPSPS del arroz (SEQ ID NO:3 y SEQ ID NO:4) se modifican de forma que el porcentaje de coincidencia sea menor que 75% comparado con SEQ ID NO:2, o con relación entre sí (tabla 8). La composición de nucleótidos y la proporción GC:AT de las secuencias polinucleotídicas para la secuencia del gen EPSPS del arroz se muestran en la tabla 9. Estos polinucleótidos pueden seleccionarse para su uso en construcciones de expresión en plantas junto con diferentes elementos reguladores, o pueden combinarse en una única planta mediante retransformación con una construcción de ADN o mediante procedimientos de cruzamiento de plantas. Se reducen los problemas de recombinación y silenciamiento de genes cuando las construcciones de ADN tienen menores niveles de ADN homólogo.

### Ejemplo 2

El maíz (*Zea mays*) ha sido genéticamente modificado para que tenga resistencia al herbicida glifosato (patente de EEUU n.º 6.040.497). Estas plantas de maíz contienen un transgén con una EPSP sintasa del maíz modificada para la tolerancia al glifosato. Los procedimientos pueden utilizarse para construir un nuevo polinucleótido artificial que codifica una EPSP sintasa del maíz que es sustancialmente diferente en cuanto a la coincidencia de secuencia con el gen de EPSP sintasa del maíz endógeno. El polinucleótido artificial de EPSP sintasa de maíz recién construido puede utilizarse como marcador seleccionable durante la selección de líneas de plantas transgénicas que pueden contener rasgos agronómicos transgénicos adicionales. Durante la producción de semillas de maíz híbridas, resulta útil que ambos progenitores sean tolerantes al glifosato utilizando transgenes que no interfieran.

Tabla 10. Construcción de polinucleótidos para una EPSP sintasa del maíz modificada (ZmEPSPS\_ZM, SEQ ID

NO:10).

1. Se retrotraduce la SEQ ID NO:8 para generar una secuencia polinucleotídica artificial utilizando la tabla de utilización de codones de *Zea mays* (tabla 3).

5 2. Se realiza el alineamiento de la secuencia con la secuencia polinucleotídica de ZmEPSPS\_Nat (SEQ ID NO:9) y la secuencia polinucleotídica artificial para determinar el grado de coincidencia de secuencia, cartografiar los marcos de lectura abierto, y seleccionar los patrones para buscar e identificar las secuencias de reconocimiento de enzimas de restricción.

10 3. Se realizan correcciones en los codones utilizados en la secuencia polinucleotídica artificial para lograr el porcentaje deseado de coincidencia de secuencia y para evitar el agrupamiento de codones idénticos. Esto resulta especialmente importante para los aminoácidos que aparecen con alta frecuencia, es decir, alanina, glicina, histidina, leucina, serina y valina. En la tabla 3 aparece la distribución aproximada de la utilización de codones en la secuencia polinucleotídica según la utilización de codones de *Zea mays*.

15 4. La secuencia polinucleotídica artificial se inspecciona para detectar regiones locales que tengan una proporción GC:AT mayor que aproximadamente 2 a lo largo de un intervalo de aproximadamente 50 nucleótidos contiguos. La secuencia polinucleotídica se ajusta según sea necesario mediante la sustitución de codones en estas regiones, de forma que la proporción GC:AT local sea menor que aproximadamente 2, y la composición polinucleotídica completa esté en el intervalo de 1,2-1,7.

5. Se siguen las etapas 6-12 de la tabla 6.

Tabla 11. Porcentaje de coincidencia de secuencia entre los polinucleótidos ZmEPSPS.

	ZmEPSPS_ZM	ZmEPSPS_Nat
ZmEPSPS_ZM	100,00	74,81
ZmEPSPS_Nat		100,00

20

25 Las secuencia de nucleótidos del gen de EPSPS de maíz también se modifica para reducir la coincidencia de secuencia entre el gen sintético y nativo, y para mantener una proporción global GC:AT típica para las monocotiledóneas. La proporción GC:AT para la secuencia ZmEPSPS\_ZM es 1,38. La coincidencia de secuencia se reduce en aproximadamente 75% entre el gen nativo (ZmEPSPS\_Nat, SEQ ID NO:9) y sintético (ZmEPSPS\_ZM, SEQ ID NO:10).

25

30 La comparación de los polinucleótidos nativos que codifican EPSPS indica que el péptido de tránsito del cloroplasto es el fragmento más divergente del gen. La similitud entre la secuencia de nucleótidos de los péptidos maduros es mayor que 88% para las enzimas de maíz y de arroz, y algunas regiones conservadas tienen una coincidencia de secuencia tan larga como 50 pb. Se ha observado silenciamiento de genes postranscripcional para secuencias tan pequeñas como 60 polinucleótidos (Sijen *et al.*, Plant Cell, 8:2277-2294, 1996; Mains, Plant Mol. Biol., 43:261-273, 2000).

30

### Ejemplo 3

35 La soja (*Glycine max*) ha sido genéticamente modificada para que tenga resistencia al herbicida glifosato mediante la expresión de EPSPS de clase II aislada a partir de *Agrobacterium* (Padgett *et al.*, Crop Sci., 35:1451-1461, 1995). Se ha identificado una secuencia del gen de EPSPS nativa de soja y se ha diseñado una secuencia polinucleotídica artificial. El polinucleótido artificial codifica una secuencia de proteína que está modificada para que produzca una enzima EPSPS resistente al glifosato (Gm\_EPSPS\_IKS, SEQ ID NO:5) reemplazando los aminoácidos T a I, R a K, y P a S, dentro del motivo GNAGTAMRP, dando como resultado una enzima EPSPS de soja modificada con el motivo GNAGIAMKS (SEQ ID NO:34), también denominada mutante IKS. La expresión de una enzima EPSPS modificada en las células de una planta mediante la transformación con un módulo de expresión en plantas de transgenes, que contiene un polinucleótido que codifica la EPSPS modificada con el motivo GNAGIAMKS, confiere a las plantas tolerancia al glifosato. Otras sustituciones de aminoácidos para la arginina (R) en el motivo también pueden incluir asparagina (N).

40

45 Tabla 12. Construcción de polinucleótidos para un gen de EPSP sintasa de soja modificada (GmEPSPS\_GM, SEQ ID NO:7).

45

1. Se retrotraduce la SEQ ID NO:5 para generar una secuencia polinucleotídica artificial utilizando la tabla de

utilización de codones de *Glycine max* (tabla 4).

2. Se realiza el alineamiento de la secuencia con la secuencia polinucleotídica de GmEPSPS\_Nat (SEQ ID NO:6) y la secuencia polinucleotídica artificial para determinar el grado de coincidencia de secuencia, cartografiar los marcos de lectura abiertos, y seleccionar los patrones para buscar e identificar las secuencias de reconocimiento de enzimas de restricción.
3. Se realizan correcciones en los codones utilizados en la secuencia polinucleotídica artificial para lograr el porcentaje deseado de coincidencia de secuencia y para evitar el agrupamiento de codones idénticos. Esto resulta especialmente importante para los aminoácidos que aparecen con alta frecuencia, es decir, alanina, glicina, histidina, leucina, serina y valina. En la tabla 4 aparece la distribución aproximada de la utilización de codones en la secuencia polinucleotídica según la utilización de codones de *Glycine max*.
4. La secuencia polinucleotídica artificial se inspecciona para detectar regiones locales que tengan una proporción GC:AT mayor que aproximadamente 2 a lo largo de un intervalo de aproximadamente 50 nucleótidos contiguos. La secuencia polinucleotídica se ajusta según sea necesario mediante la sustitución de codones en estas regiones, de forma que la proporción GC:AT local sea menor que aproximadamente 2, y la composición polinucleotídica completa esté en el intervalo de 0,9-1,3.
5. Se siguen las etapas 6-12 de la tabla 6.

Tabla 13. Comparación del porcentaje de coincidencia de secuencia de GmEPSPS modificadas a nivel de secuencia polinucleotídica.

	GmEPSPS_GM	GmEPSPS_Nat
GmEPSPS_GM	100,00	72,43
GmEPSPS_Nat		100,00

El gen de EPSPS de soja nativa se modifica utilizando una tabla de utilización de codones de la soja (tabla 4) y las condiciones descritas en la presente. La proporción relativa de GC:AT no cambia en el gen modificado, aunque la coincidencia de secuencia entre los dos se reduce al 72%.

#### Ejemplo 4

- El gen polinucleotídico *aroA* nativo aislado de *Agrobacterium* cepa CP4 (patente de EEUU n.º 5.633.435) que codifica una EPSP sintasa resistente al glifosato (SEQ ID NO:15) puede ser modificado para proporcionar un polinucleótido que tiene la utilización de codones de *Arabidopsis*, *Zea mays* o *Glycine max*. Para la expresión apropiada de CP4EPSPS para conferir tolerancia al glifosato en plantas, es necesario condensar un péptido de tránsito del cloroplasto a la secuencia codificadora de CP4EPSPS para dirigir la acumulación de la enzima hacia los cloroplastos. Se emplea habitualmente el péptido de tránsito del cloroplasto CTP2 para la expresión de este gen en plantas transgénicas (Nida *et al.*, J. Agric. Food Chem., 44:1960-1966, 1996). La secuencia de CP4EPSPS, junto con el polinucleótido CTP2 (SEQ ID NO:1) se han modificado. Otros péptidos de tránsito del cloroplasto conocidos en la técnica pueden condensarse con CP4EPSPS para dirigir la enzima a los cloroplastos.

Tabla 14. Construcción de polinucleótidos para la secuencia codificadora de *aroA*:CP4EPSP sintasa (CP4EPSPS\_AT, CP4EPSPS\_ZM, o CP4EPSPS\_GM).

1. Se coloca la secuencia del péptido de tránsito CTP2 (SEQ ID NO:11) delante de CP4EPSPS (SEQ ID NO:15) como un polipéptido de fusión. Se retrotraduce el polipéptido de fusión para producir una secuencia polinucleotídica artificial utilizando la tabla de utilización de codones de *Arabidopsis thaliana* (tabla 2), o la tabla de utilización de codones de *Zea mays* (tabla 3), o la tabla de utilización de codones de *Glycine max* (tabla 4).
2. Se realiza el alineamiento de la secuencia con la secuencia polinucleotídica de CTP2 nativa (SEQ ID NO:12) y CP4EPSPS nativa (SEQ ID NO:16) y la secuencia polinucleotídica artificial para determinar el grado de coincidencia de secuencia, cartografiar los marcos de lectura abiertos, y seleccionar los patrones para buscar e identificar las secuencias de reconocimiento de enzimas de restricción.
3. Se realizan correcciones en los codones utilizados en la secuencia polinucleotídica artificial para lograr el porcentaje deseado de coincidencia de secuencia y para evitar el agrupamiento de codones idénticos. Esto resulta

especialmente importante para los aminoácidos que aparecen con alta frecuencia, es decir, alanina, glicina, histidina, leucina, serina y valina. Se utiliza la distribución aproximada de la utilización de codones en la secuencia polinucleotídica según la utilización de codones de *Arabidopsis thaliana* (tabla 2), o la tabla de utilización de codones de *Zea mays* (tabla 3), dependiendo de qué tabla se esté utilizando.

5 4. La secuencia polinucleotídica se inspecciona para detectar regiones locales que tengan una proporción GC:AT mayor que aproximadamente 2 a lo largo de un intervalo de aproximadamente 50 nucleótidos contiguos. La secuencia polinucleotídica se ajusta según sea necesario mediante la sustitución de codones en estas regiones, de forma que la proporción GC:AT local sea menor que aproximadamente 2, y la composición polinucleotídica completa esté en el intervalo de 0,9-1,3 si se emplea la tabla 2, y de 1,2-1,7 si se emplea la tabla 3.

10 5. Se siguen las etapas 6-12 de la tabla 6.

Tabla 15. Comparación del porcentaje de coincidencia de secuencia de polinucleótidos CP4EPSPS artificiales.

	CTP2CP4_GM	CTP2CP4_AT	CTP2CP4_ZM	CTP2CP4_Syn	CTP2CP4_NAT
CTP2CP4_GM	100,00	75,66	74,12	75,15	74,37
CTP2CP4_AT		100,00	76,13	74,56*	72,93
CTP2CP4_ZM			100,00	77,76*	82,58
CTP2CP4_Syn				100,00	82,70
CTP2CP4_NAT					100,00

\* El porcentaje de coincidencia se relaciona con CP4EPSPS y no incluye el péptido de tránsito.

Tabla 16. Composición de nucleótidos y proporción GC:AT de las secuencias polinucleotídicas artificiales para la secuencia del gen CP4EPSPS.

	A	C	G	T	GC:AT
CTP2CP4_GM	382	375	442	397	1,05
CTP2CP4_AT	369	408	469	350	1,22
CTP2CP4_ZM	312	487	577	290	1,65

15

La secuencia polinucleotídica CTP2\_Nat (SEQ ID NO:12) más CP4EPSPS\_Nat (SEQ ID NO:16) denominada CTP2CP4\_Nat se compara en la tabla 15 con las secuencias polinucleotídicas artificiales denominadas CTP2CP4\_AT (CTP2\_AT, SEQ ID NO:13 condensada con CP4EPSPS\_AT, SEQ ID NO:17) y CTP2CP4\_ZM (CTP2\_AT, SEQ ID NO:14 condensada con CP4EPSPS\_ZM, SEQ ID NO:18). La secuencia polinucleotídica que es la más divergente de la secuencia nativa CTP2CP4\_NAT y CTP2CP4EPSPS\_Syn es CTP2CP4\_AT, que tiene una coincidencia de secuencia del 73% y 75%, respectivamente. La secuencia polinucleotídica CTP2CP4\_ZM, comparada con CTP2CP4\_Nat y CP4EPSPS\_Syn, tiene aproximadamente 83% y 78% de coincidencia con estas dos secuencias, respectivamente.

20

Un criterio principal para la selección de transgenes que se van a combinar en una planta es el porcentaje de coincidencia. La tabla 15 puede utilizarse para seleccionar una molécula polinucleotídica CP4EPSPS para la construcción de un módulo de expresión en plantas cuando se sabe que la planta receptora contiene más de un polinucleótido CP4EPSPS. La proporción GC:AT en CP4EPSPS nativa es de aproximadamente 1,7. La versión artificial con el sesgo de codones de *Zea mays* se produce para que tenga una proporción GC:AT muy similar. En la versión de codones de *Arabidopsis*, la proporción GC:AT disminuye hasta aproximadamente 1,2.

25

La expresión de genes también es un criterio para la selección de los transgenes que se van a expresar. La expresión de un transgén puede variar en diferentes plantas de cultivo, y por tanto resulta una ventaja tener varias secuencias codificadoras polinucleotídicas artificiales disponibles para ensayar en diferentes plantas de cultivo y genotipos, variedades o cultivares.

30

### Ejemplo 5

Se ha empleado la secuencia polinucleotídica *bar* (SEQ ID NO:20) que codifica una proteína de fosfotricina acetil transferasa (SEQ ID NO:19) para modificar genéticamente plantas para la resistencia al herbicida glifosato. Se han diseñado dos nuevas secuencias polinucleotídicas *bar*.

El alineamiento de BAR1\_Nat con los dos nuevos polinucleótidos BAR1 artificiales se muestra en la figura 4.

5      Tabla 17. Construcción de un gen polinucleotídico para BAR1\_AT (SEQ ID NO:21) y BAR1\_ZM (SEQ ID NO:22).

1. Se retrotraduce la SEQ ID NO:19 para generar una secuencia polinucleotídica artificial utilizando la tabla de utilización de codones de *Arabidopsis thaliana* (tabla 2) o la tabla de utilización de codones de *Zea mays* (tabla 3).

2. Se realiza el alineamiento de la secuencia con la secuencia polinucleotídica de BAR1\_Nat (SEQ ID NO:20) y la secuencia polinucleotídica artificial para determinar el grado de coincidencia de secuencia, cartografiar los marcos de lectura abiertos, y seleccionar los patrones para buscar e identificar las secuencias de reconocimiento de enzimas de restricción.

3. Se realizan correcciones en los codones utilizados en la secuencia polinucleotídica artificial para lograr el porcentaje deseado de coincidencia de secuencia y para evitar el agrupamiento de codones idénticos. Esto resulta especialmente importante para los aminoácidos que aparecen con alta frecuencia, es decir, alanina, glicina, histidina, leucina, serina y valina. Se utiliza la distribución aproximada de la utilización de codones en la secuencia polinucleotídica según la utilización de codones de *Arabidopsis thaliana* (tabla 2), o la tabla de utilización de codones de *Zea mays* (tabla 3), dependiendo de qué tabla se esté utilizando.

4. La secuencia polinucleotídica artificial se inspecciona para detectar regiones locales que tengan una proporción GC:AT mayor que aproximadamente 2 a lo largo de un intervalo de aproximadamente 50 nucleótidos contiguos. La secuencia polinucleotídica se ajusta según sea necesario mediante la sustitución de codones en estas regiones, de forma que la proporción GC:AT local sea menor que aproximadamente 2, y la composición polinucleotídica completa esté en el intervalo de 0,9-1,3 si se emplea la tabla 2, y de 1,2-1,7 si se emplea la tabla 3.

5. Se siguen las etapas 6-12 de la tabla 6.

La coincidencia de secuencia de los polinucleótidos artificiales BAR está en el intervalo de 73-77% (tabla 18). El polinucleótido nativo es muy rico en GC. La versión artificial (BAR1\_ZM) con el sesgo de codones de *Zea mays* ha reducido la proporción GC:AT a aproximadamente 1,3, y la versión artificial (BAR1\_AT) con el sesgo de codones de *Arabidopsis* tiene una proporción de aproximadamente 1,0 (tabla 19).

Tabla 18. Porcentaje de coincidencia de secuencia entre genes *bar* a nivel de secuencia polinucleotídica.

	BAR1_ZM	BAR1_AT	BAR1_Nat
BAR1_ZM	100,00	77,35	76,99
BAR1_AT		100,00	73,73
BAR1_Nat			100,00

30      Tabla 19. Composición de nucleótidos y proporción GC:AT de las secuencias polinucleotídicas artificiales para la secuencia del gen *bar*.

	A	C	G	T	GC:AT
BAR1_AT	139	130	144	139	1,01
BAR1_ZM	122	156	154	120	1,28

### Ejemplo 6

Este ejemplo sirve para ilustrar construcciones de ADN para la expresión de los polinucleótidos artificiales en plantas. Un módulo de expresión en plantas de ADN transgénico comprende elementos reguladores que controlan la transcripción de un ARNm a partir del módulo. Se construye un módulo de expresión en plantas para que incluya un promotor que actúe en plantas que este operablemente unido a una región conductora 5' que está operablemente unida a una secuencia de ADN de interés que está operablemente unida a una región de



terminación 3'. Estos módulos se construyen en vectores plasmídicos, que después pueden transferirse hacia el interior de las plantas mediante procedimientos de transformación mediados por *Agrobacterium* u otros procedimientos conocidos por los expertos en la técnica de la transformación de plantas. Se ilustran las siguientes construcciones de vectores plasmídicos para proporcionar ejemplos de plásmidos que contienen módulos de expresión en plantas que comprenden las moléculas polinucleotídicas artificiales.

Se sintetizan las moléculas polinucleotídicas artificiales, por ejemplo, CP4EPSPS\_AT y CP4EPSPS\_ZM utilizando cebadores solapantes. El producto de longitud completa después se amplifica con cebadores específicos de genes que contienen proyecciones con SphI (cebador directo) y EcoRI (cebador inverso). Los genes se clonan en el vector pCRII-TOPO (Invitrogen, CA). Los plásmidos resultantes pMON54949 (CP4EPSPS\_AT, figura 6) y pMON54950 (CP4EPSPS\_ZM, figura 7) contienen el polinucleótido artificial, y estos polinucleótidos se secuencian utilizando procedimientos de secuenciación de ADN para confirmar que las modificaciones están contenidas en los polinucleótidos artificiales. En la siguiente etapa, el polinucleótido artificial que codifica el péptido de tránsito del cloroplasto CTP2 se acopla al extremo 5' de los polinucleótidos CP4EPSPS. El promotor CaMV 35S con un potenciador duplicado (P-CaMVe35S) y un intrón de actina 1 del arroz (I-OsAct1) derivado de pMON30151 (figura 8) mediante una digestión con SphI y HindIII se acopla en los polinucleótidos CTP2CP4EPSPS para crear los plásmidos pMON59302 (CTP2CP4EPSPS\_AT, figura 9) y pMON59307 (CTP2CP4EPSPS\_ZM, figura 10).

Para la expresión de los nuevos polinucleótidos artificiales en plantas monocotiledóneas, los genes se colocan en módulos de expresión en plantas que contienen en el extremo 5' del polinucleótido un promotor y un intrón, una región 5' no traducida, y en el extremo 3' del polinucleótido una señal de fin de la transcripción. Para este objetivo, pMON42411 (figura 11) que contiene P-CaMV35S:en, I-HSP70, CTP2CP4\_Nat y NOS 3' se digiere con las enzimas de restricción NcoI y EcoRI. Los plásmidos pMON59302 (figura 9) y pMON59307 (figura 10) se digieren con las mismas enzimas de restricción. Los fragmentos se purifican en gel utilizando el kit de purificación en gel Qiagen y se acoplan para formar pMON58400 (CP4EPSPS\_AT, figura 12) y pMON58401 (CP4EPSPS\_ZM, figura 13). Se prepara otro vector, pMON54964 (figura 14), que contiene P-OsAct1/I-OsAct1 reemplazando P-e35S/I-Hsp70 de pMON58400 (figura 12) utilizando el fragmento HindIII/NcoI de pMON25455 (figura 15). Para crear un vector de expresión de monocotiledóneas que contiene el promotor P-FMV, pMON30152 (figura 16) se digiere con NheI, se crean extremos romos con ADN polimerasa de T4 en presencia de 4 dNTP (200  $\mu$ M) y NcoI. Se aíslan los fragmentos de ADN de CPT2CP4\_AT o CTP2CP4\_ZM a partir de pMON59302 (figura 9) o pMON59307 (figura 10), respectivamente, mediante una digestión con EcoRI, la formación de extremos romos con ADN polimerasa de T4 y una digestión con NcoI. Los fragmentos de ADN purificados en gel se acoplan y se crean los nuevos plásmidos pMON54992 (CTP2CP4\_AT, figura 17) y pMON54985 (CTP2CP4\_ZM, figura 18). En cada caso, el éxito de la construcción del plásmido se confirma mediante una digestión con endonucleasas de restricción utilizando, entre otras, ClaI (introducida en ambos polinucleótidos artificiales) y Pst I (introducida en CP4EPSPS\_ZM). El CP4EPSPS\_Nat presente en los vectores de origen tiene los sitios de restricción ClaI y dos PstI en la región codificadora en diferentes emplazamientos que en los polinucleótidos artificiales.

Para la expresión de los polinucleótidos CP4EPSPS artificiales en plantas dicotiledóneas, se emplean dos vectores de origen: pMON20999 (P-FMV/CTP2CP4\_Syn/3'E-9, figura 19) y pMON45313 (P-e35S/CTP2CP4\_Syn/3'E9, figura 20). En cada plásmido, un fragmento de ADN que contiene el polinucleótido CTP2CP4\_Syn se reemplaza por CTP2CP4\_AT o CTP2CP4\_ZM. Para crear pMON59308 (P-CaMVe35S/CTP2CP4\_AT, figura 21) o pMON59309 (P-CaMVe35S/CTP2CP4\_ZM, figura 22), el plásmido pMON45313 se digiere con NcoI y EcoRI, y los fragmentos de restricción de ADN derivados de la digestión con NcoI/EcoRI de pMON59302 (CTP2CP4\_AT, figura 9) o pMON59307 (CTP2CP4\_ZM, figura 10) se acoplan, respectivamente. Para crear pMON59313 (P-FMV/CTP2CP4\_AT/3'E9, figura 23) y pMON59396 (P-FMV/CTP2CP4\_ZM/3'E9, figura 24), el plásmido de origen pMON20999 se digiere con NcoI y BamHI para eliminar CTP2CP4\_Syn, y los fragmentos de restricción NcoI/BamHI derivados de pMON59308 (CTP2CP4\_AT, figura 21) o pMON59309 (CTP2CP4\_ZM, figura 22) se acoplan, respectivamente.

### Ejemplo 7

Los polinucleótidos artificiales se ensayan para determinar su eficacia para conferir a plantas transgénicas tolerancia al glifosato. Se transforman en maíz cinco módulos de expresión diferentes (tabla 20) con los nuevos polinucleótidos artificiales CP4EPSPS artificiales, y las plantas de maíz transgénicas resultantes se comparan con el patrón comercial (maíz Roundup Ready® 603, Monsanto Co.). El plásmido pMON25496 (figura 25) contenido en el patrón comercial tiene dos copias del polinucleótido CP4EPSPS\_Nat, con la expresión dirigida por los promotores P-CaMVe35S (P-CaMVe35S) y P-OsAct1, respectivamente. Los plásmidos que contienen los nuevos polinucleótidos CP4EPSPS artificiales solo contienen una copia del polinucleótido que se va a ensayar. La expresión de estos polinucleótidos está dirigida por el promotor P-CaMVe35S con el intrón de la proteína de choque térmico I-Hsp70 o el promotor P-FMV con un intrón de sacarosa sintasa del arroz (I-OsSS). El plásmido pMON54964 contiene el promotor de actina 1 del arroz con un primer intrón nativo (patente de EEUU n.º 5.641.876). Estos plásmidos se transforman en células de maíz mediante un procedimiento mediado por

5 *Agrobacterium* y líneas de maíz transgénico regeneradas para la selección con glifosato. Las plantas de maíz transgénico pueden producirse mediante un procedimiento de transformación mediado por *Agrobacterium*. Se emplea la cepa desactivada C58 de *Agrobacterium* (ABI) que porta una construcción binaria. Esta se transfiere a *Agrobacterium* mediante un procedimiento de cruzamiento triparental (Ditta *et al.*, Proc. Natl. Acad. Sci., 77:7347-7351). Los cultivos líquidos de *Agrobacterium* se inician a partir de disoluciones madre de glicerol, o a partir de una placa recién sembrada en estrías y cultivada durante la noche a 26 °C-28 °C con agitación (aproximadamente 150 rpm) hasta la fase de crecimiento semilogarítmica en medio LB líquido, pH 7,0, que contiene los antibióticos apropiados. Las células de *Agrobacterium* se resuspenden en el medio de inoculación (CM4C líquido), y la densidad se ajusta a una DO<sub>660</sub> de 1. Embriones de maíz HillxLH198 y Hill inmaduros de tipo II recién aislados se inoculan con *Agrobacterium* que contiene una construcción, y se cocultivan varios días en la oscuridad a 23 °C. Los embriones entonces se transfieren a un medio de retraso y se incuban a 28 °C durante varios días o más. Todos los cultivos posteriores se mantienen a esta temperatura. Los embriones se trasladan a un primer medio de selección que contiene carbenicilina 500/glifosato 0,5 mM. Dos semanas después, el tejido superviviente se traslada a un segundo medio de selección que contiene carbenicilina 500/glifosato 1,0 mM. Se subcultivan los callos supervivientes cada 2 semanas hasta que pueden identificarse acontecimientos. Esto puede llevar aproximadamente 3 subcultivos sobre glifosato 1,0 mM. Tras haber identificado los acontecimientos, se deja aumentar el volumen del tejido para la regeneración. Las plántulas se trasladan a medio MSOD en un recipiente de cultivo y se mantienen durante dos semanas. Después las plantas con raíces se trasladan a tierra. Los expertos en la técnica de la transformación del maíz pueden modificar este procedimiento para proporcionar plantas de maíz transgénicas sustancialmente idénticas que contienen las composiciones de ADN deseadas.

20 Se ensayan aproximadamente 30 líneas de maíz transgénico para cada construcción de plásmido, y en la tabla 20 se muestran la eficacia de transformación y los niveles de expresión de la enzima CP4EPSPS. Las líneas de maíz transgénico se tratan con glifosato a una proporción 4483,2 gr/hectárea (64 oz/acre) cuando son plantas jóvenes, y las plantas supervivientes se ensayan mediante ELISA de CP4EPSPS (Padgett *et al.*, Crop Sci., 35:1451-1461, 1995) para determinar los niveles de expresión de la proteína de CP4EPSPS (% de exp. de CP4) mostrados en la tabla 20, y el nivel de expresión se compara con una planta de maíz tolerante al glifosato patrón disponible en el mercado (maíz Roundup Ready® 603, Monsanto Co.) como porcentaje de la cantidad de expresión de proteínas, determinado en el patrón comercial. En general, más de 50% de las líneas de maíz sobreviven a una pulverización con 4483,2 gr/hectárea (64 oz/acre) de glifosato. Las plantas supervivientes han demostrado tener un alto nivel de expresión de CP4EPSP que varía de aproximadamente 75% a 86% del patrón comercial 603.

Tabla 20. Eficacia de transformación (TE) y expresión de CP4 (% promedio) derivadas de la transformación de diferentes construcciones de CP4-alt.

pMON	n.º de promotor/intrón	TE (%)	exp. de CP4 (%)*
58400 (CP4_AT)	P-CaMVe35S/IHsp70	5,4	75,5
58401 (CP4_ZM)	P-CaMVe35S/IHsp70	7,2	84,7
54964 (CP4_AT)	P-OsAct1	8,2	78,1
54985 (CP4_ZM)	P-FMV/IOsSS	11,5	85,7
54992 (CP4_AT)	P-FMV/IOsSS	11,5	78,2
nk603 (control)	P-OsAct1/P-e35S	-	100

\* La expresión de CP4EPSP se calcula como porcentaje del control (603) realizado en plantas que han sobrevivido a la pulverización con glifosato (4483,2 gr/hectárea (64 oz/acre)).

### Ejemplo 8

35 Se evaluaron tres construcciones plasmídicas en plantas de algodón transgénico (tabla 21). La construcción control (pMON20999) contiene P-FMV/CP4EPSPS\_Syn. Este módulo de expresión está contenido en la línea de algodón tolerante al glifosato disponible en el mercado 1445 (algodón Roundup Ready®, Monsanto Co., St. Louis, MO). Las construcciones de plásmidos, pMON59313 y pMON59396, que contienen los polinucleótidos CP4EPSPS\_AT y CP4EPSPS\_ZM, respectivamente, se ensayan para la eficacia de transformación y los niveles de la enzima CP4EPSPS con relación al módulo de expresión tolerante al glifosato comercial. Se evaluaron aproximadamente 50 líneas de algodón transgénico para cada construcción. El polinucleótido artificial CP4EPSPS\_AT dirigido por el promotor P-FMV produce un mayor porcentaje de plantas con una única inserción, y

un aumento en el nivel de expresión de la enzima CP4EPSPS con relación al módulo de expresión pMON20999, según se mide mediante ELISA.

Tabla 21. Eficacia de transformación (TE) y expresión media de de CP4EPSPS en líneas de algodón R0 derivadas de la transformación de diferentes construcciones de CP4EPSPS.

pMON	Promotor	TE (%)	exp. de CP4 (%)
20999 (CP4EPSPS_Syn)	P-FMV	15,0	100,0
59313 (CP4EPSPS_AT)	P-FMV	15,0	116,4
59396 (CP4EPSPS_ZM)	P-FMV	16,1	52,0

5

### Ejemplo 9

Se evalúan construcciones que contienen los polinucleótidos artificiales de CP4EPSPS, CP4EPSPS\_AT y CP4EPSPS\_ZM en la soja (tabla 22). Las construcciones de plásmidos contienen el promotor P-FMV para dirigir la expresión de los nuevos polinucleótidos de CP4EPSPS y se comparan con P-FMV/CP4EPSPS\_Syn contenido en pMON20999. Se producen aproximadamente 25 a 30 plantas de soja transgénicas para cada construcción. Se mide la eficacia de la transformación y los niveles de la enzima CP4EPSPS. Se mide un nivel de expresión sorprendentemente alto de proteína CP4EPSPS en plantas de soja que contienen la secuencia codificadora de CP4EPSPS\_ZM (tabla 22).

10

Tabla 22. Eficacia de transformación (TE) y expresión media de de CP4 derivadas de la transformación de diferentes construcciones de CP4EPSPS.

15

pMON	Promotor	TE (%)	exp. de CP4 (%)
20999 (CP4_Syn)	P-FMV	0,55	100,0
59313 (CP4_AT)	P-FMV	0,40	66,6
54996 (CP4_ZM)	P-FMV	0,29	242,5

### Ejemplo 10

Se transforman células de tabaco con tres construcciones plasmídicas que contienen diferentes secuencias de polinucleótidos de CP4EPSPS y se regeneran en plantas. Se evalúan aproximadamente 20 líneas transgénicas para cada construcción. La expresión de cada uno de los polinucleótidos de CP4EPSPS es dirigida por el promotor potenciador duplicado P-CaMVe35S (tabla 23). Se mide la eficacia de la transformación y los niveles de expresión de la enzima CP4EPSPS. Se demuestra que las diferentes construcciones de polinucleótidos de CP4EPSPS actúan aproximadamente igual en el tabaco transgénico con respecto a la eficacia de la transformación y la expresión.

20

Tabla 23. Eficacia de transformación (TE) y expresión media de de CP4 en líneas de tabaco R0 derivadas de la transformación de diferentes construcciones de CP4EPSPS.

25

pMON	Promotor	TE (%)	exp. de CP4 (%)
59308 (CP4EPSPS_AT)	P-CaMVe35S	35	100,0
59309 (CP4EPSPS_ZM)	P-CaMVe35S	35	91,0
54313 (CP4EPSPS_Syn)	P-CaMVe35S	35	100,0

### Ejemplo 11

Se transformó *Arabidopsis thaliana* con cuatro construcciones de plásmidos mediante infiltración al vacío (Bechtold N., *et al.*, CR Acad. Sci. Paris Sciences de la vie/life sciences, 316:1194-1199, (1993)), y la progenie V1 se evaluó para comparar la eficacia de las diferentes secuencias de polinucleótidos de CP4EPSPS y diferentes promotores

30

5 para su uso en la selección de plantas con glifosato (tabla 24). Se produjeron aproximadamente 30 plantas V1 transgénicas (+) para cada construcción. Las construcciones dirigidas por P-CaMVe35S con el potenciador duplicado (pMON45313, pMON59308, y pMON59309) no muestran una diferencia sustancial en el nivel de expresión en las hojas, según se determina mediante ELISA. Las plantas se transforman con pMON26140 que contiene CP4EPSPS\_Syn dirigido por el promotor P-FMV; estas plantas muestran el nivel de expresión más alto. Los niveles de expresión detectados en las plantas de las construcciones de ensayo se comparan con pMON26140.

Tabla 24. Evaluación de los diferentes módulos de expresión de CP4 en *Arabidopsis*.

pMON	Promotor	Plantas producidas	exp. de CP4 (%)
45313 (CP4EPSPS_Syn)	P-CaMVe35S	+	82,1
59308 (CP4EPSPS_AT)	P-CaMVe35S	+	79,3
59309 (CP4EPSPS_ZM)	P-CaMVe35S	+	77,3
26140 (CP4EPSPS_Syn)	P-FMV	+	100,0

10 **Ejemplo 12**

Se transformaron plantas de trigo con los nuevos polinucleótidos de CP4EPSPS para la eficacia de transformación y la expresión de la enzima CP4EPSPS, según se determina mediante ELISA (tabla 25). El CP4EPSPS\_ZM proporciona al menos siete veces más expresión de la enzima CP4EPSPS que CP4EPSPS\_AT. La media de la expresión de CP4EPSPS en las hojas de las plantas de trigo que contienen el polinucleótido CP4EPSPS\_ZM es aproximadamente 64% de la que se encuentra en el trigo resistente al glifosato que contiene una construcción de módulo doble, pMON30139: P-e35S/I-Hsp70/CP4EPSPS\_Nat y P-OsAct1/IOsAct1/CP4EPSPS\_Nat (documento WO/0022704).

Tabla 25. Actuación de diferentes polinucleótidos de CP4EPSPS en el trigo.

pMON	Promotor/Intrón	TE (%)	exp. de CP4 (%)
58400 (CP4EPSPS_AT)	P-e35S/I-Hsp70	0,25	9,2
58401 (CP4EPSPS_ZM)	P-e35S/I-Hsp70	0,35	64,0
30139 (CP4EPSPS_Nat)	P-e35S:P-OsAct1	-	100,0

20 **Ejemplo 13**

Este ejemplo sirve para ilustrar la detección de diferentes polinucleótidos artificiales en plantas transgénicas, de forma específica CP4EPSPS\_AT y CP4EPSPS\_ZM. Los otros polinucleótidos artificiales, OsEPSPS\_AT, OsEPSPS\_ZM, GmEPSPS\_GM, ZmEPSPS\_ZM, CTP2\_AT, CTP2\_ZM, Bar1\_AT y Bar1\_ZM, pueden ser detectados específicamente en plantas transgénicas mediante procedimientos que proporcionan una amplicón de ADN, o mediante la hibridación de una sonda de ADN con una muestra de planta. Los expertos en la técnica de la detección de ADN pueden diseñar con facilidad moléculas de cebador a partir de las secuencias polinucleotídicas artificiales para producir un procedimiento que detecte específicamente el polinucleótido artificial en una muestra de planta.

Se diseña un procedimiento de detección de ADN (reacción en cadena de polimerasa, PCR) para detectar los polinucleótidos de CP4EPSPS artificiales en plantas transgénicas. Los conjuntos exclusivos de cebadores de ADN mostrados en la tabla 26 se diseñan para amplificar un polinucleótido de CP4EPSPS específico y para proporcionar amplicones con un tamaño diferenciado. Los amplicones se diferencian suficientemente en la longitud del polinucleótido entre los diversos polinucleótidos de CP4EPSPS como para que sea fácil la separación de los amplicones mediante una electroforesis en gel de agarosa convencional. Puede detectarse la presencia de más de uno de los polinucleótidos artificiales en una planta utilizando un método de PCR de múltiplex.

Tabla 26. Secuencia de los cebadores utilizados para la detección de diferentes genes CP4 en plantas transgénicas.

Pareja de cebadores	Especificidad del gen	Producto de PCR (pb)
SEQ ID NO:24 y 25	CP4EPSPS_AT	938 (940)
SEQ ID NO:26 y 27	CP4EPSPS_ZM	595 (600)
SEQ ID NO:28 y 29	CP4EPSPS_Nat	712 (710)
SEQ ID NO:30 y 31	CP4EPSPS_Syn	443 (440)

5 Las parejas de cebadores de ADN (tabla 26) se emplean para producir un diagnóstico de amplicones para un polinucleótido de CP4EPSPS especificado contenido en una planta transgénica. Estas parejas de cebadores incluyen SEQ ID NO:24 y SEQ ID NO:25 para el polinucleótido CP4EPSPS\_AT; SEQ ID NO:26 y SEQ ID NO:27 para CP4EPSPS\_ZM; SEQ ID NO:28 y SEQ ID NO:29 para CP4EPSPS\_Nat; y SEQ ID NO:30 y SEQ ID NO:31 para la molécula polinucleotídica CP4EPSPS\_Syn. Además de estas parejas de cebadores, cualquier pareja de cebadores derivada de SEQ ID NO:17 o SEQ ID NO:18 que, cuando se usa en una reacción de amplificación de ADN, produce un diagnóstico de amplicones de ADN para el respectivo polinucleótido de CP4EPSPS, constituye un aspecto.

10 Las condiciones de amplificación para este análisis se ilustran en la tabla 27 y la tabla 28. Sin embargo, cualquier modificación de estas condiciones, que incluye el uso de fragmentos de las moléculas de ADN de SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:17, SEQ ID NO:18, o SEQ ID NO:35, o sus complementos, como moléculas de cebador, que producen un diagnóstico de amplicones de moléculas de ADN para los polinucleótidos artificiales descritos en la presente, está dentro de los conocimientos de la técnica. Las moléculas de ADN que actúan como moléculas de cebador en un procedimiento de amplificación de ADN para detectar la presencia de los anteriores polinucleótidos artificiales incluyen SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, y SEQ ID NO:31.

20 En un procedimiento para determinar la presencia de estos polinucleótidos, el análisis de una muestra de extracto de ADN de un tejido vegetal debería incluir un control positivo que se sabe que contenga los polinucleótidos artificiales, y un control de extracto de ADN negativo procedente de una planta que no sea transgénica o que no contenga el polinucleótido artificial, y un control negativo que no contenga un molde en el extracto de ADN.

25 Pueden seleccionarse otras moléculas de cebadores de ADN con una longitud suficiente a partir de SEQ ID NO:17 y SEQ ID NO:18, y pueden optimizarse las condiciones para la producción de un amplicón que puedan ser diferentes de los procedimientos mostrados en la tabla 27 y la tabla 28, pero que produzcan un diagnóstico de amplicones para los polinucleótidos artificiales. Estas secuencias de cebadores de ADN homólogas o complementarias con SEQ ID NO:17 y SEQ ID NO:18 pueden utilizarse con o sin modificaciones en los procedimientos de las tablas 27 y 28. El ensayo para el amplicón de CP4EPSPS\_AT y CP4EPSPS\_ZM puede realizarse utilizando un termociclador Stratagene Robocycler, MJ Engine, Perkin-Elmer 9700, o un termociclador de gradiente Eppendorf Mastercycler Gradient, tal como se muestra en la tabla 28, o mediante procedimientos y aparatos conocidos por los expertos en la técnica.

Tabla 27. Procedimiento de amplificación del ADN y mezcla de reacción para la confirmación de la presencia del polinucleótido de EPSPS artificial CP4EPSPS\_AT en plantas de maíz.

<b>Etapas</b>	<b>Reactivo</b>	<b>Cantidad</b>	<b>Comentarios</b>
1	agua sin nucleasa	añadir hasta un volumen final de 20 µl	-
2	10x tampón de reacción (con MgCl <sub>2</sub> )	2,0 µl	1x concentración final de tampón, 1,5 mM concentración final de MgCl <sub>2</sub>
3	disolución 10 mM de dATP, dCTP, dGTP, y dTTP	0,4 µl	concentración final de 200 µM de cada dNTP
4	cebador (SEQ ID NO:24) (resuspendido en 1x tampón TE o agua sin nucleasa hasta una concentración de 10 µM)	0,4 µl	concentración final de 0,2 µM

ES 2 420 929 T3

5	cebador (SEQ ID NO:25) (resuspendido en 1x tampón TE o agua sin nucleasa hasta una concentración de 10 µM)	0,4 µl	concentración final de 0,2 µM
6	cebador control (SEQ ID NO:32) (resuspendido en 1x tampón TE o agua sin nucleasa hasta una concentración de 10 µM)	0,2 µl	concentración final de 0,1 µM
7	cebador control (SEQ ID NO:33) (resuspendido en 1x tampón TE o agua sin nucleasa hasta una concentración de 10 µM)	0,2 µl	concentración final de 0,1 µM
8	ARNasa, sin ADNasa (500 ng/µl)	0,1 µl	50 ng/reacción
9	ADN polimerasa <i>REDTaq</i> (1 unidad/µl)	1,0 µl (se recomienda cambiar las pipetas antes de la siguiente reacción)	1 unidad/reacción
10	ADN extraído (molde): · Muestras que se van a analizar * hojas individuales * hojas reunidas (máx. de 50 hojas/grupo) · Control negativo · Control negativo · Control positivo	· 10-200 ng de ADN genómico · 200 ng de ADN genómico · 50 ng de ADN genómico de planta no transgénica · sin molde de ADN · 5 ng de ADN plasmídico	

Tabla 28. Parámetros de PCR sugeridos para diferentes termocicladores

Se mezcla con suavidad, si es necesario (no poner caliente en el termociclador), se añaden 1-2 gotas de aceite mineral sobre la parte superior de cada reacción. Se procede con la PCR en un termociclador Stratagene Robocycler, MJ Engine, Perkin-Elmer 9700, o un termociclador de gradiente Eppendorf Mastercycler Gradient utilizando los siguientes parámetros de ciclado.

5

Nota: El aparato MJ Engine o el termociclador Eppendorf Mastercycler Gradient deben hacerse funcionar en el modo calculado. Se hace funcionar el termociclador Perkin-Elmer 9700 con la velocidad de rampa ajustada al máximo.

Ciclo n.º	Ajustes: Stratagene Robocycler	
1	94 °C	3 minutos
38	94° C 60 °C 72 °C	1 minuto 1 minuto 1 minuto y 30 segundos
1	72 °C	10 minutos
Ciclo n.º	Ajustes: MJ Engine o Perkin-Elmer 9700	
1	94 °C	3 minutos
38	94° C 60 °C 72 °C	10 segundos 30 segundos 1 minuto

ES 2 420 929 T3

1	72 °C	10 minutos
<b>Ciclo n.º</b>	<b>Ajustes: Eppendorf Mastercycler Gradient</b>	
1	94 °C	3 minutos
38	94° C 60 °C 72 °C	15 segundos 15 segundos 1 minuto
1	72 °C	10 minutos

LISTADO DE SECUENCIAS

<110> Flasiński, Stanisław

<120> Procedimientos de utilización de polinucleótidos artificiales y sus composiciones para reducir el silenciamiento de transgenes.

5

<130> 11899.0235.00PC00

<140> documento US 60/396.665

<141> 18-07-2002

10

<150> documento US 06/396.665

<151> 18-07-2002

15

<160> 35

<170> PatentIn version 3.2

<210> 1

<211> 515

20

<212> PRT

<213> *Oryza sativa*

<400> 1

Met Ala Ala Thr Met Ala Ser Asn Ala Ala Ala Ala Ala Ala Val Ser  
1 5 10 15

Leu Asp Gln Ala Val Ala Ala Ser Ala Ala Phe Ser Ser Arg Lys Gln  
20 25 30

Leu Arg Leu Pro Ala Ala Ala Arg Gly Gly Met Arg Val Arg Val Arg  
35 40 45

Ala Arg Gly Arg Arg Glu Ala Val Val Val Ala Ser Ala Ser Ser Ser  
50 55 60

Ser Val Ala Ala Pro Ala Ala Lys Ala Glu Glu Ile Val Leu Gln Pro  
65 70 75 80

Ile Arg Glu Ile Ser Gly Ala Val Gln Leu Pro Gly Ser Lys Ser Leu  
85 90 95

Ser Asn Arg Ile Leu Leu Leu Ser Ala Leu Ser Glu Gly Thr Thr Val  
100 105 110

Val Asp Asn Leu Leu Asn Ser Glu Asp Val His Tyr Met Leu Glu Ala  
115 120 125

Leu Lys Ala Leu Gly Leu Ser Val Glu Ala Asp Lys Val Ala Lys Arg  
130 135 140



ES 2 420 929 T3

Ala Val Val Val Gly Cys Gly Gly Lys Phe Pro Val Glu Lys Asp Ala  
 145 150 155 160

Lys Glu Glu Val Gln Leu Phe Leu Gly Asn Ala Gly Ile Ala Met Arg  
 165 170 175

Ser Leu Thr Ala Ala Val Thr Ala Ala Gly Gly Asn Ala Thr Tyr Val  
 180 185 190

Leu Asp Gly Val Pro Arg Met Arg Glu Arg Pro Ile Gly Asp Leu Val  
 195 200 205

Val Gly Leu Lys Gln Leu Gly Ala Asp Val Asp Cys Phe Leu Gly Thr  
 210 215 220

Glu Cys Pro Pro Val Arg Val Lys Gly Ile Gly Gly Leu Pro Gly Gly  
 225 230 235 240

Lys Val Lys Leu Ser Gly Ser Ile Ser Ser Gln Tyr Leu Ser Ala Leu  
 245 250 255

Leu Met Ala Ala Pro Leu Ala Leu Gly Asp Val Glu Ile Glu Ile Ile  
 260 265 270

Asp Lys Leu Ile Ser Ile Pro Tyr Val Glu Met Thr Leu Arg Leu Met  
 275 280 285

Glu Arg Phe Gly Val Lys Ala Glu His Ser Asp Ser Trp Asp Arg Phe  
 290 295 300

Tyr Ile Lys Gly Gly Gln Lys Tyr Lys Ser Pro Gly Asn Ala Tyr Val  
 305 310 315 320

Glu Gly Asp Ala Ser Ser Ala Ser Tyr Phe Leu Ala Gly Ala Ala Ile  
 325 330 335

Thr Gly Gly Thr Val Thr Val Gln Gly Cys Gly Thr Thr Ser Leu Gln  
 340 345 350

Gly Asp Val Lys Phe Ala Glu Val Leu Glu Met Met Gly Ala Lys Val  
 355 360 365

Thr Trp Thr Asp Thr Ser Val Thr Val Thr Gly Pro Pro Arg Glu Pro  
 370 375 380

ES 2 420 929 T3

Tyr Gly Lys Lys His Leu Lys Ala Val Asp Val Asn Met Asn Lys Met  
385 390 395 400

Pro Asp Val Ala Met Thr Leu Ala Val Val Ala Leu Phe Ala Asp Gly  
405 410 415

Pro Thr Ala Ile Arg Asp Val Ala Ser Trp Arg Val Lys Glu Thr Glu  
420 425 430

Arg Met Val Ala Ile Arg Thr Glu Leu Thr Lys Leu Gly Ala Ser Val  
435 440 445

Glu Glu Gly Pro Asp Tyr Cys Ile Ile Thr Pro Pro Glu Lys Leu Asn  
450 455 460

Ile Thr Ala Ile Asp Thr Tyr Asp Asp His Arg Met Ala Met Ala Phe  
465 470 475 480

Ser Leu Ala Ala Cys Ala Asp Val Pro Val Thr Ile Arg Asp Pro Gly  
485 490 495

Cys Thr Arg Lys Thr Phe Pro Asn Tyr Phe Asp Val Leu Ser Thr Phe  
500 505 510

Val Arg Asn  
515

- <210> 2
- <211> 1548
- <212> ADN
- <213> *Oryza sativa*

5

<400> 2  
atggcgggcga ccatggcgctc caacgccgcg gctgcggcgg cggtgtccct ggaccaggcc 60  
gtggcgggcgt cggcgggcgtt ctcgctcgcg aagcagctgc ggctgcccgc cgcggcgcg 120  
gggggggatgc gggtgccgggt gcggggcgcg gggcgggcgg aggcgggtgt ggtggcgctcc 180  
gcgtcgctcgt cgtcgggtggc agcgcggcg gccaaggcgg aggagatcgt gctccagccc 240  
atcagggaga tctccggggc ggttcagctg ccaggggtcca agtcgctctc caacaggatc 300  
ctctctctct ccgccctctc cgagggcaca acagtgggtg acaacttget gaacagtgag 360  
gatgttcaact acatgcttga ggcctgaaa gcctcgggc tctctgtgga agcagataaa 420  
gttgcaaaaa gagctgtagt cgttggtgt ggtggcaagt ttctgttga gaaggatgcg 480  
aaagaggaag tgcaactctt cttggggaac gctggaatcg caatgcatc cttgacagca 540  
gccgtgactg ctgctgggtg aatgcaact tatgtgcttg atggagtgcc acgaatgagg 600

ES 2 420 929 T3

gagagaccga ttggtgactt ggttgtcggg ttgaaacaac ttggtgcgga tgtcgactgt 660  
 ttccttggca ctgaatgcc acctgttcgt gtcaagggaa ttggaggact tccctgggtggc 720  
 aaggttaagc tctctggttc catcagcagt cagtacttga gtgccttgct gatggctgct 780  
 cctttggccc ttggggatgt ggagatcgaa atcattgaca aactaatctc cattccttac 840  
 gttgaaatga cattgagatt gatggagcgt tttggtgtga aggcagagca ttctgatagt 900  
 tgggacagat tctatattaa gggagggcag aagtacaaat ctctggaaa tgcctatggt 960  
 gaaggatgat cctcaagcgc gagctatttc ttggctgggt ctgcaatcac tggaggcact 1020  
 gtgacagttc aaggttgtgg tacgaccagt ttgcagggtg atgtcaaatt tgctgaggta 1080  
 cttgagatga tgggagcaaa ggttacatgg actgacacca gtgtaaccgt aactggtcca 1140  
 ccacgtgagc cttatgggaa gaaacacctg aaagctgttg atgtcaacat gaacaaaatg 1200  
 cctgatgttg ccatgacct tgccgttggt gcactcttcg ctgatgggtcc aactgctatc 1260  
 agagatgtgg cttcctggag agtaaaggaa accgaaagga tggttgcaat tcggaccgag 1320  
 ctaacaaagc tgggagcadc ggttgaagaa ggtcctgact actgcatcat caccaccg 1380  
 gagaagctga acatcacggc aatcgacacc tacgatgatc acaggatggc catggccttc 1440  
 tccctcgctg cctgcgccga cgtgcccggt acgatcaggg accctgggtg caccgcaag 1500  
 accttcccc actacttoga cgttctaagc accttcgtca ggaactga 1548

<210> 3  
 <211> 1548  
 <212> ADN  
 <213> *Oryza sativa*

5

<400> 3  
 atggctgcaa ctatggctag taacgcagcg gctgccgctg ccgtttcctt agaccaagca 60  
 gtagcagcga gcgctgcatt ctcatcacgt aagcaactac ggctaccagc agccgctaga 120  
 ggccggcatga gagttagagt gagggctaga ggtaggcggg aggetgtagt cgtagcctcc 180  
 gcttctagca gttcgggtggc tgcgccggct gctaaggcag aggagattgt ttacaacct 240  
 attagggaaa tatcgggggc cgtacaatta cctggaagca agagcctttc caacaggatt 300  
 ctgttgcttt cagctctctc ggaggaaca acagttgtgg ataactctgtt gaatagtgag 360  
 gatgtgcact atatgctaga ggctctcaag gctctagggc tttctgtaga agcggataaa 420  
 gtagcaaac gcgcagtggg ttaggttgt ggtgggaagt tcccagttga aaaggatgct 480  
 aaggaagaag tacagctctt tctcgggaat gccgggatcg ccatgcggag tttgactgct 540  
 gcggtcacag ccgctggagg caacgcaaca tacgtcctag atgggggtgcc gagaatgcgt 600  
 gagcgtccta ttggtgatct tgccttaggt ctcaagcaac tcggcgctga cgtagattgt 660

ES 2 420 929 T3

ttcctgggta ctgagtgtcc gccagtcaga gttaaaggaa tcggtgggct gccgggcgga 720  
aaggtcaagc tgtcgggcag tatttcgagt cagtatcttt ctgctctcct gatggctgcg 780  
ccattagctt tgggagatgt tgagatcgag atcattgata aacttatatc tatcccgtat 840  
gtcgagatga ctttaagact tatggaacgg tttgggggta aggccgagca tagcgacagt 900  
tgggatcgtt totacataaa gggaggccag aagtataagt ctctgggaa tgcttatgta 960  
gaaggggatg cttcatctgc gtcttacttc cttgocgggag cggctataac tggaggaaca 1020  
gtcacagttc agggctgocg tacaacaagt ttgcaagggtg acgtgaagtt tgccgaggta 1080  
cttgaaatga tgggtgccaa agtaacgtgg acagacacat cggtgacagt tactggtcct 1140  
ccacgagaac cttacggcaa aaagcatctt aaggccgtgg atgttaatat gaataagatg 1200  
cctgacgttg ctatgacact tgccgttggt gccctttttg cagacggccc aacggcgata 1260  
cgcgatgttg catcatggcg cgtcaaggaa acggagagga tgggtggctat tcgaactgaa 1320  
ctcacciaac ttggtgcctc tgtagaggag ggccttgatt actgtatcat tacaccccct 1380  
gagaaactta acatcactgc tattgataca tacgacgac atagaatggc tatggctttc 1440  
tcaactggccg cttgtgcaga tgttctctgc acaatcagag atcctggctg tactagaaag 1500  
acgttcccga actactttga tgttctttca acattcgtcc gcaattga 1548

<210> 4  
<211> 1548  
<212> ADN  
<213> *Oryza sativa*

5

<400> 4  
atggctgcaa ctatggctag taacgcagcg gctgccgctg ccgtttcctt agaccaagca 60  
gtagcagcga gcgctgcatt ctcatcacgt aagcaactac ggctaccagc agccgctaga 120  
ggcggcatga gagttagagt gagggctaga ggtaggcggg aggctgtagt cgtagcctcc 180  
gcttctagca gttcgggtggc tgcgccggct gctaaggcag aggagattgt tttacaacct 240  
attagggaaa tatcgggggc cgtacaatta cctggaagca agagcctttc caacaggatt 300  
ctgttgcttt cagctctctc ggaggaaca acagttgtgg ataatctggt gaatagtgag 360  
gatgtgcaat atatgctaga ggctctcaag gctctagggc tttctgtaga agcggataaa 420  
gtagcaaac gcgcagtggt tgtaggttgt ggtgggaagt tcccagttga aaaggatgct 480  
aaggaagaag tacagctctt tctcgggaat gccgggatcg ccatgcggag tttgactgct 540  
gcggtcacag ccgctggagg caacgcaaca tacgtcctag atgggggtgcc gagaatgcgt 600  
gagcgtccta ttggtgatct tgtcgtaggt ctcaagcaac tcggcgctga cgtagattgt 660  
ttcctgggta ctgagtgtcc gccagtcaga gttaaaggaa tcggtgggct gccgggcgga 720

ES 2 420 929 T3

aagggtcaagc tgtcgggcag tatttcgagt cagtatcttt ctgctctcct gatggctgcg 780  
 ccattagctt tgggagatgt tgagatcgag atcattgata aacttatatc tatcccgtat 840  
 gtcgagatga ctttaagact tatggaacgg tttgggggtta aggccgagca tagcgacagt 900  
 tgggatcgtt tctacataaa gggaggccag aagtataagt ctctctgggaa tgcttatgta 960  
 gaaggggatg cttcatctgc gtcttacttc cttgcgggag cggctataac tggaggaaca 1020  
 gtcacagttc agggctgocg tacaacaagt ttgcaagggtg acgtgaagtt tgccgaggta 1080  
 cttgaaatga tgggtgccaa agtaacgtgg acagacacat cggtgacagt tactggctct 1140  
 ccacgagaac cttacggcaa aaagcatctt aaggccgtgg atgttaatat gaataagatg 1200  
 cctgacgttg ctatgacact tgccgttggt gccctttttg cagacggccc aacggcgata 1260  
 cgcgatgttg catcatggcg cgtcaaggaa acggagagga tgggtggctat tcgaactgaa 1320  
 ctcaccaaac ttggtgcctc tgtagaggag ggccctgatt actgtatcat tacaccccct 1380  
 gagaaactta acatcactgc tattgatata tacgacgac atagaatggc tatggctttc 1440  
 tcaactggccg cttgtgcaga tgttctctgc acaatcagag atcctggctg tactagaaag 1500  
 acgttcccg actactttga tgttctttca acattcgtcc gcaattga 1548

<210> 5  
 <211> 525  
 <212> PRT  
 <213> *Glycine max*

5

<400> 5  
 Met Ala Gln Val Ser Arg Val His Asn Leu Ala Gln Ser Thr Gln Ile  
 1 5 10 15  
 Phe Gly His Ser Ser Asn Ser Asn Lys Leu Lys Ser Val Asn Ser Val  
 20 25 30  
 Ser Leu Arg Pro Arg Leu Trp Gly Ala Ser Lys Ser Arg Ile Pro Met  
 35 40 45  
 His Lys Asn Gly Ser Phe Met Gly Asn Phe Asn Val Gly Lys Gly Asn  
 50 55 60  
 Ser Gly Val Phe Lys Val Ser Ala Ser Val Ala Ala Ala Glu Lys Pro  
 65 70 75 80  
 Ser Thr Ser Pro Glu Ile Val Leu Glu Pro Ile Lys Asp Phe Ser Gly  
 85 90 95

ES 2 420 929 T3

Thr Ile Thr Leu Pro Gly Ser Lys Ser Leu Ser Asn Arg Ile Leu Leu  
100 105 110

Leu Ala Ala Leu Ser Glu Gly Thr Thr Val Val Asp Asn Leu Leu Tyr  
115 120 125

Ser Glu Asp Ile His Tyr Met Leu Gly Ala Leu Arg Thr Leu Gly Leu  
130 135 140

Arg Val Glu Asp Asp Lys Thr Thr Lys Gln Ala Ile Val Glu Gly Cys  
145 150 155 160

Gly Gly Leu Phe Pro Thr Ser Lys Glu Ser Lys Asp Glu Ile Asn Leu  
165 170 175

Phe Leu Gly Asn Ala Gly Ile Ala Met Lys Ser Leu Thr Ala Ala Val  
180 185 190

Val Ala Ala Gly Gly Asn Ala Ser Tyr Val Leu Asp Gly Val Pro Arg  
195 200 205

Met Arg Glu Arg Pro Ile Gly Asp Leu Val Ala Gly Leu Lys Gln Leu  
210 215 220

Gly Ala Asp Val Asp Cys Phe Leu Gly Thr Asn Cys Pro Pro Val Arg  
225 230 235 240

Val Asn Gly Lys Gly Gly Leu Pro Gly Gly Lys Val Lys Leu Ser Gly  
245 250 255

Ser Val Ser Ser Gln Tyr Leu Thr Ala Leu Leu Met Ala Ala Pro Leu  
260 265 270

Ala Leu Gly Asp Val Glu Ile Glu Ile Val Asp Lys Leu Ile Ser Val  
275 280 285

Pro Tyr Val Glu Met Thr Leu Lys Leu Met Glu Arg Phe Gly Val Ser  
290 295 300

Val Glu His Ser Gly Asn Trp Asp Arg Phe Leu Val His Gly Gly Gln  
305 310 315 320

Lys Tyr Lys Ser Pro Gly Asn Ala Phe Val Glu Gly Asp Ala Ser Ser  
325 330 335

Ala Ser Tyr Leu Leu Ala Gly Ala Ala Ile Thr Gly Gly Thr Ile Thr

ES 2 420 929 T3

	340		345		350															
Val	Asn	Gly	Cys	Gly	Thr	Ser	Ser	Leu	Gln	Gly	Asp	Val	Lys	Phe	Ala					
		355					360					365								
Glu	Val	Leu	Glu	Lys	Met	Gly	Ala	Lys	Val	Thr	Trp	Ser	Glu	Asn	Ser					
		370				375					380									
Val	Thr	Val	Ser	Gly	Pro	Pro	Arg	Asp	Phe	Ser	Gly	Arg	Lys	Val	Leu					
		385			390					395					400					
Arg	Gly	Ile	Asp	Val	Asn	Met	Asn	Lys	Met	Pro	Asp	Val	Ala	Met	Thr					
				405					410					415						
Leu	Ala	Val	Val	Ala	Leu	Phe	Ala	Asn	Gly	Pro	Thr	Ala	Ile	Arg	Asp					
			420					425					430							
Val	Ala	Ser	Trp	Arg	Val	Lys	Glu	Thr	Glu	Arg	Met	Ile	Ala	Ile	Cys					
		435					440					445								
Thr	Glu	Leu	Arg	Lys	Leu	Gly	Ala	Thr	Val	Glu	Glu	Gly	Pro	Asp	Tyr					
		450				455					460									
Cys	Val	Ile	Thr	Pro	Pro	Glu	Lys	Leu	Asn	Val	Thr	Ala	Ile	Asp	Thr					
					470					475					480					
Tyr	Asp	Asp	His	Arg	Met	Ala	Met	Ala	Phe	Ser	Leu	Ala	Ala	Cys	Gly					
				485					490					495						
Asp	Val	Pro	Val	Thr	Ile	Lys	Asp	Pro	Gly	Cys	Thr	Arg	Lys	Thr	Phe					
			500					505					510							
Pro	Asp	Tyr	Phe	Glu	Val	Leu	Glu	Arg	Leu	Thr	Lys	His								
		515					520					525								

<210> 6  
 <211> 1578  
 <212> ADN  
 <213> *Glycine max*

5

<400> 6	atggccaag tgagcagagt gcacaatctt gctcaaagca ctcaaatttt tggccattct	60
	tccaactcca acaaactcaa atcgggtgaat tccggtttcat tgaggccacg cctttggggg	120
	gcctcaaaaat ctgcatccc gatgcataaa aatggaagct ttatgggaaa ttttaatgtg	180
	gggaagggaa attccggcgt gtttaaggtt tctgcatcgg tcgccgccgc agagaagccg	240

ES 2 420 929 T3

tcaacgtcgc cggagatcgt gttggaaccc atcaaagact tctcgggtac catcacattg 300  
ccagggtcca agtctctgtc caatcgaatt ttgcttcttg ctgctctctc tgagggaaca 360  
actgttgtag acaacttggt gtatagttag gatattcatt acatgcttgg tgcattaagg 420  
acccttggac tgcgtgtgga agatgacaaa acaaccaaac aagcaattgt tgaaggctgt 480  
gggggattgt ttcccactag taaggaatct aaagatgaaa tcaatttatt ccttggaaat 540  
gctggtatcg caatgaagtc cttgacagca gctgtggttg ctgcagggtg aatgcaagc 600  
tacgtacttg atggggtgcc ccgaatgaga gagaggccaa ttggggattt ggttgctggt 660  
cttaagcaac ttggtgcaga tgttgattgc tttcttggca caaactgtcc acctgttctg 720  
gtaaattggga agggaggact tcctggcgga aaggtgaaac tgtctggatc agttagcagt 780  
caatacttga ctgctttgct tatggcagct cctttagctc ttggtgatgt ggaaattgag 840  
attgttgata aactgatttc tgttccatat gttgaaatga ctctgaagtt gatggagcgt 900  
tttgagttt ctgtggaaca cagtggtaat tgggataggt tcttgggtcca tggaggtcaa 960  
aagtacaagt ctcttgcaa tgcttttggt gaaggtgatg cttcaagtgc cagttattta 1020  
ctagctggtg cagcaattac tgggtgggact atcactgtta atggctgtgg cacaagcagt 1080  
ttacagggag atgtaaaatt tgctgaagtt cttgaaaaga tgggagctaa ggttacatgg 1140  
tcagagaaca gtgtcactgt ttctggacca ccacgagatt tttctggctg aaaagtcttg 1200  
cgaggcattg atgtcaatat gaacaagatg ccagatggtg ccatgacact tgctgttgtt 1260  
gcactatttg ctaatggtcc cactgctata agagatgtgg caagttggag agttaaagag 1320  
actgagagga tgatagcaat ctgcacagaa ctcagaaagc taggagcaac agttgaagaa 1380  
ggtcctgatt actgtgtgat tactccacct gagaaattga atgtcacagc tatagacaca 1440  
tatgatgacc acagaatggc catggcattc tctcttgctg cttgtgggga tgttccagta 1500  
accatcaagg atcctgggtg caccaggaag acatttcctg actactttga agtccttgag 1560  
aggttaacaa agcactaa 1578

<210> 7  
<211> 1578  
<212> ADN  
<213> *Glycine max*

5

<400> 7  
atggctcagg tctctcgcgt tcataatctc gctcagagta cccagatatt cggacattcc 60  
agtaactcaa acaaactaaa gtctgtgaat agtgtatcac ttcgccctcg gctgtgggga 120  
gcaagtaaga gccgtatccc tatgcacaag aacggttcgt tcatggggaa cttaacgtc 180  
ggcaaaggaa actcaggtgt cttcaaagta agcgcacagc tagctgcggc tgagaagccc 240



ES 2 420 929 T3

agtacttctc ctgaaattgt tcttgaaccg ataaaggatt tctcaggtac gattacacta 300  
 cctggatcaa agagtctctc taatagaatt ttgttgctcg cagctctgtc cgaaggaacc 360  
 actgtagtgc ataacctcct ttatagcgaa gatatacatt atatgttggg ggcgctcaga 420  
 actcctgggc taagagttga ggacgataag actactaaac aagctatcgt cgaaggttgt 480  
 ggcgggttgt tccctacttc taaagaaagt aaagatgaga taaacttggt tcttggcaac 540  
 gcaggaatcg caatgaagag cctcaccgct gctgtcgttg cggcgggtgg taacgctagt 600  
 tacgtcttag acggcgtgcc tagaatgcga gaaagaccta tcggtgatct agtggctggc 660  
 ctaaaacagc ttggagcaga cgtcgattgt ttcttgggca caaattgccc gcccgtaga 720  
 gtgaacggga agggaggctt gccaggcggg aaggttaaac tatccggatc ggtctcgtca 780  
 cagtacctaa ctgcattgct catggccgcc ccgctcgtt tgggggacgt ggagattgaa 840  
 atcgtcgata agttgattag cgtgccttat gtggaaatga ccctcaaatt gatggagagg 900  
 ttcggagttt cggtagaaca ctccgggaat tgggatcggg ttcttgtaca cggagggcaa 960  
 aagtacaaaa gcccaggcaa tgccttcgtc gaaggggacg ctogagcgc ttcctatctc 1020  
 ctcgctggcg cagccataac cggtagcacc ataaccgtga accgctgogg cacctcatcc 1080  
 cttcaaggtg atgtaaagtt cgctgaggtc ttggagaaaa tgggcgcaaa ggtcacatgg 1140  
 tctgagaaca gcgtaaccgt gtccggacct ccagagact ttogtggtag aaaggtcctt 1200  
 aggggaatag atgtgaatat gaataagatg ccagatgtgg ctatgacgct cgctgttgtc 1260  
 gccctgttcg caaacggacc taccgcaata agggatgtcg cttcatggcg tgttaaggaa 1320  
 accgaacgga tgatcgctat ttgcaccgag ttgcgtaagc tgggtgcaac ggtggaagaa 1380  
 ggaccagact attgcgtgat aacacctcct gaaaagctca atgtgaccgc tattgacact 1440  
 tatgacgatc acagaatggc tatggcattc tcacttgctg cttgcggtga cgtgccggtt 1500  
 acgatcaagg acccaggtg tactaggaag acattcccag attactttga ggtgttggaa 1560  
 agattgacaa agcactga 1578

<210> 8  
 <211> 506  
 <212> PRT  
 <213> *Zea mays*

5

<400> 8  
 Met Ala Ala Met Ala Thr Lys Ala Ala Ala Gly Thr Val Ser Leu Asp  
 1 5 10 15  
  
 Leu Ala Ala Pro Ser Arg Arg His His Arg Pro Ser Ser Ala Arg Pro  
 20 25 30

ES 2 420 929 T3

Pro Phe Arg Pro Ala Val Arg Gly Leu Arg Ala Pro Gly Arg Arg Val  
 35 40 45

Ile Ala Ala Pro Pro Ala Ala Ala Ala Ala Ala Val Gln Ala Gly  
 50 55 60

Ala Glu Glu Ile Val Leu Gln Pro Ile Lys Glu Ile Ser Gly Thr Val  
 65 70 75 80

Lys Leu Pro Gly Ser Lys Ser Leu Ser Asn Arg Ile Leu Leu Leu Ala  
 85 90 95

Ala Leu Ser Glu Gly Thr Thr Val Val Asp Asn Leu Leu Asn Ser Glu  
 100 105 110

Asp Val His Tyr Met Leu Gly Ala Leu Arg Thr Leu Gly Leu Ser Val  
 115 120 125

Glu Ala Asp Lys Ala Ala Lys Arg Ala Val Val Val Gly Cys Gly Gly  
 130 135 140

Lys Phe Pro Val Glu Asp Ala Lys Glu Glu Val Gln Leu Phe Leu Gly  
 145 150 155 160

Asn Ala Gly Ile Ala Met Arg Ser Leu Thr Ala Ala Val Thr Ala Ala  
 165 170 175

Gly Gly Asn Ala Thr Tyr Val Leu Asp Gly Val Pro Arg Met Arg Glu  
 180 185 190

Arg Pro Ile Gly Asp Leu Val Val Gly Leu Lys Gln Leu Gly Ala Asp  
 195 200 205

Val Asp Cys Phe Leu Gly Thr Asp Cys Pro Pro Val Arg Val Asn Gly  
 210 215 220

Ile Gly Gly Leu Pro Gly Gly Lys Val Lys Leu Ser Gly Ser Ile Ser  
 225 230 235 240

Ser Gln Tyr Leu Ser Ala Leu Leu Met Ala Ala Pro Leu Ala Leu Gly  
 245 250 255

Asp Val Glu Ile Glu Ile Ile Asp Lys Leu Ile Ser Ile Pro Tyr Val  
 260 265 270

ES 2 420 929 T3

Glu Met Thr Leu Arg Leu Met Glu Arg Phe Gly Val Lys Ala Glu His  
 275 280 285

Ser Asp Ser Trp Asp Arg Phe Tyr Ile Lys Gly Gly Gln Lys Tyr Lys  
 290 300

Ser Pro Lys Asn Ala Tyr Val Glu Gly Asp Ala Ser Ser Ala Ser Tyr  
 305 310 315 320

Phe Leu Ala Gly Ala Ala Ile Thr Gly Gly Thr Val Thr Val Glu Gly  
 325 330 335

Cys Gly Thr Thr Ser Leu Gln Gly Asp Val Lys Phe Ala Glu Val Leu  
 340 345 350

Glu Met Met Gly Ala Lys Val Thr Trp Thr Glu Thr Ser Val Thr Val  
 355 360 365

Thr Gly Pro Pro Arg Glu Pro Phe Gly Arg Lys His Leu Lys Ala Ile  
 370 375 380

Asp Val Asn Met Asn Lys Met Pro Asp Val Ala Met Thr Leu Ala Val  
 385 390 395 400

Val Ala Leu Phe Ala Asp Gly Pro Thr Ala Ile Arg Asp Val Ala Ser  
 405 410 415

Trp Arg Val Lys Glu Thr Glu Arg Met Val Ala Ile Arg Thr Glu Leu  
 420 425 430

Thr Lys Leu Gly Ala Ser Val Glu Glu Gly Pro Asp Tyr Cys Ile Ile  
 435 440 445

Thr Pro Pro Glu Lys Leu Asn Val Thr Ala Ile Asp Thr Tyr Asp Asp  
 450 455 460

His Arg Met Ala Met Ala Phe Ser Leu Ala Ala Cys Ala Glu Val Pro  
 465 470 475 480

Val Thr Ile Arg Asp Pro Gly Cys Thr Arg Lys Thr Phe Pro Asp Tyr  
 485 490 495

Phe Asp Val Leu Ser Thr Phe Val Lys Asn  
 500 505

<210> 9  
 <211> 1521  
 <212> ADN

ES 2 420 929 T3

<213> *Zea mays*

<400> 9

```

atggcggcca tggcgaccaa ggccgcgcgc ggcaccgtgt cgctggacct cgcgcgcgcg      60
tcgcgcgcgc accaccgccc gagctcggcg cgcgcgcct tccgccccgc cgtccgcggg      120
ctgcgggcgc ctgggcgcgc cgtgatcgcc gcgcgcgcgc cggcggcagc ggcggcggcg      180
gtgcaggcgc gtgccagga gatcgtgctg cagcccatca aggagatctc cggcaccgtc      240
aagctgcgcg ggtccaagtc gctttccaac cggatcctcc tactcgcgcg cctgtccgag      300
gggacaacag tggttgataa cctgctgaac agtgaggatg tccactacat gctcggggcc      360
ttgaggactc ttggtctctc tgtcgaagcg gacaaagctg ccaaagagc tgtagttggt      420
ggctgtggtg gaaagtccc agttgaggat gctaaagagg aagtgcagct cttcttgggg      480
aatgctggaa tcgcaatgcg gtcottgaca gcagctgta ctgctgctgg tggaaatgca      540
acttacgtgc ttgatggagt accaagaatg agggagagac ccattggcga cttggttgtc      600
ggattgaagc agcttggtgc agatggtgat tgtttccttg gactgactg cccacctggt      660
cgtgtcaatg gaatcggagg gctacctggt ggcaaggta agctgtctgg ctccatcagc      720
agtcagtact tgagtgcctt gctgatggct gctcctttgg ctcttgggga tgtggagatt      780
gaaatcattg ataaattaat ctccattccg tacgtcgaat tgacattgag attgatggag      840
cgttttggtg tgaagcaga gcattctgat agctgggaca gattctacat taaggaggt      900
caaaaataca agtcccctaa aaatgcctat gttgaagggt atgcctcaag cgcaagctat      960
ttcttggctg gtgctgcaat tactggaggg actgtgactg tggaaagggt tggcaccacc     1020
agtttgcagg gtgatgtgaa gtttgcctgag gtactggaga tgatgggagc gaaggttaca     1080
tggaccgaga ctagcgtaac tgttactggc ccaccgcggg agccatttgg gaggaaacac     1140
ctcaaggoga ttgatgtcaa catgaacaag atgcctgatg tcgccatgac tottgcctgtg     1200
gttgccctct ttgccgatgg ccgcacagcc atcagagacg tggcttctct gagagtaaag     1260
gagaccgaga ggatggttgc gatccggacg gagctaacca agctgggagc atctggtgag     1320
gaagggccgc actactgcat catcacgcgc ccggagaagc tgaacgtgac ggcgatcgac     1380
acgtaocgac accacaggat ggccatggcc ttctcccttg ccgcctgtgc cagaggtcccc     1440
gtcaccatcc gggaccctgg gtgcaccgcg aagaccttcc ccgactactt cgatgtgctg     1500
agcactttcg tcaagaatta a                                             1521

```

5 <210> 10  
 <211> 1521  
 <212> ADN  
 <213> *Zea mays*

ES 2 420 929 T3

<400> 10

```

atggcgggcta tggccacgaa ggcagcggcc ggtacagtaa gcctcgattt ggcggccccc 60
tcccgtaggc accaccggcc aagcagtgcg aggccaccgt tcaggccagc agttcgcggt 120
cttagagcgc ctggtagaag ggttatcgca gcgccaccgg cggctgccgc tgcggcagcg 180
gtgcaggccg gcgcggaaga gatcgtccta cagoctatca aggaaatctc tggtagcgta 240
aagttaccag gcagcaaaag tcttagcaac cgaatcctgc tgttgccggc actctctgaa 300
gggaccacgg tcgtagataa tctgctcaac agcgaagacg tgcactatat gttgggtgcc 360
ctgaggacgc taggtctgtc agtgggaagc gataaggccg ccaagcgcgc tgcctcgtt 420
ggctgcggcg gtaagtccc cgtggaggac gcgaaagaag aggtgcagtt atttcttggg 480
aacgctggca tcgccatgcg gtccttacc gcagccgtca ccgctgcggg aggcaacgca 540
acttacgtgc ttgacggtgt tctcgtatg agagagcggc ccatagggga tctcgtcgtg 600
gggctcaagc agctcggggc cgacgttgat tgcttctcgc gaaccgactg cccccctgtg 660
aggggtgaacg gcatcggggg actgccagga ggcaaagtca agttgtccgg ctcaatttcc 720
tcgcagtacc tgagtgcctt gcttatggcg gcccctctgg ctctgggaga cgtcgaatt 780
gagatcattg ataagctgat ctctatccct tatgttgaga tgacactcgg tctgatggaa 840
agattcgggg tcaaagctga gcactccgat tctgggaca ggttctatat caagggcggg 900
cagaaatata agtcaccgaa gaatgcgtac gtcgagggag acgcatcgag cgcgagttac 960
ttccttgcgg gcgctgccat caccggggga accgtgacag tgggaaggctg tgggacaacg 1020
agcttgcagg gcgacgtcaa atttgctgag gtgctagaaa tgatgggcgc taaggtgact 1080
tggactgaga cgtccgtgac cgttacggga ccgccccgcg aacctttcgg ccggaagcat 1140
ctgaaagcga ttgatgtgaa catgaataag atgcgggacg tcgctatgac acttgccgtg 1200
gtggccctgt tcgctgacgg ccccaccgca atcagggatg tcgctagtgt gagggcctcaag 1260
gagacagagc gtatggtggc gatccgaacg gagctgacta aactcggggc cagtgtggag 1320
gagggcccgg attactgcat aatcacacct ccagagaagt tgaacgtcac cgtatcgcac 1380
acatacgacg atcaccggat ggcaatggcc tttagcttgg cagcgtgcgc cgaagtacct 1440
gtgactataa gagatccagg ttgcaccgcg aaaaogtttc ccgactatct cgacgtcctc 1500
tcaaccttcg tgaagaactg a 1521

```

<210> 11

<211> 76

<212> PRT

<213> *Arabidopsis thaliana*

<400> 11

5

ES 2 420 929 T3

Met Ala Gln Val Ser Arg Ile Cys Asn Gly Val Gln Asn Pro Ser Leu  
 1                    5                    10                    15

Ile Ser Asn Leu Ser Lys Ser Ser Gln Arg Lys Ser Pro Leu Ser Val  
                   20                    25                    30

Ser Leu Lys Thr Gln Gln His Pro Arg Ala Tyr Pro Ile Ser Ser Ser  
                   35                    40                    45

Trp Gly Leu Lys Lys Ser Gly Met Thr Leu Ile Gly Ser Glu Leu Arg  
                   50                    55                    60

Pro Leu Lys Val Met Ser Ser Val Ser Thr Ala Cys  
 65                    70                    75

<210> 12  
 <211> 228  
 <212> ADN  
 <213> *Arabidopsis thaliana*

5

<400> 12  
 atggcgcaag ttagcagaat ctgcaatggt gtgcagaacc catctcttat ctccaatctc 60  
 tcgaaatcca gtcaacgcaa atctccctta tcggtttctc tgaagacgca gcagcatcca 120  
 cgagcttata cgatttcgtc gtcgtgggga ttgaagaaga gtgggatgac gttaattggc 180  
 tctgagcttc gtccctcttaa ggtcatgtct tctgtttcca cggcgtgc 228

<210> 13  
 <211> 228  
 <212> ADN  
 <213> *Arabidopsis thaliana*

10

<400> 13  
 atggcccagg taagtaggat ctgtaacgga gtccaaaacc cttcactaat atcgaacctg 60  
 tcaaaaagct ctcaaagaaa gtcgccgctt tctgtatcgt tgaaaactca acagcacccg 120  
 agggcttata ccatctcaag ctccctggggt ctaaagaaaa gtggaatgac actgatcggt 180  
 agcgaactac gaccgctgaa agtcatgtcc tcagtcagca ctgcgtgc 228

<210> 14  
 <211> 228  
 <212> ADN  
 <213> *Arabidopsis thaliana*

20

<400> 14  
 atggcgcaag taagtagaat ctgcaacggc gtgcagaacc cgtcgctgat ctccaacctc 60  
 agcaagtcca gccagcggaa gtcgccgctc tcggtcagcc tcaagacca acagcacccg 120

agggcctacc ctatcagctc atcctggggc ctcaagaaga gtggcatgac gctgatcggc 180

agcgagctgc ggccactcaa ggtgatgtcc tcggtctcaa cggcgtgc 228

<210> 15

<211> 455

<212> PRT

<213> *Agrobacterium tumefaciens*

5

<400> 15

Met Leu His Gly Ala Ser Ser Arg Pro Ala Thr Ala Arg Lys Ser Ser  
1 5 10 15

Gly Leu Ser Gly Thr Val Arg Ile Pro Gly Asp Lys Ser Ile Ser His  
20 25 30

Arg Ser Phe Met Phe Gly Gly Leu Ala Ser Gly Glu Thr Arg Ile Thr  
35 40 45

Gly Leu Leu Glu Gly Glu Asp Val Ile Asn Thr Gly Lys Ala Met Gln  
50 55 60

Ala Met Gly Ala Arg Ile Arg Lys Glu Gly Asp Thr Trp Ile Ile Asp  
65 70 75 80

Gly Val Gly Asn Gly Gly Leu Leu Ala Pro Glu Ala Pro Leu Asp Phe  
85 90 95

Gly Asn Ala Ala Thr Gly Cys Arg Leu Thr Met Gly Leu Val Gly Val  
100 105 110

Tyr Asp Phe Asp Ser Thr Phe Ile Gly Asp Ala Ser Leu Thr Lys Arg  
115 120 125

Pro Met Gly Arg Val Leu Asn Pro Leu Arg Glu Met Gly Val Gln Val  
130 135 140

Lys Ser Glu Asp Gly Asp Arg Leu Pro Val Thr Leu Arg Gly Pro Lys  
145 150 155 160

Thr Pro Thr Pro Ile Thr Tyr Arg Val Pro Met Ala Ser Ala Gln Val  
165 170 175

Lys Ser Ala Val Leu Leu Ala Gly Leu Asn Thr Pro Gly Ile Thr Thr  
180 185 190

ES 2 420 929 T3

Val Ile Glu Pro Ile Met Thr Arg Asp His Thr Glu Lys Met Leu Gln  
 195 200 205

Gly Phe Gly Ala Asn Leu Thr Val Glu Thr Asp Ala Asp Gly Val Arg  
 210 215 220

Thr Ile Arg Leu Glu Gly Arg Gly Lys Leu Thr Gly Gln Val Ile Asp  
 225 230 235 240

Val Pro Gly Asp Pro Ser Ser Thr Ala Phe Pro Leu Val Ala Ala Leu  
 245 250 255

Leu Val Pro Gly Ser Asp Val Thr Ile Leu Asn Val Leu Met Asn Pro  
 260 265 270

Thr Arg Thr Gly Leu Ile Leu Thr Leu Gln Glu Met Gly Ala Asp Ile  
 275 280 285

Glu Val Ile Asn Pro Arg Leu Ala Gly Gly Glu Asp Val Ala Asp Leu  
 290 295 300

Arg Val Arg Ser Ser Thr Leu Lys Gly Val Thr Val Pro Glu Asp Arg  
 305 310 315 320

Ala Pro Ser Met Ile Asp Glu Tyr Pro Ile Leu Ala Val Ala Ala Ala  
 325 330 335

Phe Ala Glu Gly Ala Thr Val Met Asn Gly Leu Glu Glu Leu Arg Val  
 340 345 350

Lys Glu Ser Asp Arg Leu Ser Ala Val Ala Asn Gly Leu Lys Leu Asn  
 355 360 365

Gly Val Asp Cys Asp Glu Gly Glu Thr Ser Leu Val Val Arg Gly Arg  
 370 375 380

Pro Asp Gly Lys Gly Leu Gly Asn Ala Ser Gly Ala Ala Val Ala Thr  
 385 390 395 400

His Leu Asp His Arg Ile Ala Met Ser Phe Leu Val Met Gly Leu Val  
 405 410 415

Ser Glu Asn Pro Val Thr Val Asp Asp Ala Thr Met Ile Ala Thr Ser  
 420 425 430

Phe Pro Glu Phe Met Asp Leu Met Ala Gly Leu Gly Ala Lys Ile Glu



ES 2 420 929 T3

435

440

445

Leu Ser Asp Thr Lys Ala Ala  
450 455

<210> 16

<211> 1368

<212> ADN

<213> *Agrobacterium tumefaciens*

5

<400> 16

```

atgcttcacg gtgcaagcag cgggcccgca accgcccgca aatcctctgg cctttccgga      60
accgtccgca ttcccggcga caagtcgatc tcccaccggt ccttcatggt cggcgggtctc     120
gcgagcggtg aaacgcgcat caccggcctt ctggaaggcg aggacgtcat caatacgggc     180
aaggccatgc aggcgatggg cgcccgcata cgtaaggaag gcgacacctg gatcatcgat     240
ggcgtcggca atggcggcct cctggcgcct gaggcgccgc tcgatttcgg caatgccgcc     300
acgggctgcc gcctgacgat gggcctcgtc ggggtctacg atttcgacag caccttcata     360
ggcgacgcct cgctcacaaa gcgcccgatg ggccgcgtgt tgaaccgcgt gcgcgaaatg     420
ggcgtgcagg taaaatcggg agacggtgac cgtcttcccg ttaccttgcg cgggcccgaag     480
acgccgacgc cgatcaccta ccgcgtgccg atggcctccg cacaggtgaa gtcgcgccgtg     540
ctgctcgccg gcctcaacac gcccggcata acgacggtca tcgagccgat catgacgcgc     600
gatcatacgg aaaagatgct gcagggcctt ggcgccaacc ttaccgtoga gacggatgcg     660
gacggcgtgc gcaccatcog cctggaaggc cgcggaagc tcaccggoca agtcatcgac     720
gtgccggggc acccgtcctc gacggccttc ccgctggttg cggccctgct tgttccggggc     780
tccgacgtca ccatacctcaa cgtgctgatg aaccccaccc gcaccggcct catoctgacg     840
ctgcaggaaa tgggcgcoga catcgaagtc atcaaccgcg gccttgccgg cggcgaagac     900
gtggcggacc tgcgcgttcg ctctccacg ctgaagggcg tcacggtgcc ggaagaccgc     960
gcgccttcga tgatcgacga atatccgatt ctgcgtgtcg ccgccgcctt cgcggaaggg    1020
gcgaccgtga tgaacggtct ggaagaactc cgcgtcaagg aaagcgaccg cctctcggcc    1080
gtcgccaatg gcctcaagct caatggcgtg gattgcgatg agggcgagac gtcgctcgtc    1140
gtgcgtggcc gccctgacgg caaggggctc ggcaacgcct cgggcgcccgc cgtcggccacc    1200
catctcgatc accgcatcgc catgagcttc ctcgtcatgg gcctcgtgtc ggaaaacctt    1260
gtcacggtgg acgatgccac gatgatcgcc acgagcttcc cggagtcat ggacctgatg    1320
gccgggctgg gcgcgaagat cgaactctcc gatacgaagg ctgcctga                    1368

```

<210> 17

<211> 1368

ES 2 420 929 T3

<212> ADN

<213> *Agrobacterium tumefaciens*

<400> 17

5

```

atgcttcacg gagcttccatc taggccagct actgccagga agtctagcgg gctcagtggc      60
accgtgacga tccctggcga taaaagtatt tcacacagga gcttcacggt cggaggactt      120
gctagtggag agacgagaat cactggtttg cttgagggcg aagatgttat caacaccggc      180
aaggcgatgc aagcaatggg tgccagaatc cgaaaagagg gcgatacgtg gatcatcgac      240
ggtgttggtg acggaggatt gctcgcctcc gaagcgcacc ttgactttgg gaacgcagct      300
acgggggtgc gtcttactat gggactggta ggcgtgatg actttgactc taccttccatc      360
ggtgacgcga gcctcactaa gagaccaatg ggacgagtgc tgaatcccct gagggagatg      420
ggtgtccagg tgaatctga ggatgggtgat cgtcttccgg ttactctgcg aggccccaag      480
acccccacgc caatcacgta cagggttccg atggcgtcag cacagggtcaa gtcagcggta      540
ctcctggcgg gcctcaaac acctggaatc acaaccgtga ttgaacccat catgactaga      600
gaccacacgg agaagatggt gcagggtttc ggcgctaata taacggtcga aaccgacgcc      660
gacggcgtga ggacaatccg cttggagggc agaggtaaac tgactggcca agtcatcgat      720
gtgcctggag atccctcgtc cacagcgttt cccctcgtag ctgcgttgct cgtccctgga      780
tctgatgtga cgatcctgaa tgtcctcatg aatccaacta gaaccggcct catcctcaca      840
ttgcaggaga tgggtgctga catcgagggt atcaatccta ggttggcagg tggagaggat      900
gtggccgatc tgcgcgtgcg ttctagtaca ctcaaaggcg tgaccgtccc tgaggatcgc      960
gctccatcca tgatcgacga gtacccatt ctcgccgttg ctgctgcgtt tgccgagggc     1020
gcaactgtaa tgaacggcct tgaggagttg agggttaagg agagtgacag gctgtccgcg     1080
gtggcgaatg gcctgaagct aaacggcgtg gactgcgacg aagggtgaaac gtcctttgta     1140
gtccgtggtc gccagacgg gaaggggttg gggaatgctt cgggagctgc tgtggcgacg     1200
caccttgatc atagaatcgc catgtcattt ctggtgatgg gacttgtctc cgagaatccg     1260
gtgaccgttg acgatgctac catgatcgcc acctccttcc ctgagttcat ggacctcatg     1320
gcaggcttgg gggccaagat cgagctgtct gataactaagg ccgcttga                       1368

```

<210> 18

<211> 1368

<212> ADN

<213> *Agrobacterium tumefaciens*

<400> 18

10

```

atgctacacg gtgcaagcag ccggccggca accgctcgca aatcttccgg cctttcggga      60
acggtcagga ttccgggcga taagtccata tcccaccggc cgttcacggt cggcggctct      120

```

ES 2 420 929 T3

gccagcggtg agacgcgcat cacgggcctg cttgaagggtg aggacgtgat caataccggg 180  
aaggccatgc aggctatggg agcgcgtatc cgcaaggaag gtgacacatg gatcattgac 240  
ggcgttggga atggcgggtct gctcgcccct gaggcccctc tcgacttcgg caatgcccgg 300  
acgggctgca ggctcactat gggactggtc ggggtgtacg acttcgatag cacgttcac 360  
ggagacgcct cgctcacaaa ggcaccaatg ggccgcggtc tgaaccggtt gcgcgagatg 420  
ggcgtacagg tcaaatccga ggatgggtgac cgtttgcccg ttacgctgcg cgggccgaag 480  
acgcctacce cgattaccta ccgctgcca atggcatccg cccaggtcaa gtcagccgtg 540  
ctcctcgccg gactgaacac tccgggcac accacgggtga tcgagcccat catgaccagg 600  
gatcataccg aaaagatgct tcagggggtt ggcgccaacc tgacggtcga gacggacgct 660  
gacggcgta ggaccatccg ccttgagggc aggggtaaac tgactggcca agtcatcgat 720  
gttcggggag acccgtcgtc cacggccttc ccggttggtg cggcgtgct cgtgccgggg 780  
agtgacgtga ccacctgaa cgtcctcatg aaccgacca ggaccggcct gatcctcacg 840  
cttcaggaga tgggagccga catcgagggtg atcaaccgc gcctggcagg cgggtgaagac 900  
gttgccggtc tgcgcgtgcg ctcctctacc ctgaagggcg tgacggctcc ggaagatcgc 960  
gcgccgtcca tgatagacga gtatcctatt ctggccgctc ccgctgcgtt cgccgaaggg 1020  
gccacggta tgaacgggtc tgaggaactc cgcgtgaagg aatcggatcg cctgtcggcg 1080  
gtggccaatg gcctgaagct caacgggtgt gactgogacg aggggtgagac ctcaactcgtg 1140  
gtccgtggcc ggctgatgg caagggcctc ggcaacgcca gtggagcggc cgtcgccacg 1200  
cacctcgatc atcgcatcgc gatgtccttc ttgggtgatg gtctcgtctc agagaacccg 1260  
gtgaccgtcg atgacgccac gatgatagcg acgagcttcc cagagttcat ggatctgatg 1320  
gcgggcctcg gggccaagat cgaactgtct gacacgaagg ccgcttga 1368

<210> 19

<211> 183

<212> PRT

<213> *Streptomyces hygroscopicus*

5

<400> 19

Met Ser Pro Glu Arg Arg Pro Ala Asp Ile Arg Arg Ala Thr Glu Ala  
1 5 10 15

Asp Met Pro Ala Val Cys Thr Ile Val Asn His Tyr Ile Glu Thr Ser  
20 25 30

Thr Val Asn Phe Arg Thr Glu Pro Gln Glu Pro Gln Asp Trp Thr Asp  
35 40 45

ES 2 420 929 T3

Asp Leu Val Arg Leu Arg Glu Arg Tyr Pro Trp Leu Val Ala Glu Val  
50 55 60

Asp Gly Glu Val Ala Gly Ile Ala Tyr Ala Gly Pro Trp Lys Ala Arg  
65 70 75 80

Asn Ala Tyr Asp Trp Thr Ala Glu Ser Thr Val Tyr Val Ser Pro Arg  
85 90 95

His Gln Arg Thr Gly Leu Gly Ser Thr Leu Tyr Thr His Leu Leu Lys  
100 105 110

Ser Leu Glu Ala Gln Gly Phe Lys Ser Val Val Ala Val Ile Gly Leu  
115 120 125

Pro Asn Asp Pro Ser Val Arg Met His Glu Ala Leu Gly Tyr Ala Pro  
130 135 140

Arg Gly Met Leu Arg Ala Ala Gly Phe Lys His Gly Asn Trp His Asp  
145 150 155 160

Val Gly Phe Trp Gln Leu Asp Phe Ser Leu Pro Val Pro Pro Arg Pro  
165 170 175

Val Leu Pro Val Thr Glu Ile  
180

<210> 20

<211> 552

<212> ADN

<213> *Streptomyces hygroscopicus*

5

<400> 20

atgagcccag aacgacgccc ggccgacatc cgccgtgcca ccgagggcgga catgccggcg 60  
gtctgcacca tcgtcaacca ctacatcgag acaagcacgg tcaacttccg taccgagccg 120  
caggaaccgc aggactggac ggaacgacctc gtccgtctgc gggagcgcta tccttggtc 180  
gtcgccgagg tggacggcga ggtcgccggc atcgcttacg cgggcccctg gaaggcacgc 240  
aacgcctacg actggacggc cgagtcgacc gtgtacgtct cccccgcca ccagcggacg 300  
ggactgggct ccacgtctta caccacctg ctgaagtccc tggaggcaca gggcttcaag 360  
agcgtggctg ctgtcatcgg gctgcccac gaccogagcg tgcgcatgca cgagggcgctc 420  
ggatatgccc cccgcggcat gctgcggggc gccggcttca agcacgggaa ctggcatgac 480  
gtgggtttct ggcagctgga cttcagcctg ccgggtaccgc cccgtccggt cctgcccgtc 540  
accgagatct ga 552

ES 2 420 929 T3

<210> 21  
 <211> 552  
 <212> ADN  
 <213> *Streptomyces hygrosopicus*  
 5 <400> 21  
 atgagtccag aaaggagacc ggctgatatt cggagagcca ccgaagctga tatgcctgct 60  
 gtttgtacaa tcgtaaacca ttatatcgag acctcgacag ttaattttcg cactgagccg 120  
 caggagccac aggattggac ggacgatctg gtacgtttaa gagaacgtta tccgtggcta 180  
 gttgctgagg ttgacggaga agtcgctggt atagcttacg ctggaccgtg gaaagctcgt 240  
 aacgcttacg actggacagc agaatccact gtctacgtca gccctcgtca tcaaagaacc 300  
 ggattagggga gcacgttgta cactcatctt ttaaagtcac tggaggcaca aggcttcaag 360  
 tctgttggtg cagttattgg attgccaaac gatccgagtg ttcgaatgca cgaagcgctt 420  
 ggatacgtc cagcaggtat gctccgtgct gccggattca aacatggaaa ttggcacgac 480  
 gtaggttttt ggcaactgga cttttcactt cccgttcccc ctagacctgt acttccagtt 540  
 actgaaatct ag 552  
 <210> 22  
 <211> 552  
 10 <212> ADN  
 <213> *Streptomyces hygrosopicus*  
 <400> 22  
 atgtcgcctg agcgcctgcc tgctgacata agacgcgcta ccgaggcaga catgcctgct 60  
 gtttgcacca ttgtgaatca ctacatcgag acatctacgg taaacttccg cactgagcct 120  
 caagaaccgc aggattggac cgacgatctc gtgctgtca gagagcgta tccgtggctg 180  
 gttgcagagg tggacggtga agtggctggg atcgcctacg ctggaccgtg gaaggctaga 240  
 aacgcatacg attggactgc ggagtocaca gtctacgtct caccagaca tcaaagaacc 300  
 gggctcggct cgacccteta tacgatctc ctcaagtcct tagagggcga gggcttcaaa 360  
 tctgtagtgg cggtgatcgg cttgccaaac gatcccagtg tgagaatgca cgaggcactc 420  
 ggttacgctc ctagaggaat gctcagggcg gctggattca agcacggtaa ttggcacgac 480  
 gttggcttct ggcaactgga cttctctttg ccagttccac ctcgtcctgt gctacccgctc 540  
 accgaaatct ag 552  
 15 <210> 23  
 <211> 1368  
 <212> ADN  
 <213> *Agrobacterium tumefaciens*  
 20 <400> 23

ES 2 420 929 T3

atgcttcacg gtgcaagcag ccgccagca actgctcgtg agtcctctgg tctttctgga 60  
accgtccgta ttccagggtga caagtctatc tcccacaggt ccttcatggt tggagggtctc 120  
gctagcgggtg aaactcgtat caccgggtctt ttggaagggtg aagatggtat caacactggt 180  
aaggctatgc aagctatggg tgccagaatc cgtaaaggaag gtgatacttg gatcattgat 240  
gggtgttgga acgggtggact ccttgctcct gaggtcctc tcgatttcgg taacgctgca 300  
actggttgcc gtttgactat gggctctggt ggtgtttacg atttcgatag cactttcatt 360  
ggtgacgctt ctctcactaa gcgtccaatg ggtcgtgtgt tgaaccact tcgcgaaatg 420  
gggtgtgcagg tgaagtctga agacgggtgat cgtcttcag ttaccttgcg tggaccaaag 480  
actccaacgc caatcaccta cagggtacct atggcttccg ctcaagtga gtcgctggt 540  
ctgcttgctg gtctcaacac cccaggatc accactgtta tcgagccaat catgactcgt 600  
gaccacactg aaaagatgct tcaaggtttt ggtgctaacc ttaccgttga gactgatgct 660  
gacgggtgtgc gtaccatccg tottgaagg cgtggttaagc tcaccggta agtgattgat 720  
gttccagggtg atccatcctc tactgctttc ccattgggtg ctgccttgct tgttccagggt 780  
tccgacgtca ccaccttaa cgttttgatg aaccaacc gtactggtct catcttgact 840  
ctgcaggaaa tgggtgccga catcgaagtg atcaaccac gtcttgctgg tggagaagac 900  
gtggctgact tgcgtgttcg ttcttctact ttgaagggtg ttactgttcc agaagaccgt 960  
gctccttcta tgatcgacga gtatccaatt ctgctgttg cagctgcatt cgctgaagggt 1020  
gctaccgtta tgaacggttt ggaagaactc cgtgttaagg aaagcgaccg tctttctgct 1080  
gtcgcaaacg gtctcaagct caacgggtgtt gattgcatg aagggtgagac ttctctcgtc 1140  
gtgctgtgtc gtctgacgg taagggtctc ggtaacgctt ctggagcagc tgtcgtacc 1200  
cacctcgatc accgtatcgc tatgagcttc ctggttatgg gtctcgttcc tgaaaacct 1260  
gttactgttg atgatgctac tatgatcgt actagcttcc cagagttcat ggatttgatg 1320  
gctggcttg gagctaagat cgaactctcc gacactaagg ctgcttga 1368

<210> 24  
<211> 16  
<212> ADN  
<213> Secuencia artificial

5

<220>  
<223> Molécula de cebador de ADN  
<400> 24

10 catggagctt catcta 16

<210> 25  
<211> 16  
<212> ADNA  
<213> Secuencia artificial

15

<220>  
<223> Molécula de cebador de ADN

5 <400> 25  
gccttgagt gtacta 16

<210> 26  
10 <211> 16  
<212> ADN  
<213> Secuencia artificial

<220>  
15 <223> Molécula de cebador de ADN

<400> 26  
gggagcgcgt atccgc 16

20 <210> 27  
<211> 16  
<212> ADN  
<213> Secuencia artificial

25 <220>  
<223> Molécula de cebador de ADN

<400> 27  
30 ggatggtcac gtcact 16

<210> 28  
35 <211> 16  
<212> ADN  
<213> Secuencia artificial

<220>  
40 <223> Molécula de cebador de ADN

<400> 28  
cgcatcacg acggtc 16

45 <210> 29  
<211> 16  
<212> ADN  
<213> Secuencia artificial

50 <220>  
<223> Molécula de cebador de ADN

<400> 29  
55 ggcacgtcc accgtg 16

<210> 30  
60 <211> 16  
<212> ADN  
<213> Secuencia artificial

ES 2 420 929 T3

<220>  
<223> Molécula de cebador de ADN  
  
<400> 30  
5 gcaactggtt gccgtt 16  
  
<210> 31  
<211> 16  
10 <212> ADN  
<213> Secuencia artificial  
  
<220>  
<223> Molécula de cebador de ADN  
15 <400> 31  
  
atcacctgga acatca 16  
20 <210> 32  
<211> 22  
<212> ADN  
<213> *Zea mays*  
25 <400> 32  
  
cgtcaagatc ctctcacct cg 22  
30 <210> 33  
<211> 22  
<212> ADN  
<213> *Zea mays*  
  
<400> 33  
35 acaccctctc caacactctc ta 22  
  
<210> 34  
<211> 9  
40 <212> PRT  
<213> Secuencia artificial  
  
<220>  
<223> Motivo que proporciona resistencia al glifosato a una planta EPSPS  
45 <400> 34  
  
Gly Asn Ala Gly Ile Ala Met Lys Ser  
1 5  
50 <210> 35  
<211> 1596  
<212> ADN  
<213> *Agrobacterium tumefaciens*  
55 <400> 35



ES 2 420 929 T3

atggccaag ttagccgaat ctgcaacggt gtgcagaatc catcactaat ctccaacctg 60  
 tccaaatcgt cacaacgtaa gtcgccatta tctgttagct tgaagactca gcaacatcct 120  
 cgcgcatatc ctatatcaag cagttgggggt ttgaagaaat cgggtatgac cttgattgggt 180  
 tccggaactta ggccattgaa ggtgatgtct tcagttagta cagcttgcat gcttcacggt 240  
 gcttcttcca gacccgcaac ggctagaaag agttctgggt tgtctggaac cgcccgattt 300  
 ccaggagaca aaagcattag tcaccgctct ttcattgtttg gtgggctggc atctggagag 360  
 acgcgcatca ctggtcttct ggaaggagag gacgtcatca atacagggaa ggcaatgcag 420  
 gctatgggtg cccgtattcg caaggaagggt gatacttggg tcatagacgg agttgggaac 480  
 ggtggcttac ttgcaccgga ggctcctctc gactttggca acgcagccac aggggtgtaga 540  
 cttactatgg gcctcgtggg tgtttacgat ttcgattcaa cctttattgg ggatgcctct 600  
 ctactaaac gcccaatggg aagagtcctt aaccggttga gggagatggg cgtacaagtt 660  
 aagtccgagg acggcgacag attgccgctc accttgcgcg gccctaagac acccaccctt 720  
 attacttaca gggttccaat ggcattctgt caagtgaagt ccgcagttct gctcgtctga 780  
 ttgaacacac cgggtattac taccgtgatt gagccgatca tgactcgtga ccacactgag 840  
 aagatgcttc agggtttcgg tgctaacctc accggtgaaa cagacgcgga cgggtgtgagg 900  
 accattcgcc tggaggggag gggaaaactc actggtcaag tcattgacgt gcccggtgat 960  
 ccctccagca cggcgttccc actggttgcc gctcttctcg taccaggctc cgatgtgaca 1020  
 attctaaacg tcctcatgaa tcctactaga accggattga tacttacatt gcaggaaatg 1080  
 ggtgctgata ttgaagtat caatcctaga ctagccggag gtgaggacgt agctgatttg 1140  
 cgggtgaggt cttctacatt gaaagggtt accgtacctg aagatagggc accttcaatg 1200  
 attgacgagt atccaattct tgccgtcgcg gctgcctttg ctgagggcgc gaccgtgatg 1260  
 aatggactag aggagttgag agtgaaggaa tccgacagat tgagcgcagt cgctaacgga 1320  
 cttaaactca atggcgttga ttgtgatgag ggtgagacta gcttggtagt ccgtgggcca 1380  
 ccagacggaa agggtttggg caacgcttcg ggtgctgccg ttgcaactca cttggatcat 1440  
 cggatagcga tgagttttct ggtgatgggt ctcgtaagcg agaatcctgt gacagtcgac 1500  
 gatgcaacta tgatcgtctac ttccttccct gagtttatgg atttaatggc aggactaggt 1560  
 gcaaagattg aactctctga taccaaagcg gcctaa 1596

**REIVINDICACIONES**

5 1.- Una célula vegetal, una planta o su progenie tolerantes al glifosato, que comprende al menos dos polinucleótidos, en la que dichos al menos dos polinucleótidos codifican polipéptidos de EPSPS que son al menos 98% idénticos, siendo al menos uno de dichos al menos dos polinucleótidos un transgén que comprende SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:17, SEQ ID NO: 18 o SEQ ID NO:35, y dichos al menos dos polinucleótidos son sustancialmente divergentes y son menos del 85% similares en sus secuencias polinucleotídicas.

10 2.- La célula vegetal, la planta o su progenie de la reivindicación 1, que se selecciona del trigo, maíz, arroz, soja, algodón, patata, canola, césped, árboles forestales, grano de sorgo, cultivos vegetales, plantas ornamentales, cultivos de forraje y cultivos frutales.

3.- Un procedimiento de selección de células vegetales, plantas o su progenie tolerantes al glifosato de la reivindicación 1 o 2, que comprende las etapas de:

15 (a) transformar una célula vegetal con una construcción de ADN que comprende: una molécula de promotor que opera en plantas, unida operablemente a una molécula polinucleotídica artificial seleccionada de SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:17, SEQ ID NO: 18 y SEQ ID NO:35; y

(b) cultivar dicha célula vegetal en un medio selectivo que contiene glifosato para destruir selectivamente las células que no han sido transformadas con dicha construcción de ADN; y

(c) regenerar dicha célula vegetal en una planta fértil.

20

25

30

35

Figura 1A
Figura 1B
Figura 1C
Figura 1D
Figura 1E
Figura 1F
Figura 1G
Figura 1H

Figura 1

Figura 3A
Figura 3B
Figura 3C
Figura 3D
Figura 3E
Figura 3F
Figura 3G

Figura 3

Figura 5A
Figura 5B
Figura 5C
Figura 5D
Figura 5E
Figura 5F
Figura 5G

Figura 5

Figura 2A
Figura 2B
Figura 2C
Figura 2D
Figura 2E
Figura 2F

Figura 2

Figura 4A
Figura 4B
Figura 4C

Figura 4

```

1      50
OsEPSPS_Nat ATGGCGGGGA CCATGGCGTC CAACGCCGGG GCTGCGGGGG CGGTGTCCCT
OsEPSPS_AT   ....T.A. .T....TAG T....A... ..C..T. .C..T...T.
OsEPSPS_ZM   ....C.T. .G.....TAG T.....G..A .C..T.... .T..T..G..

51      100
OsEPSPS_Nat GGACCAGGCC GTGGCGGCGT CCGCGGGCGT CTCGTGCGGG AAGCAGCTGC
OsEPSPS_AT   A.....A..A ..A..A..A GC..T..A.. ..A..A..T .....A..A.
OsEPSPS_ZM   A..T.....G ..T..T..CA GC..T..C.. ..CAGTA.. ..A...T..A.

101     150
OsEPSPS_Nat GGCTGCCCGC CGCGGCGGC GGGGGGATGC GGGTGCGGGT GCGGGGCGGG
OsEPSPS_AT   ...A..A.. A..C..TA.A ..C..C...A .A..TA.A.. .A....TA.A
OsEPSPS_ZM   .T..C..G.. G..T..C..A ..T..A...A .A.....C.. T..A..CA.A

151     200
OsEPSPS_Nat GGGCGGGGG AGGCGGTGGT GGTGGCGTCC GCGTCGTCGT CGTCGGTGGC
OsEPSPS_AT   ..TA.....: ....T..A.. C..A..C... ..T..TAGCA GT.....:
OsEPSPS_ZM   ..CA.A..T. ....C..... A..C..C..G ..CAGC..... .CAGT..T..

```

Figura 1A

```

201                               250
OsEPSPS_Nat AGCGCCGGCG GCGAAGGCGG AGGAGATCGT GCTCCAGCCC ATCAGGGGAGA
OsEPSPS_AT  T.....T ..T.....A. ....T... TT.A..A..T ..T...A..A.
OsEPSPS_ZM  T.....A ..T.....C. .A..A..T.. T..G..A..G ..AC.....A.

251                               300
OsEPSPS_Nat TCTCCGGGGC GGTTCAGCTG CCAGGGTCCA AGTCGCTCTC CAACAGGATC
OsEPSPS_AT  .A..G..... C..A..AT.A ..T..AAG.. ..AGC..T.. .....T
OsEPSPS_ZM  .T..G..C.. ..G.....C ..T..C..G. ....C..G.. G..TC.T...

301                               350
OsEPSPS_Nat CTCCTCCTCT CCGCCCTCTC CGAGGGCACA ACAGTGGTGG ACAACTTGCT
OsEPSPS_AT  ..GT.G..T. .A..T..... G.....A... .....T.... T..TC..T.
OsEPSPS_ZM  ..T..G..A. :G..G..G.. ..A..G..C ..G..T..A. .T..TC..T.

351                               400
OsEPSPS_Nat GAACAGTGAG GATGTTCACT ACATGCTTGA GGCCCTGAAA GCCCTCGGGC
OsEPSPS_AT  ...T..... ..G..... T.....A.. ..T..C..G ..T..A.....
OsEPSPS_ZM  ....TCC..A ..C..G..... ..C... A..G..A... ..GT..G..CT

```

Figura 1B

```

401      OsEPSPS_Nat  TCTCTGTGGA AGCAGATAAA GTTGCAAAA  GAGCTGTAGT  CGTTGGCTGT  450
OsEPSPS_AT  .T.....A.. ...G..... ..A.....C  .C..A..G..  T..A..T...
OsEPSPS_ZM  .GAGC..T.. G..T..C... ..G..C..GC  ...A..C..  A..G..A..C

451      OsEPSPS_Nat  GGTGGCAAGT TTCCTGTTGA GAAGGATGCG AAAGAGGAAG  TGCAACTCTT  500
OsEPSPS_AT  .....G.... C..A..... A.....T  ..G..A.... .A..G.....
OsEPSPS_ZM  ..C..T.... C..A..G.. ..A..C..C  ....A..G.  T....T.G..

501      OsEPSPS_Nat  CTTGGGAAC  GCTGGAATCG CAATGCGATC CTTGACAGCA GCCGTGACTG  550
OsEPSPS_AT  TC.C.....T  ..C..G.... .C.....GAG  T.....T..T  ..G..C..A.
OsEPSPS_ZM  .....A.... .....C.... ..A..AG  .C.....T..C  ..G..C....

551      OsEPSPS_Nat  CTGCTGGTGG AAATGCAACT TATGTGCTTG ATGGAGTGCC  ACGAATGAGG  600
OsEPSPS_AT  .C.....A.. C..C.....A  ..C..C..A.  ....G..... GA.....C.T
OsEPSPS_ZM  .C..A..G.. T..C..... ..C..C..G.  .C..T..T..  ...G....C.C

```

Figura 1C

601	OsePPS_Nat	GAGAGACCGA	TTGGTGACTT	GGTTGTCGGG	TTGAAACAAC	TTGGTGCGGA	650
	OsePPS_AT	. . . C . T . . T .	. . . . . TC .	T . . C . A . . T	C . C . . G . . . .	. C . C . T . .	
	OsePPS_ZM	. . AC . . . . C .	. C . . . . TC .	C . G . A . A	C . T . . . . . T	. G . A . C . .	
651	OsePPS_Nat	TGTCGACTGT	TTCCTTGGCA	CTGAATGCCC	ACCTGTTCGT	GTCAGGGAA	700
	OsePPS_AT	C . . A . . T . . .	. . . . . G . T .	. . . . G . . T . .	G . . A . . CA . A	. . T . . A . . . .	
	OsePPS_ZM	. . A . . T . . . .	. . . T . G . . G .	. A . G . . . . .	T . . A . . G . . G	. . A . . . . . C .	
701	OsePPS_Nat	TTGGAGGACT	TCCTGGTGGC	AAGTTAAGC	TCTCTGGTTC	CATCAGCAGT	750
	OsePPS_AT	. C . . T . . G . .	G . . G . . C . A	. . . . . C . . . .	. G . . G . . CAG	T . . TTCG . . . .	
	OsePPS_ZM	. C . . G . . T . .	C . A . C . C . A	. . . . . G . . . T	. A . . A . . GAG	T . . . TC . TCG	
751	OsePPS_Nat	CAGTACTTGA	GTGCCTTGCT	GATGGCTGCT	CCTTTGGCCC	TTGGGGATGT	800
	OsePPS_AT	. . . . . TC . TT	C . . . TC . C . .	. . . . . . . . G	. A . . A . . TT	. G . A . . . . .	
	OsePPS_ZM	. . . . . C . CT	CC . GC . . T .	. . . . . A . A	. . AC . . . . . GT	. G . C . C . . C . .	

Figura 1D

```

801                               850
OsEPSPS_Nat GGAGATCGAA ATCATTGACA AACTAATCTC CATTCCCTTAC GTTGAAATGA
OsEPSPS_AT T.....G .....T.....T..A.. T..C..G..T ..C..G.....
OsEPSPS_ZM T..A..T... ..T..C..... .G..C..A.. ..C..G..T ..G..G.....

851                               900
OsEPSPS_Nat CATTGAGATT GATGGAGCGT TTTGGTGTGA AGGCAGAGCA TTCTGATAGT
OsEPSPS_AT .T..A...C. T.....A..G .....G..T. ....C..... .AGC..C...
OsEPSPS_ZM ..C..C.GC. ....AA.A .....A.... .A..T..A.. .AG...CTC.

901                               950
OsEPSPS_Nat TGGGACAGAT TCTATATTAA GGGAGGGCAG AAGTACAAAT CTCCTGGAAA
OsEPSPS_AT ....TC.T. ....C..A.. .....C.... .....T..G. ....G...
OsEPSPS_ZM .....TC.G. ....C..C.. ...C..A.... ..A..T..GA G...A..C...

951                               1000
OsEPSPS_Nat TGCCTATGTT GAAGGTGATG CCTCAAGCGC GAGCTATTTTCT TTGGCTGGTG
OsEPSPS_AT ...T.....A .....G..... .T...TCT.. .TCT..C... C.T..G..A.
OsEPSPS_ZM ...G..C..G ..G..C..C. ....TTC... C.....C.... .....A..C.

```

Figura 1E



```

1001      OsEPSPS_Nat  CTGCAATCAC TGGAGGCACT GTGACAGTTC AAGGTTGTGG TAGGACCCAGT      1050
OsEPSPS_AT   .G..T..A.. ..A..A..A..C..... .G..C..C.. ..A..A..A...
OsEPSPS_ZM   .A..T..T.. C.....T... ..T.....G. ....G..... A..A..G..C

1051      OsEPSPS_Nat  TTGCAGGGTG ATGTCAAATT TGCTGAGGTA CTTGAGATGA TGGGAGCAAA      1100
OsEPSPS_AT   .....A.... .C..G..G.. ..C..... ..A..... ..T..C...
OsEPSPS_ZM   C.....A..C. .C..A..G.. C..A..A..C ..A..... ..C..T...

1101      OsEPSPS_Nat  GGTTACATGG ACTGACACCA GTGTAACCGT AACTGGTCCA CCACGTGAGC      1150
OsEPSPS_AT   A..A..G... ..A.....AT CG..G..A.. T.....T... ..A..A..
OsEPSPS_ZM   ...G..T... ..A..T..GT CA..C..... C.....G..T ..T..G..A..

1151      OsEPSPS_Nat  CTTATGGGAA GAAACACCTG AAAGCTGTTG ATGTCAACAT GAACAAAATG      1200
OsEPSPS_AT   ...C..C.. A..G..T..T ..G..C..G. ....T..T.. ..T..G...
OsEPSPS_ZM   .C..C..T.. ..A.....T... ..G..A..G. .C..G..... ..T..G...

```

Figura 1F

```

1201                               1250
OsEPSPS_Nat CCTGATGTTG CCATGACCCT TGCCGTTGTT GCACTCTTCG CTGATGGTCC
OsEPSPS_AT   . . . . . C . . . . . T . . . . . A . . . . . . . . . . . C . . . . . T . . . . . T . . . . . A . . . . . C . . . . . C . . . . .
OsEPSPS_ZM   . . . . . A . . . . . C . . . . . . . . . . . G . . . . . . . . . . . C . . . . . A . . . . . A . . . . . C . . . . . T . . . . . G . . . . . . . . . . . A . . . . . C . . . . . A . . . . .

1251                               1300
OsEPSPS_Nat AACTGCTATC AGAGATGTGG CTTCCTGGAG AGTAAAGGAA ACCGAAAGGA
OsEPSPS_AT   . . . . . G . . . . . G . . . . . A . . . . . C . . . . . C . . . . . T . . . . . A . . . . . A . . . . . C . . . . . C . . . . . . . . . . . . . . . . . G . . . . . G . . . . . . . . . . .
OsEPSPS_ZM   T . . . . . C . . . . . . . . . . . T . . . . . G . . . . . C . . . . . T . . . . . . . . . . . A . . . . . G . . . . . T . . . . . . . . . . . T . . . . . G . . . . . C . . . . . . . . . . . G . . . . . C . . . . . . . . . . . G . . . . . C . . . . . . . . . . . G . . . . . T . . . . . G . . . . . A . . . . .

1301                               1350
OsEPSPS_Nat TGGTTGCAAT TCGGACCGAG CTAACAAGC TGGAGCATC GGTTGAAGAA
OsEPSPS_AT   . . . . . G . . . . . T . . . . . . . . . . . A . . . . . T . . . . . A . . . . . . . . . . . C . . . . . C . . . . . A . . . . . . . . . . . T . . . . . T . . . . . C . . . . . . . . . . . T . . . . . A . . . . . G . . . . . G . . . . .
OsEPSPS_ZM   . . . . . C . . . . . T . . . . . . . . . . . C . . . . . C . . . . . . . . . . . A . . . . . T . . . . . G . . . . . A . . . . . . . . . . . C . . . . . T . . . . . . . . . . . A . . . . . G . . . . . C . . . . . G . . . . . G . . . . . . . . . . .

1351                               1400
OsEPSPS_Nat GGTCCTGACT ACTGCATCAT CACCCACCG GAGAAGCTGA ACATCACGGC
OsEPSPS_AT   . . . . . C . . . . . . . . . . . T . . . . . . . . . . . T . . . . . A . . . . . C . . . . . T . . . . . . . . . . . A . . . . . T . . . . . . . . . . . T . . . . . . . . . . . T . . . . . . . . . . . T . . . . .
OsEPSPS_ZM   . . . . . A . . . . . . . . . . . . . . . . . . . . . A . . . . . G . . . . . T . . . . . C . . . . . . . . . . . A . . . . . . . . . . . C . . . . . . . . . . . A . . . . . . . . . . . C . . . . . . . . . . . A . . . . . C . . . . . C . . . . .

```

Figura 1G

```

1401      OsEPSPS_Nat  AATCGACACC TACGATGATC ACAGGATGGC CATGGCCTTC TCCCTCGCTG 1450
      OsEPSPS_AT  T..T..T..A .....C..... .T..A..... T.....T... ..A..G...C.
      OsEPSPS_ZM  T..A.....T .....C..... ..A..... G.....G... ..G.....A.

1451      OsEPSPS_Nat  CCTGGGCCGA CGTGCCCGTG ACGATCAGGG ACCCTGGTTG CACCCGCAAG 1500
      OsEPSPS_AT  .T..T..A.. T..T..T..C ..A.....A. .T.....C... T..TA.A...
      OsEPSPS_ZM  .G.....G.. ...T..G..T .....T..A. .T.....G... ..T..T..A

1501      OsEPSPS_Nat  ACCTTCCCCA ACTACTTCGA CGTTCTAAGC ACTTTCGTCA GGAAGTGA 1548
      OsEPSPS_AT  ..G.....G. ....T... T.....TTCA ..A.....C .C..T...
      OsEPSPS_ZM  ..T.....T. .T.....T..T..G...TCA ..C..T..AC .C.....

```

Figura 1H

```

1      50
ZmEPSPS_Nat ATGGCGGCCA TGGGACCAA GGCCGCCGGG GGCACCGTGT CGCTGGACCT
ZmEPSPS_ZM  .....T. ....C..G.. ...A..G..C ..T..A..AA GC..C..TT.

51      100
ZmEPSPS_Nat CGCCGGCCGG TCGCGCCGCC ACCACCGCCC GAGCTCGGGG CGCCCGCCCT
ZmEPSPS_ZM  G..G..C..C ..C..TA.G. ....G.. A...AGT... A.G..A..G.

101     150
ZmEPSPS_Nat TCCGCCCCGC CGTCCGGGGG CTGGGGGGC CTGGGGCCG CGTGATCGCC
ZmEPSPS_ZM  ..A.G..A.. A..T.....T ..TA.A.... ....TA.AA. G..T.....A

151     200
ZmEPSPS_Nat GCGCCGCCGG CGGCGGAGC GCGCGGGGG GTGCAGGGG GTGCCGAGGA
ZmEPSPS_ZM  .....A.... ....T..C.. T.....A.... .....C. .C..G..A..

201     250
ZmEPSPS_Nat GATCGTGCTG CAGCCCATCA AGGAGATCTC CGGCACCGTC AAGCTGCCGG
ZmEPSPS_ZM  .....C..A .....T.... ....A..... T..T..G..A ...T..A..A..

```

Figura 2A

```

251                               300
ZmEPSPS_Nat GGTCCAAGTC GCTTCCAAC CGGATCCTCC TACTCGCCGC CCTGTCCGAG
ZmEPSPS_ZM .CAG...AAG T...AG..... .A.....G. .GT.G..G.. A..C..T..A

301                               350
ZmEPSPS_Nat GGGACAACAG TGGTTGATAA CCTGCTGAAC AGTGAGGATG TCCACTACAT
ZmEPSPS_ZM .....C..G. .C..A..... T.....C.... ..C..A..C. .G.....T..

351                               400
ZmEPSPS_Nat GCTCGGGGCC TTGAGGACTC TTGGTCTCTC TGTCGAAGCG GACAAAGCTG
ZmEPSPS_ZM .T.G..T... C.....G. .A.....G.. A..G.....C ..T..G...C.

401                               450
ZmEPSPS_Nat CCAAAGAGC TGAGTTGTT GGCTGTGGTG GAAAGTTCCC AGTTGAGGAT
ZmEPSPS_ZM ....GC.C.. ...C..C... ..C..C..C. .T.....C.... C..G.....C

451                               500
ZmEPSPS_Nat GCTAAAGAGG AAGTGCAGCT CTTCTTGGGG AATGCTGGAA TCGCAATGCG
ZmEPSPS_ZM ..G.....A. .G.:.....T. A..TC.T... ..C.....C. ....C.....

```

Figura 2B

```

501                               550
ZmEPSPS_Nat  GTCCTTGACA GCAGCTGTTA CTGCTGCTGG TGGAAATGCA ACTTACGTGC
ZmEPSPS_ZM   ...C.T..C ...C...C..C. .C...G.. A..C..C... ..
Consenso    ...c.g..a ...c.c.c..c. .c...g.. a..a..c... ..

551                               600
ZmEPSPS_Nat  TTGATGGAGT ACCAAGAATG AGGGAGAGAC CCATTGGCGA CTTGGTTGTC
ZmEPSPS_ZM   ...C..T.. T..TC.T... ..A...C.G. ....A..G.. TC.C..C..G

601                               650
ZmEPSPS_Nat  GGATTGAAGC AGCTTGGTGC AGATGTTGAT TGTTTCCTTG GCACTGACTG
ZmEPSPS_ZM   ..GC.C.... ...C..G.. C..C..... .C.....C. .A..C.....

651                               700
ZmEPSPS_Nat  CCCACCTGTT CGTGTCAATG GAATCGGAGG GCTACCTGGT GGCAAGGTCA
ZmEPSPS_ZM   ...C.....G A.G..G..C. .C.....G.. A..G..A..A ..A.....

701                               750
ZmEPSPS_Nat  AGCTGTCTGG CTCATCAGC AGTCAGTACT TGAGTGCCTT GCTGATGGCT
ZmEPSPS_ZM   ..T....C.. ...A..TTC. TCG.....C. ....T.....G

```

Figura 2C

```

751                               800
ZmEPSPS_Nat GCTCCTTTGG CTC TTGGGGA TGTGGAGATT GAAATCATTG ATAAATTAAT
ZmEPSPS_ZM  ..C...C... ..G..A.. C..C..A... ..G..... ..GC..G..

801                               850
ZmEPSPS_Nat CTC CATTCCG TACGTCGAAA TGACATTGAG ATTGATGGAG CGTTTTGGTG
ZmEPSPS_ZM  ...T..C..T ..T..T..G. ....C..CC. TC.....A A.A..C..G..

851                               900
ZmEPSPS_Nat TGA AAGCAGA GCATTCTGAT AGCTGGGACA GATTCTACAT TAAGGGAGGT
ZmEPSPS_ZM  .C.....T.. ...C..C.... TC..... ..G.....T.. C.....C..A

901                               950
ZmEPSPS_Nat CAAA AATACA AGTCCCCTAA AAATGCCTAT GTTGAAGGTG ATGCCCTCAAG
ZmEPSPS_ZM  ..G.....T. ....A..G.. G.....G..C ..C..G..A. .C..A..G..

951                               1000
ZmEPSPS_Nat CGCAAGCTAT TTCTTGGCTG GTGCTGCAAT TACTGGAGGG ACTGTGACTG
ZmEPSPS_ZM  ...G..T..C ...C.T..G. .C.....C.. C..C..G..A ..C.....A..

```

Figura 2D

```

1001                               1050
ZmEPSPS_Nat TGGAAAGGTTG TGGCACCACC AGTTTGCAGG GTGATGTGAA GTTTGCTGAG
ZmEPSPS_ZM   . . . . . C . . . . . G . . . . . A . . . . . C . . . . . C . . . . . A . . . . .
                . . . . . C . . . . . G . . . . . C . . . . . C . . . . . C . . . . . A . . . . .

1051                               1100
ZmEPSPS_Nat GTACTGGAGA TGATGGGAGC GAAGGTTACA TGGACCCGAGA CTAGCGTAAC
ZmEPSPS_ZM   ..G..A..A. . . . . C . . . . . T . . . . . G . . . . . T . . . . . .GTC...G..
                . . . . . C . . . . . C . . . . . T . . . . . G . . . . . T . . . . . .GTC...G..

1101                               1150
ZmEPSPS_Nat TGTTACTGGC CCACCGGGG AGCCATTTGG GAGGAAACAC CTC AAGGCCGA
ZmEPSPS_ZM   C . . . . . G . . . . . A . . . . . G . . . . . C . . . . . A . . . . . T . . . . . C . . . . .
                . . . . . G . . . . . A . . . . . G . . . . . C . . . . . A . . . . . T . . . . . C . . . . .

1151                               1200
ZmEPSPS_Nat TTGATGTCAA CATGAACAAG ATGCCTGATG TCGCCATGAC TC TTGCTGTG
ZmEPSPS_ZM   . . . . . G . . . . . T . . . . . G . . . . . C . . . . . T . . . . . A . . . . . C . . . . .
                . . . . . G . . . . . T . . . . . G . . . . . C . . . . . T . . . . . A . . . . . C . . . . .

1201                               1250
ZmEPSPS_Nat GTTGCCCTCT TTGCCGATGG CCCGACAGCC ATCAGAGACG TGGCTTCCTG
ZmEPSPS_ZM   ..G.....G. .C..T..C.. . . . . C . . . . . C . . . . . A . . . . . G . . . . . T . . . . . C . . . . .
                . . . . . G . . . . . T . . . . . C . . . . . C . . . . . A . . . . . G . . . . . T . . . . . C . . . . .

```

Figura 2E



```

1251                               1300
ZmEPSPS_Nat GAGAGTAAAG GAGACCGAGA GGATGGTTGC GATCCGGACG GAGCTAACCA
ZmEPSPS_ZM   ...G..C... ..A...C .T.....G.. ..A.... ..G..T.

1301                               1350
ZmEPSPS_Nat AGCTGGGAGC ATCTGTTGAG GAAGGGCCGG ACTACTGCAT CATCAGGCCG
ZmEPSPS_ZM   .A..C..G.. CAG...G... ..G..C.... .T.....A....A..T

1351                               1400
ZmEPSPS_Nat CCGGAGAAGC TGAACGTGAC GCGGATCGAC ACGTACGACG ACCACAGGAT
ZmEPSPS_ZM   ..A.....T ..C...C.. C..T.....A......T...C....

1401                               1450
ZmEPSPS_Nat GGCCATGGCC TTCTCCCTTG CCGCCTGTGC CGAGGTCCCC GTCACCATCC
ZmEPSPS_ZM   ...A..... ..TAG.T.G. .A..G..C.. ..A..A..T ..G..T..AA

1451                               1500
ZmEPSPS_Nat GGGACCCTGG GTGCACCCGG AAGACCTTCC CCGACTACTT CGATGTGCTG
ZmEPSPS_ZM   .A..T..A.. T.....C ..A..G..T. ....T... ..C..C..C

1501                               1521
ZmEPSPS_Nat AGCACTTTCG TCAAGAATTA A
ZmEPSPS_ZM   TCA..C.... .G.....C.G .

```

Figura 2F

```

1          50
GmEPSPS_Nat ATGGCCCAAG TGAGCAGAGT GCACAATCTT GCTCAAAGCA CTCAAATTTT
GmEPSPS_GM  ....T..G. .CTCTC.C.. T..T.....C .....G..T. .C..G..A..

51          100
GmEPSPS_Nat TGGCCATTCT TCCAACCTCA ACAAACTCAA ATCGGTGAAT TCGGTTTCAT
GmEPSPS_GM  C..A.....C AGT.....A. ....A.. G..T..... AGT..A...C

101         150
GmEPSPS_Nat TGAGGCCACG CCTTTGGGG GCCTCAAAAT CTCGCATCCC GATGCATAAA
GmEPSPS_GM  .TC....T.. G..G.....A ..AAGT..GA GC...T..... T.....C..G

151         200
GmEPSPS_Nat AATGGAAGCT TTATGGGAA TTTTAAATGTG GGAAGGGAA ATTCCGGCGT
GmEPSPS_GM  ..C..TTCG. .C.....G.. C.....C..C ..C..A..... .C..A..T..

201         250
GmEPSPS_Nat GTTTAAGGTT TCTGCATCGG TCGCCGCCGC AGAGAAGCCG TCAACGTCGC
GmEPSPS_GM  C..C..A..A AGC..CAGC. .A..T..G.. T.....C AGT..T..T.

```

Figura 3A

```

251                               300
GmEPSPS_Nat CGGAGATCGT GTTGAACCC ATCAAAGACT TCTCGGGTAC CATCACATTG
GmEPSPS_GM .T..A..T.. TC.T.....G ..A..G..T. ....A..... G..T...C..A

301                               350
GmEPSPS_Nat CCAGGGTCCA AGTCTCTGTC CAATCGAATT TTGCTTCTTG CTGCTCTCTC
GmEPSPS_GM ..T..A..A. ..AG...C.. T...A..... ...T.G..C. .A.....G..

351                               400
GmEPSPS_Nat TGAGGGAACA ACTGTTGTAG ACAACTTGTT GTATAGTGAG GATATTTCATT
GmEPSPS_GM C..A.....C .....A..C. .T...C.CC. T.....C..A .....A.....

401                               450
GmEPSPS_Nat ACATGCTTGG TGCATTAAGG ACCCTTGGAC TGCGTGTGGA AGATGACAAA
GmEPSPS_GM .T...T.G.. G..GC.C..A ..T.....G. .AA.A..T.. G..C..T..G

451                               500
GmEPSPS_Nat ACAACCAAAC AAGCAATTGT TGAAGGCTGT GGGGGATTGT TTCCCACACTAG
GmEPSPS_GM ..T..T..... ....T..C.. C.....T... ..C..G..... .C..T...TC

```

Figura 3B

```

501
GmEPSPS_Nat TAAGGAATCT AAAGATGAAA TCAATTTATT CCTTGGAAT GCTGGTATCG 550
GmEPSPS_GM ...A...AG. ....G. .A..C..G.. T.....C..C ..A..A.....

551
GmEPSPS_Nat CAATGAAGTC CTTGACAGCA GCTGTGGTTG CTGCAGGTGG AAATGCAAGC 600
GmEPSPS_GM .....AG .C.C..C..T .....C.... .G..G..... T..C..T..T

601
GmEPSPS_Nat TACGTACTTG ATGGGTGCC CCGAATGAGA GAGAGGCCAA TTGGGGATTT 650
GmEPSPS_GM .....CT.A. .C.C..... TA.....C.. .A..A..T. .C..T....C.

651
GmEPSPS_Nat GGTTGCTGGT CTTAAGCAAC TTGGTGCAGA TGTTGATTGC TTTCTTGGCA 700
GmEPSPS_GM A..G.....C ..A..A..G. ....A..... C..C.....T ..CT.G....

701
GmEPSPS_Nat CAAACTGTCC ACCTGTTCGT GTAAATGGGA AGGGAGGACT TCCTGGCGGA 750
GmEPSPS_GM ....T..C.. G..C..GA.A ..G..C..... .....CT. G..A.....T

```

Figura 3C

```

751                                     800
GmEPSPS_Nat AAGGTGAAAC TGTCTGGATC AGTTAGCAGT CAATAC TTGA CTGCTTTGCT
GmEPSPS_GM   . . . . . T . . . . . A . C . . . . . G . . CTCGTCA . . G . . C . A . . . . . A . . . . .

801                                     850
GmEPSPS_Nat TATGGCAGCT CCTTTAGCTC TTGGTGATGT GGAAATTGAG ATTGTTGATA
GmEPSPS_GM   C . . . . . C . C . . GC . C . . . T . G . . G . . C . . . . G . . . . . A . . C . C . . . .

901                                     900
GmEPSPS_Nat AACTGATTTTCTGTTCCATAT GTTGAAATGA CTCTGAAGTT GATGGAGCGT
GmEPSPS_GM   .GT . . . . . AG C . G . T . . . . . G . . . . . . . C . C . A . . . . . . . A . G

951                                     950
GmEPSPS_Nat TTTGGAGTTTCTGTGGAACA CAGTGGTAAT TGGGATAGGT TC TTGGTCCA
GmEPSPS_GM   .C . . . . . . G . A . . . . . . TCC . G . . . . . C . . . . . TC . T . A . .

951                                     1000
GmEPSPS_Nat TGGAGGTCAA AAGTACAAGT CTCCTGGCAA TGCTTTTGTG GAAGGTGATG
GmEPSPS_GM   C . . . . . G . . . . . . . . . . . AA GC . A . . . . . . . C . C . C . . . . . G . . C .

```

Figura 3D

```

1001                               1050
GmEPSPS_Nat CTTCAAGTGC CAGTTATTTA CTAGCTGGTG CAGCAATTAC TGGTGGGACT
GmEPSPS_GM   ...G..C.. TTCC...C.C .C.....C. ....C..A.. C.....C..C

1051                               1100
GmEPSPS_Nat ATCACTGTTA ATGGCTGTGG CACAAGCAGT TTACAGGGAG ATGTAATAATT
GmEPSPS_GM   .A..C..G. .C.....C.. ...CTCATCC C.T..A..T. ....G...G...

1101                               1150
GmEPSPS_Nat TGCTGAAGTT CTTGAAAAGA TGGGAGCTAA GGTACATGG TCAGAGAACA
GmEPSPS_GM   C.....G..C T.G..G..A. ....C..A.. ...C..... .T.....

1151                               1200
GmEPSPS_Nat GTGTCAC TGT TCTGGACCA CCACGAGATT TTTCTGGTCG AAAAGTCTTG
GmEPSPS_GM   .C..A..C.. G..C.....T ..CA....C. ..CG....A. ...G....C.T

1201                               1250
GmEPSPS_Nat CGAGGCATTG ATGTCAATAT GAACAAGATG CCAGATGTTG CCATGACACT
GmEPSPS_GM   A.G..A..A. ....G..... ...T..... .C.....G. .T.....G...

```

Figura 3E

```

1251                               1300
GmEPSPS_Nat TGCTGTTGTT GCACTATTTG CTAATGGTCC CACTGCTATA AGAGATGTGG
GmEPSPS_GM C.....C ..C..G..C. .A..C..A.. T..C..A.... ..G.....C..C.

1301                               1350
GmEPSPS_Nat CAAGTTGGAG AGTTAAAGAG ACTGAGAGGA TGATAGCAAT CTGCACAGAA
GmEPSPS_GM .TTCA...C. T.....G..A ..C..AC... ..C..T.. T.....C..G

1351                               1400
GmEPSPS_Nat CTCAGAAAGC TAGGAGCAAC AGTTGAAGAA GGTCCCTGATT ACTGTGTGAT
GmEPSPS_GM T.GC.T.... .G..T..... G..G..... ..A..A..C. .T..C.....

1401                               1450
GmEPSPS_Nat TACTCCACCT GAGAAATTGA ATGTCACAGC TATAGACACA TATGATGACC
GmEPSPS_GM A..A..T... ..A..GC.C. ....G..C.. ...T.....T ..C..C..T..

1451                               1500
GmEPSPS_Nat ACAGAATGGC CATGGCATTG TCTCTTGCTG CTTGTGGGA TGTTCCAGTA
GmEPSPS_GM .....T..... ..A..... ..C..T.. C..G..G..T

```

Figura 3F

```
1501                                     1550
GmEPSPS_Nat ACCATCAAGG ATCCTGGTTG CACCAGGAAG ACATTTCCCTG ACTACTTTTGA
GmEPSPS_GM  ..G..... .C..A..G.. T..T..... .C...A. .T.....

1551                                     1578
GmEPSPS_Nat AGTCCTTGAG AGGTTAACAA AGCACTAA
GmEPSPS_GM  G..GT.G..A ..A..G.... ..G....G.
```

Figura 3G



```

1      50
BAR1_Nat ATGAGCCCCAG AAGACGCCC GGCAGACATC CGCCGTGCCA CCGAGGCGGA
BAR1_ZM   ...TCG..T. .G..C..T.. T..T.....A A.A..C..T. ....A...
BAR1_AT   .....T..... A.GA.A.. ..T..T.. GA.A.... ..A..T..

51      100
BAR1_Nat CATGCCGGGG GTCTGCACCA TCGTCAACCA CTACATCGAG ACAAGCACGG
BAR1_ZM   .....T..T ..T..... T..G..T.. .....TCT....
BAR1_AT   T.....T..T ..T..T.A. ....A..... T..T..... ..CTCG..A.

101     150
BAR1_Nat TCAACTTCCG TACCGAGCCG CAGGAACCGC AGGACTGGAC GGACGACCTC
BAR1_ZM   .A.....C..T.....T ..A..... T..... C.....T...
BAR1_AT   .T..T..T.. C..T..... G..A. ....T..... ..T..G

151     200
BAR1_Nat GTCCGTCTGC GGGAGCGCTA TCCCTGGCTC GTCGCCGAGG TGGACGGCGA
BAR1_ZM   ..G.....CA .A.....T.. ...G.....G ..T..A.... T...
BAR1_AT   ..A...T.AA .A..A..T.. ...G.....A ..T..T.... T.....A..

```

Figura 4A

```

201
BARI_Nat GGTGCGCGGC ATGCCTACG CGGGCCCTG GAAGGCACGC AACGCCTACG 250
BARI_ZM A..G..T..G .....

```

```

251
BARI_Nat ACTGGACGGC CGAGTCGACC GTGTACGTCT CCCCCGCCA CCAGCGGAGC 300
BARI_ZM .T.....T.. G.....C..A ..C..... .A...A.A.. T..AA.A..C
BARI_AT .....A.. A..A..C..T ..C.....A G...T..T.. T..AA.A..C
```

```

301
BARI_Nat GGACTGGGCT CCACGCTCTA CACCACCTG CTGAAGTCCC TGGAGGCACA 350
BARI_ZM ..G..C..... .G..C..... T..G..T..C ..C.....T .A.....G..
BARI_AT ...T.A..GA G.....T.G.. ...T..T..T T.A.....A. ....
```

```

351
BARI_Nat GGGCTTCAAG AGCGTGGTGG CTGTCATCGG GCTGCCCAAC GACCCGAGCG 400
BARI_ZM .....A TCT..A..G. .G..G..... CT....A.... .T..C..T.
BARI_AT A..... TCT..T..G. .A..T..T.. AT.....A.... .T.....T.
```

Figura 4B

```

401          TGGGCATGCA CGAGGGGCTC GGATATGCC CCGGGGCAT GCTGGGGCGG
BAR1_Nat          450
BAR1_ZM  ..A.A..... ..A..... .T..C..T. .TA.A..A.. ..CA.....
BAR1_AT  .T..A..... .A.....T ..C..T. .A..A..T.. ..C..T..T

451          GCCGGCTTCA AGCACGGGAA CTGGCATGAC GTGGGTTTCT GGCAGCTGGA
BAR1_Nat          500
BAR1_ZM  ..T..A.... ..A.....T.. T.....C... ..T..C.... ..A.....
BAR1_AT  .....A.... .A..T..A.. T.....C... ..A.....T. ....A.....

501          CTTCAGCCTG CCGGTACCGC CCCGTCCGGT CCTGCCCGTC ACCGAGATCT
BAR1_Nat          550
BAR1_ZM  ...TCTT.. ..A..T..A. .T.....T.. G..A..... ..A.....
BAR1_AT  ...TTCA..T ..C..T..C. .TA.A..T.. A..T..A..T ..T..A.....

551          BAR1_Nat GA
BAR1_ZM  AG
BAR1_AT  AG

```

Figura 4C

```

1      ATGGGCGCAAG TTAGCAGAAT CTGCAATGGT GTGCAGAACC CATCTCTTAT CTCCAATCTC      60
      .....A..T.....C..C.....G..G..G..G.....C...
      .....C..G..A..T..G..T..C..A.....T..A..A..A..A..G..C..C..G
      .....C.....C.....C.....T.....A..A.....C..G

61     TCGAAATCCA GTC AACGCAA ATCTCCCTTA TCGGTTTCTC TGAAGACGCA GCAGCATCCA      120
      AGC..G....C..G..G..G..G..GC.C.....CAGC..C.....C..A.....C..G
      ..A...AG.T C....A..A..G..G..GC.T ..T..A..GT ...A..T..A.....C..G
      ..C.....GT CA.....T..G..G..A....T...AGCT .....T...A.....T

121    CGAGCTTATC CGATTTTCGTC GTCGTGGGA TTGAAGAAGA GTGGGATGAC GTTAATTGGC      180
      A..G..C..C..T..CAGC..A..C.....C C.C.....C.....C..G..C...
      A..G.....C..C..AAG C..C.....T C.A.....A.....A.....AC.G..C..T
      ..C..A....T..A..AAG CAGT.....T.....AT CG..T.....C..G.....T

181    TCTGAGCTTC GTCCTCTTAA GGTCATGTCT TCTGTTTCCA CGGCGTGCAT GCTTCACGGT      240
      AGC.....G..G..A..C...G.....C..G..C..A.....A.....
      AGC..A..A..A..G..G..A.....C..A..CAG..T.....T.....T..A
      ..G..A...A..G..AT.G...G.....A...AGT..A..T.....

```

Figura 5A

```

241          GCAAGCAGCC GGCCCGCAAC CGCCCGCAAA TCCTCTGGCC TTTCCGGAAC CGTCCGCATT      300
CTP2CP4_ZM      . . . . . G . . . . . T . . . . . T . . . . . C . . . . . G . . . . . G . . . . . A . G . . . .
CTP2CP4_AT      ..TTCATCTA . . . . . A . T . . T . . A . G . . G . . TAGC . . G . . CAGT . . C . . . . . G . . . . . C
CTP2CP4_GM      ..TTCTTC.A A . . . . . . . . . . G . . TA . A . . G AGT . . . . . T . . G . . T . . . . . . . . . . T . . . .
301          CCCGGGACA AGTCGATCTC CCACCGGTCC TTCATGTTCCG GCGGTCTCGC GAGCGGTGAA      360
CTP2CP4_ZM      ..G . . . . . T . . . . . C . . A . . . . . . . . . . . G . . . . . . . . . . . T . . C . . . . . . . . . . G
CTP2CP4_AT      ..T . . . . . T . . AAGT . . T . . A . . A . . AG . . . . . . . . . . . A . A . . T . . T . . T . . A . . G
CTP2CP4_GM      ..A . A . . . . . AAGC . . TAG T . . . . . C . . T . . . . . T . . . . . T . . G . . G . . ATCT . . A . . G
361          ACGGCATCA CCGGCCTTCT GGAAGGCGAG GACGTCATCA ATACGGGCAA GGCCATGCAG      420
CTP2CP4_ZM      . . . . . . . . . . G . . . . . G . . T . . . . . T . . . . . G . . . . . C . . G . . . . . . . . . . .
CTP2CP4_AT      . . A . A . . . . . T . . TT . G . . T . . G . . . . . A . . T . . T . . . . . C . . C . . T . . . . . G . . . . . A
CTP2CP4_GM      . . . . . . . . . . T . . T . . . . . . . . . . . A . . . . . . . . . . . A . G . . . . . A . . . . .
421          GCGATGGCG CCCGCATCCG TAAGGAAGGC GACACCTGGA TCATCGATGG CGTCGGCAAT      480
CTP2CP4_ZM      ..T . . . . . A . . G . . T . . . . . C . . . . . . . . . . T . . . . . A . . . . . T . . C . . . . . T . . G . .
CTP2CP4_AT      ..A . . . . . T . . A . A . . . . . A . A . G . . . . . T . . G . . . . . C . . . . . C . . T . . T . . T . .
CTP2CP4_GM      ..T . . . . . T . . . . . T . . T . . C . . . . . . . . . . T . . T . . . . . A . C . . A . . T . . G . . ,

```

Figura 5B

```

481      CTP2CP4_NAT      540
GGCGGCCTCC TGGCGCCTGA GCGGCCGCTC GATTTCCGGCA ATGCCGCCAC GGGCTGCCCG
...T..G. .C..C.... .C..T... .C..... .G..G.. .G.....A..G
..A..AT..G. .C..T..C.. A.....A..T .C..T..G. .C..A..T.. .G.....T
..T...T..A. .T..A..G.. ..T..T... .C..T.... .C..A..... A..G..TA..A

541      CTP2CP4_NAT      600
CTGACGATGG GCCTCGTCGG GGTCTACGAT TTCGACAGCA CCTTCATCGG CGACGCCTCG
..C..T.... .A..G..... .G.....C ..T.... .G..... A.....
..T..T.... .A..G..A.. C..G..T..C ..T...TCT. ....T.... GAGC
..T..T.... ..G..... T..T..... ..TTCA. ....T..T.. G..T.....T

601      CTP2CP4_NAT      660
CTCACAAAGC GCCCGATGG CCGCGTGTG AACCCGCTGC GCGAAATGGG CGTGCAGGTG
.....A..... .A.....TC.. ..T...T... ..G..... .A.....C
...T...A .A..A..... A..A...C.. .T..C...A .G..G..... T..C.....
...T..A. ....A..... AA..A..CC.T .....T..A .G..G..... .A..A..T

661      CTP2CP4_NAT      720
AAATCGGAAG ACGGTGACCG TCCTCCCGTT ACCTTGGCGG GGCCGAAGAC GCCGACGCCG
....C..G. .T..... .T.G..... .GC..... .G..... .T..C...
...T..G. .T.....T.. ..G.....TC....A. .C..C..... C..C.....A
..G..C..G. ....C...A. AT.G.....C .....C..T..... A..C..C..T

```

Figura 5C

```

721                                     780
CTP2CP4NAT ATCACCTACC GCGTGCCGAT GGCCTCCGCA CAGGTGAAGT CCGCCGTGCT GCTCGCCGGC
CTP2CP4_ZM .T.....A...A.....C.....C.....A.....A.....C.....A
CTP2CP4_AT .G...A .G..T.....G..A...G..A...A..G..A...C..G..G...
CTP2CP4_GM .T...A .G..T..A...A..T..T...A.....A..T... ..T...T..A
781                                     840
CTP2CP4NAT CTC AACACGC CCGGCATCAC GACGGTCATC GAGCCGATCA TGACGCGCGA TCATACGGAA
CTP2CP4_ZM .G.....T. .G.....C.....G... ..C.....CA.G.. ..C.....C...
CTP2CP4_AT .....A. .T..A.....A..C..G..T ..A..C.....TA.A.. C..C.....G
CTP2CP4_GM T.G.....A. .G..T..T..T..C..G..T .....T...T... C..C..T..G
841                                     900
CTP2CP4NAT AAGATGCTGC AGGGCTTTGG CGCCAACCTT ACCGTCGAGA CGGATGCGGA CGGC GTGCGC
CTP2CP4_ZM .....T. ....G.....G.....G.....G.....C..T... ..CA.G
CTP2CP4_AT .....T... ..T..C... ..T..T..A ..G.....A. .C..C..C... ..A.G
CTP2CP4_GM .....T. ....T..C... T..T.....C .....T..A. .A..C.....T...A.G
901                                     960
CTP2CP4NAT ACCATCCGCC TGGAAGGCCG CGGCAAGCTC ACCGGCCAAG TCATCGACGT GCCGGGGCAG
CTP2CP4_ZM .....T..G...A. G..T..A..G ..T.....T... ..T...A...
CTP2CP4_AT .A.....T .....G...A. A..T..A..G ..T.....T... ..T..A..T
CTP2CP4_GM .....T..... ..G..AA. G..A..A... ..T..T.....T.....C..T..T

```

Figura 5D

```

961
CTP2CP4_NAT      1020
CGGTCCTCGA CGGCCTTCCC GCTGGTTGCG GCCCTGCTTG TTCCGGGGCTC CGACGTCACC
...G..C. ....T..... .G.....C. .G....GAG T....G...
.C..G..C. .A..G..T.. C..C..A..T ..GT....C. .C..T..A.. T..T..G..G
.C...AGC. ....G..... A.....C ..T..T..C. .A..A..... ..T..G..A

1021
CTP2CP4_NAT      1080
ATCCTCAACG TGCTGATGAA CCCACCCGC ACCGGCCTCA TCCTGACGCT GCAGGAAATG
...G.... .C..C..... .G...A.G .....G. ....C..... T.....G...
...G..T. .C..C..... T..A..TA.A .....C..AT. ....G...
.T..A.... .C..C..... T..T..TA.A .....AT.G. .A..T..AT. ....

1081
CTP2CP4_NAT      1140
GGCGCCGACA TCGAAGTCAT CAACCCGCGC CTTGCCGGCG GCGAAGACGT GGCGGACCTG
.A..... .G..G... ..T..... .G..A... ..T..... T.....T...
.T..T... ..G..T... ..T..TA.G T.G..A..T. .A..G..T.. ..C..T...
.T..T..T. T.....T... ..T..TA.A ..A.....A. .T..G..... A..T..TT..

1141
CTP2CP4_NAT      1200
CGCGTTCGCT CCTCCACGCT GAAGGGCGTC ACGGTGCCGG AGACCCGGC GCCTTCGATG
...G.... .T..C... ..T.....G .....C..... ..T..... ..G..C...
...G..T. .TAGT..A.. C..A.....G ..C..C..T. .G..T..... T..A..C...
.G..GA.G. T..T..AT. ...A..T..T ..C..A..T. ....TA.G.. A.....A...

```

Figura 5E



```

1201          1260
CTP2CP4NAT ATCGACGAAT ATCCGATTCT CGCTGTGCGCC GCCGCCTTCG CGGAAGGGGC GACCGTGATG
CTP2CP4_ZM  .A.....G. . . . .T..... G.C..... .T.G.... .C..... C..G..C...
CTP2CP4_AT  . . . . .G. .C..C..... .C..T..T .T..G..T. .C..G..C.. A..T..A...
CTP2CP4_GM  .T.....G. . . . .A..... T.C.....G .T.....T. .T..G..C.. . . . . .

1261          1320
CTP2CP4NAT AACGGTCTGG AAGAACTCCG CGTCAAGGAA AGCGACCGCC TCTCGGCCGT CGCCAATGGC
CTP2CP4_ZM  . . . . .T. .G..... . . . .G..... TCG..T.... .G.....G.. G.....
CTP2CP4_AT  . . . . .C..T. .G..GT.GA. G..T.....G .T...A.G. .G..C..G.. G..G.....
CTP2CP4_GM  .T..A..A. .G..GT.GA. A..G..... TC...A.AT .GAGC..A.. . .T..C..A

1321          1380
CTP2CP4NAT CTC AAGCTCA ATGGCGTGGA TTGGGATGAG GCGGAGACGT CGCTCGTCGT GCGTGGCCGC
CTP2CP4_ZM  .G..... .C..T..T.. C.....C... .T.....C. .A.....G.. C.....G
CTP2CP4_AT  .G.....A. .C..... .C.....C..A .T..A... .C..T..A.. C.....T...
CTP2CP4_GM  .T..A.... . . . . .T... .:..T..... .T.....TA GCT..G..A.. C.....G..A

1381          1440
CTP2CP4NAT CCTGACGGCA AGGGGCTCGG CAACGCCCTCG GGGCCGCCGG TCGCCACCCA TCTCGATCAC
CTP2CP4_ZM  . . . . .T... . . . .C..... . . . . .AGT .A..G.... . . . . .G.. C.....T
CTP2CP4_AT  .A.....G. . . . .T.G.. G..T..T... .A..T..T. .G..G..G.. C..T.....T
CTP2CP4_GM  .A.....A. . . . .TT.G.. . . . .T... .T..T.... .T..A..T.. CT.G.....T

```

Figura 5F

```

1441          CTP2CP4_NAT      1500
CGCATCGCCA TGAGCTTCCT CGTCATGGGC CTCGTGTCGG AAAACCCCTGT CACGGTGGAC
.....G.  ..TC....T. G..G.....T  ....C..A.  .G.....G.. G..C..C..T
A.A.....  ..TCA..T.. G..G.....A  ..T..C..C.  .G..T..G.. G..C..T...
..G..A..G.  ....T..T.. G..G.....T  ....AAGC.  .G..T..... G..A..C...

1501          CTP2CP4_NAT      1560
GATGCCACGA TGATCGCCAC GAGCTTCCCG GAGTTCATGG ACCTGATGGC CGGGCTGGGC
..C.....  ....A..G..  .........A  .........T..... G..C..C..G
....T..C.  ......... CTC....T..T  .........C..... A..CT.....G
....A..T.  .........T.. TTC.....T  ....T..... TT.A..... A..A..A..T

1561          CTP2CP4_NAT      1596
GCGAAGATCG AACTCTCCGA TACGAAGGCT GCCTGA
..C.....  ....G..T.. C.....C  ..T...
..C.....  ..G..G..T.. ..T.....C  ..T...
..A.....T.  .........T..  ...C..A..G  ....A.

```

Figura 5G

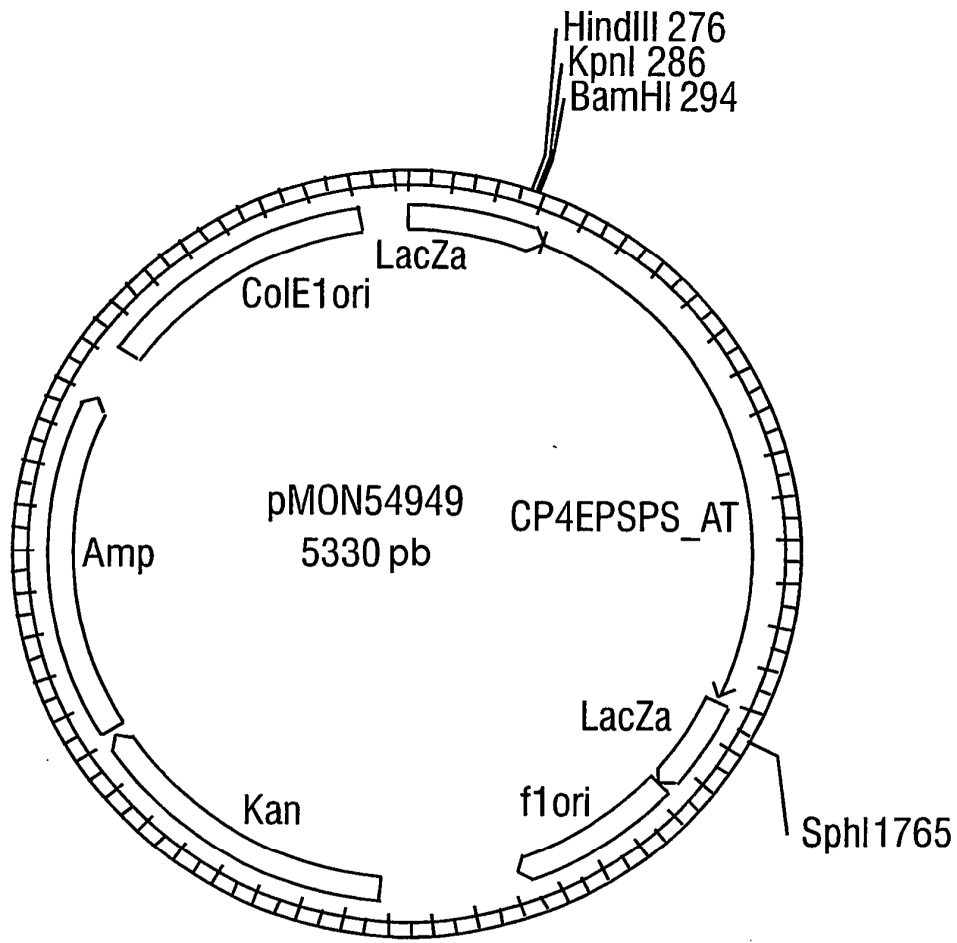


Figura 6

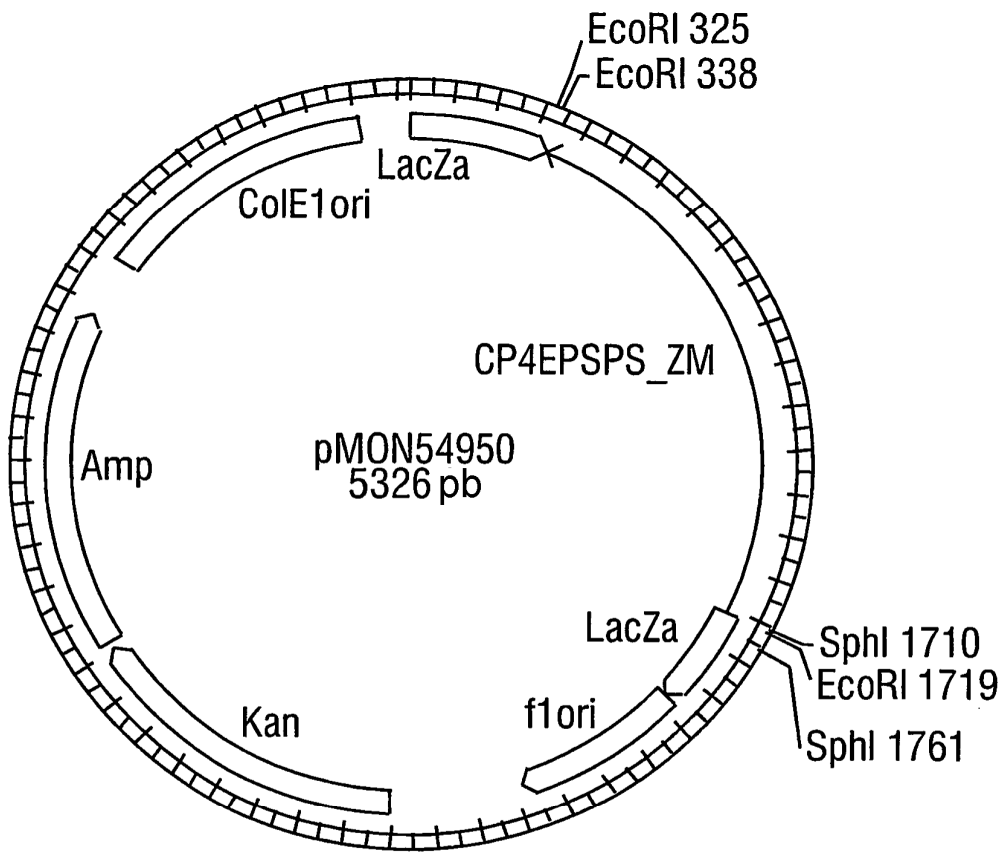


Figura 7

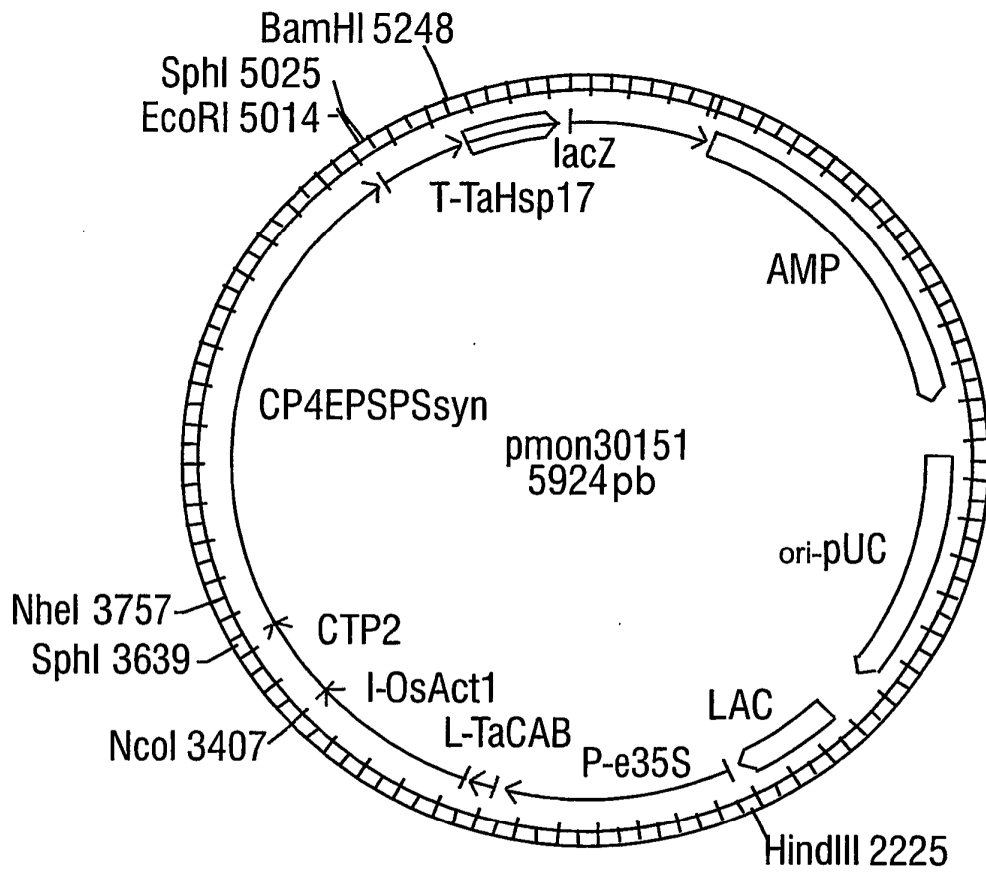


Figura 8

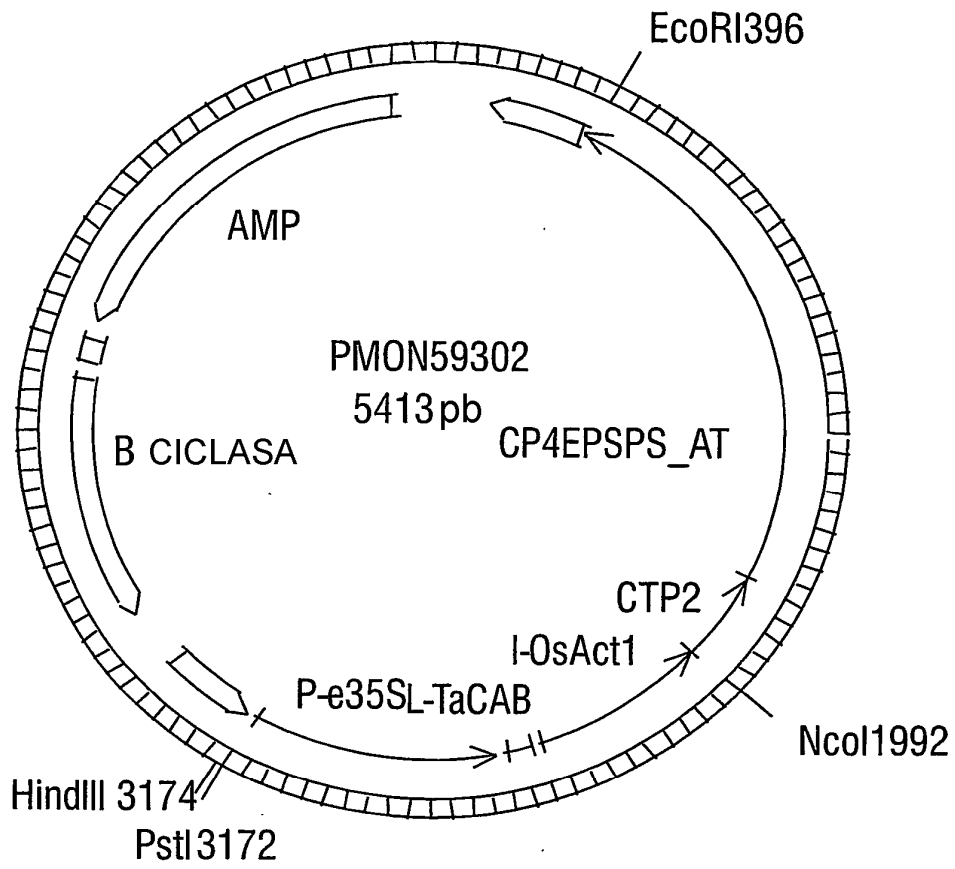


Figura 9

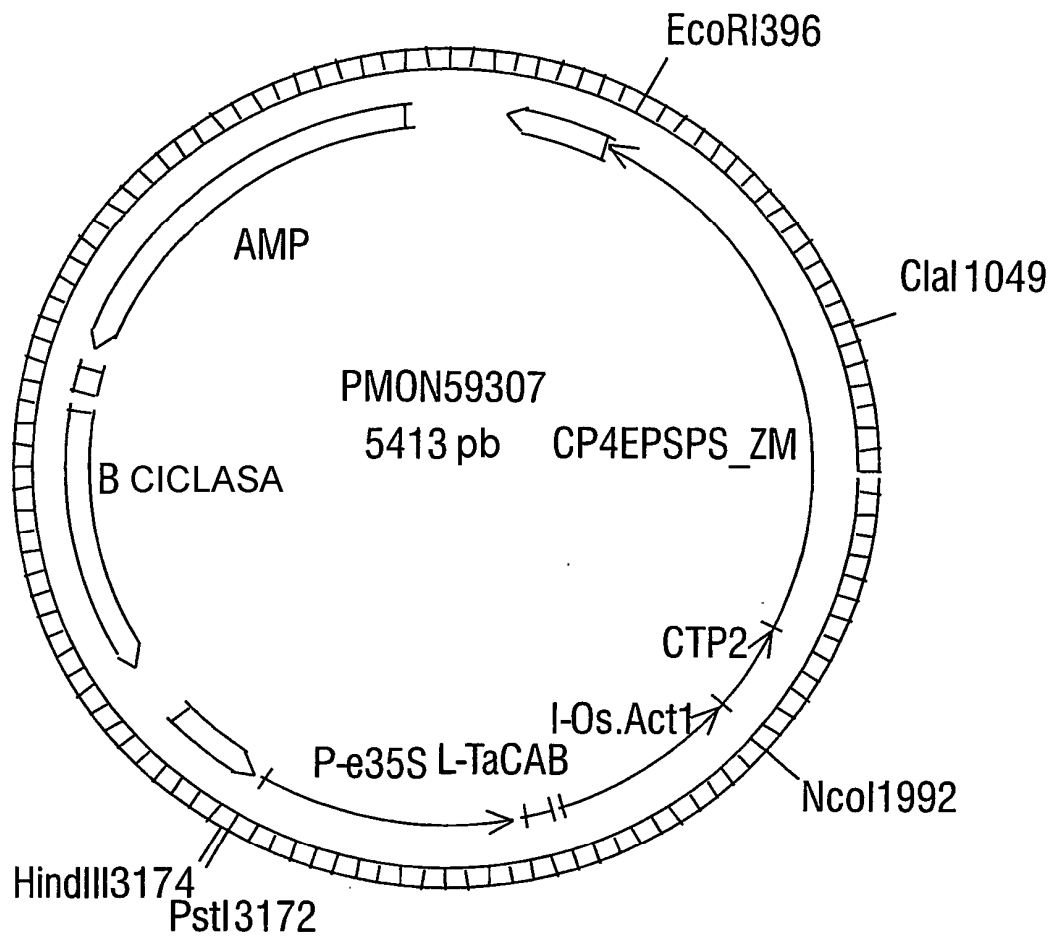


Figura 10

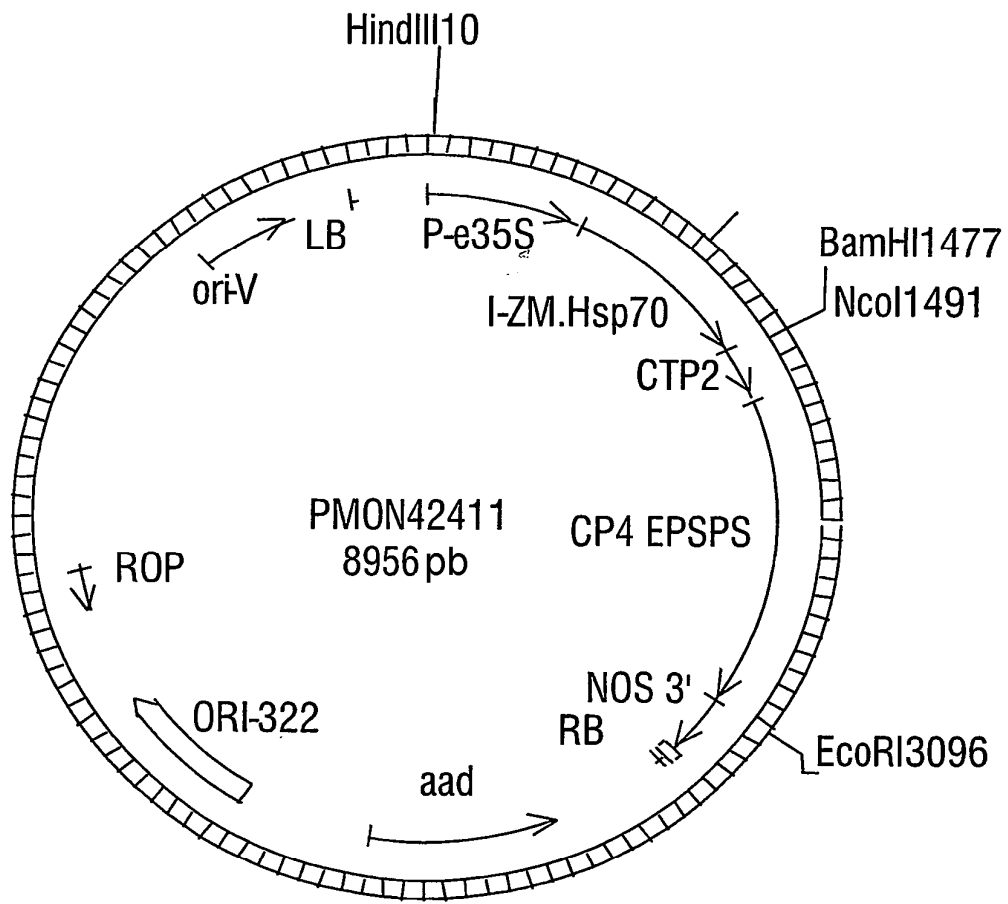


Figura 11



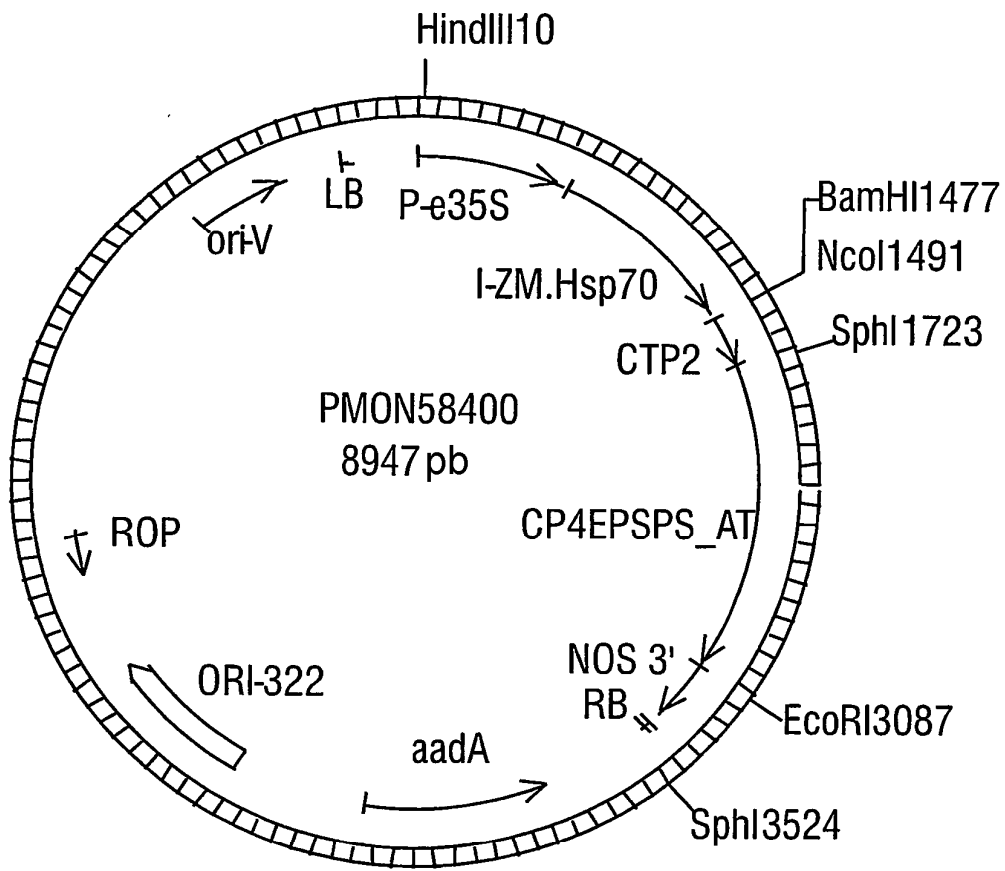


Figura 12

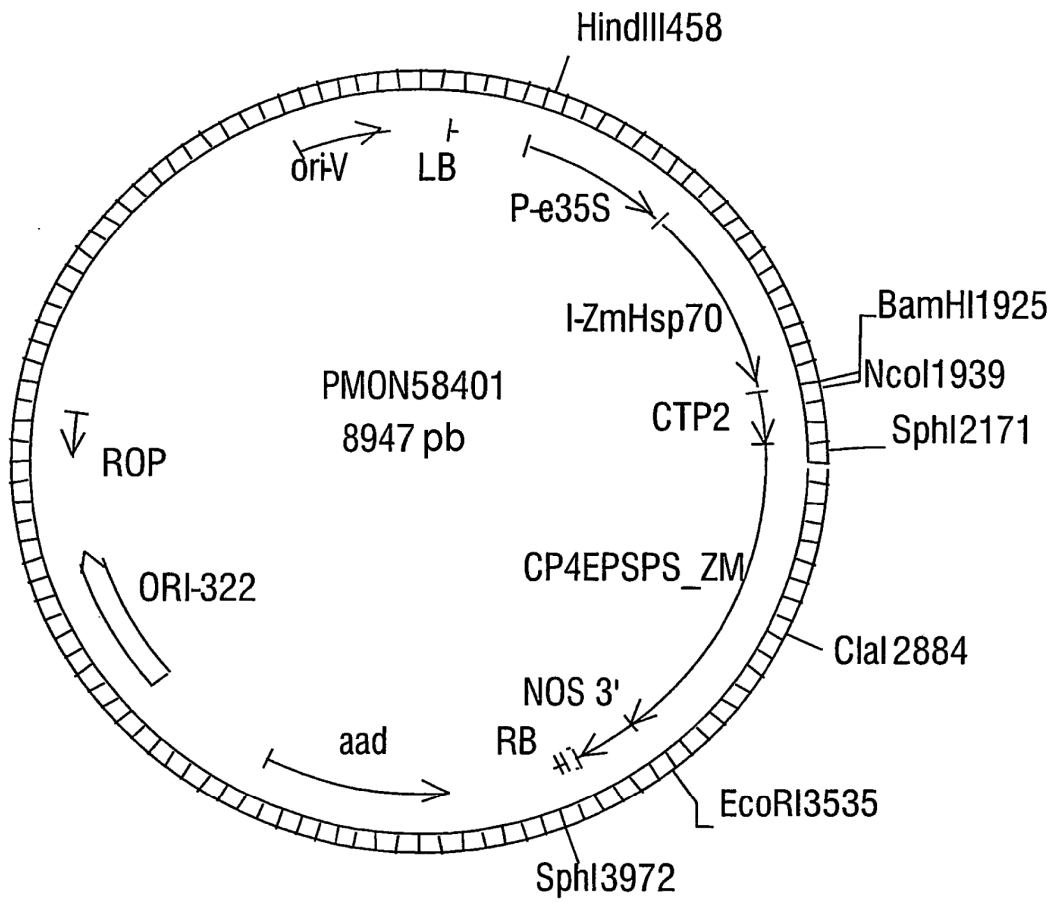


Figura 13

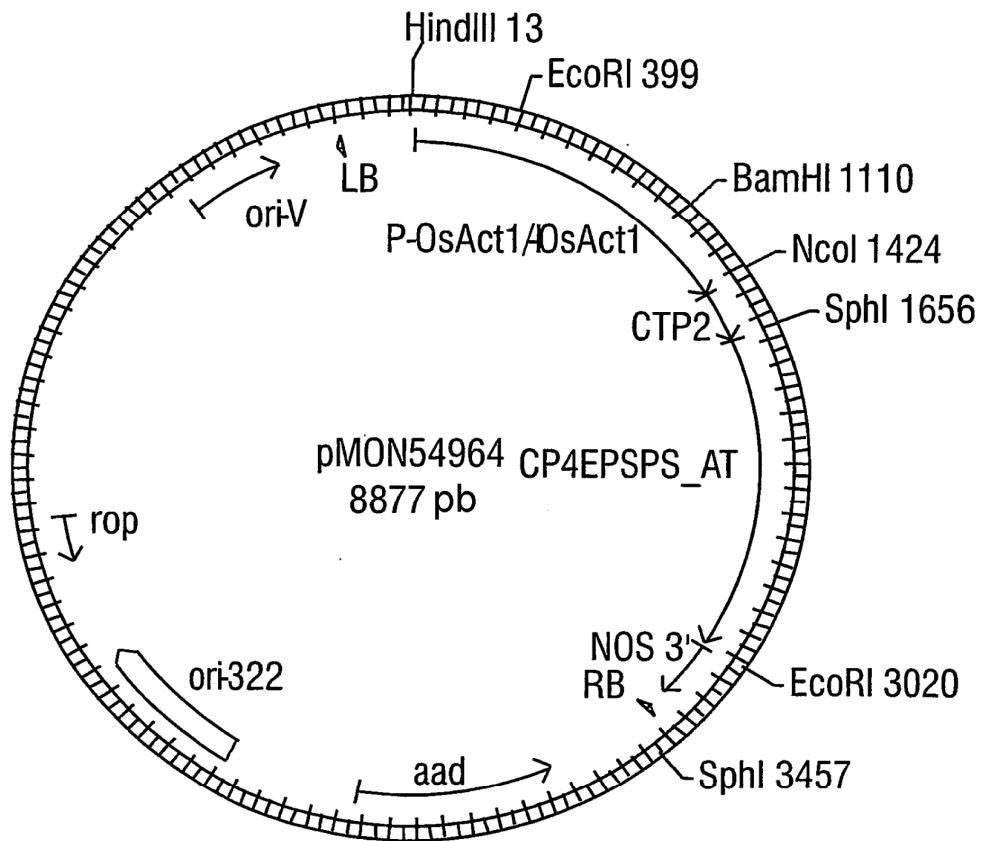


Figura 14

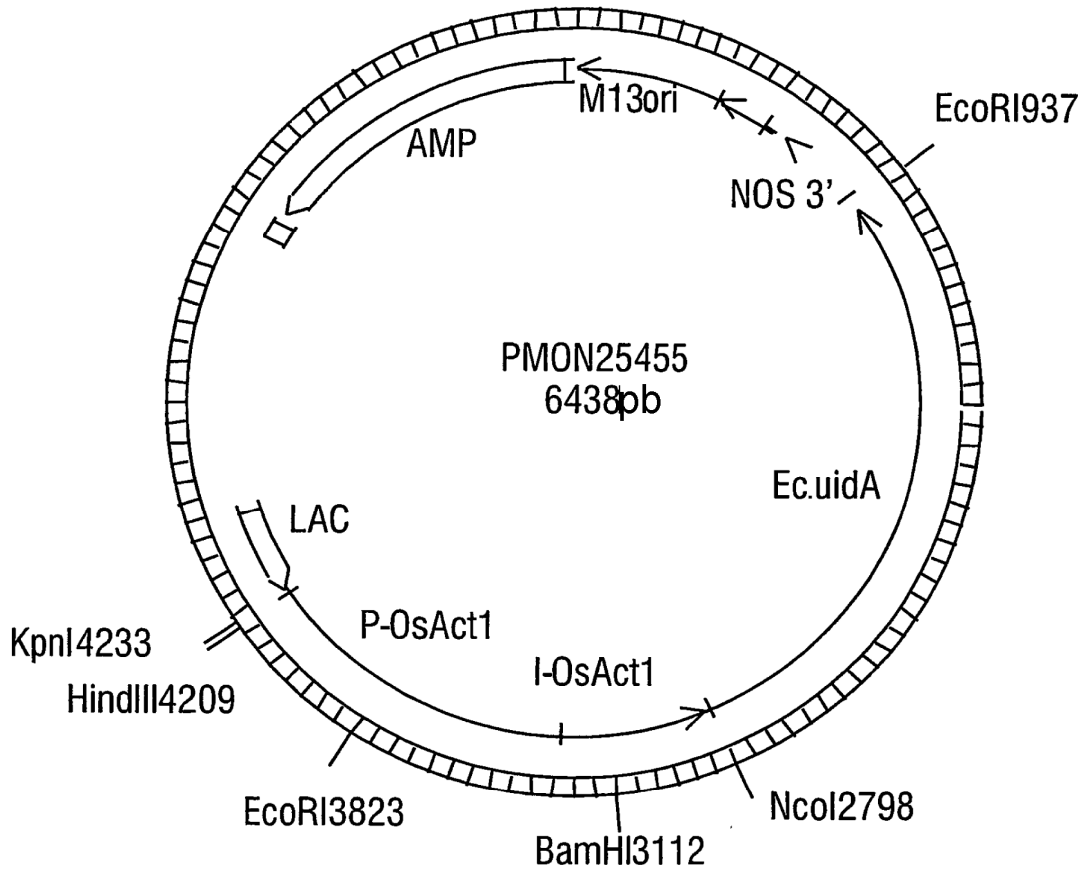


Figura 15

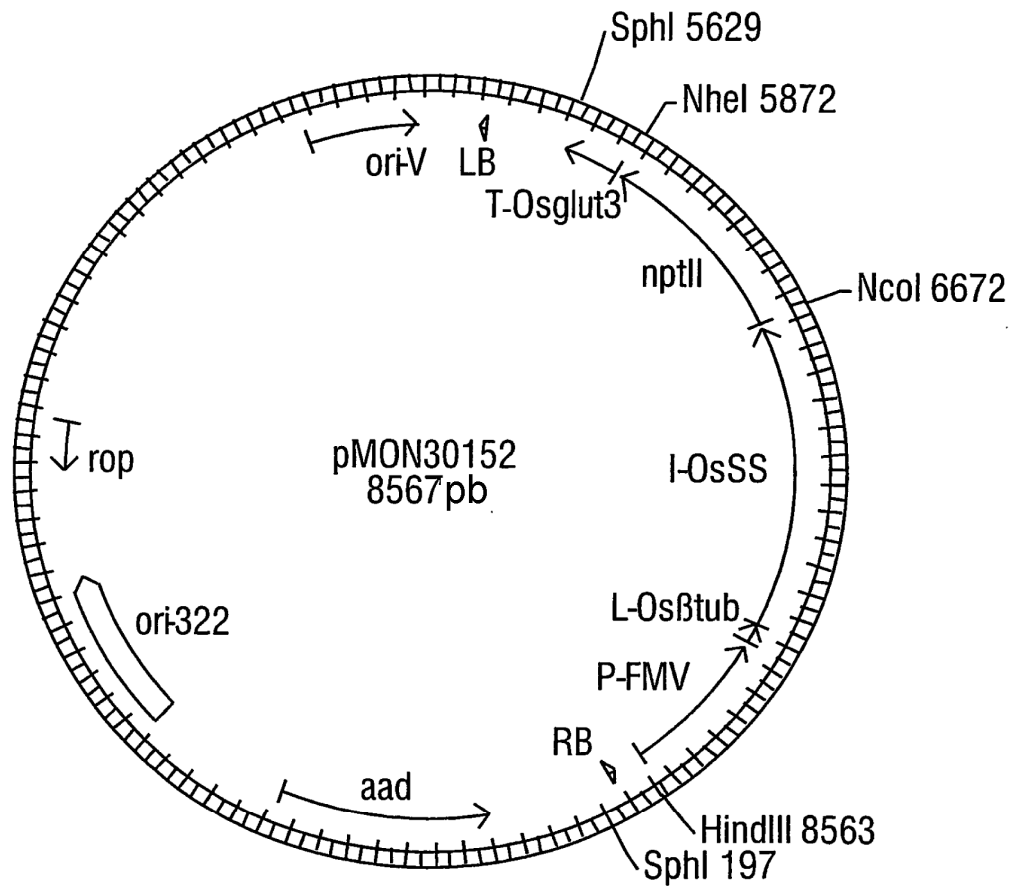


Figura 16

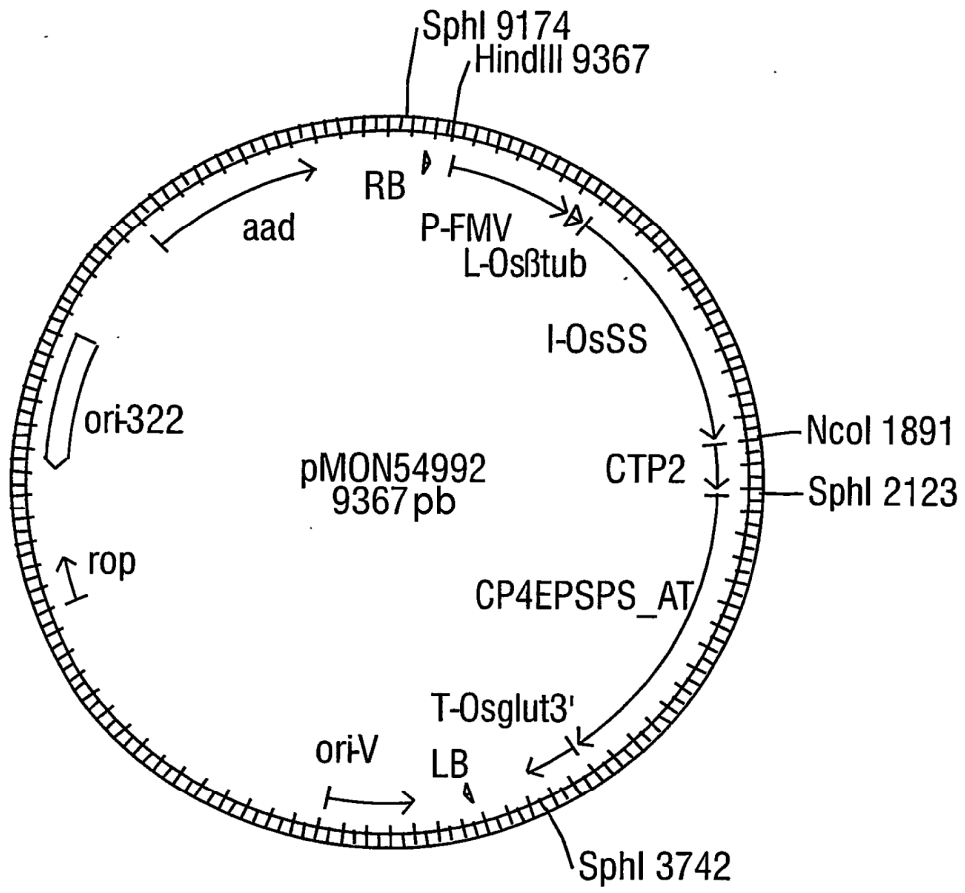


Figura 17

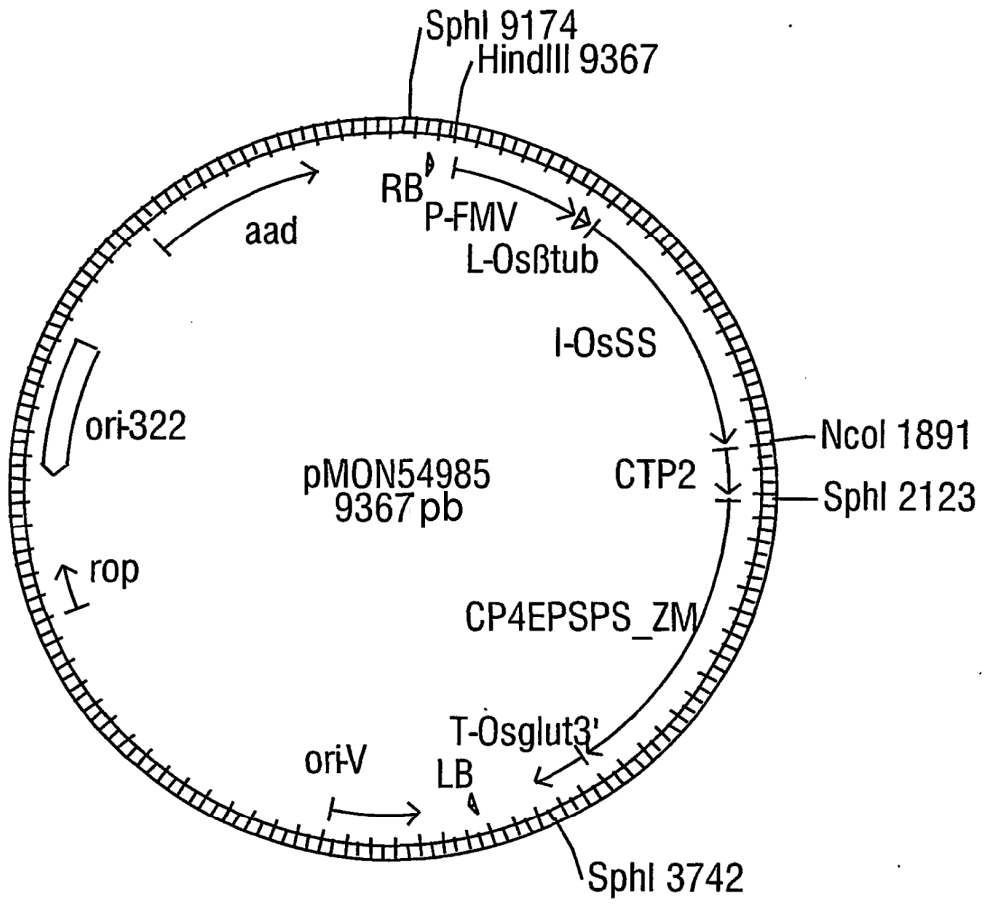


Figura 18

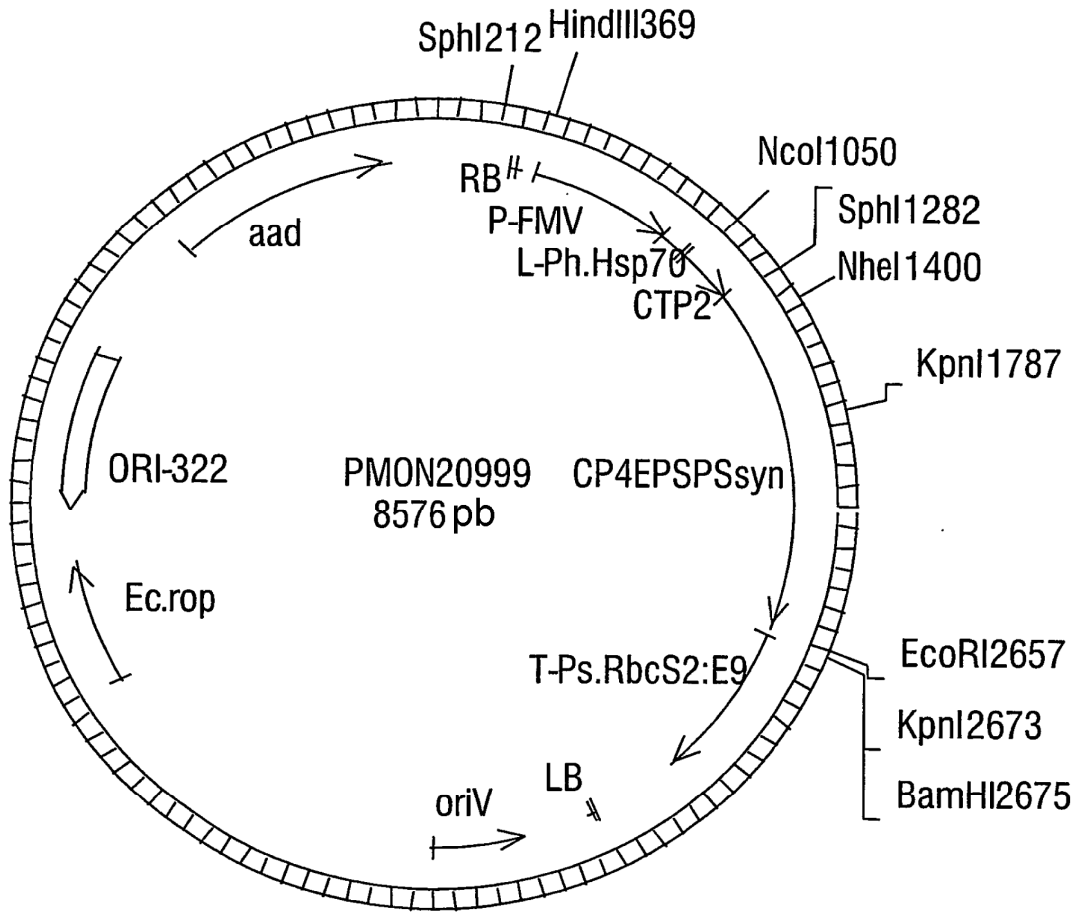


Figura 19



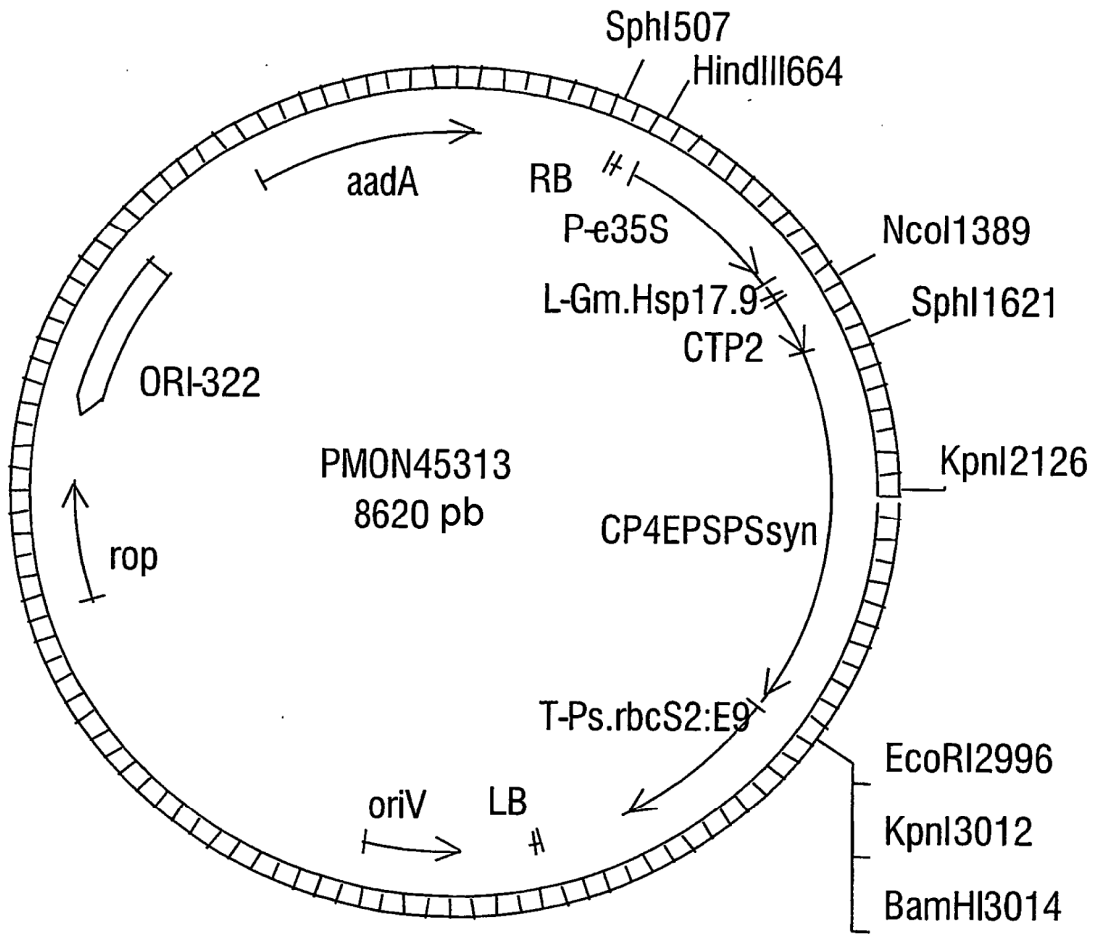


Figura 20

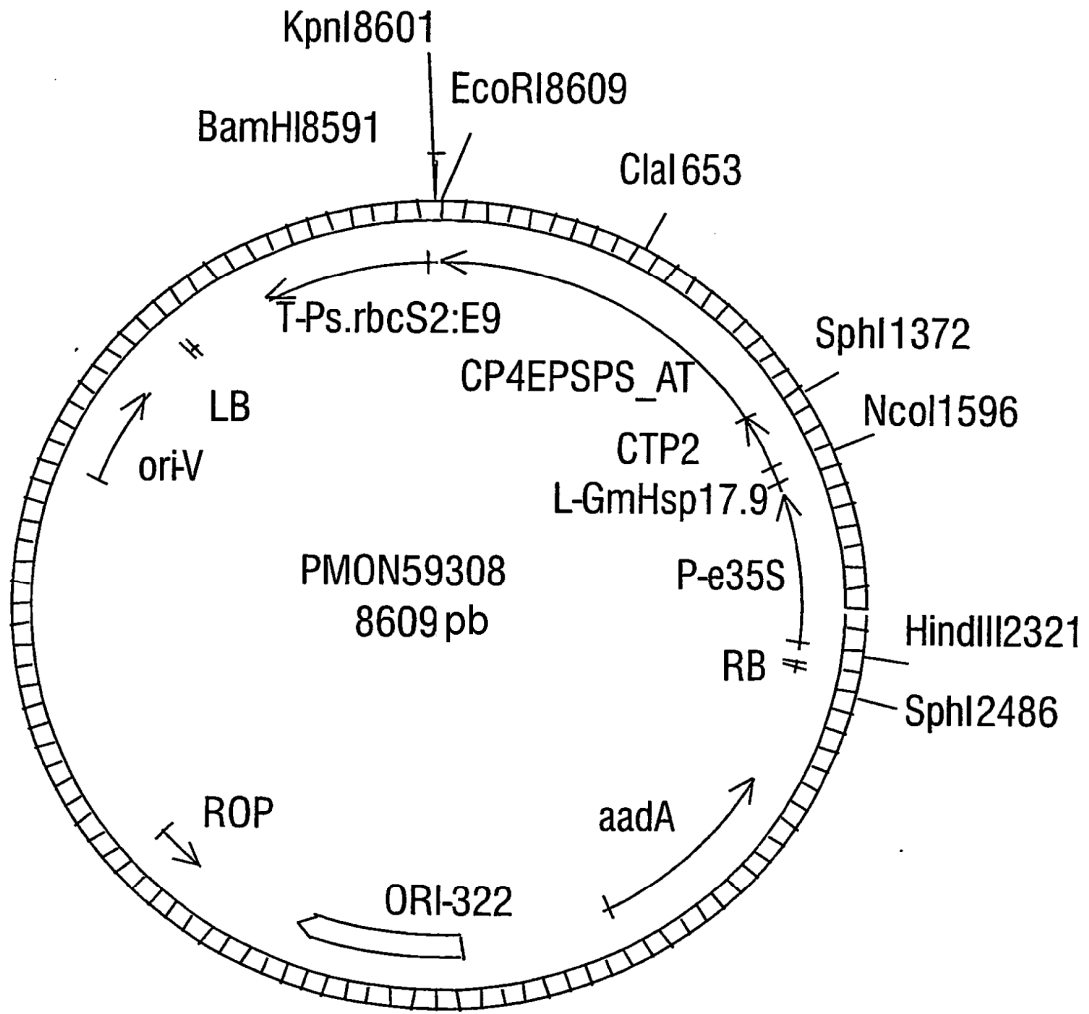


Figura 21

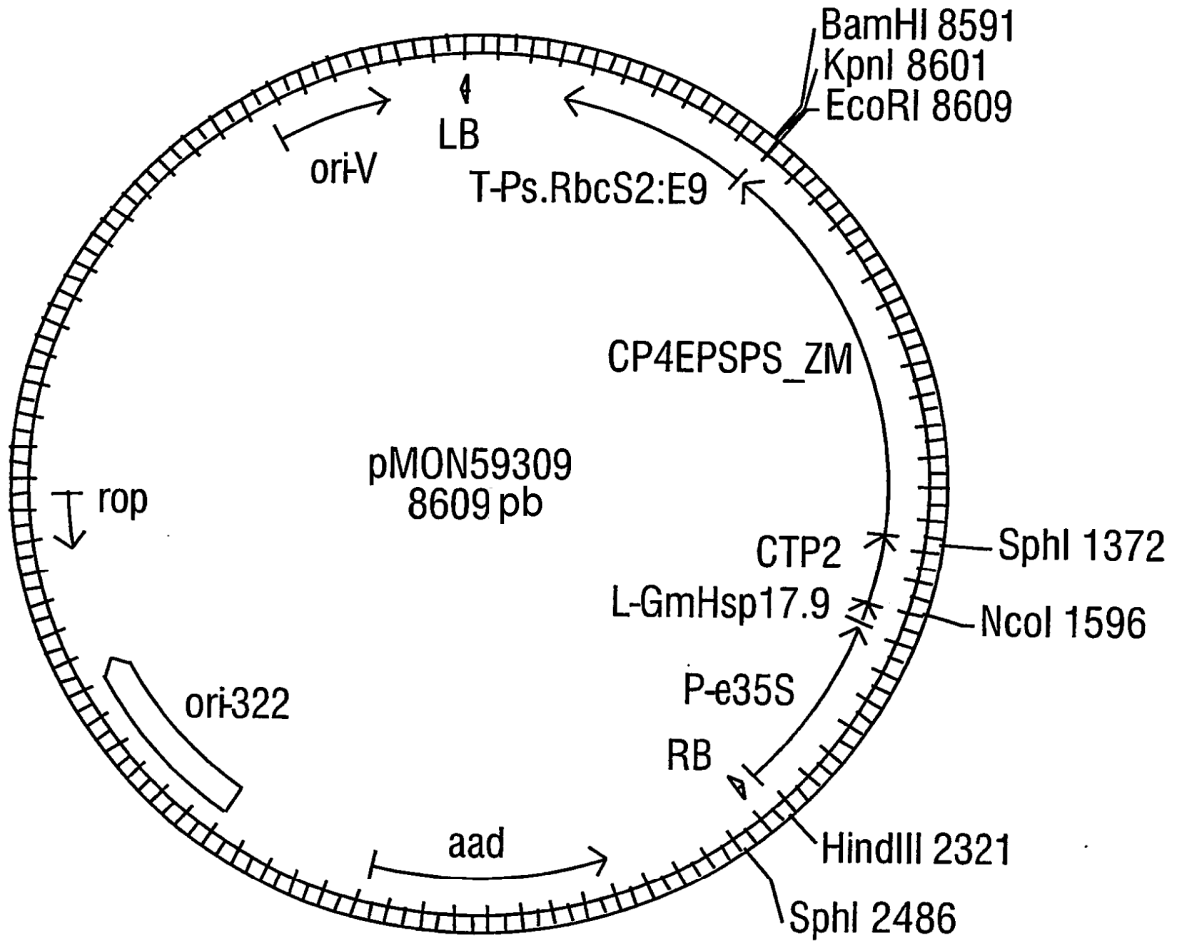


Figura 22

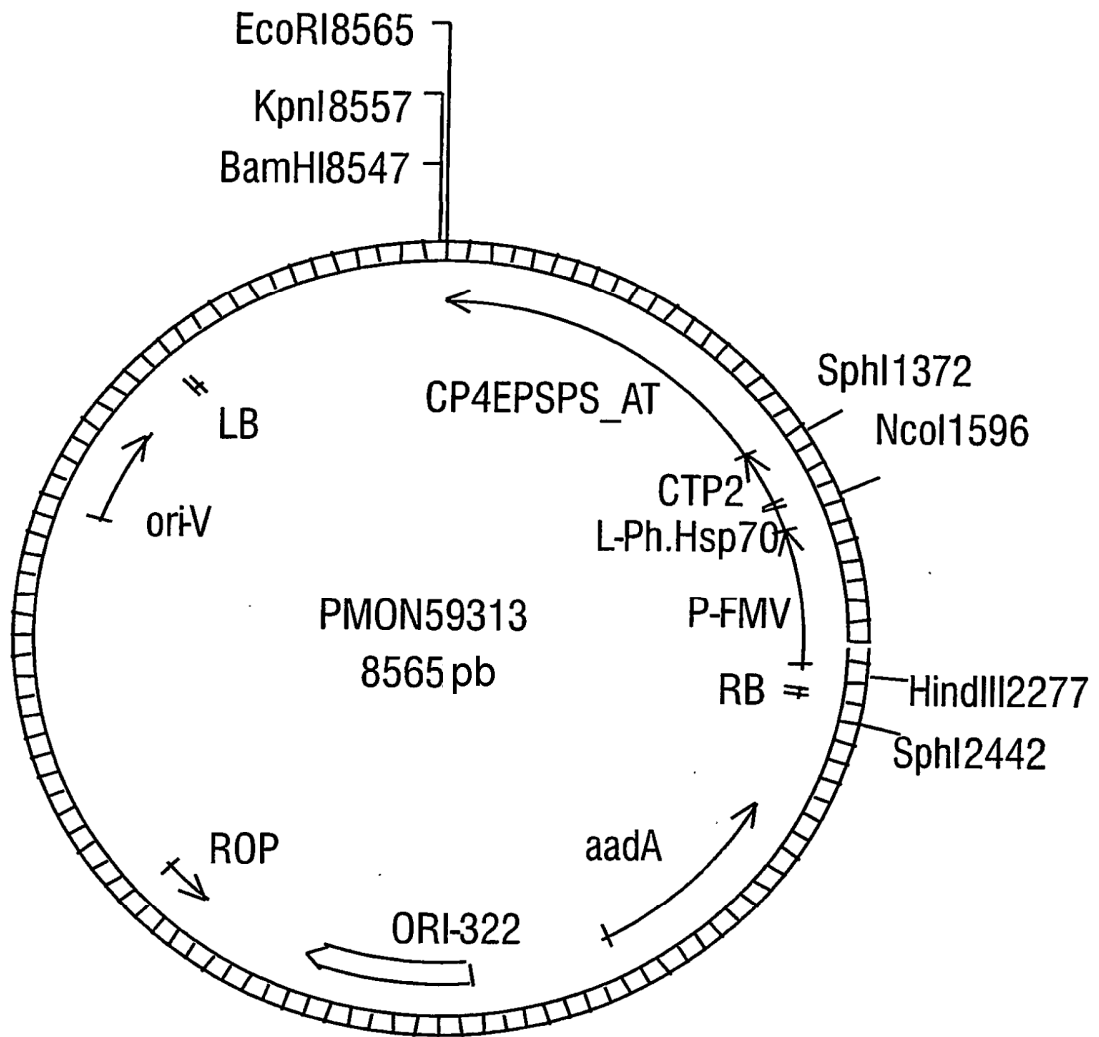


Figura 23

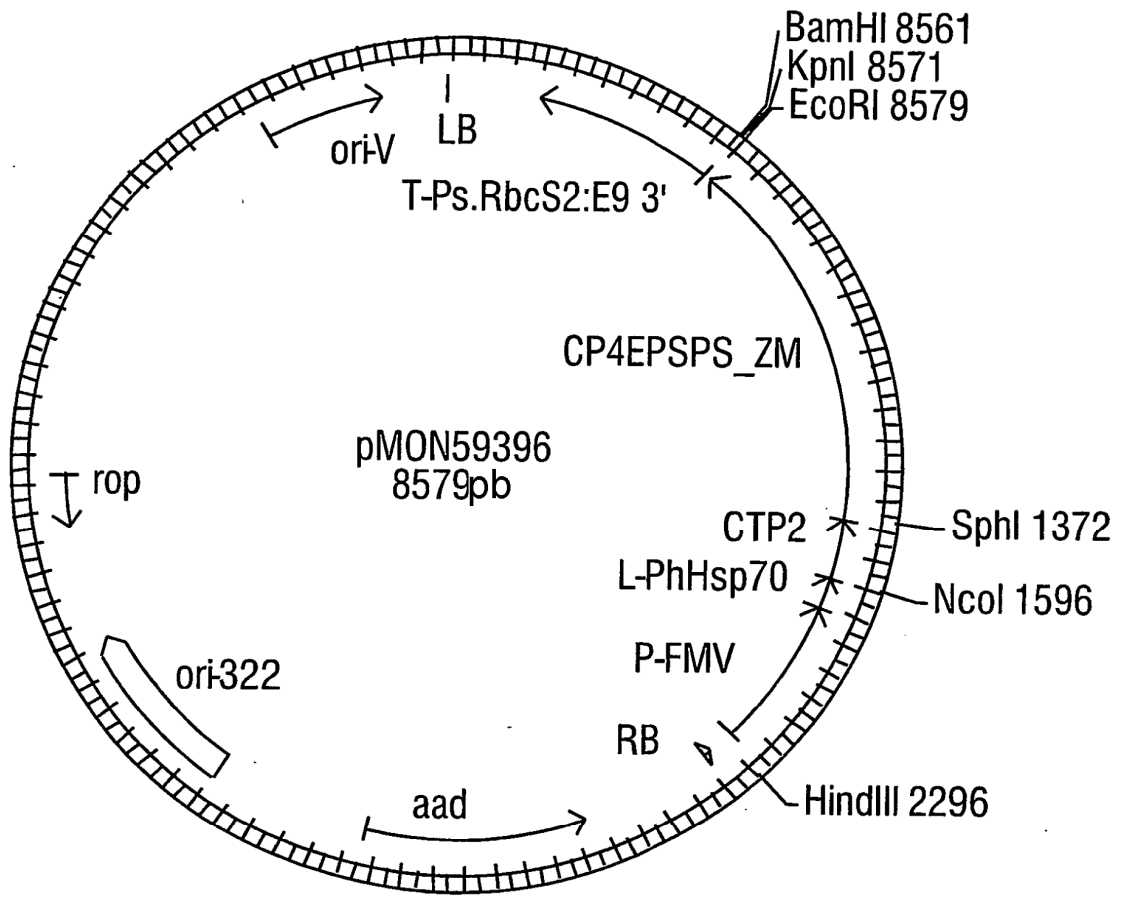


Figura 24

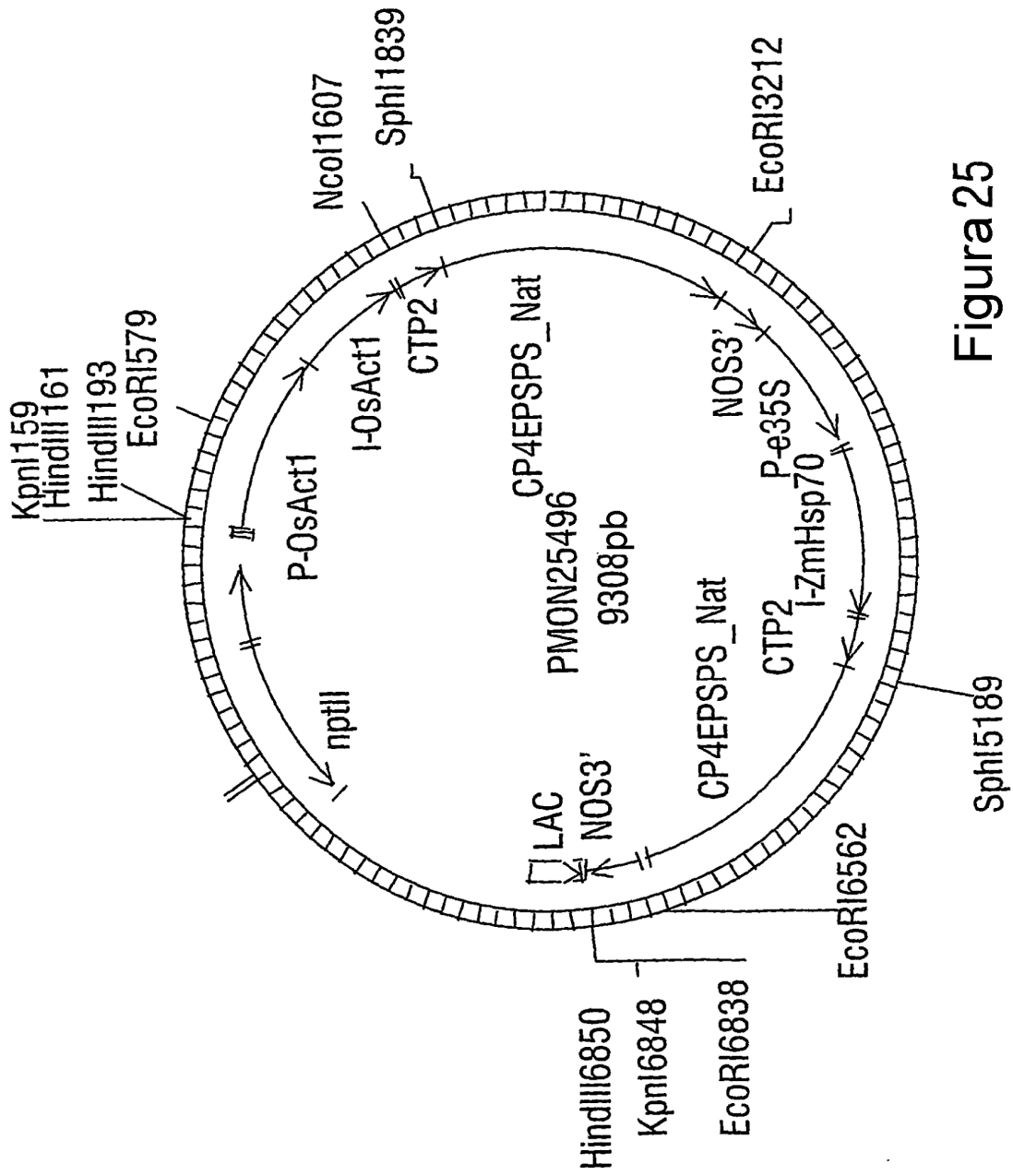


Figura 25