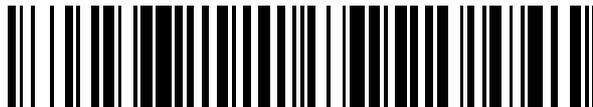


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 429 360**

51 Int. Cl.:

C12N 15/82 (2006.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.12.2001 E 01986901 (5)**

97 Fecha y número de publicación de la concesión europea: **24.07.2013 EP 1373531**

54 Título: **Método de codificar información en ácidos nucleicos de un organismo tratado por ingeniería genética**

30 Prioridad:

29.03.2001 DE 10115507

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

14.11.2013

73 Titular/es:

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

72 Inventor/es:

**GLEBA, YURI y
KLIMYUK, VICTOR**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 429 360 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

Método de codificar información en ácidos nucleicos de un organismo tratado por ingeniería genética

CAMPO DEL INVENTO

- 5 El presente invento se refiere a un método de producir un organismo vegetal tratado por ingeniería genética, que contiene un mensaje de información codificada que se refiere a una secuencia transgénica en dicho organismo tratado por ingeniería genética.

ANTECEDENTES DEL INVENTO

- 10 Un tratamiento por ingeniería genética es un proceso técnico y como tal está sometido a reglas y regulaciones específicas que se diseñan de manera tal que se aseguren unos apropiados entornos técnicos, ecológicos y económicos para el proceso propiamente dicho y para los productos resultantes. Dichas reglas están bien establecidas y consagradas en otros sectores de la tecnología y, entre otras finalidades, están destinadas a asegurar una calidad y una reproducibilidad altas del proceso y de los productos resultantes, y a proteger al público en general, a los consumidores, a los productores así como al medio ambiente. Los organismos modificados por ingeniería genética (GM = acrónimo de Genetically modified) (GMO's) son unos productos especiales creados por los seres humanos en el hecho de que ellos son autorreplicantes. Así, cualquier material transgénico liberado dentro del entorno tiene un potencial de persistir durante un periodo de tiempo muy largo. La comunidad científica es acusada con frecuencia de ser fascinada excesivamente por un descubrimiento científico y de perjudicar temerariamente a la salud pública y al medio ambiente, por abogar para una temprana puesta en circulación de organismos modificados genéticamente, tales como plantas GM. Desde 1994, se han cultivado alrededor de 3,5 trillones de plantas transgénicas en los EE.UU, y no se encontró ninguna evidencia de efectos negativos. Por lo demás, unas "buenas prácticas de tecnología" concernientes al proceso de manipulación genética y a los organismos transgénicos como productos están muy lejos de ser maduras, y deberían realizarse esfuerzos adicionales. Unos de tales elementos de procesos biotecnológicos más avanzados es el de proporcionar unos apropiados datos de información técnica, incluyendo una marcación y un registro de los productos.

- Se han abogado varios grupos para marcar productos alimenticios que contienen organismos GM (GMO's). La reciente decisión de la Unión Europea para frenar la moratoria sobre plantas transgénicas es un movimiento positivo en la dirección correcta (Schiermeier, Q. 2001, *Nature*, 409, 967-968). Sin embargo, los Estados que de hecho están detrás de la moratoria solicitaron adicionales reglas sobre la rastreabilidad y la marcación de productos GM.

- La marcación de unos objetos que incluyen materiales líquidos y sólidos, objetos valiosos, etc., con fragmentos de ácidos nucleicos ha sido divulgada en un cierto número de patentes y solicitudes de patentes (véanse los documentos de solicitud de patente internacional WO9014441; de patente europea EP408424; WO9117265; WO9404918; WO9416902). Dos solicitudes de patente (véanse los documentos WO9617954; WO0059917) reivindicán, además de objetos no vivientes, el uso de ADN y biopolímeros para marcar objetos vivientes. Sin embargo, estas publicaciones no proporcionan la enseñanza de una marcación segura con respecto a organismos transgénicos, en particular las plantas y los animales transgénicos/os se trataron por ingeniería genética de manera tal que se separase la parte funcional y la parte técnica (informativa) del inserto de ADN en un cromosoma de un anfitrión. El método de marcación de la técnica anterior no afronta ni el problema de perder la marcación durante reproducciones múltiples del GMO ni el problema de una alta frecuencia de pérdida o de corrupción del contenido informativo de la marcación.

El documento de solicitud de patente europea EP 1 045 037 A1 se refiere a un sistema de identificación de células de plantas.

- Por lo tanto, un problema de este invento es proporcionar unos métodos que permitan una marcación genética no ambigua y segura de los GMO's.

Otro problema del invento es proporcionar métodos que permitan el rastreo de organismos transgénicos y GM liberados en el medio ambiente y de los productos derivados de los mismos.

SUMARIO DEL INVENTO

- 50 Estos objetivos se consiguen mediante un método de producir un organismo vegetal tratado por ingeniería genética, de acuerdo con la reivindicación 1.

El cartografiado desde una secuencia de ADN hasta un mensaje de información es singular mientras que el cartografiado desde un mensaje de información hasta una secuencia de ADN no es singular, de manera tal que se proporcionan una versatilidad y un ajuste de la secuencia de ADN.

Además, dicho esquema de codificación es de manera preferible redundante y/o hace posible la detección y la corrección de errores con el fin de proporcionar una estabilidad de la información codificada durante un período de tiempo relevante para la práctica.

5 Más aun, dicha secuencia de ADN funcional y dicha secuencia de ADN no funcional permanecen de manera preferible engarzadas durante la reproducción del organismo durante un período de tiempo relevante para la práctica.

El invento describe además un método que permite una marcación genética y un rastreo del transgén que interesa añadiendo cerca de él y engarzada a él una secuencia de ADN especializada, cuya secuencia desarrolla solamente una función de marcación y no tiene ninguna otra funcionalidad.

10 La memoria descriptiva describe además un método de marcación genética que no afecta desfavorablemente a la idoneidad del organismo transgénico y no afecta desfavorablemente al rendimiento del organismo transgénico que interesa. Además, la marcación del invento (una secuencia de ADN no funcional) es genéticamente estable y está engarzada establemente con el transgén que interesa, y no establece ningún riesgo ecológico medible.

15 Esta memoria descriptiva describe además un organismo transgénico obtenido u obtenible de acuerdo con el método.

Según el mejor saber y entender de los autores de este invento, los organismos vegetales transgénicos obtenidos aquí y descritos en el invento son los primeros organismos modificados genéticamente que son marcados genéticamente por incorporación dentro de un organismo de una secuencia de ADN que no tiene ninguna otra función en el organismo distinta de la de proporcionar una información técnica acerca del material transgénico.

20 Además, se describen unos vectores para incorporar un ADN en un organismo de acuerdo con este método.

Se describe un método para analizar un organismo tratado por ingeniería genética producido de acuerdo con el método del invento, comprendiendo dicho método por lo menos una de las siguientes etapas:

- (a) emplear el método de PCR (reacción en cadena de la polimerasa) para amplificar la secuencia de ADN no funcional,
- 25 (b) usar una sonda que tiene una secuencia complementaria con la secuencia de ADN no funcional,
- (c) usar un anticuerpo para la detección inmunológica de un polipéptido expresado a partir de la secuencia de ADN no funcional,
- (d) secuenciar la secuencia de ADN no funcional, y
- (e) leer un mensaje con información usando dicho esquema de codificación previamente definido.

30 De acuerdo con el método del invento, se introducen por lo menos dos secuencias en un organismo vegetal, una secuencia de ADN funcional y otra que no es funcional.

La secuencia de ADN funcional contiene un gen o fragmento de gen que interesa, p.ej. para conferir al organismo un rasgo útil. La funcionalidad de la secuencia funcional corresponde esencialmente a la razón para tratar a dicho organismo por ingeniería genética.

35 La secuencia de ADN no funcional no es requerida para la función del organismo ni para la función de la secuencia de ADN funcional, aunque ella se puede solapar parcialmente con la secuencia funcional. El mensaje de información está relacionado con la secuencia de ADN funcional por el hecho de que contiene una información concerniente a la secuencia funcional, cuya información indica la presencia de la secuencia funcional. Una información concerniente a la secuencia funcional puede incluir una fecha, un sitio, un productor, el nombre de una
40 compañía, un número de registro o de serie, una referencia a una base de datos o una anotación en una base de datos que contiene más información concerniente a la secuencia funcional, una marca registrada, etc.. Una finalidad importante de la secuencia no funcional es permitir que un GMO encontrado en el medio ambiente o en el sitio comercial sea rastreado de retorno al productor, la fecha y el sitio de construcción o liberación, etc.

45 El esquema de codificación previamente definido proporciona un cartografiado desde una pluralidad de posibles mensajes de información hasta una pluralidad de secuencias de ADN. Cada posible mensaje de información puede ser cartografiado para dar una secuencia de ADN por la aplicación del esquema de codificación previamente determinado. El esquema de codificación puede permitir un cartografiado singular o no singular. El cartografiado desde una secuencia de ADN hasta un mensaje de información es singular y el cartografiado desde un mensaje de información hasta una secuencia de ADN no es singular. En el sentido más amplio, este cartografiado puede ser
50 cualquier relación binaria entre el conjunto o la pluralidad de mensajes de información permisibles y el conjunto o la pluralidad de secuencias de ADN o (a) uno o varios subconjunto(s) de las mismas.

Hay muchos esquemas de codificación posibles que son concebibles para la finalidad de este invento. Una posibilidad particularmente oportuna sería que el esquema de codificación sea definido por un libro de códigos que asocie cada posible mensaje de información (p.ej. el nombre de una compañía) con una o más correspondientes secuencias de ADN. El libro de códigos puede tener o no cualquier estructura inherente. En otras formas de realización del invento, se usa un esquema de codificación previamente definido (que especifica cómo un gran número de posibles mensajes de información han de ser cartografiados para dar unas correspondientes secuencias de ADN) en lugar de un libro de códigos previamente definido, que proporciona el cartografiado previamente definido por separado para cada mensaje de información individual. En cualquier caso, el número de posibles mensajes de información que se pueden codificar es de manera preferible bastante grande y, en algunas formas de realización, puede ser mayor que 10 o mayor que 1.000 o mayor que 100.000.

El mensaje de información se compone de una secuencia de caracteres alfanuméricos. Estos pueden ser asignados a unas bases o a unos multipletes de bases. De manera preferible se usan unos multipletes de bases tales como los dobletes, tripletes, cuartetos o quintetos. Son sumamente preferidos los tripletes de bases. Ventajosamente, el multiplete de bases tiene una capacidad de codificación más alta que el número de signos o caracteres que se han de codificar. La redundancia del esquema de codificación obtenido de esta manera proporciona una más alta estabilidad de la información codificada, p.ej. cuando se analiza un organismo tratado por ingeniería genética o durante la reproducción de un organismo, una versatilidad más amplia y un ajuste de la secuencia de nucleótidos a otras circunstancias, tales como las relacionadas con el tratamiento por ingeniería genética del organismo (p.ej. sitios de restricción). Además de la redundancia del esquema de codificación, existen otros medios que son capaces de mejorar la impermeabilidad a los errores del método del invento, proporcionando una detección y/o corrección hasta de un número dado de errores. Éstos incluyen p.ej. el uso de unos códigos de bloques tales como códigos binarios y códigos cíclicos, códigos intercalados, códigos de Hamming cíclicos y códigos convolucionales (véase p.ej. la obra Informatik-Handbuch (manual de informática), coordinadores de edición: Pechenberger y Pomberger, 2ª edición, Múnich, Viena, editorial Hanser, 1999). La información codificada debería ser estable durante un periodo de tiempo que sea relevante para la práctica.

Si se usan multipletes de bases para codificar unos signos del mensaje de información, existen múltiples marcos de lectura. En este caso, se puede usar más de un marco de lectura (ORF) para codificar el mensaje de información.

Con el fin de hacer que el método del invento sea tan ampliamente aplicable como sea posible, el esquema de codificación deberá estar convenido generalmente y de esta manera ser previamente definido. De manera preferible el convenio del esquema de codificación es de amplitud nacional. De manera más preferible, es supranacional. La incorporación de una secuencia de ADN no funcional de acuerdo con el invento puede convertirse entonces en un patrón internacional que ayuda a la seguridad de los GMO's.

Las secuencias de ADN, la funcional y la no funcional, del organismo tratado por ingeniería genética de acuerdo con el invento deberían estar engarzadas y permanecer de esta manera en el proceso de reproducción del organismo. Esto se puede conseguir por una proximidad física de dichas secuencias. Ellas están situadas en el mismo cromosoma. De manera preferible, ellas deberían estar suficientemente próximas con el fin de reducir al mínimo la probabilidad de una separación durante una reproducción debida a una recombinación de ADN tal como un cruce, durante un período de tiempo que sea relevante para la práctica. Un período de tiempo relevante para la práctica puede ser definido de manera aproximada como 200, de manera preferible 100 y por lo menos 50 generaciones del organismo. Puede ser útil en la práctica tener a la secuencia funcional y a la secuencia no funcional directamente en sucesión dentro de dicho organismo, opcionalmente separadas por una secuencia espaciadora. Tal como se ha señalado con anterioridad, estas secuencias pueden solaparse particularmente.

Una estrecha proximidad se consigue con la máxima facilidad incorporando simultáneamente la secuencia funcional y la no funcional dentro del organismo. Sin embargo, el invento comprende también el caso en el que la secuencia no funcional es incorporada dentro de un organismo que ya sido tratado por ingeniería genética, de manera preferible por subsiguiente incorporación dirigida hacia un objetivo de una secuencia funcional o viceversa.

La secuencia de ADN no funcional es provista de manera preferible de por lo menos una secuencia de reconocimiento previamente definida. La(s) secuencia(s) de reconocimiento puede(n) estar situada(s) dentro de la secuencia no funcional. De manera preferible, la secuencia no funcional está flanqueada en por lo menos un lado por una(s) secuencia(s) de reconocimiento previamente definida(s). Dicha(s) secuencia(s) de reconocimiento previamente definida(s) permite(n) la identificación de dicha secuencia de ADN no funcional. De manera sumamente preferible, la secuencia de ADN no funcional está flanqueada por ambos lados por unas secuencias de reconocimiento. Ella puede luego ser analizada mediante una amplificación por PCR usando unos cebadores complementarios con dichas secuencias de reconocimiento previamente definidas, seguidos por una secuenciación de la secuencia no funcional. Las secuencias de reconocimiento flanqueadoras se pueden usar también para determinar la iniciación y/o la terminación de la secuencia no funcional. Además, dicha(s) secuencia(s) de reconocimiento previamente definida(s) se puede(n) diseñar de manera tal que se acomode el reconocimiento conjunto y por separado de múltiples secuencias de ADN no funcionales en un organismo tratado por ingeniería

genética. Esto es importante cuando resultan corrientes unos organismos múltiplemente tratados o marcados por ingeniería genética.

5 La(s) secuencia(s) de reconocimiento o flanqueadora(s) es (son) de manera preferible definida(s) previamente, es decir es (son) convenidas generalmente de una manera similar a como más arriba se ha descrito para el esquema de codificación. La secuencia no funcional puede entonces ser analizada y el mensaje de información puede ser leído a partir de la secuencia de bases de la secuencia de ADN no funcional mediante la aplicación del esquema de codificación por cualquier persona con experiencia ordinaria en la biología molecular. Alternativamente, la presencia de una secuencia de ADN no funcional puede ser analizada p.ej. usando una sonda de ADN dirigida hacia una secuencia de reconocimiento o hacia (una porción de) la secuencia no funcional propiamente dicha.

10 Otros medios de analizar la secuencia de ADN no funcional incluyen el uso de un anticuerpo para un polipéptido expresado a partir de dicha secuencia de ADN no funcional. En este caso, la secuencia no funcional puede comprender una casete de expresión completa. Alternativamente, dicho polipéptido puede ser traducido por vía de un elemento de IRES (acrónimo de Internal Ribosome Entry Site = sitio de entrada de ribosoma interno) que es funcional en dicho organismo, a partir de un ARNm (mensajero) bicistrónico funcionalmente, que codifica tanto el transgén que interesa como también dicho polipéptido. Por estos medios, la secuencia no funcional puede ser detectada en los organismos o en los productos derivados de los mismos tales como un alimento (elaborado).

15 El proceso de analizar la secuencia no funcional puede ser la simple detección de la presencia de una secuencia no funcional de acuerdo con el invento en un organismo (p.ej. por uso de una sonda de ADN) o puede implicar la determinación o el reconocimiento de la secuencia de base de la secuencia no funcional o de una porción de la misma. El proceso de analizar puede implicar además la aplicación del esquema de codificación con el fin de descodificar la secuencia de ADN no funcional.

20 La secuencia de reconocimiento puede ser añadida a la secuencia de ADN no funcional después de una aplicación del esquema de codificación al mensaje de información. Alternativamente, el esquema de codificación puede ser diseñado de manera tal que una secuencia que tiene la función de la secuencia de reconocimiento sea añadida a, o incorporada directamente en, la secuencia no funcional después de una aplicación del esquema de codificación al mensaje de información. En esta forma de realización, la expresión "secuencia de ADN no funcional" comprende las funciones de la secuencia de ADN no funcional y la función de la secuencia de reconocimiento.

25 El método del invento puede ser aplicado a todos los GMO's vegetales. Sin embargo se prefiere aplicarlo a plantas superiores. Son sumamente preferidas las plantas útiles superiores.

30 También se describe un método de codificar un mensaje de información en una molécula de ADN, que comprende incorporar en dicha molécula de ADN una secuencia de ADN no funcional, en donde dicha secuencia de ADN no funcional corresponde al resultado de la aplicación de un esquema de codificación previamente definido al mensaje de información, proporcionando dicho esquema de codificación previamente definido un cartografiado desde una pluralidad de mensajes de información posibles para dar una pluralidad de secuencias de ADN.

35 Los principios más arriba bosquejados para dicho método de producir un organismo tratado por ingeniería genética se aplican también a este método de codificar un mensaje de información en una molécula de ADN, cuando sea pertinente.

40 Dicha molécula de ADN puede ser un plásmido o un ADN cromosomal. Ella puede estar aislada o contenida en un organismo. De manera preferible, ella puede ser mantenida dentro de un organismo. De manera sumamente preferible, dicho mensaje de información o dicha secuencia de ADN no funcional puede ser una marca registrada.

También se describe un método de marcar una molécula de ADN con una marca registrada, que comprende
 (a) seleccionar una secuencia de nucleótidos capaz de funcionar con una marca registrada y
 (b) incorporar dicha secuencia de nucleótidos en dicha molécula de ADN de manera tal que ella pueda ser detectada y reconocida como una referencia al origen de la molécula de ADN.

45 Los vectores para incorporar un ADN dentro de un organismo de acuerdo con el método de producir un organismo tratado por ingeniería genética pueden ser diseñados de acuerdo con los principios conocidos de biología molecular. Estos vectores deberían ser capaces de introducir de una manera estable una secuencia de ADN que interesa dentro del genoma de un organismo. Dichos vectores pueden ser p.ej. de origen vírico. Se puede usar por ejemplo una incorporación en la planta del ADN que interesa, mediada por *Agrobacterium tumefaciens*. Son conocidos varios otros métodos de transformación en la técnica tales como métodos balísticos o de cañones de partículas.

Breve descripción de los dibujos

La Fig. 1 representa un vector para codificar y poner a punto una información no biológica.

5 La zona rellena del vector corresponde a las regiones conservadas, diseñadas para amplificar la región de codificación (no rellena) con unos cebadores A y B. Los cebadores C y D están diseñados para poner al día o reemplazar la información dentro de la región de codificación.

La Fig. 2 representa el vector pIC4100 que codifica una información no biológica para la transformación de plantas.

La Fig. 3 muestra la secuencia procedente del vector pIC4100 que codifica una información no biológica con presentación esquemática de los ORFs (acrónimo de Open Read Frames = marcos de lectura abiertos) de iniciación y terminación (programa DNASTAR).

10 La Fig. 4 representa un vector para ensayar la frecuencia de recombinación dentro de una información técnica de codificación por ADN.

SM – marcador seleccionable;

CSM – marcador de selección contra-seleccionable o negativo;

MCS – múltiples sitios de clonación;

15 A y B –cebadores de PCR para amplificar un ADN “técnico”.

La Fig. 5 representa el vector binario pICBV10.

Descripción detallada del invento

20 Aquí, nosotros proponemos marcar unos organismos vegetales GM propiamente dichos, incluyendo un mensaje de información (técnica) codificado por un ADN en un genoma de GMO's, de manera preferible cerca de una secuencia de ADN funcional tal como un transgén. Dicho mensaje de información técnica puede estar basado en un código no genético o en un esquema de codificación, y puede ser recuperado y leído con facilidad. Él puede almacenar diversas cantidades de información. En una forma preferida de realización, dicha marcación podría ser adaptada como un patrón de codificación universal de GMO's, definiendo de esta manera aún más el entorno legal en un tratamiento por ingeniería genética y protegiendo de esta manera al público, al entorno y así como a los productores.

Esquema de codificación / lenguajes

La información (técnica) propuesta del mensaje de información será codificada en forma de un ADN, por lo tanto, estamos limitados al lenguaje de almacenamiento de ADN de cuatro letras. Para hacer que el lenguaje del mensaje de información resulte lo más propicio para el usuario que sea posible, se puede escoger la lengua inglesa. En este caso, un esquema de codificación que traduce las cuatro letras de un ADN (A, T, C, G) en un lenguaje que tiene 26 letras del alfabeto latino, 10 números arábigos y uno o más caracteres de espaciado / terminación, es decir un total de 37-38 caracteres (alfanuméricos). Es evidente que nosotros estamos siendo servidos óptimamente por un esquema de codificación que traduce tripletes de nucleótidos, de manera muy similar a un código genético de Nature. Dicho código tiene la capacidad de codificar $4^3 = 64$ caracteres, un número que es más que suficiente para las finalidades del invento. El código propuesto será por lo tanto también redundante en cierta medida, lo que reduce al mínimo los errores de lectura y proporciona una estabilidad más alta durante la reproducción de un organismo. Además, esto permite evitar la creación de unas secuencias no funcionales que contengan características indeseadas tales como unos sitios de restricción, que complican al proceso de tratamiento por ingeniería genética, o unas secuencias repetitivas que pueden afectar a la estabilidad del inserto de ADN. Si el lenguaje está basado en palabras de la lengua inglesa, el esquema de codificación se vuelve incluso más redundante, puesto que unos mensajes corrompidos con símbolos alfanuméricos cambiados pueden ser todavía reconocidos debido a la redundancia inherente de los lenguajes naturales. Un ejemplo de un esquema de codificación que toma en consideración lo anterior está dado en la Tabla 1. Otro ejemplo está dado en la Tabla 2, en donde unos tripletes que codifican el mismo carácter alfanumérico difieren en la totalidad de las tres bases, dando una versatilidad adicional y unos medios para ajustar la secuencia de ADN no funcional a las circunstancias en cuestión.

De manera sumamente preferible, el esquema de codificación es degenerado con el fin de permitir la codificación de un mensaje de información deseado o de un carácter alfanumérico por más de una secuencia de nucleótidos o por un multiplete de bases, respectivamente. La degeneración del esquema de codificación puede ser sustancialmente aumentada en comparación con el uso de tripletes de bases, usando cuadrupletes o quintetes de bases. Se prefiere el uso de cuadrupletes de bases.

Aunque nosotros consideramos que un esquema de codificación altamente estructurado, tal como el esquema que convierte codones en forma de tripletes en valores alfanuméricos, es un esquema de codificación preferido, son también considerados en este invento otros esquemas de codificación que se basan en el uso de unas secuencias de nucleótidos idiosincráticas, asignadas a unos usuarios o unas claves de datos específicos/as. Notablemente, el esquema de codificación del invento no está de ninguna manera restringido al Código Genético ni a esquemas de

codificación análogos. El número de esquemas de codificación, que se pueden usar de acuerdo con los principios del invento, es virtualmente ilimitado e incluye también el uso de dos o más diferentes esquemas de codificación, que pueden ser combinados.

El mensaje de información

5 Para definir un contenido y una longitud útiles de un mensaje, se pueden establecer analogías con otros productos existentes en el mercado. En el respaldo de una máquina, por ejemplo de un ordenador, una persona encuentra usualmente datos tales como el nombre del fabricante o de la compañía, la fecha de producción, el sitio de producción, el modelo de producto y el número de serie. Suponiendo que la situación en un tratamiento por ingeniería genética es generalmente similar (posiblemente, nosotros deseáramos añadir un nombre de un propietario de un producto registrado, que puede ser o no el productor, y una sección de mensaje adicional que podría ser reservada para un próximo futuro), nosotros sacamos la conclusión de que una suficiente longitud del mensaje podría ser para el próximo futuro del orden de 3-10 palabras, teniendo cada una de éstas una suficiente capacidad de codificación de manera tal que se refleje la complejidad del mundo industrial. Sin embargo, son posibles asimismo unos mensajes sustancialmente más largos. La información codificada por un ADN podría proporcionar de esta manera suficiente información general disponible para cualquier persona, así como unas referencias a una información almacenada en otros medios, tales como una base de datos completa del productor, o una base de datos especial de una agencia gubernamental. Suponiendo unas longitudes medias de las palabras usualmente usadas en marcaciones de productos técnicos, llegamos a una longitud total de un mensaje de aproximadamente más o menos 100 caracteres alfanuméricos que corresponden a 300 nucleótidos si se usan unos tripletes de bases. Sin embargo, se puede codificar también unos mensajes de información mucho más largos.

Por lo tanto, la secuencia de ADN no funcional puede contener una información relevante para el organismo tratado por ingeniería genética, que ha sido producido por el método del invento. Dicha información relevante para un producto puede comprender una información relacionada con el ADN funcional seleccionado a partir del siguiente grupo o cualquier otra información: una marca registrada, una referencia a una base de datos, una fecha, un lugar, el nombre del productor o propietario, así como una información acerca de la presencia de la secuencia de ADN funcional.

Iniciación / terminación de la lectura

Para leer de una manera técnicamente simple y confiable un mensaje de información codificado por un ADN, se necesitan de manera preferible unas señales de iniciación y terminación convenientes y universales. Tomando en cuenta el estado actual de la especialidad de manipulación de ADN, se está óptimamente servido por conjuntos singulares de cortas secuencias de reconocimiento previamente definidas (con una longitud del orden de 18-30 nucleótidos) que permiten una fácil amplificación del ADN de mensaje en cuestión (secuencia de ADN no funcional) y que son convenientes de manera preferible de un modo amplio y declarados como patrones. Unos cebadores de amplificación podrían permitir también que una persona determine el comienzo y el final de la región del ADN que codifica el mensaje de información. Dichas secuencias de reconocimiento previamente definidas y los respectivos cebadores tienen de manera preferible una longitud suficiente como para ser singulares en el organismo tratado por ingeniería genética.

La secuencia de ADN no funcional, que tiene dicha(s) secuencia(s) de reconocimiento previamente definida(s), se compone de manera preferible de un tramo continuo de ADN. Alternativamente, la secuencia de ADN no funcional se puede componer de dos o más tramos de ADN, es decir que ella puede ser interrumpida por otras secuencias de ADN. Dicha secuencia de ADN no funcional puede ser interrumpida por la secuencia de ADN funcional o por una parte de la misma. Por lo tanto, la secuencia de ADN funcional puede estar flanqueada en ambos lados por unos tramos o unas porciones de la secuencia de ADN no funcional. Además, las secuencias de reconocimiento previamente definidas de la secuencia de ADN no funcional pueden flanquear a la secuencia de ADN funcional de manera tal que la secuencia de ADN funcional pueda ser amplificada por una PCR y analizada por secuenciación, opcionalmente en común con la secuencia de ADN no funcional. Dicha secuenciación de la secuencia de ADN funcional puede proporcionar una información adicional acerca de la secuencia de ADN funcional. Esta forma de realización puede ser combinada ventajosamente con la forma de realización, en la que la secuencia de ADN no funcional tiene dos, tres o más segmentos que (i) codifican la misma información y (ii) tienen unas secuencias de nucleótidos, que de manera preferible son lo suficientemente diferentes de modo que no causen una recombinación (véase más abajo). La secuencia de ADN funcional puede ser entonces flanqueada por dichos segmentos de dicha secuencia de ADN no funcional.

Múltiples mensajes

Los organismos pueden ser modificados genéticamente más de una vez, lo cual puede dar como resultado más de una marcación o una secuencia de ADN no funcional. El problema de una marcación múltiple es resuelto con facilidad creando diferentes conjuntos de secuencias de iniciación de mensajes, una que es degenerada y detectará todas las posibles iniciaciones, y una serie de otras que son menos degeneradas o singulares. Cualquier descodificación mediante una PCR podría comenzar entonces con secuencias degeneradas y luego, cuando los organismos múltiplemente modificados y marcados se convierten en una cuestión corriente, podrían incluir múltiples cebadores menos degenerados. Las regiones conservadas de los mensajes deberán ser suficientemente largas

para servir como sitios de cebadura para un análisis por PCR, p.ej. deberán tener por lo menos 18 a 20 pares de bases. Sin embargo, unas regiones conservadas innecesariamente largas pueden causar unos indeseados sucesos de recombinación en el caso de que múltiples mensajes estén situados en el mismo lugar.

Estabilidad genética del mensaje

5 Se desea que la secuencia de ADN no funcional para codificar el mensaje de información (información técnica) permanezca estable y engarzada a la secuencia de ADN funcional (p.ej. un transgén que interesa) a lo largo de muchas generaciones, y que no haya ningún deterioro de la calidad de la información a lo largo de un período de tiempo relevante en la práctica. De manera preferible, la estabilidad genética de la secuencia de ADN no funcional es tal que el mensaje de información codificado persiste por lo menos a lo largo de tantas reproducciones del organismo tratado por ingeniería genética como la función de la secuencia de ADN funcional. Para cumplir estos requisitos, la secuencia de nucleótidos de la secuencia de ADN no funcional, opcionalmente en unión con una secuencia de ADN espaciadora que se puede usar entre la secuencia de ADN no funcional y la secuencia de ADN funcional, se escoge de manera preferible de tal manera que no se exhiba esencialmente ningún perfil propenso a una recombinación, notablemente en comparación con el genoma de una especie de anfitrión, y/o que no se exhiba esencialmente ninguna homología perjudicial con el genoma del anfitrión.

Más específicamente, la secuencia de ADN no funcional es de manera preferible tratada por ingeniería genética de manera tal que ella no sea capaz de formar estructuras secundarias y de manera tal que ella no contenga secuencias conocidas como “puntos calientes” de recombinación o mutación, p.ej. estructuras de “horquilla de pelo” o palíndromos o repeticiones directas/invertidas. Similarmente, una secuencia espaciadora de ADN, que se puede usar entre las secuencias de ADN, la no funcional y la funcional, ha de ser tratada por ingeniería genética usando las mismas precauciones. Con este fin, se pueden usar búsquedas de diferentes bases de datos de secuencias de nucleótidos (por ejemplo una Búsqueda Blast) para identificar cualesquiera putativos “puntos calientes de recombinación o mutación” o una homología con genes residenciales o con un transgén del organismo anfitrión en la secuencia de ADN no funcional y/o en la secuencia de ADN espaciadora. Una lista de diferentes herramientas de búsqueda y análisis para la biología molecular está disponible de http://www.biophys.uni-duesseldorf.de/BioNet/Pedro/research_tools.html. Un ejemplo de tal búsqueda se presenta en el EJEMPLO 3. Para reducir aún más la frecuencia de una recombinación entre las secuencias de ADN, la no funcional y la funcional, la distancia física entre las dos ha de ser mantenida corta, de manera preferible tan corta como sea técnicamente posible. La distancia entre las secuencias de ADN, la funcional y la no funcional, es mantenida más corta que 10.000 nucleótidos, de manera preferible más corta que 1.000 nucleótidos.

Se describe además un método de medir experimentalmente la frecuencia de recombinación y la mutabilidad de la secuencia de ADN no funcional (véase el EJEMPLO 4). De acuerdo con este método, la secuencia de nucleótidos de la secuencia de ADN no funcional, opcionalmente en unión con una secuencia de ADN espaciadora que puede ser introducida entre la secuencia de ADN no funcional y la secuencia de ADN funcional, es definida como un miembro de un conjunto de secuencias de ensayo identificables con el siguiente ensayo:

- (a) crear unas moléculas de ADN, cada una de las cuales comprenda una secuencia de ensayo entre una secuencia marcadora seleccionable y una secuencia marcadora contra-seleccionable;
- (b) cultivar unas células transformadas con dichas moléculas de ADN en la presencia de un primer agente que corresponde a la secuencia marcadora seleccionable y de un segundo agente que corresponde a la secuencia marcadora contra-seleccionable, en unas condiciones en las que ambos agentes son efectivos;
- (c) aislar las células que sobreviven en la etapa (b);
- (d) seleccionar una secuencia de ensayo entre las células aisladas en la etapa (c) que no muestran una mutación o una frecuencia de recombinación significativamente aumentada en comparación con otras regiones de los cromosomas transformados de dichas células.

45 De acuerdo con este método, se crean dos o más secuencias de ADN no funcionales que difieren en sus secuencias de nucleótidos para un deseado mensaje de información, haciendo uso de la degeneración del esquema de codificación; de esta manera se crea un conjunto de secuencias de ensayo. Cada miembro del conjunto de secuencias de ensayo es sometido entonces al anterior ensayo. Para dicho método de producir un organismo tratado por ingeniería genética, un miembro de dicho conjunto de secuencias de ensayo se puede seleccionar como una secuencia de ADN no funcional, que exhibe una frecuencia de recombinación y/o una mutabilidad deseada o baja. Dicho ensayo se puede usar también para identificar una secuencia de ADN espaciadora apropiada o una combinación apropiada de una secuencia de ADN no funcional y de una secuencia de ADN espaciadora.

Dicho ensayo es diseñado de manera preferible de manera tal que la situación en el organismo que ha de ser modificado genéticamente de acuerdo con el método del invento se simule lo mejor que sea posible. Por otro lado, se evitan de manera preferible los largos tiempos de generación de muchos organismos multicelulares, p.ej. un año en el caso de muchas especies de plantas útiles. Por lo tanto, dicho ensayo se realiza de manera preferible con células en un cultivo de células en vez de con organismos multicelulares. Dicho ensayo se lleva a cabo de manera más preferible con células bacterianas, notablemente con células de *E. coli*, por motivos de simplicidad. Sin embargo, también se pueden escoger células de plantas monocelulares tales como *Chlamydomonas reinhardtii* notablemente si el organismo que ha de ser tratado por ingeniería genética es una planta.

Una posibilidad adicional de aumentar la protección del mensaje de información contra mutaciones de sentidos erróneos o aleatorias en la secuencia de ADN no funcional es aumentar la redundancia del mensaje de información, p.ej. empleando una secuencia de ADN no funcional que tenga dos, tres o más segmentos que (i) codifiquen la misma información y (ii) de manera preferible tengan unas secuencias de nucleótidos que sean lo suficientemente diferentes para que no se cause una recombinación. Dichos segmentos pueden ser unos intrones. Dichos segmentos pueden estar en la misma hebra de ADN o en diferentes hebras de ADN y se pueden leer en una diferente orientación. El uso de diferentes segmentos de secuencias de ADN no funcionales que tienen diferentes secuencias de nucleótidos y codifican el mismo mensaje de información, es permitido por la degeneración del esquema de codificación. Un suceso de mutación o recombinación que corrompa al mensaje de información en uno de dichos segmentos dejará una o más copias no corrompidas de dicho mensaje de información codificado por un segmento no mutado.

En el EJEMPLO 2, la degeneración del esquema de codificación de la Tabla 1 se emplea para duplicar el mensaje de información con el fin de aumentar el nivel de protección contra “sentidos erróneos” aleatorios, sin crear repeticiones de secuencias. Se describen unas “repeticiones de información” (regiones conservadas) que son codificadas por (segmentos de) repeticiones invertidas imperfectas con una identidad entre secuencias de un 68 %. Sin embargo, un 32 % de faltas de concordancia distribuidas uniformemente entre estas repeticiones hacen que sea próxima a cero la posibilidad de interacción entre estas repeticiones en condiciones fisiológicas. El uso de las repeticiones de información permite detectar la gran mayoría de las potenciales mutaciones de ADN: tanto sustituciones de nucleótidos como también deleciones internas.

Por uso del esquema de identificación dado como ejemplo en la Tabla 2, la identidad entre secuencias de repeticiones de información puede ser reducida aun más al mismo tiempo que se aumentan la libertad y la versatilidad globales para codificar el mensaje de información. Se puede obtener una degeneración particularmente alta del esquema de codificación si se usan unos cuadrupletes de bases para codificar los caracteres alfanuméricos.

Otra alternativa aquí considerada es el uso de diferentes códigos para generar “unas repeticiones de información”, evitando de esta manera en absoluto la creación de cualesquiera repeticiones de ácidos nucleicos. El uso de repeticiones de información se puede acomodar con facilidad, puesto que la longitud global del mensaje está situada plenamente dentro de unos límites manejables técnicamente.

Una posibilidad adicional de aumentar la impermeabilidad a los errores del mensaje de información es el uso de multipletes de, preferiblemente, bases idénticas en vez de una única base. Como un ejemplo se puede usar la secuencia de bases GGGCCCGGG en vez de la secuencia de bases GCG para codificar el carácter B de acuerdo con la Tabla 1. La duplicación, triplicación o cuadruplicación de una base son unas alternativas preferidas de esta forma de realización. Se prefiere la triplicación. Dicho esquema de codificación tiene la ventaja de que las mutaciones puntuales que pueden producirse en el curso de la reproducción del organismo tratado por ingeniería genética se pueden identificar y corregir con facilidad al descodificar el mensaje de información a partir de la secuencia de ADN no funcional. Además, el marco de lectura de la secuencia de ADN no funcional es fácil de identificar, incluso si se ha producido un desplazamiento del marco debido a la inserción o la deleción de mutaciones.

En otra forma aún más preferida de realización, la secuencia de ADN no funcional puede ser diseñada como un intrón o una inteína. Se prefieren los intrones. Dicho intrón o dicha inteína se puede situar de manera preferible dentro de la secuencia de ADN funcional relacionada. Éste o ésta se puede colocar dentro de una porción altamente conservada o funcionalmente crucial en grado sumo del mismo, lo cual es sumamente protector, notablemente para conservar el engarce entre la secuencia de ADN funcional y la secuencia de ADN no funcional. También es posible usar varios (p.ej. tres) intrones diferentes con diferentes secuencias de bases, pero que de manera preferible codifican el mismo mensaje de información como se permite por la degeneración de la codificación. Las secuencias de reconocimiento previamente definidas, que se pueden usar para una amplificación por PCR de la secuencia de ADN no funcional, pueden ser una parte de dichos intrones. Unos métodos de tratar por ingeniería genética los intrones del grupo I y del grupo II son conocidos en la especialidad. Se pueden encontrar detalles en los siguientes documentos y en las referencias allí citadas: Ayre y colaboradores (1999) Proc. Nat. Acad. Sci. USA 96, 3507-3512; Valadkhan y Manley (2001) Nature 413, 701-707; Hagen y Cech (1999) EMBO J. 18, 6491-6500; y documentos de patentes US6143503; US6010904; US6015794). En el caso de una inteína como secuencia de ADN no funcional, la secuencia de ADN no funcional expresada, o una porción de la misma, es un polipéptido y puede ser detectada con un anticuerpo. Un conjunto de anticuerpos específicos, cada uno de los cuales puede reconocer a un polipéptido previamente definido, se puede proporcionar p.ej. como un estuche de partes. Cada reacción detectada entre un polipéptido y un anticuerpo puede constituir un mensaje de información previamente acordado. Unos métodos de tratar por ingeniería genética a inteínas son conocidos en la especialidad. Se pueden encontrar detalles en los siguientes documentos y en las referencias allí citadas: Lew y colaboradores (1998) J. Biol. Chem. 273, 15887-15890; Liu (2000) Annu. Rev. Genet. 34, 61-76; Kenneth y colaboradores (1998) Proc. Natl. Acad. Sci. USA 95, 3543-3548).

En el procedimiento del invento, dicha incorporación dentro de dicho organismo de dicha secuencia de ADN no funcional no debería afectar desfavorablemente a la estabilidad genética de la secuencia de ADN no funcional y/o de la secuencia de ADN funcional y/o el engarce entre la secuencia de ADN no funcional y la secuencia de ADN funcional. Con este fin, se describe un ensayo experimental, en el que la secuencia de ADN no funcional es definida como un miembro de un conjunto de secuencias de ensayo identificables con el siguiente ensayo:

(a) crear unas moléculas de ADN que comprenden dicha secuencia de ADN funcional, estando dicha secuencia de ADN no funcional conectada con la secuencia de ADN funcional opcionalmente a través de una secuencia de ADN espaciadora;

(b) introducir dichas moléculas de ADN dentro de unas células para crear células transgénicas y organismos transgénicos que comprenden dicha secuencia de ADN funcional conectada con la secuencia de ADN no funcional opcionalmente a través de una secuencia de ADN espaciadora.

(c) analizar el rendimiento de dicha secuencia de ADN funcional conectada con la secuencia de ADN no funcional opcionalmente a través de una secuencia de ADN espaciadora, siendo definido dicho rendimiento por la segregación Mendeliana regular esperada de dicha secuencia entre la progenie sexual y por el mantenimiento de pleno engarce entre dicha secuencia de ADN funcional y dicha secuencia de ADN no funcional en el proceso de reproducción sexual de dichos organismos transgénicos, y

(d) identificar una secuencia de ADN no funcional opcionalmente con una secuencia de ADN espaciadora que no afecta a la estabilidad genética de dicha secuencia de ADN funcional ni al engarce genético entre la secuencia de ADN no funcional y la secuencia de ADN funcional.

20 Neutralidad biológica/seguridad ecológica de la secuencia de ADN no funcional

A. Efecto de la secuencia de ADN no funcional sobre el rendimiento de la secuencia de ADN funcional (transgén)

La proximidad física de la secuencia de ADN no funcional y del ADN que codifica el transgén que interesa (secuencia de ADN funcional) puede afectar al rendimiento del transgén que interesa en el organismo tratado por ingeniería genética de muchas maneras, la mayor parte de las cuales no serían deseables. Algunos de los parámetros que tienen un interés especial a este respecto incluyen: la estabilidad genética del inserto de ADN combinado; el efecto directo o indirecto de la secuencia de ADN no funcional sobre la transcripción y la traducción del transgén que interesa, en particular una inhibición del nivel de expresión; el potencial de disparar el silenciamiento del transgén; etc. La mayor parte de estos efectos se pueden vigilar o medir por una evaluación directa de dos grupos de organismos transgénicos: (a) los que tienen tanto el transgén que interesa como también la secuencia de ADN no funcional, incluyendo opcionalmente un espaciador entre medias, y (b) los que tienen solamente el transgén. Para evitar la mayor parte de los problemas de esta clase, aquí también deberían realizarse de antemano unas búsquedas en bases de datos de secuencias, con el fin de reducir al mínimo la creación inadvertida de unas secuencias tales como las que imitan a secuencias víricas que son conocidas por disparar el silenciamiento en eucariotas. En casos especiales, la secuencia de ADN espaciadora y la no funcional pueden incluso tener un efecto deseable, p.ej. una función como la de un intensificador o aislador de la expresión.

Además, se describe un ensayo experimental para analizar el efecto de la secuencia de ADN no funcional sobre el rendimiento de la secuencia de ADN funcional, que comprende las siguientes etapas:

a) crear unas moléculas de ADN que comprenden dicha secuencia de ADN funcional (i) sin y (ii) con dicha secuencia de ADN no funcional conectada con la secuencia de ADN funcional opcionalmente a través de una secuencia de ADN espaciadora;

(b) introducir dichas moléculas de ADN dentro de unas células para crear células transgénicas y organismos transgénicos que comprenden dicha secuencia de ADN funcional (i) sin y (ii) con dicha secuencia de ADN no funcional conectada con la secuencia de ADN funcional opcionalmente a través de una secuencia de ADN espaciadora.

(c) comparar el rendimiento de dicha secuencia de ADN funcional en los dos conjuntos de células transgénicas y/o de organismos transgénicos, siendo definido dicho rendimiento por al menos uno de los siguientes parámetros: capacidad de transcripción y traducción, nivel de traducción, sensibilidad al silenciamiento, mutabilidad de secuencias; y

(d) identificar una secuencia de ADN no funcional, opcionalmente con una secuencia de ADN espaciadora, que no afecta al rendimiento de dicha secuencia de ADN funcional.

Usando la degeneración del esquema de codificación se pueden crear varias secuencias de ADN no funcionales, que difieren en su secuencia de nucleótidos pero que codifican el mismo mensaje de información deseado. La secuencia de ADN no funcional que exhibe unas propiedades sumamente apropiadas en el anterior ensayo se puede seleccionar para dicho procedimiento de producir un organismo tratado por ingeniería genética.

Este ensayo se lleva a cabo de manera preferible con células y/o con organismos de la especie que ha de ser tratada por ingeniería genética mediante el procedimiento del invento. Si se han de evitar unos largos periodos de tiempo de generación de los organismos de dicha especie, dicho ensayo se puede llevar a cabo en un cultivo de células.

B. Neutralidad genética y fisiológica de la secuencia de ADN no funcional

En este invento, la presencia de la secuencia de ADN no funcional no debería afectar desfavorablemente a la idoneidad biológica del organismo tratado por ingeniería genética que la contiene. En una forma preferida de realización del invento, la secuencia de ADN no funcional no debería ser expresada, p.ej. transcrita y/o traducida (excepto para los casos en donde nosotros deseamos que el mensaje técnico sea traducido para dar un polipéptido o que por lo menos sea transcrito). Con el fin de evitar cualquier traducción, nosotros proponemos flanquear a la secuencia de ADN no funcional por seis codones de detención de la traducción (tres por cada hebra de ADN, uno por cada marco de lectura), haciendo imposible de esta manera la creación de un producto de traducción incluso si la secuencia de ADN no funcional será transcrita por un proceso de transcripción "fugaz". Es aconsejable usar unas búsquedas frente a diferentes bases de datos de secuencias disponibles, por ejemplo mediante una búsqueda Blast Search en cuanto a la presencia de cualesquiera putativos promotores, intensificadores de la transcripción o de la traducción o bien de una homología con genes y transgenes residenciales en la secuencia de ADN no funcional (una lista de diferentes herramientas de búsqueda y análisis para biología molecular está disponible de http://www.biophys.uni-duesseldorf.de/BioNet/Pedro/research_tools.html). En el EJEMPLO 3, la búsqueda se realiza con el fin de evitar la presencia, en la secuencia de ADN no funcional, de una homología con un fragmento de ARN del anfitrión, que es complementario con un ARNm residente que inadvertidamente podría afectar a la expresión de un gen residencial y consiguientemente a la idoneidad del anfitrión. Sin embargo la búsqueda en las bases de datos está restringida en la mayor parte de los casos a unos organismos con genomas completamente secuenciados y los resultados de las búsquedas pueden no detectar todas las homologías existentes para otros organismos. Para afrontar este problema, la secuencia de ADN no funcional se puede usar también como una sonda de hibridación para un análisis por hibridación Southern o dot-blot (puntos - borrones) de un ADN anfitrión en diferentes condiciones de diferente rigurosidad (Sambrook, Fritsch & Maniatis, 1989, Molecular Cloning [Clonación molecular], Cold Spring Harbor Laboratory Press, N.Y.).

Además se puede llevar a cabo un ensayo experimental análogo al que se ha descrito más arriba dentro del párrafo (A) con el fin de identificar unos efectos indeseados de la secuencia de ADN no funcional sobre la idoneidad biológica del organismo tratado por ingeniería genética. El rendimiento de dichas células transgénicas y/o de dichos organismos transgénicos es evaluado mediante por lo menos uno de los siguientes parámetros en la etapa (c): rendimiento, productividad general, resistencia a estreses bióticos y abióticos, etc.

C. Seguridad ecológica

Las modificaciones consideradas en este invento deberían dar como resultado unos organismos tratados por ingeniería genética que no tengan componentes de propiedades que los harían ecológicamente peligrosos o inseguros. La probabilidad de uno de tales efectos inadvertidos es extremadamente remota, puesto que la longitud relativamente corta del ADN, que se necesita en la práctica para una marcación técnica, hace improbable que dicha secuencia de ADN no funcional cambie el rendimiento del organismo de una manera tan espectacular que se cree una plaga o "supermaleza" ecológicamente superior. También, las longitudes de la secuencia de ADN no funcional considerada en este invento son demasiado cortas para que ésta sea capaz de manifestar cualquier función genética o biológica independiente en un organismo (tal como la de funcionar como un promotor o convertirse en una entidad autorreplicante, un vector similar a un virus), o la de adquirir dichas funciones al realizar una transferencia horizontal desde una clase de organismos biológicos a otra.

También se describe un ensayo experimental para averiguar la seguridad ecológica de la secuencia de ADN no funcional. Este ensayo se puede llevar a cabo de una manera análoga a la de los ensayos más arriba descritos dentro de los párrafos (A) y (B). El rendimiento de dichas células transgénicas y/o de dichos organismos transgénicos se compara en la etapa (c), siendo definido dicho rendimiento mediante por lo menos uno de los siguientes parámetros: preferencia de cruce con las plantas autóctonas locales de la misma especie o con especies de hierbas y malezas relacionadas. Dicho ensayo se lleva a cabo de manera preferible con plantas en un invernadero.

Consideraciones técnicas

Técnicamente, la síntesis de un tramo de codificación de ADN (secuencia de ADN no funcional) se puede ejecutar gradualmente por uso de unos conjuntos de cebadores que se solapan. Esto no requiere la síntesis de unos cebadores muy largos y se puede realizar en cualquier laboratorio de biología molecular normal. El producto final de la PCR, de manera preferible flanqueado por dos sitios de restricción que se cortan raramente (p.ej. Not1), puede ser clonado dentro de un pequeño vector plásmido de alto número de copias (p.ej. de la familia pUC o pBS), tal como se muestra en la Figura 1. En este vector, el mensaje de información codificado en el tramo de ADN puede ser puesto al día con facilidad usando dos cebadores habituales fosforilados, que no se solapan pero que comienzan desde la misma posición en dos orientaciones opuestas (cebadores C y D, Fig 1). Estos cebadores pueden introducir todos los necesarios cambios dentro del tramo de codificación. El producto de PCR que lleva dichos cambios puede ser tratado con un fragmento Klenow de la polimerasa 1 de ADN, renovadamente ligado y transformado en el seno de *E. coli*. El tramo de codificación puede ser vuelto a clonar con facilidad dentro de cualquier vector que interese, usando técnicas clásicas de biología molecular. El tramo de ADN que codifica una información no biológica puede ser detectado en un organismo modificado genéticamente (o incluso un producto procedente de dicho organismo, puesto que éste siempre tiene trazas de ADN) por uso de unos cebadores

complementarios con las secuencias de reconocimiento flanqueadoras de la secuencia de ADN no funcional (cebadores A y B, Fig. 1). Teniendo en cuenta que una mala concordancia con el molde en el extremo 3' del cebador no permite que se desarrolle la PCR, una prolongación de los cebadores A y B dentro de la región de núcleo de codificación variable (de la secuencia no funcional) permite la detección de dos o más diferentes marcaciones en el mismo GMO o en su derivado. Los productos de PCR pueden ser subclonados o secuenciados directamente y la información no biológica puede ser descodificada, permitiendo de esta manera la identificación de un producto GM, también en un producto derivado de un GMO.

Adición de un fragmento de mensaje que puede ser traducido para dar un polipéptido

Para facilitar aún más la detección de un GMO, la región que codifica el mensaje de información puede codificar también un segmento que contenga una información que es expresable por la maquinaria genética del organismo vivo. Dicha información, después de una transcripción y de una traducción, produce un corto polipéptido universal fácilmente detectable por técnicas inmunológicas rápidas y sensibles, y que es indicativo de la naturaleza transgénica del organismo en cuestión. En este enfoque, las cuestiones tales como la neutralidad biológica del polipéptido añadido deberían de ser afrontadas de antemano, de una manera muy similar a como se ha bosquejado en los precedentes párrafos.

En resumen, la neutralidad biológica, la seguridad ecológica y la estabilidad genética del mensaje se pueden afrontar por el método de diseño y los análisis de secuencias que se han descrito más arriba y se describirán en los EJEMPLOS 3 y 4. El mensaje técnico puede ser sometido a ensayos adicionales usualmente requeridos para organismos transgénicos. Dichos ensayos tienen usualmente un alto grado de similaridad y las diferencias dependen en muy gran manera de la naturaleza de los transgenes que se han de ensayar, por ejemplo, una planta que resiste a los herbicidas es evaluada de una manera diferente que una planta que está protegida contra las plagas de insectos. Un efecto de la secuencia de ADN no funcional sobre la expresión de transgenes puede ser evaluado por simple comparación de unos organismos transgénicos que difieren solamente por la presencia o ausencia de un mensaje técnico. Las averiguaciones por los técnicos desarrolladores de productos son realizadas de acuerdo con los principios desarrollados por expertos en el medio ambiente por todo el globo terrestre (U.S. National Research Council [Consejo de Investigación Nacional de los EE.UU.], 1989, *Field Testing Genetically Modified Organisms: Framework for Decisions [Ensayo en el campo de organismos modificados genéticamente: Marco para las decisiones]*. Committee on Scientific Evaluation of the Introduction of Genetically Modified Microorganisms and Plants into the Environment [Comité sobre la evaluación científica de la introducción en el medio ambiente de microorganismos y plantas que se han modificado genéticamente]. National Academy Press, Washington, D.C.; Organization for Economic Cooperation and Development [Organización para la cooperación y el desarrollo económico], 1992, *Safety considerations for biotechnology [Consideraciones de seguridad para la biotecnología]*, OECD, Paris, 50 pp.; Gobierno del Canadá, 1994, *Assessment criteria for determining environmental safety of plants with novel traits*. [Criterios de averiguación para determinar la seguridad medioambiental de plantas con nuevos rasgos] Dir9408, 16 de Diciembre de 1994. Plant Products Division [División de productos vegetales], Plant Industry Directorate [Directorio de la industria vegetal], Agriculture and Agri-food Canadá). Los resultados de estas averiguaciones son enviados a la EPA, a la USDA, a la Food and Drug Administration y a otras corporaciones reguladoras a lo largo del mundo, que evalúan los datos al determinar si un producto biotecnológico debería ser aprobado definitivamente para la comercialización. Diferentes parámetros pueden ser incluidos en dichas averiguaciones.

Por ejemplo un paquete de datos clásicos, revisado por el Animal and Plant Health Inspection Service (APHIS) del US Department of Agriculture típicamente incluye una información en los siguientes sectores (véase <http://www.cast-science.org/biotechnology/20001011.htm>):

Análisis de la composición. Los investigadores buscan cualquier cambio en la composición o en la fisiología de las plantas que pudiera tener un efecto ecológico. Un aumento en los agentes tóxicos existentes, por ejemplo, provocaría preocupaciones acerca de la vida salvaje que pudiera alimentarse con la planta cultivada.

Germinación/dormición. Estos estudios, realizados en una amplia gama de regímenes de temperatura, examinan si la semilla biotecnológica difiere de la semilla convencional. Una semilla con dormición aumentada, por ejemplo, puede tener una capacidad aumentada para sobrevivir en la tierra durante el invierno, lo cual podría conducir a un efecto negativo tal como una herbosidad.

Longevidad de bancos de semillas. Estos estudios determinan si la semilla sobrevive en diferentes profundidades en la tierra, en almacenamiento o en una colección de referencia.

Crecimiento y reproducción. Se comparan diversas características de crecimiento y reproducción de la planta biotecnológica y la planta convencional. Una planta de maíz, por ejemplo, sería examinada en cuanto al recuento de emergencias tempranas, los días hasta el desprendimiento de polen, los días hasta la formación de barbas, la altura de las mazorcas, la altura de las plantas, los días hasta alcanzar la madurez, el número de las mazorcas caídas, la humedad de los granos, el peso de ensayo, el rendimiento y otros factores. Muchas de estas características son

evaluadas rutinariamente para nuevas variedades de plantas. Se podrían averiguar cualesquiera diferencias para determinar si ellas podrían dar a la planta modificada una ventaja ecológica o afectar a la vida salvaje.

Cruce. Estas averiguaciones específicas de plantas útiles examinan si la inserción de un gen aumenta la capacidad de la planta para convertirse en una maleza. Por ejemplo una planta de soja no tiene parientes silvestres en los EE.UU., por lo que el cruce con parientes silvestres no es un problema en los EE.UU. para nuevas variedades de soja desarrolladas mediante biotecnología. En los casos en donde existen plantas silvestres compatibles, el APHIS considera si la transferencia de un rasgo específico, tal como el de protección contra insectos, podría dar a la planta silvestre una ventaja con respecto a otras plantas. También, basándose en la averiguación del APHIS, la EPA ha prohibido la plantación de plantas útiles de Bt en unas regiones geográficas aisladas en las que se considera posible el cruce con parientes silvestres (p.ej. parientes silvestres de algodón en Hawái). A la vista de los recientes datos que han demostrado una transferencia de transgenes de maíz a partir de variedades transgénicas a plantas autóctonas tradicionales en Méjico (Quist & Chapela, *Nature*, 414, 541-542 (2001), puede ser también importante una evaluación específica del patrón de cruces.

Averiguación de la idoneidad. Muchas plantas útiles agrícolas, a causa de los años de crianza cruzada tradicional, no son idóneas para sobrevivir fuera de un entorno agrícola altamente manipulado. En la averiguación de una planta cultivada biotecnológica, el APHIS considera si la introducción de un gen podría mejorar la capacidad de una planta para resultar establecida y ganar la competencia con la vegetación silvestre, en otras palabras, para convertirse en una hierba.

Potencial alelopático. En la naturaleza, algunas plantas pueden tener un efecto herbicida sobre otras plantas. El APHIS examina si la introducción de un nuevo gen en una planta cultivada proporciona dicha capacidad alelopática.

Observaciones en el campo. Múltiples años de pruebas en el campo en múltiples sitios proporcionan oportunidades para observar cualesquiera efectos inesperados sobre la tierra, el agua, la vida salvaje u otra vida vegetal. Sin embargo, creemos que la secuencia de ADN no funcional del invento, propiamente dicha, no requerirá dicho análisis a fondo, puesto que su expresión al nivel del ARN o de las proteínas puede ser controlada de una manera eficiente por el diseño de dicha secuencia. Usualmente, un ensayo requerido de un organismo transgénico servirá también como ensayo para una secuencia de ADN no funcional. Pueden ser necesarios estudios toxicológicos solamente en unos casos específicos de secuencias de ADN no funcionales, que pueden ser traducidas para dar un producto.

Se describe además un método de detectar miembros de un conjunto de organismos diversamente tratados por ingeniería genética, en donde dicho conjunto se define de la siguiente manera:

- (i) cada miembro contiene por lo menos una secuencia de ADN funcional;
- (ii) diferentes miembros contienen diferentes secuencias de ADN funcionales;
- (iii) cada secuencia de ADN funcional está flanqueada en su lado situado secuencia arriba por una primera secuencia de reconocimiento previamente definida y en su lado situado secuencia abajo por una segunda secuencia de reconocimiento previamente definida,
- (iv) todas las primeras secuencias de reconocimiento previamente definidas en dicho conjunto son idénticas;
- (v) todas las segundas secuencias de reconocimiento previamente definidas en dicho conjunto son idénticas,
- (vi) dichas primeras y segundas secuencias de reconocimiento previamente definidas están adaptadas para una amplificación por PCR de dicha secuencia de ADN funcional;
- (vii) dicha secuencia de ADN funcional y dicha secuencia de ADN no funcional permanecen engarzadas durante la reproducción del organismo durante un período de tiempo relevante para la práctica;

con lo que dicho método está caracterizado por las siguientes etapas:

- (a) someter a una muestra de ADN de dicho organismo tratado por ingeniería genética a una amplificación por PCR usando unos cebadores que corresponden a dichas primeras y segundas secuencias de reconocimiento previamente definidas, para obtener un producto de PCR;
- (b) determinar si se ha obtenido o no en la etapa (a) un producto de PCR;
- (c) interpretar un resultado positivo de la etapa (b) como una evidencia de la presencia de un miembro del conjunto más arriba definido en dicha muestra; y
- (d) secuenciar opcionalmente el producto de PCR de la etapa (a) para identificar la secuencia de ADN funcional.

Este método permite detectar diferentes organismos tratados por ingeniería genética mediante una amplificación por PCR usando los mismos cebadores. Dicha PCR permite amplificar un transgén que está flanqueado en ambos lados por unas secuencias de reconocimiento previamente definidas. La formación de un producto de PCR identifica al organismo genéticamente dotado de un transgén. Dicho transgén puede ser identificado por secuenciación del producto de PCR. Dichas secuencias de reconocimiento previamente definidas están convenidas generalmente de manera preferible, tal como se ha descrito para dicho método de producir una planta transgénica. Las secuencias de reconocimiento previamente definidas con una secuencia idéntica se usan como secuencias flanqueadoras de un transgén en un conjunto de organismos tratados por ingeniería genética de una manera diversa. Cada una de dichas secuencias de reconocimiento previamente definidas tiene una posibilidad despreciable de presentarse de un modo

5 natural en el genoma de un organismo. Los miembros de dicho conjunto difieren de manera preferible en el hecho de que ellos contienen diferentes transgenes. De manera alternativa o adicional, dichos miembros de dicho conjunto pueden diferir aún más en el hecho de que ellos pertenecen a diferentes familias taxonómicas, a diferentes géneros pero a la misma familia, a diferentes especies pero al mismo género, o a diferentes variedades pero a la misma especie. Si dichos miembros pertenecen a la misma especie, ellos son modificados con diferentes transgenes. Los organismos de este método son de manera preferible plantas, notablemente plantas útiles superiores. El uso de unas idénticas secuencias de reconocimiento previamente definidas en organismos tratados por ingeniería genética de manera diversa, permite el uso de los mismos cebadores para detectar un miembro de dicho conjunto. También se proporciona un organismo tratado por ingeniería genética detectable por dicho método de detección.

10 **TABLA 1**

USO DE CODONES PROPUESTO PARA CODIFICAR EN EL ALFABETO INGLÉS Y NÚMEROS

Carácter :	Codones en forma de tripletes	
A	GAT	
B	GCG	GCA
C	GTC	GTT
D	AGT	AGC
E	GCC	GCT
F	AAA	AAC
G	ACG	ACA
H	TGT	TGC
I	ACC	ACT
J	CCG	CCA
K	TAT	TAC
L	TCG	
M	CCT	CCC
N	TCC	TCT
O	GTG	
P	GTA	
Q	CGC	
R	AGA	AGG

S	ATA	
T	ATC	ATT
U	CTG	CTA
V	CTC	CTT
W	TTG	
X	TTA	
Y	TAA	
Z	TAG	
“espacio“	ATG	TTC
0	CAT	CAC
1	CAA	CAG
2	AAT	AAG
3	GAA	GAG
4	TGA	TGG
5	CGT	CGA
6	GGT	TTT
7	GGC	GAC
8	GGA	TCA
9	CGG	GGG

TABLA 2

USO DE CODONES PROPUESTO PARA CODIFICAR EN EL ALFABETO INGLÉS Y NÚMEROS

Carácter :	Codones en forma de tripletes	
A	GAT	
B	GCG	AGC
C	GTC	CGA
D	AGT	GCA
E	GCC	AGG
F	AAA	TGC
G	ACG	CTA
H	TGT	CCC

I	ACC	GTT
J	CCG	TAC
K	TAT	CCA
L	TCG	
M	CCT	TTC
N	TCC	CAG
O	GTG	
P	GTA	
Q	CGC	
R	AGA	GAG
S	ATA	
T	ATC	TCT
U	CTG	ACT
V	CTC	GCT
W	TTG	
X	TTA	
Y	TAA	
Z	TAG	
"espacio"	ATG	CAC
0	CAT	GGG
1	CAA	TTT
2	AAT	TGG
3	GAA	CTT
4	TGA	ATT
5	CGT	AAG
6	GGT	TCA
7	GGC	ACA
8	GGA	AAC
9	CGG	GAC

EJEMPLO 1Protocolo de codificación

5 El código biológico propuesto en la Tabla 1 se puede usar para codificar una información de carácter no biológico. Considerando que hay por lo menos 37 caracteres alfanuméricos que han de ser codificados (26 letras del alfabeto) en el caso de la lengua inglesa, un signo de "espacio" (espaciado) y 10 números) y solamente están disponibles 64 codones en forma de tripletes, la elección de la designación de uno o dos codones para el carácter está basada en la esperada frecuencia de uso para el carácter específico. Como resultado, algunas letras tienen solamente un codón en forma triplete. El signo de "espacio" tiene dos codones como el carácter usado con la máxima frecuencia.

10 Sin embargo, nosotros recomendamos usar el primer codón del signo de "espacio" (ATG) al comienzo de una información codificada y el segundo codón (TTC) al final de una información codificada. La elección del signo de "espacio" en el centro de una información codificada deberá ser de libre elección y la preferencia puede ser dictada solamente por la necesidad de evitar la creación de sitios de restricción o de otras secuencias indeseadas.

EJEMPLO 2

15 Generación y análisis de plantas transgénicas que llevan la información de carácter no biológico

En este ejemplo, se emplea el esquema de codificación de la Tabla 1. Un ADN sintético fue sintetizado por la compañía Genart (de Regensburg, Alemania) como un "producto de ligación". El producto traducido se lee como: "I.CAN.GENETICS.NEC.VERO. TERRAE.FERRE.OMNES.OMNIA.POSSUNT"- seguido (en un marco en dirección opuesta) por "SCITENEG .NAC. I". La traducción de la expresión en latín NEC.VERO.TERRAE.FERRE.OMNES.OMNIA.POSSUNT" significa "tampoco cualquier tierra puede llevar cualquier fruto (Vergil, "Georgica", 2, 109). Las dos secuencias flanqueadoras con un texto idéntico ("I.CAN.GENETICS") codificado en ambas hebras de ADN representan a una parte de secuencia conservada que codifica una información técnica.

25 El producto de ligación (que se compone de un producto, obtenido por PCR, de oligos ensamblados) se clonó como un fragmento de BamHI-EcoRI en pICBV10 (Fig. 5), dando como resultado el plásmido pIC4100. El inserto fue comprobado por secuenciación. El plásmido fue inmovilizado dentro de la cepa de *Agrobacterium* GV3101 y usado para la transformación.

30 El ADN T del plásmido pIC4100 (Fig. 2) fue introducido en plantas *Arabidopsis thaliana* (Col-0) como ha sido descrito por Bent y colaboradores, (1994, Science, 285. 1856-1860). Se cosecharon semillas las tres semanas después de una infiltración en vacío, se esterilizaron y se escrutaron para descubrir transformantes en un medio de GM + 1 % de glucosa (Valvekens y colaboradores, 1988, Proc. Natl. Acad. Sci. USA, 85, 5536-5540.) que contenía 50 mg/l de kanamicina. El ADN T de pIC4100 fue introducido también en *Nicotiana tabacum* usando un protocolo de transformación de discos de hojas mediada por *Agrobacterium* (Horsh y colaboradores, 1985, Science, 227, 1229-1231). Unas plantas de tabaco transgénicas fueron recuperadas aplicando una selección con kanamicina. La transformación de *Arabidopsis thaliana* y de *Nicotiana tabacum* se realizó también con el plásmido pICBV10 (Fig. 5), que contiene un marcador seleccionable pero no contiene ningún mensaje técnico codificado por un ADN. Se obtuvieron veinte transformantes primarios por cada construcción artificial para cada especie de planta (en total 80 plantas transgénicas) y se usaron para un análisis comparativo con el fin de determinar el efecto de un mensaje técnico codificado por un ADN sobre la idoneidad de las plantas, la expresión de transgenes, la estabilidad, etc. Los experimentos de transformación con dos diferentes plásmidos (pIC4100 y pICBV10) se realizaron en paralelo, usando el mismo material vegetal cultivado en unas condiciones idénticas. La selección en cuanto a transformantes se llevó a cabo también en unas condiciones idénticas para los plásmidos, el testigo (pICBV10) y el ensayado (pIC4100). Los experimentos demostraron que la frecuencia de transformación era comparable para los vectores testigo y ensayado. También, no hubo diferencias fenotípicas entre las plantas testigos y las plantas que eran portadoras de la información técnica. Esto demuestra que no hay ningún efecto detectable del ADN que codifique información técnica acerca de la frecuencia de transformación y la idoneidad de la planta, comparadas con las de plantas testigos.

50 Hemos ensayado también la estabilidad del ADN que codifica una información técnica dentro de anfitriones bacterianos y vegetales. El siguiente conjunto de cebadores se usó para amplificar un fragmento de ADN de 231 pb (pares de bases):

F1 (5'-GTC GAC TGA ATA TAA TGC GCA AAC TG-3')

R1 (5'-GAA TTC TAA ATC TGA TGT GCA AAT TAA GG-3')

55 Como un molde para la PCR se usó 1 µl de una suspensión de bacterias o un pequeño trozo de un tejido de planta por 50 µl de una mezcla para la PCR. El tejido de planta para la PCR fue preparado tal como se describió por Klimyuk & colaboradores (1993, Plant J., 3, 493-494). La suspensión de *E.coli* fue tomada para una PCR después de un crecimiento durante dos horas en un medio LB recientemente inoculado, con una apropiada selección (con 50

µg/ml de carbenicilina). La suspensión de *Agrobacterium* fue tomada a partir de un cultivo durante una noche (LB con 50 µg/ml de rifampicina, 50 µg/ml de carbenicilina). Las condiciones de PCR fueron como sigue: 94°C durante 20 s (segundos); 60°C durante 20 s; 72°C durante 1 min; 35 ciclos.

5 A pesar de la presencia de unas repeticiones invertidas imperfectas que flanquean al mensaje, el fragmento de ADN que interesa fue detectado en todos los clones analizados de *E. coli* y *Agrobacterium tumefaciens* así como en todos los transformantes primarios que llevaban el ADN T de pIC4100 (se analizaron 20 transformantes primarios de tabaco y 20 transformantes primarios de Arabidopsis). Se llevó a cabo un análisis de secuencias de los productos de la PCR a partir de un tejido de planta. No se detectaron errores en el ADN que codifica el mensaje. El mismo análisis de secuencias se llevó a cabo también usando unos productos de la PCR a partir de una auto-progenie de dos transformantes primarios de Arabidopsis. No se detectaron cambios en la secuencia de ADN.

EJEMPLO 3

Análisis de un ADN que codifica información técnica acerca de la presencia de secuencias indeseadas

15 La manera más simple de detectar la presencia de unas secuencias que son idénticas o similares a las secuencias de las bases de datos consiste en realizar una búsqueda Blast. Hemos usado la frecuencia mostrada en la Fig. 3 para buscar frente a bases de datos públicas. Los resultados de las búsquedas BlastN y BlastX se presentan en el Apéndice 1. Los resultados de la búsqueda BlastN revelaron la presencia de unos cortos tramos de secuencias de ADN (de 19-21 pb) idénticos a las secuencias de mamíferos (secuencias de ADNc), de *Drosophila melanogaster* y de *Aspergillus parasiticus*. Sin embargo, no se detectó ninguna analogía para la secuencia del ADN de la planta. Solamente la búsqueda BlastX reveló una homología significativa con una proteína hipotética de *Arabidopsis thaliana* y comprende el tramo de ADN que codifica 41 residuos de aminoácidos. Sin embargo, este tramo de ADN técnico contiene tres codones de detención que rompen la continuidad del putativo producto de traducción en diferentes posiciones, haciendo de esta manera imposible la síntesis de una proteína artificial con homología para una proteína hipotética de *Arabidopsis*.

EJEMPLO 4

Medición de la frecuencia de recombinación dentro del ADN que codifica una información técnica

30 Es aconsejable medir la frecuencia de recombinación en la región abarcada por el ADN que codifica una información técnica y el ADN de engarzador que separa al ADN con información técnica con respecto de uno o varios transgén(es) tales como un ADN funcional. Esto puede ayudar a evitar la creación accidental de cualesquiera "puntos calientes" de recombinación. El esquema general de una construcción artificial diseñada para esta finalidad se muestra en la Fig. 4. La secuencia de ADN que se ha de ensayar puede ser insertada entre un marcador seleccionable (un gen que confiere resistencia a antibióticos) y un marcador contra-seleccionable (gen condicionalmente letal que confiere una sensibilidad a un producto químico). Cualquier gen bacteriano que confiera resistencia a antibióticos (ampicilina, gentamicina, tetraciclina, kanamicina, neomicina, higromicina, etc.) se podría escoger para servir como marcador seleccionable. Como una opción del marcador contra-seleccionable se puede usar el gen *sacB* de *Bacillus subtilis* que confiere sensibilidad a la sacarosa (Steinmetz y colaboradores, 1983, *Mol. Gen. Genet.*, 191, 138-144; Recorbet y colaboradores, 1993, *Appl. Environ. Microbiol.*, 59, 1361-1366). Los sucesos de recombinación se pueden seleccionar por medio de bacterias en crecimiento en unos medios que contienen el antibiótico apropiado y sacarosa. Solamente sobrevivirán las células bacterianas que hayan perdido por recombinación o tengan una forma inactiva (mutante) del gen *sacB*. Dichas colonias sobrevivientes se pueden usar para un análisis por PCR (cebadores A y B, Fig. 4) con el fin de comprobar la frecuencia de sucesos de recombinación con implicación del fragmento de ADN que codifica una información técnica. Esta frecuencia se puede comparar con la frecuencia de recombinación en la parte restante del vector, con el fin de determinar la presencia de un "punto caliente" de recombinación dentro de la región analizada.

Apéndice 1

Búsqueda BlastN

Secuencias que producen alineaciones significativas:	Nota (bits)	Valor de N
gi [16877788 [gb] BC017130.1] BC017130 Mus musculus, clon MGC	<u>42</u>	0,21
gi [13435767 [gb] BC004742.1] BC004742 Mus musculus, clon IMAG	<u>42</u>	0,21
gi [971274 [dbj] D32249.1]RATNDAP1 Rattus norvegicus ARNm para	<u>42</u>	0,21
gi [16604068 [gb] AC097466.2 Homo sapiens cromosoma 4 clon	<u>38</u>	3,2
55 gi [15451474 [gb] AC009343.8 Drosophila melanogaster, cromos	<u>38</u>	3,2
gi [10727654 [gb] AE003828.2] AE003828 Drosophila melanogaster	<u>38</u>	3,2

gi [3900825 [gb] AC005991.1] AC005991 Aspergillus parasiticus c _38 3,2

_Alineaciones

- 5 > gi [16877788 [gb] BC017130.1] BC017130 Mus musculus, clon MGC: 27629
 IMAGEN: 4506259, ARNm, cds completa
 Longitud = 2595
 Nota = 42,1 bits (21), Esperada = 0,21
 Identities = 21/21 (100 %)
 Hebra = Más / Menos
Petición: 208 atttgcacatcagatttagaa 228
 |||
 10 **Sujeto: 2112 atttgcacatcagatttagaa 2092**
-
- 15 > gi [13435767 [gb] BC004742.1] BC004742 Mus musculus, clon IMAGEN: 3499845, ARNm, cds parcial
 Longitud = 1688
 Nota = 42,1 bits (21), Esperada = 0,21
 Identities = 21/21 (100 %)
 Hebra = Más / Menos
Petición: 208 atttgcacatcagatttagaa 228
 |||
Sujeto: 1216 atttgcacatcagatttagaa 1196
- 20 > gi [971274 [dbj] D32249.1]RATNDAP1 Rattus norvegicus ARNm para la proteína 1 asociada con la neurodegeneración, cds completa
 Longitud = 4758
 Nota = 42,1 bits (21), Esperada = 0,21
 Identities = 21/21 (100 %)
 Hebra = Más / Menos
Petición: 208 atttgcacatcagatttagaa 228
 |||
Sujeto: 2338 atttgcacatcagatttagaa 2318
- 25 > gi [16604068 [gb] AC097466.2] Homo sapiens cromosoma 4 clon RP11-26A2 secuencia completa
 Longitud = 170321
 Nota = 38,2 bits (19), Esperada = 3,2
 Identities = 19/19 (100 %)
 Hebra = Más / Menos
Petición: 131 agtcataagccatgaacat 149
 |||
Sujeto: 7185 agtcataagccatgaacat 7203
- 30 > gi [15451474 [gb] AC009343.8] Drosophila melanogaster, cromosoma 2R, región 47A/47B, clon de BAC BACR08L17, secuencia completa
 Longitud = 182494
 Nota = 38,2 bits (19), Esperada = 3,2
 Identities = 19/19 (100 %)
 Hebra = Más / Menos
Petición: 167 cagcgaatacctaagc 185
 |||
Sujeto: 61278 cagcgaatacctaagc 61260
- 40 > gi [10727654 [gb] AE003828.2] AE003828 Drosophila melanogaster andamio genómico 142000013386047
 Sección 15 de 52, secuencia completa
 Longitud = 259934
 Nota = 38,2 bits (19), Esperada = 3,2
 Identities = 19/19 (100 %)
 Hebra = Más / Menos
Petición: 167 cagcgaatacctaagc 185
 |||
Sujeto: 53026 cagcgaatacctaagc 53008
- 45

Gi [3900825 [gb] AC005991.1] AC005991 Aspergillus parasiticus clon sp0, secuencia completa

Longitud = 37754

Nota = 38,2 bits (19), Esperada = 3,2

Identidades = 19/19 (100 %)

Hebra = Más / Menos

5

Petición: 22 aactgaggagaaaacgaga 40

|||||

Sujeto: 11194 aactgaggagaaaacgaga 11176

Búsqueda BlastX

Secuencias que producen alineaciones significativas:

Nota E
(bits)

10

Valor

gi [15234968 [ref] NP_195630.1] (NC_003075) prote hipotética

32 1,8

Alineaciones

15

gi [15234968 [ref] NP_195630.1] (NC_003075) proteína hipotética [Arabidopsis thaliana]

gi [7487352 [pir] [T08567] proteína hipotética T22F8.90 - Arabidopsis thaliana

gi [4914431 [emb] CAB43634.1] (AL050351) proteína hipotética [Arabidopsis thaliana]

gi [7270902 [emb] CAB80852.1] (AL161594) proteína hipotética [Arabidopsis thaliana]

Longitud = 277

20

Nota = 31,5 bits (70), Esperado = 1,5

Identidades = 19/60 (31 %) Positivos = 32/60 (52 %), Intersticios = 1/60 (1 %)

Marco = +2

Petición: 23 TEEKTRQYVHETSVK*SEPEQSDAQS-NLNAENEP*MSHKP*TSHDQPHQRIPNEQMVSP 199

TE+KT++ + E VK S+PE+ QS ++N E + + HK T + Q+ S+

Sujeto: 163 TEKKTkRIISeKkVqSkPEkLTKQSTSVNREkQSEVEHKDITMTIEkQNLTEkRQIQSY 222

REIVINDICACIONES

1. Un método de producir un organismo vegetal tratado por ingeniería genética, por incorporación simultánea dentro de dicho organismo
- 5 (a) de una secuencia de ADN funcional que contiene un gen o un fragmento de gen; y
 (b) de una secuencia de ADN no funcional no requerida para la función del organismo ni para la función de la secuencia de ADN funcional;
- en donde
- 10 (i) la secuencia de ADN no funcional es provista por cartografiado de un mensaje de información, que se compone de una secuencia de caracteres alfanuméricos para dar una secuencia de ADN de acuerdo con un esquema de codificación previamente definido;
- (ii) dicho mensaje de información está relacionado con dicha secuencia de ADN funcional por el hecho de que contiene una información concerniente a la secuencia de ADN funcional, cuya información indica la presencia de la secuencia de ADN funcional:
- 15 (iii) dicho esquema de codificación previamente definido proporciona un cartografiado desde una pluralidad de posibles mensajes de información para dar una pluralidad de secuencias de ADN;
- (iv) el cartografiado desde una secuencia de ADN hasta un mensaje de información es singular mientras que el cartografiado desde un mensaje de información hasta una secuencia de ADN no es singular;
- 20 (v) en donde la secuencia de ADN no funcional y la secuencia de ADN funcional son incorporadas en el mismo cromosoma,
- en el que la distancia entre la secuencia de ADN funcional y la secuencia de ADN no funcional es más corta que 10.000 nucleótidos para reducir la frecuencia de recombinación entre la secuencia de ADN no funcional y la secuencia de ADN funcional.
2. El método de acuerdo con la reivindicación 1, en el que la secuencia de ADN no funcional y la secuencia de ADN funcional están situadas lo suficientemente próximas en un cromosoma como para impedir una separación por recombinación.
- 25 3. El método de acuerdo con una de las reivindicaciones 1 ó 2, en el que la secuencia de ADN no funcional es un intrón.
4. El método de acuerdo con la reivindicación 3, en el que el intrón está situado dentro de la secuencia de ADN funcional relacionada.
- 30 5. El método de acuerdo con la reivindicación 4, en el que el intrón está situado dentro de una porción altamente conservada de la secuencia de ADN funcional relacionada.
6. El método de acuerdo con una cualquiera de las reivindicaciones 1 hasta 5, en el que la secuencia de ADN no funcional, opcionalmente en conjunción con una secuencia de ADN espaciadora situada entre la secuencia de ADN no funcional y la secuencia de ADN funcional, se escoge de manera tal que ella no sea capaz de formar estructuras secundarias y/o de tal manera que ella esté libre de "puntos calientes" de recombinación o mutación.
- 35 7. El método de acuerdo con una de las reivindicaciones 1 hasta 6, en el que la secuencia de ADN no funcional es provista de por lo menos una secuencia de reconocimiento previamente definida, que permite una identificación y/o un análisis de dicha secuencia de ADN no funcional.
- 40 8. El método de acuerdo con la reivindicación 7, en el que la secuencia de ADN no funcional está flanqueada por lo menos en un lado por la secuencia de reconocimiento.
9. El método de acuerdo con la reivindicación 7 u 8, en el que dicha(s) secuencia(s) de reconocimiento previamente definida(s) está(n) diseñada(s) de manera tal que acomoda(n) un reconocimiento conjunto así como por separado de múltiples secuencias de ADN no funcionales en el organismo.
- 45 10. El método de acuerdo con una de las reivindicaciones 1 hasta 9, en el que el ADN no funcional existe en forma de dos o más segmentos o intrones que
- (i) codifican el mismo mensaje de información, y
- 50 (ii) tienen unas secuencias de nucleótidos que son lo suficientemente diferentes como para no causar recombinaciones, como se permiten por la redundancia del esquema de codificación para obtener una estabilidad aumentada de la información.
11. El método de acuerdo con la reivindicación 10, en el que dichos segmentos están en la misma hebra o en diferentes hebras de ADN y se pueden leer en diferentes orientaciones.

12. El método de acuerdo con una de las reivindicaciones 1 hasta 11, en el que la secuencia de ADN funcional está flanqueada por ambos lados por unas porciones de la secuencia de ADN no funcional.
- 5 13. El método de acuerdo con una de las reivindicaciones 7 hasta 12, en el que dichas secuencias de reconocimiento previamente definidas están situadas de manera tal que permiten una amplificación por PCR de la secuencia de ADN funcional o de porciones de la misma, usando unos cebadores complementarios con dichas secuencias de reconocimiento previamente definidas.
14. El método de acuerdo con una cualquiera de las reivindicaciones 1 hasta 13, en el que la secuencia de ADN no funcional contiene una secuencia de ADN adicional que codifica un polipéptido expresable para una detección rápida.
- 10 15. El método de acuerdo con una de las reivindicaciones 1 hasta 14, en el que la información comprende además una marca registrada; una referencia a una base de datos; una fecha, un lugar y/o el nombre de un productor o de un propietario.

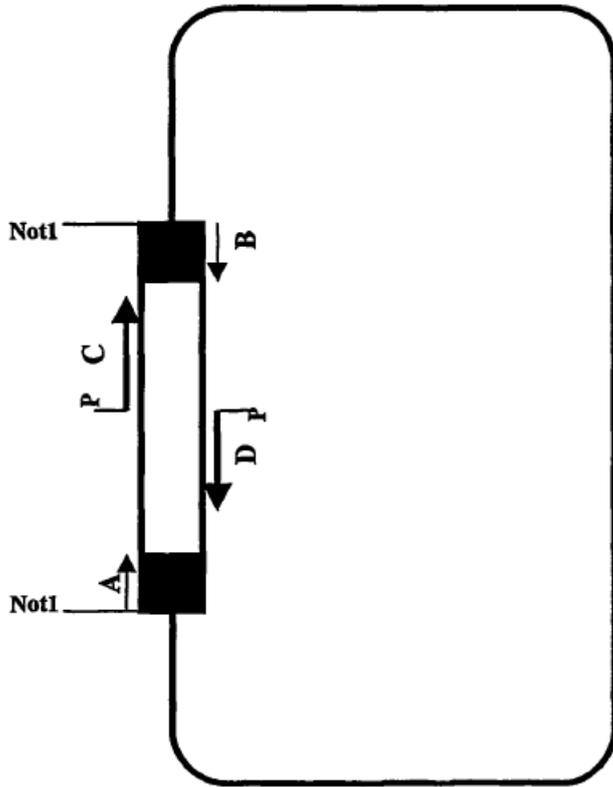


Figura 1

pIC4100

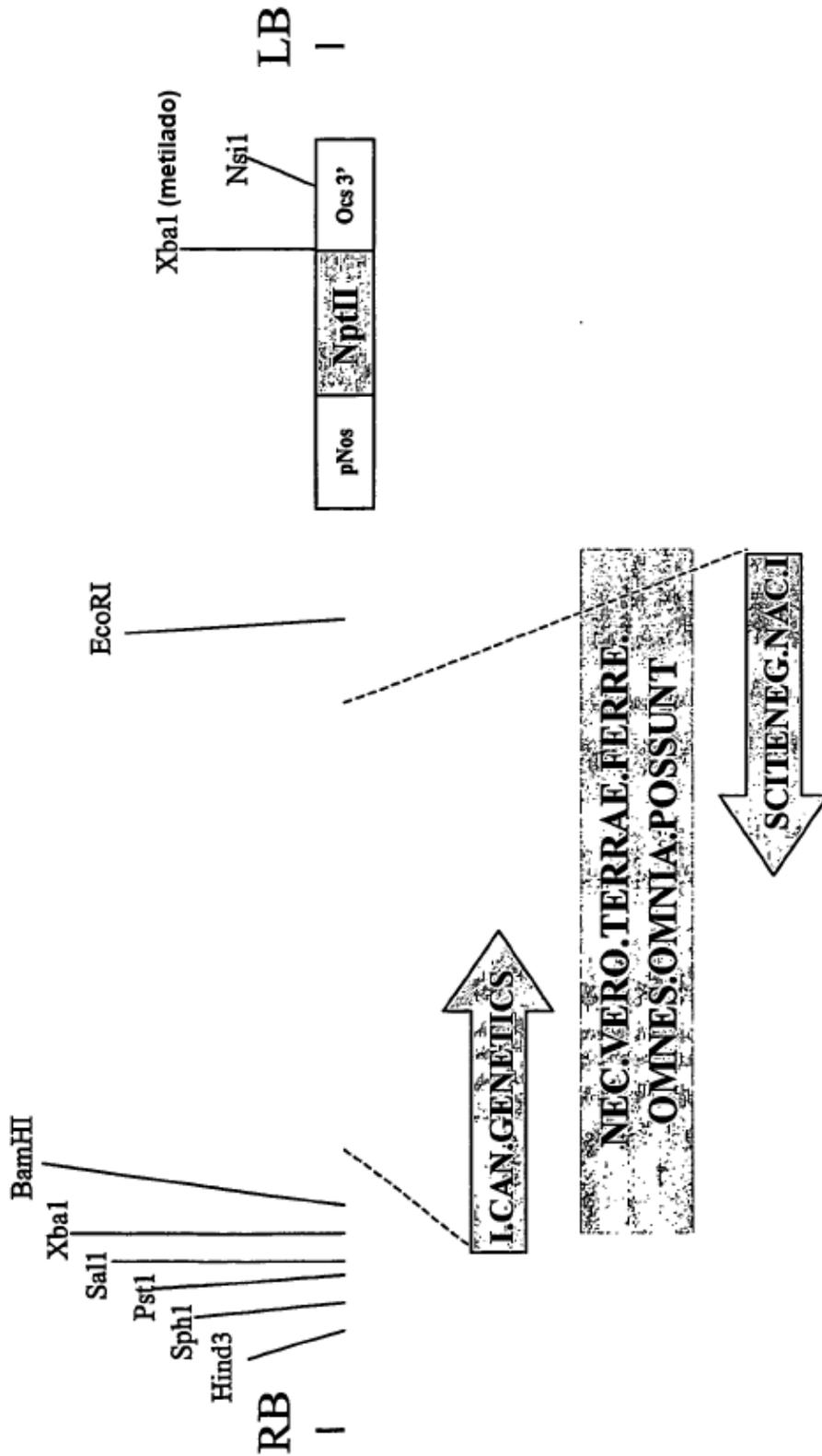


Figura 2

1

GTCGACTGAA TATAATGCGC AACTGAGGA GAAAACGAGA CAATATGTTT ATGAAACGAG
 TGTTAAGTAG AGCGAGCCTG AACAGAGCGA CGCGCAGAGT AATTTGAACG CCGAGAATGA
 GCCATGAATG AGTCATAAGC CATGAACATC GCATGACCAG CCTCATCAGC GAATACCTAA
 TGAGCAGATG GTTTCATTCT CGCCTTAATT TGCACATCAG ATTTAGAATT C

231

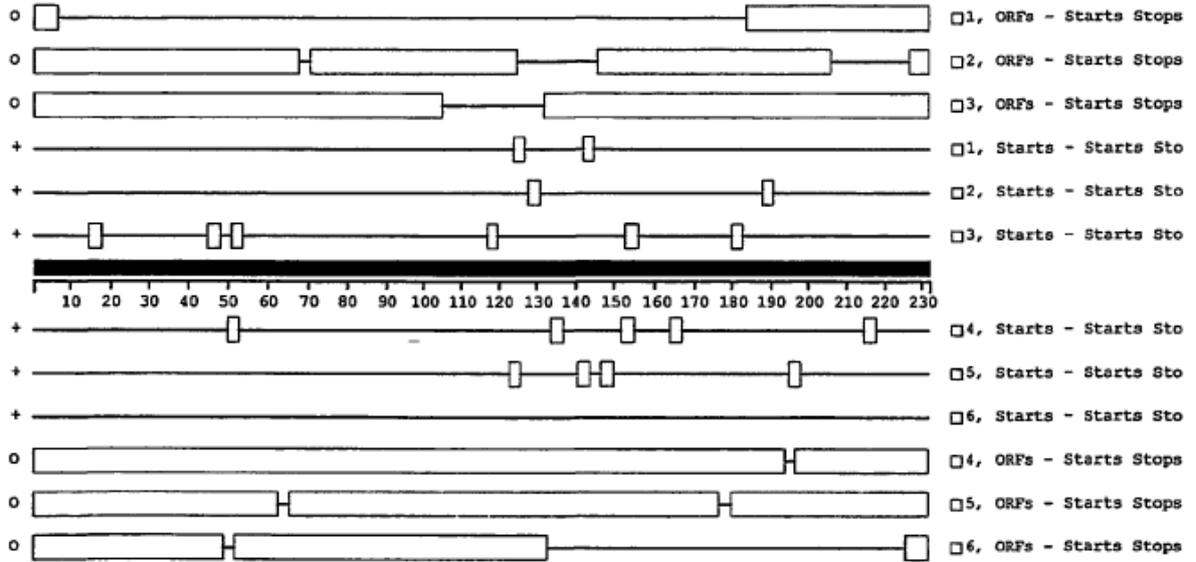


Figura 3

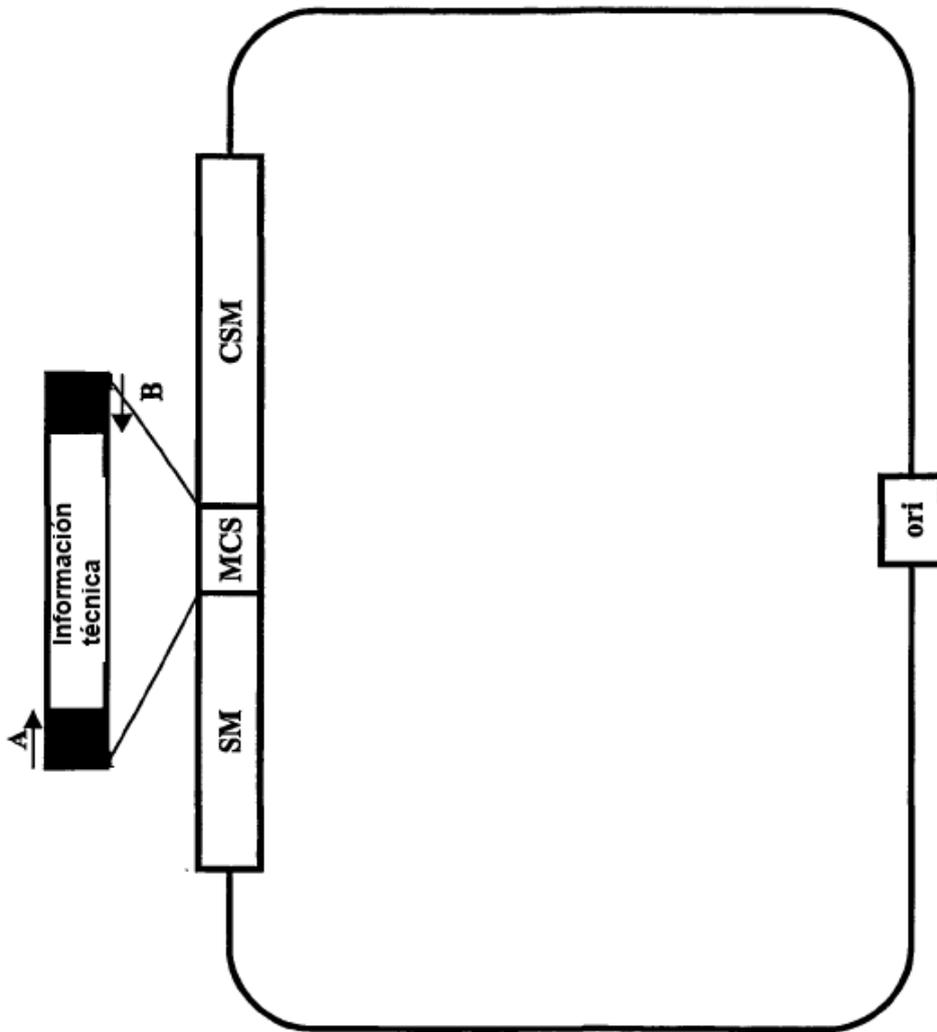


Figura 4

pICBV10

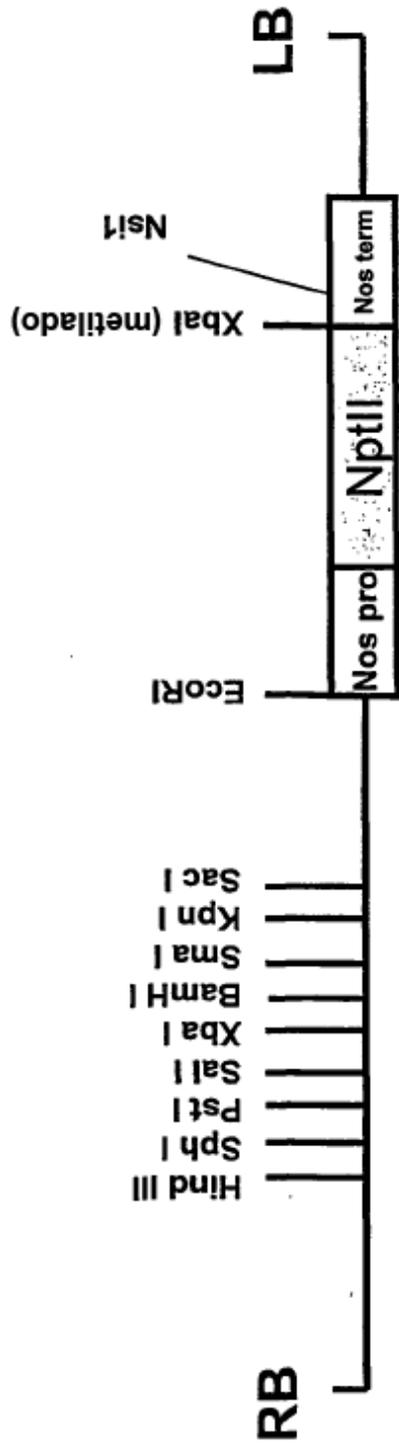


Fig. 5