

(12)

SOLICITUD de PATENTE

(43) Fecha de publicación: **10/05/2006** (51) Int. Cl.7: **C12N 15/05**
(22) Fecha de presentación: **08/11/2004**
(21) Número de solicitud: **PA04011042**

(71) Solicitante:
**CENTRO DE INVESTIGACION Y DE ESTUDIOS
AVANZADOS DEL INSTITUTO POLITECNICO
NACIONAL
Av. Instituto Politécnico Nacional, Numero
2508 07360 Distrito Federal MX**

(72) Inventor(es):
**ANDRES ADOLFO ESTRADA LUNA
Km. 9.7 Libramiento Norte, Carretera Irapuato-Leon,
Apartado Postal 629 Irapuato Guanajuato 36500 MX**

(74) Representante:
**ALBERTO ENRIQUE NAVA GARCES
Av. Instituto Politecnico Nacional, Numero
2508 Distrito Federal 07360 MX**

(54) Título: **SECUENCIAS NUCLEOTIDICAS Y METODOS PARA LA EXPRESION ESPECIFICA DE GENES EN EL
GAMETOFITO FEMENINO, CELULAS REPRODUCTIVAS FEMENINAS, GRANO DE POLEN Y/O LAS
CELULAS REPRODUCTIVAS MASCULINAS DE LAS PLANTAS.**

(54) Title: **NUCLEOTIDE SEQUENCES AND METHODS FOR THE SPECIFIC EXPRESSION OF GENES IN THE
FEMALE GAMETOPHYTE, FEMALE REPRODUCTIVE CELLS, POLLEN GRAIN AND/OR MALE
REPRODUCTIVE CELLS OF PLANTS.**

(57) **Resumen**

El invento presenta la localizacion y los patrones de expresion de promotores (secuencias nucleotidicas regulatorias) que actuan de manera especifica en el gametofito femenino, los gametos femeninos, el gametofito masculino, y/o los gametos masculinos de Arabidopsis thaliana. Se presentan tambien los metodos que fueron utilizados para la identificacion de dichos promotores asi como los metodos para utilizar estos promotores en construcciones de expresion y manipular el desarrollo reproductivo de las plantas. Los promotores aislados e identificados de la invencion, permiten regular la expresion genica de genes seleccionados en plantas, de manera que tales genes puedan ser eficientemente transcritos en las celulas reproductivas femeninas o masculinas de estas. Asi mismo, dichos promotores pueden utilizarse para impedir la transferencia de rasgos transgenicos a las plantas a traves del polen en campo, para proporcionar nutrientes adicionales a la semilla y para conferir resistencia a patogenos y enfermedades en la semilla.

(57) **Abstract**

The invention relates to the localisation and expression patterns of promoters (regulatory nucleotide sequences) which act in a specific manner in the female gametophyte, the female gametes, the male gametophyte and/or the male gametes of Arabidopsis thaliana. The invention also relates to the methods that were used to identify the aforementioned promoters as well as the methods of using said promoters in expression constructions and to manipulate the reproductive development of plants. According to the invention, the identified, isolated promoters can be used to regulate the genic expression of selected genes in plants, so that such genes can be efficiently transcribed in the female or male reproductive cells of same. Said promoters can be used to prevent the transfer of transgenic traits to plants through in-field pollen in order to provide the seed with additional nutrients and to confer resistance to pathogens and diseases to the seed.

Secuencias nucleotídicas y métodos para la expresión específica de genes en el gametofito femenino, células reproductivas femeninas, grano de polen y/o células reproductivas masculinas de las plantas

5 **Campo del invento.**

Este invento se obtiene en el campo de la ingeniería genética y la biología molecular de plantas. De manera más específica, el invento se relaciona con la caracterización de secuencias nucleotídicas novedosas específicas al gametofito femenino, a los gametos femeninos, al gametofito masculino, y/o de los gametos masculinos de las plantas. Estas secuencias permiten la expresión específica temporal y espacial de productos génicos en el gametofito masculino o femenino, o las células reproductivas masculinas o femeninas de plantas transgénicas.

Antecedentes.

15 La expresión de genes (expresión génica) depende de una serie de procesos que se originan en la molécula de ADN. El control y la regulación de la expresión génica pueden ocurrir a partir de diferentes mecanismos. Se considera generalmente que el inicio de la transcripción de un gene (o gen) constituye el control predominante de la expresión génica. Los controladores transcripcionales (mejor conocidos como
20 promotores) son generalmente secuencias nucleotídicas cortas que se encuentran presentes en el borde 5' o la región río arriba de un gen transcrito. Existen secuencias promotoras que afectan la expresión génica en respuesta a estímulos ambientales, a la disponibilidad de nutrientes, o a condiciones adversas que incluyen altas temperaturas, anaerobiosis o la presencia de metales pesados en el
25 sustrato de crecimiento de las plantas. Existen también secuencias de ADN que controlan la expresión génica durante el desarrollo de un organismo, y esto de manera específica en ciertas células o tejidos.

30 Los promotores contienen las señales que requiere la enzima ARN polimerasa para iniciar la transcripción de manera a que se obtenga la síntesis proteica. Existen proteínas nucleares de unión al ADN que interactúan de manera específica con la secuencia del promotor para promover la formación de un complejo transcripcional

y eventualmente iniciar el proceso de expresión génica. La región que contiene todos los elementos necesarios para garantizar la producción de suficientes niveles de transcrito puede tener entre 100 pares de bases y hasta 1,000 pares de bases de longitud.

5

La existencia de promotores que confieren actividad en las células de las plantas ha sido reportada en la literatura, incluyendo los promotores de la nopalina sintasa (NOS) y la sintasa de octopina (OCS; ambos presentes en los plásmidos de *Agrobacterium tumefaciens*), los promotores 19S y 35S del virus del mosaico de la coliflor, el promotor inducible por luz de la sub-unidad pequeña de la ribulosa carboxilasa bifosfato (ssRUBISCO), y el promotor de la sucrosa sintasa. Todos estos promotores han sido utilizados para crear diferentes tipos de construcciones de ADN que han sido expresadas en plantas ¹.

10

15

Sin embargo, los promotores más valiosos son los que actúan de manera específica en ciertos tejidos o células. Este tipo de promotores puede activar secuencias de genes solo en tejidos específicos en los cuales la expresión de una proteína es deseada. Este tipo de promotores puede también ser expresado solo durante cierto período del desarrollo de la planta. Por ejemplo, el promotor E8 del jitomate solo está transcripcionalmente activo durante la etapa de maduración del fruto, y por lo tanto puede ser utilizado para activar la expresión génica durante el proceso de maduración del jitomate ². La actividad del promotor E8 no se limita a la fruta del jitomate, si no que se piensa que es compatible con cualquier proceso biológico que pueda ser activado por el etileno. Otros promotores que actúan durante el desarrollo de la fruta son el fragmento 1.45 ³ y el promotor de la poligalacturonasa de jitomate ⁴.

20

25

30

Tal vez los promotores específicos más importantes son aquellos que regulan la expresión génica en embriones y en semillas y en sus precursores celulares, las células reproductivas masculinas o femeninas contenidas en el gametofito masculino o femenino. Este tipo de promotores puede ser utilizado para modular el tamaño de la semilla, para producir semillas sin necesidad de fecundación (por

partenogénesis), para activar el desarrollo autónomo del embrión o del endospermo, o sencillamente para obtener información sobre cuando y en donde ocurre la regulación génica en la fase reproductiva del ciclo de vida de las plantas. Este tipo de promotores incluye secuencias reguladoras de genes que se expresan en el tejido materno como la cubierta de la semilla, el pericarpio o el óvulo. Uno de estos promotores es el promotor del gene que codifica la subunidad α' de la β -conglucina de soya⁵ que se expresa en el endospermo prematuro de la semilla. Algunos promotores adicionales incluyen el promotor del gen Napin⁶ para dirigir la expresión de un gen que aumente la producción de aceite en la semilla; el promotor DC3 que se expresa en la fase temprana y media del desarrollo del embrión⁷; el promotor del gen que codifica una faseolina⁸ y que solo se expresa en *Phaseolus vulgaris* durante el desarrollo de la semilla.

Como se desprende de los párrafos anteriores, existe en el campo de la biotecnología agrícola y la biología molecular de plantas una necesidad de promotores específicos que permitan dirigir la expresión de ácidos nucleicos en las células reproductivas contenidas en el gametofito masculino y el gametofito femenino de las plantas.

20 Descripción de las figuras.

- Figura 1. Se observa la secuencia del promotor pFM1.
- Figura 2. Se observa la secuencia del promotor pES1.
- Figura 3. Se observa la secuencia del promotor pPo1.
- Figura 4. Se observa la secuencia del promotor pPo2.
- 25 Figura 5. Se observa el fragmento de ADN aislado para la clonación del promotor pFM1.
- Figura 6. Se observan los patrones de actividad de cada uno de los promotores que se describen en este invento. Cada una de las secuencias promotoras de la invención fue fusionada al gen reportero *uidA* (GUS) y plantas de *Arabidopsis thaliana* fueron transformadas con dicha construcción. El tejido de las plantas transformantes fue analizado para detectar la presencia de la proteína reportera (azul). (a) – (c), actividad
- 30

del promotor pFM1 en el gametofito femenino y los gametos femeninos; el promotor pFM1 se activa al inicio de la gametogénesis femenina y se mantiene en el gametofito y los gametos femeninos (d), actividad del promotor pES1 en el gametofito femenino y los gametos femeninos. (e); (f), actividad del promotor pPo2 en el gametofito masculino y los gametos masculinos.

5

Figura 7. Se observa el fragmento de ADN aislado para la clonación del promotor pES1.

Figura 8. Se observa el fragmento de ADN aislado para la clonación del promotor pPo1.

10

Figura 9. Se observa el fragmento de ADN aislado para la clonación del promotor pPo2.

Figura 10. Se observa el efecto del promotor pFM1 en el desarrollo del gametofito femenino. El gametofito femenino detiene su desarrollo cuando se altera por la activación de una secuencia RNAi bajo control del promotor pFM1 que se describe en este invento.

15

Objetivos de la invención.

El objetivo principal de la invención que aquí se presenta es proporcionar elementos reguladores que dirijan la expresión de secuencias nucleotídicas en las células del gametofito femenino, de las células reproductivas femeninas (los gametos femeninos), el gametofito masculino y/o las células reproductivas masculinas (los gametos masculinos) de las plantas.

20

Otro objetivo de este invento es proporcionar unidades transcripcionales que permitan la expresión de genes durante la formación del gametofito femenino, de las células reproductivas femeninas (los gametos femeninos), del gametofito masculino y/o de las células reproductivas masculinas (los gametos masculinos) de las plantas.

25

30

Otro objetivo de este invento es proporcionar vehículos para la transformación de células de plantas, incluyendo vectores virales o plásmidos que incluyan las nuevas secuencias reguladoras que se describen en este invento.

- 5 Otro objetivo de este invento es proporcionar células bacterianas que contengan los vectores descritos en el párrafo anterior y que permitan el mantenimiento, la replicación y la transformación de plantas.

Otro objetivo de este invento es proporcionar construcciones de expresión, secuencias y células transformadas para la creación de plantas transgénicas.

- 10 Otro objetivo de este invento es proporcionar materiales genéticos que puedan ser utilizados en programas de mejoramiento que sirvan para producir cultivos comerciales a partir de genes exógenos deseables.

- 15 Otro objetivo de este invento es proporcionar secuencias que sirvan como etiquetas para el mapeo fino de genes de plantas de *Arabidopsis* así como para establecer ensayos de activación o inactivación de genes durante el desarrollo del gametofito masculino y femenino.

- 20 Estos y otros objetivos serán ampliamente descritos y justificados en los próximos párrafos.

Descripción de la invención.

- 25 El invento que aquí se describe incluye el aislamiento y la caracterización de nuevas secuencias nucleotídicas reguladoras que permiten la expresión de genes durante el desarrollo del gametofito masculino y femenino (incluyendo los gametos masculinos y los gametos femeninos) de *Arabidopsis* y de otras especies. El invento también incluye el control de la expresión espacial y temporal de secuencias nucleotídicas utilizando los elementos reguladores que aquí se describen.
- 30

El invento también incluye construcciones de expresión que incluyan uno o varios de los promotores descritos en este invento, un gen estructural cuya expresión sea deseable en células vegetales, y una señal de poliadenilación o de paro transcripcional. Este tipo de construcciones puede ser incluida en vectores plasmídicos o virales utilizados para la transformación de células de planta.

El invento también incluye células bacterianas transformadas para el mantenimiento y la replicación del vector arriba descrito, así como células vegetales transformadas y plantas transgénicas o todo germoplasma derivado a partir de estas.

A otro nivel, el invento también incluye la identificación y localización de nuevas secuencias nucleotídicas que se expresan o regulan la expresión que permite la activación o inactivación de genes durante el desarrollo del gametofito masculino y femenino. Dichas secuencias pueden ser utilizadas como marcadores de mapeo, para la manipulación de la expresión génica durante la gametogénesis masculina o femenina, y para ensayos y protocolos que pretendan silenciar genes en el gametofito masculino o femenino, o durante el desarrollo de la semilla.

I. Definiciones.

Para propósitos de esta solicitud los siguientes términos tendrán la definición que a continuación se detalla. Las unidades, los prefijos y los símbolos podrán ser mencionados utilizando su abreviatura aceptada por el Sistema Internacional (SI). A menos que sea explícitamente indicado, la secuencia de los ácidos nucleicos está escrita de izquierda a derecha en la orientación 5' a 3'; la secuencia de aminoácidos está escrita de izquierda a derecha en la orientación del borde amino al borde carboxi-terminal. Los rangos numéricos son incluyentes de los números que definen el rango e incluyen cada uno de los números íntegros que definen el rango. Los aminoácidos pueden ser designados ya sea por sus conocidos símbolos de tres letras o por sus símbolos de una sola letra que han sido recomendados por la Comisión de Nomenclatura de la IUPAC-IUB. De la misma manera, los nucleótidos pueden ser designados por su comúnmente aceptado código de letra única. Los

términos de software, eléctricos y electrónicos son utilizados como fueron definidos en el Nuevo Diccionario IEEE para Términos Eléctricos y Electrónicos⁹.

5 Para fines de esta solicitud de patente, por "gametofito femenino" se entiende el conjunto organizado de células reproductivas femeninas y sus células precursoras en las plantas. En la mayoría de las plantas, incluyendo *Arabidopsis thaliana*, las células del saco embrionario se diferencian a partir de cuatro esquemas de desarrollo distintos para formar ya sea las sinérgidas (2), la ovocélula (1), la célula central (1), o las antípodas (numero variable). La ovocélula y las sinérgidas están
10 organizadas en un triángulo localizado en un polo de la célula central. Las sinérgidas son células altamente especializadas que participan en la atracción del tubo polínico y el transporte de las células espermáticas hacia la célula huevo y la célula central. Por causas desconocidas una de las dos sinérgidas empieza a degenerar antes de la fecundación. La ovocélula se localiza junto a las sinérgidas, y
15 su núcleo se encuentra en la parte apical, junto a la membrana plasmática que la separa de la célula central. Su citoplasma está altamente vacuolizado y presenta una actividad metabólica inferior a la de las sinérgidas. A diferencia de las demás células del saco embrionario, la célula central contiene dos núcleos que se fusionan antes de la fecundación. El origen de estos dos núcleos se encuentra en los dos
20 polos del saco embrionario (micropilar y apical). La diferenciación de las tres antípodas tiene lugar en la región apical.

Para fines de esta solicitud de patente, por "célula reproductiva femenina" o "gameto femenino" se entiende cualquiera de estas células: las sinérgidas, la
25 ovocélula, la célula central o las células antípodas.

Para fines de esta solicitud de patente, por "gametofito masculino" se entiende el conjunto organizado de células reproductivas masculinas y sus células precursoras característico de las plantas. En la antera, las microspora madre antes de la primera
30 división meiótica. En las gramíneas, una pared celular se produce entre los productos de la meiosis I; la segunda división es perpendicular a la primera y las cuatro microsporas forman una tetradá. En muchas otras familias, la formación de

la pared celular solo inicia después de la meiosis II. Durante la maduración del grano de polen, el núcleo haploide en cada una de las microsporas se divide (por mitosis) de manera asimétrica para formar un núcleo vegetativo de tamaño predominante, y un núcleo denso y a menudo alargado (llamado el núcleo generativo), pero más pequeño. La diferenciación del núcleo generativo da lugar a la formación de una pequeña célula con membrana plásmica doble. La célula generativa contiene muy pocos organelos, incluyendo mitocondrias y plastos. Se considera, sin evidencias experimentales claras, que la ausencia de organelos se debe a una distribución asimétrica del citoplasma durante la división mitótica del núcleo haploide. En la mayoría de las especies, la célula generativa y su núcleo se dividen antes de la diseminación del polen para formar un grano trinucleado, en el cual existen dos células espermáticas y un núcleo vegetativo cuya función sigue siendo una gran incógnita pero motivo de abundante especulación. Se cree que actúa durante como control y guía del crecimiento del tubo polínico, y durante su crecimiento, se le encuentra a menudo en estrecha asociación con las membranas citoplásmicas de una de las células espermáticas. En la mayoría de las gramíneas (ej. maíz), la división mitótica de la célula generativa ocurre después de iniciada la germinación del grano de polen, dentro del tubo polínico.

Para fines de esta solicitud de patente, por "célula reproductiva masculina" o "gameto masculino" se entiende una célula vegetativa o una célula espermática contenida dentro del polen de las plantas.

Por "amplificación" se entiende la construcción de múltiples copias de una secuencia de ácido nucleico o múltiples copias complementarias a la secuencia de un ácido nucleico utilizando al menos una secuencia de ácido nucleico como marco de referencia inicial. Los sistemas de amplificación incluyen la reacción de la polimerasa en cadena (*polymerase chain reaction* – por sus siglas en inglés PCR), la reacción en cadena dependiente de ligasa (LCR), la amplificación dependiente de secuencia de ácido nucleico (NASBA, Canteen, Mississauga, Ontario), el sistema *Q-Beta Replicase*, el sistema de amplificación dependiente de transcripción (*Transcriptional-dependent Amplification System* –TAS, por sus siglas en inglés), y

el sistema de amplificación por desplazamiento de hebra (*Strand Displacement Amplification* –SDA, por sus siglas en inglés) ¹⁰. El producto de amplificación se denomina un amplicón.

5 La “orientación antisentido” hace referencia a una secuencia polinucleotídica de doble hebra que se encuentra funcionalmente ligada a un promotor en una orientación en la cual se transcribe la orientación antisentido de dicha molécula. La hebra antisentido es suficientemente complementaria a un producto endógeno de transcripción como para que la traducción del producto del transcrito endógeno sea
10 inhibida.

Una “región cromosómica” hace referencia a la longitud de un cromosoma que puede ser medida por referencia al segmento lineal de ADN que dicha región incluye. La región puede ser también definida a partir de dos secuencias únicas de
15 ADN o marcadores moleculares.

El término “variantes modificadas conservativamente” se aplica tanto a secuencias nucleotídicas como a secuencias de aminoácidos. En referencia a secuencias nucleotídicas específicas, variantes modificadas conservativamente se refiere a
20 aquellos ácidos nucleicos que codifican variantes idénticas o modificadas conservativamente de las secuencias de aminoácidos. A causa de la degeneración del código genético un número grande de ácidos nucleicos funcionalmente idénticos codifican cualquier tipo de proteína. Por ejemplo, los codones GCA, GCC, GCG y GCU codifican todos el aminoácido alanina. Estas variaciones del ácido nucleico
25 son variaciones silenciosas y representan variantes modificadas conservativamente. Por lo mismo, toda secuencia de ácido nucleico que codifique un polipéptido en este documento describe todas las variaciones silenciosas posibles del ácido nucleico. De la misma manera, toda variación silenciosa de una secuencia de ácido nucleico que codifique un polipéptido queda implícitamente
30 incluida en cada secuencia polipeptídica de este documento.

En cuanto a las secuencias de aminoácidos, se reconoce que sustituciones individuales, deleciones o adiciones a un ácido nucléico, a un péptido, a un polipéptido o a una secuencia proteica que altere, adicione o elimine un solo aminoácido o un porcentaje pequeño de aminoácidos en la secuencia codificada es también una “variante modificada conservativamente” en la cual la alteración resulta en la substitución de un aminoácido por un aminoácido bioquímicamente similar. Por lo tanto, cualquier número de residuos de aminoácidos seleccionados a partir del número de íntegros que van del 1 al 15 puede ser alterado de la manera descrita aquí arriba. Por ejemplo, 1, 2, 3, 4, 5, 7 o 10 alteraciones pueden ser creadas. Las variantes conservativamente modificadas tienen generalmente una actividad biológica similar a la de la secuencia polipeptídica no modificada a partir de la cual fueron derivadas. Por ejemplo, la especificidad del substrato, la actividad enzimática, o las propiedades de unión entre un receptor y su elicitador son generalmente del 30% al 90% de la actividad de la proteína nativa y de su substrato nativo. Las tablas de substituciones conservativas son ampliamente conocidas en el área.

Los siguientes 6 grupos contienen aminoácidos que representan substituciones conservativas dentro de cada uno de los grupos:

- 1) Alanina (A), Serina (S), Treonina (T).
- 2) Acido Aspártico (D), Acido Glutámico (E)
- 3) Aspargina (N), Glutamina (Q)
- 4) Arginina (R), Lisina (K)
- 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V)
- 6) Fenilalanina (F), Tirosina (Y), Triptofano (W)

El término “codificado” o “que codifica” en referencia a un ácido nucléico específico se refiere a la información de traducción de la proteína correspondiente. Un ácido nucléico que codifica una proteína puede incluir secuencias no traducidas (intrones, por ejemplo) contenidas dentro de regiones traducidas del ácido nucléico, o puede no incluir este tipo de secuencias (como en un cDNA, por ejemplo). La información de la proteína codificada está especificada en el uso de codones. De manera

general se utiliza el código genético universal para determinar el código de traducción de una secuencia de ácido nucleico en secuencia de aminoácidos. Sin embargo, existen variantes del código universal en la información genética contenida dentro de las mitocondrias de algunas plantas, animales u hongos, del organismo bacterioide *Mycoplasma capricolum*, el organismo ciliado *Macronucleus* pueden ser incluidas dentro del organismo correspondiente.

Cuando el ácido nucleico es generado o alterado de manera sintética, se puede aprovechar las preferencias de codones del organismo hospedero en el cual se pretende expresar el ácido nucleico. Por ejemplo, aunque las secuencias nucleotídicas del invento que aquí se describe pueden ser expresadas tanto en especies de plantas monocotiledóneas como dicotiledóneas, las secuencias pueden ser modificadas tomando en cuenta las preferencias de codones y la preferencia de contenido GC que se ha demostrado difieren en monocotiledóneas y dicotiledóneas¹¹. Por lo tanto, el codón predominante en maíz para un aminoácido específico puede ser derivado de la secuencia conocida de un gen de maíz.

El término "secuencia de longitud completa" (del inglés "full-length sequence") de un polinucleótido específico o su proteína codificada hace referencia a la secuencia completa de la cadena de aminoácidos de una proteína nativa (no sintética), endógena y en su forma biológicamente activa. Los métodos que se utilizan para determinar si una secuencia es de longitud completa son ampliamente conocidos y como ejemplo podemos mencionar hibridaciones de tipo "Northern" o "Western", extensión de primers, o protección de ribonucleasa¹². La comparación con secuencias de longitud completa homologas (ortólogas o parálogas) puede ser también utilizada para identificar la secuencia de longitud completa de secuencias incluidas en el invento que aquí se describe. Las secuencias consenso generalmente presentes en el borde 3' o 5' de las regiones no traducidas de una molécula de RNA mensajero (mRNA) ayudan a la identificación de la secuencia de longitud completa de un polinucleótido. Por ejemplo, la secuencia consenso ANNNNAUGG, en la cual el codón subrayado representa la metionina presente en el borde N-terminal ayuda a determinar si el polinucleótido tiene un borde 5'

terminal completo. Las secuencias consenso en el borde 3' tales como las secuencias de poliadenilación, ayudan a determinar si el borde 3' terminal está completo.

5 El término "heterólogo" se refiere aquí a un ácido nucleico que deriva de una especie diferente o, en caso de que derive de la misma especie, un ácido nucleico que se encuentra sustancialmente modificado de su forma nativa. Por ejemplo, un promotor que se encuentre operativamente ligado a un gen estructural heterólogo pertenece a una especie diferente a partir de la cual el gen estructural fue
10 originalmente obtenido siempre y cuando se origine de una intervención humana deliberada. En caso de pertenecer a la misma especie, uno a varios genes heterólogos deben estar sustancialmente modificados de su forma original. Una proteína heteróloga puede originarse a partir de una especie distinta, o de la misma especie siempre y cuando se origine a partir de una intervención humana
15 deliberada.

El término "célula huésped" hace referencia a una célula que contiene un vector y que asegura la replicación y/o la expresión de dicho vector. Las células huésped pueden ser células procarióticas como las de *e. Coli*, o eucarióticas como las de levadura,
20 insecto, anfibio o células de mamífero. Preferentemente, las células huésped son de plantas monocotiledóneas o dicotiledóneas. Una célula huésped preferenciada en monocotiledóneas es la célula de maíz.

El término "complejo de hibridación" hace referencia a una estructura doble hebra de ácido nucleico formado por dos secuencias de ácidos nucleicos de hebra
25 sencilla hibridados conjuntamente de manera selectiva.

El término "introducido" en referencia al acto de insertar un ácido nucleico en una célula significa "transfectar" o "transformar" e incluye la incorporación de ácidos
30 nucleicos en una célula eucariota o procariota en la cual el ácido nucleico puede ser incorporado en el genoma de la célula (en el ADN de un cromosoma, de un

plásmido, un plástido o una mitocondria), o puede ser convertido en un replicón autónomo, o expresado de manera transiente.

El término "aislado" se refiere a un material (ácido nucléico o una proteína) que se encuentra:

(1) sustancialmente o completamente libre de los componentes que normalmente lo acompañan o interactúan con él en su forma natural. El material aislado puede opcionalmente comprender otro material que no se encuentra asociado con el aislado en su forma natural; o bien

(2) en caso de que el material se encuentre en su medio natural, si el material ha sido alterado de manera sintética por una intervención humana deliberada que modifica su composición o lo asigna a un lugar específico de la célula (por ejemplo, un organelo) diferente al lugar en el que se encuentra en su entorno natural. La alteración que da lugar a la forma sintética del material puede ser dirigida al material (ácido nucléico y/o proteína) o depender de una remoción del entorno natural. Por ejemplo, un ácido nucléico natural puede ser aislado si se le altera o si se le transcribe a partir de ADN que ha sido previamente alterado a partir de una intervención humana deliberada realizada en la célula de la cual el ácido nucléico se originó¹³. De la misma manera, un ácido nucléico natural (por ejemplo, un promotor) se encuentra aislado si se le introduce por medios no naturales en un locus del genoma que no le es nativo. Los ácidos nucléicos que son "aislados" siguiendo esta definición pueden ser denominados como ácidos nucléicos "heterólogos".

El término "localizado en una región cromosómica definida por" en referencia a marcadores moleculares específicos hace referencia a un segmento cromosómico delimitado por los marcadores correspondientes, e incluye dichos marcadores.

El término "marcador" hace referencia a un locus en un cromosoma que sirve para identificar una posición única en dicho cromosoma. Un "marcador polimórfico" hace referencia a un marcador que puede tener varias formas (alelos) de manera que las

diferentes formas del marcador, cuando están presentes en un par de cromosomas homólogos, permiten que se le pueda dar seguimiento a la transmisión de cada uno de los cromosomas del par. Un genotipo puede ser definido por uno o varios marcadores polimórficos o no.

5

El término "ácido nucleico" o "nucleótido" hace referencia a un polímero de desoxirribonucleótidos o ribonucleótidos en su forma de hebra sencilla o doble hebra y comprende aquellas moléculas análogas que tienen la naturaleza esencial de los nucleótidos naturales de poder hibridarse a ácidos nucleicos de hebra sencilla de una manera similar a la de los nucleótidos naturales.

10

El término "biblioteca de ácidos nucleicos" se refiere a una colección de moléculas de ARN o ADN que comprenden y representan de manera sustancial la integridad de la fracción transcrita del genoma de un organismo específico. Ejemplos de construcciones de bibliotecas, ya sean bibliotecas genómicas o bibliotecas de cDNA, son conocidas en la literatura¹⁴.

15

El término "operativamente ligado" hace referencia a un ligamiento funcional entre un promotor y una segunda secuencia, en la cual la secuencia del promotor inicia y permite la transcripción de la secuencia de ADN correspondiente a la segunda secuencia. De manera general, operativamente ligado significa que las secuencias ligadas son contiguas, y en el caso de que sea necesario juntar dos secuencias codificando las proteínas correspondientes, se trata de secuencias contiguas y en el mismo marco de lectura.

20

25

En este documento, el término "planta" puede hacer referencia a organismos vegetales completos, partes de plantas u órganos (por ejemplo, hojas, tallos, raíces, óvulos etc.), células de planta, o semillas y crías de planta. Células de planta en este caso incluyen sin limitación células obtenidas a partir de o presentes en: semillas, cultivos en suspensión, embriones, regiones meristemáticas, tejido de callo, hojas, raíces, botones, meristemas, gametofitos, gametos, esporofitos, polen y microsporas. Células de plantas también puede incluir células modificadas como

30

protoplastos obtenidos a partir de los tejidos arriba mencionados. El tipo de plantas que pueden ser utilizado para el presente invento es generalmente tan amplio como el grupo de Angiospermas que pueden ser modificadas por técnicas de transformación tanto en dicotiledóneas como en monocotiledóneas. Plantas particularmente importantes son el maíz, la soya, el girasol, el sorgo, la canola, el trigo, la alfalfa, el algodón, el tomate, la papa, el arroz, la cebada, el mijo y Arabidopsis.

El término "polinucleótido" hace referencia a un desoxirribonucleótido, un ribopolinucleótido, o sus análogos que tengan las propiedades esenciales de un ribonucleótido natural, como es el hecho de que hibriden, bajo condiciones de hibridación astringente, a esencialmente las mismas secuencias nucleotídicas que los nucleótidos que ocurren de manera natural, o que permitan la traducción en los mismos aminoácidos que los nucleótidos naturales. Un polinucleótido puede ser de longitud completa o un segmento de secuencia de un gen estructural o un gen regulatorio nativo o heterólogo. De no ser que se mencione explícitamente, el término incluye la secuencia especificada así como su secuencia complementaria. Por ello, moléculas de ADN o de ARN que contengan segmentos que han sido modificados para aumentar su estabilidad o por otras razones son también "polinucleótidos" para efectos del presente invento. Adicionalmente, moléculas de ADN o de ARN que incluyan bases nucleotídicas poco frecuentes, como inosina, o bases nucleotídicas modificadas, como las bases tritiladas por citar solo dos ejemplos, serán también consideradas "polinucleótidos" para efectos del presente invento. Se desprende de este párrafo que una gran variedad de modificaciones han sido realizadas a moléculas de ADN o de ARN, y que dichas modificaciones sirven para múltiples propósitos. El término polinucleótido se utiliza aquí para designar formas de polinucleótidos modificadas químicamente, enzimáticamente o metabólicamente, así como las formas químicas del ADN y del ARN características de los virus y de las células sencillas o complejas.

30

Los términos "polipéptido", "péptido", y "proteína" se utilizan aquí de manera intercambiable para designar a un polímero de aminoácidos. Los términos hacen

referencia a polímeros de aminoácidos en el cual uno o mas de los aminoácidos es un análogo de un aminoácido natural correspondiente. La naturaleza esencial de dichos análogos es que cuando están incorporados en una proteína, dicha proteína puede ser reconocida de manera específica por anticuerpos diseñados para reconocer esa misma proteína cuando está compuesta exclusivamente de aminoácidos naturales. Los términos "polipéptido", "péptido", y "proteína" incluyen también, sin limitaciones, la glicosilación, la unión con lípidos, la sulfatación, la gama carboxilación de residuos que contengan ácido glutámico, la hidroxilación y la ribosilación de ADP. Es importante mencionar que los polipéptidos no son completamente lineales. Por ejemplo, los polipéptidos pueden estar ramificados como resultado de la ubiquitinación, pueden ser circulares, con o sin ramificaciones como resultado de eventos post-traduccionales tales como procesos naturales y eventos causados por la manipulación humana que no ocurren de manera natural. Polipéptidos circulares y/o ramificados pueden ser sintetizados por procesos naturales que no dependen de la traducción y por procesos completamente sintéticos. Adicionalmente, este invento incluye también las variantes terminales de aminoácidos con o sin metionina de cada una de las proteínas referentes al presente invento.

El término "promotor" hace referencia a una región de ADN localizada río arriba del codón de inicio de la transcripción y que está involucrada en el reconocimiento y la unión de una RNA polimerasa y de otras proteínas necesarias para el inicio de la transcripción. Un "promotor de plantas" es una secuencia derivada de plantas o de otros organismos pero capaz de iniciar la transcripción en células de plantas. Algunos ejemplos de promotores de plantas incluyen aquellos obtenidos a partir del ADN de plantas, de virus de plantas y de bacterias que contienen genes que se expresan en plantas como *Agrobacterium* o *Rhizobium*. Ejemplos de promotores que están bajo control de estados de desarrollo incluyen promotores que de manera preferencial inician la transcripción en ciertos tejidos, tales como hojas, raíces o semillas. Este tipo de promotores es denominado "tejido-preferentes". Aquellos promotores que solo inician la transcripción en ciertos tejidos se denominan "promotores específicos". Un promotor específico de "tipo celular" es un promotor

que solo dirige la expresión en ciertos tipos de células localizada en uno o varios órganos, como por ejemplo las células de la vasculatura en hojas y raíces. Un promotor "inducible" o "reprimible" es un promotor que responde a señales de control que emite del medio ambiente. Algunos ejemplos de condiciones ambientales que pueden tener un efecto sobre la transcripción dependiente de promotores inducibles son la presencia de condiciones anaeróbicas, o el efecto de la luz. Los promotores tejido-específicos, los promotores tejido-preferentes, los promotores específicos de tipo celular y los promotores inducibles constituyen la clase de promotores "no constitutivos" Los promotores "constitutivos" son aquellos que dirigen la expresión de manera sistémica y bajo la mayoría de las condiciones ambientales.

El término "recombinante" hace referencia a una célula o a un vector que ha sido modificado por la introducción de ácido nucléico o que una célula es derivada de otra célula que fue modificada por dicha introducción. Por ejemplo, células recombinantes expresan genes que no se encuentran en forma idéntica dentro de la forma nativa (no recombinante) de la célula o que expresan genes nativos que de otra manera se sub-expresarían, se expresarían de manera anormal, no se expresarían para nada como resultado de la intervención humana. El término recombinante no incluye la alteración de la célula o del vector por eventos ocurridos en forma natural (por ejemplo, mutación espontánea, o transformación, transducción o transposición) como todos aquellos que ocurren sin la intervención humana.

El término "caset de expresión" o "construcción de expresión" se refiere a una construcción de ácido nucléico (generada en forma recombinante o sintética) que contiene una serie de elementos de ácidos nucléicos que permiten la transcripción de un ácido nucléico particular en una célula huésped. El caset de expresión recombinante puede ser incorporado en un plásmido, en un cromosoma, en ADN mitocondrial, en ADN plátídico, en un virus, o en un fragmento de ácido nucléico. Típicamente, la porción del caset de expresión recombinante de un vector de

expresión incluye, entre otras secuencias, el ácido nucléico que va a ser transcrito y un promotor.

5 Los términos "residuo" o "residuo de aminoácido" se usan de manera intercambiable para referirse a un aminoácido que está incorporado en una proteína, un polipéptido o un péptido. El aminoácido puede ser natural o puede incluir aminoácidos sintéticos análogos de aminoácidos naturales que puedan funcionar de manera similar a los aminoácidos naturales.

10 El término "hibrida selectivamente" hace referencia a la hibridación (en condiciones de hibridación astringentes) de una secuencia de ácido nucléico a una secuencia blanco específica de ácido nucléico que pueda ser distinguida por detección de su hibridación a una secuencia de ácido nucléico no-blanco, lo que sirve para excluir de manera sustancial ácidos nucléicos no blanco. Las secuencias que hibridan selectivamente típicamente tienen 90% de identidad de secuencia compartida, y
15 preferentemente 100% de identidad de secuencia compartida.

El término "condiciones astringentes" o "condiciones astringentes de hibridación" hace referencia a condiciones en las cuales una sonda hibrida con su secuencia blanco a un nivel superior que la de otras secuencias (es decir, al menos 2 veces superior al fondo). Las condiciones astringentes son dependientes de la naturaleza de la secuencia y pueden variar en función de las circunstancias. Controlando las condiciones de astringencia y de lavado, secuencias blanco con 100% de homología a la sonda pueden ser identificadas. Alternativamente, las condiciones
20 de astringencia pueden ser ajustadas para permitir ciertos apareamientos no homólogos en las secuencias de manera a que niveles de homología inferior puedan ser detectados. Generalmente una sonda tiene menos de 1000 nucleótidos de longitud y opcionalmente menos de 500 nucleótidos.

30 Las condiciones de astringencia generalmente serán aquellas en las que la concentración de sal sea inferior a 1.0 M de iones Na, y típicamente comprendidas entre 0.01 y 1.0 M de concentración de iones Na (u otras sales) a un pH de 7.0 a

8.3 y una temperatura de al menos 30°C para sondas cortas (de 10 a 50 nucleótidos) y de al menos 60°C para sondas largas (mayores que 550 nucleótidos). Las condiciones astringentes pueden también ser obtenidas a partir de agentes desestabilizantes como la formamida. Ejemplos de condiciones de hibridación astringentes incluyen la hibridación con una solución reguladora de 30 a 35% de formamida, 1 M NaCl, 1% de SDS (dodecil sulfato de sodio) a 37°C, y un lavado en 1X a 2X SSC (20X SSC= 3.0 M NaCl/0.3 M citrato trisódico) a una temperatura de 50 a 55°C. Ejemplos de condiciones moderadas de astringencia incluyen hibridaciones en 40-45% formamida, 1 M NaCl, 1% SDS a 37°C, y un lavado en 0.1X SSC a una temperatura de 60 a 65°C.

La especificidad está normalmente determinada por los lavados después de la hibridación, y los factores críticos son la fuerza iónica y la temperatura de la solución de lavado final. Para híbridos ADN-ADN, el valor de T_m puede ser aproximado a partir de la ecuación de Meinkoth y Wahl¹⁵: $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ formamida}) - 500/L$; en donde M es la molaridad de los cationes monovalentes, %GC es el porcentaje de los nucleótidos guanosina y citosina en el ADN, %formamida es el porcentaje de formamida en la solución de hibridación, y L es la longitud del híbrido en pares de bases. El T_m es la temperatura (bajo condiciones determinadas de fuerza iónica y pH) a la cual el 50% de la secuencia blanco hibrida a una sonda que le es perfectamente homóloga. El T_m se reduce de 1°C por cada 1% de disparidad de apareamiento entre la sonda y su blanco; por lo tanto, las condiciones de hibridación o de lavado pueden ser ajustadas para hibridar con secuencias deseadas. Por ejemplo, si se busca detectar secuencias que tienen solo 90% de homología con la sonda, el valor de T_m puede ser disminuido en 10°C. Generalmente las condiciones de astringencia pueden ser seleccionadas para ser 5°C inferiores que el valor de T_m correspondiente a la secuencia específica y su complemento bajo condiciones determinadas de pH y de fuerza iónica. Sin embargo, condiciones de astringencia severa pueden utilizar una hibridación o lavados de 1 a 4°C inferiores al T_m ; condiciones de astringencia moderada pueden utilizar hibridaciones o lavados de 11 a 20°C inferiores al T_m .

Utilizando la ecuación arriba mencionada las condiciones de hibridación, la composición de las soluciones de lavado y el T_m deseado, aquellos que han sido rutinariamente entrenados en técnicas de biología molecular podrán entender que las variaciones en las condiciones de astringencia de una hibridación o de las soluciones de lavado están claramente descritas. Si el grado aceptable de apareamiento erróneo entre las dos secuencias resulta en un valor de T_m inferior a 45°C (solución acuosa) o 32°C (solución de formamida) es preferente incrementar la concentración de SSC de manera a que un valor de temperatura superior pueda ser utilizado durante la hibridación y/o los lavados. Una guía extensa de las técnicas de hibridación de ácidos nucleicos puede ser encontrada en la literatura¹⁶.

El término "planta transgénica" hace referencia a una planta que contiene en su genoma un polinucleótido heterólogo. Generalmente dicho polinucleótido heterólogo está integrado de manera estable y se transmite a la descendencia de dicha planta. El polinucleótido heterólogo puede estar integrado en el genoma en forma aislada o como parte de un vector de recombinación. El término "transgénico" incluye aquí cualquier célula, línea celular, callo, tejido, parte de la planta o planta, para la cual el genotipo ha sido alterado por la presencia de un ácido nucleico heterólogo, incluyendo aquellos elementos transgénicos que fueron creados por transformación como aquellos creados por cruza sexuales o propagación asexual a partir del transgénico inicial. El término "transgénico" no hace referencia aquí a la alteración del genoma (cromosomal o extra-cromosomal) por métodos convencionales de mejoramiento genético o por eventos naturales tales como cruza aleatorias, la infección viral no recombinante, la transformación bacteriana no recombinante, la transposición no recombinante o la mutación espontánea.

El término "vector" hace referencia a un ácido nucleico utilizado para la transfección de una célula huésped en la cual puede ser insertado un polinucleótido. Los vectores son a menudo replicones. Los vectores de expresión permiten la transcripción de un ácido nucleico que les ha sido insertado.

Los términos que se indican a continuación son utilizados para describir la relación que tienen las secuencias de dos o más ácidos nucleicos o polinucleótidos: (a) "secuencia de referencia", (b) "intervalo de comparación", (c) "identidad de la secuencia". (d) porcentaje de identidad de la secuencia" y (e) "identidad sustancial".

5

(a) El término "secuencia de referencia" es una secuencia definida que se utiliza como base para la comparación de secuencias. Una secuencia de referencia puede ser una porción o una secuencia completa; por ejemplo, un segmento de un cDNA de longitud completa, o bien la longitud completa del cDNA o un gen entero.

10

15

(b) El término "intervalo de comparación" hace referencia a un segmento específico y contiguo de una secuencia polinucleotídica que puede ser comparado a una secuencia de referencia y para el cual la porción comparativa de la secuencia polinucleotídica puede incluir elementos adicionales o ausentes (por ejemplo, "gaps") al ser comparada con la secuencia de referencia (que no tiene elementos adicionales o ausentes) de manera a que el alineamiento de las dos secuencias sea óptimo. Por lo general el intervalo de comparación es de al menos 20 nucleótidos contiguos y a menudo puede incluir más de 100 nucleótidos. Aquellos que conocen como realizar este tipo de análisis entienden que para evitar una similitud elevada con la secuencia de referencia a causa de "gaps" en la secuencia polinucleotídica, valores de penalización por el "gap" pueden ser utilizados. Se conocen bien varios métodos de alineamiento de secuencias. El alineamiento óptimo de secuencias para comparación puede ser obtenido a partir del algoritmo de homología local de Smith y Waterman¹⁷, por el algoritmo de Needleman y Wunsch¹⁸, por el método de búsqueda de similaridad de Pearson y Lipman¹⁹, por la implementación computarizada de estos algoritmos, incluyendo pero no limitada a: CLUSTAL en el programa PC/Gene de Intelligenetics, Mountain View, California; GAP, BESTFIT, BLAST, FASTA y TFASTA en el paquete denominado Wisconsin Genetics Software, Genetics Computer Group (GCG), 575 Science Dr., Madison

20

25

30

Wisconsin, USA; el programa CLUSTAL de Higgins y Sharp²⁰. La familia de programas BLAST que pueden ser utilizados para búsquedas comparativas de secuencias incluye: BLASTN para búsqueda comparativa de secuencias nucleotídicas comparadas a las secuencias nucleotídicas contenidas en bases de datos del dominio público; BLASTP para búsqueda comparativa de secuencias proteicas comparadas a las secuencias proteicas contenidas en bases de datos del dominio público; TBLASTN para búsqueda de homologías entre una secuencia proteica y una secuencia nucleotídica; y TBLASTX para una comparación de una secuencia nucleotídica con un conjunto de bases de datos de secuencias nucleotídicas²¹.

A menos que se indique otra cosa de manera explícita, los valores de identidad o similitud de una secuencia que se indiquen en este documento se refieren a los valores obtenidos con la versión BLAST 2.0 usando los parámetros "default"²². El software para realizar estos análisis es del dominio público y se puede acceder a él u obtener a través del sitio de internet del Centro Nacional de Información Biotecnológica de los Estados-Unidos de Norteamérica. El algoritmo empieza por identificar pares de secuencias con alto grado de similitud a partir de la identificación de palabras cortas con una longitud de W en la secuencia de búsqueda; dichas palabras deben ser idénticas o muy similares a la de un umbral de valor T cuando se alinean con una palabra de la misma longitud contenida en una base de datos del dominio público. Esta similitud inicial entre dos secuencias da lugar al inicio de una serie de búsquedas para encontrar secuencias de mayor longitud y alto grado de similitud. A T se le denomina el umbral de valor de vecindad de palabra²³. Las palabras encontradas se extienden en ambas direcciones para cada una de las secuencias siempre y cuando el valor de alineamiento acumulativo siga creciendo. Para secuencias nucleotídicas los valores acumulativos se calculan usando los parámetros M (valor de premio por un par de residuos coincidentes; siempre será superior a 0) y N (valor de castigo para residuos no coincidentes; siempre será inferior a 0). Para secuencias de aminoácidos, se utiliza un valor de matriz para calcular el valor acumulativo.

La extensión de las palabras encontradas se detiene en cada una de las dos direcciones cuando: a) el valor de alineamiento acumulativo decae por la cantidad X inferior al máximo valor obtenido; b) el valor acumulativo llega a ser inferior o igual a cero debido a la acumulación de uno o más residuos de alineamiento con valor negativo; o c) el final de cada secuencia es alcanzado. Los parámetros W, T y X del algoritmo BLAST determinan la sensibilidad y la velocidad del alineamiento de secuencias. El programa BLASTN (para secuencias nucleotídicas) utiliza como valores "default" una longitud de palabra W de 11, un valor de espectancia E de 10, un umbral de corte de 100, M=5, N=-4, y efectúa la comparación de ambas hebras de secuencias nucleotídicas²⁴.

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también efectúa un análisis estadístico de la similitud entre dos secuencias²⁵. Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad mínima de suma (P(N)), que proporciona una indicación de la probabilidad con la cual la concordancia entre dos secuencias nucleotídicas o de aminoácidos pueda ocurrir al azar.

Las búsquedas BLAST asumen que las proteínas pueden ser modeladas como secuencias aleatorias. Sin embargo, muchas proteínas reales comprenden regiones de secuencias no aleatorias que pueden estar enriquecidas con uno o más amino-ácidos. Esas regiones de complejidad baja pueden ser alineadas a partir de proteínas no relacionadas aunque otras regiones de dichas proteínas sean completamente diferentes y no presenten similitud. Algunos programas que funcionan como filtros de baja complejidad pueden ser utilizados para reducir ese tipo de alineamientos. Por ejemplo, el programa SEG²⁶ y XNU²⁷. Estos tipos de filtros pueden ser utilizados de manera individual o en combinaciones.

30

(c) Los términos "identidad de secuencia" o "identidad" en el contexto de dos secuencias de ácidos nucleicos o de polipéptidos hace referencia a los

residuos en ambas secuencias que son los mismos cuando se alinean para buscar la mayor correspondencia dentro de una ventana comparativa específica. Cuando el porcentaje de la identidad de secuencia se usa en referencia a proteínas se reconoce que las posiciones de los residuos que no son idénticos a menudo difieren por sustituciones de aminoácidos conservativos, en los cuales los aminoácidos se substituyen por otros aminoácidos que tienen propiedades bioquímicas similares (carga o hidrofobicidad) y por lo tanto no cambia las propiedades funcionales de la molécula. Si las secuencias difieren en por la naturaleza conservativa de la sustitución, el porcentaje de identidad de secuencia puede ser ajustado hacia arriba para corregir el efecto de la naturaleza conservativa de la sustitución. Se dice que aquellas secuencias que difieren en sustituciones conservativas tienen "similaridad de secuencia" o "similaridad". La tarea de realizar este tipo de ajustes es rutinaria para aquellos con destrezas en el arte. Generalmente involucra valorar una sustitución conservativa como una discrepancia de secuencias parcial y no completa, incrementando así el porcentaje de identidad de secuencia. Por ejemplo, cuando a un aminoácido idéntico se le asigna un valor de 1 y a una sustitución no conservativa se le asigna un valor de cero, a una sustitución conservativa se le asigna un valor comprendido entre cero y 1. El valor de las sustituciones se calcula utilizando el algoritmo de Meyers y Miller²⁸, tal y como fue implementado en el programa PC/GENE (Intellegentics, Mountain View, California, USA).

(d) El término "porcentaje de secuencia de identidad" hace referencia al valor determinado a partir de la comparación de dos secuencias óptimamente alineadas a partir de una ventana comparativa específica, y en la cual la porción de secuencia polinucleotídica en la ventana comparativa puede incluir adiciones o deleciones (es decir, ausencias) cuando se le compara a la secuencia de referencia (que no incluye adiciones o deleciones) para el alineamiento óptimo de dos secuencias. El porcentaje se calcula determinando el número de posiciones a partir del cual la base nucleotídica o el residuo de amino-ácidos aparece en ambas secuencias para obtener el

número de posiciones coincidentes, dividiendo dicho número por el número total de posiciones en la ventana comparativa y multiplicando por 100 para obtener el porcentaje de secuencia de identidad.

- 5 (e) El término "identidad substancial" de secuencias polinucleotídicas significa que un polinucleótido comprende una secuencia que tiene al menos 70%, preferentemente al menos 80%, aún mas preferentemente 90% e idealmente 95% de identidad de secuencia cuando se le compara con una secuencia de referencia utilizando uno de los programas de alineamiento antes
- 10 mencionados. Se reconoce que estos valores pueden ser apropiadamente ajustados para determinar la identidad correspondiente de proteínas codificadas por 2 secuencias nucleotídicas y tomando en cuenta la degeneración de codones o la similitud de aminoácidos. Para este propósito la identidad substancial de secuencias de aminoácidos significa normalmente
- 15 una identidad de al menos 60%, o preferentemente 70%, 80%, 90% e idealmente 95%.

Otra indicación de que secuencias nucleotídicas son substancialmente idénticas es el hecho de que dos moléculas hibriden una con la otra bajo

20 condiciones astringentes de hibridación. Sin embargo, los ácidos nucleicos que no hibridan el uno con el otro bajo condiciones astringentes siguen siendo substancialmente idénticos si los polipéptidos que codifican son substancialmente idénticos. Esto puede ocurrir cuando, por ejemplo, una copia del ácido nucleico es creada usando el máximo codón de

25 degeneración permitido por el código genético. Una indicación de que dos secuencias de ácidos nucleicos son substancialmente idénticas es que el polipéptido codificado en el primer ácido nucleico sea inmunológicamente idéntico al polipéptido codificado en el segundo ácido nucleico.

30 El término "identidad substancial" en referencia a péptidos indica que un péptido comprende una secuencia con por lo menos 70%, preferentemente

80% o 85%, aún más preferentemente 90%, e idealmente 95% de identidad de secuencia con una secuencia de referencia en una ventana comparativa específica. De manera opcional, el alineamiento óptimo es efectuado utilizando el algoritmo de alineamiento de Needleman y Wunsch²⁹. Una indicación de que dos secuencias de péptidos son substancialmente idénticas es que un péptido sea inmunológicamente idéntico al segundo péptido. Por ejemplo, dos péptidos serán substancialmente idénticos si solo difieren por una substitución conservativa.

10 **Descripción detallada.**

En esta invención se muestra que los solicitantes han identificado promotores aislados a partir del DNA de Arabidopsis con capacidad para regular la expresión génica en plantas de manera que un gen seleccionado pueda ser eficientemente transcrito en las células reproductivas femeninas o masculinas de las plantas, y/o en el gametofito femenino o masculino (el grano de polen) en un momento y en un lugar preciso, y en la cantidad necesaria para obtener el efecto deseado. Promotores que dirigen la expresión temporal y espacial en el óvulo, en el gametofito femenino, en el polen, o en las células espermáticas son particularmente importantes y pueden ser usados con una gran variedad de protocolos transgénicos, para alterar los mecanismos reproductivos e inducir eventualmente apomixis, o formación de semillas asexuales genéticamente idénticas a la planta madre. Otra aplicación importante de estos promotores es la posibilidad de usarlos para implementar tecnologías que impidan la transferencia de rasgos transgénicos a través del polen en campo. También pueden ser utilizados para proporcionar nutrientes adicionales a la semilla y para conferir resistencia a patógenos y enfermedades en la semilla, o para proporcionarle otros rasgos de alto valor agregado, por citar algunos ejemplos.

La invención se obtuvo generando plantas transgénicas que contienen elementos "trampa génica" (gene trap) o "detector de potenciadores" (enhancer trap)³⁰. A partir de estas tecnologías se identificaron 2 regiones en el DNA genómico de Arabidopsis thaliana que confieren actividad transcripcional en tejido reproductivo

diploide o haploide de esta planta. En un caso la actividad se restringe a los tejidos femeninos (el gametofito femenino y los gametos femeninos) y en el otro al tejido masculino (el grano de polen y los gametos masculinos). Utilizando esta información y técnicas de biología molecular rutinarias ampliamente conocidas en el

5 área, la actividad específica de estas regiones reguladoras fue confirmada en fusiones de dichas regiones reguladoras con el gen reportero *uidA* (GUS) que codifica una beta-glucoronidasa en *Escherichia coli*. Estas secuencias pueden también ser utilizadas como marcadores moleculares para identificar otros genes u otras regiones regulatorias así como marcadores para mapeo molecular.

10

En general promotores específicos como los que se describen para el óvulo, el gametofito femenino, el grano de polen o los gametos masculinos pueden ser aislados utilizando procedimientos ya conocidos. Estas secuencias de control transcripcional están generalmente asociadas con genes que se activan

15 únicamente en los órganos, tejidos y tiempos deseados. En una planta normal, cada órgano o tejido contiene miles de RNA mensajeros (mRNAs) que están ausentes en otros tejidos u órganos³¹. Los mRNAs se aíslan para obtener sondas que se utilizan para posteriormente para obtener la secuencia genómica apropiada que contiene las secuencias reguladoras correspondientes. En éste sentido, un

20 ejemplo de la utilización de cDNA asociado con mRNA específico se ha reportado en la fruta del aguacate³². En breve, en este estudio el mRNA se aisló del fruto del aguacate durante su proceso de maduración y fue utilizado para construir una biblioteca de cDNA. Las clonas deseadas de la biblioteca fueron identificadas a partir de su hibridación con RNA radiactivamente marcado y aislado a partir de

25 frutos maduros; dichas sondas no hibridaron con RNA radiactivamente marcado y aislado a partir de frutos no maduros. Muchas de las clonas obtenidas representaron mRNAs codificados por genes que están transcripcionalmente activos solamente en frutos maduros.

30

Otro método importante para identificar promotores específicos y que fue utilizado para obtener la invención que aquí se presenta, y que permite identificar promotores activos en células únicas es el método de trampas génicas (gene trap) y

detector de potenciadores (enhancer traps)³³. Este método fue inicialmente desarrollado en *Drosophila* y rápidamente adaptado para ratones y para plantas³⁴.

5 En este tipo de estrategias se utilizan elementos transponibles (o transposones) que se modifican para servir como trampas génicas o detector de potenciadores que se integran de manera aleatoria en el genoma de las plantas para generar individuos transgénicos que son posteriormente analizados para identificar genes que se localizan en la vecindad inmediata del sitio de inserción. A partir de la identificación del momento y del sitio (órgano, tejido o célula) en donde se activa la expresión del gene reportero presente en el transposón, se identifica rutinariamente la secuencia aledaña a la inserción por medio de TAIL-PCR³⁵. A partir de esta secuencia se identifica el gen etiquetado y su promotor correspondiente. En algunas líneas el elemento detector de potenciadores puede haberse insertado en la región codificante del gen etiquetado. En otras el elemento puede estar directamente insertado en el promotor del gen correspondiente. Los promotores que dirigen la actividad del gen en óvulo, el gametofito femenino, el grano de polen o las células espermáticas pueden ser aislados y evaluados utilizando técnicas rutinarias de biología molecular como las que aquí se describen. Un ejemplo del método inicia con la identificación de unidades transcripcionales que rodean el sitio de inserción del transposón, el aislamiento subsecuente de secuencias reguladoras localizadas en *cis*, y la fusión de dichas secuencias con un gen reportero como la beta-glucoronidasa GUS³⁶. A partir de dichas fusiones se generan plantas transgénicas y se identifica cual de los fragmentos confiere actividad en los tejidos y el momento deseado.

25

Los solicitantes de esta patente proporcionan en este documento datos suficientes para caracterizar los promotores que fueron identificados: la demostración de que actúan de manera específica en los óvulos, los gametos femeninos, el grano de polen o los gametos masculinos, y que fueron aislados a partir de su identificación por medio de trampas génicas y las estrategias que a continuación se describen:

30

Después de haber identificado líneas de trampas génicas que tenían patrones de expresión en el gametofito femenino o masculino de *Arabidopsis thaliana*, se identificó el sitio de inserción del transposón activo por medio de TAIL-PCR. El transposón se puede haber insertado en regiones codificantes en la región promotora de unidades transcripcionales predichas a partir de programas y algoritmos establecidos para el análisis del genoma de *Arabidopsis thaliana*. En ambos casos, la identificación de secuencias reguladoras (promotores) se enfoca al gen identificado por el sitio de la inserción del transposón. En la mayoría de los casos las secuencias regulatorias han sido identificadas en los primeros 2 o 3 Kb inmediatamente adyacentes al extremo 5' de la región codificante del gen correspondiente³⁷. Por lo tanto, la región localizada en el extremo 5' de la región codificante del gen es amplificada por PCR y clonada frente a un gen reportero apropiado (GUS o GFP). En algunos casos excepcionales, secuencias regulatorias han sido identificadas en intrones o en regiones 3'³⁸. Si la región 5' no confiere el patrón de expresión esperado, posiblemente intrones o regiones 3' del mismo gen pueden ser clonadas frente a un promotor mínimo para monitorear nuevamente la expresión de un gen reportero (como por ejemplo el promotor mínimo del gen 35 utilizado en las construcciones detectoras de potenciadores³⁹). Si la inserción del transposón se localiza en una región localizada entre dos genes, el mismo análisis puede ser efectuado para determinar cual de los dos genes se activa en un patrón de expresión idéntico al que se identificó por medio de trampas génicas.

A partir de la información que aquí se presenta, las plantas transgénicas no son necesarias para obtener el invento ya que la localización de los promotores y su secuencia se muestran en las figuras 1 a 4, y la descripción de los patrones de actividad de cada uno de los promotores se ilustra en la Figura 6.

Una vez que los promotores han sido identificados y aislados, pueden ser utilizados para una multitud de aplicaciones para transformar plantas por estrategias de ingeniería genética. La selección de genes estructurales, del vector de transferencia y del método de transformación puede ser efectuada por cualquier persona

conocedora del arte y por lo tanto todos estos aspectos entran en los límites de protección de este invento.

GENES ESTRUCTURALES.

5 De la misma manera, a través del invento que aquí se presenta, genes de interés agronómico pueden ser expresados en plantas transformadas. De manera particular, plantas pueden ser transformadas por métodos de ingeniería genética para expresar varios fenotipos de interés agronómico cuyas características deben manifestarse en el gametofito femenino, los gametos femeninos, el grano de polen
10 o los gametos masculinos. Algunos ejemplos no excluyentes se incluyen a continuación.

1. Genes que confieren resistencia a plagas o enfermedades como por ejemplo:

15 A) Genes de resistencia a enfermedades. Las defensas de las plantas se activan de manera específica a partir de la interacción del producto de un gen de resistencia a enfermedades presente en la planta (llamados R) y del producto de un gen de avirulencia (Avr) presente en el patógeno. Una variedad de planta puede ser transformada con un gen
20 de resistencia a enfermedades para generar plantas transgénicas resistentes a un patógeno específico⁴⁰.

25 B) Una proteína de *Bacillus thuringiensis*⁴¹, que reportaron el aislamiento y secuenciación de un gen que codifica una Bt δ -endotoxina. Además, moléculas de DNA que codifican δ -endotoxinas pueden ser compradas en American Type Cell Culture Collection (Rockville MD, USA), bajo los números de catalogo Nos. 40098, 67136, 31995 y 31998.

30 C) Una lectina⁴², que reportó la secuencia nucleotídica de varios genes de unión a lectinas mediadas por manosa en *Clivia miniata*.

- 5
- D) Una proteína de unión a vitaminas, como la avidina⁴³. Se muestra el uso de la avidina y de sus homólogos como larvicidas contra insectos.
- E) Un inhibidor de enzimas como puede ser un inhibidor de proteasas o de amilasas⁴⁴.
- 10
- F) Una hormona específica de insecto o ferhormona como un ecdisteroide o una hormona juvenil, o una de sus variantes. Un ejemplo de ello es la expresión de una esterasa juvenil en baculovirus⁴⁵.
- 15
- G) Un péptido o neuropéptido específico de insecto que al ser expresado provoca una interrupción fisiológica en una plaga⁴⁶, como por ejemplo una alostatina identificada en *Diploptera puntata*". En la literatura se reportan genes que codifican neurotoxinas paráliticas específicas de insectos⁴⁷.
- 20
- H) Un veneno específico para insectos producido en la naturaleza por una víbora o una avispa etc. En la literatura se reporta la expresión heteróloga en plantas de un gen que codifica un péptido insecto-tóxico de escorpión⁴⁸.
- 25
- I) Una enzima responsable de la hiper-acumulación de un monoterpeno, un sesquiterpeno, un esteroide, el ácido hidroxámico, un derivado del fenilpropanol o de otra molécula no proteica que tiene actividad insecticida.
- 30
- J) Una enzima involucrada en la modificación de una molécula biológicamente activa (incluyendo modificaciones post-traduccionales); por ejemplo, una enzima glicolítica, una enzima proteolítica, una enzima lipolítica, una nucleasa, una ciclasa, una

transaminasa, una esterasa, una hidrolasa, una fosfatasa, una cinasa, una fosforilasa, una polimerasa, una elastasa, una quitinasa o una glucanasa, ya sean naturales o sintéticas. Scott reporta la secuencia nucleotídica de un gen que codifica una callasa⁴⁹. Moléculas que contienen secuencias codificantes de quitinasas pueden ser obtenidas, como ejemplo, por los métodos descritos por Kramer⁵⁰, quien muestra la secuencia nucleotídica de una secuencia de cDNA que codifica una quitinasa de gusano barrenador de tabaco, y por Kawalleck⁵¹, que reportan la secuencia nucleotídica del gen ubi4-2 que codifica una poliubiquitina.

K) Una molécula que estimule la señalización transduccional. Por ejemplo, ver el reporte de Botella⁵² que describen un cDNA correspondiente a un gen que codifica una calmodulina de frijol, o el de Griess⁵³, que reporta la secuencia nucleotídica de una clona de cDNA de calmodulina de maíz.

L) Un péptido hidrofóbico. Como por ejemplo derivados de péptidos de taquiplesina que inhiben el crecimiento de patógenos fúngicos de plantas⁵⁴ y péptidos sintéticos antimicrobianos que confieren resistencia a enfermedades⁵⁵.

M) Una permeasa membranal generadora o bloqueadora de la formación de canales celulares. Por ejemplo, ver el reporte de Jaynes⁵⁶ correspondiente a la expresión heteróloga de una cecropina lítica que confiere resistencia contra *Pseudomonas solanacearum* a plantas de tabaco.

N) Una proteína invasiva viral o un complejo de toxinas derivadas. Por ejemplo, la acumulación de proteínas de la cobertura viral en células transformadas de planta confiere resistencia contra infecciones virales⁵⁷. Este tipo de resistencia ha sido en plantas de alfalfa a las

que se les confirió resistencia contra el virus del mosaico del tabaco, el virus del mosaico del pepino, el virus X de la papa, el virus Y de la papa, o el virus mosaico del tabaco, como ejemplos.

5 O) Un anticuerpo específico de insectos o una inmunotoxina derivada de ella. Por ejemplo, un anticuerpo cuyo blanco es una función metabólica en el intestino del insecto que sea letal para este último⁵⁸. Inactivación enzimática de tabaco transgénico a partir de la producción de fragmentos de anticuerpo de cadena única.

10

P) Un anticuerpo específico de virus. Ver por ejemplo, el reporte de Tavladoraki⁵⁹ que muestra como plantas transgénicas que expresan genes que codifican anticuerpos recombinantes están protegidas contra ataques virales.

15

Q) Una proteína que detenga el desarrollo de un patógeno y que sea producida por un patógeno o un parásito. Por ejemplo, la endo 1-1-4 D-poligalacturonasa de origen fúngico facilita la colonización de hongos y la pérdida de nutrientes de la planta al solubilizar la homo 1-4D-galacturonasa en la pared celular⁶⁰. La clonación y caracterización de un gen que codifica una proteína inhibidora de endopoligalacturonasa se describe en Toubaart⁶¹.

20

R) Una proteína que detenga el desarrollo de un patógeno y que sea producido por la planta. Por ejemplo Logemann⁶² muestra que plantas transgénicas que expresan el gen inactivador del ribosoma en cebada tienen una resistencia elevada a enfermedades fúngicas.

25

2. Genes que confieren resistencia a un herbicida, como por ejemplo:

30

A) Un herbicida que inhibe el punto de crecimiento de los meristemos, como es la imidazalinona o sulfonilurea. Ejemplos de genes que

confieren resistencia a este tipo de herbicidas codifican para versiones mutantes de la enzima ALS o AHAS como lo describe Miki⁶³.

5 B) Gliofosato (resistencia conferida por una mutante de 5-enolpiruvyl-3-
fosfiquimato sintasa (EPSP) y los genes *aroA*, y otros compuestos
fosforados como son el glufosinato (genes que codifican una
fosfinotricina acetil transferasa (PAT) o una fosfinotricina acetil
transferasa de *Streptomyces hygroscopicus* conocido como el gen
10 *bar*), y ácidos piridinoxi – o fenoxi propiónicos y ciclohexonas (genes
que codifican inhibidores de ACCasas). Shah y colaboradores⁶⁴
liberaron la secuencia nucleotídica de una forma de EPSP que
confiere resistencia a gliofosfatos. La secuencia nucleotídica de una
mutante del gen *aroA* puede ser obtenida a partir de lo reportado por
Comai⁶⁵.

15

C) Un herbicida que inhibe la fotosíntesis como triazina (genes *psbA* y
gs+) y benzonitrilo (gen de la nitrilasa). Ver por ejemplo Przibilla⁶⁶
donde se describe la transformación de *Chlamydomonas* con
plásmidos que codifican mutantes del gen *psbA*. La secuencia
nucleotídica de los genes que codifican nitrilasas fue reportada por
20 Stalker⁶⁷ y las moléculas que contienen estos genes se encuentran
bajo el numero de acceso ATCC Nos. 53435, 67441 y 67442. La
clonación y expresión del gen que codifica una S-transferasa esta
descrita por Hayes⁶⁸.

25

3. Genes que confieren o contribuyen a obtener un rasgo de alto valor agregado como los siguientes ejemplos no excluyentes:

A) Modificaciones al metabolismo de los ácidos grasos, como por
ejemplo, al transformar con un gen anti-sentido que codifica a una
30 desaturasa que aumente el contenido de ácido esteárico en la
planta⁶⁹.

B) Una reducción en el contenido de fitatos:

5 a) Introduciendo un gen que codifica una fitasa y que contribuye a la degradación del fitato, aumentando el fosfato libre en la planta transformante. Ver por ejemplo Van Hartingsveldt⁷⁰ para el reporte de una secuencia nucleotídica de un gene que codifica una fitasa de *Aspergillus niger*.

10 b) Un gen que reduce el contenido de fitato. En el maíz, esto puede ser logrado al clonar y re-introducir DNA asociado con un solo alelo responsable de mutantes que se caracterizan por un nivel bajo de fitato⁷¹.

15 C) Modificaciones del contenido en carbohidratos, por ejemplo, al transformar plantas con un gen que codifica una enzima que altera el patrón de ramificación del almidón. Ver por ejemplo el reporte de Shiroza⁷² donde se utiliza la secuencia nucleotídica de un gen que codifica una fructosiltransferasa de *Streptococcus mutans*, de Steinmetz⁷³ donde se utiliza la secuencia nucleotídica de una levansucrasa de *Bacillus subtilis*), de Elliot⁷⁴ donde se utiliza una secuencia nucleotídica de un gen que codifica una invertasa de tomate), o de Fischer⁷⁵ que reporta la secuencia nucleotídica de un gen que codifica una enzima de ramificación del almidón del endospermo.

20 D) Inducción de muerte celular programada del polen transgénico. Un rasgo transgénico transferido se comporta como un carácter dominante cuya herencia puede resultar indeseable en ciertas condiciones. En el caso de plantas que permiten la polinización cruzada, este tipo de transferencia involuntaria puede tener consecuencias legales y ecológicas que comprometan seriamente la implementación de cultivos transgénicos en cualquier país del mundo. En numerosas especies, la transferencia horizontal de un rasgo transgénico a plantas no-transgénicas de la misma especie puede causar graves confusiones entre plantas destinadas a la producción industrial y plantas destinadas al consumo humano. Adicionalmente,

25

30

E)

F)

G)

H)

I)

5

10

15

parientes cercanos a plantas de cultivo se pueden ver también afectados por cruzas inter-específicas que les transfieran a los híbridos F1 características transgénicas indeseables, lo que puede representar graves riesgos para programas de mejoramiento que utilizan métodos de introgresión. Este último problema es particularmente importante en países como México que representan el centro de origen de numerosas especies de interés agrícola. Para eliminar la transferencia horizontal de rasgos transgénicos se pueden utilizar genes gametofíticos letales que, ligados en una construcción molecular al carácter transgénico, eliminen la posibilidad de que granos de polen que contienen el alelo transgénico participen en la formación de semillas. El gen gametofítico letal estaría presente únicamente en el genoma los granos de polen que contienen la construcción transgénica (50% de los granos de polen). El resto del polen (50%) no portará el rasgo transgénico, y al ser completamente viable podrá asegurar la polinización cruzada (y la auto-polinización), dando lugar a la formación de semillas libres de rasgos transgénicos.

20

4. Genes que controlan el crecimiento y desarrollo de la semilla, como por ejemplo:

25

30

- A) Genes que controlan la proliferación celular y el crecimiento del embrión y/o del endospermo como pueden ser reguladores del ciclo celular⁷⁶, genes que afectan la proliferación celular en el embrión y en el endospermo para aumentar el tamaño de la semilla, generar cambio de biomasa o minimizar el contenido de ciertos compuestos en la semilla⁷⁷.
- B) Genes que promueven la proliferación autónoma del endospermo en ausencia de fecundación, una componente esencial de la apomixis (o reproducción asexual a partir de semillas), como MEA, FIS2 o FIE (ver referencias del párrafo anterior).

C) Genes que promueven la formación del embrión, otra de las componentes de la apomixis (reproducción asexual a partir de semillas) tal y como SOMATIC EMBRYOGENESIS RELATED KINASE (SERK)⁷⁸ o LEAFY COTYLEDONS.

5

5. Genes que generan muerte celular programada del polen transgénico para evitar la transferencia de rasgos transgénicos en plantas de cultivo, como por ejemplo:

10

A) El gen que codifica para una RNAsa T1⁷⁹ y el gen que codifica para la cadena A de la toxina difteria de *Pseudomona aeruginosa*⁸⁰. Se ha demostrado ampliamente que la expresión de cada uno de estos dos genes es capaz de provocar la muerte de células vegetales con gran especificidad⁸¹.

15

PROMOTORES.

Los promotores que aquí se describen y que son motivo de esta solicitud de patente pueden ser utilizados en asociación con versiones naturales de secuencias transcritas de un gen estructural deseado o con cualquier otra secuencia transcrita que sea crítica para la formación o la función de genes.

20

Puede ser también deseable el que se incluyan secuencias de intrones en las construcciones que contengan el promotor ya que la inclusión de secuencias de intrones puede resultar en la amplificación de la expresión o de la especificidad del promotor. Puede resultar ventajoso el unir secuencias de DNA a una secuencia de promotor que contiene el primer intrón y el primer exón de un polipéptido que sea

25

único a las células y los tejidos críticos para la formación de semillas.

30

Adicionalmente, ciertas regiones de un promotor pueden ser unidas a regiones de un segundo promotor diferente al primero para obtener la actividad deseada, dando lugar a un promotor quimérico. Promotores sintéticos que regulan la expresión de genes también pueden ser utilizados.

El sistema de expresión también puede ser optimizado adicionalmente empleando elementos suplementarios como terminadores transcripcionales o elementos potenciadores.

5 **OTROS ELEMENTOS REGULADORES.**

Además de la secuencia de un promotor, una construcción debe también contener una región terminadora de transcripción adyacente al gen estructural deseado, y esto de manera a proveer de una terminación transcripcional eficiente. La región terminadora o la señal de poliadenilización puede ser obtenida a partir del mismo
10 gen al cual pertenece el promotor utilizado o a partir de la región terminadora de otros genes. Las secuencias de poliadenilización incluyen pero no están limitadas a la señal de terminación del gen que codifica la octopina sintasa de *Agrobacterium tumefaciens*⁸², o la señal terminadora del gen que codifica la nopalina sintasa⁸³.

15 **GENES MARCADORES.**

Moléculas de ADN recombinante que contengan cualquiera de las secuencias y de los promotores que se describen en esta solicitud de patente pueden contener genes marcadores adicionales que codifiquen un producto génico de selección que le confiere a las células de las plantas resistencia a un agente químico o a un estrés
20 fisiológico, o que confieren una característica fenotípica distintiva a células de la planta transformadas con moléculas de DNA recombinante que pueden ser reconocidas por medio de un agente de selección. Uno de los marcadores de selección rutinariamente utilizados es el gen que codifica la neomicina fosfotransferasa (NPTII), que confiere resistencia a kanamicina y al antibiótico G-
25 418. Células transformadas con estos marcadores de selección pueden ser seleccionadas al determinar la presencia in-vitro de resistencia a kanamicina en las plántulas, utilizando técnicas ampliamente descritas en la literatura o determinando la presencia del mRNA correspondiente al gen NPTII utilizando un Northern blot a
30 partir de tejido de la planta transformada. La reacción de la polimerasa en cadena (PCR) puede también ser utilizada para identificar la presencia de un transgen utilizando

estrategias rutinarias de RT-PCR para monitorear la expresión de gen marcador. Otros marcadores de selección que son rutinariamente utilizados son el gen de resistencia a ampicilina, el gen de resistencia a tetraciclina y el gen de resistencia a higromicina. Las células de plantas transformadas pueden ser inducidas para diferenciarse en estructuras de plantas que eventualmente pueden dar lugar a plantas completas. Se entiende que un gen marcador también puede ser endógeno a una planta, como es el caso de genes de pigmentación del endospermo que pueden ser utilizados como genes marcadores en el maíz.

10 **TRANSFORMACION.**

Una proteína de ADN recombinante que sea diseñada para inhibir la expresión o para inducir la expresión y que contenga las secuencias de ADN y/o los promotores que aquí se describen puede ser integrada en el genoma de una planta al introducirla en las células de dicha plantas por algunos de los métodos ampliamente conocidos y sus variantes. Generalmente la molécula de ADN recombinante se introduce inicialmente en un vector apropiado y dicho vector es utilizado para introducir la molécula de DNA recombinante en una célula de la planta.

El uso del virus del mosaico de la coliflor (CaMV)⁸⁴ y algunos geminivirus como vectores⁸⁵ han sido sugeridos como vectores, pero el vector más exitoso se ha desarrollado con *Agrobacteria* sp⁸⁶.

Métodos para el uso de *Agrobacterium* basados en sistemas de transformación han sido ya descritos para muchas especies de plantas diferentes. Generalmente las cepas de bacterias contienen versiones modificadas del plásmido Ti de manera que el ADN sea transferido a la planta huésped sin que provoque la formación subsecuente de tumores. Estos métodos dependen de la inserción dentro de los bordes del plásmido Ti del ADN que se desea sea insertado en el genoma de la planta y que incluye un marcador de selección, generalmente un gen marcador como alguno de los arriba descritos. Numerosos

tejidos vegetales pueden servir de blanco para transformación con *Agrobacterium*, y varios han sido específicamente descritos para ser utilizados en las plantas Crucíferas, familia a la cual pertenece *Arabidopsis*. Estos tejidos incluyen capas finas de células⁸⁷, hipocotilos⁸⁸, discos en hojas⁸⁹, cotiledones⁹⁰ y embriones⁹¹, o hasta plantas enteras usando métodos de infiltración disponibles para *Arabidopsis* y *Medicago* sp. pero que se estandarizan continuamente para otras especies de plantas. Se entiende, sin embargo, que hay otros métodos de transformación que pueden ser más convenientes para otros cultivos.

10 Otros métodos que han sido utilizados para introducir moléculas recombinantes en plantas incluyen la inoculación directa con DNA, a través de liposomas, la electroporación⁹² y la micro-inyección⁹³. La posibilidad de usar microproyectiles y un cañón u otro tipo de instrumento para forzar partículas cubiertas con DNA dentro de células ha recibido considerable atención⁹⁴.

15 Es a menudo conveniente contemplar en el marco del invento que aquí se presenta la posibilidad de generar plantas homocigotas parentales que requieren más de un evento de transformación con el mismo producto. También se puede contemplar que una célula de planta sea transformada con una molécula de DNA recombinante que contenga al menos dos secuencias diferentes o transformada con más de una molécula recombinante en eventos de transformación distintos. Las moléculas de DNA pueden estar físicamente ligadas en el mismo vector, o separadas y en vectores distintos. Una célula puede ser simultáneamente transformada con más de un vector siempre y cuando cada vector tenga un marcador único de selección.

20 Alternativamente, una célula puede ser transformada con más de un vector permitiendo un paso de regeneración intermedia después de la transformación con el primer vector. Adicionalmente, puede ser posible realizar una cruce sexual a partir de plantas o de líneas de plantas que contengan diferentes secuencias de DNA o diferentes moléculas de DNA recombinante y seleccionar en los descendientes de dicha cruce plantas que contengan ambas secuencias deseables de DNA endógeno o recombinante.

30

La expresión de moléculas de DNA recombinante en plantas transformantes a partir de los promotores y secuencias que aquí se describen puede ser monitoreada usando técnicas de Northern o Southern Blot y métodos de PCR que son conocidas por todas aquellas personas que practican la biología molecular.

5

La regeneración de plantas viables a partir de células individuales que han sido transformadas ha sido realizada con éxito para un gran número de especies de plantas. Por ejemplo, la regeneración ha sido demostrada en dicotiledóneas para al menos las siguientes especies: manzana (*Malus pumila*)⁹⁵, frambuesa y sus híbridos derivados (*Rubus sp.*)⁹⁶, zanahoria (*Daucus carota*)⁹⁷, coliflor (*Brassica oleracea*)⁹⁸, pepino (*Cucumis sativus*)⁹⁹, apio (*Apium graveolens*)¹⁰⁰, berenjena (*Solanum melonoea*)¹⁰¹, lechuga (*Lactuca sativa*)¹⁰², papa (*Solanum tuberosum*)¹⁰³, soya (*Glycine canescens*)¹⁰⁴; fresa (*Fragaria x anannassa*)¹⁰⁵, tomate (*Lycopersicon esculentum*)¹⁰⁶, nuez (*Junglans regia*)¹⁰⁷, melón (*Cucumis melo*)¹⁰⁸, vid (*Vitis vinifera*)¹⁰⁹, mango (*Magnifera indica*)¹¹⁰; y para las siguientes especies monocotiledóneas: arroz (*Oryza sativa*)¹¹¹, maíz¹¹² y *Secale cereale*¹¹³.

Adicionalmente, la regeneración de plantas completas a partir de células que no han sido necesariamente transformadas ha sido exitosa en chabacano (*Prunus armeniaca*)¹¹⁴; espárrago (*Asparagus officinalis*)¹¹⁵; plátano (*Musa acuminata*)¹¹⁶, frijol (*Phaseolus vulgaris*)¹¹⁷, cereza (*Prunus sp.*)¹¹⁸, vid (*Vitis vinifera*)¹¹⁹, mango (*Mangifera indica*)¹²⁰, melón (*Cucumis melo*)¹²¹, oca (*Abelmoschus esculentus*)¹²², cebolla (*Allium sp.*)¹²³, naranja (*Citrus senensis*)¹²⁴, papaya (*Carrica papaya*)¹²⁵, durazno (*Prunus persica* y *Prunus domestica*)¹²⁶, pera (*Pyrus comunis*)¹²⁷, piña (*Ananas comosus*)¹²⁸, sandía (*Citrullus vulgaris*)¹²⁹ y trigo (*Triticum aestivum*)¹³⁰.

Las plantas regeneradas se transfieren al suelo y se cultivan de forma convencional. Después de que la construcción se incorpora de manera estable en el genoma de las plantas transgénicas, se puede transferir a otras plantas de la misma especie o de especies filogenéticamente cercanas a partir de una cruce

30

sexual. Varias técnicas de mejoramiento pueden ser utilizadas, dependiendo de la especie que se quiera cruzar.

5 Puede resultar útil generar un cierto número de plantas transformadas a partir de cualquier construcción de DNA recombinante para recuperar plantas que no presenten efectos indeseables debido a la posición del inserto transgénico. Puede ser también deseable obtener plantas que contengan más de una copia de la molécula recombinante.

10 De acuerdo con la incorporación preferencial, la planta transgénica que se utiliza para la producción comercial de una proteína recombinante es el maíz. En otra incorporación preferencial de este invento, la biomasa de interés son las semillas. Para el número relativamente reducido de plantas que muestran niveles elevados de expresión, un mapa genético puede ser generado principalmente a partir de la

15 estrategia de Fragmentos de Restricción de Longitud Polimórfica (RFLPs, por sus siglas en inglés), a partir de la reacción de polimerasa en cadena (PCR), y de secuencias repetidas sencillas (SSRs) que permiten identificar la localización cromosómica de la molécula de DNA recombinante. Para ejemplos de estas metodologías, ver Glick y Thompson¹³¹. La información del mapa referente a la

20 localización cromosómica es útil para protección de la propiedad intelectual de un individuo transgénico en plantas. Si se efectúa una propagación no autorizada de dicha planta transgénica cruzándola con otro tipo de germoplasma, el mapa de la región de integración puede ser comparado con otros mapas de plantas sospechosas, para determinar si las plantas sospechosas de haber sido

25 propagadas sin autorización tiene parientes en común con las plantas transgénicas protegidas. La comparación de mapas se efectúa a partir de hibridaciones, RFLPs, SSRs, PCRs y experimentos de secuenciación, todas técnicas y tecnologías ampliamente conocidas en el medio.

30 Como se indica más arriba, puede resultar deseable producir líneas de plantas transgénicas que sean homocigotas para un gen en particular. En algunas especies esto se puede realizar rutinariamente a partir del cultivo *in-vitro* de anteras o de

microsporas. Esto es particularmente importante en el caso de la canola, *Brassica napus*¹³². Utilizando estas técnicas, es posible producir una línea haploide que tiene integrado el gen de interés y duplicar el número cromosómico espontáneamente o partir del uso de colchicina. Esto da lugar a una planta homocigota para el gen de interés, lo que puede ser fácilmente probado si el gen insertado incluye un marcador de selección apropiado para la detección de plantas transgénicas. De manera alternativa, las plantas pueden ser auto-fecundadas, dando lugar a una mezcla de semilla que contiene 25% de homocigotas, 50% de heterocigotas y 25% de nulias (no transgénicas) para el gen de interés. Aunque resulta relativamente sencillo identificar plantas que contienen y que no contienen el gen de interés, en la práctica se puede también identificar las plantas homocigotas y heterocigotas a partir de un análisis tipo Southern en el cual se presta especial atención a la cantidad de DNA genómico que se carga en el gel antes de su hibridación. Se sugiere verificar los resultados del análisis por Southern Blot dejando que cada transformante independiente se auto-fecunde, ya que cualquier evidencia adicional de homocigosis puede obtenerse a partir del simple hecho de que, si la planta es homocigota para el gen insertado, todas las plantas que resulten de la semilla generada por auto-fecundación heredarán y contendrán el gen, mientras que si la planta es heterocigota, ciertas plantas resultantes de la semilla generada por auto-fecundación no heredarán ni contendrán el gen de interés. Por lo mismo, a partir de una simple auto-fecundación se pueden seleccionar plantas homocigotas transgénicas.

La producción de plantas parentales transgénicas permite la generación de híbridos y semillas que contengan un componente proteico modificado. Las plantas transgénicas homocigotas parentales se mantienen de manera que cada uno de los padres contenga ya sea la secuencia de la primera o de la segunda molécula recombinante ligada a un promotor activo. En este esquema también se incorporan las ventajas de cultivar un híbrido, incluida la ventaja de tener varias características agronómicamente interesantes combinadas en un mismo individuo.

Los siguientes ejemplos sirven para ilustrar mejor el invento que aquí se describe y no pretenden limitar el invento de ninguna manera.

EJEMPLOS

- 5 En nuestro laboratorio se identificó, aisló y caracterizó una serie de promotores de *Arabidopsis* con actividad en el saco embrionario. Estos promotores fueron identificados a través del uso de líneas gene-trap y enhancer-trap. En las cuales se buscaron líneas transposantes con esquemas de expresión de *GUS* asociado al
- 10 saco embrionario en desarrollo y en líneas que presentaran defectos a nivel reproductivo. Se determinó el sitio de inserción del transposón y se identificó la posible secuencia-regulatoria responsable del esquema de expresión observado. Dicha secuencia fue fusionada a *GUS* y en las plantas transformantes T1 se determinó su esquema de expresión.
- 15 EL promotor pFM1 (FUNCTIONAL MEGASPORE-1) ver SEQ ID1, fue identificado a partir de una línea “detector de potenciadores”, o en inglés “enhancer-trap”. Esta línea presenta un esquema de expresión de *GUS* que inicia a partir de la megaspora funcional y se prolonga hasta el saco embrionario maduro. Mediante TAIL-PCR se determinó que el elemento Ds se insertó en una región intergénica, a
- 20 788 pb del gen At4g12250 (nucleotide sugar epimerase) y a aproximadamente 2000 pb de gen At4g12260 (similar a transposable element Ac-like). Para identificar la secuencia regulatoria responsable del esquema de expresión observado en la línea transposante se amplificó mediante PCR una secuencia de 880 pb, que comprende la región del nucleótido -817 al +63, tomando como referencia el codón de inicio
- 25 (ATG) (ver figura 5).

- Los iniciadores utilizados en sentido y en antisentido (ver secuencias subrayadas en la figura 5) se les adicionaron los sitios de restricción HindIII y BamHI respectivamente (en negritas). Esta secuencia se clonó en el vector PCRII-TOPO.
- 30 Este fragmento se digirió con las enzimas de restricción mencionadas, se purificó y se clonó en el vector pBI101. El vector pBI101 posee un sitio de clonación múltiple

río arriba del gen GUS, lo que permitió generar una fusión transcripcional; promotor-pFM1::GUS.

5 Se transformaron plantas de *Arabidopsis thaliana* ecotipo *Columbia*, mediante la técnica de floral dipping. Tinciones de GUS realizadas a plantas transformantes T1 presentaron un esquema de expresión idéntico a la línea transposante (Figura 6). El esquema de expresión de la construcción pFM1-GUS presenta actividad después de meiosis e inicia en la megaspora funcional hasta el saco embrionario maduro y se restringe al núcleo polar y al aparato huevo, no se detectó actividad en las 10 células antípodas (Figura 6a a 6c). Después de fecundación, se observa expresión de GUS en estado globular, en la célula del suspensor más cercanas al micrópilo (datos no mostrados).

El promotor pES1 (EMBRYO SAC 1), ver SEQ ID2, fue identificado a partir de otra 15 línea “detector de potenciadores” o “enhancer-trap”. Esta línea presenta un esquema de expresión de GUS durante megagametogénesis a partir del estado 4-8 nuclear del saco embrionario hasta el saco embrionario maduro donde es activo en las células antípodas, núcleo polar y aparato huevo (Figura 1d). Mediante TAIL-PCR se determinó que el elemento Ds se insertó en la región 5'-no traducida (5'- 20 UTR) del gen At2g41050 que codifica para una proteína hipotética que presenta homología con una proteína transmembranal de *Saccharomyces cerevisiae* (nº de acceso: YDR352W). Para identificar la secuencia regulatoria responsable del esquema de expresión observado en la línea transposante se amplificó mediante PCR una secuencia de 600 pb que comprende la región del nucleótido -597 al -3, 25 a partir del codón de inicio (ATG) (ver figura 7). Los iniciadores utilizados en sentido y en antisentido se muestran como secuencias subrayadas en la figura 7.

Los promotores pPo1 (ver SEQ ID3) y pPo2 (ver SEQ ID4) (POLLEN PROMOTER 1 and 2) se identificaron a partir de 2 líneas “detector de potenciadores” o “enhancer trap” adicionales mostraron patrones de expresión del gen reportero en el grano de 30 polen (el gametofito masculino) y las células espermeáticas (los gametos masculinos) (Figura 1e y 1f). En una de ellas la inserción se encontró en la región

3'UTR de un gen que codifica una manosa deshidrogenasa (At4g37970), y en la otra se encontró dentro de la región regulatoria 5' de un gen que codifica un factor de transcripción del tipo *WUSCHEL* (At3g11260). Con la ayuda de iniciadores específicos se amplificaron las secuencias mostradas en las figuras 8 y 9.

5

Las secuencias mostradas en las figuras 8 y 9 se clonaron en el vector PCRII-TOPO. Utilizando los sitios de restricción XbaI y BamHI presentes en este vector se liberó y purificó esta secuencia y se clonó en el vector pBI101. El vector pBI101 posee un sitio de clonación múltiple río arriba del gen GUS, lo que permite generar una fusión transcripcional; promotor-pES1::GUS. Se transformaron plantas de *Arabidopsis thaliana* ecotipo *Columbia*, mediante la técnica de floral dipping. Tinciones de GUS realizadas a plantas transformantes T1 presentaron un esquema de expresión idéntico a la línea transposante. El esquema de expresión de la construcción pES1-GUS mostró actividad específica de GUS a partir de los estados 4 y 8-nuclear, hasta el saco embrionario maduro y una expresión adicional en pequeñas zonas de la raíz.

Se transformaron plantas de *Arabidopsis thaliana* ecotipo *Columbia*, mediante la técnica de "floral dipping" (transformación floral). Tinciones de GUS realizadas en plantas transformantes T1 mostraron que la actividad de esta secuencia es de tipo constitutiva, ya que se observó tinción de GUS en todos tejidos a nivel de plántula y en inflorescencias presenta actividad en todas las células del óvulo.

Adicionalmente, el promotor pFM1 (ver SEQ ID1) se puede utilizar de manera exitosa para alterar el desarrollo del gametofito femenino y las células reproductivas femeninas en *Arabidopsis thaliana*. Para ello, se utilizó una secuencia del gen *CHR11* de *Arabidopsis thaliana* clonada frente al promotor pFM1 en orientación sentido y anti-sentido, de manera a silenciar post-transcripcionalmente el gen *CHR11* por la estrategia conocida como RNA intereferente (RNAi). *CHR11* codifica una proteína remodeladora de cromatina esencial para las divisiones nucleares que dan lugar a la formación de los gametos femeninos en el gametofito femenino. La plantas transformantes bajo el control del promotor pFM1 mostraron gametofitos

femeninos alterados (Figura 10) pero ningún otro defecto en su crecimiento vegetativo o reproductivo, demostrando que el promotor pFM1 confiere activación específica de genes en el gametofito femenino.

5 Referencias.

1. PCT W084/02913 Rogers et al.
2. Deikman and Fischer, *EMBO J*, 1988; Giovannoni et al., *The Plant Cell* (1989) 1:3.
3. Bird et al., *Plant Mol. Biol.* p 651-663 (1988).
- 10 4. U.S Patent 5,023,179 de Lam et al.
5. Walling et al. *PNAS* 83:2123-2127.
6. U.S Patent 5,110,728.
7. *Plant Physiology* 1992, 100(2) 576-581.
8. U.S. Patent 5,504,200.
- 15 9. *New IEEE Standard Dictionary of electrical and Electronic Terms* (5th edition, 1993).
10. *Diagnostic Molecular Microbiology: Principles and Applications* D.H Persing et al., Ed., American Society for Microbiology, Washington D.C (1993).
11. Murray et al. *Nucl. Acid. Res.* 17:447-498 – 1989.
- 20 12. *Plant Molecular Biology: A Laboratory Manual* Clark Ed., Springer-Verlag Berlin-1997.
13. Compounds and Methods for Site Directed Mutagenesis in Eukaryotic Cells, Kmiec, U.S. Patent No. 5,565,350; In Vivo Homologous Sequence Targeting in Eukaryotic Cells; Zarlino et al.; PCT/US93/03868.
- 25 14. Berger and Kimmel, Guide to Molecular Biology techniques, Methods in Enzymology, Vol. 152, Academic Press, Inc., San Diego CA (Berger); Sambrook et al. Molecular Cloning – A Laboratory Manual, 2nd Ed., Vol 1-3 (1999); y Current Protocols in Molecular Biology, FM Ausubel et al., Eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. Y John
- 30 Wiley and Sons, Inc. (1994).
15. *Anal. Biochem* 138:267-284 (1984).

16. *Laboratory Techniques in Biochemistry and Molecular Biology- Hybridization with Nucleic Acid Probes*, Part I, Chapter 2, Ausubel et al. Eds, Greene Publishing and Wiley-Interscience, New York (1995).
17. *Adv. Appl. Math.* 2:482 (1981);
18. *J. Mol. Biol.* 48:443 (1970);
- 5 19. *Proc. Natl. Acad. Sci. USA* 85:2444 (1988);
20. *CABIOS* 5:151-153 (1989); Corpet et al., *Nucleic Acids Research* 16:10881-90 (1988); Huang et al., *Computer Applications in the Biosciences* 8:155-65 (1992) y Pearson et al., *Methods in Molecular Biology* 24:307-331 (1994).
21. *Current Protocols in Molecular Biology*, Chapt 19, Ausubel et al. Eds Grenne
10 Publishing and Wiley-Interscience, New York (1995).
22. Altschul et al. *Nucleic Acids Research* 25:3389-3402 1997.
23. Altschul et al., *Nucleic Acids Research* 25:3389-3402 1997.
24. Henikoff y Henikoff 1989 – *Proc. Natl. Acad. Sci. USA* 89:10915.
25. Karlin y Altschul, *Proc. Natl. Acad. Sc.* 90:5787-5873 – 1993.
- 15 26. Wooten and Federhen, *Comput. Chem.*, 17:149-163 1993.
27. Claverie and States, *Comput. Chem.* 17:191-201 1993.
28. *Comput. Appl. Biol. Sci.* 4:11-17 1988.
29. *J. Mol. Biol.* 48:443 1970.
30. O’Kane C and Ghering WJ 1987; “Detection in situ of genomic regulatory
20 elements in *Drosophila*”, *PNAS* 84, 9123-9127; Sundaresan et al., 1995
“Patterns of gene action revealed by enhancer trap and gene trap transposable
elements” *Genes and Development* 9:1797-1810.
31. Goldberg en Phil. Trans. Royal Soc. London 1986: B-314-343.
32. Christoffersen et al., *Plant Molecular Biology* 1984 3:385.
- 25 33. O’Kane C and Ghering WJ 1987; “Detection in situ of genomic regulatory
elements in *Drosophila*”, *PNAS* 84, 9123-9127; Sundaresan et al., 1995
“Patterns of gene action revealed by enhancer trap and gene trap transposable
elements” *Genes and Development* 9:1797-1810.
- 30 34. Wilson C, pearson RK, Bellen HJ, O’Kane CJ, Grossniklaus U, Ghering WJ
1989; “P-element mediated enhancer detection: An efficient method for isolating
and characterizing developmentally regulated genes in *Drosophila*” *Genes &*

- Development 3, 1301-1313; Skarnes WC, 1990 "entrapments vectors: a new tool for mammalian genetics", Biotechnology 8, 827-831; Topping JF, Wei W, Lindsey K (1991) "functional tagging of regulatory elements in the plant genome" Development 112, 1009-1019; Sundaresan V., Springer P., Volpe T.,
- 5 Haward S, Jones J, Dean C, Ma H., Martienssen R (1995); Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements" Genes & Dev., 9 1797-1810).
35. Liu YG, Mitsukawa N., Oosumi T, Whittier RF 1995; "Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR" Plant J., 8, 457-463.
- 10 36. Jefferson RA, Kavanagh TA, Bevan MW 1987; GUS-fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. Trends in Genetics 11: 328-329.
37. Dwyer KG, Kandasamy MK, Mahosky DI, Acciai J, Kudish BI, Mille JE, Nasrallah M, Nasrallah JB 1994; " A superfamily of S-locus related sequences in Arabidopsis: diverse structures and expression patterns Plant Cell 6: 1829-1843; Thoma S, Hecht U, Kippers A, Botella J, de Vries S, Sommerville C 1994; "Tissue specific expression of the carrot gene *EP2* lipid transfer protein gene. Plant Cell 3:907-921; Xia Y, Nikolau BJ, Schnable PS 1996; "Cloning and
- 15 20 characterization of *CER2*, an Arabidopsis gene that affects cuticular wax accumulation Plant Cell 8:1291-1304.
38. Gidekel, M, Jimenez B, Herrera-Estrella L 1996; The first intron of the Arabidopsis thaliana gene coding for elongation factor 1-beta contains an enhancer-like element Gene 170: 201-206; Sieburth LE, Meyerowitz EM 1997;
- 25 Molecular dissection of the AGAMOUS control region shows that *cis* elements for spatial regulation are located intragenically Plant Cell 9: 355-365; Daniel SG, Becker WM 1995 "Transgenic analysis of the 5' and 3' flanking regions of the NADH-dependent hydroxypyruvate reductase gene from *Cucumis sativus* L. Plant Mol Biol. 28:821-836.
- 30 39. Sundaresan V., Springer P., Volpe T., Haward S, Jones J, Dean C, Ma H., Martienssen R (1995); Patterns of gene action in plant development revealed by

- enhancer trap and gene trap transposable elements" *Genes & Dev.*, 9 1797-1810.
40. Jones et al. 1994 "Cloning of the tomato *Cf-9* gene for resistance gene to *Cladosporium fulvum*; ; Martin et al., 1993 "Tomato *Pto* gene for resistance to *Pseudomonas syringae* pv. *Tomato* encodes a protein kinase" *Science* 262:1432; Mindrinos et al. 1994 "Arabidopsis *RSP2* gene for resistance to *Pseudomonas syringae*" *Cell* 78: 1089.
- 5 41. Geiser et al. *Gene* 48: 109 (1986).
42. Van Damme, *Plant Molec. Biol.* 24: 25 (1994).
43. PCT US93/06487.
- 10 44. Abe et al., *J. Biol. Chem.* 262: 16793 (1987) "Nucleotide sequence of rice cysteine proteinase inhibitor"; Huub et al., *Plant Molec. Biol.* 21: 985 (1993) "Nucleotide sequence of cDNA encoding tobacco proteinase inhibitor I", y Sumitani et al., *Biosci Biotech. Biochem.* 57: 1243 (1993) "Nucleotide sequence of *Streptomyces nitrosporeus* alpha-amylase inhibitor".
- 15 45. Hammock et al., *Nature* 344: 458 (1990).
46. Regan et al., *J. Biol. Chem.* 269: 9 (1994)"Expression cloning yields DNA coding for insect diuretic hormone receptor" y de Pratt et al. *Biochem. Biophys. Res. Comm.* 163: 1243 (1989).
47. USAPat No. 5,266,317.
- 20 48. Pang et al., *Gene* 116: 165 (1992).
49. PCT WO 93/02197.
50. *Insect Biochem. Molec. Biol.* 23: 691 (1993).
51. *Plant Molec. Biol.* 21: 673 (1993).
52. *Plant Molec Biol.* 24:757 (1994).
- 25 53. *Plant Physiol.* 104: 1467 (1994).
54. PCT WO95/16776.
55. PCT WO95/18855.
56. *Plant Sci* 89: 43 (1993).
57. Beachy et al., *Ann Rev. Phytopathol.* 28: 451 (1990).
- 30 58. Taylor et al., Abstract #497 7th Intl. Symposium on Molecular Plat-Microbe Interactions (Edinburgo, Escocia 1994).
59. *Nature* 366: 469 (1993).

60. Lamb et al., *Bio/Technology* 10:1436 (1992).
61. *Plant J.* 2:367 (1992).
62. *Bio/Technology* 10:305 (1992).
63. *Theor Appl. Genet.* 80: 449 (1990).
- 5 64. USPat. 4,940,835.
65. USPat. 4,769,061.
66. *Plant Cell* 3:169 (1991).
67. USPat. 4,810,648.
68. *Biochem. J.* 285:173 (1992).
- 10 69. Knultzon et al. *Proc. Natl. Acad. Sci. USA* 89, 2624 (1992).
70. *Gene* 127: 87 (1993).
71. Raboy et al. *Maydica* 35: 383 (1990).
72. *J. Bacteriol.* 170: 810 (1988).
73. *Mol Gen. Genet.* 200: 220 (1985).
- 15 74. *Plant Molec. Biol.* 21:515 (1993).
75. *Plant Physiol.* 102: 1045 (1993).
76. Bogre et al., "regulation of cell division and the cytoskeleton by mitogen-activated protein kinases in higher plants" *results probl. Cell. Differ.* 27:95-117 (2000); Huntley RP and Murray JA "The Plant Cell Cycle" *Curr Opin. Plant biol.* 2:440-446 (1999), MEDEA (MEA) (Grossniklaus et al. 1998 "Maternal control of embryogenesis by MEDEA, a Polycomb group gene in Arabidopsis" *Science* 280: 446-50 (1998), FERTILIZATION-INDEPENDENT SEEDS2 (FIS2) (Luo et al., "genes controlling fertilization-independent seed development in Arabidopsis thaliana" *proc Natl. Academ sci USA* 96:296-301 (1999) and FERTILIZATION-INDEPENDENT ENDOSPERM (FIE) (Ohad et al., "Mutations in FIE, a WD Polycomb group gene, allow endosperm development without fertilization" *Plant Cell* 11:407-16 (1999).
- 20 77. USPat 6,239,327. J-Ph. Vielle-Calzada y U. Grossniklaus.
- 25 78. Schmidt et al., "A leucine-rich repeat containing receptor-like kinase marks plant cells competent to form embryos" *Development* 124:2049-62 (1997).
- 30

79. Mariani C, de Beuckeleer M, Truettner J, Leemans J and Goldberg RB. 1990. "Induction of male sterility in plants by a chimeric ribonuclease gene". *Nature* 347:737-741.
80. Koning A, Jones A, Fillatti JJ, Comai L and Lassner MW. 1992 "Arrest of embryo development in *Brassica napus* mediated by modified *Pseudomonas aeruginosa* exotoxin A". *Plant Mol. Biol.* 18:247-258; Palmitier RD, Behringer RR, Quaife CJ, Maxwell F, Maxwell IH and Brinster RL. 1987. "Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene". *Cell* 50:435-443).
81. Czako M, Jyan-Chyun J, Herr JMJ and Marton L. 1992. Differential manifestation of seed mortality induced by seed specific expression of the gene for diphtheria toxin A chain in *Arabidopsis* and tobacco. *Mol. Gen. Genet.* 235:33-40.
82. Gielen et al., *EMBO J.* (1984) 3:835-846.
83. Depicker et al., *Mol. Appl. Genet.* (1982) 1:1561-1573.
84. Howell SH et al., 1980. *Science* 208:1265.
85. Goodman RM, 1981, *J. Gen. Virol.* 54:9.
86. Horsch et al., 1995 *Science* 227:1229-1231.
87. Charest PJ et al., 1988 *Theor Appl. Genet* 75:438-444).
88. DeBlock M et al. *Plant Physiol.* 91:694-701.
89. *Plant Cell. Rep.* 6:321-325.
90. Moloney MM et al., 1989, *Plant Cell Reports* 8:238-242.
91. Neuhaus G et al., 1987 *Theor Appl. Genet.* 75:30-36.
92. Guerche P. 1987 *Plant Science* 52:111-116.
93. Neuhaus G., 1987 *Theor Appl. Genet.* 75: 30-36.
94. Klein TM 1987 *Nature* 327:70-73.
95. James et al. *Plant Cell Reports* (1989) 7:658.
96. Grahamn et al. *Plant Cell, Tissue and Organ Culture* (1990) 20:35.
97. Thomas et al. *Plant Cell Reports* (1989) 8:354.
98. Srivastava et al., *Plant Cell Reports* (1988) 7:100.
99. Trulson et al., *Theor Appl. Genet.* (1986) 73:11.
100. Catlin et al., *Plant Cell Reports* (1988) 7:100.
101. Guri and Sink, *J Plant Physiol.* (1988) 133:52.

102. Michelmore et al., Plant cell reports (1987) 6:439.
103. Sheerman and Bevan Plant Cell reports (1989) 8:238.
104. Rech et al., Plant Cell Reports (1989) 8:33.
105. Nehra et al., Plant cell Reports (1990) 9:10.
- 5 106. McCormick et al., Plant Cell Reports (1986) 5:81.
107. McGranahan et al. Plant Cell Reports (1990) 8:512.
108. Fang et al., 86th Annual Meeting of the American Society for Horticultural Science Hort. Sci. (1989) 24:89.
109. Mathews et al., Symposium on Plant gene Transfer, UCLA J Cell Biochem
10 Suppl. 13D:255.
110. Mathews et al., J Cell Biochem Suppl. 13D:264.
111. Shimamoto et al., Nature (1989) 338:274.
112. Rhodes et al., Science (1988) 240:204.
113. de la Peña et al., Nature (1987) 325:274.
- 15 114. Pieterse, Plat Cell Tissue and Organ Culture (1989) 19:175.
115. Elmer et al., J. Amer. Soc. Hort. Sci., (1989) 114:1019.
116. Escalant and Teisson, Plant Cell Reports (1989) 7:665.
117. McClean and Grafton Plant Science (1989) 60:117.
118. Ochatt et al., Plant cell reports (1988) 7:393.
- 20 119. Matsuta and Hirabayashi. Plant Cell Reports (1989) 7:684.
120. deWald et al., J. Amer. Soc. Hort. Sci (1989) 114:712.
121. Moreno et al., Plnt Sci. Letters (1985) 34:195.
122. Roy and Mangat, Plant Science (1989) 60:77.
123. Lu et al., Plant Cell Reports (1989) 7:626.
- 25 124. Hidaka and Kajikura, Scientia Horticulturae (1988) 34:85.
125. Litz and Conover, Plant Science Letters (1982) 26:153.
126. Mante et al., Plant Cell Tissue and Organ Culture (1989) 19:1.
127. Chevreau et al., Plant Cell Reports (1989) 7:587.
128. DeWald et al., Plant Cell Reports (1988) 7:535.
- 30 129. Srivastava et al., Plant Cell Reports (1989) 8:300.
130. Redway et al., Plant Cell Reports (1990) 8:714.

131. Methods in Plant Molecular Biology and Biotechnology 269-284 (CRC Press, Boca Ratón, Florida).
132. Keller and Armstrong, Z. flanzenzucht 80:100-108, 1978.

Listado de secuencias**SEQ ID 1****Nombre del Promotor: pFM1****Localización: cromosoma 4; coordenadas 7290780 - 7291659****Gen correspondiente: locus At4g12250****Secuencia:**

catactagca	tgtatccaca	tagtattggt	cacattataa	gtagaaaatt	cttattaataa	60
aataaaaaaa	tcactacatt	ttttaatgca	atthttattgt	ttacctccta	aaaaacagat	120
caatttcata	aaataaaaaa	taactcgtaa	taaaattcct	caaaaataac	cgatttttcta	180
ttattttactt	ttttgtcaat	cgatataaat	tagaactaaa	aaaagttaat	cccgaagat	240
attattttata	tatatgacat	gatacgatcg	atthttatatac	tattttttatc	attattaccc	300
ccgttattttc	acattttctta	taaccttttg	ttattgtata	atthtttcttt	ttattatcaa	360
tcggaacaaa	gaaatacaga	agttaaaagc	aaaagtagta	aatattatta	aacaattatt	420
gcgtgtataaa	atagtgtata	tgtgatcgtg	tacaaaagta	catgcataac	atgatcaata	480
tagaatgcaa	tacaactata	caaatgtcaa	caatctacaa	acacgtttta	actttgaaaa	540
gcaatctaaa	attcatgaat	ctaaatatac	tctctggcag	ctttcttatt	accatatctt	600
aaagattgaa	ccaaataaat	aaatttacet	cattaaccaa	attaaacct	ccacatttaa	660
tttaattaca	acaagtaagt	aaaagatcgt	ctctttcttc	ttcccggaa	aatcctctca	720
tttttgactt	gctcaatttt	gcttccttaa	gacccaaaaa	tccaaacct	accctaactg	780
acattttctc	cgattttctaa	atcctctctc	ggtgaaaatg	tctcaccttg	atgatcttcc	840
ttctactccc	ggaaagtaca	aaaccgataa	agttccacc			879

SEQ ID 2**Nombre del Promotor: pES1****Localización: cromosoma 2; coordenadas 17132704 - 17133304****Gen correspondiente: locus At2g41050****Secuencia:**

cccggataat	tcttccaagt	ataaacccgt	cgactcattc	attcacgggt	ttcgtttctt	60
tcttctggaa	aagggtgaaa	cttttttctg	tttttttttt	ataataaaag	gtgacaactt	120
tctcttccat	tgatgtgtaa	acgaatagtc	gaatacgaat	tcgtttctct	taatgatta	180
gggaatcgat	ttctcttgct	ttttatgaaa	ccggatattt	caggctgcga	atagattatt	240
ttaacaatgg	tgttgatatg	tgaatgtcga	aaccagattt	tgtgtaacgt	tttcgatttt	300
gttgctttta	atggaagaat	ctgaatcctg	aggggtgtgat	ttgggttttg	ggggaatatt	360
cactttgtgg	ggtctttttg	tttgtgaatc	tagtatattg	aattattgtc	ggaattagaa	420
atctttttta	attctctagg	attcgttttc	tactctttca	aatgtggtca	ttctagtgtg	480
aatgtttttc	cttttctgct	tctaacactt	ttttctagtt	tgcacaatca	aaaagacggt	540
acctttgctc	attaagggac	ttgtttgatt	tttgggttta	acaggttttt	ggctgtgaag	600

SEQ ID 3**Nombre del promotor: pPo1****Localización: Cromosoma 3; Coordenadas: 3526561-3527612****Gen correspondiente: Locus At3g11260****Secuencia:**

```

aatggctttt tacaaatatt ttggtgtagg acttatatta tacatgtgtg tggcgaacct 60
tctggttcta tctattaaat gccttttctt tttgtagttt tttatgattc agaattcaga 120
tatttgattt tgtaatatta tatttttcta taaaagagaa taaaatttca aatttcctgt 180
ctacacttgc cgaccaacgt tcctcaagtg tcacatcttc tgtgcaagag agacaaaact 240
acaatctggt cacacaatgt catgatggtt ataattaatc ttatcatctt cattcatatc 300
aaactaatat cggccatcca tccgtctcaa taatgtaatt gcatacaatc tttgttttagg 360
acttgctagc tatttcaact tattattgat tacagattag atctacaatt tgtaaagaac 420
ttctaacttc aaaacaaata caaaaaaaag aaagtgaatg taaaatttct agtgtcataa 480
tttggaatag ttgaagcatt gaagttacaa tatcgaaact taaaatattg aactagata 540
gtacattacc accaaggctt tctgattatt cttatcatgt tatatgaata gtaagaaatt 600
acgtagaacg caatttaagt gtgtgcgagt caccacaaat taaaggagaa caaattcaat 660
gtttcaatcg ctggttccga tatacaactt atgcatgcca tggcatgaat gcgagatcca 720
attatatact aatataatac cacaaaaaca tatgtaaata tagatggaac agaagcctag 780
ataggttagg ataaagaaaa cgatcaaact tgcaaagatc agtctctccc aaatcccca 840
aaaaaacaaa gcatgcattt cgtaataaac aactcatcat aaaacgacac ataactcgaa 900
aacctctcct cccgacattt catcaattca tttctctttt tatttttcgaa aagatgaaaa 960
cttaataaat tattatgaac aatcttacta tatatatgga tatatacaca ggccctaaaa 1020
cgtaaaacag ttgaggactt tacatctgaa catg 1054

```

SEQ ID 4**Nombre del Promotor: pPo2****Localización: Cromosoma 4; Coordenadas: 17849276- 17849669****Gen correspondiente: Locus At4g37970****Secuencia:**

```

tatttggcct attttagttt ttcttttggt attaatgtaa aactaatcat tcgataagtt 60
catcaagttc ttcatacttt agagtatttt taattaaaaa caaaaagtg gccaaaactg 120
atztataaag catatagtta tatcaaatag tacaataatc acgatgcatg atatatattgt 180
tagtatatga acattataag taatatgttt atgttaaata tgtaaagaaa aaatacatat 240
atgtaagtca acttctgatt ggtatgagag acctaaagtc aaaacgatat ttctcaaagc 300
aaacgtcagc gtttagcccc atttatgttc tcaactcttt ctatataaaa agaaaggtag 360
tctagctcgc ttaattgttc gaaacaaagg gagtgagaga tg 402

```

Revindicaciones.

1. Una secuencia nucleotídica reguladora aislada y purificada a partir de Arabidopsis, y que tenga las siguientes características:
capaz de dirigir la expresión en una célula de planta en el gametofito masculino, el gametofito masculino, los gametos femeninos y/o los gametos masculinos; y que sea identificable a partir de una de las secuencias identificadas con el número de identificación SEQ ID 1, 2, 3, 4.
2. Una construcción de expresión que comprende una secuencia nucleotídica en acorde con la reivindicación 1, y que ha sido ligada de manera operativa a un gen estructural.
3. Un vector capaz de transformar o transfectar una célula huésped, vector que comprende una construcción de expresión como la descrita en la reivindicación 2.
4. El vector de la reivindicación 3 en el cual dicho vector es un plásmido.
5. El vector de la reivindicación 3 en el cual dicho vector es una componente viral.
6. Una célula eucariótica o procariótica transformada o transfectada con el vector descrito en la reivindicación 3.
7. La célula huésped de la reivindicación 6 en el cual dicha célula es una célula vegetal.
8. Una secuencia regulatoria purificada y aislada de Arabidopsis capaz de dirigir la expresión en una célula vegetal, y cuya secuencia se caracteriza por lo siguiente:
 - a) Dirige la expresión de una unidad transcripcional en el gametofito femenino, los gametos femeninos, el gametofito masculino y/o los gametos masculinos.
 - b) Se encuentra localizada en algunos de los siguientes intervalos cromosómicos:
En el cromosoma 2; entre las coordenadas nucleotídicas 17132704 y 17133304 (SEQ ID1)
En el cromosoma 3, entre las coordenadas nucleotídicas 3526561 y 3527612 (SEQ ID2).
En el cromosoma 4, entre las coordenadas nucleotídicas 17849276 y 17849669 (SEQ ID3)

En el cromosoma 4; entre las coordenadas nucleotídicas 7290780 y 7291659
(SEQ ID4)

9. Una construcción que comprenda una secuencia nucleotídica que corresponda a alguna de las descritas en la reivindicación 8, y operativamente ligada a un gen estructural.
10. Un vector capaz de transformar o transfectar una célula huésped, y que comprenda una construcción de expresión como la descrita en la reivindicación 9.
11. El vector de la reivindicación 10 en el cual dicho vector es un plásmido.
12. El vector de la reivindicación 10 en el cual dicho vector es una componente viral.
13. Una célula eucariótica o procariótica transformada o transfectada con el vector descrito en la reivindicación 10.
14. La célula huésped descrita en la reivindicación 13 en el cual dicha célula es una célula vegetal.
15. Una secuencia regulatoria purificada y aislada de Arabidopsis capaz de dirigir la expresión en una célula vegetal, y cuya secuencia se caracteriza por lo siguiente:
 - a) Dirige la expresión de una unidad transcripcional en el gametofito femenino, los gametos femeninos, el gametofito masculino y/o los gametos masculinos.
 - b) Se caracteriza por conferir expresión específica a la unidad transcripcional que contenga las secuencias genómicas identificadas con el número de identificación SEQ ID 1, 2, 3, 4, o que contenga secuencias estrechamente ligadas (secuencias que incluyan o sean vecinas) a las secuencias genómicas identificadas con el número de identificación SEQ ID 1, 2, 3, 4.
16. Una construcción de expresión que comprenda una secuencia nucleotídica como la descrita en la reivindicación 15, y que sea operativamente ligada a un gen estructural.
17. Un vector capaz de transformar o transfectar una célula huésped, y que comprenda una construcción de expresión como la descrita en la reivindicación 16.
18. El vector de la reivindicación 17 en el cual dicho vector es un plásmido.

19. El vector de la reivindicación 17 en el cual dicho vector es una componente viral.
20. Una célula eucariótica o procariótica transformada o transfectada con el vector descrito en la reivindicación 17.
- 5 21. La célula huésped descrita en la reivindicación 13 en el cual dicha célula es una célula vegetal.
22. Un método para identificar una secuencia regulatoria que sea capaz de dirigir la expresión al gametofito femenino, los gametos femeninos, el gametofito masculino, y/o los gametos masculinos de las plantas y que comprenda la
10 identificación de una secuencia regulatoria a partir de una secuencia de etiquetamiento seleccionado a partir del grupo de secuencias genómicas identificadas con el número de identificación SEQ ID 1, 2, 3, 4.
23. El método descrito en la reivindicación 22 y que comprende la identificación de
15 unidad transcripcional vecina a cualquiera de las regiones identificadas con el número de identificación SEQ ID 1, 2, 3, 4.
24. Una composición nucleotídica novedosa de Arabidopsis que comprenda una secuencia seleccionada partir del grupo que comprende las secuencias identificadas con el número de identificación SEQ ID 1, 2, 3, 4.

Resumen.

El invento presenta la localización y los patrones de expresión de promotores (secuencias nucleotídicas regulatorias) que actúan de manera específica en el gametofito femenino, los gametos femeninos, el gametofito masculino, y/o los gametos masculinos de *Arabidopsis thaliana*. Se presentan también los métodos que fueron utilizados para la identificación de dichos promotores así como los métodos para utilizar estos promotores en construcciones de expresión y manipular el desarrollo reproductivo de las plantas. Los promotores aislados e identificados de la invención, permiten regular la expresión génica de genes seleccionados en plantas, de manera que tales genes puedan ser eficientemente transcritos en las células reproductivas femeninas o masculinas de éstas. Así mismo, dichos promotores pueden utilizarse para impedir la transferencia de rasgos transgénicos a las plantas a través del polen en campo, para proporcionar nutrientes adicionales a la semilla y para conferir resistencia a patógenos y enfermedades en la semilla.

FIGURA 1

CA	TACTAGCATG	TATCCACATA	GTATTGTTCA	CATTATAAGT	AGAAAATTCT
70	80	90	100	110	120
TATTAAAAAA	TAAAAAATC	ACTACATTTT	TTAATGCAAT	TTTATTGTTT	ACCTCCTAAA
130	140	150	160	170	180
AAACAGATCA	ATTTTCATAAA	ATAAAAAATA	ACTCGTAATA	AAATTCCTCA	AAAATAACCG
190	200	210	220	230	240
ATTTTGTATT	ATTTACTTTT	TTGTCAATCG	ATATAAATTA	GAACTAAAAA	AAGTTAATCC
250	260	270	280	290	300
CGAAAGATAT	TATTTATATA	TATGACATGA	TACGATCGAT	TTTATATCTA	TTTTTATCAT
310	320	330	340	350	360
TATTACCCCC	GTTATTTTAC	ATTTCTTATA	ACCTTTTGTT	ATTGTATAAT	TTTTCTTTTT
370	380	390	400	410	420
ATTATCAATC	GGAACAAAGA	AATACAGAAG	TTAAAAGCAA	AAGTAGTAAA	TATTATTAAA
430	440	450	460	470	480
CAATTATTGC	GTGTAAAAAT	AGTGTATATG	TGATCGTGTA	CAAAAGTACA	TGCATAACAT
490	500	510	520	530	540
GATCAATATA	GAATGCAATA	CAACTATACA	AATGTCAACA	ATCTACAAAC	ACGTTTTAAC
550	560	570	580	590	600
TTTGAAAAGC	AATCTAAAAT	TCATGAATCT	AAATATACTC	TCTGGCAGCT	TTCTTATTAC
610	620	630	640	650	660
CATATCTTAA	AGATTGAACC	AAATAAATAA	ATTTACATCA	TTAACCAAAT	TAAACCATCC
670	680	690	700	710	720
ACATTTAATT	TAATTACAAC	AAGTAAGTAA	AAGATCGTCT	CTTTCTTCTT	TCCCGGAAAA
730	740	750	760	770	780
TCCTTCTATT	TTTGACTTGC	TCAATTTTGC	TTCTTAAGA	CCCAAAAATC	CAAACCTAAC
790	800	810	820	830	840
CCTAATCGAC	ATTTTCTCCG	ATTTCTAAAT	CCTCCTCCGG	TGAAAATGTC	TCACCTTGAT
850	860	870	880	890	900
GATCTTCCTT	CTACTCCCGG	AAAGTACAAA	ACCGATAAAG	TTCCACC	

FIGURA 2

CCCGGATAAT	TCTTCCAAGT	ATAAACCCGT	CGACTCATT	ATTCACGGGT	TTCGTTTCTT
70	80	90	100	110	120
TCTTCTGGAA	AAGGTGACAA	CTTTTTTCTG	TTTTTTTTTT	ATAATAAAAG	GTGACAACCT
130	140	150	160	170	180
TCTCTTCCAT	TGATGTGTAA	ACGAATAGTC	GAATACGAAT	TCGTTTCTCT	TAATGTATTA
190	200	210	220	230	240
GGGAATCGAT	TTCTCTTGCT	TTTTATGAAA	CCGGATATTT	CAGGCTGCGA	ATAGATTATT
250	260	270	280	290	300
TTAACAATGG	TGTTGATATG	TGAATGTCGA	AACCAGATTT	TGTGTAACGT	TTTCGATTTT
310	320	330	340	350	360
GTTGCTTTTA	ATGGAAGAAT	CTGAATCCTG	AGGGTGTGAT	TTGGGTTTTG	GGGGAATATT
370	380	390	400	410	420
CACTTTGTGG	GGTCTTTTTG	TTTGTGAATC	TAGTATATTG	AATTATTGTC	GGAATTAGAA
430	440	450	460	470	480
ATCTTTTTTA	ATTCTCTAGG	ATTCGTTTTC	TACTCTTTCA	AATGTGGTCA	TTCTAGTGTA
490	500	510	520	530	540
AATGTTTTTC	CTTTTCGTCT	TCTAACACTT	TTTTCTAGTT	TGCACAATCA	AAAAGACGTT
550	560	570	580	590	600
ACCTTTGCTC	ATTAAGGGAC	TTGTTTGATT	TTTGGGTTTA	ACAGGTTTTT	GGCTGTGAAG

FIGURA 3

AATGGCTTTTTACAAATATTTTGGTGTAGGACTTATATTATACATGTGTGTGGCGAACCTTCTGGTTCTATCTATTAAAT
GCCTTTTCTTTTTGTAGTTTTTATGATTCAGAATTCAGATATTTGATTTTGTAAATATTATATTTTCTTATAAAAGAGAA
TAAAATTTCAAATTTCTGTCTACACTTGCCGACCAACGTTCTCAAGTGTACATCTTCTGTGCAAGAGAGACAAAAC
ACAATCTGTTACACAATGTCATGATGTTTATAATTAATCTTATCATCTTCATTCAATCAAACATAATATCGGCCATCCA
TCCGTCTCAATAATGTAATTGCATACAATCTTGTFTTAGGACTTGCTAGCTATTTCAACTTATTATTGATTACAGATTAG
ATCTACAATTTGTAAAGAACTTCTAACTTCAAAACAAATACAAAAAAGAAAAGTGAATGTAAAATTTCTAGTGTATAA
TTTGGAATAGTTGAAGCATTGAAGTTACAATATCGAACGTTAAAATATTGACACTAGATAGTACATTACCACCAAGGCTT
TCTGATTATTCTTATCATGTTATATGAATAGTAAGAAATTACGTAGAACGCAATTTAAGTGTGTGCGAGTCACCACAAAT
TAAAGGAGAACAAATTCATGTTTCAATCGCTGGTTCGGATATACAACCTTATGCATGCCATGGCATGAATGCGAGATCCA
ATTATATACTAATATATACACACAAAAACATATGTAAATATAGATGGAACAGAAGCCTAGATAGGTTAGGATAAAGAAAA
CGATCAAATCTGCAAAGATCAGTCTCTCCCAAATCCCCAAAAAACAAGCATGCATTTTCGTAATAAACAACTCATCAT
AAAACGACACATAACTCGAAAACCTCTCTCCCGACATTTTATCAATTCATTTCTTTTTATTTTCGAAAAGATGAAAA
CTTAATAAATTTATTATGAACAATCTTACTATATATATGGATATATACACAGGCCCTAAAACGTAAAACAGTTGAGGACTT
TACATCTGAACATG

FIGURA 4

TATTTGGCCTATTTTAGTTTTCTTTTGTATTAAATGTAAAATAATCATTGATAAGTTCATCAAGTTCTTCATACTTT
AGAGTATTTTAAATTAACAAAAAGTGGCCAAAAGTATTTATAAAGCATATAGTTATATCAAATAGTACAATAATC
ACGATGCATGATATATTTGTTAGTATATGAACATTATAAGTAATATGTTTATGTTAAATATGTTAAGAAAAATACATAT
ATGTAAGTCAACTTCTGATTGGTATGAGAGACCTAAAGTCAAAACGATATTTCTCAAACGAAACGTCAGCGTTTAGCCCC
ATTTATGTTCTCACTCTTTTCTATATAAAAAAGAAAGGTACTCTAGCTCGCTTAATTGTTTCAAACAAGGGAGTGAGAGA
TG

FIGURA 5

10	20	30	40	50	60
TAAAGCTTCA	TACTAGCATG	TATCCACATA	GTATTGTTCA	CATTATAAGT	AGAAAATTCT
70	80	90	100	110	120
TATTAAAAAA	TAAAAAATC	ACTACATTTT	TTAATGCAAT	TTTATTGTTT	ACCTCCTAAA
130	140	150	160	170	180
AAACAGATCA	ATTTCATAAA	ATAAAAAATA	ACTCGTAATA	AAATTCCTCA	AAAATAACCG
190	200	210	220	230	240
ATTTTGTATT	ATTTACTTTT	TTGTCAATCG	ATATAAATTA	GAACTAAAAA	AAGTTAATCC
250	260	270	280	290	300
CGAAAGATAT	TATTTATATA	TATGACATGA	TACGATCGAT	TTTATATCTA	TTTTTATCAT
310	320	330	340	350	360
TATTACCCCC	GTTATTTTAC	ATTTCTTATA	ACCTTTTGTT	ATTGTATAAT	TTTTCTTTTT
370	380	390	400	410	420
ATTATCAATC	GGAACAAAGA	AATACAGAAG	TTAAAAGCAA	AAGTAGTAAA	TATTATTAAA
430	440	450	460	470	480
CAATTATTGC	GTGTAAAAAT	AGTGTATATG	TGATCGTGTA	CAAAGTACA	TGCATAACAT
490	500	510	520	530	540
GATCAATATA	GAATGCAATA	CAACTATACA	AATGTCAACA	ATCTACAAAC	ACGTTTTAAC
550	560	570	580	590	600
TTTGAAAAGC	AATCTAAAAT	TCATGAATCT	AAATATACTC	TCTGGCAGCT	TTCTTATTAC
610	620	630	640	650	660
CATATCTTAA	AGATTGAACC	AAATAAATAA	ATTTACATCA	TTAACCAAAT	TAAACCATCC
670	680	690	700	710	720
ACATTTAATT	TAATTACAAC	AAGTAAGTAA	AAGATCGTCT	CTTTCTTCTT	TCCCGGAAAA
730	740	750	760	770	780
TCCTTCTATT	TTTGACTTGC	TCAATTTTGC	TTCCCTTAAGA	CCCAAAAATC	CAAACCTAAC
790	800	810	820	830	840
CCTAATCGAC	ATTTTCTCCG	AFTTCTAAAT	CCTCCTCCGG	TGAAAATGTC	TCACCTTGAT
850	860	870	880	890	900
GATCTTCCTT	CTACTCCCGG	AAAGTACAAA	ACCGATAAAG	TTCCACCGGA	TCCTA.

FIGURA 6

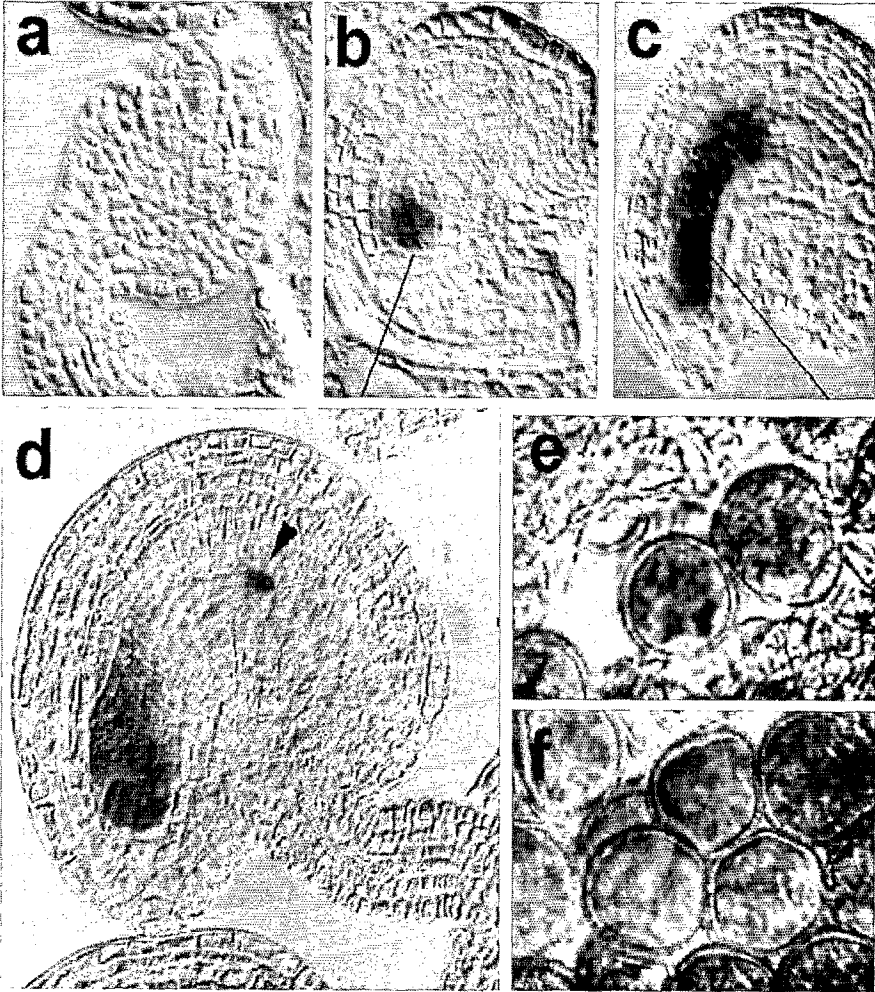


FIGURA 7

10	20	30	40	50	60
<u>CCCGGATAAT</u>	<u>TCTTCCAAGT</u>	ATAAACCCGT	CGACTCATT	ATTCACGGGT	TTCGTTTCTT
70	80	90	100	110	120
TCTTCTGGAA	AAGGTGACAA	CTTTTTTCTG	TTTTTTTTTT	ATAATAAAAG	GTGACAACTT
130	140	150	160	170	180
TCTCTTCCAT	TGATGTGTAA	ACGAATAGTC	GAATACGAAT	TCGTTTCTCT	TAATGTATTA
190	200	210	220	230	240
GGAATCGAT	TTCTCTTGCT	TTTTATGAAA	CCGGATATTT	CAGGCTGCCA	ATAGATTATT
250	260	270	280	290	300
TTAACAATGG	TGTTGATATG	TGAATGTCGA	AACCAGATTT	TGTGTAACGT	TTTCGATTTT
310	320	330	340	350	360
GTTGCTTTTA	ATGGAAGAAT	CTGAATCCTG	AGGGTGTGAT	TTGGGTTTTG	GGGGAATATT
370	380	390	400	410	420
CACTTTGTGG	GGTCTTTTTG	TTTGTGAATC	TAGTATATTG	AATTATTGTC	GGAATTAGAA
430	440	450	460	470	480
ATCTTTTTTA	ATTCTCTAGG	ATTCGTTTTC	TACTCTTTCA	AATGTGGTCA	TTCTAGTGTA
490	500	510	520	530	540
AATGTTTTTC	CTTTTCGTCT	TCTAACACTT	TTTTCTAGTT	TGCACAATCA	AAAAGACGTT
550	560	570	580	590	600
ACCTTTGCTC	ATTAAGGGAC	TTGTTTGATT	TTTGGGTTTA	<u>ACAGGTTTTT</u>	<u>GGCTGTGAAG</u>

FIGURA 8

AATGGCTTTTTACAAATATTTTGGTGTAGGACTTATATTATACATGTGTGTGGCGAACCTTCTGGTCTATCTATTAAAT
GCCTTTTCTTTTTGTAGTTTTTTATGATTCAGAATTCAGATATTTGATTTTGTAATATTATATTTTCTTATAAAAAGAGAA
TAAAATTTCAAATTTCTGTCTACACTTGCCGACCAACGTTCTCAAGTGTACATCTTCTGTGCAAGAGAGACAAAAC
ACAATCTGTTACACAATGTCATGATGTTTATAATTAATCTTATCATCTTCATTCATATCAAACATAATATCGGCCATCCA
TCCGTCTCAATAATGTAATTGCATACAATCTTTGTTTAGGACTTGCTAGCTATTTCAACTTATTATTGATTACAGATTAG
ATCTACAATTTGTAAAGAACTTCTAACTTCAAACAAATACAAAAAAGAAAGTGAATGTAAAATTTCTAGTGTCAATA
TTTGGAAATAGTTGAAGCATTGAAGTTACAATATCGAACGTTAAAATATTGACACTAGATAGTACATTACCACCAAGGCTT
TCTGATTATTCTTATCATGTTATATGAATAGTAAGAAATTACGTAGAACGCAATTTAAGTGTGTGCGAGTCACCACAAAT
TAAAGGAGAACAAATTTCAATGTTTCAATCGCTGGTTCGGATATACAACCTTATGCATGCCATGGCATGAATGCGAGATCCA
ATTATATACTAATATATACACACAAAAACATATGTAAATATAGATGGAACAGAAGCCTAGATAGGTTAGGATAAAGAAAA
CGATCAAATCTGCAAAGATCAGTCTCTCCCAAATCCCCAAAAAACAAGCATGCAATTTCTGTAATAAACAACTCATCAT
AAAACGACACATAACTCGAAAACCTCTCTCCCGACATTTTCAATTCATTTCTCTTTTTATTTTCGAAAAGATGAAAA
CTTAATAAATTATTATGAACAATCTTACTATATATATGGATATATACACAGGCCCTAAAACGTAAAACAGTTGAGGACTT
TACATCTGAACATG

FIGURA 9

TATTTGGCCTATTTTAGTTTTCTTTTGTATTAAATGTAAAACTAATCATTTCGATAAGTTCATCAAGTTCCTTCATACTTT
AGAGTATTTTAAATTA AAAACAAAAAGTGGCCAAAAGTATTTATAAAGCATATAGTTATATCAAATAGTACAATAATC
ACGATGCATGATATATTTGTTAGTATATGAACATTATAAGTAATATGTTTATGTTAAATATGTTAAGAAAAATACATAT
ATGTAAGTCAACTTCTGATTGGTATGAGAGACCTAAAAGTCAAAACGATATTTCTCAAACGAAACGTCAGCGTTTAGCCCC
ATTTATGTTCTCACTCTTTTCTATATAAAAAGAAAGGTACTCTAGCTCGCTTAATTGTTCGAAACAAAGGGAGTGAGAGA
TG

FIGURA 10

