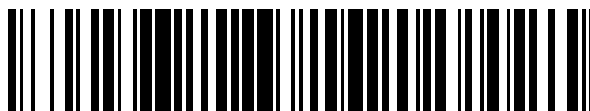


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 398 142**

51 Int. Cl.:

**C07K 14/415** (2006.01)

**C12N 15/82** (2006.01)

**A01H 5/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **17.01.2007 E 07797081 (2)**

97 Fecha y número de publicación de la concesión europea: **10.10.2012 EP 1973935**

54 Título: **Elementos de control de la expresión de la familia de Lemnaceae**

30 Prioridad:

**17.01.2006 US 759308 P**

**03.10.2006 US 848961 P**

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

**13.03.2013**

73 Titular/es:

**SYNTHON BIOPHARMACEUTICALS B.V. (100.0%)  
Microweg 22 PO Box 7071  
6503 GN Nijmegen, NL**

72 Inventor/es:

**DICKEY, LYNN F.;  
COX, KEVIN M. y  
PEELE, CHARLES G.**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 398 142 T3**

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

**DESCRIPCIÓN**

Elementos de control de la expresión de la familia de Lemnaceae

Campo de la invención

La presente invención se refiere a composiciones y métodos para intensificar la expresión génica en plantas.

5 Antecedentes de la invención

Las lentejas de agua son los únicos miembros de la familia de monocotiledóneas Lemnaceae. Los cinco géneros y 38 especies son todas plantas de agua dulce, que flotan libremente, pequeñas cuyo ámbito geográfico abarca todo el mundo (Landolt (1986) Biosystematic Investigation on the Family of Duckweeds: The Family of Lemnaceae - A Monograph Study (Geobotanischen Institut ETH, Stiftung Rubel, Zurich)). Aunque son las plantas morfológicamente más reducidas conocidas, la mayor parte de las lentejas de agua tienen todos los tejidos y órganos de plantas mucho más grandes, incluyendo raíces, tallos, flores, semillas y frondas. Las especies de lentejas de agua se han estudiado exhaustivamente y existe una bibliografía sustancial que detalla su ecología, sistemática, ciclo de vida, metabolismo, susceptibilidad a enfermedades y plagas, su biología reproductiva, estructura genética, y biología celular (Hillman (1961) Bot. Review 27:221; Landolt (1986) Biosystematic Investigation on the Family of Duckweeds: The Family of Lemnaceae - A Monograph Study (Geobotanischen Institut ETH, Stiftung Rubel, Zurich)).

El hábito de crecimiento de las lentejas de agua es ideal para los métodos de cultivo microbiano. La planta proliferan rápidamente a través de brotación vegetativa de nuevas frondas, de una manera macroscópica análoga a la propagación asexual en levadura. Esta proliferación se produce por medio de brotación vegetativa de células meristemáticas. La región meristemática es pequeña y se encuentra sobre la superficie ventral de la fronda. Las células meristemáticas se encuentran en dos bolsillos, uno a cada lado del nervio central de la fronda. La pequeña región del nervio central también es el lugar a partir del cual se origina la raíz y surge el tallo que conecta cada fronda con su fronda madre. El bolsillo meristemático está protegido por una solapa de tejido. Las frondas brotan alternativamente de estos bolsillos. El tiempo de duplicación varía entre especies y es tan breve como 20-24 horas (Landolt (1957) Ber. Schweiz. Bot. Ges. 67:271; Chang et al. (1977) Bull. Inst. Chem. Acad. Sin. 24:19; Datko y Mudd (1970) Plant. Physiol. 65:16; Venkataraman et al. (1970) Z. Pflanzenphysiol. 62: 316).

El cultivo intensivo de la lenteja de agua da como resultado las mayores tasas de acumulación de biomasa por unidad de tiempo (Landolt y Kandler (1987) The Family of Lemnaceae - A Monographic Study Vol. 2: Phytochemistry, Physiology, Application, Bibliography (Veröffentlichungen des Geobotanischen Institutes ETH, Stiftung Rubel, Zurich)), oscilando la acumulación de peso seco de 6 a 15% del peso húmedo (Tillberg et al. (1979) Physiol. Plant. 46:5; Landolt (1957) Ben. Schweiz. Bot. Ges. 67:271; Stomp, datos no publicados). Se ha informado de que el contenido de proteína de numerosas especies de lenteja de agua desarrolladas en condiciones variables oscila de 15 a 45% de peso seco (Chang et al. (1977) Bull. Inst. Chem. Acad. Sin. 24:19; Chang y Chui (1978) Z. Pflanzenphysiol. 89:91; Porath et al. (1979) Aquatic Botany 7:272; Appenroth et al. (1982) Biochem. Physiol. Pflanz. 177:251). Utilizando estos valores, el nivel de producción de proteína por litro de medio en la lenteja de agua es del mismo orden de magnitud que los sistemas de expresión génica de levaduras.

Los cultivos de plantas de lenteja de agua de o nódulos de lenteja de agua se pueden transformar eficazmente con una casete de expresión que contiene una secuencia de nucleótidos de interés por medio de cualquiera de los numerosos métodos incluyendo la transferencia génica mediada por Agrobacterium, el bombardeo balístico, o la electroporación. Los transformantes de lenteja de agua estables se pueden aislar transformando las células de la lenteja de agua tanto con la secuencia de nucleótidos de interés como con un gen que confiere resistencia a un agente de selección, seguido de cultivo de las células transformadas en un medio que contiene el agente de selección. Véase la Patente de los Estados Unidos Núm. 6.040.498 de Stomp et al.

El análisis por delección del promotor rbcS de lenteja de agua fue descrito por Rolfe et al. (1991) en Proc. Natl. Acad. Sci. USA 88:2683.

Un sistema de expresión génica de lenteja de agua proporciona la tecnología esencial que sería útil para numerosas aplicaciones de investigación y comerciales. Para la investigación sobre la biología molecular vegetal en conjunto, un sistema vegetal diferenciado que se puede manipular con la facilidad de la levadura en el laboratorio proporciona un sistema muy rápido en el cual analizar los papeles evolutivos y fisiológicos de genes aislados. Para la producción comercial de polipéptidos valiosos, un sistema basado en la lenteja de agua tiene numerosas ventajas sobre los sistemas de cultivo microbianos o celulares existentes. Las plantas demuestran un procesamiento post-traduccional que es similar al de las células de mamíferos, superando un problema principal asociado con la producción en células microbianas de polipéptidos de mamífero biológicamente activos, y ha sido demostrado por otros que los sistemas vegetales tienen la capacidad de ensamblar proteínas de múltiples subunidades, una capacidad de la que a menudo carecen los sistemas microbianos (Hiatt (1990) Nature 334:469). El aumento a escala de la biomasa de lenteja de agua a los niveles necesarios para la producción comercial de proteínas recombinantes es más rápido y tiene un coste más eficaz que el aumento a escala similar de las células de mamífero, y a diferencia de otros

sistemas de producción vegetal sugeridos, por ejemplo, soja y tabaco, la lenteja de agua se puede desarrollar en recipientes de producción de biomasa completamente integrados y controlados, haciendo bastante más fácil la integración del sistema en la infraestructura industrial de producción de proteínas existente.

5 Por consiguiente, sigue existiendo la necesidad de composiciones y métodos optimizados para expresar proteínas de interés en lenteja de agua.

Breve resumen de la invención

10 Se proporcionan composiciones y métodos para la regulación de la expresión génica en una planta. Las composiciones incluyen secuencias de nucleótidos novedosas para elementos de control de la expresión (p. ej., promotores e intrones) aislados del gen de ubicuitina de Lemnaceae. Los elementos para el control de la expresión de la invención inician la transcripción de las secuencias de nucleótidos heterólogas conectadas operablemente en plantas. Más concretamente, las composiciones de la invención incluyen el elemento para el control de la expresión mostrado en el SEQ ID NO: 1, y las variantes y fragmentos del mismo.

15 También se proporciona un péptido señal de quitinasa de Lemnaceae novedoso mostrado en el SEQ ID NO: 16, codificado por una secuencia mostrada en el SEQ ID NO: 15, y las variantes y fragmentos del mismo. La secuencia que codifica el péptido señal se pueden conectar operablemente a una secuencia codificante para un polipéptido de interés para dirigir la secreción extracelular del polipéptido codificado.

20 Se proporcionan constructos de expresión (p. ej., casetes y vectores) que comprenden un elemento de control de la expresión de la invención conectado operablemente a una secuencia de nucleótidos heteróloga de interés. Se proporcionan adicionalmente plantas, células vegetales y nódulos transformados establemente que tienen un constructo de expresión de la invención.

25 Las composiciones de la invención encuentran uso en los métodos dirigidos a la expresión de secuencias de nucleótidos de interés en una planta o célula o nódulo vegetal. Los métodos de la invención incluyen la introducción en una planta o célula o nódulo vegetal de un constructo de expresión que tiene un elemento de control de la expresión de ubicuitina de Lemnaceae (p. ej., como se muestra en el SEQ ID NO: 1), o una variante o fragmento del mismo, conectado operablemente a una secuencia de nucleótidos de interés.

Los métodos incluyen la introducción en una planta o célula o nódulo vegetal de un constructo de expresión que tiene una secuencia que codifica un péptido señal de quitinasa de Lemnaceae (p. ej., como se muestra en el SEQ ID NO: 15), o una variante o fragmento del mismo, conectado operablemente a la secuencia codificante de un polipéptido de interés.

30 En algunas realizaciones, los métodos de la invención están dirigidos a la producción de un polipéptido codificado por una secuencia de nucleótidos de interés en un sistema de expresión vegetal (p. ej., un sistema de expresión de lenteja de agua). El sistema de expresión vegetal de la presente invención se optimiza para producir elevados niveles de la secuencia de polipéptidos de interés. De este modo, la invención incluye métodos para la expresión de una secuencia de nucleótidos de interés en plantas que están transformadas con constructos de expresión para la expresión de la secuencia de nucleótidos de interés, donde estas secuencias de nucleótidos se modifican para intensificar su expresión en plantas.

35 Estos y otros aspectos de la invención se describen con más detalle en la descripción de la invención dada más abajo.

Descripción detallada de la invención

40 La presente invención proporciona composiciones y métodos dirigidos a ácidos nucleicos novedosos para elementos de control de la expresión en plantas que regulan la transcripción de secuencias de nucleótidos heterólogas en plantas. Específicamente, las composiciones de la invención comprenden elementos de control de la expresión aislados del gen de la ubicuitina de Lemnaceae, incluyendo el elemento para el control de la expresión mostrado en el SEQ ID NO: 1, y las variantes y fragmentos del mismo, que se definen más abajo en la presente memoria. La secuencia del promotor individual (SEQ ID NO: 4) en este elemento para el control de la expresión también encuentra uso en la regulación de la transcripción en plantas. Asimismo se proporciona un péptido señal de quitinasa de L. minor novedoso (SEQ ID NO: 16) y la correspondiente secuencia codificante (SEQ ID NO: 15), y las variantes y fragmentos de la misma.

45 Según se utiliza en la presente memoria, "ácido nucleico" incluye la referencia a un polímero desoxirribonucleotídico o ribonucleotídico en forma de hebra sencilla o doble, y a menos que se limite de otro modo, incluye análogos conocidos (p. ej., ácidos péptido nucleicos) que tienen la naturaleza esencial de los nucleótidos naturales ya que hibridan con ácidos nucleicos de hebra sencilla de una manera similar a los nucleótidos de origen natural.

50 La invención incluye composiciones de ácidos nucleicos aisladas o sustancialmente purificadas. Una molécula de ácido nucleico "aislada" o "purificada" está sustancialmente o esencialmente libre de los componentes que

normalmente lo acompañan o que interaccionan con la molécula de ácido nucleico o la proteína encontrada en su entorno natural. De este modo, una molécula de ácido nucleico aislada o purificada está sustancialmente libre de otro material celular, o medio de cultivo cuando es producido mediante técnicas recombinantes, o sustancialmente libre de precursores químicos u otros agentes químicos cuando es sintetizada químicamente. Preferiblemente, un ácido nucleico "aislado" está libre de secuencias (preferiblemente secuencias que codifican proteínas) que naturalmente flanquean el ácido nucleico (esto es, secuencias localizadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo del cual deriva el ácido nucleico. Por ejemplo, en diferentes realizaciones, la molécula de ácido nucleico aislada puede contener menos de aproximadamente 5 kb, 4 kb, 3kb, 2 kb, 1 kb, 0,5 kb, o 0,1 kb de secuencias de nucleótidos que flanquean naturalmente la molécula de ácido nucleico en el ADN genómico de la célula de la cual deriva el ácido nucleico.

Las composiciones de la invención incluyen moléculas de ácido nucleico aisladas que comprenden secuencias de nucleótidos del elemento de control de la expresión mostrado en el SEQ ID NO: 1, y las variantes y fragmentos del mismo, que se definen en la presente memoria más abajo. Por "elemento de control de la expresión" se entiende una región reguladora de ADN que comprende normalmente una caja TATA capaz de dirigir la ARN polimerasa II para iniciar la síntesis de ARN en el sitio de inicio de la transcripción apropiado para una secuencia codificante concreta. Un elemento de control de la expresión puede comprender adicionalmente otras secuencias de reconocimiento generalmente situadas aguas arriba o 5' con respecto a la caja TATA, que influyen (p. ej., intensifican) en la velocidad de inicio de la transcripción. Además, un elemento de control de la expresión puede comprender adicionalmente secuencias situadas generalmente aguas abajo o 3' con respecto a la caja TATA, que influyen (p. ej., intensifican) en la velocidad de inicio de la transcripción.

Se reconoce que al haber identificado las secuencias de nucleótidos para las regiones de los elementos de control de la expresión descritas en la presente memoria, está dentro del estado de la técnica aislar e identificar elementos reguladores adicionales en la región no traducida 5' (UTR) aguas arriba de las regiones de los elementos de control de la expresión concretas identificadas en la presente memoria. De este modo, por ejemplo, las regiones de los elementos de control de la expresión descritas en la presente memoria pueden comprender adicionalmente elementos reguladores adicionales tales como los responsables de la expresión en tejidos y temporal de la secuencia codificante, intensificadores, y similares. Véanse concretamente la Patente Australiana Núm. AU-A-77751/94 y las Patentes de los Estados Unidos Núms. 5.466.785 y 5.635.618.

Los elementos de control de la expresión fueron aislados de los genes de ubicuitina, r-histona y quitinasa para varios miembros de la familia Lemnaceae, y de ese modo son referidos como "elementos de control de la expresión de Lemnaceae". El SEQ ID NO: 1 muestra el elemento de control de la expresión de ubicuitina de Lemna minor completo, incluyendo tanto el promotor plus 5' UTR (nucleótidos 1-1625) como el intrón (nucleótidos 1626-2160). El SEQ ID NO: 2 muestra el elemento de control de la expresión de ubicuitina de Spirodella polyrrhiza completo, incluyendo tanto el promotor plus 5' UTR (nucleótidos 1-1041) como el intrón (nucleótidos 1042-2021). El SEQ ID NO: 3 muestra el elemento de control de la expresión de ubicuitina de Lemna aequinoctialis completo, incluyendo tanto el promotor plus 5' UTR (nucleótidos 1-964) como el intrón (nucleótidos 965-2068). El SEQ ID NO: 4 muestra la porción plus 5' UTR del promotor del elemento de control de la expresión de ubicuitina de L. minor. El SEQ ID NO: 5 muestra la porción plus 5' UTR del promotor del elemento de control de la expresión de ubicuitina de S. polyrrhiza. El SEQ ID NO: 6 muestra la porción plus 5' UTR del promotor del elemento de control de la expresión de ubicuitina de L. aequinoctialis. El SEQ ID NO: 7 muestra la porción del intrón del elemento de control de la expresión de ubicuitina de L. minor. El SEQ ID NO: 8 muestra la porción del intrón del elemento de control de la expresión de ubicuitina de S. polyrrhiza. El SEQ ID NO: 9 muestra la porción del intrón del elemento de control de la expresión de ubicuitina de L. aequinoctialis.

El SEQ ID NO: 13 muestra el elemento de control de la expresión de r-histona de Lemna minor completo, incluyendo el promotor plus 5' UTR. El SEQ ID NO: 14 muestra el elemento de control de la expresión de quitinasa de Lemna minor completo, incluyendo el promotor plus 5' UTR. El SEQ ID NO: 15 muestra la secuencia codificante del péptido señal de quitinasa de L. minor. El SEQ ID NO: 16 muestra el péptido señal de quitinasa de L. minor.

Se considera que las secuencias del promotor plus 5' UTR individuales mostradas en los SEQ ID NO: 4-6, 13 y 14, y las variantes y fragmentos biológicamente activos de las mismas, se pueden utilizar para regular la transcripción de secuencias de nucleótidos de interés conectadas operablemente en plantas. De un modo similar, una o más secuencias de intrones mostradas en los SEQ ID NO: 7-9, y los fragmentos o variantes biológicamente activos de las mismas, pueden ser conectados operablemente a un promotor de interés, incluyendo un promotor mostrado en los SEQ ID NO: 4, 5, 6, 13, o 14 con el fin de intensificar la expresión de una secuencia de nucleótidos que está conectada operablemente a ese promotor.

También se describen los fragmentos y variantes de los elementos de control de la expresión descritos, la secuencia codificante del péptido señal, y el péptido señal codificado. Por "fragmento" en el contexto de un elemento de control de la expresión se entiende una porción del elemento de control de la expresión completo, tal como una porción de uno cualquiera de los elementos de control de la expresión mostrados en los SEQ ID NO: 1-3, 13 y 14. Los fragmentos de un elemento de control de la expresión conservan la actividad biológica y por lo tanto incluyen fragmentos capaces de iniciar o intensificar la expresión de una secuencia de nucleótidos conectada operablemente.

De este modo, por ejemplo, se puede utilizar menos de un elemento de control de la expresión completo descrito en la presente memoria para conducir la expresión de una secuencia de nucleótidos de interés conectada operablemente, tal como una secuencia de nucleótidos que codifica una proteína heteróloga. Los ejemplos no limitantes, específicos de tales fragmentos de un elemento de control de la expresión incluyen: (i) las secuencias de nucleótidos mostradas en uno cualquiera de los SEQ ID NO: 4-9 (como se describe en la presente memoria más arriba); (ii) truncamientos 5' del elemento de control de la expresión de ubicuitina de *L. minor* (SEQ ID NO: 1), tales como los nucleótidos 1288-2160 del SEQ ID NO: 1 (promotor truncado LmUbq Núm. 1, como el que se encuentra en el constructo Egs22 descrito más adelante en la presente memoria) y los nucleótidos 1132-2160 del SEQ ID NO: 1 (promotor truncado LmUbq Núm. 2, como el que se encuentra en el constructo Egs23 descrito más adelante en la presente memoria); (iii) truncamientos 5' del elemento de control de la expresión de r-histona de *L. minor* (SEQ ID NO: 13), tales como los nucleótidos 461-1808 del SEQ ID NO: 13 (LmHIS (461-1808), como el que se encuentra en el constructo Egs19 descrito más adelante en la presente memoria) y los nucleótidos 805-1808 del SEQ ID NO: 13 (LmHIS (805-1808), como el que se encuentra en el constructo Egs20 descrito más adelante en la presente memoria); y (iv) truncamientos 5' del elemento de control de la expresión de quitinasa de *L. minor* (SEQ ID NO: 14), tales como los nucleótidos 51-1338 del SEQ ID NO: 14 (LmCHT (51-1338), como el que se encuentra en los constructos Egs24 y Egs25 descritos más adelante en la presente memoria).

Según se utiliza en la presente memoria, "secuencia completa" en referencia a una secuencia de nucleótidos especificada significa que tiene la secuencia de ácido nucleico entera de una secuencia nativa. Por "secuencia nativa" se entiende una secuencia endógena, esto es, una secuencia no diseñada encontrada en el genoma de un organismo.

De este modo, un fragmento de un elemento de control de la expresión de Lemnaceae puede funcionar como una porción biológicamente activa del elemento de control de la expresión. Una porción biológicamente activa de un elemento de control de la expresión se puede preparar aislando una porción de uno de los elementos de control de la expresión de la invención y evaluando la actividad (p. ej., la capacidad para iniciar o intensificar la transcripción) de esa porción del elemento de control de la expresión. Las moléculas de ácido nucleico que son fragmentos de un elemento de control de la expresión comprenden al menos 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1200, 1500, 1800, o 2000 nucleótidos contiguos, o hasta el número de nucleótidos presente en los elementos de control de la expresión completos descritos en la presente memoria (esto es, 2160 nucleótidos para el SEQ ID NO: 1, 2021 nucleótidos para el SEQ ID NO: 2, 2068 nucleótidos para el SEQ ID NO: 3, 1808 nucleótidos para el SEQ ID NO: 13, y 1338 nucleótidos para el SEQ ID NO: 14).

Los nucleótidos de tales fragmentos comprenderán normalmente la secuencia de reconocimiento TATA del elemento de control de la expresión concreto. Tales fragmentos se pueden obtener por medio del uso de enzimas de restricción para escindir los elementos de control de la expresión de origen natural descritos en la presente memoria; sintetizando una secuencia de nucleótidos a partir de la secuencia de origen natural de la secuencia de ADN del elemento de control de la expresión; o se pueden obtener por medio del uso de la tecnología de la reacción en cadena de la polimerasa (PCR). Véase concretamente, Mullis et al. (1987) *Methods Enzymol.* 155:335-350, y Erlich, ed. (1989) *PCR Technology* (Stockton Press, Nueva York). Las variantes de estos fragmentos de los elementos de control de la expresión, tales como aquellas resultantes de la mutagénesis dirigida al sitio, también están recogidos en las composiciones de la presente invención.

Se pretende que "fragmento" en el contexto de una secuencia que codifica un péptido señal y un péptido señal codificado represente una porción de la secuencia codificante o una porción del péptido señal codificado de ese modo. Con respecto a las secuencias codificantes, los fragmentos de una secuencia de nucleótidos pueden codificar fragmentos de polipéptidos que conservan la actividad biológica del polipéptido nativo, en este caso, el péptido señal de quitinasa de *L. minor* nativo. De este modo, un fragmento funcional del péptido señal de quitinasa de *L. minor* dirige el movimiento de una proteína madura de interés a través de la ruta secretora de una célula vegetal. Los fragmentos de una secuencia de nucleótidos codificante pueden oscilar de al menos aproximadamente 20 nucleótidos, aproximadamente 25 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 75 nucleótidos, y hasta la secuencia de nucleótidos entera que codifica el péptido señal de quitinasa de *L. minor* (esto es, hasta 84 nucleótidos del SEQ ID NO: 15).

Por "variantes" se entienden las secuencias que tienen una similitud sustancial con un elemento de control de la expresión descrito en la presente memoria (p. ej., los SEQ ID NO: 1-3, 13 y 14) o un fragmento de las mismas (p. ej., las secuencias mostradas en los SEQ ID NO: 4-9), o con una secuencia codificante de un péptido señal (p. ej., SEQ ID NO: 15) o un péptido señal (p. ej., SEQ ID NO: 16) o un fragmento del mismo. Para las secuencias de nucleótidos, las variantes de origen natural tales como estas se pueden identificar con el uso de mecanismos de la biología molecular bien conocidos, como, por ejemplo, los mecanismos de PCR y de hibridación como los esbozados más abajo. Las secuencias de nucleótidos variantes también incluyen secuencias de nucleótidos derivadas sintéticamente, tales como las generadas, por ejemplo, mediante el uso de la mutagénesis dirigida al sitio. Generalmente, las variantes de una secuencia de nucleótidos concreta de la invención, incluyendo las variantes de cualquiera de los SEQ ID NO: 1, 4 y 15, tendrán una identidad de secuencia de al menos 95%, 96%, 97%, y más preferiblemente aproximadamente 98%, 99% o más con esa secuencia de nucleótidos concreta como se determina

mediante los programas de alineamiento de secuencias descritos en la presente memoria más adelante utilizando los parámetros por defecto. Las variantes biológicamente activas también están incluidas en la presente invención. Las variantes biológicamente activas incluyen, por ejemplo, los elementos de control de la expresión nativos, o la secuencia que codifica el péptido señal nativo, de la invención que tienen una o más sustituciones, deleciones o inserciones de nucleótidos.

Según se utiliza en la presente memoria, "identidad de secuencia" o "identidad" en el contexto de dos secuencias de ácido nucleico o dos secuencias de polipéptidos hace referencia a los residuos de las dos secuencias que son iguales cuando son alineados para una correspondencia máxima a lo largo de una ventana de comparación especificada. Por "ventana de comparación" se entiende un segmento contiguo y especificado de una secuencia de polinucleótidos/polipéptidos, donde la secuencia de polinucleótidos/polipéptidos en la ventana de comparación puede comprender adiciones o deleciones (esto es, espacios) en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para el alineamiento óptimo de las dos secuencias. Generalmente, la ventana de comparación tiene al menos 20 nucleótidos/aminoácidos contiguos de longitud, y opcionalmente puede tener 30, 40, 50, 100 nucleótidos/aminoácidos, o más.

Los métodos de alineamiento de secuencias para su comparación son bien conocidos en la técnica. De este modo, la determinación del porcentaje de identidad de secuencia entre dos secuencias cualesquiera se puede completar utilizando un algoritmo matemático. Los ejemplos no limitantes de tales algoritmos matemáticos son el algoritmo de Myers y Miller (1988) CABIOS 4:11-17; el algoritmo de alineamiento local de Smith et al. (1981) Adv. Appl. Math. 2:482; el algoritmo de alineamiento global de Needleman y Wunsch (1970) J. Mol. Biol. 48:443-453; el método de búsqueda del alineamiento local de Pearson y Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; el algoritmo de Karlin y Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264, modificado como en Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5877.

Se pueden utilizar implementaciones por ordenador de estos algoritmos matemáticos para la comparación de secuencias para determinar la identidad de secuencia. Tales implementaciones incluyen, pero no están limitadas a: CLUSTAL en el programa PC/Gene (asequible de Intelligenetics, Mountain View, California); el programa ALIGN (Versión 2.0) y GAP, BESTFIT, BLAST, FASTA, y TFASTA del GCG Wisconsin Genetics Software Package, Versión 10 (asequible de Accelrys Inc., 9685 Scranton Road, San Diego, California, EEUU). Los alineamientos que utilizan estos programas se pueden realizar empleando los parámetros por defecto.

El programa CLUSTAL está bien descrito por Higgins et al. (1988) Gene 73:237-244 (1988); Higgins et al. (1989) CABIOS 5:151-153; Corpet et al. (1988) Nucleic Acids Res. 16:10881-90; Huang et al. (1992) CABIOS 8:155-65; y Pearson et al. (1994) Meth. Mol. Biol. 24:307-331. El programa ALIGN se basa en el algoritmo de Myers y Miller (1988) CABIOS 4:11-17. Se pueden utilizar una tabla de residuos en peso PAM120, una penalización de la longitud del espacio de 12, y una penalización del espacio de 4 con el programa ALIGN cuando se comparan secuencias de aminoácidos. Los programas BLAST de Altschul et al (1990) J. Mol. Biol. 215:403 se basan en el algoritmo de Karlin y Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264. Las búsquedas de nucleótidos BLAST se pueden llevar a cabo con el programa BLASTN, puntuación = 100, longitud de palabra = 12, para obtener secuencias de nucleótidos homólogas a una secuencia de nucleótidos de la invención. Las búsquedas de proteínas BLAST se pueden llevar a cabo con el programa BLASTX, puntuación = 50, longitud de palabra = 3, para obtener secuencias de aminoácidos homólogas a una proteína o polipéptido de la invención. Para obtener alineamientos con espacios con fines comparativos, se puede utilizar Gapped BLAST (en BLAST 2.0) como describen Altschul et al. (1997) Nucleic Acids Res. 25:3389. Alternativamente, se puede utilizar PSI-BLAST (en BLAST 2.0) para llevar a cabo una búsqueda repetida que detecte las relaciones lejanas entre moléculas. Véase Altschul et al. (1997) Nucleic Acids Res. 25:3389. Cuando se utilizan BLAST, Gapped BLAST, PSI-BLAST, se pueden emplear los parámetros por defecto de los respectivos programas (p. ej., BLASTN para las secuencias de nucleótidos y BLASTX para las proteínas).

GAP utiliza el algoritmo de Needleman y Wunsch (1970) J. Mol. Biol. 48:443-453, para encontrar el alineamiento de dos secuencias completas que maximiza el número de emparejamientos y minimiza el número de espacios. GAP considera todos los posibles alineamientos y posiciones de espacios y crea el alineamiento con el mayor número de bases emparejadas y los mínimos espacios. Esto permite la provisión de una penalización por creación de espacios y una penalización por prolongación del espacio en las unidades de bases emparejadas. GAP debe sacar provecho del número de penalizaciones por creación de espacios de los emparejamientos para cada espacio que inserta. Si se selecciona una penalización por prolongación del espacio mayor de cero, GAP debe, además, sacar provecho para cada espacio insertado de la longitud de las veces el espacio con respecto a la penalización por extensión del espacio. Los valores de penalización por creación de espacios y los valores de penalización por prolongación del espacio por defecto en la Versión 10 del GCG Wisconsin Genetics Software Package para las secuencias de proteínas son de 8 y 2, respectivamente. Para las secuencias de nucleótidos la penalización por creación de espacios por defecto es de 50 mientras la penalización por prolongación del espacio por defecto es de 3. Las penalizaciones por creación de espacios y por prolongación de espacios se pueden expresar como un número entero seleccionado del grupo de números enteros que consiste en 0 a 200. De este modo, por ejemplo, las penalizaciones por creación de espacios y por prolongación de espacios pueden ser 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 o mayores.

GAP representa un miembro de la familia de mejores alineamientos. Puede haber muchos miembros de esta familia, pero ningún otro miembro tiene una calidad mejor. GAP presenta cuatro cifras de mérito para los alineamientos: Calidad, Razón, Identidad, y Similitud. La calidad es la medida maximizada con el fin de alinear las secuencias. La razón es la calidad dividida por el número de bases del segmento más corto. El Porcentaje de Identidad es el porcentaje de los símbolos que realmente coinciden. Porcentaje de Similitud es el porcentaje de símbolos que son similares. Los símbolos que están al otro lado de los espacios se ignoran. Una similitud se puntúa cuando el valor de la matriz de puntuación para un par de símbolos es mayor o igual a 0,50, el umbral de similitud. La matriz de puntuación utilizada en la Versión 10 del GCG Wisconsin Genetics Software Package es BLOSUM62 (véase Henikoff y Henikoff (1992) Proc. Natl. Acad. Sci. USA 89:10915). El alineamiento también se puede realizar manualmente mediante inspección.

Un indicio alternativo de que dos moléculas de ácido nucleico están íntimamente relacionadas es que las dos moléculas hibridan entre sí en condiciones restrictivas. Las condiciones restrictivas dependen de la secuencia y son diferentes bajo diferentes parámetros medioambientales. Generalmente, las condiciones restrictivas se seleccionan para que sean aproximadamente 5°C a 20°C más bajas que el punto de fusión térmico (Tm) para la secuencia específica a una fuerza iónica y un pH definidos. La Tm es la temperatura (a una fuerza iónica y un pH definidos) a la cual 50% de la secuencia diana hibrida con una sonda perfectamente emparejada. Las condiciones para la hibridación de ácidos nucleicos y el cálculo de las restricciones se pueden encontrar, por ejemplo, en Sambrook et al. (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) y Tijssen (1993) *Hybridization With Nucleic Acid Probes, Parte I: Theory and Nucleic Acid Preparation* (Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Ltd., NY, NY).

Para los fines de la presente invención, "las condiciones restrictivas" incluyen condiciones en las cuales la hibridación solamente se producirá si hay menos de 25% de emparejamientos erróneos entre la molécula de hibridación y la secuencia diana. Las "condiciones restrictivas" se pueden interrumpir a niveles concretos de restricción para una definición más precisa. De este modo, según se utiliza en la presente memoria, las condiciones "restrictivas moderadas" son aquellas en las cuales las moléculas con más de 25% de emparejamientos erróneos no hibridarán; las condiciones "restrictivas medias" son aquellas en las cuales las moléculas con más de 15% de emparejamientos erróneos no hibridarán, y las condiciones "altamente restrictivas" son aquellas en las cuales las secuencias con más de 10% de emparejamientos erróneos no hibridarán. Las condiciones "muy altamente restrictivas" son aquellas en las cuales las secuencias con más de 6% de emparejamientos erróneos no hibridarán.

La actividad del elemento de control de la expresión para cualquiera de los elementos de control de la expresión de Lemnaceae, o fragmentos o variantes de los mismos, se puede analizar utilizando una variedad de mecanismos bien conocidos por los expertos en la técnica, incluyendo, por ejemplo, el análisis por transferencia Northern, las mediciones de la actividad del informador tomadas de fusiones transcripcionales, y similares. Véase, por ejemplo, Sambrook et al. (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Alternativamente, los análisis del elemento de control de la expresión se pueden basar en la medición de los niveles de un gen informador tal como  $\beta$ -glucuronidasa (GUS), proteína fluorescente verde (GFP), o similares producidos bajo el control de un elemento de control de la expresión, o un fragmento o variante del mismo. Véase, por ejemplo, la Patente de los Estados Unidos Núm. 6.072.050, incorporada en la presente memoria como referencia. La actividad del péptido señal de quitinasa de *L. minor*, o los fragmentos o variantes del mismo, se pueden analizar de un modo similar utilizando una variedad de mecanismos bien conocidos por los expertos en la técnica, incluyendo aquellos que detectan la actividad del péptido señal de quitinasa, o un fragmento o variante del mismo, para dirigir la secreción extracelular de un polipéptido de interés.

Los métodos para la mutagénesis y las alteraciones de la secuencia de nucleótidos son bien conocidos en la técnica. Véanse, por ejemplo, Kunkel (1985) Proc. Natl. Acad. Sci. USA 82:488-492; Kunkel et. al. (1987) *Methods in Enzymol.* 154:367-382; Patente de los Estados Unidos Núm. 4.873.192 Walker y Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, NY) y las referencias allí citadas.

Los elementos de control de la expresión de Lemnaceae de la presente invención, y las variantes o fragmentos de los mismos, cuando se ensamblan en un constructo nucleotídico de manera que el elemento de control de la expresión está conectado operablemente a una secuencia de nucleótidos de interés, permiten la expresión de la secuencia de nucleótidos conectada operablemente en una planta o célula o nódulo vegetal (p. ej., una planta de lenteja de agua o una célula o un nódulo de una planta de lenteja de agua, por ejemplo del género *Spirodela*, del género *Wolffia*, del género *Wolffiella*, del género *Landoltia*, o del género *Lemna*). Por "conectado operablemente" se entiende que la transcripción o la traducción de la secuencia de nucleótidos de interés está bajo la influencia del elemento de control de la expresión. De esta manera, las secuencias de nucleótidos para los elementos de control de la expresión de la invención se proporcionan en casetes o vectores de expresión junto con la secuencia de nucleótidos de interés, típicamente una secuencia de nucleótidos heteróloga, para su expresión en la planta o la célula o nódulo vegetal. Por "secuencia de nucleótidos heteróloga" se entiende una secuencia que no está conectada operablemente de forma natural al elemento de control de la expresión. Si bien esta secuencia de nucleótidos es heteróloga con respecto al elemento de control de la expresión, puede ser homóloga, o nativa, o heteróloga, o foránea, con respecto al anfitrión vegetal.

Se admite que se pueden utilizar los elementos de control de la expresión de la invención, o las variantes o fragmentos de los mismos, para dirigir la expresión de la respectiva secuencia codificante nativa. Tales constructos pueden cambiar los niveles de expresión del polipéptido nativo en una planta o célula vegetal. De este modo, se puede alterar el fenotipo de la planta o la célula vegetal.

5 Según se utiliza en la presente memoria, "vector" hace referencia a una molécula de ADN tal como un plásmido, cósmido o bacteriófago para introducir un constructo nucleotídico, por ejemplo, una casete de expresión, en una célula anfitriona. Los vectores de clonación contienen típicamente uno o un pequeño número de sitios de reconocimiento para endonucleasas de restricción en los cuales se pueden insertar secuencias de ADN foráneas de una manera definible sin pérdida de la función biológica esencial del vector, así como un gen marcador, como se describe más adelante en la presente memoria, que sea adecuado para su uso en la identificación y selección de las células transformadas con el vector de clonación.

10 Según se utiliza en la presente memoria, el término "vegetal" incluye la referencia a plantas completas, órganos de plantas (p. ej., hojas, tallos, raíces, etc.), semillas, células vegetales, y progenie de las mismas. Se debe entender que las partes de las plantas transgénicas están dentro del alcance de la invención al comprender, por ejemplo, células vegetales, protoplastos, tejidos, callos, embriones así como flores, óvulos, tallos, frutos, hojas, raíces, puntas radiculares, y similares que se originan en plantas transgénicas o su progenie previamente transformadas con una molécula de ADN de la invención y que por lo tanto consisten, al menos en parte, en células transgénicas. Según se utiliza en la presente memoria, el término "célula vegetal" incluye, sin limitación, células de semillas, embriones, regiones meristemáticas, tejido de callo, hojas, raíces, brotes, gametofitos, esporofitos, polen, y microsporas. La clase de plantas que se puede utilizar en los métodos de la invención es generalmente tan amplia como la clase de plantas superiores susceptible de técnicas de transformación, incluyendo plantas tanto monocotiledóneas como dicotiledóneas. Tales plantas incluyen, por ejemplo, la lenteja de agua.

15 El término "lenteja de agua" hace referencia a miembros de la familia Lemnaceae. Esta familia está dividida actualmente en cinco géneros y 38 especies de lenteja de agua como sigue: género *Lemna* (*L. aequinoctialis*, *L. disperma*, *L. ecuadoriensis*, *L. gibba*, *L. japonica*, *L. minor*, *L. miniscule*, *L. obscura*, *L. perpusilla*, *L. tener-u*, *L. trisulca*, *L. turionifera*, *L. valdiviana*); género *Spirodela* (*S. intermedia*, *S. polirrhiza*, *S. punctata*); género *Wolffia* (*Wa. angusta*, *Wa. arrhiza*, *Wa. australina*, *Wa. borealis*, *Wa. brasiliensis*, *Wa. columbiana*, *Wa. elongata*, *Wa. globosa*, *Wa. microscopica*, *Wa. neglecta*); género *Wolffiella* (*Wl. caudata*, *Wl. denticulata*, *Wl. gladiata*, *Wl. hyalina*, *Wl. lingulata*, *Wl. repunda*, *Wl. rotunda*, y *Wl. neotropica*) y género *Landoltia* (*L. punctata*). Cualquier otro género o especie de Lemnaceae, si existiera, también sería un aspecto de la presente invención. La especie *Lemna* se puede clasificar utilizando el esquema taxonómico descrito por Landolt (1986) en *Biosystematic Investigation on the Family, of Duckweeds: The family of Lemnaceae - A Monograph Study* (Geobotanischen Institut ETH, Stiftung Rubel, Zurich).

20 El término "nódulo de lenteja de agua" según se utiliza en la presente memoria hace referencia a tejido de lenteja de agua que comprende células de lenteja de agua donde al menos aproximadamente 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, o 100% de las células son células diferenciadas. Una "célula diferenciada", según se utiliza en la presente memoria, es una célula con al menos una característica fenotípica (p. ej., una morfología celular distintiva o la expresión de un ácido nucleico o proteína marcadores) que la distingue de las células no diferenciadas o de las células encontradas en otros tipos de tejidos. Las células diferenciadas de cultivo de nódulos de lenteja de agua descrito en la presente memoria forman una superficie lisa en mosaico de células interconectadas fusionadas en sus paredes celulares adyacentes, con nódulos que han empezado a organizarse en primordios de frondas dispersados por todo el tejido. La superficie del tejido del cultivo de nódulos tiene células epidérmicas conectadas entre sí a través de plasmodesmos.

25 En algunas realizaciones, se proporcionan casetes o vectores de expresión que comprenden un elemento de control de la expresión de Lemnaceae, o una variante o fragmento del mismo, conectado operablemente a una secuencia de nucleótidos de interés para la expresión del polipéptido codificado por la secuencia de nucleótidos de interés en una planta o célula o nódulo vegetal. La secuencia de nucleótidos de interés conectada operablemente puede ser cualquier secuencia cuya expresión en una planta o célula o nódulo vegetal sea deseable. La secuencia de nucleótidos de interés será típicamente una secuencia de nucleótidos heteróloga, como se define en la presente memoria. Las secuencias de nucleótidos heterólogas ilustrativas de interés incluyen, pero no están limitadas a, secuencias que codifican polipéptidos de mamífero, tales como insulina, hormona de crecimiento, interferón interferón  $\beta$ ,  $\beta$ -glucocerebrosidasa,  $\beta$ -glucuronidasa, proteína de retinoblastoma, proteína p53, angiostatina, leptina, eritropoyetina, factor estimulador de colonias de granulocitos macrófagos, plasminógeno, anticuerpos monoclonales, fragmentos Fab, anticuerpos de cadena sencilla, citoquinas, receptores, vacunas humanas, vacunas animales, péptidos, y albúmina de suero.

30 Los términos "polipéptido", "péptido", y "proteína" se utilizan indistintamente en la presente memoria para hacer referencia a un polímero de residuos de aminoácido. Los términos se aplican a polímeros de aminoácidos en los cuales uno o más residuos de aminoácido son análogos químicos artificiales de los correspondientes aminoácidos de origen natural, así como a polímeros de aminoácidos de origen natural. Según se utilizan en la presente memoria, los términos "que codifica" o "codificado" cuando se utilizan en el contexto de un ácido nucleico especificado



significan que el ácido nucleico comprende la información requisito para dirigir la traducción de la secuencia de nucleótidos a una proteína especificada. La información por medio de la cual una proteína es codificada viene especificada por el uso de codones. Un ácido nucleico que codifica una proteína puede comprender secuencias no traducidas (p. ej., intrones) dentro de regiones traducidas del ácido nucleico o pueden carecer de tales secuencias no traducidas intermedias (p. ej., como en el ADNc).

En un ejemplo específico, no limitante, se obtiene lenteja de agua transformada mediante transformación con una casete de expresión que comprende un elemento de control de la expresión de ubicuitina de Lemnaceae (p. ej., como se muestra en el SEQ ID NO: 1H), un fragmento del mismo (p. ej., como se muestra en el SEQ ID NO: 4), o una variante de estas secuencias conectado operablemente a una secuencia de nucleótidos heteróloga de interés.

Se proporciona una casete de expresión de la invención con una pluralidad de sitios de restricción para la inserción de la secuencia de nucleótidos que codifica la proteína de interés para que esté bajo la regulación transcripcional del elemento de control de la expresión. La casete de expresión puede codificar un único gen de interés. Alternativamente, la casete de expresión puede codificar dos o más genes de interés.

Las casetes de expresión descritas en la presente memoria incluyen en la dirección 5'-3' de la transcripción, una región de inicio de la transcripción y la traducción (p. ej., un elemento de control de la expresión de la invención o una variante o fragmento biológicamente activo del mismo), una secuencia de nucleótidos de interés, y una región de terminación de la transcripción y la traducción funcional en plantas. Se puede utilizar cualquier secuencia de terminación adecuada conocida en la técnica de acuerdo con la presente invención. La región de terminación puede ser nativa con la región de inicio de la transcripción, puede ser nativa con la secuencia de nucleótidos de interés, o puede derivar de otra fuente. Las regiones de terminación convenientes se encuentran disponibles del plásmido Ti de *A. tumefaciens*, tales como las regiones de terminación de la octopina sintetasa y de la nopalina sintetasa. Véanse también Guerineau et al. (1991) *Mol. Gen. Genet.* 262:141; Proudfoot (1991) *Cell* 64:671; Sanfacon et al. (1991) *Genes Dev.* 5:141; Mogen et al. (1990) *Plant Cell* 2:1261; Munroe et al. (1990) *Gene* 91:151; Ballas et al. (1989) *Nucleic Acids Res.* 17:7891; y Joshi et al. (1987) *Nucleic Acids Res.* 15:9627. Son secuencias de terminación ilustrativas adicionales la secuencia de terminación de la subunidad pequeña de RubP carboxilasa de guisante, la secuencia de terminación de 35S del Virus del Mosaico de la coliflor, y el terminador de ubicuitina de muchas especies de plantas. Otras secuencias de terminación adecuadas serán evidentes para los expertos en la técnica.

Generalmente, la casete de expresión comprenderá un gen marcador seleccionable para la selección de células o tejidos transformados. Los genes marcadores seleccionables incluyen genes que codifican resistencia a antibióticos, tales como aquellos que codifican la neomicina fosfotransferasa II (NEO), la neomicina fosfotransferasa III y la higromicina fosfotransferasa (HPT), así como genes que confieren resistencia a compuestos herbicidas. Los genes de resistencia a herbicidas generalmente codifican una proteína diana modificada insensible al herbicida o una enzima que degrada o detoxifica el herbicida en la planta antes de que actúe. Véanse, DeBlock et al. (1987) *EMBO J.* 6:2513; DeBlock et al. (1989) *Plant Physiol.* 91:691; Fromm et al. (1990) *BioTechnology* 8:833; Gordon-Kamm et al. (1990) *Plant Cell* 2:603; y Frisch et al. (1995) *Plant Mol. Biol.* 27:405-9. Por ejemplo, se ha obtenido resistencia a herbicidas de glifosato o sulfonilurea utilizando genes que codifican las enzimas diana mutantes, 5-enolpiruvilshikimato-3-fosfato sintasa (EPSPS) y acetolactato sintasa (ALS). La resistencia a glufosinato amónico, boromoxinilo, y 2,4-diclorofenoxiacetato (2,4-D) se ha obtenido utilizando genes bacterianos que codifican fosfinotricina acetiltransferasa, una nitrilasa, o una 2,4-diclorofenoxiacetato monooxigenasa, que detoxifican los respectivos herbicidas.

Para los fines de la presente invención, los genes marcadores seleccionables incluyen, pero no están limitados a, genes que codifican la neomicina fosfotransferasa II (Fraley et al. (1986) *CRC Critical Reviews in Plant Science* 4:1), la neomicina fosfotransferasa III (Frisch et al. (1995) *Plant Mol. Biol.* 27:405-9), la cianamida hidratasa (Maier-Greiner et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:4250); la aspartato quinasa; la dihidrodipicolinato sintasa (Perl et al. (1993) *BioTechnology* 11:715); el gen bar (Toki et al. (1992) *Plant Physiol.* 100:1503; Meagher et al. (1996) *Crop Sci.* 36:1367); la triptófano descarboxilasa (Goddijn et al. (1993) *Plant Mol. Biol.* 22:907); la neomicina fosfotransferasa (NEO; Southern et al. (1982) *J. Mol. Appl. Gen.* 1:327); la higromicina fosfotransferasa (HPT o HYG; Shimizu et al. (1986) *Mol Cell. Biol.* 6:1074); la dihidrofolato reductasa (DHFR; Kwok et al. (1986) *Proc. Natl. Acad. Sci. USA* 83:4552); la fosfinotricina acetiltransferasa (DeBlock et al. (1987) *EMBO J.* 6:2513); la ácido 2,2-dicloropropionico deshalogenasa (Buchanan-Wollatron et al. (1989) *J. Cell. Biochem.* 13D:330); la acetohidroxiácido sintasa (Patente de los Estados Unidos Núm. 4.761.373 de Anderson et al.; Haughn et al. (1988) *Mol. Gen. Genet.* 221:266); la 5-enolpiruvil-shikimato-fosfato sintasa (aroA; Comai et al. (1985) *Nature* 317:741); la haloarilnitrilasa (documento WO 87/04181 de Stalker et al.); la acetil-coenzima A carboxilasa (Parker et al. (1990) *Plant Physiol.* 92:1220); la dihidropteroato sintasa (sull); Guerineau et al. (1990) *Plant Mol. Biol.* 15:127); y el polipéptido del fotosistema II de 32 kDa (psbA; Hirschberg et al. (1983) *Science* 222:1346 (1983).

También están incluidos los genes que codifican la resistencia a gentamicina (p. ej., aacCI, Wohlleben et al. (1989) *Mol. Gen. Genet.* 217:202-208); cloramfenicol (Herrera-Estrella et al. (1983) *EMBO J.* 2:987); metotrexato (Herrera-Estrella et al. (1983) *Nature* 303:209; Meijer et al. (1991) *Plant Mol. Biol.* 16:807); higromicina (Waldron et al. (1985) *Plant Mol. Biol.* 5:103; Zhijian et al. (1995) *Plant Science* 108:219; Meijer et al. (1991) *Plant Mol. Bio.* 16:807); estreptomycin (Jones et al. (1987) *Mol. Gen. Genet.* 210:86); espectinomycin (Bretagne-Sagnard et al. (1996)

Transgenic Res. 5:131); bleomicina (Hille et al. (1986) Plant Mol. Biol. 7:171); sulfonamida (Guerineau et al. (1990) Plant Mol. Bio. 15:127); bromoxinilo (Stalker et al. (1988) Science 242:419); 2,4-D (Streber et al. (1989) BioTechnology 7:811); fosfotricina (DeBlock et al. (1987) EMBO J. 6:2513); espectinomicina (Bretagne-Sagnard y Chupeau, Transgenic Research 5:13).

5 El gen bar confiere resistencia a herbicidas para los herbicidas de tipo glufosinato, tales como fosfotricina (PPT) o bialafos, y similares. Otros marcadores seleccionables que se podrían utilizar en los constructos de expresión incluyen, pero no están limitados a, el gen PAT, también para la resistencia a bialafos y fosfotricina, el gen ALS para la resistencia a imidazolinona, el gen HPH o HYG para la resistencia a higromicina, el gen de EPSP sintasa para la resistencia a glifosato, el gen Hml para la resistencia a la toxina Hc, y otros agentes selectivos utilizados rutinariamente y conocidos por los expertos en la técnica. Véanse Yarranton (1992) Curr. Opin. Biotech. 3:506; Christopherson et al. (1992) Proc. Natl. Acad. Sci. USA 89:6314; Yao et al. (1992) Cell 71:63; Reznikoff (1992) Mol. Microbiol. 6:2419; Barkley et al. (1980) The Operon 177-220; Hu et al. (1987) Cell 48:555; Brown et al. (1987) Cell 49:603; Figge et al. (1988) Cell 52:713; Deuschle et al. (1989) Proc. Natl. Acad. Sci. USA 86:5400; Fuerst et al. (1989) Proc. Natl. Acad. Sci. USA 86:2549; Deuschle et al. (1990) Science 248:480; Labow et al. (1990) Mol. Cell. Biol. 10:3343; Zambretti et al. (1992) Proc. Natl. Acad. Sci. USA 89:3952; Baim et al. (1991) Proc. Natl. Acad. Sci. USA 88:5072; Wyborski et al. (1991) Nuc. Acids Res. 19:4647; Hillenand-Wissman (1989) Topics in Mol. And Struc. Biol. 10:143; Degenkolb et al. (1991) Antimicrob. Agents Chemother. 35:1591; Kleinschmidt et al. (1988) Biochemistry 27:1094; Gatz et al. (1992) Plant J. 2:397; Gossen et al. (1992) Proc. Natl. Acad. Sci. USA 89:5547; Oliva et al. (1992) Antimicrob. Agents Chemother. 36:913; Hlavka et al. (1985) Handbook of Experimental Pharmacology 78; y Gill et al. (1988) Nature 334:721.

La lista anterior de genes marcadores seleccionables no pretende ser limitante. En la presente invención se puede utilizar cualquier gen marcador seleccionable letal o no letal.

25 En algunas realizaciones, la presente invención proporciona la modificación de la secuencia de nucleótidos expresada de interés para intensificar su expresión en la planta de interés. Los métodos se encuentran disponibles en la técnica para sintetizar secuencias de nucleótidos con codones preferidos por las plantas. Véanse, por ejemplo, las Patentes de los Estados Unidos Núms. 5.380.831 y 5.436.391 Perlak et al. (1991) Proc. Natl. Acad. Sci. USA 15:3324; Iannacome et al. (1997) Plant Mol. Biol. 34:485; y Murray et al. (1989) Nucleic Acids. Res. 17:477.

30 Por ejemplo, cuando la planta de interés es la lenteja de agua, una de tales modificaciones es la síntesis de la secuencia de nucleótidos de interés utilizando los codones preferidos por la lenteja de agua. Los codones preferidos se pueden determinar a partir de los codones de máxima frecuencia en las proteínas expresadas en la lenteja de agua. De este modo, la frecuencia de uso de un codón concreto de lenteja de agua se puede determinar analizando el uso de codones en un grupo de secuencias codificantes de lenteja de agua. Numerosas secuencias codificantes de lenteja de agua son conocidas por los expertos en la técnica; véanse por ejemplo, las secuencias contenidas en la base de datos GenBank®, a la que se puede acceder a través del sitio de la red para el Centro Nacional para Información de Biotecnología, una división de la Biblioteca Nacional de Medicina, que está localizada en Bethesda, Maryland. Las tablas que muestran la frecuencia de uso de codones basadas en las secuencias contenidas en la publicación más reciente de GenBank® se pueden encontrar en el sitio de la red para Kazusa DNA Research Institute en Chiba, Japón. Esta base de datos es descrita por Nakamura et al. (2000) en Nucleic Acids Res. 28:292.

40 Se admite que los genes que han sido optimizados para determinar la expresión en lenteja de agua y otras monocotiledóneas o dicotiledóneas pueden ser utilizados en los métodos de la invención. Véanse, p. ej., las Patentes Europeas EP 0 359 472, EP 0 385 962, el documento WO 91/16432; Perlak et al. (1991) Proc. Natl. Acad. Sci. USA 88:3324; Iannacome et al. (1997) Plant Mol. Biol. 34:485; Murray et al. (1989) Nucleic Acids Res. 17:477; y similares. Adicionalmente se admite que toda o cualquier parte de la secuencia génica puede ser optimizada o sintética. En otras palabras, también se pueden utilizar secuencias totalmente optimizadas o parcialmente optimizadas. Por ejemplo, 70%, 75%, 80%, 85%, 87%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 49%, o 100% de los codones pueden ser codones preferidos por la planta, por ejemplo, codones preferidos por la lenteja de agua. De este modo, en algunas realizaciones, la secuencia de nucleótidos que codifica el polipéptido de interés comprende entre 70-100% de codones preferidos por la lenteja de agua. En una realización, entre 90 y 96% de los codones son codones preferidos por la lenteja de agua. La secuencia codificante de la secuencia de nucleótidos de interés puede comprender codones utilizados con una frecuencia de al menos 17% en la lenteja de agua. El uso de codones en Lemna gibba (Tabla 1) y Lemna minor (Tabla 2) se muestra más abajo. En algunas realizaciones, se utilizan la Tabla 1 o la Tabla 2 para seleccionar los codones preferidos por la lenteja de agua.

55

## ES 2 398 142 T3

Tabla 1: Uso de codones en Lemna gibba de la Publicación 139\* de GenBank®

Aminoácido	Codón	Número	/1000	Fracción
Gly	GGG	57,00	28,89	0,35
Gly	GGA	8,00	4,05	0,05
Gly	GGT	3,00	1,52	0,02
Gly	GGC	93,00	47,14	0,58
Glu	GAG	123,00	62,34	0,95
Glu	GAA	6,00	3,04	0,05
Asp	GAT	6,00	3,04	0,08
Asp	GAC	72,00	36,49	0,92
Val	GTG	62,00	31,42	0,47
Val	GTA	0,00	0,00	0,00
Val	GTT	18,00	9,12	0,14
Val	GTC	51,00	25,85	0,39
Ala	GCG	44,00	22,30	0,21
Ala	GCA	14,00	7,10	0,07
Ala	GCT	14,00	7,10	0,07
Ala	GCC	139,00	70,45	0,66
Arg	AGG	16,00	8,11	0,15
Arg	AGA	11,00	5,58	0,10
Ser	AGT	1,00	0,51	0,01
Ser	AGC	44,00	22,30	0,31
Lys	AAG	116,00	58,79	1,00

ES 2 398 142 T3

Aminoácido	Codón	Número	/1000	Fracción
Lys	AAA	0,00	0,00	0,00
Asn	AAT	2,00	1,01	0,03
Asn	AAC	70,00	35,48	0,97
Met	ATG	67,00	33,96	1,00
Ile	ATA	4,00	2,03	0,06
Ile	ATT	0,00	0,00	0,00
Ile	ATC	63,00	31,93	0,94
Thr	ACG	19,00	9,63	0,25
Thr	ACA	1,00	0,51	0,01
Thr	ACT	6,00	3,04	0,08
Thr	ACC	50,00	25,34	0,66
Trp	TGG	45,00	22,81	1,00
End	TGA	4,00	2,03	0,36
Cys	TGT	0,00	0,00	0,00
Cys	TGC	34,00	17,23	1,00
End	TAG	0,00	0,00	0,00
End	TAA	7,00	3,55	0,64
Tyr	TAT	4,00	2,03	0,05
Tyr	TAC	76,00	38,52	0,95
Leu	TTG	5,00	2,53	0,04
Leu	TTA	0,00	0,00	0,00
Phe	TTT	4,00	2,03	0,04
Phe	TTC	92,00	46,63	0,96

ES 2 398 142 T3

Aminoácido	Codón	Número	/1000	Fracción
Ser	TCG	34,00	17,23	0,24
Ser	TCA	2,00	1,01	0,01
Ser	TCT	1,00	0,51	0,01
Ser	TCC	59,00	29,90	0,42
Arg	CGG	23,00	11,66	0,22
Arg	CGA	3,00	1,52	0,03
Arg	CGT	2,00	1,01	0,02
Arg	CGC	50,00	25,34	0,48
Gln	CAG	59,00	29,90	0,86
Gln	CAA	10,00	5,07	0,14
His	CAT	5,00	2,53	0,26
His	CAC	14,00	7,10	0,74
Leu	CTG	43,00	21,79	0,35
Leu	CTA	2,00	1,01	0,02
Leu	CTT	1,00	0,51	0,01
Leu	CTC	71,00	35,99	0,58
Pro	CCG	44,00	22,30	0,31
Pro	CCA	6,00	3,04	0,04
Pro	CCT	13,00	6,59	0,09
Pro	CCC	80,00	40,55	0,56

## ES 2 398 142 T3

Tabla 2: Uso de codones en Lemna minor de la Publicación 139\* de GenBank®

Aminoácido	Codón	Número	/1000	Fracción
Gly	GGG	8,00	17,39	0,22
Gly	GGA	11,00	23,91	0,31
Gly	GGT	1,00	2,17	0,03
Gly	GGC	16,00	34,78	0,44
Glu	GAG	25,00	54,35	0,78
Glu	GAA	7,00	15,22	0,22
Asp	GAT	8,00	17,39	0,33
Asp	GAC	16,00	34,78	0,67
Val	GTG	21,00	45,65	0,53
Val	GTA	3,00	6,52	0,07
Val	GTT	6,00	13,04	0,15
Val	GTC	10,00	21,74	0,25
Ala	GCG	13,00	28,26	0,32
Ala	GCA	8,00	17,39	0,20
Ala	GCT	6,00	13,04	0,15
Ala	GCC	14,00	30,43	0,34
Arg	AGG	9,00	19,57	0,24
Arg	AGA	11,00	23,91	0,30
Ser	AGT	2,00	4,35	0,05
Ser	AGC	11,00	23,91	0,26
Lys	AAG	13,00	28,26	0,68
Lys	AAA	6,00	13,04	0,32
Asn	AAT	0,00	0,00	0,00

ES 2 398 142 T3

Aminoácido	Codón	Número	/1000	Fracción
Asn	AAC	12,00	26,09	1,00
Met	ATG	9,00	19,57	1,00
Ile	ATA	1,00	2,17	0,08
Ile	ATT	2,00	4,35	0,15
Ile	ATC	10,00	21,74	0,77
Thr	ACG	5,00	10,87	0,28
Thr	ACA	2,00	4,35	0,11
Thr	ACT	2,00	4,35	0,11
Thr	ACC	9,00	19,57	0,50
Trp	TGG	8,00	17,39	1,00
End	TGA	1,00	2,17	1,00
Cys	TGT	1,00	2,17	0,12
Cys	TGC	7,00	15,22	0,88
End	TAG	0,00	0,00	0,00
End	TAA	0,00	0,00	0,00
Tyr	TAT	1,00	2,17	0,12
Tyr	TAC	7,00	15,22	0,88
Leu	TTG	3,00	6,52	0,08
Leu	TTA	1,00	2,17	0,03
Phe	TTT	6,00	13,04	0,25
Phe	TTC	18,00	39,13	0,75
Ser	TCG	11,00	23,91	0,26
Ser	TCA	4,00	8,70	0,09

ES 2 398 142 T3

Aminoácido	Codón	Número	/1000	Fracción
Ser	TCT	6,00	13,04	0,14
Ser	TCC	9,00	19,57	0,21
Arg	CGG	4,00	8,70	0,11
Arg	CGA	4,00	8,70	0,11
Arg	CGT	0,00	0,00	0,00
Arg	CGC	9,00	19,57	0,24
Gln	CAG	11,00	23,91	0,73
Gln	CAA	4,00	8,70	0,27
His	CAT	0,00	0,00	0,00
His	CAC	6,00	13,04	1,00
Leu	CTG	9,00	19,57	0,24
Leu	CTA	4,00	8,70	0,11
Leu	CTT	4,00	8,70	0,11
Leu	CTC	17,00	36,96	0,45
Pro	CCG	8,00	17,39	0,29
Pro	CCA	7,00	15,22	0,25
Pro	CCT	5,00	10,87	0,18
Pro	CCC	8,00	17,39	0,29

También se pueden realizar otras modificaciones en la secuencia de nucleótidos de interés para optimizar su expresión en una planta. Estas modificaciones incluyen, pero no están limitadas a, la eliminación de secuencias que codifican señales de poliadenilación espúreas, señales de sitios de empalme exón-intrón, repeticiones de tipo transposón, y otras de tales secuencias caracterizadas que pueden ser deletéreas para la expresión génica. El contenido de G-C de la secuencia se puede ajustar a niveles medios para un anfitrión celular dado, calculados mediante la referencia a genes conocidos expresados en la célula anfitriona. Cuando sea posible, la secuencia puede ser modificada para evitar estructuras de ARNm secundarias en horquilla pronosticadas.

5

Existen diferencias conocidas entre las secuencias de nucleótidos en el contexto del inicio de la traducción óptima para los codones de inicio de la traducción en animales y plantas, y la composición de estas secuencias de nucleótidos en el contexto del inicio de la traducción puede influir en la eficacia del inicio de la traducción. Véanse, por ejemplo, Lukaszewicz et al. (2000) *Plant Science* 154:89-98; y Joshi et al. (1997) *Plant Mol. Biol.* 35:993-1001. En algunas realizaciones de la presente invención, la secuencia de nucleótidos en el contexto del inicio de la traducción para el codón de inicio de la traducción de la secuencia de nucleótidos de interés puede ser modificada para intensificar la expresión en lenteja de agua. En una realización, la secuencia de nucleótidos se modifica de

10



manera que los tres nucleótidos directamente aguas arriba del codón de inicio de la traducción de la secuencia de nucleótidos de interés sean "ACC". En una segunda realización, estos nucleótidos son "ACA".

Además de los elementos de control de la expresión descritos en la presente memoria para iniciar o intensificar la expresión de una secuencia de nucleótidos heteróloga en una planta, la expresión de una secuencia de nucleótidos de interés también puede ser intensificada por medio del uso opcional de diferentes elementos reguladores. "Elemento regulador" según se utiliza en la presente memoria, hace referencia a una secuencia de nucleótidos, ya sea ADN o ARN, normalmente aguas arriba (5') de la secuencia codificante de un gen estructural, incluyendo secuencias de control de la transcripción tales como secuencias líder, promotores, intensificadores o represores de la traducción y la transcripción, y determinantes de la estabilidad y la inestabilidad del ARNm. Las secuencias encontradas dentro de los intrones también pueden regular la expresión de la región codificante de interés. Los elementos reguladores también se pueden encontrar 3' con respecto al sitio de inicio de la transcripción, o en las regiones transcritas. Los diferentes elementos reguladores se pueden conectar operablemente a otros elementos reguladores. "Secuencia líder" según se utiliza en la presente memoria hace referencia a la porción de un ácido nucleico localizada en el extremo 5' del ARNm, extendiéndose desde el sitio CAP 5' hasta el codón de inicio de la traducción de la proteína AUG. La secuencia líder es importante en el inicio de la traducción y en la regulación de la expresión génica.

Por ejemplo, se pueden utilizar adicionalmente una o más secuencias líder combinadas para intensificar la expresión de la secuencia de nucleótidos diana. Los líderes de la traducción son conocidos en la técnica e incluyen, pero no están limitados a, líderes de picornavirus, p. ej., líder de EMCV (región no codificante 5' de encefalomiocarditis; Elroy-Stein et al. (1989) Proc. Natl. Acad. Sci USA 86:6126); líderes de polivirus, p. ej., líder de TEV (Virus del Grabado del Tabaco; Gallie et al. (1995) Gene 165:233); proteína de unión de la cadena pesada de inmunoglobulina humana (BiP; Macajak y Sarnow (1991) Nature 353:90); líder no traducido del ARNm de la proteína de la envoltura del virus del mosaico de la alfalfa (AMV RNA 4; Jobling y Gehrke (1987) Nature 325:622); líder del virus del mosaico del tabaco (TMV; Gallie (1989) Molecular Biology of RNA, 23:56); líder del virus del mosaico (Tomashevskaya et al. (1993) J. Gen. Virol. 74:2717-2724); región no traducida 5' Fed-1 (Dickey (1992) EMBO J. 11:2311-2317); región no traducida 5' RbcS (Silverthorne et al. (1990) J. Plant. Mol. Biol. 15:49-58); y líder del virus del moteado clorótico del maíz (MCMV; Lommel et al. (1991) Virology 81:382). Véase también, Della-Cioppa et al. (1987) Plant Physiology 84:965. También se ha demostrado que las secuencias líder que comprenden una secuencia intrónica vegetal, incluyendo la secuencia intrónica del gen de la deshidrogenasa 1 de maíz, el gen de la catalasa del ricino, o el gen de la ruta del triptófano PAT1 de Arabidopsis, incrementan la eficacia traduccional en plantas (Callis et al. (1987) Genes Dev. 1:1183-1200; Mascarenhas et al. (1990) Plant Mol. Biol. 15:913-920).

Los intrones de la ubicuitina de Lemnaceae descritos anteriormente en la presente memoria (es decir, como se muestra en los SEQ ID NO: 7-9) se pueden utilizar con promotores distintos de sus respectivos promotores de ubicuitina para intensificar la expresión de una secuencia de nucleótidos de interés conectada operablemente. El promotor utilizado con los intrones de ubicuitina puede ser cualquier promotor adecuado para su uso en la planta de interés, incluyendo los promotores de r-histona y quitinasa de Lemnaceae novedosos descritos en los SEQ ID NO: 13 y 14, respectivamente. Otros promotores adecuados se pueden obtener de una variedad de fuentes, tales como plantas o virus con ADN de plantas. Los promotores útiles incluyen aquellos aislados del grupo de los caulimovirus, tales como los promotores de los transcritos 19S y 35S del virus del mosaico de la coliflor (CaMV19S y CaMV35S). Otros promotores útiles incluyen el promotor de CaMV35S intensificado (eCaMV35S) como describen Kat et al. (1987) en Science 236:1299-1302, y el promotor de la subunidad pequeña de la ribulosa 1,5-bisfosfato carboxilasa oxigenasa (RUBISCO). Los ejemplos de otros promotores adecuados son el promotor de la actina de arroz; el promotor de ciclofilina; el promotor de ADH1 (Callis et al. (1987) Gene Dev. 1:1183-1200); el promotor de patatina de Clase I (Bevan et al. (1986) Nucl. Acids Res. 14:4675-4638); el promotor de ADP glucosa pirofosforilasa; el promotor de  $\beta$ -conglucininina (Tiemey et al. (1987) Planta 172:356-363); el promotor E8 (Deikman et al. (1988) Embo J. 7:3315-3320); el promotor 2AII (Pear et al. (1989) Plant Mol. Biol. 13:639-651); y el promotor de quitinasa ácida (Samac et al. (1990) Plant Physiol. 93:907-914).

Se admite que se puede utilizar en la presente invención cualquiera de las modificaciones de secuencias de nucleótidos que intensifican la expresión descritas más arriba, incluyendo una única modificación o cualquier posible combinación de modificaciones.

En algunas realizaciones, las composiciones y métodos de la invención se utilizan en un sistema de expresión vegetal, por ejemplo, un sistema de expresión de lenteja de agua, y la secuencia de nucleótidos heteróloga de interés es una proteína secretada. Las proteínas secretadas son traducidas normalmente a partir de polipéptidos precursores que incluyen un "péptido señal" que interacciona con una proteína receptora sobre la membrana del retículo endoplásmico (RE) para dirigir la translocación de la cadena del polipéptido en crecimiento a través de la membrana y al retículo endoplásmico para la secreción desde la célula. Este péptido señal es escindido a menudo del polipéptido precursor para producir un polipéptido "maduro" que carece del péptido señal. De esta manera, un polipéptido biológicamente activo es expresado en una planta, por ejemplo, lenteja de agua, a partir de un constructo de expresión que tiene un elemento de control de la expresión de la invención, o una variante o fragmento biológicamente activo del mismo, conectado operablemente a una secuencia de nucleótidos de interés que está

conectada operablemente adicionalmente con una secuencia de nucleótidos que codifica un péptido señal que dirige la secreción del polipéptido al medio de cultivo. Un "polipéptido biológicamente activo" hace referencia a un polipéptido que tiene la capacidad de realizar una o más funciones biológicas o un conjunto de actividades normalmente atribuidas al polipéptido en un contexto biológico. Los péptidos señal de plantas que dirigen la translocación de la proteína al retículo endoplásmico (para la secreción al exterior de la célula) son conocidos en la técnica. Véase, por ejemplo, la Patente de los Estados Unidos Núm. 6.020.169.

En una realización, el péptido señal es el péptido señal de quitinasa de *L. minor* novedoso mostrado en el SEQ ID NO: 16, o una variante o fragmento del mismo, y el constructo de expresión incluye una secuencia de nucleótidos que codifica este péptido señal conectado operablemente a una secuencia de nucleótidos de interés. En algunas realizaciones, esta secuencia que codifica el péptido señal es la secuencia mostrada en el SEQ ID NO: 15.

Se admite que el péptido señal de quitinasa de *L. minor* de la invención, o las variantes o fragmentos del mismo, se pueden utilizar para dirigir la secreción extracelular de cualquier polipéptido codificado de interés. De esta manera; la secuencia que codifica el péptido señal del SEQ ID NO: 15, o una variante o fragmento del mismo, se pueden incorporar a cualquier constructo de expresión de manera que estén conectados operablemente en un marco de lectura apropiado a un promotor de interés y una secuencia de nucleótidos que codifica un polipéptido de interés. Dicho constructo de expresión puede ser introducido en una planta o célula o nódulo vegetal para proporcionar la expresión y la secreción extracelular del polipéptido codificado de interés.

Alternativamente, se puede utilizar un péptido señal de mamífero para dirigir los polipéptidos recombinantes expresados en plantas diseñadas por ingeniería genética, por ejemplo, lenteja de agua, para su secreción. Se ha demostrado que las células vegetales reconocen péptidos señal de mamíferos que eligen como diana el retículo endoplásmico, y que estos péptidos señal pueden dirigir la secreción de los polipéptidos no solamente a través de la membrana plasmática si no también a través de la pared celular vegetal. Véanse, por ejemplo, las Patentes de los Estados Unidos Núms. 5.202.422 y 5.639.947.

El polipéptido secretado puede ser cosechado del medio de cultivo por medio de cualquier método convencional conocido en la técnica y purificado mediante cromatografía, electroforesis, diálisis, extracción disolvente-disolvente, y similares.

Los métodos de la invención implican la introducción de un constructo de expresión en una planta o célula o nódulo vegetal. Por "introducción" se entiende presentar a la planta un constructo de expresión de tal manera que el constructo tiene acceso al interior de una célula de la planta. Los métodos de la invención no dependen de un método concreto para la introducción de un constructo de expresión en una planta, solamente de que el constructo de expresión tenga acceso al interior de al menos una célula de la planta. Los métodos para la introducción de constructos de expresión en plantas son conocidos en la técnica incluyendo, pero no limitados a, métodos de transformación estables, métodos de transformación transitorios, y métodos mediados por virus.

Por "transformación estable" se entiende que una secuencia de nucleótidos introducida en una planta se integra en el genoma de la planta y es capaz de ser heredada por la progenie de la misma. Por "transformación transitoria" se entiende que una secuencia de nucleótidos (p. ej., una secuencia de nucleótidos contenida en un constructo de expresión) introducida en una planta no se integra en el genoma de la planta.

Las secuencias de nucleótidos de la invención se pueden introducir en plantas o células o nódulos vegetales poniendo en contacto las plantas o células o nódulos vegetales con un virus o con ácidos nucleicos virales. Generalmente, tales métodos implican la incorporación de una secuencia de nucleótidos de la invención a una molécula de ADN o ARN viral. Los métodos para introducir secuencias de nucleótidos en plantas o células o nódulos vegetales y expresar una proteína codificada en ella, que incluyen moléculas de ADN o ARN virales, son conocidos en la técnica. Véanse, por ejemplo, las Patentes de los Estados Unidos Núms. 5.889.191, 5.889.190, 5.866.785, 5.589.367, y 5.316.931.

Los protocolos de transformación así como los protocolos para introducir secuencias de nucleótidos en plantas pueden variar dependiendo del tipo de planta o célula o nódulo vegetal, esto es, monocotiledónea o dicotiledónea, elegido como diana para la transformación. Los métodos adecuados para introducir secuencias de nucleótidos en plantas o células o nódulos vegetales incluyen microinyección (Crossway et al. (1986) *Biotechniques* 4:320-334), electroporación (Riggs et al. (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606), transformación mediada por *Agrobacterium* (Patentes de los Estados Unidos Núms. 5.563.055 y 5.981.840, ambas las cuales se incorporan a la presente memoria como referencia), transferencia génica directa (Paszowski et al. (1984) *EMBO J.* 3:2717-2722), y aceleración de partículas balísticas (véanse, p. ej., las Patentes de los Estados Unidos Núms. 4.945.050; 5.879.918; 5.886.244; y 5.932.782 (cada una de las cuales se incorpora a la presente memoria como referencia); y Tomes et al. (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," en *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg y Phillips (Springer-Verlag, Berlín); McCabe et al. (1988) *Biotechnology* 6:923-926). Las células que han sido transformadas se pueden hacer crecer en plantas de acuerdo con los métodos convencionales. Véase, por ejemplo, McCormick et al. (1986) *Plant Cell Reports* 5:81-84.

En algunas realizaciones, las plantas de lenteja de agua o las células o nódulos de lenteja de agua transformadas establemente expresan polipéptidos biológicamente activos que no pueden ser producidos comercialmente por los sistemas de expresión génica existentes, debido a limitaciones de coste o logísticas, o ambas. Por ejemplo, algunas proteínas no pueden ser expresadas en sistemas de mamífero debido a que la proteína interfiere en la viabilidad celular, la proliferación celular, la diferenciación celular, o el ensamblaje de proteínas en células de mamífero. Tales proteínas incluyen, pero no están limitadas a, proteína del retinoblastoma, p53, angiostatina, y leptina. La presente invención se puede emplear ventajosamente para producir proteínas reguladoras de mamífero; es poco probable dada la gran distancia evolutiva entre las plantas superiores y los mamíferos que estas proteínas interfieran en los procesos reguladores de la lenteja de agua. También se puede utilizar lenteja de agua transgénica para producir grandes cantidades de proteínas tales como albúmina de suero (en particular, albúmina de suero humana), hemoglobina, y colágeno, que cuestionan las capacidades de producción de los sistemas de expresión existentes.

Adicionalmente, se pueden diseñar sistemas de plantas superiores para producir proteínas multiméricas biológicamente activas (p. ej., anticuerpos monoclonales, hemoglobina, oxidasa P450, y colágeno, y similares) bastante más fácilmente de lo que pueden los sistemas de mamífero. Un enfoque ilustrativo para producir proteínas multiméricas biológicamente activas en la lenteja de agua utiliza un constructo de expresión que contiene los genes que codifican todas las subunidades del polipéptido. Véanse, p. ej., During et al. (1990) *Plant Mol. Biol.* 15:281 y van Engelen et al. (1994) *Plant Mol. Biol.* 26:1701. Este constructo es introducido después en células de lenteja de agua utilizando cualquier método de transformación conocido, tal como un bombardeo balístico o transformación mediada por *Agrobacterium*. Este método da como resultado líneas celulares clónicas que expresan todos los polipéptidos necesarios para ensamblar la proteína multimérica. Una variación de este enfoque consiste en elaborar constructos génicos sencillos, mezclar el ADN de estos constructos entre sí, liberar después esta mezcla de ADN en células vegetales utilizando el bombardeo balístico o la transformación mediada por *Agrobacterium*. Como una variación adicional, algunos o todos estos constructos pueden codificar más de una subunidad de la proteína multimérica (esto es, de manera que haya menos clones de lenteja de agua que cruzar que el número de subunidades de la proteína multimérica). Alternativamente, cada clon de lenteja de agua expresa al menos una de las subunidades de la proteína multimérica, y los clones de lenteja de agua que secretan cada subunidad se cultivan juntos y la proteína multimérica se ensambla en el medio a partir de las diferentes subunidades secretadas. En algunos casos, puede ser deseable producir no todas las subunidades de una proteína multimérica, o incluso una única subunidad de la proteína, en una planta de lenteja de agua o un cultivo de nódulos de lenteja de agua transformados, por ejemplo, para procedimientos industriales o químicos o para fines de diagnóstico, terapéuticos, o de vacunación.

Los siguientes ejemplos se ofrecen con fines ilustrativos, pero no como limitación.

## SECCIÓN EXPERIMENTAL

### Ejemplo 1: Vectores de Expresión

Los vectores de expresión utilizados en los ejemplos descritos más abajo incluyen Egs05, Egs07, Egs11, Egs22, Egs23, Egs46, Egs50, Egs51, Egs19, Egs20, Egs24, Egs25, IFN53, e IFN54. Egs05 y Egs07 son vectores de expresión no modificados que comprenden un promotor de control conectado operablemente a la secuencia codificante de  $\beta$ -glucuronidasa (GUS) de *E. coli*, cada uno con un gen marcador seleccionable diferente. Egs11 comprende el elemento de control de la expresión de ubicuitina de *L. minor* completo (SEQ ID NO: 1) conectado operablemente a la secuencia codificante de GUS, con un gen marcador seleccionable. Egs22 y Egs23 son constructos similares, pero utilizan versiones truncadas del elemento de control de la expresión de ubicuitina de *L. minor*. En Egs22, los nucleótidos 1288-2160 del SEQ ID NO: dirigen la expresión de la secuencia codificante de GUS conectada operablemente. En Egs23, los nucleótidos 1132-2160 del SEQ ID NO: dirigen la expresión de esta secuencia codificante de GUS. Egs46 es similar a Egs 11, pero comprende un gen marcador seleccionable diferente.

Egs50 comprende el elemento de control de la expresión de ubicuitina de *S. polirrhiza* completo (SEQ ID NO: 2) conectado operablemente a la secuencia codificante de GUS, con un gen marcador seleccionable. De un modo similar, Egs51 comprende el elemento de control de la expresión de ubicuitina de *L. aequinoctialis* completo (SEQ ID NO: 3) conectado operablemente a la secuencia codificante de GUS, con un gen marcador seleccionable.

Egs19 comprende los nucleótidos 461-1808 del elemento de control de la expresión de r-histona de *L. minor* (SEQ ID NO: 13) conectado operablemente a la secuencia codificante de GUS, con un gen marcador seleccionable. En Egs20, los nucleótidos 805-1808 del SEQ ID NO: 13 dirigen la expresión de la secuencia codificante de GUS.

Egs24 comprende los nucleótidos 51-1338 del elemento de control de la expresión de quitinasa de *L. minor* (SEQ ID NO: 14) conectado operablemente a la secuencia codificante de GUS, con un gen marcador seleccionable. Egs25 comprende los nucleótidos 51-1338 del elemento de control de la expresión de quitinasa de *L. minor* (SEQ ID NO: 14) conectado operablemente al intrón ADH1 de maíz y a la secuencia codificante de GUS, con un gen marcador seleccionable.

Los vectores de expresión IFN53 y IFN54 contienen cada uno el súper promotor AmasPmas, el elemento de control

de la expresión de RbcS SSU5B de *L. gibba* (SEQ ID NO: 12), y el intrón ADH1 de maíz conectado operablemente a un gen de interferón alfa-2b con los codones optimizados, con una secuencia señal de alfa amilasa con los codones optimizados (IFN53) o la secuencia señal de quitinasa de *L. minor* (SEQ ID NO: 15; IFN54).

#### Ejemplo 2: Transformación de Lenteja de Agua

5 Las frondas de lenteja de agua o los cultivos de nódulos de lenteja de agua (obtenidos de la cepa 8627 de *Lemna minor* en estos ejemplos) se transformaron con los constructos de expresión descritos más arriba utilizando métodos de transformación mediada por *Agrobacterium*. Se utiliza la cepa C58Z707 de *Agrobacterium tumefaciens*, una cepa C58 con una gama amplia de anfitriones, desarmada (Hepburn et al. (1985) *J. Gen. Microbiol.* 131:2961-2969) para la transformación en estos ejemplos. Los constructos de expresión descritos más arriba se movilizaron en *A. tumefaciens* mediante electroporación, o mediante un procedimiento de apareamiento triparental utilizando *E. coli* MM294 que albergaba el plásmido movilizado pRK2013 (Hoekema et al. (1983) *Nature* 303:179-180; Ditta et al. (1980) *Proc. Natl. Acad. Sci. USA* 77:7347-7350). Las cepas C58Z707 que comprenden los constructos de expresión descritos más arriba se siembran en estría sobre medio mínimo AB (Chilton et al. (1974) *Proc. Nat. Acad. Sci. USA* 71:3672-3676) o en medio YEB o LB (1 g/L de extracto de levadura, 5 g/L de extracto de ternera, 5 g/L de peptona, 5 g/L de sacarosa, 0,5 g/L de MgSO<sub>4</sub>) que contenía estreptomina a 500 mg/L, espectinomicina a 50 mg/L y sulfato de kanamicina a 50 mg/L y se hicieron crecer durante la noche a 28°C.

Los cultivos de nódulos de lenteja de agua para la transformación se produjeron como sigue. Se separaron frondas de lenteja de agua, las raíces cortaron con un escalpelo estéril, y las frondas se colocaron, con la cara ventral hacia abajo, sobre medio de Murashige y Skoog (número de catálogo M-5519; Sigma Chemical Corporation, St. Louis, MO) pH 5,6, con un suplemento de ácido 2,4-diclorofenoxiacético 5 µM, 1-Fenil-3(1,2,3-tiadiazol-5-il)urea tiazuron 0,5 µM (Sigma P6186), sacarosa al 3%, Bacto-agar Difco 0,4 (Fisher Scientific), y Gelrite al 0,15% (Sigma). Las frondas se hicieron crecer durante 5-6 semanas. En este momento, aparecieron los nódulos (masas de células amarillentas, pequeñas), generalmente de la parte central de la cara ventral. Este tejido nodular se desprendió de la fronda madre y se cultivó en medio de Murashige y Skoog con un suplemento de sacarosa al 3%, Bacto-agar Difco al 0,4%, Gelrite al 0,15%, ácido 2,4-diclorofenoxiacético 1 µM, y benciladenina 2 µM.

Los cultivos de nódulos de lenteja de agua se transformaron como sigue. Se hizo crecer la cepa de *Agrobacterium tumefaciens* apropiada sobre agar patata dextrosa o agar YEB o LB con 50 mg/L de kanamicina y acetosiringona 100 µM, y se resuspendió en medio de Murashige y Skoog con un suplemento de manitol 0,6 M y acetosiringona 100 µM. El tejido de cultivo de nódulos se inoculó sumergiendo la solución de bacteria resuspendida durante 1-2 minutos, se secó para eliminar el exceso de fluido, y se cultivó en placa sobre medio de co-cultivo que consistía en medio de Murashige y Skoog con un suplemento de auxina y citoquinina optimizado para promover el crecimiento de los nódulos y acetosiringona 100 µM. Véase, Yamamoto et al. (2001) *In Vitro Cell Dev. Biol. Plant* 37:349-353.

Para la selección, se transfirió el tejido de cultivo de nódulos a medio de regeneración; 0,5 X medio de Schenk y Hildebrandt con un suplemento de sacarosa al 1%, Bacto-Agar Difco al 0,4%, Gelrite al 0,15%, 500 mg/L de cefotaxima, y 6 mg/L de geneticina y se cultivó durante aproximadamente 6-8 semanas bajo iluminación continua (20-40 µM/m<sup>2</sup>·seg). El tejido nodular se transfirió cada 7 días a medio de cultivo de nueva aportación. La selección es completa cuando el tejido nodular muestra un crecimiento vigoroso sobre el agente de selección.

#### Ejemplo 3: Expresión Transitoria de GUS de *E. coli* en Lenteja de agua

Se evaluó la expresión de GUS transitoria en cultivos de nódulos de lenteja de agua transformados con los constructos Egs05, Egs07, Egs11, Egs22, Egs23, Egs46, Egs50, Egs51, Egs 19, Egs20, Egs24, y Egs25. Todos los constructos fueron capaces de dirigir la expresión fuerte de GUS, determinada por medio de una tinción de 24 horas (Tabla 3).

Adicionalmente, se llevaron a cabo análisis de enzima GUS en cultivos de nódulos de lenteja de agua transformados con los constructos Egs07, Egs46, Egs50, y Egs51. Las 36 líneas transgénicas Egs07 promediaron 1,345% de GUS, las 29 líneas transgénicas Egs46 promediaron 2,320% de GUS, las 4 líneas transgénicas Egs50 promediaron 4,008% de GUS, y las 8 líneas transgénicas Egs51 promediaron 6,682% de GUS.

Tabla 3: Expresión de GUS Transitoria en Callo

Vector de Ensayo	Promotor	Tinción
Egs05	control	++++
Egs07	control	++++
Egs11	LmUBQ	++++
Egs22	LmUBQ (trunc núm. 1)	+++
Egs23	LmUBQ (trunc núm. 2)	++++
Egs46	LmUBQ	++++
Egs50	SpUBQ	++++
Egs51	LaUBQ	+++
Egs19	LmHIS (461-1808)	++
Egs20	LmHIS (805-1808)	+
Egs24	LmCHT (51-1338)	+
Egs25	LmCHT (51-1338) + intrón ADH1	++
tinción 24 horas; clasificado en una escala de 1 a 4.		

Ejemplo 4: Expresión de Interferón en Lenteja de Agua

5 Se produjeron varios cientos de líneas de lenteja de agua transgénicas utilizando los constructos IFN53 y IFN54 y con posterioridad se escrutaron para determinar la expresión de interferón mediante ELISA. Se observaron niveles similares de expresión de interferón para los dos constructos. IFN53: máxima expresión, 1735,66 ng/ml; expresión media, 362,04 ng/ml. IFN54: máxima expresión, 1173,81 ng/ml; expresión media, 347,40 ng/ml.

LISTADO DE SECUENCIAS

<110> Dickey, Lynn F.  
Cox, Kevin M.  
5 Peele, Charles G.

<120> ELEMENTOS DE CONTROL DE EXPRESIÓN DE LA FAMILIA LEMNACEAE

<130> 40989/322154  
10

<150> US 60/848,961  
<151> 10-03-2006

<150> US 60/759,308  
15 <151> 17-01-2006

<160> 16

<170> Versión PatentIn 3.3  
20

<210> 1  
<211> 2160  
<212> ADN  
<213> Lemna minor  
25

<400> 1

agtcgagtga tatgaaatct tggatgaagaa ggatcggaga acggaccggg tgaggcaagg 60  
 ataattctgc tgtaaattc gagagcaaga cacctgcaat tcaagaatcg agtggcaatt 120  
 aatatagcag gatgatctgg aaggtagatc ctgccatcg aatgatccaa acatcaacac 180  
 taggatcata ccgttaacaa taatgaatga aaaagtagaa gatgacgaag ttgaagtgat 240  
 gaccaaaaac tttgaaaatt ccaaccgtat ggccggaatc agtgtgaaga aaatcgaat 300  
 caaatactct aatggatcgg attgttattc tggaggcaaa tctgaaactt cgaggatagg 360  
 atttaatcca cgcaagtaat aatttgaac tcagaaggag aaaaaaaac taaaatagag 420  
 aagaagagat ctcaagaag ccgtgagcac gagacgaacg agaagaggtg aagcaccagt 480  
 cagaggaaaa caccaaaatt agagaaatag cacgaacatt aaagcacaga tccgcgccgc 540  
 aaacccgaaa gacgaaaaat agagccaaac gaaacctaa taatcgatct gcacaaaaaa 600  
 aaaaaaaaaa aactttgaga agagccgcga aattacccta gaatcctcag aactggccgg 660  
 acgagagaag cgctcgatcg aaacccaaca taaaacctt tccaacggca aattactccg 720  
 caaaaccgca aaaataaaca aatcaacga tcacgagaag gtgcaagggc aaaaagaggc 780  
 agtgcgatcg agagtctacc tgaatcgtcg ggcgaaaagg cgagcccacc gacgaacgct 840  
 ccctctagaa cctggagatg cggcgagaga gaaggaaaga tcttcggtgg gtgatgctcg 900  
 ctatttatcg caagagagt agagagatct tcttcggcgg cggatttctg gcatctagcg 960  
 tttaacctca ccgcccagtg ctcacatcct tcttctcata tttgaatatt taattaacaa 1020  
 atgaatcagt catTTTTctt taatttttaa tccccgaga gggcaatgtt ggtatcaaaa 1080  
 attatttagg aaaaattaat tacacgaata atcggatTTT tccctTTTT taattaattt 1140  
 ctaattttgg aaaaggaaag aaaaatttta ggggtatgga gggcaagaat gaaatattac 1200

aaaattaggg gtttttgcgt aatttattat atttaataaa gaaagtcgaa tattcccatc 1260  
 cgattggtag ttgaaagggg cggaaaggcc tcgggggttc tagagatttc tacattatc 1320  
 tcgtttttgt cgccaagaag gtgggcaatt atgtttcatg ccttaacttc ttctttttgt 1380  
 gggaaactc ttattcttag tacaaaagaa aagagtatat gcataaataa gatgaaaaat 1440  
 gggtttattc gagatttcta cgtcatgtgt gactcgctta ggaaatatcg ccgaaaccta 1500  
 acaaaggcgg tacgctcete tcccecgacc tataaataga gacctttgcc tcgtctttct 1560  
 caactcaage atttctgtat gatccttctc tttccgogga agctctcgcg ccagttgatc 1620  
 gcaaggatg cgtctttcct cettgtgatt cgatctttct gttggctaga tctggctctat 1680  
 tgatctgctc tattgatctg gtctatttat cgctgcacg ggatctattg atccgtatgt 1740  
 tgatttggga tccgtaggtt ggtttgatc ggagactgcg atttgattct tgtgatttcg 1800  
 cttggatttc ggaaatcggg gtggttgaag tcgtgogatc ttttagatct gctccttttt 1860  
 ttatttgcta ttttatattt acgttgttta tgatcoggga ttattttgat tcgtttattc 1920  
 gagatccatg ccgtttaact cgttctttgt gctccgatct ttgcgatacg tcggtegttc 1980  
 tagatccgtt cactagggta gttttaagtt ctttgagctt gatttatatg gatttgctgt 2040  
 tttccaggaa aaatttatgc gcgattetta cgcccgttc cccattttac tttaggtcgt 2100  
 gaattctttt gatctgagaa tgatgaatct gscatgtacc ttccggtttg taatttgcag 2160

<210> 2

<211> 2021

<212> ADN

5 <213> Spirodela polyrrhiza

<400> 2



ES 2 398 142 T3

**caaataaaga gatggacaga taatgagatg aattagaaaa aaaaaattcg tgttgtaaga 60**  
**tagaataactt gctatctact gatgaatgca gttcagtttt cctcacgatac ttaaagatcg 120**  
**cgcactatcc tcagcttcac tctggaaatt ttgattctct tcttctctc agcagcctcg 180**  
**actctgtcta gggtttcgta caatcggacg ccattctaca tgaatcgagc acagggaatg 240**  
**aagacaatta ggagatcctc gatgtcctcc gacttaactg catgacttga cggggaagat 300**  
**ctcgagcagg gaagcgacgc ctctccggag gactcgctc gccgagagga cctcctccgc 360**  
**gacacggacc atggcctcca cggggtagaa gctggccctg ttctttatc tcttgaggat 420**  
**catcggccga agcctccga aatccatccc cgaggagttag aatctcgct gcaggaagca 480**  
**tctgtcgaga tcctcgccga ggcgccggag ataacctgcc ggcgcccca tggcgccggg 540**  
**gacggagcac caccacggag aagaagaacc ctaacccaag gcattaacga agttgcgcag 600**  
**attatacaaa agccctcaaa tatcttcat tttctattc actgatacat tttcattatt 660**  
**gtatatgagt gtttatttaa attattccgt attagaaaag cacctccaga acccgacaaa 720**  
**atagggtgac gtcateatgg tgtcatgacc gcccaacagc cgcagattta aaatcgggtg 780**

atgagtgcgg ccacgccacg aaagcgatgg gccttcgctg atgccgtgag aatccatctg 840  
 acataaagta aacggcgccg tcagtattga cggcgtatga cacgtggaaa gaagctattg 900  
 gttcacgcat cggtggttcc gctagcctcc gtcgaccgct agtactataa atacggtccc 960  
 gaggcctcct caccactcgc acatatectc tttgttttcc tetccgtgaa agaagcgagg 1020  
 aagcgcgctg tctctcccaa ggtaaggagc agatctcttt gatcgttttt gttcttcttt 1080  
 tgttttgttt tttttttctg cggatcttcg gttgcatcat gccttggctg tttttattag 1140  
 tttaggatat cctcgtttgg atctgagccg atcatatatg ttaaagggtg tgttcgatct 1200  
 ctttgttcat tttcgcataa aaaggatgta tccttttgat gtgaggcgat cttctatggt 1260  
 taagactttg ttcggtctat tgatcatttc tgttcttcgt ttttgagttt tttctgccc 1320  
 atatcgcatc atccctaggt ttttgctttg gttaggatgc atcctttgga tttgagccga 1380  
 tctcccttgg ttaaggctgt gtctgttgca gaggagaaag tctgtcgagg tccttatgca 1440  
 ggctttgtcc agatgcgctg gctctctcat gctatgaatt tatgttttga gaactcctcc 1500  
 cggtttttct agatccggat ttgaagtatt cattgcggtt ccccttcggt tttatgtatt 1560  
 tctcgagttg atttggcca tgatcgtgtt ctgtccagat ctctcttgat atggatgaga 1620  
 tattcgttac ctctttcaaa catcgggtgga tgttcttttt agtcttgget cacctttatc 1680  
 tagaaattaa ttttcggttt gaaaccctcg cttgttaagg tgatgtattc cttctttata 1740  
 gatttcggtg tgttatttct taacggtgat ctgtccgata catgtgttgc acctcttgtt 1800  
 ttctgtgtaa tcctctgtga attataatta tgttttgaaa acgtaactaa gtaaggggca 1860  
 tgttccccgt ttaaaacttt tgttctatca atttgtggtt aatagatcct gatttgtggt 1920  
 cgccttattc tgtctttaat cgtggatttt atttatcttg agcgcgtcct tttcttttaa 1980  
 aatcatgtgt ttaaccttcc agtcgtcata tgttccatca g 2021

<210> 3

<211> 2068

<212> ADN

5 <213> Lemna aequinoctialis

<400> 3

ES 2 398 142 T3

agtgtaccaa tattttaaac cctacattta tcattcttta ttcattattg ccataagtta 60  
atgaatattg aaattcaaat acgcgcaaga tgtcaatata gatogaatat gaataccaga 120  
tataaatca aaaatcaaat atcaaattaa taaagatata aaatattgaa tccagaagca 180  
ataaagaata tcactattaa tatcaaaata togatttgaa gttcaaaaat tgggtccatt 240  
aggagccaag accgatcatg atccgatact gatatacaata tctgtagctc agtggctagg 300  
cccctcaatt tgcttggccg aaggcagtgt acaaaacctg gotctcgcaa gggcaagaa 360  
agagtctttc ccaaaaaaaaa aaaaatcgaa cccatttgta gtatccaata tttggattga 420  
cataagatac caaacataa agtactaacc acccaatctt ataattaatc aagatttata 480

tcacatccaa tatcaagatc cgatatcaat acctagaccg gtaaacccta atttactctt 540  
ccccctcta aaaatttcca ataaatatct ccacatattt aactattaaa aattgataa 600  
gagataggcc ctagccctaa gtccatacat ataaccactc tctatgaaa gtcctattaa 660  
atgacgtcat ttatttattt attgccggtt ggctgctcca cagccgcaat ttaatggatg 720  
gctgacacgg cacgaaaccg acgggcggtg ccgtgggaat aattctagag taaacctaac 780  
ggcgccggtt actttgacgg tggcgaagac gcgtggggat aggtggttgg tccgctgac 840  
ggcgccggtt cagcccgctg accttgagcc gagactataa atcgaggcga agggatgagc 900  
tttgccattg cgttctctct ctgttcactc ctgaaattcg ggccggaatcc ttcttctctc 960  
caaggtatgg gcctcgatct ttctgtttca atcgagtttt gatcttcggt ttggccgca 1020  
tcggtgtttt ctttgtattg tgaataaatc ottgataaga aaacctagg tttgtgacc 1080  
tgttgacgga tgcgtgcgga tctgttattt gtcttttagg cgattttctc ttgtttgtaa 1140  
tagtttatca taaccagatg aacatggatc aagtcgattt gactattttt ttctgtgaaa 1200  
ttagcccgaa atcctttttt ttggtttgag ccttgatatt tetatataat tcgatttgat 1260  
tttttgtttt ctctcgctc tgatgcttcc tcttgactcc tgattaaatt tttgctacgg 1320  
aaacctaga tgcgagatc tgttgacaga ttctggcaa tctgttttta tcataatcsg 1380  
atgaacgcaa attaagtcga tttggttttt ctctgaaatt aggggggaaa ctcttatag 1440  
tatgagcctc gatatttcta taatagtcga tttgatttcc tcttgctccc tgattcaatt 1500  
tttggtgccg aaacctaga tattgtaatc tgtttacgga tgcttgccga tctgattttt 1560  
aatattgtga tctattgacg gatgctcgtc gatctggttg ttttgatttc tcatgcctt 1620  
atacggcgat ttgattcggc gattaaaaat tttcaattct tttaaaaaa atattaagat 1680  
tttcaacggt tcaaattatt tcatagatcg gcacaaatac ttttcatcag attcctcctg 1740  
atgtgatggt ttgtgtttaa aatctgttga agatcagaga ttctattagg tcaccgatat 1800  
aatcttctct gtttattctg cgatcggtgc ttacaaacco tatttctac ggtgattaat 1860  
tatttttaat ctctageta gcgtaaatat atattttttt aatttgatct ttgcattagt 1920  
ttctctcttt tatttgctat taattgtaac cgatgctaca aaacatcaga ttttttttcc 1980  
caattcgttg tcatcattat agaaaacttt tatctgatat ttttaatcgt cattaatata 2040  
attttcaatt tattattttc ccttgacg 2068

<210> 4  
<211> 1625  
<212> ADN  
5 <213> Lemna minor

<400> 4

agtcgagtga tatgaaatct tggggaagaa ggatcggaga acggaccggg tgaggcaagg 60  
ataattctgc tgttaaattc gagagcaaga cacctgcaat tcaagaatcg agtggcaatt 120

aatatagcag gatgatctgg aaggtagatc ctgcccacgc aatgatccaa acatcaacac 180  
taggatcata ccgttaacaa taatgaatga aaaagtagaa gatgacgaag ttgaagtgat 240  
gaccanaaac tttgaaaatt ccaaccgtat ggccggaatc agtgtgaaga aaatcgaat 300  
caaatactct aatggatcgg attgttattc tggaggcnaa tctgaaactt cgaggatagg 360  
atttaatcca cgcaagtaat aatttgaac tcagaaggag aaaaaaac taaaatagag 420  
aagaagagat ctcaaagaag ccgtgagcac gagacgaacg agaagaggta aagcaccagt 480  
cagagggaaa caccaaaatt agagaaatag cacgaacatt aaagcacaga tccgcgccgc 540  
aaaccgaaa gacgaaaaat agagccaaac gaaaccctaa taatcgatct gcacaaaaaa 600  
aaaaaaaaaa aactttgaga agagccgcga aattacccta gaatcctcag aactggccgg 660  
acgagagaag cgctcgatcg aaaccaaca taaaaccctt tccaacggca aattactcgg 720  
cnaaacccga aaaataaaca aatcaacga tcacgagaag gtgcaagggc aaaaagaggc 780  
agtgcgatcg agagtctacc tgaatcgtcg gcgcaaaagg cgagcccacc gacgaacgct 840  
ccctctagaa cctggagatg cggcgagaga gaaggaaaga tcttcggtgg gtgatgctcg 900  
ctatttatcg caagagagtt agagagatct tcttcggcgg cggatttctg gcactatagc 960  
ttaacctca ccgccagtg ctcacatcct tcttctcata tttgaatatt taattaacaa 1020  
atgaatcagt ctttttctt taatttttaa ttcccggaga gggcaatggt ggtatcaaaa 1080  
attatttagg aaaaattaat tacacgaata atcggatttt tccctttttt taattaattt 1140  
ctaattttgg aaaaggaaag aaaaatttta ggggtatgga gggcaagaat gaaatattac 1200  
aaaattaggg gtttttgcgt aatttattat atttaataaa gaaagtcgaa tattcccatc 1260  
cgattggtag ttgaaagggg ccgaaaggcc tcgggggttc tagagatttc tacattattc 1320  
tcgtttttgt cgccaagaag gtgggcaatt atgtttcatg ccttaacttc ttctttttgt 1380  
gggaatactc ttattcttag tacaaaagaa aagagtatat gcataaataa gatgaaaaat 1440  
gggtttattc gagatttcta cgtcatgtgt gactcgttta ggaatatcg ccgaaacctc 1500  
acaaaggcgg tacgctctc tccccgacc tataaataga gacctttgcc tcgtctttct 1560  
caactcaagc atttctgtat gatccttctc tttccgcgga agctctcgcg ccagttgatc 1620  
gcaag 1625

- <210> 5
- <211> 1041
- <212> ADN
- 5 <213> Spirodela polyrrhiza
- <400> 5

caataaaga gatggacaga taatgagatg aattagaaaa aaaaaattcg tgttgtaaga 60  
 tagaataactt getatctact gatgaatgca gttcagtttt cctcagatc ttaagatcg 120  
 cgcactatcc tcagcttcac tctggaaatt ttgattctct tcttctgctc agcagcctcg 180

actctgtcta gggtttcgta caatcggacg ccattctaca tgaatcgagc acagggaatg 240  
 aagacaatta ggagatcctc gatgtcctcc gacttacttg catgacttga cggggaagat 300  
 ctcgagcagg gaagcgacgc ctctccggag gactcgcctc gccgagagga cctcctccgc 360  
 gacacggacc atggcctcca cggggtagaa gctggcctg ttctttatc tcttgaggat 420  
 catcgccga agcctccga aatccatccc cgaggagtag aatctcgctc gcaggaagca 480  
 tctgtcgaga tcctcgccga ggcggcgag atacctcgcc ggcgcccga tggcgccggg 540  
 gacggagcac caccacggag aagaagaacc ctaaccaag gcattaacga agttgcgcag 600  
 attatacaaa agcctcaaa tatcttcat tttctattc actgatacat tttcattatt 660  
 gtatatgagt gtttatttaa attattccgt attagaaaag cacctccaga acccgacaaa 720  
 atagggtgac gtcacatgg tgatcagacc gcccaacagc cgcagattta aaatcgggtg 780  
 atgagtgcgg ccacgccacg aaagcgatgg gccttcgtcg atgccgtgag aatccatctg 840  
 acataaagta aacggcgccg tcagtattga cggcgtatga cacgtggaaa gaagctattg 900  
 gttcaogcat cggtggttcc gctagcctcc gtcgaccgct agtactataa atacggtccc 960  
 gaggcctect caccactcgc acatatactc tttgttttcc tctccgtgaa agaagcgagg 1020  
 aagcgcgtcg tctctcccaa g 1041

<210> 6

<211> 964

<212> ADN

5 <213> Lemna aequinoctialis

<400> 6

agtgtacca	tattttaaac	cctacattta	tcattcttta	ttcattattg	ccataagtta	60
atgaatattg	aaattcaaat	acgcgcaaga	tgtcaatata	gatcgaatat	gaataccaga	120
tataaaatca	aaaatcaaat	atcaaatata	taaagatata	aaatattgaa	tccaaaagca	180
ataaagaata	tcactattaa	tatcaaaaata	tcgatttgaa	gttcaaaaat	tgggtccatt	240
aggagccaag	accgatcatg	atccgatact	gatatacaata	tctgtagctc	agtggctagg	300
ccoctcaatt	tgcttgccg	aaggcagtgt	acaaaacctg	gctctcgcaa	gggcaaagaa	360
agagtctttc	ccaaaaaaa	aaaatcgaa	cccatttgta	gtatccaata	tttgattga	420
cataagatac	caaacataa	agtactaacc	acccaatott	ataattaate	aagatttata	480
tcacatccaa	tatcaagatc	cgatatcaat	acctagaccg	gtaaacccta	atttactott	540
cccccteta	aaaatttcca	ataaatatct	ccacatattt	aactattaaa	aaattgataa	600
gagataggcc	ctagccctaa	gtcctaacat	ataaccactc	tctatgaaaa	gtcctattaa	660
atgacgtcat	ttatttattt	attgccggtt	ggctgctcca	cagccgcaat	ttaatggatg	720
gctgacaacg	cacgaaaccg	acgggcggtg	ccgtgggaat	aattctagag	taaacctaac	780
ggcgccgtta	actttgacgg	tggcgaagac	gcgtggggat	aggtggttgg	tccgcgtgac	840
ggcgccggtt	cagcccgctc	accttgagcc	gagactataa	atcgaggcga	agggatgagc	900
tttgccattg	cgttcttctt	ctgttcatct	ctgaaattcg	ggcggaatcc	ttcttcttct	960
caag						964

- <210> 7
- <211> 535
- <212> ADN
- <213> Lemna minor

5

<400> 7

gtatgcgtct	ttctccttg	tgattcgatc	ttctgttgg	ctagatctgg	tctattgatc	60
tgctctattg	atctggteta	tttatcgctg	catcgggatc	tattgatccg	tatggtgatt	120
tgggatccgt	aggttggttt	ggatcggaga	ctgcgatttg	attcttgtga	tttcgcttgg	180
atctcgaaa	tcggtgtggt	tgaagtcgtg	cgatctttta	gatctgctcc	tttttttatt	240
tgetatttta	tatttacggt	gtttatgatc	gcggattatt	ttgattcggt	tattcgagat	300
ccatgcggtt	taactegtto	tttggtctcc	gatctttgcg	atacgtcggg	cgttctagat	360
ccgttcaacta	ggttagtttt	aagttctttg	agcttgattt	atatggattt	gctgttttcc	420
aggaaaaatt	tatgcgcgat	tcttacgccc	gtttcccat	tttactttag	gtcgtgaatt	480
cttttgatct	gagaatgatg	aatctgacat	gtaccttccg	gtttgtaatt	tgcag	535

ES 2 398 142 T3

<210> 8  
 <211> 980  
 <212> ADN  
 5 <213> Spirodela polyrrhiza

<400> 8

```

gtaaggagca gatctctttg atcgtttttg ttcttctttt gttttgtttt ttttttctgc      60
ggatcttcgg ttgcatcatg ccttggetgt ttttattagt ttaggatata ctogtttggg      120
tctgagccga tcatatatgt taaagggtgt gttegatctc tttgttcatt ttogcatgaa      180
aaggatgtat ccttttgatg tgaggcgatc ttctatggtt aagaactttgt tgggtctatt      240
gatcatttct gttcttegtt tttgagtttt tttctgogga tategcatca tccctagggtt      300
tttgctttgg ttaggatgca tectttggat ttgagccgat ctcccttggg taaggctgtg      360
tctgttgacg aggagaaagt ctgtcgaggt ccttatgcag gctttgtcca gatgcgcgtg      420
ctctctcatg ctatgaattt atgttttgag aactcctccc ggtttttcta gatccggatt      480
tgaagtattc attgcgggtc cecttcgggt ttatgtattt ctcgagttga tttggtocat      540
gatcgtgttc tgtccagatc tctcttgata tggatgagat attcgttacc tctttcaaac      600
atcgggtgat gttcttttta gtcttggtc acctttatct agaaattaat tttcggtttg      660
aaaccctgc ttgttaaggt gatgtattcc ttctttatag atttcgggtg gttattttott      720
aacgggtgac tgtccgated atgtgttgea cctcttgttt tctgtgtaat cctctgtgaa      780

ttataattat gttttgaaaa cgtacttaag taaggggcat gttccccgtt taaaactttt      840
gttctataca tttgtgggta atagatcctg atttgtgggc gccttattct gtctttaate      900
gtggatttta tttatcttga gcgcgtcctt ttcttttaaa atcatgtgtt taacctttca      960
gtcgtcatat gttccatcag                                          980
    
```

10 <210> 9  
 <211> 1104  
 <212> ADN  
 <213> Lemna aequinoctialis

15 <400> 9



**gtatgggect egatctttct gtttcaatcg agttttgato ttcgttttgg eggcgatcgg 60**  
**tgttttcttt gtattgtgaa taaatccttg ataagaaaac cctaggtttt gtgacctggt 120**  
**gacggatgcg tgcggatctg ttatttgtct tttaggegat tttctcttgt ttgtaatagt 180**  
**ttatcataac cagatgaaca tggatcaagt cgatttgact tattttttct gtgaaattag 240**  
**gccgaaatcc ttttttttgg tttgagcctt gatatttcta tataattcga tttgattttt 300**  
**tgttttcttc tgcgtctgat gctttctctt gactectgat taaatttttg ctacggaaac 360**  
**cctagatgtc gagatctggt gacagattct ggcaaatctg tttttatcat aatcagatga 420**  
**acgcaaatta agtcgatttg gtttttctct gaaattaggg gggaaactcc ttatagtatg 480**  
**agcctcgata tttctataat agtcgatttg attttctctt gcctctgat tcaatttttg 540**  
**gtgcggaaac cctagatatt gtaatctggt tacggatgct tgcggatctg atttttaata 600**  
**ttgtgatcta ttgacggatg ctcgtagatc tgggtgttt gatttcttca tgccttatac 660**  
**ggcgatttga ttcggcgatt aaaaattttc aattctttta aaaaaaatat taagattttc 720**  
**aacgtttcaa attatttcat agatcggcac aaatactttt catcagattc ctctgatgt 780**  
**gatggtttgt gtttaaaatc tgttgaagat atcagattct attaggtcac cgatataatc 840**  
**ttctctgttt attctgcgat cgggtgettac aaacctatt tctacgggtg attaattatt 900**  
**tttaatctcc tagctagcgt aatatatat ttttttaatt tgatctttgc attagtttcc 960**  
**tccttttatt tgctattaat tgtaaccgat gctacaaaac atcagatttt ttttcccaat 1020**  
**tcgttgtcat cattatagaa aacttttata tgatattttt aatcgteatt aatataattt 1080**  
**tcaatttatt attttccctt gcag 1104**

<210> 10  
 <211 > 64  
 <212> ADN  
 <213> Lemna giba

5

<400> 10

**aagcacgagc tgagcgagaa ttcggggagg ctgagtcgaa gaggaagaga gaagtaggta 60**  
**cgcc 64**

<210> 11  
 <211 > 58  
 <212> ADN  
 <213> Lemna gibba

10

<400> 11

actcgcaagt ggagagagga tccgagcgtc cagtgagagg aagagagagg gaggcgcg 58

15

<210> 12  
 <211 > 62  
 <212> ADN  
 <213> Lemna giba

20

<400> 12

**aaactcccga ggtgagcaag gatccggagt cgagcgcgaa gaagagaaag agggaaagcg 60**

**cg 62**

<210> 13

<211> 1808

5 <212> ADN

<213> Lemna minor

<220>

<221> característica\_misc

10 <222> (598)..(598)

<223> n e s a , c , g , o t

<400> 13

**agattgttgt gagagagcga cagcacctgg agcgcgggaa gctcgccgat cgccggagga 60**

**actatgcctt gaagcgcatt cccgtgcaag ctaatgtgca gaatggagct gcagttcgtg 120**

**agagcggctg gaagagtccc ttcgagaaga ttcagtgcca gccagaggta aagaagatcc 180**

**tgaagcccgc caatgtcgac cgggatcgtc cctgtgaggc gattgaacga gagattgatg 240**

**aaactggagac gaggagaggc ggagagactg octggaatct cgccggagaa gccattggag 300**

**gagggatcga ggtaccgaag ccccgccgga agagcgggag ggattccgcc ggagaaattg 360**

**ttgccagcga ggtttaggac ctgagggatg caagattgga agggaaatct ggccggaaggt 420**

**cgccggagag gcggttggtc tgcaagaaga gggatcggag ttggcggaga gaggagaatg 480**

**cgcccggaat gctgccattg aacgcgttgc cgcgagact gagccggcgg agttgagaaa 540**

**gatcggcgag gcgcccggag atcggaccag agagatggag aagagggagg cggagctnca 600**

**cgacgcggga atcgccggag gtggaacaga gaacgccgcg ccatgagcac ggagcggaaa 660**

**gcattgaggc gttccagtca gagagagccc cgtgccgata gtaaagggag aagcggaaact 720**

**tctccagcgc tctgatctec gccagggatt ccggagggat cttctgctcc tgccccgcgg 780**

**tggaaatgag gaagagaatg aagagaagga agtggaggag agaggacgcc atggtagcag 840**

**aggaaggtct ggcttgatct cccgacgatt cctctctcat cagtgaaca agagaataag 900**

**agggatcggc attcttgga ggtacagagg gaagttgata aaagagaggc tccgggaaga 960**

**agcagatggc ggggaagacg aaacatggcg cctgacaaca taggctatca taggataatc 1020**

15

ES 2 398 142 T3

caetctccct	ctgtctttct	ctctgtttct	ttctctctct	ctttatctct	ctctctctct	1080
cgaacatttc	acacattttt	gggcctgttc	tttggcgtag	ctagcccttt	cttgggccat	1140
atthtgaggc	caaggccaat	catgctgagc	cacgttgcac	ggcgggagga	ccccatctat	1200
ccattccgtc	agttcctggt	ttttgggaca	atctgaacag	tacataaacc	acgggctcgg	1260
gcttggggcc	gcgaaaage	ccgtccgtac	aattttctga	cgtacaatat	taatttecca	1320
gaaaagaaaa	ttcattaaaa	aataactatg	tcaccacgca	actcgtgac	tagatataag	1380
gcaggcaact	ttcccgtgac	agtcggactt	gtggggttcc	cttgacggcg	ccgccgtcta	1440
ctgacggcgc	ttgaatgacg	tcattattac	tatatttaac	atatccgaga	ataaatggct	1500
ccctggatcc	ccgtgaaatg	gtcacgtcat	ccacgcgggc	tttgacttcg	cctaggtcgc	1560
cacgttgggg	cctcataacg	tgaagctga	cgtggctatc	attgcctcga	tggcgatctt	1620
gaacgaacca	cataaccaat	ctgagacggc	gacgaggatt	ctgtgttttc	ctcggggatc	1680
tctggccgtc	cgatgaacta	acacgcacta	tttcaaaaag	ggggttaaac	acgattgtta	1740
ggtttctttg	agtcttccac	atctccgctc	cacgatcttg	tacctcttct	tgcgacgatc	1800
taetcgcc						1808

- <210> 14
- <211> 1338
- <212> ADN
- 5 <213> Lemna minor
- <400> 14

ES 2 398 142 T3

ttcgagtttg gtgttcttga atatTTTTTC ccattTTTTT tttctatttt tgagcttttc 60  
 tccgtctttg cggggatctt gtgtcggaga ggtggcccggt gccactggcg gtctacgggc 120  
 ggtttctctg gtaggttagg ctgttggaag tegtcatctc ccccaaaggc agcaggcggc 180  
 ggcgctccag cgtaaggcct aacgcgcgt ctgtcggaga tcgcctagcg tgatcagggc 240  
 ttcgggaacc tcccaggggg gagcacgatt tctctccatt cctctcttct tgatgtcttt 300  
 ggcggcttcg gtatgggaag ggatcaactg gtagcagcga agcagatggt ccaccatcat 360  
 gaagtcaatt aatacattta aatgagatg ttgatgggga attaaaataa gtaaaacaaa 420  
 atattgcaag tgctcatgtt tcaaacctcc ttgtcttagt tcttaagttc acatcacttg 480  
 aggtttcaca cctccatgcc ctagttcttg ccttggttct tgaattcaca ccacttaagg 540  
 gaatatttaa atatttaaaa tttatgttta actgattaag aagattgaac ctgtcggaga 600  
 gacgacattt agtttcttct actaggaaaa aaaattatte gagagaaagg gccaaagttt 660  
 cacacctcca taccctactt ctttaattca caccactcaa gggaatattt agatatttaa 720  
 attttatatt taactgatta agaagattgc acctaacgag gagacgatat ttagtttctt 780  
 ccactagaaa aaaaataaaa atactaaatt attcgagagg aaggcccga tgttcagcc 840  
 gccagtaaat attatccagg aaaaaaatt atcctatcca cgacaaaacc ttcttcacaa 900

aatacgataa tcagaataat gaatatgatg ttttgaaaat tcccgatagg gtttggcgag 960  
 atggcgtegt cagctagggg agaggtgaca gcgataggcg ctgaaaaact gccgcagcag 1020  
 acagagagcg ataaatgctg cacgcgcttt ctcttcaactt gggteagtgg gcgagttcga 1080  
 aggagacttg agtctcccct gggccagctc gttgaacggt tctacgcggg taatccatcc 1140  
 ttcggcgaac gcgagacccc gtctggtect ccgccgagtc ttcgatctgc tggaaaatgg 1200  
 caataatttc tctgatctcc cegtctttgg caaccagatg gcgtgtagac tggggctctc 1260  
 ttccacggac actaaataag taggaggcag ogaatggagt gtgaatcggc tcccctttc 1320  
 gtcttctctt ttgaagtc 1338

- <210> 15
- <211> 84
- <212> ADN
- 5 <213> Lemna minor

<400> 15

atggetgctt tgaacaatgc tegtcttate gcgctcttct tcagcttggg gttcgtactc 60  
 tccctgcctt ggctgcecca cgga 84

- 10 <210> 16
- <211> 28
- <212> PRT

<213> Lemna minor

<400> 16

Met Ala Ala Leu Asn Asn Ala Arg Leu Ile Ala Leu Phe Phe Ser Leu  
1 5 10 15

Gly Phe Val Leu Ser Leu Pro Trp Pro Ala His Gly  
20 25

5

**REIVINDICACIONES**

1. Una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos seleccionada del grupo que consiste en:
  - a) una secuencia de nucleótidos que comprende la secuencia mostrada en el SEQ ID NO: 1 o 4;
  - 5 b) una secuencia de nucleótidos que comprende al menos 350 nucleótidos contiguos de la secuencia mostrada en el SEQ ID NO: 1 o 4, en donde dicha secuencia de nucleótidos inicia la transcripción en una célula vegetal, y en donde dicha secuencia de nucleótidos comprende una caja TATA;
  - c) una secuencia de nucleótidos que comprende un fragmento funcional de la secuencia mostrada en el SEQ ID NO: 1 o 4, en donde dicho fragmento inicia la transcripción en una célula vegetal;
  - 10 d) una secuencia de nucleótidos que comprende una secuencia que tiene una identidad de secuencia de al menos 95% con la secuencia mostrada en el SEQ ID NO: 1 o 4, en donde dicha secuencia de nucleótidos inicia la transcripción en una célula vegetal; y
  - e) una secuencia de nucleótidos que hibrida en condiciones restrictivas con un complemento de una secuencia de nucleótidos de a), en donde dicha secuencia de nucleótidos inicia la transcripción en una célula vegetal.
  - 15
2. Un constructo de expresión que comprende la molécula de ácido nucleico de la reivindicación 1.
3. El constructo de expresión de la reivindicación 2, que comprende adicionalmente una secuencia de nucleótidos heteróloga de interés conectada operablemente.
4. Un método para expresar una secuencia de nucleótidos en una planta o célula o nódulo vegetal, comprendiendo dicho método introducir en la planta o célula o nódulo vegetal un constructo de expresión, comprendiendo dicho constructo de expresión un elemento de control de la expresión conectado operablemente a una secuencia de nucleótidos heteróloga de interés, en donde dicho elemento de control de la expresión comprende una secuencia de nucleótidos seleccionada del grupo que consiste en:
  - a) una secuencia de nucleótidos que comprende la secuencia mostrada en el SEQ ID NO: 1 o 4;
  - 25 b) una secuencia de nucleótidos que comprende al menos 350 nucleótidos contiguos de la secuencia mostrada en el SEQ ID NO: 1 o 4, en donde dicha secuencia de nucleótidos inicia la transcripción en una célula vegetal, y en donde dicha secuencia de nucleótidos comprende una caja TATA;
  - c) una secuencia de nucleótidos que comprende un fragmento funcional de la secuencia mostrada en el SEQ ID NO: 1 o 4, en donde dicho fragmento inicia la transcripción en una célula vegetal;
  - 30 d) una secuencia de nucleótidos que comprende una secuencia que tiene una identidad de secuencia de al menos 95% con la secuencia mostrada en el SEQ ID NO: 1 o 4, en donde dicha secuencia de nucleótidos inicia la transcripción en una célula vegetal; y
  - e) una secuencia de nucleótidos que hibrida en condiciones restrictivas con un complemento de una secuencia de nucleótidos de a), en donde dicha secuencia de nucleótidos inicia la transcripción en una célula vegetal.
  - 35
5. El constructo de expresión de la reivindicación 3 o el método de la reivindicación 4, en donde dicho constructo de expresión comprende adicionalmente una secuencia codificante de un péptido señal conectada operablemente que dirige la secreción de un polipéptido codificado por dicha secuencia de nucleótidos heteróloga de interés al medio de cultivo.
6. El constructo de expresión de la reivindicación 5, o el método de la reivindicación 5, en donde dicho péptido señal comprende una secuencia de aminoácidos seleccionada del grupo que consiste en:
  - a) la secuencia mostrada en el SEQ ID NO: 16;
  - b) una secuencia que tiene una identidad de secuencia de al menos 90% con la secuencia mostrada en el SEQ ID NO: 16, en donde dicha secuencia dirige la secreción de dicho polipéptido al medio de cultivo; y
  - 45 c) un fragmento funcional de la secuencia mostrada en el SEQ ID NO: 16, en donde dicha secuencia dirige la secreción de dicho polipéptido al medio de cultivo.
7. El constructo de expresión de la reivindicación 6, o el método de la reivindicación 6, en donde dicha secuencia codificante de dicho péptido señal comprende una secuencia de nucleótidos seleccionada del grupo que consiste en:

- a) la secuencia mostrada en el SEQ ID NO: 15;
- b) una secuencia que tiene una identidad de secuencia de al menos 95% con la secuencia mostrada en el SEQ ID NO: 15; y
- c) un fragmento de la secuencia mostrada en el SEQ ID NO: 15.

- 5 8. El constructo de expresión de una cualquiera de las reivindicaciones 3 ó 5 a 7, o el método de una cualquiera de las reivindicaciones 4 a 7, en donde dicha secuencia de nucleótidos heteróloga de interés codifica un polipéptido de mamífero.
- 10 9. El constructo de expresión de la reivindicación 8, o el método de la reivindicación 8, en donde dicho polipéptido de mamífero se selecciona del grupo que consiste en insulina, hormona del crecimiento, interferón interferón  $\beta$ ,  $\beta$  - glucocerebrosidasa,  $\beta$ -glucuronidasa, proteína de retinoblastoma, proteína p53, angiostatina, leptina, eritropoyetina, factor estimulador de colonias de granulocitos macrófagos, plasminógeno, anticuerpos monoclonales, fragmentos Fab, anticuerpos de cadena sencilla, citoquinas, receptores, vacunas humanas, vacunas animales, péptidos, albúmina de suero, y combinaciones de los mismos.
- 15 10. Una planta o célula o nódulo vegetal transformados que comprenden el constructo de expresión de una cualquiera de las reivindicaciones 2 a 3 ó 5 a 9.
11. Una planta o célula o nódulo vegetal transformados de la reivindicación 10, o el método de una cualquiera de las reivindicaciones 4 a 9, en donde dicha planta o célula o nódulo vegetal es un monocotiledónea.
- 20 12. Una planta o célula o nódulo vegetal transformados de la reivindicación 11, o el método de la reivindicación 11, en donde dicha monocotiledónea es de un género seleccionado del grupo que consiste en el género Spirodela, el género Wolffia, el género Wolffiella, el género Landoltia, y el género Lemna.
13. La planta o célula o nódulo vegetal transformados de la reivindicación 12, o el método de la reivindicación 12, en donde dicha monocotiledónea es un miembro de una especie seleccionada del grupo que consiste en Lemna minor, Lemna miniscula, Lemna aequinoctialis, y Lemna gibba.
- 25 14. La planta o célula o nódulo vegetal transformados de la reivindicación 10, o el método de una cualquiera de las reivindicaciones 4 a 9, en donde dicha planta o célula o nódulo vegetal es una dicotiledónea.
15. El método de una cualquiera de las reivindicaciones 4 a 9, en donde dicha secuencia de nucleótidos heteróloga de interés comprende los codones preferidos por la lenteja de agua en la secuencia codificante de dicha secuencia de nucleótidos heteróloga de interés.
- 30 16. El método de la reivindicación 15, en donde dicha secuencia de nucleótidos heteróloga de interés comprende entre 70-100% de los codones preferidos por la lenteja de agua.