

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 974 294**

51 Int. Cl.:

C07K 14/415 (2006.01)

A01H 1/08 (2006.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.10.2010** **E 19177524 (6)**

97 Fecha y número de publicación de la concesión europea: **14.02.2024** **EP 3560951**

54 Título: **Generación de plantas haploides y fitomejoramiento mejorado**

30 Prioridad:

06.10.2009 US 24899609 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
26.06.2024

73 Titular/es:

**THE REGENTS OF THE UNIVERSITY OF
CALIFORNIA (100.0%)
1111 Franklin Street, 12th Floor
Oakland, CA 94607, US**

72 Inventor/es:

**CHAN, SIMON y
MARUTHACHALAM, RAVI**

74 Agente/Representante:

PONS ARIÑO, Ángel

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 974 294 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

Generación de plantas haploides y fitomejoramiento mejorado

5 Antecedentes de la invención

Aunque los programas de fitomejoramiento en todo el mundo han hecho un progreso considerable en el desarrollo de nuevos cultivares con resistencias a enfermedades mejoradas, rendimientos mejorados y otros rasgos útiles, el mejoramiento en su conjunto depende de la selección de numerosas plantas para identificar nuevas características deseables. Con frecuencia, deben cultivarse y evaluarse cantidades muy grandes de progenie de cruces durante varios años con el fin de seleccionar una o unas pocas plantas con una combinación deseada de rasgos.

El mejoramiento convencional de plantas diploides con frecuencia requiere la selección y el retrocruce de una gran cantidad de plantas para conseguir el genotipo deseado. Una solución al problema de seleccionar grandes cantidades de progenie ha sido producir plantas haploides, cuyos cromosomas pueden duplicarse usando colchicina u otros medios para conseguir instantáneamente plantas homocigóticas doblemente haploides.

Por tanto, pueden conseguirse mejoras marcadas en la economía del mejoramiento a través de la producción de haploides duplicados, puesto que la selección y otras eficiencias de procedimiento pueden mejorarse notablemente mediante el uso de progenies de mejoramiento verdadero (homocigóticas). Con sistemas de producción de haploides duplicados, la homocigosidad se consigue en una generación. Por tanto, el mejorador puede eliminar los numerosos ciclos de endogamia necesarios mediante métodos convencionales para conseguir niveles prácticos de homocigosidad. De hecho, la homocigosidad verdadera para todos los rasgos ni siquiera puede conseguirse mediante métodos de mejoramiento convencionales.

Breve resumen de la invención

La presente invención proporciona nuevas formas de producción de organismos haploides.

En particular, la invención proporciona una planta que comprende una o dos copias de un alelo de un gen CENH3 endógeno mutado, en donde el gen CENH3 endógeno mutado está mutado para reducir la función de CENH3.

Sumario de enseñanzas técnicas

En algunos aspectos, la divulgación proporciona una planta transgénica que comprende un casete heterólogo de expresión transgénica, comprendiendo el casete de expresión un promotor unido operativamente a un polinucleótido que codifica un polipéptido CENH3, CENPC, MIS12, NDC80 o NUF2 recombinantemente alterado, en donde en el evento el polipéptido recombinantemente alterado se expresa en una primera planta que tiene un gen CENH3, CENPC, MIS12, NDC80 o NUF2 y la primera planta se cruza con una planta de tipo silvestre, al menos el 0,1 % de la progenie resultante es haploide.

En algunos aspectos de la divulgación, uno o dos alelos de la secuencia codificante genómica de CENH3, CENPC, MIS12, NDC80 o NUF2 endógena de la planta están inactivados o desactivados. En algunos aspectos de la divulgación, todos los alelos de la secuencia codificante genómica de CENH3, CENPC, MIS12, NDC80 o NUF2 endógena de la planta están inactivados o desactivados. En algunos aspectos de la divulgación, la planta, cuando se cruza con una planta de tipo silvestre, genera al menos un 0,1 % (o, por ejemplo, un 0,5, 1, 2, 5, 10, 20 % o más) de progenie haploide.

En algunos aspectos de la divulgación, el polipéptido es un polipéptido CENH3 recombinantemente alterado. En algunos aspectos de la divulgación, el polipéptido comprende una secuencia de aminoácidos heteróloga de al menos 5 aminoácidos ligada a una proteína que comprende un dominio de plegamiento de histona CENH3, en donde la secuencia de aminoácidos es heteróloga al dominio de plegamiento de histona CENH3. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga se une directamente al dominio de plegamiento de histona CENH3 y el polipéptido carece de un dominio de cola de CENH3. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga se une al dominio de plegamiento de histona CENH3 a través de una secuencia de proteína interpuesta. En algunos aspectos de la divulgación, la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 no CENH3. En algunos aspectos de la divulgación, el dominio de cola de histona H3 no CENH3 comprende una secuencia de aminoácidos al menos un 70 % idéntica a la SEQ ID NO: 95 o un fragmento de la misma de al menos 20 aminoácidos de longitud.

En algunos aspectos de la divulgación, la secuencia de proteína interpuesta comprende un dominio de cola de CENH3. En algunos aspectos de la divulgación, la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 y un dominio de cola de histona CENH3 heterólogo. En algunos aspectos de la divulgación, el dominio de cola de CENH3 es heterólogo con respecto al dominio de plegamiento de histona CENH3.

En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga tiene al menos 10 aminoácidos de

longitud. En algunos aspectos de la divulgación, la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 y un dominio de cola de histona CENH3 heterólogo. la secuencia de aminoácidos heteróloga comprende proteína verde fluorescente. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga altera los centrómeros. En algunos aspectos de la divulgación, el dominio de plegamiento de histona CENH3 se selecciona entre el grupo que consiste en las SEQ ID NO: 49-94.

En algunos aspectos de la divulgación, el polipéptido comprende un dominio de cola no de CENH3 unido a un dominio de plegamiento de histona CENH3.

En algunos aspectos de la divulgación, el polipéptido comprende un dominio de plegamiento de histona CENH3 y un dominio de cola de CENH3 truncado, en donde el extremo amino del dominio de cola se trunca con respecto al dominio de cola endógeno de la planta. En algunos aspectos de la divulgación, el dominio de cola de CENH3 truncado carece de tres o más aminoácidos amino terminales del dominio de cola endógeno. En algunos aspectos de la divulgación, una secuencia de aminoácidos heteróloga se une al extremo amino del dominio de cola truncado. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga tiene al menos 10 aminoácidos de longitud. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga comprende proteína fluorescente verde. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga altera centrómeros. En algunos aspectos de la divulgación, el dominio de plegamiento de histona CENH3 se selecciona del grupo que consiste en SEQ ID NO: 49-94.

En algunos aspectos de la divulgación, el polipéptido es un polipéptido CENPC, MIS12, NDC80 y NUF2 recombinantemente alterado.

La presente divulgación también proporciona un ácido nucleico aislado que comprende un polinucleótido que codifica un polipéptido, en donde el polipéptido comprende:

un dominio de cola de histona H3 no CENH3 unido a un dominio de plegamiento de histona CENH3; o
un dominio de cola de CENH3 truncado unido a un dominio de plegamiento de histona CENH3, en donde el extremo amino del dominio de cola está truncado.

La presente divulgación también proporciona una planta que comprende un CENH3 silenciado o una o dos copias de un alelo de un gen desactivado, inactivado o mutado de CENH3 endógena.

En algunos aspectos de la divulgación, la planta que expresa una proteína CENH3 endógena es el progenitor de polen del cruce.

En algunos aspectos de la divulgación, la planta que expresa una proteína CENH3 endógena es el progenitor de óvulo del cruce.

En algunos aspectos de la divulgación, el método comprende adicionalmente convertir al menos una planta haploide seleccionada en una planta doblemente haploide.

Se desvela un método para producir una planta transgénica que comprende un casete de expresión transgénico heterólogo, comprendiendo el casete de expresión un promotor unido operativamente a una secuencia de polinucleótido que codifica un polipéptido CENH3, CENPC, MIS12, NDC80 o NUF2 recombinantemente alterado, en donde en el evento el polipéptido recombinantemente alterado se expresa en una primera planta que tiene un gen CENH3, CENPC, MIS12, NDC80 o NUF2 endógeno inactivado correspondiente y la primera planta se cruza con una planta de tipo silvestre, al menos el 0,1 % de la progenie resultante es haploide, comprendiendo el método,

transformar células vegetales con un ácido nucleico que comprende el casete de expresión; y
seleccionar transformantes que comprenden el ácido nucleico, produciendo de este modo la planta.

En algunos aspectos, la presente divulgación proporciona un polinucleótido aislado que codifica un polipéptido, en donde el polipéptido comprende:

una secuencia de aminoácidos de al menos 5 aminoácidos unidos a una proteína que comprende un dominio de plegamiento de histona CENH3, en donde la secuencia de aminoácidos es heteróloga con respecto al dominio de plegamiento de histona CENH3; o
una proteína que comprende un dominio de plegamiento de histona CENH3 y un dominio de cola CENH3 truncado, en donde el extremo amino del dominio de cola está truncado.

En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga se une directamente al dominio de plegamiento de histona CENH3. En algunos aspectos de la divulgación, el polipéptido carece de un dominio de cola CENH3.

En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga se une al dominio de plegamiento de

histona CENH3 a través de una secuencia de proteína interpuesta. En algunos aspectos de la divulgación, la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 no CENH3. En algunos aspectos de la divulgación, la secuencia de proteína interpuesta comprende un dominio de cola de CENH3. En algunos aspectos de la divulgación, el dominio de cola de CENH3 es heterólogo con respecto al dominio de plegamiento de histona CENH3.

5 En algunos aspectos de la divulgación, el dominio de cola de histona H3 no CENH3 comprende una secuencia de aminoácidos al menos un 70 % idéntica a la SEQ ID NO: 95 o un fragmento de la misma de al menos 20 aminoácidos de longitud. En algunos aspectos de la divulgación, la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 y un dominio de cola de histona CENH3 heterólogo.

10 En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga tiene al menos 3, 5, 10, 15, 20, 30 o 50 aminoácidos de longitud, careciendo opcionalmente de una estructura secundaria fija.

En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga comprende proteína verde fluorescente.

15 En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga altera los centrómeros.

En algunos aspectos de la divulgación, el dominio de plegamiento de histona CENH3 se selecciona entre el grupo que consiste en las SEQ ID NO: 49-94.

20 En algunos aspectos de la divulgación, el polipéptido comprende una proteína que comprende un dominio de plegamiento de histona CENH3 y un dominio de cola de CENH3 truncado, en donde el extremo amino del dominio de cola está truncado.

25 En algunos aspectos de la divulgación, el dominio de cola de CENH3 truncado carece de al menos 1, 2, 3, 4, 5, 6, 10, 15 o 20 aminoácidos amino terminales del dominio de cola endógeno. En algunos aspectos de la divulgación, una secuencia de aminoácidos heteróloga se une al extremo amino del dominio de cola truncado. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga tiene al menos 3, 5, 10, 15, 20, 30 o 50 aminoácidos de longitud. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga comprende proteína verde

30 fluorescente. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga altera los centrómeros.

En algunos aspectos de la divulgación, el dominio de plegamiento de histona CENH3 se selecciona entre el grupo que consiste en las SEQ ID NO: 49-94.

35 La presente divulgación también proporciona un casete de expresión que comprende cualquiera de los polinucleótidos enumerados anteriormente, en donde el casete de expresión comprende un promotor unido operativamente al polinucleótido que codifica un polipéptido. En algunos aspectos, la divulgación proporciona un vector que comprende el casete de expresión.

40 En algunos aspectos de la divulgación, la divulgación proporciona una planta que comprende el casete de expresión.

En algunos aspectos de la divulgación, el dominio de cola de histona heterólogo comprende un dominio de cola de histona H3 o un dominio de cola de histona CENH3 heterólogo.

45 En algunos aspectos de la divulgación, el polipéptido comprende un dominio de cola de histona H3 y un dominio de cola de histona CENH3.

En algunos aspectos de la divulgación, la planta comprende una CENH3 silenciada o una o dos copias de un alelo de un gen CENH3 endógeno desactivado o mutado.

50 En algunos aspectos de la divulgación, el casete de expresión se integra en el cromosoma de la planta.

La presente divulgación también proporciona una planta que comprende una CENH3 silenciada o una o dos copias de un alelo de un gen CENH3 endógeno desactivado o mutado.

55 La presente divulgación también proporciona un método para generar una planta haploide.

En algunos aspectos de la divulgación el método comprende adicionalmente convertir al menos una planta haploide seleccionada en una planta doblemente haploide.

60 Otros aspectos de la divulgación quedarán claros a partir del resto del texto en el presente documento.

Definiciones

65 Una secuencia de gen o proteína "endógena" se refiere a una secuencia no recombinante de un organismo ya que la secuencia aparece en el organismo antes de la mutación de la secuencia inducida por el ser humano. Una secuencia

"mutada" se refiere a una secuencia alterada por seres humanos. Los ejemplos de mutaciones inducidas por seres humanos incluyen la exposición de un organismo a una dosis alta de un mutágeno químico, radiológico o de inserción con el fin de seleccionar mutantes, así como la alteración recombinante de una secuencia. Los ejemplos de alteraciones recombinantes inducidas por seres humanos pueden incluir, por ejemplo, fusiones, inserciones, supresiones y/o cambios en la secuencia.

El término "promotor" se refiere a regiones o secuencias ubicadas corriente arriba y/o corriente abajo desde el inicio de la transcripción y que están implicadas en el reconocimiento y la unión de la ARN polimerasa y otras proteínas para iniciar la transcripción. Un "promotor vegetal" es un promotor capaz de iniciar la transcripción en células vegetales. Un promotor vegetal puede ser, pero no necesariamente tiene que ser, una secuencia de ácido nucleico aislada originariamente de una planta.

La expresión "unido operativamente" se refiere a una unión funcional entre una secuencia de control de la expresión de ácido nucleico (tal como un promotor, o una serie de sitios de unión de factor de transcripción) y una segunda secuencia de ácido nucleico, en donde la secuencia de control de la expresión dirige la transcripción del ácido nucleico correspondiente a la segunda secuencia.

El término "planta" incluye plantas enteras, brotes de órganos/estructuras vegetativas (por ejemplo, hojas, tallos y tubérculos), raíces, flores y órganos/estructuras florales (por ejemplo, brácteas, sépalos, pétalos, estambres, carpelos, anteras y óvulos), semilla (incluyendo embrión, endospermo y cubierta de semilla) y fruto (el ovario maduro), tejido vegetal (por ejemplo, tejido vascular y tejido terrestre) y células (por ejemplo, células protectoras, óvulos y tricomas) y progenie de la misma. La clase de plantas que puede usarse es generalmente tan amplia como la clase de plantas superiores e inferiores susceptibles de técnicas de transformación, incluyendo las angiospermas (plantas monocotiledóneas y dicotiledóneas), gimnospermas, helechos y algas multicelulares. Incluye plantas de una diversidad de niveles de ploidía, incluyendo aneuploides, poliploides, diploides, haploides y hemicigóticas.

Una secuencia polinucleotídica o polipeptídica es "heteróloga con respecto a" un organismo o una segunda secuencia si se origina a partir de una especie extraña o, si es de la misma especie, se modifica con respecto a su forma original. Por ejemplo, un promotor unido operativamente a una secuencia codificante heteróloga se refiere a una secuencia codificante de una especie diferente de la que derivó el promotor o, si es de la misma especie, una secuencia codificante que no se asocia de forma natural al promotor (por ejemplo, una secuencia codificante modificada mediante ingeniería genética o un alelo de un ecotipo o variedad diferente). En otro ejemplo, un dominio de cola de CENH3 de una primera especie es heterólogo con respecto a un dominio de plegamiento de histona CENH3 de una segunda especie.

"Recombinante" se refiere a un polinucleótido manipulado por el ser humano o una copia o complemento de un polinucleótido manipulado por el ser humano. Por ejemplo, un casete de expresión recombinante que comprende un promotor unido operativamente a un segundo polinucleótido puede incluir un promotor que sea heterólogo con respecto al segundo polinucleótido como resultado de la manipulación humana (por ejemplo, mediante métodos descritos en Sambrook et al., *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, (1989) o *Current Protocols in Molecular Biology*, Volúmenes 1-3, John Wiley & Sons, Inc. (1994-1998)). En otro ejemplo, un casete de expresión recombinante puede comprender polinucleótidos combinados de manera que sea extremadamente improbable que los polinucleótidos se encuentren en la naturaleza. Por ejemplo, los sitios de restricción o secuencias de vector plasmídico manipulados por el ser humano pueden flanquear o separar el promotor del segundo polinucleótido. Un experto reconocerá que los polinucleótidos pueden manipularse de muchas maneras y no se limitan a los ejemplos anteriores.

Se usa un "transgén" como se entiende el término en la técnica y se refiere a un ácido nucleico heterólogo introducido en una célula mediante manipulación molecular humana del genoma de la célula (por ejemplo, mediante transformación molecular). Por tanto, una "planta transgénica" es una planta que comprende un transgén, es decir, es una planta modificada genéticamente. La planta transgénica puede ser la planta inicial en la que se introdujo el transgén, así como la progenie de la misma cuyo genoma contiene el transgén.

El término "correspondiente" como se usa en el presente documento significa "respectivo". Por ejemplo, cuando se dice que una planta contiene una copia alterada de forma recombinante de una proteína seleccionada entre A, B y C, y que la planta también contiene una copia endógena mutada "correspondiente" del gen seleccionado entre un gen que codifica A, B o C, si la planta contiene una proteína A alterada de forma recombinante, la copia endógena mutada correspondiente también sería A. Como alternativa, si la planta contiene una proteína B alterada de forma recombinante, la copia endógena mutada correspondiente también sería B, etc.

La frase "ácido nucleico" o "secuencia polinucleotídica" se refiere a un polímero mono o bicatenario de bases desoxirribonucleotídicas o ribonucleotídicas leído del extremo 5' al extremo 3'. Los ácidos nucleicos también pueden incluir nucleótidos modificados que permitan la lectura correcta a través de una polimerasa, y/o la formación de dúplex bicatenarios, y no alteren significativamente la expresión de un polipéptido codificado por ese ácido nucleico.

La frase "ácido nucleico codificante de secuencia" se refiere a un ácido nucleico que dirige la expresión de una proteína

o un péptido específicos. Las secuencias de ácido nucleico incluyen tanto la secuencia de cadena de ADN que se transcribe en ARN como la secuencia de ARN que se traduce en proteína. Las secuencias de ácido nucleico incluyen tanto las secuencias de ácido nucleico de longitud completa como las secuencias de longitud no completa derivadas de las secuencias de longitud completa. Ha de comprenderse además que la secuencia incluye los codones degenerados de la secuencia o secuencias nativas que pueden introducirse para proporcionar preferencia de codón en una célula hospedadora específica.

La frase "célula hospedadora" se refiere a una célula de cualquier organismo. Las células hospedadoras de ejemplo derivan de plantas, bacterias, levaduras, hongos, insectos u otros animales. Se conocen bien en la técnica métodos para introducir secuencias polinucleotídicas en diversos tipos de células hospedadoras.

Un "casete de expresión" se refiere a una construcción de ácido nucleico, que cuando se introduce en una célula hospedadora (por ejemplo, una célula vegetal), da como resultado la transcripción y/o traducción de un ARN o polipéptido, respectivamente. Un casete de expresión puede dar como resultado la transcripción sin traducción, por ejemplo, cuando se transcribe un ARNip u otro ARN que no codifica proteínas.

Se dice que dos secuencias de ácido nucleico o polipéptidos son "idénticos" si la secuencia de nucleótidos o restos de aminoácidos, respectivamente, en las dos secuencias es la misma cuando se alinean para una correspondencia máxima como se describe a continuación. La expresión "complementario a" se usa en el presente documento para indicar que la secuencia es complementaria a toda o una porción de una secuencia polinucleotídica de referencia.

Son ejemplos de algoritmos que son adecuados para determinar el porcentaje de identidad de secuencia y la similitud de secuencia los algoritmos BLAST y BLAST 2.0, que se describen en Altschul et al., *Nuc. Acids Res.* 25: 3389-3402 (1977) y Altschul et al., *J. Mol. Biol.* 215: 403-410 (1990), respectivamente. El software para realizar análisis BLAST está disponible públicamente en la Web a través del Centro Nacional de Información Biotecnológica de los EE.UU. (www.ncbi.nlm.nih.gov/). Este algoritmo implica identificar en primer lugar pares de secuencias de alta puntuación (HSP, por sus siglas en inglés) identificando palabras cortas de longitud W en la secuencia de consulta, que coincide con o satisface alguna puntuación umbral de valor positivo T cuando se alinea con una palabra de la misma longitud en una secuencia de la base de datos. T se denomina el umbral de puntuación de palabras adyacentes (Altschul *et al.*, citado anteriormente). Estos aciertos de palabras adyacentes iniciales actúan como semillas para iniciar búsquedas para encontrar HSP más largos que los contengan. Los aciertos de palabras se prolongan en ambas direcciones a lo largo de cada secuencia tanto cuanto pueda aumentarse la puntuación de alineación acumulativa. Las puntuaciones acumulativas se calculan usando, para secuencias nucleotídicas, los parámetros M (puntuación de recompensa para un par de restos emparejados; siempre > 0) y N (valor de penalización para restos no emparejados; siempre < 0). Para las secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulativa. La prolongación de los aciertos de palabras en cada dirección se detiene cuando: la puntuación de alineación acumulativa queda fuera por una cantidad X de su valor máximo conseguido; la puntuación acumulativa llega a cero o por debajo, debido a la acumulación de una o más alineaciones de restos de puntuación negativa; o se alcanza el final de cualquier secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad de la alineación. El programa BLASTN (para secuencias nucleotídicas) usa por defecto una longitud de palabra (W) de 11, una expectativa (E) o 10, M = 5, N = -4 y una comparación de ambas cadenas. Para las secuencias de aminoácidos, el programa BLASTP usa como valores predeterminados una longitud de palabra de 3 y una expectativa (E) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff, *Proc. Natl. Acad. Sci. USA* 89: 10915, (1989)) alineaciones (B) de 50, expectativa (E) de 10, M = 5, N = -4 y una comparación de ambas cadenas.

El algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, *Proc. Natl. Acad. Sci. USA* 90: 5873-5787, (1993)). Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), que proporciona una indicación de la probabilidad por la que sucedería un emparejamiento entre dos secuencias nucleotídicas o de aminoácidos por casualidad. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es inferior a aproximadamente 0,2, más preferentemente inferior a aproximadamente 0,01 y mucho más preferentemente inferior a aproximadamente 0,001.

El "porcentaje de identidad de secuencia" se determina comparando dos secuencias alineadas de manera óptima a lo largo de una ventana de comparación, en donde la porción de la secuencia polinucleotídica en la ventana de comparación puede comprender adiciones o supresiones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones o supresiones) para la alineación óptima de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que se encuentra la base de ácido nucleico o el resto de aminoácido idénticos en ambas secuencias para producir el número de posiciones emparejadas, dividiendo el número de posiciones emparejadas entre el número total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para producir el porcentaje de identidad de secuencia.

La expresión "identidad sustancial" de secuencias polinucleotídicas significa que un polinucleótido comprende una secuencia que tiene al menos un 25 % de identidad de secuencia con una secuencia de referencia designada. Como alternativa, el porcentaje de identidad puede ser cualquier número entero del 25 % al 100 %, por ejemplo, al menos:

25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 99 % en comparación con una secuencia de referencia usando los programas que se describen en el presente documento; preferentemente BLAST usando parámetros de referencia, como se describe a continuación. Un experto reconocerá que los valores de porcentaje de identidad anteriores pueden ajustarse adecuadamente para determinar la identidad correspondiente de proteínas codificadas por dos secuencias nucleotídicas teniendo en cuenta la degeneración de codones, la similitud de aminoácidos y el posicionamiento del marco de lectura. La identidad sustancial de las secuencias de aminoácidos para estos fines normalmente significa una identidad de secuencia de al menos el 40 %. El porcentaje de identidad de los polipéptidos puede ser cualquier número entero del 40 % al 100 %, por ejemplo, al menos el 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 99 %. En algunos aspectos de la divulgación, polipéptidos que son "sustancialmente similares" comparten secuencias como se ha indicado anteriormente, excepto por que las posiciones de restos que no son idénticas pueden diferir por cambios conservadores de aminoácidos. Las sustituciones conservadoras de aminoácidos se refieren a la intercambiabilidad de restos que tienen cadenas laterales similares. Por ejemplo, un grupo de aminoácidos que tiene cadenas laterales alifáticas es glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos que tiene cadenas laterales de hidroxilo alifático es serina y treonina; un grupo de aminoácidos que tiene cadenas laterales que contienen amida es asparagina y glutamina; un grupo de aminoácidos que tiene cadenas laterales aromáticas es fenilalanina, tirosina y triptófano; un grupo de aminoácidos que tiene cadenas laterales básicas es lisina, arginina e histidina; un grupo de aminoácidos que tiene cadenas laterales que contienen sulfuro es cisteína y metionina. Son grupos conservadores de ejemplo de sustitución de aminoácidos: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina, ácido aspártico-ácido glutámico y asparagina-glutamina.

Breve descripción de los dibujos

La Figura 1 ilustra una alineación de secuencia de diversas proteínas CENH3 (H3.3 de *A. thaliana* = SEQ ID NO: 1; H3.3 humana = SEQ ID NO: 2; *C. albicans* = SEQ ID NO: 106; Humana = SEQ ID NO: 107; *A. thaliana* = SEQ ID NO: 10; Chopo = SEQ ID NO: 11; Arroz = SEQ ID NO: 108).

Descripción detallada

I. Introducción

La presente invención se basa, en parte, en el sorprendente descubrimiento de que la eliminación de una CENH3 endógena en combinación con la expresión de una proteína heteróloga que comprende una CENH3 alterada da como resultado una planta que tiene propiedades útiles para el mejoramiento. Por ejemplo, cuando una planta que carece de una proteína CENH3 endógena y expresa una proteína que comprende (enumerados del extremo amino al extremo carboxilo) un marcador GFP, un dominio de cola no de CENH3 y un dominio de plegamiento de histona CENH3, se cruza con una planta que tiene una proteína CENH3 endógena, una porción de la progenie resultante carece de todos los cromosomas derivados de la planta parental que expresa una versión alterada de CENH3. Por tanto, la invención permite la producción de progenie haploide. Las plantas haploides son útiles, por ejemplo, para mejorar y acelerar el mejoramiento.

CENH3 es un miembro del complejo del cinetocoro, la estructura proteínica en los cromosomas donde las fibras del huso se fijan durante la división celular. Sin pretender limitar el alcance de la invención, se cree que los resultados observados se deben en parte a la generación de una proteína del cinetocoro que actúa más débilmente que el tipo silvestre, dando como resultado complejos del cinetocoro funcionales (por ejemplo, en la mitosis), pero que dan como resultado cromosomas que se segregan de forma relativamente escasa durante la meiosis con respecto a los cromosomas que también contienen complejos del cinetocoro de tipo silvestre del otro progenitor. Esto da como resultado complejos del cinetocoro funcionales cuando la proteína alterada es la única isoforma en la célula, pero cromosomas que se segregan de forma relativamente escasa durante la mitosis cuando el progenitor con cinetocoros alterados se cruza con un progenitor con complejos del cinetocoro de tipo silvestre. Además de CENH3, otras proteínas del cinetocoro incluyen, por ejemplo, CENPC, MCM21, MIS12, NDC80 y NUF2. En consecuencia, la presente divulgación proporciona plantas, hongos o animales (o células de los mismos) que expresan una proteína del cinetocoro mutada recombinante (CENH3, CENPC, MCM21, MIS12, NDC80 y NUF2) que altera el centrómero y/o plantas, hongos o animales (o células de los mismos) en los que al menos una o ambas copias de un alelo del gen endógeno de CENH3 se ha inactivado, mutado para reducir o eliminar su función, o silenciado. Como se explica con más detalle a continuación, la proteína del cinetocoro mutada puede mutarse de muchas maneras diferentes, incluyendo, pero sin limitación, como una proteína "tailswap" (intercambio de cola, en inglés), que comprende un dominio de plegamiento de histona CENH3 y una secuencia amino terminal heteróloga. La presente divulgación también proporciona métodos para generar una planta haploide.

II. PROTEÍNAS DEL CINETOCORO

A. PROTEÍNAS CENH3

Las proteínas CENH3 son una clase de proteínas bien caracterizadas que son variantes de las proteínas histonas H3 y que son proteínas especializadas asociadas al centrómero. Las proteínas CENH3 se caracterizan por un dominio de

cola variable, que no forma una estructura secundaria rígida, y un dominio de plegamiento de histona conservado constituido por tres regiones α -helicoidales conectadas por secciones de bucle. Pueden encontrarse características estructurales y funcionales adicionales de las proteínas CENH3 en, por ejemplo, Cooper et al., *Mol Biol Evol.* 21(9): 1712-8 (2004); Malik et al., *Nat Struct Biol.* 10(11): 882-91 (2003); Black et al., *Curr Opin Cell Biol.* 20(1): 91-100 (2008).

5 Las proteínas CENH3 son una de las proteínas que forman el complejo del cinetocoro.

Se ha identificado una amplia diversidad de proteínas CENH3. Véanse, por ejemplo, las SEQ ID NO: 1-48. Se apreciará que la lista anterior no pretende ser exhaustiva y que las secuencias hay disponibles secuencias de CENH3 adicionales a partir de estudios genómicos o pueden identificarse a partir de bases de datos genómicas o mediante técnicas de laboratorio bien conocidas. Por ejemplo, cuando la CENH3 de una planta particular u otra especie de organismo no está fácilmente disponible en una base de datos, puede identificarse y clonarse la secuencia génica de CENH3 del organismo usando cebadores, que opcionalmente se degeneran, basándose en regiones conservadas de otras proteínas CENH3 conocidas.

15 La puesta en práctica de la presente invención generalmente empleará métodos convencionales de química, bioquímica, biología molecular, biología celular, genética, inmunología y farmacología, dentro de la habilidad en la técnica. Dichas técnicas se explican por completo en la bibliografía. Véanse, por ejemplo, Gennaro, A. R., ed. (1990) *Remington's Pharmaceutical Sciences*, 18ª ed., Mack Publishing Co.; Hardman, J. G., Limbird, L. E. y Gilman, A. G., eds. (2001) *The Pharmacological Basis of Therapeutics*, 10ª ed., McGraw-Hill Co.; Colowick, S. et al., eds., *Methods In Enzymology*, Academic Press, Inc.; Weir, D. M. y Blackwell, C. C., eds. (1986) *Handbook of Experimental Immunology*, Vol. I-IV, Blackwell Scientific Publications; Maniatis, T. et al., eds. (1989) *Molecular Cloning: A Laboratory Manual*, 2ª edición, Vol. I-III, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al., eds. (1999) *Short Protocols in Molecular Biology*, 4ª edición, John Wiley & Sons; Ream et al., eds. (1998) *Molecular Biology Techniques: An Intensive Laboratory Course*, Academic Press; Newton, C. R. y Graham, A., eds. (1997) *PCR (Serie Introduction to Biotechniques)*, 2ª ed., Springer Verlag.

i. Dominio de plegamiento de histona CENH3

30 Como se ha indicado anteriormente, el dominio de plegamiento de histona CENH3 se conserva entre proteínas CENH3 de diferentes especies. El dominio de plegamiento de histona CENH3 puede distinguirse por tres regiones α -helicoidales conectadas por secciones de bucle. Aunque se apreciará que la ubicación exacta del dominio de plegamiento de histona variará en las proteínas CENH3 de otras especies, se encontrará en el extremo carboxilo de una proteína CENH3 endógena (de tipo silvestre). Por tanto, en algunos aspectos de la divulgación, una proteína CENH3 puede identificarse en una proteína endógena por tener un dominio carboxilo terminal sustancialmente similar (por ejemplo, de al menos un 30 %, 40 %, 50 %, 60 %, 70 %, 85 %, 90 %, 95 % o más de identidad) a cualquiera de las SEQ ID NO: 49-94. Una alineación de una selección de proteínas CENH3 se proporciona en la Figura 1 e ilustra áreas de conservación en el dominio de plegamiento de histona.

40 El límite entre el dominio de cola y el dominio de plegamiento de histona de las proteínas CENH3 está en, dentro o cerca (es decir, dentro de 5, 10, 15, 20 o 25 aminoácidos de la "P" de) la secuencia PGTVAL conservada (SEQ ID NO: 114). La secuencia PGTVAL (SEQ ID NO: 114) es de aproximadamente 81 aminoácidos del extremo N terminal de la proteína CENH3 de *Arabidopsis*, aunque la distancia desde el extremo N de diferentes proteínas CENH3 endógenas varía. Véase, por ejemplo, el listado de secuencias. Por tanto, en algunos aspectos de la divulgación, la región de plegamiento de histona de CENH3 empleada en las proteínas tailswap incluye todos los aminoácidos C-terminales de una proteína CENH3 endógena (o una proteína sustancialmente similar a la secuencia endógena) hasta e incluyendo la PGTVAL (SEQ ID NO: 114). Las SEQ ID NO: 49-94 reflejan esta opción. En otros aspectos, las proteínas tailswap de la divulgación pueden comprender más o menos de la secuencia de CENH3. Por ejemplo, en algunos aspectos de la divulgación, tailswap comprenderá la secuencia C-terminal de una proteína CENH3, pero solo hasta un aminoácido 5, 10, 15, 20 o 25 aminoácidos en la dirección C-terminal de la "P" de la secuencia PGTVAL conservada (SEQ ID NO: 114). En algunos aspectos de la divulgación, tailswap comprenderá la secuencia C-terminal de una proteína CENH3, pero solo hasta un aminoácido 5, 10, 15, 20 o 25 aminoácidos en la dirección N-terminal desde la "P" de la secuencia PGTVAL conservada (SEQ ID NO: 114).

ii. Dominio de cola de histona CENH3

55 Aunque el dominio de plegamiento de histona de CENH3 evoluciona más rápidamente que el de H3 convencional, los dominios de plegamiento de histona CENH3 y H3 todavía pueden alinearse. Por el contrario, los dominios de cola N-terminales de CENH3 son altamente variables incluso entre especies estrechamente relacionadas. Los dominios de cola de histona (incluyendo los dominios de cola de CENH3) son flexibles y no estructurados, como lo demuestra su falta de densidad fuerte de electrones en la estructura del nucleosoma determinada por cristalografía de rayos X (Luger et al., *Nature* 389 (6648): 251-60 (1997)).

iii. Proteínas CENH3 mutadas

65 Puede introducirse cualquier cantidad de mutaciones de CENH3 en una proteína CENH3 para generar una proteína CENH3 mutada (incluyendo, pero sin limitación, una alterada de forma recombinante) capaz de generar plantas

haploides cuando se expresa en una planta que carece o que tiene una expresión suprimida de, una proteína CENH3 endógena, y cuando la planta transgénica resultante se cruza con una planta que expresa una proteína CENH3 de tipo silvestre. Las proteínas CENH3 mutantes activas pueden identificarse, por ejemplo, mediante mutagénesis aleatoria, mediante mutagénesis dirigida a aminoácidos simples o múltiples, mediante generación de supresiones completas o parciales de dominio de proteína, mediante fusión con secuencias de aminoácidos heterólogas o mediante combinaciones de los mismos. Las proteínas CENH3 mutantes "activas" se refieren a proteínas, que cuando se expresan en una planta en la que CENH3 se desactiva o inactiva, dan como resultado plantas viables, plantas viables que cuando se cruzan con una planta de tipo silvestre, producen una progenie haploide a una frecuencia más que normal (por ejemplo, de al menos el 0,1, 0,5, 1, 5, 10, 20 % o más). Las proteínas CENH3 mutadas pueden someterse a ensayo fácilmente mediante la expresión recombinante de la proteína CENH3 mutada en una planta que carece de proteína CENH3 endógena, cruzando la planta transgénica (como masculina o femenina, dependiendo de la fertilidad) con una proteína CENH3 de tipo silvestre de expresión vegetal y después seleccionando para determinar la producción de progenie haploide.

En algunos aspectos de la divulgación, la proteína CENH3 mutada es idéntica a una proteína CENH3 endógena pero para 1, 2, 3, 4, 5, 6, 7, 8 o más (por ejemplo, 1-2, 1-4, 1-7) aminoácidos. Por ejemplo, en algunos aspectos de la divulgación, la proteína endógena de tipo silvestre de la planta es idéntica o sustancialmente idéntica a cualquiera de las SEQ ID NO: 1-48 y la proteína CENH3 mutada difiere de la proteína CENH3 endógena en 1, 2, 3, 4, 5, 6, 7, 8 o más (por ejemplo, 1-2, 1-4, 1-7) aminoácidos.

En algunos aspectos de la divulgación, la proteína CENH3 mutada contiene un dominio de plegamiento de histona CENH3 idéntico al dominio de plegamiento de histona CENH3 de una proteína CENH3 endógena pero para 1, 2, 3, 4, 5, 6, 7, 8 o más (por ejemplo, 1-2, 1-4, 1-7) aminoácidos. Por ejemplo, en algunos aspectos de la divulgación, el dominio de plegamiento de histona CENH3 de tipo silvestre endógeno de la planta es idéntico o sustancialmente idéntico a cualquiera de las SEQ ID NO: 49-94 y la proteína CENH3 mutada contiene un dominio de plegamiento de histona CENH3 que difiere del dominio de plegamiento de histona de proteína CENH3 endógena en 1, 2, 3, 4, 5, 6, 7, 8 o más (por ejemplo, 1-2, 1-4, 1-7) aminoácidos.

Se cree que los mutantes activos de CENH3 incluyen, por ejemplo, proteínas que comprenden:

- una secuencia de aminoácidos heteróloga (que incluye pero sin limitación GFP) unida a un dominio de cola truncado o completo de CENH3 o un dominio de cola no de CENH3, cualquiera de los cuales se une a un dominio de plegamiento de histona CENH3; o
- un dominio de cola truncado de CENH3, un dominio de cola de CENH3 heteróloga o un dominio de cola no de CENH3, cualquiera de los cuales se une a un dominio de plegamiento de histona CENH3.

En algunos aspectos de la divulgación, la proteína CENH3 mutada comprende una fusión de una secuencia de aminoácidos heteróloga amino terminal con el dominio de plegamiento de histona de una proteína CENH3. Generalmente, el dominio de plegamiento de histona será idéntico o al menos sustancialmente idéntico a la proteína CENH3 endógena con respecto al organismo en el que se expresará la proteína CENH3 mutada. En algunos aspectos de la divulgación, la proteína CENH3 mutada incluirá un dominio de cola de histona, que puede ser, por ejemplo, un dominio de cola no de CENH3 o un dominio de cola de CENH3.

Se cree que una gran cantidad de secuencias de aminoácidos diferentes, cuando se une a una proteína que comprende un dominio de plegamiento de histona CENH3 y una secuencia que puede funcionar como o reemplazar un dominio de cola de histona, puede usarse de acuerdo con la presente divulgación. En algunos aspectos de la divulgación, la secuencia heteróloga se une directamente al dominio de plegamiento de histona CENH3. En algunos aspectos de la divulgación, la secuencia heteróloga se une como una secuencia de aminoácidos interpuesta al dominio de plegamiento de histona CENH3. En algunos aspectos de la divulgación, la secuencia de aminoácidos interpuesta es un dominio de cola de CENH3 intacto o truncado. En algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga, en combinación con el dominio de plegamiento de histona, será suficiente para prevenir la letalidad asociada a la pérdida de CENH3 endógena, pero alterará suficientemente los centrómeros para permitir la producción de progenie haploide, como se analiza en el presente documento. Por tanto, en algunos aspectos de la divulgación, la secuencia de aminoácidos heteróloga comprenderá una porción que es, o imita la función de, un dominio de cola de histona y, opcionalmente, también puede comprender una secuencia de aminoácidos voluminosa que altera la función del centrómero. En algunos aspectos de la divulgación, al menos una porción de la secuencia de aminoácidos heteróloga de la proteína CENH3 mutada comprende cualquier secuencia de aminoácidos de al menos 10, 20, 30, 40, 50, por ejemplo, 10-30, 10-50, 20-50, 30-60 aminoácidos, que opcionalmente carecen de una estructura secundaria estable (por ejemplo, que carecen de enrollamiento, hélices o láminas beta). En algunos aspectos de la divulgación, el dominio de cola tiene menos del 90, 80 o 70 % de identidad con el dominio de cola (por ejemplo, los 135 aminoácidos N-terminales) de la proteína CENH3 endógena con respecto al organismo en el que se expresará la proteína CENH3 mutada. En algunos aspectos de la divulgación, el dominio de cola de la proteína CENH3 mutada comprende el dominio de cola de una proteína histona no CENH3, incluyendo, pero sin limitación, una proteína histona H3. En algunos aspectos de la divulgación, el dominio de cola de la proteína CENH3 mutada comprende el dominio de cola de una proteína histona no CENH3 endógena con respecto al organismo en el que se expresará la proteína CENH3 mutada. En algunos aspectos de la divulgación, el dominio de cola de la proteína CENH3 mutada comprende

el dominio de cola de una cola de CENH3 homóloga u ortóloga (de una especie vegetal diferente). Por ejemplo, se ha descubierto que la GFP fusionada con un dominio de cola de CENH3 de maíz unido a un dominio de plegamiento de histona CENH3 de *Arabidopsis* es activa.

5 Como se ha indicado anteriormente, en algunos aspectos de la divulgación, el dominio de cola de una histona H3 (que no debe confundirse con una histona CENH3) se usa como la porción del dominio de cola de la proteína CENH3 mutada (estos aspectos de la divulgación en ocasiones se denominan proteínas "tailswap"). Los dominios de cola de H3 vegetales están bien conservados en diversos organismos. Por ejemplo, un dominio de cola H3 común de plantas es la SEQ ID NO: 95. Por tanto, en algunos aspectos de la divulgación, la porción de cola heteróloga de la proteína
10 tailswap comprenderá una secuencia de aminoácidos sustancialmente idéntica (por ejemplo, al menos un 70, 80, 90, 95 o 100 % idéntica) a la SEQ ID NO: 95, o un fragmento de la misma de al menos 15, 20, 25, 30, 35 o 40 aminoácidos de longitud.

15 En algunos aspectos, las proteínas CENH3 mutadas de la divulgación carecerán de al menos una porción (por ejemplo, al menos 5, 10, 15, 20, 25, 30 o más aminoácidos) de la región N-terminal de CENH3 endógena y, por tanto, en algunos aspectos de la divulgación, tendrá un dominio de cola de CENH3 truncado en comparación con una proteína CENH3 endógena de tipo silvestre. Las proteínas CENH3 mutadas pueden, o no, unirse a una secuencia heteróloga.

20 Opcionalmente, la secuencia de aminoácidos heteróloga puede comprender, o comprender adicionalmente, una o más secuencias de aminoácidos en el extremo amino y/o carboxilo y/o unir los dominios de cola y de plegamiento de histona. Por ejemplo, en algunos aspectos de la divulgación, la proteína CENH3 mutada (por ejemplo, una tailswap u otra proteína mutada CENH3) comprende una secuencia de aminoácidos heteróloga unida al extremo amino del dominio de cola. En algunos aspectos de la divulgación, la secuencia heteróloga se une al extremo amino de una proteína CENH3 de otro tipo silvestre de otro modo, en donde la secuencia heteróloga interfiere con la función del
25 centrómero. Por ejemplo, se ha encontrado, por ejemplo, que la proteína verde fluorescente, cuando se une a CENH3 de tipo silvestre, altera suficientemente los centrómeros para permitir la producción de progenie haploide. Se cree que la secuencia heteróloga puede ser cualquier secuencia que altere la capacidad de la proteína CENH3 para mantener la función del centrómero. Por tanto, en algunos aspectos de la divulgación, la secuencia heteróloga comprende una secuencia de aminoácidos de al menos 5, 10, 15, 20, 25, 30, 50 o más kD.

30 En algunos aspectos de la divulgación, la proteína CENH3 mutada comprenderá un dominio de proteína que actúa como un marcador detectable o seleccionable. Por ejemplo, una proteína marcadora seleccionable de ejemplo es fluorescente o un producto génico de resistencia a antibióticos o herbicidas. Los dominios de proteína seleccionables o detectables son útiles para controlar la presencia o ausencia de la proteína CENH3 mutada en un organismo.

35 B. PROTEÍNAS DEL CINETOCORO NO DE CENH3

Se cree que otras proteínas que constituyen el complejo del cinetocoro también pueden mutarse y expresarse en una planta que de otro modo no expresa la proteína del complejo del cinetocoro endógeno correspondiente para dar como
40 resultado una planta viable que, cuando se cruza con una planta de tipo silvestre que tiene un complejo del cinetocoro de tipo silvestre, genera progenie haploide a una frecuencia determinada (por ejemplo, al menos el 0,1, 0,5, 1, 5, 10, 20, % o más). Los miembros no de CENH3 de ejemplo del complejo del cinetocoro incluyen, por ejemplo, CENPC, MCM21, MIS12, NDC80 y NUF2.

45 Pueden identificarse proteínas del complejo del cinetocoro no de CENH3 mutadas activas (por ejemplo, CENPC, MCM21, MIS12, NDC80 o NUF2), por ejemplo, mediante mutagénesis aleatoria, mutagénesis dirigida a aminoácidos únicos o múltiples, mediante generación o supresiones de dominio de proteína completas o parciales, mediante fusión con secuencias de aminoácidos heterólogas o combinaciones de las mismas. Las proteínas del complejo del cinetocoro no de CENH3 mutantes "activas" se refieren a proteínas, que cuando se expresan en una planta en la que
50 la proteína del complejo del cinetocoro no de CENH3 correspondiente se desactiva o se inactiva, dan como resultado plantas viables, que cuando se cruzan con una planta de tipo silvestre, producen progenie haploide a una frecuencia superior a la normal (por ejemplo, al menos el 1, 5, 10, 20 % o más). En algunos aspectos de la divulgación, los polipéptidos CENPC, MCM21, MIS12, NDC80 o NUF2 mutados activos son sustancialmente idénticos a las SEQ ID NO: 96, 97, 98, 99 o 100, respectivamente. Pueden someterse a ensayo proteínas del complejo del cinetocoro no de CENH3 mutadas (por ejemplo, CENPC, MCM21, MIS12, NDC80 o NUF2) fácilmente mediante expresión recombinante de la proteína del complejo del cinetocoro no de CENH3 mutada en una planta que carece de proteína del complejo del cinetocoro no de CENH3 endógena, cruzando la planta transgénica (como masculina o femenina, dependiendo de la fertilidad) con una planta que expresa una proteína del complejo del cinetocoro no de CENH3 de tipo silvestre y después seleccionando para determinar la producción de progenie haploide.

60 En algunos aspectos de la divulgación, la proteína del complejo del cinetocoro no de CENH3 mutada es idéntica a una proteína del complejo del cinetocoro no de CENH3 endógena pero para 1, 2, 3, 4, 5, 6, 7, 8 o más (por ejemplo, 1-2, 1-4, 1-7) aminoácidos. Por ejemplo, en algunos aspectos de la divulgación, la proteína de tipo silvestre endógena de la planta es idéntica o sustancialmente idéntica a cualquiera de las SEQ ID NO: 96, 97, 98, 99 o 100 y la proteína del
65 complejo del cinetocoro no de CENH3 mutada difiere de la proteína del complejo del cinetocoro no de CENH3 endógena en 1, 2, 3, 4, 5, 6, 7, 8 o más (por ejemplo, 1-2, 1-4, 1-7) aminoácidos.

Opcionalmente, la secuencia de aminoácidos heteróloga puede comprender una o más secuencias de aminoácidos en el extremo amino y/o carboxilo y/o unir los dominios de cola y de plegamiento de histona. Por ejemplo, en algunos aspectos de la divulgación la proteína del complejo del cinetocoro no de CENH3 mutada comprende una secuencia de aminoácidos heteróloga unida a un extremo amino de la proteína del complejo del cinetocoro no de CENH3. La secuencia heteróloga puede ser cualquier secuencia. En algunos aspectos de la divulgación, la secuencia heteróloga se une al extremo amino de una proteína del complejo del cinetocoro no de CENH3 de tipo silvestre de otro modo, en donde la secuencia heteróloga interfiere con la función del centrómero. En algunos aspectos de la divulgación, la secuencia heteróloga comprende una secuencia de aminoácidos de al menos 5, 10, 15, 20, 25, 30, 50 o más kD.

En algunos aspectos de la divulgación, la proteína del complejo del cinetocoro no de CENH3 mutada comprenderá un dominio de proteína que actúa como un marcador detectable o seleccionable. Por ejemplo, una proteína marcadora seleccionable de ejemplo es fluorescente o un producto génico de resistencia a antibióticos o herbicidas. Los dominios de proteína seleccionables o detectables son útiles para controlar la presencia o ausencia de la proteína del complejo del cinetocoro no de CENH3 mutada en un organismo.

III. GENERACIÓN DE ORGANISMOS

La presente divulgación proporciona organismos que no expresan, o expresan a niveles reducidos (por ejemplo, menos del 90, 80, 70, 60, 50, 40, 30, 20 o 10 % de los niveles de tipo silvestre), una proteína CENH3 endógena o una proteína del complejo del cinetocoro no CENH3 y opcionalmente que expresa una proteína CENH3 mutada correspondiente o una proteína del complejo del cinetocoro no CENH3. Generalmente, la falta de una proteína del complejo del cinetocoro es letal, a menos que se complemente al menos parcialmente con una proteína del complejo del cinetocoro mutada como se describe en el presente documento. Se cree que hay varias formas de hacer un organismo que carezca o tenga una expresión reducida de una proteína del complejo del cinetocoro endógena pero que exprese una versión mutada de esa proteína.

En algunos aspectos de la divulgación, puede generarse una mutación de CENH3 en un gen de CENH3 endógeno (o una proteína del complejo del cinetocoro no CENH3) que reduzca o elimine la actividad o expresión de CENH3, o puede generarse una proteína del complejo del cinetocoro inactivada (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2). En estos aspectos de la divulgación, puede generarse un organismo heterocigótico para la inactivación o mutación del gen y puede introducirse un casete de expresión para la expresión de la proteína del complejo del cinetocoro mutada heteróloga correspondiente en el organismo. Después puede seleccionarse la progenie del heterocigoto que sea homocigótica para la mutación o la inactivación, pero que comprenda la proteína del complejo del cinetocoro mutada heteróloga expresada de forma recombinante. En consecuencia, la divulgación proporciona plantas, células vegetales u otros organismos en que uno o ambos alelos de CENH3 se inactivan o se mutan para que carezcan de manera significativa o esencialmente completa de actividad CENH3, es decir, suficientes para inducir letalidad embrionaria sin una expresión complementaria de una proteína del complejo del cinetocoro como se describe en el presente documento (por ejemplo, una proteína tailswap). La divulgación también proporciona plantas, células vegetales u otros organismos en los cuales uno o ambos alelos de un gen del complejo del cinetocoro no CENH3 están desactivados o mutados para carecer significativa o esencialmente de forma completa de actividad de la proteína del complejo del cinetocoro no CENH3, es decir, lo suficiente para inducir la letalidad del embrión sin una expresión complementaria de una proteína del complejo del cinetocoro mutada como se describe en el presente documento. En las plantas que tienen más de un conjunto diploide de cromosomas (por ejemplo, tetraploides), todos los alelos pueden inactivarse, mutar o desactivarse.

Como alternativa, puede introducirse el casete de expresión que codifica una proteína del complejo del cinetocoro mutada (por ejemplo, incluyendo, pero sin limitación, una proteína tailswap) en un organismo con un conjunto intacto de alelos de la proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) y después puede silenciarse el gen de la proteína del complejo del cinetocoro endógena (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) de cualquier manera conocida en la técnica. Como ejemplo, puede introducirse o expresarse un ARNip o microARN en el organismo que reduzca o elimine la expresión de la proteína del complejo del cinetocoro endógena (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2).

Idealmente, el ARNip silenciador u otro agente silenciador se selecciona para silenciar el gen de la proteína del complejo del cinetocoro endógena (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) pero no interfiere sustancialmente con la expresión de la proteína del complejo del cinetocoro mutada (por ejemplo, una proteína tailswap). En situaciones donde CENH3 endógena ha de inactivarse, esto puede conseguirse, por ejemplo, dirigiendo el ARNip a la sección codificante de la cola N-terminal o porciones no traducidas, o el ARNm de CENH3, dependiendo de la estructura de la proteína del complejo del cinetocoro mutada. Como alternativa, el transgén de proteína del complejo del cinetocoro mutada puede diseñarse con el uso de codones novedosos, de manera que carezca de homología de secuencia con el gen de proteína del complejo del cinetocoro endógena y con el ARNip silenciador.

IV. REDUCCIÓN O ELIMINACIÓN DE LA EXPRESIÓN DE PROTEÍNA DEL COMPLEJO DEL CINETOCORO ENDÓGENA

Pueden usarse varios métodos para inhibir, mutar o inactivar la expresión de una proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) en plantas. Por ejemplo, puede usarse tecnología antisentido convenientemente para inactivar la expresión génica. Para conseguir esto, un segmento de ácido nucleico del gen deseado se clona y se une operativamente a un promotor de manera que se transcribirá la cadena antisentido de ARN. Después, el casete de expresión se transforma en plantas y se produce la cadena antisentido de ARN. En células vegetales, se ha sugerido que el ARN antisentido inhibe la expresión génica evitando la acumulación de ARNm que codifica el polipéptido de interés, véase, por ejemplo, Sheehy et al., *Proc. Nat. Acad. Sci. USA*, 85: 8805-8809 (1988); Pnueli et al., *The Plant Cell* 6: 175-186 (1994); e Hiatt et al., Patente de los EE.UU. N.º 4.801.340.

La secuencia de ácido nucleico antisentido transformada en plantas será sustancialmente idéntica a al menos una porción del gen o genes endógenos que han de reprimirse. La secuencia, sin embargo, no tiene que ser perfectamente idéntica para inhibir la expresión. Por tanto, una molécula de ácido nucleico antisentido o sentido que codifica solo una porción de la proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) o una porción del ARNm de la proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) (incluyendo, pero sin limitación, porciones no traducidas del ARNm) puede ser útil para producir una planta en la que la expresión de proteína del complejo del cinetocoro se suprime. Los vectores de la presente divulgación se diseñan opcionalmente de manera que el efecto inhibitorio se aplique solo a una proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) y no afecte a la expresión de otros genes. En situaciones donde la CENH3 endógena ha de inactivarse, un método para conseguir este objetivo es dirigir la secuencia antisentido a secuencias de CENH3 (por ejemplo, secuencias de ARNm de cola o no traducidas) que no se encuentran en otras proteínas dentro de una familia de genes que presentan homología u homología sustancial con el gen de CENH3.

Para la supresión antisentido, tampoco es necesario que la secuencia introducida sea de longitud completa con respecto al producto de transcripción primario o al ARNm totalmente procesado. Generalmente, puede usarse una homología más alta para compensar el uso de una secuencia más corta. Además, no es necesario que la secuencia introducida tenga el mismo patrón de intrones o exones, y la homología de segmentos no codificantes puede ser igualmente eficaz. Por ejemplo, puede usarse una secuencia de entre aproximadamente 30 o 40 nucleótidos y, en algunos aspectos de la divulgación, deben usarse nucleótidos de longitud aproximadamente completa, aunque puede usarse una secuencia de al menos aproximadamente 20, 50, 100, 200 o 500 nucleótidos.

Las moléculas de ARN catalítico o ribozimas también pueden usarse para inhibir la expresión de un gen de la proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2). Es posible diseñar ribozimas que se emparejan específicamente con prácticamente cualquier ARN diana y escinden la cadena principal del fosfodiéster en una ubicación específica, inactivando funcionalmente de este modo el ARN diana. Cuando se realiza esta escisión, la propia ribozima no se altera y, por tanto, es capaz de reciclarse y escindir otras moléculas, convirtiéndola en una verdadera enzima. La inclusión de secuencias de ribozimas dentro de ARN antisentido les confiere actividad de escisión de ARN, aumentando de este modo la actividad de las construcciones.

Se han identificado varias clases de ribozimas. Una clase de ribozimas deriva de una serie de pequeños ARN circulares que son susceptibles de autoescisión y replicación en plantas. Los ARN se replican solos (ARN viroides) o con un virus auxiliar (ARN satélite). Los ejemplos incluyen los ARN del viroide de las manchas solares del aguacate y los ARN satélite del virus de la mancha anular del tabaco, el virus del estriado transitorio de la alfalfa, el virus del moteado aterciopelado del tabaco, el virus del moteado de *solanum nodiflorum* y el virus del moteado del trébol subterráneo. El diseño y uso de ribozimas específicas de ARN diana se describe en Haseloff et al. *Nature*, 334: 585-591 (1988).

Otro método de supresión es la supresión del sentido (también conocida como cosupresión). Se ha demostrado que la introducción de casetes de expresión en los que se configura un ácido nucleico en la orientación sentido con respecto al promotor es un medio eficaz para bloquear la transcripción de genes diana. Para un ejemplo del uso de este método para modular la expresión de genes endógenos, véase, Napoli et al., *The Plant Cell* 2: 279-289 (1990); Flavell, *Proc. Natl. Acad. Sci., USA* 91: 3490-3496 (1994); Kooter y Mol, *Current Opin. Biol.* 4: 166-171 (1993); y las Patentes de los EE.UU. N.º 5.034.323, 5.231.020 y 5.283.184.

Generalmente, cuando se desea la inhibición de la expresión, se produce cierta transcripción de la secuencia introducida. El efecto puede producirse cuando la secuencia introducida no contiene ninguna secuencia codificante per se, sino solo secuencias intrónicas o no traducidas homólogas a las secuencias presentes en la transcripción primaria de la secuencia endógena. La secuencia introducida generalmente será sustancialmente idéntica a la secuencia endógena que se pretende reprimir. Esta identidad mínima normalmente será superior a aproximadamente el 65 % de la secuencia de la de proteína del complejo del cinetocoro diana (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2), pero una identidad más alta puede ejercer una represión más eficaz de la expresión de las secuencias endógenas. En algunos aspectos de la divulgación, se usan secuencias con una identidad sustancialmente mayor, por ejemplo, se usan al menos de aproximadamente el 80, al menos aproximadamente el 95 % o el 100 % de identidad. Igual que con la regulación antisentido, el efecto puede diseñarse y someterse a ensayo para no afectar significativamente a la expresión de otras proteínas dentro de una familia similar de genes que presentan homología u homología sustancial.

Para la supresión sentido, tampoco es necesario que la secuencia introducida en el casete de expresión, que necesita una identidad inferior a la absoluta, sea de longitud completa, con respecto al producto de transcripción primario o el ARNm totalmente procesado. Esto puede preferirse para evitar la producción simultánea de algunas plantas que son sobreexpresadoras. Una identidad mayor en una secuencia más corta que la de longitud completa compensa una secuencia más larga y menos idéntica. Además, no es necesario que la secuencia introducida tenga el mismo patrón de intrones o exones y la identidad de los segmentos no codificantes será igualmente eficaz. En algunos aspectos de la divulgación, se usa una secuencia de los intervalos de tamaño indicados anteriormente para la regulación antisentido, es decir, 30-40, o al menos aproximadamente 20, 50, 100, 200, 500 o más nucleótidos.

La expresión de genes endógenos también puede suprimirse por medio de interferencia de ARN (ARNi) (y, de hecho, la cosupresión puede considerarse un tipo de ARNi), que usa un ARN bicatenario que tiene una secuencia idéntica o similar a la secuencia del gen diana. La ARNi es el fenómeno en el que cuando un ARN bicatenario que tiene una secuencia idéntica o similar a la del gen diana se introduce en una célula, se suprimen las expresiones tanto del gen exógeno insertado como del gen endógeno diana. El ARN bicatenario puede formarse a partir de dos ARN complementarios separados o puede ser un solo ARN con secuencias internamente complementarias que forman un ARN bicatenario. Aunque todavía se desconocen los detalles completos del mecanismo de la ARNi, se considera que el ARN bicatenario introducido se escinde inicialmente en fragmentos pequeños, que después sirven como índices del gen diana de alguna manera, degradando de este modo el gen diana. Se sabe que la ARNi también es eficaz en plantas (véase, por ejemplo, Chuang, C. F. y Meyerowitz, E. M., *Proc. Natl. Acad. Sci. USA* 97: 4985 (2000); Waterhouse et al., *Proc. Natl. Acad. Sci. USA* 95:13959-13964 (1998); Tabara et al. *Science* 282: 430-431 (1998); Matthew, *Comp. Funct. Genom.* 5: 240-244 (2004); Lu, et al., *Nucleic Acids Research* 32(21): e171 (2004)). Por ejemplo, para conseguir la supresión de la expresión de una proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) usando ARNi, un ARN bicatenario que tiene la secuencia de un ARNm que codifica la proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2), o una secuencia sustancialmente similar del mismo (incluyendo los modificados mediante ingeniería genética para no traducir la proteína) o fragmento del mismo, se introduce en una planta u otro organismo de interés. Después, las plantas/organismos resultantes pueden seleccionarse para un fenotipo asociado a la proteína diana (opcionalmente en presencia de la expresión de una proteína tailswap para evitar la letalidad) y/o controlando los niveles de ARN en estado estacionario para las transcripciones que codifican la proteína. Aunque no es necesario que los genes utilizados para ARNi sean totalmente idénticos al gen diana, pueden ser al menos un 70 %, 80 %, 90 %, 95 % o más idénticos a la secuencia del gen diana (por ejemplo, secuencias de CENH3 como se describen en el presente documento). Véase, por ejemplo, la Publicación de Patente de los EE.UU. N.º 2004/0029283 para un ejemplo de una secuencia de ARNip no idéntica utilizada para suprimir la expresión génica. Las construcciones que codifican una molécula de ARN con una estructura de tallo-bucle que no está relacionada con el gen diana y que se ubica distalmente con respecto a una secuencia específica para el gen de interés también pueden usarse para inhibir la expresión del gen diana. Véase, por ejemplo, la Publicación de Patente de los EE.UU. N.º 2003/0221211.

Los polinucleótidos de ARNi pueden abarcar el ARN diana de longitud completa o pueden corresponder a un fragmento del ARN diana. En algunos casos, el fragmento tendrá menos de 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1.000 nucleótidos correspondientes a la secuencia diana. Además, en algunos aspectos de la divulgación, estos fragmentos tienen al menos, por ejemplo, 10, 15, 20, 50, 100, 150, 200 o más nucleótidos de longitud. En algunos casos, los fragmentos para su uso en ARNi serán al menos sustancialmente similares a las regiones de una proteína diana que no aparecen en otras proteínas en el organismo o pueden seleccionarse para que tengan la menor similitud posible con otros transcritos de organismos, por ejemplo, seleccionados mediante comparación con las secuencias en el análisis de bases de datos de secuencias disponibles públicamente.

Los vectores de expresión que expresan continuamente ARNip en células transfectadas de forma transitoria y estable se han modificado mediante ingeniería genética para expresar ARN de horquilla pequeños, que se procesan *in vivo* en moléculas de ARNip capaces de realizar un silenciamiento específico de genes (Brummelkamp et al., *Science* 296: 550-553 (2002) y Paddison, et al., *Genes & Dev.* 16: 948-958 (2002)). Se analiza en más detalle el silenciamiento génico postranscripcional mediante ARN bicatenario en Hammond et al. *Nature Rev Gen* 2: 110-119 (2001), Fire et al. *Nature* 391: 806-811 (1998) y Timmons y Fire *Nature* 395: 854 (1998).

Un experto en la materia reconocerá que el sentido (incluyendo, pero sin limitación, ARNip) o el transcrito antisentido deben dirigirse a secuencias con la mayor variación entre miembros de la familia donde el objetivo es dirigir solo a un miembro de la familia de histonas (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2).

Otra forma más de suprimir la expresión de un gen vegetal endógeno es mediante la expresión recombinante de un microARN que suprime una diana (por ejemplo, un gen de CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2). Los microARN artificiales son ARN monocatenarios (por ejemplo, de entre 18-25 monómeros, generalmente 21 monómeros), que normalmente no se encuentran en plantas y que se procesan a partir de precursores de miARN endógenos. Sus secuencias se diseñan de acuerdo con los determinantes de la selección de la diana de miARN vegetal, de manera que el microARN artificial silencie específicamente su gen o genes diana y generalmente se describen en Schwab et al, *The Plant Cell* 18: 1121-1133 (2006) así como los métodos basados en internet para diseñar dichos microARN como se describe en el mismo. Véase también, la Publicación de Patente de los EE.UU. N.º

2008/0313773.

Los métodos para introducir mutaciones genéticas en genes de plantas y seleccionar plantas con rasgos deseados y pueden usarse para introducir mutaciones o inactivar una proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2). Por ejemplo, pueden tratarse semillas u otro material vegetal con un polinucleótido de inserción mutagénico (por ejemplo, transposón, ADN-T, etc.) o una sustancia química, de acuerdo con técnicas convencionales. Dichas sustancias químicas incluyen, pero sin limitación, las siguientes: sulfato de dietilo, etilenimina, metanosulfonato de etilo y N-nitroso-N-etilurea. Como alternativa, puede usarse radiación ionizante de fuentes tales como rayos X o rayos gamma. Entonces, pueden identificarse plantas que han mutado una proteína del complejo del cinetocoro (por ejemplo, CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2), por ejemplo, por fenotipo o mediante técnicas moleculares.

Las cadenas de proteínas modificadas también pueden diseñarse fácilmente utilizando diversas técnicas de ADN recombinante bien conocidas por los expertos en la materia y descritas, por ejemplo, en Sambrook *et al.*, citado anteriormente. También puede usarse hidroxilamina para introducir mutaciones de base únicas en la región codificante del gen (Sikorski et al., *Meth. Enzymol.*, 194: 302-318 (1991)). Por ejemplo, las cadenas pueden variar con respecto a la secuencia de origen natural en el nivel de estructura primaria en sustituciones, adiciones y supresiones de aminoácidos. Estas modificaciones pueden usarse en varias combinaciones para producir la cadena de proteína modificada final.

Como alternativa, puede usarse recombinación homóloga para inducir modificaciones o inactivaciones de genes dirigidas dirigiendo específicamente un gen de proteína del complejo del cinetocoro (por ejemplo, el gen *CENH3*, *CENPC*, *MCM21*, *MIS12*, *NDC80* o *NUF2*) *in vivo* (véase, en general, Grewal y Klar, *Genetics*, 146: 1221-1238 (1997) y Xu et al., *Genes Dev.*, 10: 2411-2422 (1996)). La recombinación homóloga se ha demostrado en plantas (Puchta et al., *Experientia*, 50: 277-284 (1994); Swoboda et al., *EMBO J.*, 13: 484-489 (1994); Offringa et al., *Proc. Natl. Acad. Sci. USA*, 90: 7346-7350 (1993); y Kempin et al., *Nature*, 389: 802-803 (1997)).

En la aplicación de la tecnología de recombinación homóloga a los genes de la divulgación, se hacen mutaciones *in vitro* en porciones seleccionadas de secuencias de genes de una proteína del complejo del cinetocoro (incluyendo las regiones 5' corriente arriba, 3' corriente abajo e intragénicas) tales como las que se desvelan en el presente documento y después se introducen en la planta deseada usando técnicas convencionales. Puesto que se sabe que la eficacia de la recombinación homóloga depende de los vectores utilizados, el uso de vectores de direccionamiento genético dicistrónico como se describe en Mountford et al., *Proc. Natl. Acad. Sci. USA*, 91: 4303-4307 (1994); y Vaulont et al., *Transgenic Res.*, 4: 247-255 (1995) se usa convenientemente para aumentar la eficiencia de selección para la expresión del gen *CENH3* alterado en plantas transgénicas. El gen mutado interactuará con el gen de tipo silvestre diana de manera que la recombinación homóloga y el reemplazo dirigido del gen de tipo silvestre se produzcan en las células vegetales transgénicas, dando como resultado la supresión de la actividad proteínica del complejo del cinetocoro.

V. PREPARACIÓN DE VECTORES RECOMBINANTES

Para usar secuencias aisladas en las técnicas anteriores, se preparan vectores de ADN recombinante adecuados para la transformación de células vegetales. Se conocen bien técnicas para transformar una amplia diversidad de especies de plantas superiores y se describen en la bibliografía técnica y científica, por ejemplo, Weising et al., *Ann. Rev. Genet.* 22: 421-477 (1988). Una secuencia de ADN que codifica el polipéptido deseado, por ejemplo, las fusiones de proteína tailswap como se describen en el presente documento y/o ARNip, antisentido u otras construcciones silenciadoras, se combinará con secuencias reguladoras del inicio de la transcripción y la traducción que dirigirán la transcripción de la secuencia a partir del gen en los tejidos previstos de la planta transformada.

Por ejemplo, puede emplearse un fragmento promotor vegetal que dirigirá la expresión del gen en todos los tejidos de una planta regenerada. Como alternativa, el promotor vegetal puede dirigir la expresión del polinucleótido de la divulgación en un específico (promotores específicos de tejido), órgano (promotores específicos de órgano) tejido o puede estar de otro modo bajo un control ambiental más preciso (promotores inducibles). Los ejemplos de promotores específicos de tejido bajo control del desarrollo incluyen promotores que inician la transcripción solo en determinados tejidos, tales como frutos, semillas, flores, pistilos o anteras. Los promotores adecuados incluyen los de genes que codifican proteínas de almacenamiento o la proteína de membrana del cuerpo lipídico, la oleosina.

Si se desea una expresión adecuada del polipéptido, debe incluirse una región de poliadenilación en el extremo 3' de la región codificante. La región de poliadenilación puede derivar del gen natural, de una diversidad de otros genes vegetales o de ADN-T.

El vector que comprende las secuencias (por ejemplo, promotores o regiones codificantes) de genes de la divulgación que se describen en el presente documento también puede comprender, por ejemplo, un gen marcador que confiera un fenotipo seleccionable en células vegetales. Por ejemplo, el marcador puede codificar resistencia a biocidas, particularmente resistencia a antibióticos, tal como resistencia a la kanamicina, G418, bleomicina, higromicina o resistencia a herbicidas, tal como resistencia a clorosulfurón o Basta.

Promotores constitutivos

Puede emplearse un promotor o un fragmento activo del mismo, que dirigirá la expresión de un ácido nucleico que codifica una proteína de fusión de la divulgación, en todas las células o tejidos transformados, por ejemplo, como los de una planta regenerada. Dichos promotores se denominan en el presente documento promotores "constitutivos" y son activos en la mayoría de las condiciones ambientales y estados de desarrollo o diferenciación celular. Los ejemplos de promotores constitutivos incluyen los de virus que infectan plantas, tales como la región de inicio de la transcripción 35S del virus del mosaico de la coliflor (CaMV) (véase, por ejemplo, Dagless, *Arch. Virol.* 142: 183-191 (1997)); el promotor 1' o 2' derivado del ADN-T de *Agrobacterium tumefaciens* (véase, por ejemplo, Mengiste citado anteriormente (1997); O'Grady, *Plant Mol. Biol.* 29: 99-108 (1995)); el promotor del virus del mosaico del tabaco; el promotor del virus del mosaico Figwort (véase, por ejemplo, Maiti, *Transgenic Res.* 6: 143-156 (1997)); promotores de actina, tales como el promotor del gen de actina de *Arabidopsis* (véase, por ejemplo, Huang, *Plant Mol. Biol.* 33: 125-139 (1997)); promotores del gen de la alcohol deshidrogenasa (Adh) (véase, por ejemplo, Millar, *Plant Mol. Biol.* 31: 897-904 (1996)); *ACT11* de *Arabidopsis* (Huang et al., *Plant Mol. Biol.* 33: 125-139 (1996)), *Cat3* de *Arabidopsis* (N.º de GenBank U43147, Zhong et al., *Mol. Gen. Genet.* 251: 196-203 (1996)), el gen que codifica la proteína desaturasa del transportador de estearoil-acilo de *Brassica napus* (N.º de Genbank X74782, Solocombe et al., *Plant Physiol.* 104: 1167-1176 (1994)), *Gpc1* de maíz (N.º de GenBank X15596, Martinez et al., *J. Mol. Biol.* 208: 551-565 (1989)), *Gpc2* de maíz (N.º de GenBank U45855, Manjunath et al., *Plant Mol. Biol.* 33: 97-112 (1997)), otras regiones de inicio de la transcripción de diversos genes vegetales conocidas por los expertos. Véase también Holtorf, "Comparison of different constitutive and inducible promoters for the overexpression of transgenes in *Arabidopsis thaliana*", *Plant Mol. Biol.* 29: 637-646 (1995). Los promotores constitutivos adicionales incluyen, por ejemplo, los promotores del gen de la poliubiquitina de *Arabidopsis thaliana*, UBQ3 y UBQ10, (Norris et al., *Plant Mol. Biol.* 21: 895 (1993)), también son útiles para dirigir la expresión génica.

Promotores inducibles

Puede usarse opcionalmente un promotor inducible para controlar (1) la expresión de un micro ARN artificial, ARNip u otro polinucleótido silenciador, (2) y activar simultáneamente la expresión de la proteína transgénica mutada (por ejemplo, tailswap) o (3) ambos (1) y (2). Esto tendría la ventaja de tener una planta normal (por ejemplo, una que podría tener una mayor fertilidad) hasta la inducción, que después crearía gametos listos para inducir haploides.

Promotores específicos de tejido

Una alternativa es regular negativamente la proteína endógena (por ejemplo, mediante silenciamiento génico) en un tejido específico (por ejemplo, al menos en los gametofitos maduros (ya sea polen o saco embrionario)) y reemplazarla solo en este tejido con un promotor específico que impulsa la expresión de una proteína tailswap. En algunos aspectos de la divulgación, se usa el mismo promotor específico de tejido para impulsar un micro ARN artificial, ARNip u otro polinucleótido silenciador y el transgén codificante de tailswap de rescate.

VI. PRODUCCIÓN DE PLANTAS TRANSGÉNICAS O CÉLULAS VEGETALES

Pueden introducirse construcciones de ADN de la divulgación en el genoma de la planta hospedadora deseada mediante una diversidad de técnicas convencionales. Por ejemplo, la construcción de ADN puede introducirse directamente en el ADN genómico de la célula vegetal usando técnicas tales como la electroporación y la microinyección de protoplastos de células vegetales o las construcciones de ADN pueden introducirse directamente en el tejido vegetal usando métodos biolísticos, tales como el bombardeo de partículas de ADN. Como alternativa, las construcciones de ADN pueden combinarse con regiones flanqueantes de ADN-T adecuadas e introducirse en un vector hospedador convencional de *Agrobacterium tumefaciens*. Las funciones de virulencia del hospedador *Agrobacterium tumefaciens* dirigirán la inserción de la construcción y el marcador adyacente en el ADN de la célula vegetal cuando la célula es infectada por la bacteria.

Se conocen en la técnica técnicas de microinyección y se describen bien en la bibliografía científica y de patentes. La introducción de construcciones de ADN usando precipitación con polietilenglicol se describe en Paszkowski et al., *Embo J.* 3: 2717-2722 (1984). Se describen técnicas de electroporación en Fromm et al., *Proc. Natl. Acad. Sci. USA* 82: 5824 (1985). Se describen técnicas de transformación biolística en Klein et al., *Nature* 327: 70-73 (1987).

Se describen bien técnicas de transformación mediadas por *Agrobacterium tumefaciens*, que incluyen el desarme y el uso de vectores binarios, en la bibliografía científica. Véanse, por ejemplo, Horsch et al., *Science* 233: 496-498 (1984) y Fraley et al., *Proc. Natl. Acad. Sci. USA* 80: 4803 (1983).

Pueden cultivarse células vegetales transformadas que derivan de cualquiera de las técnicas de transformación anteriores para regenerar una planta completa que posea el genotipo transformado y, por tanto, el fenotipo deseado, tal como el aumento de la resistencia a enfermedades en comparación con una planta de control que no se transformó o se transformó con un vector vacío. Dichas técnicas de regeneración se basan en la manipulación de determinadas fitohormonas en un medio de crecimiento de cultivo tisular, que normalmente se basa en un marcador de biocida y/o

herbicida que se ha introducido junto con las secuencias nucleotídicas deseadas. Se describe la regeneración de plantas a partir de protoplastos cultivados en Evans et al., *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, págs. 124-176, MacMillan Publishing Company, Nueva York, 1983; y Binding, *Regeneration of Plants, Plant Protoplasts*, págs. 21-73, CRC Press, Boca Ratón, 1985. También puede obtenerse regeneración a partir de callos de plantas, explantes, órganos o partes de los mismos. Dichas técnicas de regeneración se describen en general en Klee et al., *Ann. Rev. of Plant Phys.* 38: 467-486 (1987).

Los ácidos nucleicos y los polipéptidos codificados de la divulgación pueden usarse para conferir las características que se describen en el presente documento, incluyendo la capacidad de generar progenie haploide, como se describe en el presente documento, esencialmente en cualquier planta. Por tanto, la invención tiene uso en una amplia gama de plantas, incluyendo dicotiledóneas o monocotiledóneas, incluyendo, por ejemplo, especies de los géneros *Asparagus, Atropa, Avena, Brassica, Citrus, Citrullus, Capsicum, Cucumis, Cucurbita, Daucus, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum, Lolium, Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Oryza, Panicum, Pennisetum, Persea, Pisum, Pyrus, Prunus, Raphanus, Secale, Senecio, Sinapis, Solanum, Sorghum, Trigonella, Triticum, Vitis, Vigna* y *Zea*.

VII. MÉTODOS DE MEJORAMIENTO MEJORADO

El cruce de plantas que carecen de una proteína del complejo del cinetocoro endógena y expresan una proteína del complejo del cinetocoro mutada activa como se describe en el presente documento (por ejemplo, una proteína tailswap u otra proteína CENH3 mutada o del complejo del cinetocoro no CENH3) ya sea como un progenitor de polen o de óvulos con una planta que expresa una proteína del complejo del cinetocoro endógena (por ejemplo, proteína CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) dará como resultado al menos algo de progenie (por ejemplo, al menos el 0,1 %, 0,5 %, 1 %, 5 %, 10 %, 20 % o más) que son haploides y comprenden solo cromosomas de la planta que expresa la proteína del complejo del cinetocoro. Por tanto, la presente divulgación permite la generación de plantas haploides que tengan todos sus cromosomas de una planta de interés cruzando la planta de interés con una planta que expresa de forma transgénica la proteína del complejo del cinetocoro mutada y recogiendo la semilla haploide resultante.

Como se ha indicado anteriormente, la planta que expresa una proteína CENH3 endógena de tipo silvestre puede cruzarse como progenitor masculino o femenino. Un aspecto único de la presente divulgación es que permite la generación de una planta (u otro organismo) que tiene solo los cromosomas nucleares del progenitor masculino y un citoplasma del progenitor femenino con mitocondrias y plastidios asociados, cuando el progenitor de tailswap es el progenitor masculino.

Aunque las plantas que carecen de un gen de CENH3 endógeno y que expresan una proteína CENH3 mutada compuesta de GFP-cola de histona H3-dominio de plegamiento de histona CENH3 tienen fertilidad masculina limitada, se ha descubierto que las plantas que carecen de un gen de CENH3 endógeno y que expresan las dos, una proteína CENH3 mutada compuesta por GFP-cola de histona H3-dominio de plegamiento de histona CENH3 y GFP-CENH3 de tipo silvestre, dan como resultado plantas con mayor fertilidad masculina, haciéndolas convenientes para su uso como progenitor masculino, así como femenino, en el cruce. En general, la divulgación proporciona la expresión de dos o más proteínas del complejo del cinetocoro mutadas diferentes en una planta (por ejemplo, una planta que carece de expresión de la proteína o proteínas del complejo del cinetocoro endógenas correspondientes).

Una vez generadas, las plantas haploides pueden usarse para una diversidad de esfuerzos útiles, incluyendo, pero sin limitación, la generación de plantas doblemente haploides, que comprenden una copia duplicada exacta de los cromosomas. Estas plantas doblemente haploides son de uso particular para acelerar el fitomejoramiento, por ejemplo. Se conoce una amplia diversidad de métodos para generar organismos doblemente haploides a partir de organismos haploides.

Pueden tratarse células haploides somáticas, embriones haploides, semillas haploides o plantas haploides producidas a partir de semillas haploides con un agente de duplicación cromosómica. Pueden regenerarse plantas doblemente haploides homocigóticas a partir de células haploides poniendo en contacto las células haploides, incluyendo, pero sin limitación, el calo haploide, con agentes de duplicación cromosómica, tales como colchicina, herbicidas antimicrotúbulos u óxido nitroso para crear células homocigóticas doblemente haploides.

Se desvelan métodos de duplicación cromosómica en, por ejemplo, las Patentes de los EE.UU. N.º 5.770.788; 7.135.615 y las Publicaciones de Patente de los EE.UU. N.º 2004/0210959 y 2005/0289673; Antoine-Michard, S. et al., *Plant Cell, Tissue Organ Cult.*, Cordrecht, Países Bajos, Kluwer Academic Publishers 48 (3): 203-207 (1997); Kato, A., *Maize Genetics Cooperation Newsletter* 1997, 36-37; y Wan, Y. et al., *Trends Genetics* 77: 889-892 (1989). Wan, Y. et al., *Trends Genetics* 81: 205-211 (1991). Los métodos pueden implicar, por ejemplo, poner en contacto la célula haploide con óxido nitroso, herbicidas antimicrotúbulos o colchicina. Opcionalmente, los haploides pueden transformarse con un gen heterólogo de interés, si se desea.

Pueden cruzarse plantas doblemente haploides con otras plantas para generar F1, F2 o generaciones posteriores de plantas con los rasgos deseados.

VIII. Organismos no vegetales

Se cree que el contenido de la divulgación es también funcional en organismos no vegetales que no tengan cromosomas sexuales no emparejados. Los expertos en la materia pueden generar de este modo una proteína del complejo del cinetocoro mutada (incluyendo, pero sin limitación, una proteína tailswap) basada en una secuencia de proteína del complejo del cinetocoro de un organismo particular (por ejemplo, proteína CENH3, CENPC, MCM21, MIS12, NDC80 o NUF2) e inactivar el gen de proteína del complejo del cinetocoro endógeno correspondiente según sea apropiado para esos organismos. Los ejemplos de organismos no vegetales para los que se cree que la divulgación es aplicable incluyen, pero sin limitación, levadura y otros hongos, así como a animales que carezcan de cromosomas sexuales no emparejados (por ejemplo, XY) u otros cromosomas para los que los haploides no son viables.

Ejemplos

Los siguientes ejemplos se ofrecen para ilustrar, pero no para limitar la invención reivindicada.

La producción de plantas haploides que heredan cromosomas de un solo progenitor puede acelerar en gran medida el fitomejoramiento (Dunwell, J. M., *Plant Biotechnol J* en prensa; Forster, B. P. et al., *Trends Plant Sci* 12: 368-75 (2007); Forster, B. P. y Thomas, W. T. B. en *Plant Breeding Reviews* (ed. Janick, J.) 57-88 (John Wiley & Sons, Inc., 2005)). Los haploides generados a partir de un individuo heterocigótico y convertidos en diploide crean estirpes homocigóticas instantáneas, evitando generaciones de endogamia. Generalmente se usan dos métodos para producir haploides: En primer lugar, pueden regenerarse células de gametofito cultivadas en plantas haploides (Guha, S. y Maheshwari, S. C., *Nature* 204: 497 (1964)), pero muchas especies y genotipos son recalcitrantes a este proceso (Forster, B. P. et al., *Trends Plant Sci* 12: 368-75 (2007); Wedzony, M. et al. en *Advances in Haploid Production in Higher Plants* (eds. Touraev, A., Forster, B. P. y Jain, S. M.) 1-33 (Springer, 2009)). En segundo lugar, pueden inducirse haploides a partir de cruces interespecíficos raros, en los que se elimina un genoma parental después de la fertilización (Bains, G. S. y Howard, H. W., *Nature* 166: 795 (1950); Barclay, I. R., *Nature* 256: 410-411 (1975); Burk, L. G. et al., *Science* 206: 585 (1979); Clausen, R. E. y Mann, M. C., *Proc Natl Acad Sci USA* 10: 121-124 (1924); Hougas, H. W. y Peloquin, S. J., *Nature* 180: 1209-1210 (1957); Kasha, K. J. y Kao, K. N., *Nature* 225: 874-6 (1970)). La base molecular para la eliminación del genoma no se comprende, pero una teoría postula que los centrómeros de las dos especies parentales interactúan de manera desigual con el huso mitótico, provocando una pérdida selectiva de cromosomas (Bennett, M. D. et al., *Chromosoma* 54: 175-200 (1976); Finch, R. A., *Chromosoma* 88: 386-393 (1983); Laurie, D. A. y Bennett, M. D., *Genome* 32: 953-961 (1989)). En este caso se demuestra que puede generarse fácilmente *Arabidopsis thaliana* haploide a través de semillas mediante manipulación de una única proteína de centrómero, la histona específica de centrómero CENH3/CENP-A. Cuando los mutantes con *cenh3* anulado que expresan proteínas CENH3 alteradas se cruzan con el tipo silvestre, los cromosomas del mutante se eliminan, produciendo progenie haploide. Los haploides se convierten espontáneamente en diploides fértiles a través de no reducción meiótica, lo que permite perpetuar su genotipo. Pueden generarse haploides maternos y paternos a través de cruces recíprocos. La eliminación del genoma mediada por centrómeros también se ha aprovechado para convertir una *Arabidopsis* tetraploide natural en un diploide, reduciendo su ploidía para simplificar el mejoramiento. Como CENH3 es universal en eucariotas, el método de los presentes inventores puede extenderse para producir haploides en cualquier especie vegetal.

Los centrómeros son los loci cromosómicos que se unen a microtúbulos del huso para mediar la herencia fiel del genoma durante la división celular. Se especifican epigenéticamente mediante la incorporación de CENH3 (CENP-A en seres humanos, *HTR12* en *A. thaliana* (Talbert, P. B. et al., *Plant Cell* 14: 1053-66 (2002))), una variante de histona H3 que reemplaza la H3 convencional en nucleosomas centroméricos (Henikoff, S. y Dalal, Y., *Curr Opin Genet Dev* 15: 177-84 (2005))). Se aisló *Cenh3-1*, un mutante embriotóxico letal nulo en *A. thaliana* que permite reemplazar totalmente la CENH3 nativa con variantes modificadas. Las plantas *cenh3-1* complementadas con CENH3 marcada con proteína verde fluorescente transgénica (*GFP-CENH3*) tienen un fenotipo de tipo silvestre. *cenh3-1* también puede rescatarse mediante "*GFP-tailswap*", un transgén en el que el dominio de cola N-terminal hipervariable de CENH3 se reemplazó con la cola de la H3 convencional, usando la variante H3.3 (codificada por *Atg13370*). *GFP-tailswap* se marcó en su extremo N con GF y contenía la cola N-terminal de H3 fusionada con el dominio de plegamiento de histona de CENH3 de la siguiente manera:
cola de H3: MARTKQSARKSHGGKAPTKQLATKAARKSAPTTGGVKKPHRFR (SEQ ID NO: 95) unida al dominio de histona CENH3:

PGTVALKEIRHFQKQTNLLIPAASFIREVRSITHMLAPPQINRWTAELVALQEAAED

YLVGLFSDSMLCAIHARRVTLMRKDFELARRLGKGRPW (SEQ ID NO:109).

Las plantas "*GFP-tailswap*" (*cenh3-1* rescatado mediante un transgén *GFP-tailswap*) mostraron una mitosis precisa, ya que no se detectó aneuploidía en las células somáticas. Sin embargo, las plantas *GFP-tailswap* eran estériles tras florecer, lo que indica que pueden tener un defecto específico en la meiosis. *GFP-tailswap* era principalmente estéril

masculina, aunque podría usarse como donante de polen si se agrupasen muchas anteras. Cuando se cruzaron como hembras con un macho de tipo silvestre, las plantas *GFP-tailswap* fueron un 60-70 % tan fértiles como las de tipo silvestre.

- 5 Cuando se polinizó *GFP-tailswap* por el tipo silvestre, se observaron varios fenotipos inusuales en la progenie F1. En primer lugar, un 80-95 % de los óvulos fertilizados abortaron precozmente en el desarrollo, produciendo semillas inviables (Tabla 1).

Tabla 1 | Las plantas haploides contienen solo el genoma nuclear de su progenitor de tipo silvestre.

Cruce	Semillas /silicua	% de semillas normales	plantas totales analizadas	haploides (%)	diploides (%)	aneuploides (%)
Col-0 TS x Col-0 TS	52 ± 6 (n = 23)	99,5	224	0 (0)	224 (100)	0 (0)
<i>GFP-tailswap</i> x <i>GFP-tailswap</i>	0,6 (n = 1206)	80	213	0 (0)	197 (92)	16 (8)
<i>GFP-tailswap</i> x Col-0 TS	32 ± 9 (n = 40)	12	67	23 (34)	23 (34)	21 (32)
Col-0 TS x <i>GFP-tailswap</i>	nd	nd	116	5 (4)	99 (85)	12 (11)
<i>GFP-tailswap</i> x <i>Ler</i> TS	30 ± 4 (n = 22)	23	127	32* (25)	32 (25)	63 (50)
<i>GFP-tailswap</i> x <i>Ws-0</i> TS	23 ± 5 (n = 14)	8	22	10* (45)	7 (32)	5 (28)
<i>GFP-tailswap</i> x <i>C24/Ler</i> TS	28 ± 5 (n = 13)	30	117	34† (29)	39 (33)	44 (38)
<i>C24/Ler</i> estéril masculina x <i>GFP-tailswap</i>	22 ± 14 (n = 18)	63	226	12† (5)	206 (91)	8 (4)
<i>GFP-CENH3</i> x <i>GFP-CENH3</i>	53 ± 4 (n = 21)	99	209	0 (0)	209 (100)	0 (0)
<i>GFP-CENH3</i> x TS	54 ± 7 (n = 18)	67	164	8 (5)	109 (66)	47 (29)
TS x <i>GFP-CENH3</i>	48 ± 6 (n = 13)	96	112	0 (0)	108 (96)	4 (4)
				diploides (%)	triploides (%)	aneuploides (%)
<i>GFP-tailswap</i> x <i>Wa-1</i> (tetraploide)	21 ± 6 (n = 96)	1,8	41	11 (27)	0	30 (73)

En segundo lugar, aunque se esperaba que la descendencia viable fuera diploides heterocigóticos para *cenH3-1* y hemicigóticos para el transgén *GFP-tailswap*, 10 de 16 plantas tenían solamente *CENH3* de tipo silvestre y carecían de *GFP-tailswap*. Cada una de estas plantas era estéril a pesar de tener un genotipo de tipo silvestre. Además, cruzar *GFP-tailswap* con un mutante masculino de *quartet* también produjo descendencia F1 estéril (3/5 plantas) que mostró el fenotipo mutante *quartet* de polen fusionado, a pesar del hecho de que *quartet* es recesivo y se esperaba que el progenitor *GFP-tailswap* transmitiese un alelo *QUARTET* de tipo silvestre. Estas sorprendentes observaciones sugirieron que la progenie estéril había perdido cromosomas de su madre *GFP-tailswap* y, por tanto, tenía menos cromosomas que *A. thaliana* diploide (2n = 10). Se examinó el cariotipo de estas plantas y se descubrió que eran haploides que contenían solo cinco cromosomas.

Como los centrómeros controlan la herencia cromosómica, se razonó que los cromosomas que entraron en el cigoto que contenía la variante *GFP-tailswap* de *CENH3* se desagregarían y se perderían, creando plantas haploides con cromosomas solo de su progenitor de tipo silvestre. Para confirmar esto, se cruzaron plantas *GFP-tailswap* (en el acceso Col-0) con varios accesos polimórficos y haploides F1 genotipados para marcadores en los cinco cromosomas de *A. thaliana* (Tabla 1). Independientemente del progenitor de tipo silvestre utilizado, las plantas haploides contenían invariablemente solo cromosomas de tipo silvestre (haploides paternos), lo que indica que se eliminó el genoma de *GFP-tailswap* (se genotipó un total de 42 haploides). Adicionalmente, los resultados de los presentes inventores demuestran que el proceso de inducción de haploides mediante la eliminación del genoma mediada por centrómero es independiente del genotipo del progenitor de tipo silvestre.

La eliminación del genoma inducida por las alteraciones de *CENH3* no es específica del transgén *GFP-tailswap*. El

cruce de mutantes de *cenh3-1* complementados por *GFP-CENH3* con el tipo silvestre también produjo plantas haploides, pero a una frecuencia inferior a la de *GFP-tailswap* (Tabla 1). No se observó progenie haploide de plantas autofertilizadas *GFP-tailswap* o *GFP-CENH3* (Tabla 1). Los resultados de los presentes inventores sugieren que perturbaciones generales en la estructura del centrómero son suficientes para impedir la segregación cromosómica durante la mitosis cigótica, creando un embrión haploide cuando los cromosomas que contienen CENH3 mutante compiten con el tipo silvestre en el mismo huso.

Los haploides se generan de manera eficiente a partir de un cruce *GFP-tailswap* x tipo silvestre, que comprende un 25-45 % de descendencia viable (Tabla 1). La progenie restante era híbridos diploides o híbridos aneuploides que mostraban los fenotipos de desarrollo típicos de plantas de *A. thaliana* con más de 10 cromosomas (Henry, I. M. et al., *Genetics* 170: 1979-88 (2005)) (Tabla 1). La aneuploidía también podría explicar el alto nivel de aborto de semillas en un cruce de *GFP-tailswap* x tipo silvestre, ya que algunos embriones con cariotipos desequilibrados pueden ser inviables.

Los haploides uniparentales pueden contener el genoma de su progenitor femenino o masculino. También se obtuvieron haploides cruzando una hembra de tipo silvestre con *GFP-tailswap* como donante de polen (Tabla 1). En este caso, la progenie haploide es de origen puramente materno. El genotipado del genoma plastídico mostró que los haploides tanto maternos como paternos contenían el citoplasma de su progenitor materno. Se hicieron haploides maternos o paternos usando plantas *GFP-tailswap* como progenitor masculino o femenino respectivamente en un cruce con tipo silvestre. La proporción de haploides y aneuploides fue mucho menor si una hembra de tipo silvestre se cruzó con un macho *GFP-tailswap* (Tabla 1). Se presume que si se expresa *CENH3* antes en el desarrollo del genoma materno (tipo silvestre), podría incorporarse CENH3 de tipo silvestre en los cromosomas paternos derivados de *GFP-tailswap*, evitando la eliminación del genoma en un cruce tipo silvestre x *GFP-tailswap*.

Las plantas haploides de *A. thaliana* son morfológicamente similares a las diploides, pero son comparativamente más pequeñas en tamaño. Precocemente en el desarrollo vegetativo, los haploides tienen hojas de roseta más estrechas. Después de la floración, los haploides producen más hojas a partir de meristemas secundarios. Las flores haploides son más pequeñas que las flores diploides, siguiendo la tendencia general de que el tamaño de la flor aumenta con la ploidía en *A. thaliana*. Los haploides son generalmente estériles. Contienen una sola copia de cada cromosoma y no pueden someterse a un emparejamiento homólogo en la meiosis, dando como resultado gametos que no contienen un complemento completo de cromosomas. Las plantas haploides maternas y paternas tenían una morfología adulta similar. Esto es coherente con el hecho de que toda la impresión documentada en *A. thaliana* ocurre en el endospermo efímero, una estructura confinada a la semilla.

Para aprovechar el potencial de los haploides en la mejora de los cultivos, su genoma debe duplicarse para generar diploides fértiles (haploides duplicados) (Forster, B. P., et al. *Trends Plant Sci* 12: 368-75 (2007)). Una inspección minuciosa de haploides de *A. thaliana* reveló que silicuas al azar tenían una o dos semillas. Cada planta haploide produjo un total de 50-2500 semillas dependiendo de la accesión parental de tipo silvestre (Tabla S1).

Tabla S1. Número de semillas diploides espontáneas producidas por haploides de *A. thaliana*

Planta	Col-0	Ler	Ws-0
1	54	951	1662
2	68	2115	293
3	91	352	2343
4	214	520	1532
5	349	325	2679
6	421	101	215
7	537	219	913
8	121	1013	
9	134	424	
10	85	630	
11	99	1346	
Media	197,5454545	726,9090909	1376,714286
Desviación típica	164,1294389	594,8035734	955,1793399

Una mayoría (95 %) de estas semillas parecía normal y dio lugar a diploides fértiles. Para abordar cómo los haploides dieron lugar a semillas diploides, se analizó la segregación cromosómica durante la meiosis masculina haploide. Durante la profase I, los cinco cromosomas permanecieron separados como univalentes, que se alinearon adecuadamente en la metafase I. En la anafase I, la mayoría de los meiocitos mostraron una segregación reductora desequilibrada (4-1, 3-2, etc.). La meiosis II en estos casos dio lugar a tétradas aneuploides. En una pequeña minoría de células en anafase I, los 5 univalentes migraron hacia un solo polo (segregación 5-0). En la meiosis II posterior, las cromátidas hermanas se segregaron por igual, dando lugar a díadas haploides y gametos viables. Por tanto, se supone que la no reducción ocasional durante la meiosis haploide tanto masculina como femenina produjo haploides

duplicados a través de autofertilización, de forma coherente con observaciones anteriores (Chase, S. S., *Botanical Review* 35: 117-167 (1969); Jauhar, P. P. et al., *Crop Science* 40: 1742-1749 (2000)). En raras ocasiones, se observó duplicación cromosómica espontánea en tejidos somáticos de plantas haploides de *A. thaliana*; una rama lateral de la inflorescencia principal (2 de 78 plantas) o una silicua aleatoria (6 de 78 plantas) mostraron un conjunto completo de semillas. El inhibidor de la polimerización de microtúbulos colchicina también induce la duplicación cromosómica somática en *A. thaliana* haploide y los brotes diploides que se regeneran después del tratamiento muestran un conjunto completo de semillas. Aunque se han producido haploides de *A. thaliana* a través de cultivo de anteras (Avetisov, V. A., *Genetika*, 12: 17-25 (1976)), los diploides espontáneos recuperados en estos experimentos eran supuestamente estériles (Scholl, R. y Amos, J. A., *Z Pflanzenphysiol* 96: 407-414 (1980)) y el método no se ha adoptado ampliamente. La facilidad de generar haploides a través de semilla mediante la alteración de CENH3 y de convertir haploides en diploides permite la generación a gran escala de haploides duplicados en *A. thaliana*.

Muchos cultivos comerciales son poliploides (Udall, J. A. y Wendel, J. F., *Crop Sci*, 46: S3-S14 (2006)), pero el análisis genético de poliploides es tedioso. La reducción de la ploidía de estos cultivos facilitará el mejoramiento fácil, por lo que se sometió a ensayo si la eliminación del genoma mediada por centrómeros podría reducirse a escala de un tetraploide a un diploide. *A. thaliana* es predominantemente diploide, pero existen accesos tetraploides (Henry, I. M. et al., *Genetics* 170: 1979-88 (2005)). Se cruzó *GFP-tailswap* con el tetraploide natural Warschau-1 (Wa-1) y aunque se abortó más del 98 % de las semillas, la progenie F1 viable incluía plantas diploides sintéticas que contenían solo cromosomas de Wa-1 (Tabla 1). Por tanto, es posible extender la eliminación del genoma mediada por centrómeros para reducir a la mitad la ploidía de los poliploides.

Se hipotetizó anteriormente que la incompatibilidad de centrómeros provoca la eliminación selectiva del genoma en cruces entre especies (Bennett, M. D. et al., *Chromosoma* 54: 175-200 (1976); Finch, R. A., *Chromosoma* 88: 386-393 (1983); Laurie, D. A. y Bennett, M. D., *Genome* 32: 953-961 (1989); Heppich, S. et al., *Theor Appl Genet* 61: 101-104 (1982); Jin, W. et al., *Plant Cell* 16: 571-81 (2004)), pero no se sabía cómo podrían manipularse los centrómeros para conseguir esto. Se estableció una base práctica para la ingeniería de eliminación del genoma mediante la alteración de CENH3, una proteína esencial para la función del centrómero en todos los eucariotas. El hecho de que los haploides se produjesen con transgenes tanto *GFP-tailswap* como *GFP-CENH3* sugiere que múltiples alteraciones diferentes de la proteína pueden inducir la eliminación del genoma en otras plantas. Las plantas de *A. thaliana* que coexpresan proteínas de tipo silvestre y proteínas *GFP-tailswap* o *GFP-CENH3* no actúan como inductores haploides. Por tanto, el método de los presentes inventores actualmente se basa en reemplazar la CENH3 nativa con una variante alterada. Podría usarse una mutación *cenh3* o un método de silenciamiento génico tal como el ARN de interferencia para reducir o eliminar la función endógena de *CENH3* en una especie nueva.

Se han descrito estirpes inductoras de haploides en las gramíneas (Coe, E. H., *American Naturalist* 93: 381-382 (1959); Hagberg, A. y Hagberg, G., *Hereditas* 93: 341-343 (1980); Kermicle, J. L., *Science* 166: 1422-1424 (1969)), pero se desconoce su base genética, a excepción del *gametofito indeterminado (ig)* del maíz (Evans, M. M., *Plant Cell* 19: 46-62 (2007)). El efecto de *ig* puede limitarse al maíz, porque las mutaciones en el ortólogo de *ig* de *A. thaliana* AS2 no fenocopia su efecto (Ori, N. et al., *Development* 127: 5523-32 (2000)). El proceso de los presentes inventores tiene ventajas clave sobre los métodos actuales para producir plantas haploides. 1) No se necesita cultivo de tejidos, eliminando una fuente importante de dependencia del genotipo. 2) El mismo inductor produce haploides maternos y paternos. 3) Cruce de un mutante *cenh3* mientras la hembra transfiere el genoma nuclear del progenitor masculino a un citoplasma heterólogo. Esto podría acelerar la producción de estirpes estériles masculinas citoplasmáticas para producir semillas híbridas. 4) Se produce eliminación del genoma entre progenitores que son isogénicos, excepto por las alteraciones de CENH3, evitando barreras de fertilidad inherentes a cruces amplios.

La eliminación del genoma inducida por cambios en CENH3 probablemente se produce durante las primeras mitosis cigóticas, cuando los centrómeros de los dos progenitores se cargan con diferentes poblaciones de proteínas CENH3. La expresión de genes *CENH3* tanto de tipo silvestre como mutantes en ciclos celulares posteriores debería igualar rápidamente la cantidad de las dos proteínas en los centrómeros individuales. La mitosis cigótica es normal en plantas *GFP-tailswap* y *GFP-CENH3*, porque no se observaron haploides de plantas autofertilizadas. Además, las plantas *GFP-CENH3* tienen un fenotipo de tipo totalmente silvestre. Las diferencias sutiles en la unión al ADN del centrómero, el ensamblaje del cinetocoro o el acoplamiento a los microtúbulos del huso pueden ser suficientes para ralentizar la segregación de los cromosomas que contienen CENH3 alterada, dando como resultado la eliminación del genoma. Los puntos de control del ciclo celular en las plantas deben estar suficientemente relajados para permitir que los cromosomas mutantes y de tipo silvestre se segreguen de manera diferencial y, presumiblemente, para permitir la citocinesis sin segregación cromosómica completa. El mecanismo preciso de eliminación del genoma en los experimentos de los presentes inventores sigue siendo desconocido.

Las secuencias de ADN del centrómero y la proteína CENH3 evolucionan rápidamente y se han propuesto diferencias de centrómero para crear barreras de especies (Henikoff, S. et al., *Science* 293: 1098-102 (2001)). Aunque los experimentos de los presentes inventores usaron proteínas marcadoras, indican que los cambios en CENH3 pueden inducir una pérdida cromosómica específica en un cigoto híbrido.

Resumen de métodos

Materiales vegetales.

cenh3-1 es una transición de G a A en el nucleótido 161 con respecto a ATG = +1 y muta un aceptor de corte y empalme conservado en el segundo intrón. Los transgenes *GFP-CENH3* y *GFP-tailswap* contenían un GFP N-terminal y usaban el promotor y terminador de *CENH3* endógena. La ubicación del transgén *GFP-tailswap* se determinó mediante TAIL-PCR, permitiendo a los presentes inventores determinar si el transgén era homocigótico o hemicigótico. La estirpe estéril masculina C24/Ler fue un regalo del Dr. Luca Comai (Universidad de California, Davis). La esterilidad masculina fue conferida por el transgén A9-barnasa. Las plantas se cultivaron con 16 horas de luz/8 horas de oscuridad a 20 grados C.

Preparación de ADN genómico y genotipado.

La preparación del ADN genómico y el genotipado por PCR se realizaron usando métodos convencionales.

Análisis citogenético.

Para analizar la progresión meiótica y determinar la ploidía, se prepararon extensiones de cromosomas mitóticos y meióticos a partir de anteras de acuerdo con los protocolos publicados.

Materiales vegetales

Se aisló *cenh3-1* mediante el procedimiento TILLING (Comai, L. y Henikoff, S., *Plant J* 45: 684-94 (2006)). La población TILLING se creó mutagenizando *Arabidopsis thaliana* en la accesión Col-0 con etilmetanosulfonato, usando protocolos convencionales. Se aisló *Cenh3-1* mediante TILLING usando el ensayo de escisión del heterodúplex CEL1, con cebadores de PCR específicos para el gen *CENH3/HTR12*.

Se predice que *cenh3-1* altera el corte y empalme normal de *CENH3*, porque muta un sitio aceptor de corte y empalme conservado al comienzo del segundo exón codificante. Se predice que la traducción de un ARNm que contiene el primer exón codificante cortado y empalmado a una ubicación incorrecta dentro de *CENH3* producirá solo 18 aminoácidos correctos. Como el dominio de plegamiento de histona de *CENH3* comienza en el resto de aminoácido 82, se cree que *cenh3-1* es un alelo nulo (esto es respaldado por su fenotipo embriotóxico letal).

La clonación de los transgenes *GFP-CENH3* y *GFP-tailswap* y la construcción de las estirpes *cenh3-1 GFP-CENH3* y *cenh3-1 GFP-tailswap* complementadas se describe en otra parte (Ravi, Comai, Sundaresan, Chan et al, manuscrito en preparación). Las secuencias de cebadores y los detalles completos están disponibles bajo petición.

Para cruzar el tipo silvestre como la hembra con *GFP-tailswap* como el macho, se usó un microscopio de disección para observar directamente la deposición de polen en el estigma (*GFP-tailswap* es principalmente estéril masculino). La cantidad de polen viable en flores individuales de *GFP-tailswap* varía. Se seleccionaron flores que mostraban claramente mayores cantidades de polen y se polinizaron con más de 60 anteras (10 flores *GFP-tailswap*) por estigma de tipo silvestre para conseguir el conjunto de semillas indicado en la Tabla 1. Usando un optivisor (lente de aumento) y aproximadamente 12 anteras (2 flores *GFP-tailswap*) por estigma de tipo silvestre, se obtuvo un conjunto de semillas mucho menor por silicua.

El porcentaje de semillas normales se determinó mediante inspección visual usando un microscopio de disección.

Las semillas de cruces *GFP-tailswap* x tipo silvestre se sembraron en placas 1 x MS que contenían sacarosa al 1 % para maximizar la eficiencia de germinación, particularmente de semillas que tenían un aspecto anormal. Las semillas de germinación tardía eran frecuentemente haploides.

El mutante *quartet* utilizado fue *qrt1-2* (Francis, K. E. et al., *Plant Physiol* 142: 1004-13 (2006)).

La esterilidad masculina en la estirpe C24/Ler fue conferida por el transgén A9-barnasa (Bushell, C. et al., *Plant Cell* 15: 1430-42 (2003); Paul, W. et al., *Plant Mol Biol* 19: 611-22 (1992)).

En el experimento *GFP-tailswap* x Wa-1, la progenie del cruce *GFP-tailswap* x Wa-1 que contenía solo cromosomas Wa-1 se confirmó como diploide usando extensiones de cromosomas. Las plantas que eran heterocigóticas para algunos cromosomas (marcadores Col-0 y Wa-1) y homocigóticas para otros cromosomas (marcadores Wa-1 solamente) se puntuaron como aneuploides. No se descubrieron descendientes triploides (heterocigóticos para marcadores en todos los cromosomas). Un subconjunto de plantas se cariotipó adicionalmente por medio de extensiones de cromosomas para confirmar la aneuploidía.

Análisis citogenético.

Las extensiones de cromosomas mitóticos y meióticos de anteras se prepararon de acuerdo con protocolos publicados (Ross, K. J. et al., *Chromosome Res* 4: 507-16 (1996)).

Tratamiento con colchicina

El tratamiento con colchicina de plantas haploides en desarrollo usó un protocolo publicado anteriormente con modificaciones menores (Josefsson, C. et al., *Curr Biol* 16: 1322-8 (2006)). Se preparó una solución de colchicina al 0,25 %, Silwet al 0,2 % y se colocó una gota de 20 µl en el meristemo antes de la floración. Las plantas se enfermaron transitoriamente después del tratamiento con colchicina. Tras la recuperación, aparecieron inflorescencias fértiles de meristemas secundarios que indicaron una duplicación exitosa de los cromosomas. Las plantas haploides también pueden tratarse después de la floración, aunque la tasa de éxito es considerablemente menor.

Ejemplo 2

***GFP-tailswap* de maíz crea haploides F1 en un cruce con el tipo silvestre**

Se creó una quimera en la que la cola de CENH3 de *A. thaliana* de CENH3 se reemplaza con el dominio de cola de CENH3 (SEQ ID NO: 102) de maíz (*Zea mays*), generando de este modo una fusión de la cola de CENH3 de maíz y el dominio de plegamiento de histona de CENH3 de *A. thaliana* y se transformó la fusión en heterocigotos *cenh3-1*. Como cabía esperar, esta proteína *GFP-tailswap* de maíz se dirigió a cinetocoros y rescató el fenotipo embriotóxico letal de *cenh3-1*. Las plantas complementadas eran más estériles que las plantas complementadas con *GFP-tailswap*, pero tenían una fertilidad limitada cuando se usaban como la hembra. Cuando las hembras de *cenh3-1 GFP-tailswap* de maíz se cruzaron con machos de tipo silvestre, se descubrieron 2 haploides, 3 diploides y 5 aneuploides entre un total de 10 progenies F1.

***mCherry-tailswap* crea haploides F1 en un cruce de tipo silvestre**

Se creó un transgén en el que el marcador GFP en *GFP-tailswap* se reemplazó con un marcador mCherry N-terminal (mCherry (SEQ ID NO: 105) es una versión monomérica de la proteína roja fluorescente DsRed). Del extremo N al extremo C, esta proteína contiene mCherry, el dominio de cola de H3.3 de *Arabidopsis thaliana* y el dominio de plegamiento de histona de CENH3 de *Arabidopsis thaliana*. Los transgenes *mCherry-tailswap* se transformaron en heterocigotos *cenh3-1*. Cuando plantas *cenh3-1 mCherry-tailswap* complementadas se cruzaron como hembras con machos de tipo silvestre, se observaron 1 haploide, 6 aneuploides y 4 diploides de 11 progenies F1.

La construcción mCherry-tailswap se preparó como un vector CP 169 pCAMBIA 1300 con un promotor HTR. El inserto incluía un sitio Mlu I seguido del mcherry Sal I Xba I N terminal seguido del terminador HTR12. El fragmento Tailswap de H3 se sintetizó solapando PCR y se digirió con Sall + Xba I y se clonó en CP169 para preparar la construcción mcherrytailswap.

Un transgén *tailswap* sin marcador GFP complementa una mutación *cenh3-1*. Las plantas complementadas crean haploides F1 en un cruce de tipo silvestre

Se creó un transgén en el que se retiró el marcador GFP en *GFP-tailswap*. Del extremo N al extremo C, esta proteína contiene el dominio de cola de H3.3 de *Arabidopsis thaliana* y el dominio de plegamiento de histona de CENH3 de *Arabidopsis thaliana*. Los transgenes *tailswap* se transformaron en heterocigotos *cenh3-1*. Cuando se cruzaron plantas *cenh3-1 tailswap* complementadas como hembras con machos de tipo silvestre, se observaron 4 haploides, 27 aneuploides y 67 diploides de 95 progenies F1.

La coexpresión de diferentes variantes de CENH3 crea propiedades deseables en una cepa de eliminación del genoma.

La planta de *GFP-tailswap* descrita anteriormente (plantas mutantes de *cenh3-1* rescatadas por un transgén *GFP-tailswap*) es un inductor haploide muy eficiente, pero es difícil de cruzar como donante de polen porque es principalmente estéril masculina. *GFP-CENH3* (plantas mutantes de *cenh3-1* rescatadas por un transgén *GFP-CENH3*) es un inductor haploide más débil pero es mucho más fértil. Se descubrió que la coexpresión de *GFP-CENH3* y *GFP-tailswap* en plantas *cenh3-1* produciría un polen más viable que *GFP-tailswap*, pero aún induciría la eliminación del genoma cuando estas plantas se cruzaran con diploides o tetraploides de tipo silvestre. De hecho, las plantas *cenh3-1* que portaban transgenes *GFP-CENH3* y *GFP-tailswap* (*GEM*; Eliminación del Genoma provocada por una Mezcla de variantes de *cenh3*, por sus siglas en inglés) produjeron abundante polen para cruces, aunque la viabilidad del polen era aún inferior a la de tipo silvestre.

El cruce de hembras *GEM* con machos de tipo silvestre produjo 2 haploides F1 entre 50 progenies. Cuando las hembras de tipo silvestre se cruzaron con machos *GEM*, se encontró un haploide entre 104 progenies.

Las plantas *GEM* son una mejora importante sobre *GFP-tailswap* o *GFP-CENH3* cuando el progenitor de tipo silvestre es un tetraploide que tiene gametos diploides. Cuando las plantas *GEM* se cruzaron como machos o hembras con tipo silvestre tetraploide, los cromosomas del progenitor *GEM* se eliminaron en un subconjunto de la progenie F1 (Tabla 3). *GEM* es fértil como macho o hembra y muestra una eliminación eficiente del genoma cuando se cruza con un

progenitor tetraploide con gametos diploides.

Tabla 2. Los cruces entre *GEM* y plantas de tipo silvestre diploides producen la eliminación de genoma.

cruce (♀×♂)	Planta total analizada	Triploide	Aneuploide	Plantas diploides uniparentales*
4n de tipo silvestre x <i>GEM</i>	85	53	27	5
<i>GEM</i> x 4n de tipo silvestre	84	12	57	15

5 **Tabla 3. Los cruces entre *GEM* y plantas de tipo silvestre tetraploides producen eliminación del genoma.**

cruce (♀×♂)	Planta total analizada	Diploide	Aneuploide	Plantas haploides uniparentales*
2n de tipo silvestre x <i>GEM</i>	104	62	18	1
<i>GEM</i> x 2n de tipo silvestre	50	36	12	2

Métodos de construcción de *GFP-tailswap* de maíz

- 10 El transgén de CENH3 tailswap de maíz se construyó fusionando en fase de lectura la cola N-terminal de CENH3 de maíz (correspondiente a 1-61aa) y dominio de plegamiento de histona de CENH3 de *Arabidopsis* (correspondiente a 82-179 aa) mediante PCR de superposición. El dominio de cola N terminal de maíz (206 pb) se amplificó a partir de ADNc de maíz usando las combinaciones de cebadores CP 384 (5'-NNNN^{gtcgac}ATGGCTCGAACCAAGCACCA-3' (SEQ ID NO: 110), el sitio *Sall* está en cursiva) y CP 572 (5'-CAACGGTTCCTGGCCTCCAGCGGTGGC-3' (SEQ ID NO: 111)). El HFD de *Arabidopsis* (950 pb) se amplificó a partir de ADN genómico usando combinaciones de
- 15 cebadores CP 571 (5'-GCCACCGCTGGAGGCCAGGAACCGTTG-3' (SEQ ID NO: 112)) y CP 375 (5'-NNNN tctagaTCACCATGGTCTGCCTTTTCCTCC-3' (SEQ ID NO: 113), el sitio *XbaI* está en cursiva). Los fragmentos resultantes se purificaron en gel y se usaron como molde para fusionarlos en una PCR de superposición usando combinaciones de cebadores CP 384 y CP 375. El fragmento resultante de 1,15 kb se clona como un fragmento *Sall-XbaI* en un vector binario CP 93 (derivado de pCambia 1300). El vector CP 93 contiene la secuencia codificante GFP
- 20 corriente arriba en fase de lectura con el sitio *Sall-XbaI* y su expresión es controlada por las secuencias reguladoras 5' y 3' del gen de CENH3 de *Arabidopsis*.

REIVINDICACIONES

1. Una planta que comprende una o dos copias de un alelo de un gen de CENH3 endógena mutado, en donde el gen de CENH3 endógena mutado se muta para reducir la función CENH3.
- 5 2. La planta de la reivindicación 1, en donde el gen de CENH3 endógena mutado codifica un polipéptido CENH3 mutante que es idéntico al polipéptido CENH3 de tipo silvestre salvo por 1 o 2 diferencias de aminoácidos.

Figura 1

H3.3 de <i>A. thaliana</i>	MAR TKDSARKSHGG	KAPTKD	20
H3.3 humana	MAR TKDTARKSTGG	KAPRKD	20
<i>C. albicans</i>	MARISGQSSGROTGGGTSAAETROQREELRRQRELRLQOQDQAEFRODQDQDYRTEDSPLY		60
Humana	MGP RRRSRKPLAP	RRRSPTSPT	21
<i>A. thaliana</i>	MARTKHRTVSOPRNOT	DAAGASSQA	27
Chopo	MARTKHIVARKKARSPK	RSD-ASPSTP	26
Arroz	MARTKHIPAVRKSKALPK	KKLOFERS	25
* .			
H3.3 de <i>A. thaliana</i>	LATKAARKSAPTTG	GVVK	38
H3.3 humana	LATKAARKSAPSTG	GVVK	38
<i>C. albicans</i>	EAATSSSRYSQFGTYRNQPGDWDHASSLPRRTTETRP	EVNRIVP	106
Humana	ETPGPSRRRPSLG	ASSH	38
<i>A. thaliana</i>	AGPTTTPTRGGEGGDNTQDINPTTSPATGTRRGAK		63
Chopo	RPTSSRTIRPOANG QD	GSSTQR	48
Arroz	PRPSKAQRAGGGTGTSAITRSAAGTSASGTRRQOI		60
: .			
H3.3 de <i>A. thaliana</i>	PHRFRTGTVALREIRKYQKSTELLNRKLPFQRLVREIAQDFK	TD	82
H3.3 humana	PHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFK	TD	82
<i>C. albicans</i>	R YKKRYRPGTKALREIRQYQKSTDLLIRKLPFARLVREISLDFVGPSYG		155
Humana	DHSRRROGWLKEIRKLQKSTHLLIRKLPFSRLAREICVKFTRG	VD	83
<i>A. thaliana</i>	RSRQAMPRCQKSYRYRRTGTVALKEIRHFQKQINLLIPAASFIREVRSITHMLAPP	QI	122
Chopo	QRKHREIRSGTVALREIRQYQKTWRPLIPAASFIRCVRMITQEFSSR	EV	96
Arroz	K ORKPHIRRTGTVALREIRKFQKTELLIPFAPFSRLVREITDFEYSK	DV	109
: * * : * * : * * * * *			
H3.3 de <i>A. thaliana</i>	LRFQSHAVLALQEAEEAYLVGLFEDTNLCATHAKRVTIMPKDVQLARRIRAERA		136
H3.3 humana	LRFQSAATGALQEASEAYLVGLFEDTNLCATHAKRVTIMPKDIQLARRIRGERA		136
<i>C. albicans</i>	LRWQSNATLALQEASESFLIHLLEDTNLCATHAKRVTIMQKDQLARRIRGQSWIL		211
Humana	ENWQAQALLALQEAEEAFLVHLFEDAYLLTLHAGRVTLFPKDVQLARRIRGLEEGLG		140
<i>A. thaliana</i>	NRWTAELVALQEAAEDYLVGLFSDSMLCATHARRVTLMRKDFELARRLGKGRPW		178
Chopo	NRWTAELVALQEAAEDFLVHLFEDGMLCATHAKRVTLMKKDFELARRLGKGRPW		152
Arroz	SRWTLEALLALQEAEEYHVDIFEVSNLCATHAKRVTIMQKDMQLARRIGGR	RPW	164
: . * : * : * * : * : : * : * * * : * : * * * : .			