

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 774 722**

51 Int. Cl.:

A01H 1/08 (2006.01)

C07K 14/415 (2006.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.10.2010 PCT/US2010/051483**

87 Fecha y número de publicación internacional: **14.04.2011 WO11044132**

96 Fecha de presentación y número de la solicitud europea: **05.10.2010 E 10822533 (5)**

97 Fecha y número de publicación de la concesión europea: **08.01.2020 EP 2486135**

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

30 Prioridad:

06.10.2009 US 248996 P

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

22.07.2020

73 Titular/es:

**THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (100.0%)
1111 Franklin Street, 5th Floor
Oakland, CA 94607, US**

72 Inventor/es:

**CHAN, SIMON y
MARUTHACHALAM, RAVI**

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 774 722 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.

25

Breve resumen de la invención

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

En algunas realizaciones, la invención 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 que comprende una secuencia de aminoácidos heteróloga de al menos 5 aminoácidos unidos a un extremo amino terminal de una proteína que comprende un dominio de plegamiento de histona CENH3, en donde la secuencia de aminoácidos heteróloga es heteróloga con respecto al dominio de plegamiento de histona CENH3 y en donde todos los alelos de la secuencia codificante genómica de CENH3 endógena de la planta están inactivados o desactivados.

En algunas realizaciones, 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.

40

En algunas realizaciones, 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 algunas realizaciones, la secuencia de aminoácidos heteróloga se une al dominio de plegamiento de histona CENH3 a través de una secuencia de proteína interpuesta. En algunas realizaciones, la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 no CENH3. En algunas realizaciones, 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.

45

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

50

En algunas realizaciones, la secuencia de aminoácidos heteróloga tiene al menos 10 aminoácidos de longitud. En algunas realizaciones, la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 y un dominio de cola de histona CENH3 heterólogo. En algunas realizaciones, la secuencia de aminoácidos heteróloga comprende proteína verde fluorescente. En algunas realizaciones, la secuencia de aminoácidos heteróloga altera los centrómeros. En algunas realizaciones, el dominio de plegamiento de histona CENH3 se selecciona entre el grupo que consiste en las SEQ ID NO: 49-94.

55

60

En algunas realizaciones, el polipéptido comprende un dominio de cola no de CENH3 unido a un dominio de plegamiento de histona CENH3.

En algunas realizaciones, 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 algunas realizaciones, el dominio de cola de CENH3 truncado carece de tres o más

65

aminoácidos amino terminales del dominio de cola endógeno. En algunas realizaciones, una secuencia de aminoácidos heteróloga se une al extremo amino del dominio de cola truncado.

5 La presente invención 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 de al menos cinco aminoácidos unido a un extremo amino de un dominio de plegamiento de histona CENH3.

10 La presente invención también proporciona una planta que comprende una o dos copias de un alelo de un gen mutado de CENH3 endógena, en donde la CENH3 endógena mutada se muta para reducir la función CENH3 y en donde dicha planta no se obtiene exclusivamente mediante un proceso esencialmente biológico.

15 La presente invención también proporciona un método para generar una planta de progenie F1 que tiene la mitad de la ploidía de una planta parental que expresa una proteína CENH3 endógena, comprendiendo el método cruzar la planta parental con la planta del párrafo inmediatamente anterior y seleccionar la progenie F1 generada a partir de la etapa de cruce que tiene la mitad de la ploidía de la planta parental.

En algunas realizaciones, la planta que expresa una proteína CENH3 endógena es el progenitor de polen del cruce.

20 En algunas realizaciones, la planta que expresa una proteína CENH3 endógena es el progenitor de óvulo del cruce.

En algunas realizaciones, el método comprende adicionalmente convertir al menos una planta haploide seleccionada en una planta doblemente haploide.

25 La invención también proporciona un método para hacer una planta transgénica como se ha definido anteriormente, comprendiendo el método transformar células vegetales con un ácido nucleico que comprende el casete de expresión; y seleccionar transformantes que comprenden el ácido nucleico, haciendo de este modo la planta transgénica.

30 En algunas realizaciones, la presente invención proporciona un polinucleótido aislado que codifica un polipéptido, en donde el polipéptido comprende:

una secuencia de aminoácidos de al menos 5 aminoácidos unidos a un extremo amino de 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 y en donde la secuencia de aminoácidos heteróloga se une al dominio de plegamiento de histona CENH3 a través de una secuencia de proteína interpuesta, en donde la 35 secuencia de proteína interpuesta comprende un dominio de cola de histona H3 no CENH3 o un dominio de cola de CENH3 heteróloga con respecto al dominio de plegamiento de histona CENH3. En algunas realizaciones, 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 algunas realizaciones, la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 y un dominio de cola de histona CENH3 40 heterólogo.

En algunas realizaciones, la secuencia de aminoácidos heteróloga tiene al menos 10, 15, 20, 30 o 50 aminoácidos de longitud, careciendo opcionalmente de una estructura secundaria fija.

45 En algunas realizaciones, la secuencia de aminoácidos heteróloga comprende proteína verde fluorescente.

En algunas realizaciones, la secuencia de aminoácidos heteróloga altera los centrómeros.

50 En algunas realizaciones, el dominio de plegamiento de histona CENH3 se selecciona entre el grupo que consiste en las SEQ ID NO: 49-94.

55 En algunas realizaciones, 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.

60 En algunas realizaciones, 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 algunas realizaciones, una secuencia de aminoácidos heteróloga se une al extremo amino del dominio de cola truncado. En algunas realizaciones, la secuencia de aminoácidos heteróloga tiene al menos 10, 15, 20, 30 o 50 aminoácidos de longitud. En algunas realizaciones, la secuencia de aminoácidos heteróloga comprende proteína verde fluorescente. En algunas realizaciones, la secuencia de aminoácidos heteróloga altera los centrómeros.

65 En algunas realizaciones, el dominio de plegamiento de histona CENH3 se selecciona entre el grupo que consiste en las SEQ ID NO: 49-94.

La presente invención también proporciona plantas que comprenden 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 el presente documento también se desvela un vector que comprende el casete de expresión.

- 5 En algunas realizaciones, el dominio de cola de histona heterólogo comprende un dominio de cola de histona H3 o un dominio de cola de histona CENH3 heterólogo.

En algunas realizaciones, el polipéptido comprende un dominio de cola de histona H3 y un dominio de cola de histona CENH3.

10

En algunas realizaciones, el casete de expresión se integra en el cromosoma de la planta.

La presente invención también proporciona un método para generar una planta haploide. En algunas realizaciones, el método comprende cruzar una planta que expresa una proteína CENH3 endógena con la planta como se describe en el presente documento (por ejemplo, que expresa una proteína tailswap); y seleccionar la progenie haploide F1 generada a partir de la etapa de cruce.

15

En algunas realizaciones, la planta que expresa una proteína CENH3 endógena es el progenitor de polen del cruce.

20

En algunas realizaciones, la planta que expresa una proteína CENH3 endógena es el progenitor de óvulo del cruce.

En algunas realizaciones, el método comprende adicionalmente convertir al menos una planta haploide seleccionada en una planta doblemente haploide.

25

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

Definiciones

30

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.

35

40

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.

45

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.

50

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, tejido terrestre y similares) y células (por ejemplo, células protectoras, óvulos, tricomas y similares) y progenie de la misma. La clase de plantas que puede usarse en el método de la invención 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 hemigóticas.

55

60

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.

65

"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) de 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, el posicionamiento del marco de lectura y similares. 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 algunas realizaciones, 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, en el presente documento también se desvelan plantas, hongos o animales (o células de los mismos) que expresan proteínas del cinetocoro CENPC, MCM21, MIS12, NDC80 o NUF2 mutadas recombinantes que alteran el centrómero. En el presente documento también se desvelan 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 invención proporciona métodos para generar una planta haploide cruzando una planta que expresa una proteína CENH3 mutada y que no expresa una proteína CENH3 endógena, con una planta que expresa una proteína CENH3 endógena.

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). 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.

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*, 18^a ed., Mack Publishing Co.; Hardman, J. G., Limbird, L. E. y Gilman, A. G., eds. (2001) *The Pharmacological Basis of Therapeutics*, 10^a 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*, 2^a edición, Vol. I-III, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al., eds. (1999) *Short Protocols in Molecular Biology*, 4^a 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)*, 2^a ed., Springer Verlag.

i. Dominio de plegamiento de histona CENH3

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 algunas realizaciones, 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.

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 algunas realizaciones, 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 otras realizaciones, las proteínas tailswap de la invención pueden comprender más o menos de la secuencia de CENH3. Por ejemplo, en algunas realizaciones, 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 algunas realizaciones, 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

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

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 algunas realizaciones, 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 algunas realizaciones, 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 algunas realizaciones, 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 algunas realizaciones, 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.

- 5 En algunas realizaciones, 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 algunas realizaciones, 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 invención. En algunas realizaciones, la secuencia heteróloga se une directamente al dominio de plegamiento de histona CENH3. En algunas realizaciones, la secuencia heteróloga se une como una secuencia de aminoácidos interpuesta al dominio de plegamiento de histona CENH3. En algunas realizaciones, la secuencia de aminoácidos interpuesta es un dominio de cola de CENH3 intacto o truncado. En algunas realizaciones, 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 algunas realizaciones, 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 algunas realizaciones, 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 algunas realizaciones, 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 algunas realizaciones, 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 algunas realizaciones, 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 algunas realizaciones, 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.

Como se ha indicado anteriormente, en algunas realizaciones, 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 (estas realizaciones 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 algunas realizaciones, la porción de cola heteróloga de la proteína 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.

En algunas realizaciones, las proteínas CENH3 mutadas de la invención 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 algunas realizaciones, 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.

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 algunas realizaciones, 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 algunas realizaciones, 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 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 algunas realizaciones, la secuencia heteróloga comprende una secuencia de aminoácidos de al menos 5, 10, 15, 20, 25, 30, 50 o más kD.

En algunas realizaciones, 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.

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 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 que se desvelan en el presente documento incluyen, por ejemplo, CENPC, MCM21, MIS12, NDC80 y NUF2.

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 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 el presente documento se desvelan polipéptidos CENPC, MCM21, MIS12, NDC80 o NUF2 mutados activos 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.

En la presente 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, la proteína de tipo silvestre endógena de la planta puede ser idéntica o sustancialmente idéntica a cualquiera de las SEQ ID NO: 96, 97, 98, 99 o 100 y la proteína del complejo del cinetocoro no de CENH3 mutada puede diferir 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, la proteína del complejo del cinetocoro no de CENH3 mutada puede comprender 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. La secuencia heteróloga puede unirse 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. La secuencia heteróloga puede comprender una secuencia de aminoácidos de al menos 5, 10, 15, 20, 25, 30, 50 o más aminoácidos.

En la presente divulgación, la proteína del complejo del cinetocoro no de CENH3 mutada puede comprender 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 DE LA INVENCION

La presente invención proporciona plantas 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 y que expresa una proteína CENH3 mutada correspondiente. 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. Sin pretender limitar el alcance de la invención, se cree que hay varias formas de hacer una planta 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 algunas realizaciones, puede generarse una mutación de CENH3 en un gen de CENH3 endógeno que reduzca o elimine la actividad o expresión de CENH3, o puede generarse un gen inactivado del gen de CENH3. En estas realizaciones, puede generarse un organismo heterocigótico para la inactivación o mutación del gen y puede introducirse un casete de expresión para la expresión de la CENH3 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 CENH3 mutada heteróloga expresada de forma recombinante. En consecuencia, la invención proporciona plantas en las 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 CENH3 mutada como se describe en el presente documento (por ejemplo, una proteína tailswap). En las plantas que tienen más de un conjunto diploide de cromosomas (por ejemplo, tetraploides), todos los alelos pueden inactivarse, mutar o desactivarse.

- 5 Como alternativa, puede introducirse el casete de expresión que codifica una proteína CENH3 mutada (por ejemplo, incluyendo, pero sin limitación, una proteína tailswap) en una planta con alelos CENH3 intactos y después puede silenciarse la CENH3 endógena 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 CENH3.
- 10 Idealmente, el ARNip silenciador u otro agente silenciador se selecciona para silenciar la proteína CENH3 endógena pero no interfiere sustancialmente con la expresión de la proteína CENH3 mutada (por ejemplo, una proteína tailswap). 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. 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
- 15 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

- 20 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
- 25 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.

- 30 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 CENH3, o una porción del ARNm de CENH3 (incluyendo, pero sin limitación, porciones no traducidas del ARNm) puede ser útil para producir una planta en la que la expresión de proteína del
- 35 complejo del cinetocoro se suprime. Los vectores que se desvelan en el presente documento se diseñan opcionalmente de manera que el efecto inhibitor se aplique solo a una proteína CENH3 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
- 40 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

45 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 algunas realizaciones, deben usarse nucleótidos de longitud aproximadamente completa, aunque puede usarse una secuencia de al menos aproximadamente 20, 50, 100, 200 o 500 nucleótidos.

- 50 Las moléculas de ARN catalítico o ribozimas también pueden usarse para inhibir la expresión de un gen CENH3. 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
- 55 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).

60

- 65 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.

5
10
15
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 CENH3 diana, pero una identidad más alta puede ejercer una represión más eficaz de la expresión de las secuencias endógenas. En algunas realizaciones, 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.

20
25
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 algunas realizaciones, 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.

30
35
40
45
50
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 proteína CenH3 usando ARNi, un ARN bicatenario que tiene la secuencia de un ARNm que codifica la proteína CENH3, 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 ARNi 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.

55
60
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 algunas realizaciones, 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.

65
Los vectores de expresión que expresan continuamente ARNi 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 ARNi 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).

5 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).

10 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). 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.

20 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). 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), por ejemplo, por fenotipo o mediante técnicas moleculares.

25 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)).

30 Por ejemplo, las cadenas pueden variar con respecto a la secuencia de origen natural en el nivel de estructura primaria en sustituciones, adiciones, supresiones de aminoácidos y similares. Estas modificaciones pueden usarse en varias combinaciones para producir la cadena de proteína modificada final.

35 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*) *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)).

40 En la aplicación de la tecnología de recombinación homóloga a los genes, 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 Vaultont 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

55 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.

65 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 invención 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.

- 5 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.

10 El vector que comprende las secuencias (por ejemplo, promotores o regiones codificantes) de genes 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.

15 *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 invención, 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.

40 *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.

45 *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 algunas realizaciones, 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.

55 **VI. PRODUCCIÓN DE PLANTAS TRANSGÉNICAS O CÉLULAS VEGETALES**

Pueden introducirse construcciones de ADN de la invención 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.

65 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).

5 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).

10 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).

15 Los ácidos nucleicos y los polipéptidos codificados de la invención 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
25 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*.

30 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) ya sea como un progenitor de polen o de óvulos con una planta que
35 expresa una proteína del complejo del cinetocoro endógena (por ejemplo, proteína CENH3) 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 invención 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
40 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 invención es que permite la generación de una planta que tiene solo los cromosomas nucleares del progenitor masculino y un citoplasma del
45 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
50 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 invención proporciona la expresión de dos o más proteínas del complejo del cinetocoro mutadas diferentes en una planta (por ejemplo, una
55 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
60 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
65 doblemente haploides homocigóticas a partir de células haploides poniendo en contacto las células haploides, incluyendo, pero sin limitación, el callo 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 invención es aplicable a 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 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

5 YLVGLFSDSMLCAIHARRVTLMRKDFELARRLGGKGRPW (SEQ ID NO:109).

10 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.

15 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 Ler TS	30 ± 4 (n = 22)	23	127	32* (25)	32 (25)	63 (50)
<i>GFP-tailswap</i> x Ws-0 TS	23 ± 5 (n = 14)	8	22	10* (45)	7 (32)	5 (28)
<i>GFP-tailswap</i> x C24/Ler TS	28 ± 5 (n = 13)	30	117	34† (29)	39 (33)	44 (38)
C24/Ler 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 Wa-1 (tetraploide)	21 ± 6 (n = 96)	1,8	41	11 (27)	0	30 (73)

20 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 meristemos 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	

(continuación)

Planta	Col-0	Ler	Ws-0
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.

15 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 hemocigótico. 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)).

5 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.

15 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

20 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 meristemos 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

30 ***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.

40 ***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.

50 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*.

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

60 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.

65 **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

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

25 Métodos de construcción de *GFP-tailswap* de maíz

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'-NNNNgtcgacATGGCTCGAACCAAGCACCA-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 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* 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*.

40 LISTADO DE SECUENCIAS

<110> Chan, Simon Maruthachalam, Ravi Los Regentes de la Universidad de California

<120> Generación de plantas haploides y fitomejoramiento mejorado

<130> 023070-200610PC

<140> WO Todavía sin asignar

<141> Todavía sin asignar

<150> US 61/248.996

<151> 10/06/2009

<160> 114

<170> FastSEQ para Windows versión 4.0

<210> 1

ES 2 774 722 T3

<211> 136
 <212> PRT
 <213> Arabidopsis thaliana

5 <220>
 <223> histona 3.3 (H3.3) de arabidopsis, 1 similar a histona 3
 <400> 1

```

Met Ala Arg Thr Lys Gln Ser Ala Arg Lys Ser His Gly Gly Lys Ala
 1          5          10          15
Pro Thr Lys Gln Leu Ala Thr Lys Ala Ala Arg Lys Ser Ala Pro Thr
      20          25          30
Thr Gly Gly Val Lys Lys Pro His Arg Phe Arg Pro Gly Thr Val Ala
      35          40          45
Leu Arg Glu Ile Arg Lys Tyr Gln Lys Ser Thr Glu Leu Leu Asn Arg
 50          55          60
Lys Leu Pro Phe Gln Arg Leu Val Arg Glu Ile Ala Gln Asp Phe Lys
 65          70          75          80
Thr Asp Leu Arg Phe Gln Ser His Ala Val Leu Ala Leu Gln Glu Ala
      85          90          95
Ala Glu Ala Tyr Leu Val Gly Leu Phe Glu Asp Thr Asn Leu Cys Ala
      100         105         110
Ile His Ala Lys Arg Val Thr Ile Met Pro Lys Asp Val Gln Leu Ala
      115         120         125
Arg Arg Ile Arg Ala Glu Arg Ala
 130          135
    
```

10
 <210> 2
 <211> 136
 <212> PRT
 15 <213> *Homo sapiens*

<220>
 <223> histona humana 3.3 (H3.3), histona H3, familia 3A (H3F3A), histona H3, familia 3B (H3F3B)

20 <400> 2

```

Met Ala Arg Thr Lys Gln Thr Ala Arg Lys Ser Thr Gly Gly Lys Ala
 1          5          10          15
Pro Arg Lys Gln Leu Ala Thr Lys Ala Ala Arg Lys Ser Ala Pro Ser
      20          25          30
Thr Gly Gly Val Lys Lys Pro His Arg Tyr Arg Pro Gly Thr Val Ala
      35          40          45
Leu Arg Glu Ile Arg Arg Tyr Gln Lys Ser Thr Glu Leu Leu Ile Arg
 50          55          60
Lys Leu Pro Phe Gln Arg Leu Val Arg Glu Ile Ala Gln Asp Phe Lys
 65          70          75          80
Thr Asp Leu Arg Phe Gln Ser Ala Ala Ile Gly Ala Leu Gln Glu Ala
      85          90          95
Ser Glu Ala Tyr Leu Val Gly Leu Phe Glu Asp Thr Asn Leu Cys Ala
      100         105         110
Ile His Ala Lys Arg Val Thr Ile Met Pro Lys Asp Ile Gln Leu Ala
      115         120         125
Arg Arg Ile Arg Gly Glu Arg Ala
 130          135
    
```

25 <210> 3
 <211> 125

ES 2 774 722 T3

<212> PRT
<213> *Physcomitrella patens*

5 <220>
<223> histona H3 de musgo

<400> 3

```

Met Ala Arg Arg Lys Thr Thr Pro Val His Gly Asn His Arg Ala Ser
 1          5          10          15
Thr Ser Ser Val Gly Gly Ala Ala Val Arg Pro Arg Lys Pro His Arg
          20          25          30
Trp Arg Pro Gly Thr Lys Ala Leu Gln Glu Ile Arg His Tyr Gln Lys
          35          40          45
Thr Cys Asp Leu Leu Ile Pro Arg Leu Pro Phe Ala Arg Tyr Val Lys
          50          55          60
Glu Ile Thr Met Met Tyr Ala Ser Asp Val Ser Arg Trp Thr Ala Glu
65          70          75          80
Ala Leu Thr Ala Leu Gln Glu Ala Thr Glu Asp Tyr Met Cys His Leu
          85          90          95
Phe Glu Asp Thr Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile
          100          105          110
Met Pro Lys Asp Leu Gln Leu Ala Arg Arg Leu Arg Gly
          115          120          125

```

10 <210> 4
<211> 164
<212> PRT
<213> *Pinus taeda*

15 <220>
<223> histona H3 de pino taeda

20 <400> 4

```

Met Val Arg Arg Lys Thr Val Pro Pro Arg Lys Lys Ser Gly Ser Gly
 1          5          10          15
Asn Ala Ala Ser Thr Ser Gly Val Gly Val Ser Thr Pro Gly Ser Ala
          20          25          30
Gly Glu Arg Gly Glu Arg Arg Gly Ser Ala Arg Leu Ala Ser Thr Pro
          35          40          45
Gly Ser Asp Ala Ser Pro Ser Ala Pro Ser Gly Arg Lys Pro His Arg
          50          55          60
Phe Arg Pro Gly Thr Val Ala Leu Arg Glu Ile Lys Arg Tyr Gln Lys
65          70          75          80
Ser Phe Glu Leu Leu Ile Pro Ser Leu Pro Phe Ala Arg Ile Val Arg
          85          90          95
Glu Leu Thr Met Tyr Tyr Ser Gln Val Val Ser Arg Trp Ala Ala Glu
          100          105          110

Ala Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Ile Val His Leu
          115          120          125
Phe Glu Asp Thr Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile
          130          135          140
Met Pro Arg Asp Leu Arg Leu Ala Arg Arg Leu Arg Gly Gly Gly Leu
145          150          155          160
Asp Arg Pro Trp

```

<210> 5
<211> 177

ES 2 774 722 T3

<212> PRT

<213> *Boecheera holboellii*

<220>

5 <223> CENH3 de berro de roca de Holboell

<400> 5

```

Met Ala Arg Thr Lys His Leu Ala Thr Arg Ser Arg Pro Arg Asn Gln
 1          5          10          15
Thr Asp Ala Thr Ala Ser Ser Ser Gln Ala Ala Gly Pro Ser Thr Asn
          20          25          30
Pro Thr Thr Arg Gly Ser Glu Gly Glu Asp Ala Ala Gln Glu Thr Thr
          35          40          45
Pro Thr Thr Ser Pro Ala Thr Gly Arg Lys Lys Gly Ala Lys Arg Ala
 50          55          60
Arg Tyr Ala Arg Pro Gln Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys
 65          70          75          80
Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Tyr Phe Gln Lys Ser Ile
          85          90          95
Asn Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Gln Val Arg Ser Ile
          100          105          110
Thr His Ala Leu Ala Pro Pro Gln Ile Thr Arg Trp Thr Ala Glu Ala
          115          120          125
Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe
 130          135          140
Ser Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met
 145          150          155          160
Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro
          165          170          175
Trp

```

10

<210> 6

<211> 177

<212> PRT

<213> *Boecheera stricta*

15

<220>

<223> CENH3 de berro de roca de Drummond

<400> 6

20

ES 2 774 722 T3

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

Asn Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Gln Val Arg Ser Ile
 100 105 110
 Thr His Ala Leu Ala Pro Pro Gln Ile Thr Arg Trp Thr Ala Glu Ala
 115 120 125
 Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe
 130 135 140
 Ser Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Ile Thr Leu Met
 145 150 155 160
 Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro
 165 170 175

Trp

<210> 7
 <211> 180
 <212> PRT
 <213> *Lepidum virginicum*

5

<220>
 <223> Proteina similar a histona H3 centromérica de mastuerzo de Virginia (CENH3)

10

<400> 7

Met Ala Arg Thr Lys Arg Tyr Ala Ser Arg Pro Gln Arg Pro Arg Asn
 1 5 10 15
 Gln Thr Asp Val Thr Val Pro Ser Ser Pro Ala Ala Gly Pro Ser Thr
 20 25 30
 Asn Pro Thr Arg Arg Asp Ser Glu Gly Glu Gly Gly Asp Asp Ala Gln
 35 40 45
 Gln Thr Val Pro Thr Thr Ser Pro Ala Ser Ile Ser Lys Lys Ala Ser
 50 55 60
 Lys Lys Asn Arg Lys Ala Thr Pro Gln Ser Ser Lys Lys Lys Thr Tyr
 65 70 75 80
 Arg Tyr Lys Pro Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln
 85 90 95
 Lys Ser Thr His Leu Leu Ile Pro Ala Ala Ala Phe Ile Arg Glu Val
 100 105 110
 Arg Cys Ile Thr Gln Ala Val Ala Pro Pro Gln Ile Ser Arg Trp Thr
 115 120 125
 Ala Glu Ala Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Val Val
 130 135 140
 Gly Leu Leu Ser Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val
 145 150 155 160
 Thr Leu Met Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys
 165 170 175
 Gly Arg Pro Trp
 180

ES 2 774 722 T3

<210> 8
 <211> 172
 <212> PRT
 <213> *Cardamine flexuosa*

5

<220>
 <223> proteína similar a histona H3 centromérica de berro amargo del bosque (CENH3)

<400> 8

10

Met Ala Arg Thr Lys His Phe Pro Asn Arg Thr Arg Pro Arg Asn Gln
 1 5 10 15
 Thr Asp Ala Thr Thr Pro Ala Ala Gly Pro Ser Thr Arg Thr Thr Arg
 20 25 30
 Ala Asn Gln Gly Glu Glu Thr Gln Gln Thr Asn Pro Thr Thr Ser Pro
 35 40 45

Ala Thr Ser Lys Lys Lys Gly Ala Lys Arg Thr Arg Arg Asp Met Pro
 50 55 60
 Gln Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro Gly Thr Val Ala
 65 70 75 80
 Leu Arg Glu Ile Arg His Phe Gln Lys Ser Thr Asn Leu Leu Ile Pro
 85 90 95
 Ala Ala Ser Phe Ile Arg Gln Val Arg Ser Ile Thr Gln Met Tyr Ala
 100 105 110
 Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu Ala Leu Val Ala Leu Gln
 115 120 125
 Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser Asp Ser Met Leu
 130 135 140
 Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Arg Lys Asp Phe Glu
 145 150 155 160
 Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
 165 170

15

<210> 9
 <211> 139
 <212> PRT
 <213> *Hordeum vulgare*

20

<220>
 <223> CENH3 de cebada

<400> 9

ES 2 774 722 T3

```

Met Ala Arg Thr Lys Lys Thr Val Ala Ala Lys Glu Lys Arg Pro Pro
1      5      10      15
Cys Ser Lys Ser Glu Pro Gln Ser Gln Pro Lys Lys Lys Glu Lys Arg
20      25      30
Ala Tyr Arg Phe Arg Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Lys
35      40      45
Tyr Arg Lys Ser Thr Asn Met Leu Ile Pro Phe Ala Pro Phe Val Arg
50      55      60
Leu Val Arg Asp Ile Ala Asp Asn Leu Thr Pro Leu Ser Asn Lys Lys
65      70      75      80
Glu Ser Lys Pro Thr Pro Trp Thr Pro Leu Ala Leu Leu Ser Leu Gln
85      90      95
Glu Ser Ala Glu Tyr His Leu Val Asp Leu Phe Gly Lys Ala Asn Leu
100     105     110
Cys Ala Ile His Ser His Arg Val Thr Ile Met Leu Lys Asp Met Gln
115     120     125
Leu Ala Arg Arg Ile Gly Thr Arg Ser Leu Trp
130     135

```

5 <210> 10
 <211> 178
 <212> PRT
 <213> Arabidopsis thaliana

10 <220>
 <223> proteína centromérica similar a histona H3 de arabidopsis (CENH3)

<400> 10

```

Met Ala Arg Thr Lys His Arg Val Thr Arg Ser Gln Pro Arg Asn Gln
1      5      10      15
Thr Asp Ala Ala Gly Ala Ser Ser Ser Gln Ala Ala Gly Pro Thr Thr
20      25      30
Thr Pro Thr Arg Arg Gly Gly Glu Gly Gly Asp Asn Thr Gln Gln Thr
35      40      45
Asn Pro Thr Thr Ser Pro Ala Thr Gly Thr Arg Arg Gly Ala Lys Arg
50      55      60

Ser Arg Gln Ala Met Pro Arg Gly Ser Gln Lys Lys Ser Tyr Arg Tyr
65      70      75      80
Arg Pro Gly Thr Val Ala Leu Lys Glu Ile Arg His Phe Gln Lys Gln
85      90      95
Thr Asn Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Glu Val Arg Ser
100     105     110
Ile Thr His Met Leu Ala Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu
115     120     125
Ala Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu
130     135     140
Phe Ser Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu
145     150     155     160
Met Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg
165     170     175
Pro Trp

```

15 <210> 11
 <211> 152
 <212> PRT
 <213> Populus trichocarpa

20 <220>

ES 2 774 722 T3

<223> histona H3 de álamo negro

<400> 11

```

Met Ala Arg Thr Lys His Pro Val Ala Arg Lys Arg Ala Arg Ser Pro
 1           5           10           15
Lys Arg Ser Asp Ala Ser Pro Ser Thr Pro Arg Thr Pro Thr Ser Ser
           20           25           30
Arg Thr Arg Pro Gln Ala Asn Gly Gln Gln Gly Ser Ser Thr Gln Arg
           35           40           45
Gln Arg Lys Lys His Arg Phe Arg Ser Gly Thr Val Ala Leu Arg Glu
 50           55           60
Ile Arg Gln Tyr Gln Lys Thr Trp Arg Pro Leu Ile Pro Ala Ala Ser
 65           70           75           80
Phe Ile Arg Cys Val Arg Met Ile Thr Gln Glu Phe Ser Arg Glu Val
           85           90           95
Asn Arg Trp Thr Ala Glu Ala Leu Val Ala Ile Gln Glu Ala Ala Glu
           100          105          110
Asp Phe Leu Val His Leu Phe Glu Asp Gly Met Leu Cys Ala Ile His
           115          120          125
Ala Lys Arg Val Thr Leu Met Lys Lys Asp Phe Glu Leu Ala Arg Arg
           130          135          140
Leu Gly Gly Lys Gly Arg Pro Trp
145           150

```

5

<210> 12

<211> 166

<212> PRT

10 <213> *Triticum aestivum*

<220>

<223> CENH3 de trigo

15 <400> 12

```

Met Ala Arg Thr Lys His Pro Ala Val Arg Lys Thr Lys Ala Leu Pro
 1           5           10           15
Lys Lys Gln Leu Gly Thr Arg Pro Ser Ala Gly Thr Pro Arg Arg Gln
           20           25           30
Glu Thr Asp Gly Ala Gly Thr Ser Ala Thr Pro Arg Arg Ala Gly Arg
           35           40           45
Ala Ala Ala Pro Gly Ala Ala Glu Gly Ala Thr Gly Gln Pro Lys Gln
 50           55           60

Arg Lys Pro His Arg Phe Arg Pro Gly Thr Val Ala Leu Arg Glu Ile
 65           70           75           80
Arg Lys Tyr Gln Lys Ser Val Asp Phe Leu Ile Pro Phe Ala Pro Phe
           85           90           95
Val Arg Leu Ile Lys Glu Val Thr Asp Phe Phe Cys Pro Glu Ile Ser
           100          105          110
Arg Trp Thr Pro Gln Ala Leu Val Ala Ile Gln Glu Ala Ala Glu Tyr
           115          120          125
His Leu Val Asp Val Phe Glu Arg Ala Asn His Cys Ala Ile His Ala
           130          135          140
Lys Arg Val Thr Val Met Gln Lys Asp Ile Gln Leu Ala Arg Arg Ile
 145           150          155          160
Gly Gly Arg Arg Leu Trp
           165

```

<210> 13

20 <211> 170

ES 2 774 722 T3

<212> PRT
 <213> *Oryza sativa*

<220>
 5 <223> histona 3 centromérica de arroz (CENH3)

<400> 13

```

Met Ala Arg Thr Lys His Pro Ala Val Arg Lys Ser Lys Ala Glu Pro
 1          5          10          15
Lys Lys Lys Leu Gln Phe Glu Arg Ser Pro Arg Pro Ser Lys Ala Gln
 20          25          30
Arg Ala Gly Gly Gly Thr Gly Thr Ser Ala Thr Thr Arg Ser Ala Ala
 35          40          45
Gly Thr Ser Ala Ser Gly Thr Pro Arg Gln Gln Thr Lys Gln Arg Lys
 50          55          60
Pro His Arg Phe Arg Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Lys
 65          70          75          80
Phe Gln Lys Thr Thr Glu Leu Leu Ile Pro Phe Ala Pro Phe Ser Arg
 85          90          95
Leu Val Arg Glu Ile Thr Asp Phe Tyr Ser Lys Asp Val Ser Arg Trp
100          105          110
Thr Leu Glu Ala Leu Leu Ala Leu Gln Glu Ala Ala Glu Tyr His Leu
115          120          125
Val Asp Ile Phe Glu Val Ser Asn Leu Cys Ala Ile His Ala Lys Arg
130          135          140
Val Thr Ile Met Gln Lys Asp Met Gln Leu Ala Arg Arg Ile Gly Gly
145          150          155          160
Arg Arg Pro Trp Asn Leu Asn Ser Leu Arg
165          170
    
```

10 <210> 14
 <211> 167
 <212> PRT
 <213> *Luzula nivea*

15 <220>
 <223> variante de histona H3 específica centromérica de luzula (CENH3)

20 <400> 14

ES 2 774 722 T3

```

Met Ala Arg Thr Lys His Phe Pro Gln Cys Ser Arg His Pro Lys Lys
 1          5          10          15
Gln Arg Thr Ala Ala Gly Glu Ala Gly Ser Ser Val Ile Ala Lys Gln
          20          25          30
Asn Ala Pro Ala Lys Thr Gly Asn Ala Ser Ser Ile Thr Asn Ser Thr
          35          40          45

Pro Ala Arg Ser Leu Lys Lys Asn Lys Ala Ser Lys Arg Gly Glu Lys
 50          55          60
Thr Gln Ala Lys Gln Arg Lys Met Tyr Arg Tyr Arg Pro Gly Thr Val
65          70          75          80
Ala Leu Arg Glu Ile Arg Lys Leu Gln Lys Thr Thr Asp Leu Leu Val
          85          90          95
Pro Lys Ala Ser Phe Ala Arg Leu Val Lys Glu Ile Thr Phe Gln Ser
          100          105          110
Ser Lys Glu Val Asn Arg Trp Gln Ala Glu Ala Leu Ile Ala Leu Gln
          115          120          125
Glu Ala Ser Glu Cys Phe Leu Val Asn Leu Leu Glu Ser Ala Asn Met
          130          135          140
Leu Ala Ile His Ala Arg Arg Val Thr Ile Met Lys Lys Asp Ile Gln
          145          150          155          160
Leu Ala Arg Arg Ile Gly Ala
          165

```

5 <210> 15
 <211> 176
 <212> PRT
 <213> *Arabidopsis arenosa*

10 <220>
 <223> CENH3 de berro de arena

<400> 15

```

Met Ala Arg Thr Lys His Phe Ala Thr Arg Thr Gly Ser Gly Asn Arg
 1          5          10          15
Thr Asp Ala Asn Ala Ser Ser Ser Gln Ala Ala Gly Pro Thr Thr Thr
          20          25          30
Pro Thr Thr Arg Gly Thr Glu Gly Gly Asp Asn Thr Gln Gln Thr Asn
          35          40          45
Pro Thr Thr Ser Pro Ala Thr Gly Gly Arg Arg Pro Arg Arg Ala Arg
          50          55          60
Gln Ala Met Pro Arg Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro
65          70          75          80
Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Gln Thr Asn
          85          90          95
Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Gln Val Arg Ser Ile Thr
          100          105          110
His Ala Leu Ala Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu Ala Leu
          115          120          125
Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Ile Gly Leu Phe Ser
          130          135          140
Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Arg
          145          150          155          160
Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
          165          170          175

```

15 <210> 16
 <211> 157
 <212> PRT
 <213> *Zea mays*

ES 2 774 722 T3

<220>
 <223> histona H3 centromérica de maíz (CENH3)

5 <400> 16

```

Met Ala Arg Thr Lys His Gln Ala Val Arg Lys Thr Ala Glu Lys Pro
 1      5      10      15
Lys Lys Lys Leu Gln Phe Glu Arg Ser Gly Gly Ala Ser Thr Ser Ala
      20      25      30
Thr Pro Glu Arg Ala Ala Gly Thr Gly Gly Arg Ala Ala Ser Gly Gly

          35          40          45
Asp Ser Val Lys Lys Thr Lys Pro Arg His Arg Trp Arg Pro Gly Thr
 50      55      60
Val Ala Leu Arg Glu Ile Arg Lys Tyr Gln Lys Ser Thr Glu Pro Leu
 65      70      75      80
Ile Pro Phe Ala Pro Phe Val Arg Val Val Arg Glu Leu Thr Asn Phe
      85      90      95
Val Thr Asn Gly Lys Val Glu Arg Tyr Thr Ala Glu Ala Leu Leu Ala
      100      105      110
Leu Gln Glu Ala Ala Glu Phe His Leu Ile Glu Leu Phe Glu Met Ala
      115      120      125
Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile Met Gln Lys Asp
      130      135      140
Ile Gln Leu Ala Arg Arg Ile Gly Gly Arg Arg Trp Ala
      145      150      155
    
```

10 <210> 17
 <211> 154
 <212> PRT
 <213> *Sorghum bicolor*

15 <220>
 <223> CENH3 de sorgo

20 <400> 17

```

Met Ala Arg Thr Lys His Gln Ala Val Arg Lys Leu Pro Gln Lys Pro
 1      5      10      15
Lys Lys Lys Leu Gln Phe Glu Arg Ala Gly Gly Ala Ser Thr Ser Ala
      20      25      30
Thr Pro Arg Arg Asn Ala Gly Thr Gly Gly Gly Ala Ala Arg Gly
      35      40      45
Glu Asp Leu Phe Lys Lys His Arg Trp Arg Ala Gly Thr Val Ala Leu
 50      55      60
Arg Glu Ile Arg Lys Tyr Gln Lys Ser Thr Glu Pro Leu Ile Pro Phe
 65      70      75      80
Ala Pro Phe Val Arg Val Val Lys Glu Leu Thr Ala Phe Ile Thr Asp
      85      90      95
Trp Arg Ile Gly Arg Tyr Thr Pro Glu Ala Leu Leu Ala Leu Gln Glu
      100      105      110
Ala Ala Glu Phe His Leu Ile Glu Leu Phe Glu Val Ala Asn Leu Cys
      115      120      125
Ala Ile His Ala Lys Arg Val Thr Val Met Gln Lys Asp Ile Gln Leu
      130      135      140
Ala Arg Arg Ile Gly Gly Arg Arg Trp Ser
      145      150
    
```

ES 2 774 722 T3

<210> 18
 <211> 150
 <212> PRT
 <213> *Cichorium intybus*

5

<220>
 <223> CENH3 de achicoria

<400> 18

10

```

Met Ala Arg Thr Lys Gln Pro Ala Lys Arg Ser Trp Gly Asn Arg Lys
 1          5          10          15
Ser Ser Gln Ser Arg Ala Ser Thr Ser Thr Ser Thr Ser Thr Pro Arg
          20          25          30
Lys Ser Pro Arg Lys Asp Pro Gly Arg Thr Gly Glu Arg Arg Gln Gln
          35          40          45
Lys Pro His Arg Phe Lys Pro Gly Ala Gln Ala Leu Arg Glu Ile Arg
 50          55          60

Arg Leu Gln Lys Thr Val Asn Leu Leu Ile Pro Ala Ala Pro Phe Ile
65          70          75          80
Arg Thr Val Lys Glu Ile Ser Asn Tyr Ile Ala Pro Glu Val Thr Arg
          85          90          95
Trp Gln Ala Glu Ala Ile Gln Ala Leu Gln Glu Ala Ala Glu Asp Tyr
          100         105         110
Leu Val Gln Leu Phe Glu Asp Ser Met Leu Cys Ser Ile His Ala Lys
          115         120         125
Arg Val Thr Leu Met Lys Lys Asp Trp Glu Leu Ala Arg Arg Leu Thr
          130         135         140
Lys Lys Gly Gln Pro Trp
          145         150
    
```

<210> 19
 <211> 153
 <212> PRT
 <213> *Cycas rumphii*

15

<220>
 <223> CENH3 de sagú reina

20

<400> 19

ES 2 774 722 T3

```

Met Ala Arg Lys Lys Ala Ser Thr Pro Arg Lys Lys Thr Gly Thr Ala
 1      5      10      15
Ala Ser Thr Ser Ala Val Glu Ser Pro Pro Ser Gly Val Asn Gln Thr
      20      25      30
Ala Arg Ala Arg Arg Ser Val Gly Gly Val Ala Pro Gly Ala Pro Arg
      35      40      45
Thr Pro Gln Ala Ser Thr Asn Val Gly Thr Pro Arg Arg Pro His Arg
      50      55      60
Phe Arg Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Arg Tyr Gln Lys
      65      70      75      80
Ser Phe Glu Leu Leu Ile Pro Ala Leu Pro Phe Ala Arg Asn Val Arg
      85      90      95
Glu Leu Thr Leu His His Ser Arg Glu Val His Arg Trp Thr Ala Glu
      100      105      110
Ala Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Ile Val His Leu
      115      120      125
Phe Glu Asp Thr Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile
      130      135      140
Met Pro Lys Asp Met His Leu Ala Arg
      145      150

```

5 <210> 20
 <211> 154
 <212> PRT
 <213> *Allium cepa*

10 <220>
 <223> CENH3 de cebolla

<400> 20

```

Met Ala Arg Thr Lys Gln Met Ala His Lys Lys Leu Arg Arg Lys Leu
 1      5      10      15
Asn Val Asp Glu Ala Gly Pro Ser Thr Pro Val Thr Arg Ser Thr Glu
      20      25      30
Val Asn Pro Lys Ser Ser Arg Pro Thr Pro Ile Thr Glu Asp Arg Gly
      35      40      45
Thr Gly Ala Arg Lys Lys His Arg Phe Arg Pro Gly Thr Val Ala Leu
      50      55      60
Arg Glu Ile Arg Lys Tyr Gln Lys Thr Ala Glu Leu Leu Ile Pro Ala
      65      70      75      80
Ala Pro Phe Ile Arg Leu Val Arg Glu Ile Thr Asn Leu Tyr Ser Lys
      85      90      95
Glu Val Thr Arg Trp Thr Pro Glu Ala Leu Leu Ala Ile Gln Glu Ala
      100      105      110
Ala Glu Phe Phe Ile Ile Asn Leu Leu Glu Glu Ala Asn Leu Cys Ala
      115      120      125
Ile His Ala Lys Arg Val Thr Leu Met Gln Lys Asp Ile Gln Leu Ala
      130      135      140
Arg Arg Ile Gly Gly Ala Arg His Phe Ser
      145      150

```

15 <210> 21
 <211> 199
 <212> PRT
 <213> *Malus domestica*

20 <220>
 <223> CENH3 de manzana

<400> 21

```

Met Ala Arg Ile Lys His Thr Ala His Lys Lys Ser Val Ala Arg Lys
 1      5      10      15
Ser Ser Thr Pro Lys Glu Ala Ala Ala Gly Thr Gly Gly Thr Ser Ala
 20      25      30
Ala Ser Pro Ala Lys Gln Pro Glu Pro Ser Ala Pro Trp Arg Arg Ser
 35      40      45
Glu Arg Ser Ser Gln Arg Thr Ser Glu Ser Gln Glu Gln Gln Glu Pro
 50      55      60
Glu Thr Asn Ala Gln Ala Thr Pro Gln Ser Lys Lys Gln Lys Gln Ser
 65      70      75      80
Glu Arg Asn Pro Gln Thr Pro Gln Ser Lys Lys Gln Lys Pro Ser Glu
 85      90      95
Arg Asn Pro Pro Pro Thr Gln Lys Lys Lys Trp Arg Tyr Arg Pro Gly
 100     105     110
Thr Val Ala Leu Arg Glu Ile Arg Tyr Tyr Gln Lys Thr Trp Asn Leu
 115     120     125
Ile Ile Pro Ala Ala Pro Phe Ile Arg Thr Val Arg Glu Ile Ser Ile
 130     135     140
Asn Met Ser Lys Asp Pro Val Arg Trp Thr Pro Glu Ala Leu Gln Ala
 145     150     155     160
Ile Gln Glu Ala Ala Glu Asp Phe Leu Val Arg Leu Phe Glu Asp Ser
 165     170     175
Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Lys Lys Asp
 180     185     190
Leu Glu Leu Ala Arg Arg Ile
 195

```

5

<210> 22
 <211> 150
 <212> PRT
 <213> *Lactuca sativa*

10

<220>
 <223> CENH3 de lechuga

15

<400> 22

ES 2 774 722 T3

Met Ala Arg Thr Lys Gln Pro Ala Lys Arg Ser Trp Gly Lys Arg Gln
 1 5 10 15
 Ser Ala Gly Ala Ser Thr Ser Thr Ser Thr Ser Thr Pro Arg Lys Ser
 20 25 30
 Pro Arg Lys Asp Pro Gly Ser Ser Gly Thr Gly Gln Arg Gln Lys Gln
 35 40 45
 Lys Pro His Arg Phe Lys Pro Gly Thr Gln Ala Leu Arg Glu Ile Arg
 50 55 60

Arg Leu Gln Lys Thr Val Asn Leu Leu Ile Pro Ala Ala Pro Phe Ile
 65 70 75 80
 Arg Thr Val Lys Glu Ile Ser Asn Tyr Ile Ala Pro Glu Val Thr Arg
 85 90 95
 Trp Gln Ala Glu Ala Leu Gln Ala Leu Gln Glu Ala Ala Glu Asp Tyr
 100 105 110
 Ile Val Gln Leu Phe Glu Asp Ser Met Leu Cys Ser Ile His Ala Lys
 115 120 125
 Arg Val Thr Leu Met Lys Lys Asp Met Glu Leu Ala Arg Arg Leu Thr
 130 135 140
 Lys Lys Gly Gln Pro Trp
 145 150

5 <210> 23
 <211> 145
 <212> PRT
 <213> *Carthamus tinctorius*

10 <220>
 <223> CENH3 de cártamo
 <400> 23

Met Ala Arg Thr Lys Gln Pro Ala Lys Arg Ser Ser Gly Lys Arg Asp
 1 5 10 15
 Ala Arg Pro Ser Thr Ser Thr Pro Thr Pro Arg Pro Ser Ala Arg Lys
 20 25 30
 Asn Pro Glu Ser Ser Gly Ala Gly Asp Gly Gln Arg Arg His Arg Tyr
 35 40 45
 Arg Pro Gly Thr Gln Ala Leu Arg Glu Ile Arg Arg Leu Gln Lys Thr
 50 55 60
 Val Asn Leu Leu Ile Pro Ala Ala Pro Phe Ile Arg Thr Val Lys Glu
 65 70 75 80
 Ile Ser Asn Tyr Ile Ala Pro Glu Val Thr Arg Trp Gln Ala Glu Ala
 85 90 95
 Leu Gln Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Ile Gln Leu Phe
 100 105 110
 Glu Asp Ser Met Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu Met
 115 120 125
 Lys Lys Asp Trp Glu Leu Ala Arg Arg Leu Gly Lys Lys Gly Gln Pro
 130 135 140
 Trp
 145

15 <210> 24
 <211> 142
 <212> PRT
 <213> *Helianthus exilis*

20 <220>
 <223> CENH3 de girasol serpentina

ES 2 774 722 T3

<400> 24

```

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

```

5

```

          85          90          95
Leu Gln Glu Ala Ala Glu Asp Tyr Leu Ile Gln Leu Phe Glu Asp Ser
          100          105          110
Met Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Lys Lys Asp
          115          120          125
Trp Glu Leu Ala Arg Arg Ile Gly Lys Lys Gly Gln Pro Trp
130          135          140

```

<210> 25

<211> 152

10 <212> PRT

<213> *Gossypium hirsutum*

<220>

<223> CENH3 de algodón americano

15

<400> 25

```

Met Ser Arg Thr Lys His Thr Ala Ala Lys Lys Pro Arg Arg Lys Pro
 1          5          10          15
Ser Ala Ala Ala Ala Ala Ser Pro Ala Thr Ala Ser Pro His Thr Arg
          20          25          30
Ser Val Thr Ala Lys Lys Thr Gly Gly Pro Ala Thr Pro Thr Pro Gly
          35          40          45
Lys Ser Lys Arg Pro His Arg Phe Arg Ala Gly Thr Arg Ala Leu Gln
 50          55          60
Glu Ile Arg Lys Tyr Gln Lys Thr Ser Asn Leu Leu Val Pro Ala Ala
65          70          75          80
Ser Phe Ile Arg Glu Val Arg Ala Ile Ser Tyr Arg Phe Ala Pro Asp
          85          90          95
Ile Asn Arg Trp Gln Ala Glu Ala Leu Val Ala Ile Gln Glu Ala Glu
100          105          110
Asp Tyr Leu Ile Gln Leu Phe Gly Asp Ala Met Leu Cys Ala Ile His
115          120          125
Ala Lys Arg Val Thr Leu Met Lys Lys Asp Ile Gln Leu Ala Arg Arg
130          135          140
Leu Gly Gly Met Gly Gln Pro Trp
145          150

```

20

<210> 26

<211> 155

<212> PRT

<213> *Glycine max*

25

<220>

ES 2 774 722 T3

<223> CENH3 de soja

<400> 26

```

Met Ala Arg Val Lys His Thr Pro Ala Ser Arg Lys Ser Ala Lys Lys
 1          5          10          15
Gln Ala Pro Arg Ala Ser Thr Ser Thr Gln Pro Pro Pro Gln Ser Gln
          20          25          30
Ser Pro Ala Thr Arg Glu Arg Arg Arg Ala Gln Gln Val Glu Pro Gln
          35          40          45
Gln Glu Pro Glu Ala Gln Gly Arg Lys Lys Arg Arg Asn Arg Ser Gly
          50          55          60
Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Arg Ser Cys Glu Leu
65          70          75          80
Leu Ile Pro Ala Ala Pro Phe Ile Arg Cys Val Lys Gln Ile Thr Asn
          85          90          95
Gln Phe Ser Thr Glu Val Ser Arg Trp Thr Pro Glu Ala Val Val Ala
          100          105          110
Leu Gln Glu Ala Ala Glu Glu Tyr Leu Val His Leu Phe Glu Asp Gly
          115          120          125

Met Leu Cys Ala Ile His Ala Arg Arg Ile Thr Leu Met Lys Lys Asp
130          135          140
Ile Glu Leu Ala Arg Arg Leu Gly Gly Ile Gly
145          150          155

```

5

<210> 27

<211> 153

<212> PRT

10 <213> *Cucumis melo*

<220>

<223> CENH3 de melón

15 <400> 27

```

Met Ala Arg Ala Arg His Pro Val Gln Arg Lys Ser Asn Arg Thr Ser
 1          5          10          15
Ser Gly Ser Gly Ala Ala Leu Ser Pro Pro Ala Val Pro Ser Thr Pro
          20          25          30
Leu Asn Gly Arg Thr Gln Asn Val Arg Lys Ala Gln Ser Pro Pro Ser
          35          40          45
Arg Thr Lys Lys Lys Ile Arg Phe Arg Pro Gly Thr Val Ala Leu Arg
          50          55          60
Glu Ile Arg Asn Leu Gln Lys Ser Trp Asn Leu Leu Ile Pro Ala Ser
65          70          75          80
Cys Phe Ile Arg Ala Val Lys Glu Val Ser Asn Gln Leu Ala Pro Gln
          85          90          95
Ile Thr Arg Trp Gln Ala Glu Ala Leu Val Ala Leu Gln Glu Ala Ala
          100          105          110
Glu Asp Phe Leu Val His Leu Phe Glu Asp Thr Met Leu Cys Ala Ile
          115          120          125
His Ala Lys Arg Val Thr Ile Met Lys Lys Asp Phe Glu Leu Ala Arg
          130          135          140
Arg Leu Gly Gly Lys Gly Arg Pro Trp
145          150

```

20 <210> 28

<211> 147

<212> PRT

<213> *Solanum chacoense*

ES 2 774 722 T3

<220>

<223> CENH3 de patata chaco

5 <400> 28

```

Met Ala Arg Thr Lys His Leu Ala Lys Arg Ser Arg Thr Lys Pro Ser
 1          5          10          15
Val Ala Ala Gly Pro Ser Ala Thr Pro Ser Thr Pro Thr Arg Lys Ser
          20          25          30
Pro Arg Ser Ala Pro Ala Thr Ser Val Pro Lys Pro Lys Gln Lys Lys
          35          40          45
Arg Tyr Arg Pro Gly Ser Val Ala Leu Arg Glu Ile Arg His Phe Gln
          50          55          60
Lys Thr Trp Asn Leu Val Ile Pro Ala Ala Pro Phe Ile Arg Leu Val
65          70          75          80
Arg Glu Ile Ser His Phe Phe Ala Pro Gly Val Thr Arg Trp Gln Ala
          85          90          95
Glu Ala Leu Ile Ala Ile Gln Glu Ala Ala Glu Asp Phe Leu Val His
          100          105          110
Leu Phe Glu Asp Ala Met Leu Cys Ala Ile His Ala Lys Arg Val Thr
          115          120          125
Leu Met Lys Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly
          130          135          140
Gln Pro Trp
    
```

145

<210> 29

10 <211> 144

<212> PRT

<213> *Solanum lycopersicum*

<220>

15 <223> CENH3 de tomate

<400> 29

```

Met Ala Arg Thr Lys His Leu Ala Lys Arg Ser Arg Thr Thr Ser Ala
 1          5          10          15
Ala Pro Ser Ala Thr Pro Ser Thr Pro Ser Arg Lys Ser Pro Arg Ser
          20          25          30
Ala Pro Ala Thr Ser Val Gln Lys Pro Lys Gln Lys Lys Arg Tyr Arg
          35          40          45
Pro Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Thr Trp
          50          55          60
Asp Leu Leu Ile Pro Ala Ala Pro Phe Ile Arg Leu Val Arg Glu Ile
65          70          75          80
Ser His Phe Tyr Ala Pro Gly Val Thr Arg Trp Gln Ala Glu Ala Leu
          85          90          95
Ile Ala Ile Gln Glu Ala Ala Glu Asp Phe Leu Val His Leu Phe Glu
          100          105          110
Asp Ala Met Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Lys
          115          120          125
Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Gln Pro Trp
          130          135          140
    
```

20

<210> 30

<211> 156

<212> PRT

<213> *Nicotiana tabacum*

ES 2 774 722 T3

<220>

<223> variante de histona H3 específica de centrómero de tabaco alotetraploide (CENH3-1)

5 <400> 30

```

Met Ala Arg Thr Lys His Leu Ala Leu Arg Lys Gln Ser Arg Pro Pro
 1          5          10          15
Ser Arg Pro Thr Ala Thr Arg Ser Ala Ala Ala Ala Ala Ser Ser Ala
          20          25          30
Pro Gln Ser Thr Pro Thr Arg Thr Ser Gln Arg Thr Ala Pro Ser Thr
          35          40          45
Pro Gly Arg Thr Gln Lys Lys Lys Thr Arg Tyr Arg Pro Gly Thr Val
 50          55          60
Ala Leu Arg Glu Ile Arg Arg Phe Gln Lys Thr Trp Asp Leu Leu Ile
 65          70          75          80
Pro Ala Ala Pro Phe Ile Arg Leu Val Lys Glu Ile Ser His Phe Phe
          85          90          95
Ala Pro Glu Val Thr Arg Trp Gln Ala Glu Ala Leu Ile Ala Leu Gln
          100          105          110
Glu Ala Ala Glu Asp Phe Leu Val His Leu Phe Asp Asp Ser Met Leu
          115          120          125
Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Lys Lys Asp Phe Glu
          130          135          140
Leu Ala Arg Arg Leu Gly Gly Lys Ala Arg Pro Trp
          145          150          155

```

<210> 31

10 <211> 157

<212> PRT

<213> *Nicotiana tabacum*

<220>

15 <223> variante de histona H3 específica de centrómero de tabaco alotetraploide (CENH3-2)

<400> 31

```

Met Ala Arg Thr Lys His Leu Ala Leu Arg Lys Gln Ser Arg Pro Pro
 1          5          10          15
Ser Arg Pro Thr Ala Thr Arg Ser Ala Ala Ala Ala Ala Ser Ser Ser
          20          25          30
Ala Pro Gln Ser Thr Pro Thr Arg Thr Ser Gln Arg Thr Ala Pro Ser
          35          40          45
Thr Pro Gly Arg Thr Gln Lys Lys Lys Thr Arg Tyr Arg Pro Gly Thr
 50          55          60
Val Ala Leu Arg Glu Ile Arg Arg Phe Gln Lys Thr Trp Asn Leu Leu
 65          70          75          80
Ile Pro Ala Ala Pro Phe Ile Arg Leu Val Lys Glu Ile Ser Tyr Phe
          85          90          95
Phe Ala Pro Glu Val Thr Arg Trp Gln Ala Glu Ala Leu Ile Ala Leu
          100          105          110
Gln Glu Ala Ala Glu Asp Phe Leu Val His Leu Phe Asp Asp Ser Met
          115          120          125
Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Lys Lys Asp Phe
          130          135          140
Glu Leu Ala Arg Arg Leu Gly Gly Lys Ala Arg Pro Trp
          145          150          155

```

20

<210> 32

<211> 156

<212> PRT

<213> *Nicotiana tomentosiformis*

ES 2 774 722 T3

<220>

<223> variante de histona H3 específica centromérica de tabaco diploide (CENH3)

5 <400> 32

```

Met Ala Arg Thr Lys His Leu Ala Leu Arg Lys Gln Ser Arg Pro Pro
 1      5      10      15
Ser Arg Pro Thr Ala Thr Arg Ser Ala Ala Ala Ala Ser Ser Ala
      20      25      30
Pro Gln Ser Thr Pro Thr Arg Thr Ser Gln Arg Thr Ala Pro Ser Thr
      35      40      45
Pro Gly Arg Thr Gln Lys Lys Lys Thr Arg Tyr Arg Pro Gly Thr Val
 50      55      60
Ala Leu Arg Glu Ile Arg Arg Phe Gln Lys Thr Trp Asp Leu Leu Ile
 65      70      75      80
Pro Ala Ala Pro Phe Ile Arg Leu Val Lys Glu Ile Ser His Phe Phe
      85      90      95
Ala Pro Glu Val Thr Arg Trp Gln Ala Glu Ala Leu Ile Ala Leu Gln
      100      105      110
Glu Ala Ala Glu Asp Phe Leu Val His Leu Phe Asp Asp Ser Met Leu
      115      120      125
Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Lys Lys Asp Phe Glu
 130      135      140
Leu Ala Arg Arg Leu Gly Gly Lys Ala Arg Pro Trp
 145      150      155

```

<210> 33

10 <211> 158

<212> PRT

<213> *Vitis vinifera*

<220>

15 <223> CENH3 de uva de vino europea

<400> 33

```

Met Thr Arg Thr Lys His Leu Ala Arg Lys Ser Arg Asn Arg Arg Arg
 1      5      10      15
Gln Phe Ala Ala Thr Pro Ala Ser Pro Ala Ser Ala Gly Pro Ser Ser
      20      25      30
Ala Pro Pro Arg Arg Pro Thr Arg Thr Ala Thr Asp Ala Ser Pro Ser
      35      40      45
Thr Ala Gly Ser Gln Gly Gln Arg Lys Pro Phe Arg Tyr Arg Pro Gly
 50      55      60
Thr Val Ala Leu Arg Glu Ile Arg Arg Phe Gln Lys Thr Thr His Leu
 65      70      75      80
Leu Ile Pro Ala Ala Pro Phe Ile Arg Thr Val Arg Glu Ile Ser Tyr
      85      90      95
Phe Phe Ala Pro Glu Ile Ser Arg Trp Thr Ala Glu Ala Leu Val Ala
      100      105      110
Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val His Leu Phe Glu Asp Ala
      115      120      125
Met Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Lys Lys Asp
 130      135      140
Trp Glu Leu Ala Arg Arg Ile Gly Gly Lys Gly Gln Pro Trp
 145      150      155

```

20

<210> 34

<211> 157

<212> PRT

<213> *Nicotiana glauca*

ES 2 774 722 T3

<220>

<223> variante de histona H3 específica centromérica de tabaco del bosque (CENH3)

5 <400> 34

```

Met Ala Arg Thr Lys His Leu Ala Leu Arg Lys Gln Ser Arg Pro Pro
 1           5           10           15
Ser Arg Pro Thr Ala Thr Arg Ser Ala Ala Ala Ala Ala Ser Ser Ser
           20           25           30
Ala Pro Gln Ser Thr Pro Thr Arg Thr Ser Gln Arg Thr Ala Pro Ser
           35           40           45
Thr Pro Gly Arg Thr Gln Lys Lys Lys Thr Arg Tyr Arg Pro Gly Thr
 50           55           60
Val Ala Leu Arg Glu Ile Arg Arg Phe Gln Lys Thr Trp Asn Leu Leu
65           70           75           80
Ile Pro Ala Ala Pro Phe Ile Arg Leu Val Lys Glu Ile Ser Tyr Phe
           85           90           95
Phe Ala Pro Glu Val Thr Arg Trp Gln Ala Glu Ala Leu Ile Ala Leu
           100          105          110
Gln Glu Ala Ala Glu Asp Phe Leu Val His Leu Phe Asp Asp Ser Met
           115          120          125
Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Lys Lys Asp Phe
130           135          140
Glu Leu Ala Arg Arg Leu Gly Gly Lys Ala Arg Pro Trp
145           150          155

```

<210> 35

10 <211> 177

<212> PRT

<213> *Crucihimalaya himalaica*

<220>

15 <223> histona centromérica de berro de roca del Himalaya (CENH3)

<400> 35

```

Met Ala Arg Thr Lys His Phe Ala Thr Arg Ser Arg Pro Arg Asn Gln
 1           5           10           15
Thr Asp Ala Thr Ala Ser Ala Ser Gln Ala Thr Gly Pro Ser Thr Asn
           20           25           30
Pro Thr Thr Arg Gly Ser Glu Gly Glu Asp Ala Ala Arg Gly Thr Asn
           35           40           45
Pro Thr Thr Ser Pro Ala Thr Gly Arg Lys Lys Gly Val Lys Arg Ala
 50           55           60
Arg His Ala Met Pro Gln Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys
65           70           75           80
Ala Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Asn Thr
           85           90           95
Asn Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Gln Val Lys Ser Ile
           100          105          110
Thr Tyr Ala Val Ala Pro Pro Gln Ile Thr Arg Trp Thr Ala Glu Ala
           115          120          125
Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe
130           135          140
Ser Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met
145           150          155
Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro
           165          170          175
Trp

```

20

<210> 36

ES 2 774 722 T3

<211> 176
 <212> PRT
 <213> *Arabidopsis lyrata*

5 <220>
 <223> histona centromérica de berro de roca de hoja de lira (CENH3)

<400> 36

```

Met Ala Arg Thr Lys His Phe Ala Thr Lys Ser Arg Ser Gly Asn Arg
 1          5          10          15
Thr Asp Ala Asn Ala Ser Ser Ser Gln Ala Ala Gly Pro Thr Thr Thr
          20          25          30
Pro Thr Thr Arg Gly Thr Glu Gly Gly Asp Asn Thr Gln Gln Thr Asn
          35          40          45
Pro Thr Thr Ser Pro Ala Thr Gly Gly Arg Arg Pro Arg Arg Ala Arg
 50          55          60
Gln Ala Met Pro Arg Val Ser Gln Asn Lys Pro Tyr Arg Tyr Lys Pro
 65          70          75          80
Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Gln Thr Asn
          85          90          95
Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Gln Val Arg Ser Ile Thr
          100          105          110
His Ala Leu Ala Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu Ala Leu
          115          120          125
Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser
          130          135          140
Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Arg
 145          150          155          160
Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
          165          170          175
    
```

10

<210> 37
 <211> 170
 <212> PRT
 <213> *Capsella bursapastoris*

15

<220>
 <223> histona centromérica de bolsa de pastor (CENH3)

20

<400> 37

ES 2 774 722 T3

```

Met Ala Arg Thr Lys His Phe Ala Thr Arg Ser Gly Pro Arg Thr Pro
 1          5          10          15
Ala Val Ala Ser Ser Ser Gln Ala Ala Val Pro Ser Ser Ser Pro Ala
          20          25          30
Thr Arg Gly Arg Val Gly Val Asp Ala Ala Ala Gln Gln Pro Thr Pro
          35          40          45
Ala Thr Ser Pro Ala Thr Ala Lys Lys Lys Gly Ala Lys Arg Ala Arg
          50          55          60
Phe Gly Arg Pro Gln Gly Ser Gln Lys Lys Lys Pro Tyr Arg Tyr Arg
65          70          75          80
Pro Gly Thr Val Ala Leu Arg Glu Ile Arg His Tyr Gln Lys Gly Thr
          85          90          95
Ser Leu Leu Ile Pro Ala Ala Ala Phe Ile Arg Gln Val Arg Ser Ile
          100          105          110
Thr Asn Ala Val Ala Pro Arg Glu Val Asn Arg Trp Thr Ala Glu Ala
          115          120          125
Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Phe Leu Val Gly Leu Phe
          130          135          140
Ser Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met
145          150          155          160
Arg Lys Asp Phe Asp Leu Ala Arg Arg Leu
          165          170

```

<210> 38

<211> 178

5 <212> PRT

<213> *Raphanus sativus*

<220>

<223> proteína similar a histona H3 centromérica de rábano (CENH3)

10

<400> 38

```

Met Ala Arg Thr Lys His Phe Ala Ser Arg Ala Arg Asp Arg Asn Gln
 1          5          10          15
Pro Asn Ala Ala Ala Ala Ala Ala Gly Pro Ser Ala Thr Pro Thr Arg
          20          25          30
Arg Gly Ser Ser Gln Gly Glu Glu Ala Gln Gln Thr Thr Pro Thr Thr
          35          40          45
Thr Ser Pro Ala Thr Thr Ala Ser Gly Arg Lys Lys Lys Thr Phe Arg Tyr
          50          55          60
Thr Thr Gln Ala Met Pro Lys Ser Ser Lys Lys Lys Thr Phe Arg Tyr
65          70          75          80
Lys Pro Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Ser
          85          90          95
Thr Lys Leu Leu Ile Pro Ser Ala Pro Phe Ile Arg Glu Val Arg Ser
          100          105          110
Ile Thr His Asn Leu Ala Ala Ala Tyr Val Thr Arg Trp Thr Ala Glu
          115          120          125
Ala Leu Ile Ala Leu Gln Glu Ala Ala Glu Asp Phe Leu Val Gly Leu
          130          135          140
Phe Ser Asp Ala Met Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu
145          150          155          160
Met Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg
          165          170
Pro Phe

```

<210> 39

<211> 164

<212> PRT

<213> *Eruca sativa*

15

ES 2 774 722 T3

<220>

<223> proteína similar a histona H3 centromérica de rúcula (CENH3)

5 <400> 39

```

Met Ala Arg Thr Lys His Phe Ala Ser Arg Ala Arg Asp Arg Asn Arg
 1      5      10      15
Asn Asn Ala Thr Ala Ser Ser Ser Ala Ala Ala Ala Ala Gly Pro
      20      25      30
Ser Ala Thr Pro Thr Arg Arg Gly Ser Arg Gln Gly Gly Gly Gly Gly
      35      40      45
Gly Gly Val Glu Ala Gln Gln Gly Ser Asn Lys Lys Lys Lys Ser Phe
      50      55      60
Arg Tyr Lys Pro Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln
65      70      75      80
Lys Thr Thr Lys Leu Leu Ile Pro Ala Ala Thr Phe Ile Arg Leu Val
      85      90      95
Arg Ser Ile Thr Leu Asp Arg Ala Lys Pro Gln Val Thr Arg Trp Thr
      100     105     110
Ala Glu Ala Leu Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val
      115     120     125
Gly Leu Phe Ser Asp Ser Met Leu Cys Ala Ile His Ala Lys Arg Val
      130     135     140
Thr Leu Met Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys
145     150     155     160
Gly Arg Pro Trp
    
```

<210> 40

10 <211> 173

<212> PRT

<213> *Olimarabidopsis pumila*

<220>

15 <223> proteína 1 similar a histona H3 centromérica de cohete enano (CENH3-1)

<400> 40

```

Met Ala Arg Thr Lys His Asn Ala Ile Arg Ser Arg Asp Arg Thr Gly
 1      5      10      15
Ala Thr Ala Ser Ser Gln Ala Ala Gly Pro Ser Thr Asn Pro Thr
      20      25      30
Ala Gly Gly Ser Glu Asp Ala Ala Gln Gln Thr Thr Pro Thr Thr Ser
      35      40      45
Pro Ala Thr Gly Ser Lys Lys Arg Ala Lys Arg Ala Arg Gln Ala Met
      50      55      60
Pro Arg Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro Gly Thr Val
65      70      75      80
Ala Leu Arg Glu Ile Arg His Phe Gln Lys Thr Thr Ser Leu Leu Leu
      85      90      95
Pro Ala Ala Pro Phe Ile Arg Gln Val Arg Ser Ile Ser Ser Ala Leu
      100     105     110
Ala Pro Arg Glu Ile Thr Arg Trp Thr Ala Glu Ala Leu Val Ala Leu
      115     120     125
Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser Asp Ser Met
      130     135     140
Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Arg Lys Asp Phe
145     150     155     160

      Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
              165              170
    
```

20

ES 2 774 722 T3

<210> 41
 <211> 174
 <212> PRT
 <213> *Olimarabidopsis pumila*

5

<220>
 <223> proteína 2 similar a histona H3 centromérica de cohete enano (CENH3-2)

<400> 41

10

```

Met Thr Arg Thr Lys His Thr Val Ile Lys Ser Ser Arg Pro Leu Asp
 1          5          10          15
Arg Thr Asp Ala Ser Ser Ser Gln Ala Ala Gly Pro Ser Thr Asn Pro
          20          25          30
Thr Ala Gly Ser Ser Gly Asp Ala Ala Gln Gln Thr Thr Pro Thr Thr
          35          40          45
Ser Pro Ala Thr Gly Ser Thr Lys Arg Ala Lys Arg Ala Arg Gln Ala
 50          55          60
Met Pro Arg Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro Gly Thr
 65          70          75          80
Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Thr Thr Ser Phe Leu
          85          90          95
Ile Pro Ala Ala Pro Phe Ile Arg Gln Val Arg Ser Ile Ser Ser Ala
          100          105          110
Leu Ala Pro Thr Gln Ile Thr Arg Trp Thr Ala Glu Ala Leu Val Ala
          115          120          125
Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser Asp Ser
 130          135          140
Met Leu Cys Ala Ile His Ala Lys Arg Val Thr Leu Met Arg Lys Asp
 145          150          155          160
Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
          165          170
  
```

<210> 42
 <211> 174
 <212> PRT
 <213> *Turritis glabra*

15

<220>
 <223> proteína similar a histona H3 de mostaza torre (CENH3)

20

<400> 42

ES 2 774 722 T3

```

Met Ala Arg Thr Lys His Phe Ala Thr Arg Ser Arg Pro Arg Asn Gln
 1      5      10      15
Thr Asp Ser Ser Gln Ala Ala Gly Pro Ser Thr Asn Pro Thr Thr
      20      25      30
Gly Gly Ser Glu Gly Gly Asp Ala Ala Gln Gln Thr Thr Pro Thr Thr
      35      40      45
Ser Pro Ala Thr Gly Arg Lys Lys Arg Ala Lys Arg Ala Lys Gln Ala
 50      55      60
Met Pro Gln Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro Gly Thr
65      70      75      80
Ile Ala Leu Arg Glu Ile Arg Tyr Phe Gln Lys Asn Thr Asn Leu Leu
      85      90      95
Ile Pro Ala Ala Ser Phe Ile Arg Glu Val Arg Ser Ile Thr His Ala
      100      105      110
Leu Ala Pro Pro Gln Ile Ser Arg Trp Thr Ala Glu Ala Leu Val Ala
      115      120      125
Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser Asp Ser
130      135      140

Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Arg Lys Asp
145      150      155      160
Phe Glu Leu Ala Arg Arg Ile Gly Gly Lys Gly Arg Pro Trp
      165      170

```

5 <210> 43
 <211> 176
 <212> PRT
 <213> *Arabidopsis halleri*

10 <220>
 <223> proteína 1 similar a histona H3 (CENH3-1)

<400> 43

```

Met Ala Arg Thr Lys His Phe Ala Ile Lys Ser Arg Ser Gly Asn Arg
 1      5      10      15
Thr Asp Ala Asn Ala Ser Ser Ser Gln Ala Ala Gly Pro Thr Thr Thr
      20      25      30
Pro Thr Thr Arg Gly Thr Glu Gly Gly Asp Asn Thr Gln Gln Thr Asn
      35      40      45
Pro Thr Thr Ser Pro Ala Thr Gly Gly Arg Arg Pro Arg Arg Ala Arg
 50      55      60
Gln Ala Met Pro Arg Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro
65      70      75      80
Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Gln Thr Asn
      85      90      95
Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Gln Val Arg Ser Ile Thr
      100      105      110
His Ala Leu Ala Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu Ala Leu
      115      120      125
Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser
130      135      140
Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Arg
145      150      155      160
Lys Asp Phe Glu Leu Thr Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
165      170

```

15 <210> 44
 <211> 176
 <212> PRT
 <213> *Arabidopsis halleri*

ES 2 774 722 T3

<220>

<223> proteína 2 similar a histona H3 (CENH3-2)

5 <400> 44

```

Met Ala Arg Thr Lys His Phe Val Thr Arg Lys Gly Ser Gly Asn Arg
 1      5      10      15
Thr Asp Phe Asp Ala Asn Ala Ser Ser Gln Ala Ala Gly Pro Thr
 20      25      30
Lys Thr Pro Thr Thr Arg Gly Thr Glu Gly Gly Asp Asn Thr Gln Gln
 35      40      45
Thr Thr Ser Pro Ala Thr Gly Gly Arg Arg Gly Pro Arg Arg Ala Arg
 50      55      60
Gln Ala Met Pro Arg Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro
 65      70      75      80
Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Gln Thr Asn
 85      90      95
Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Gln Val Arg Ser Ile Thr
 100     105     110
His Ala Leu Ala Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu Ala Leu
 115     120     125
Val Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser

```

```

          130          135          140
Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Arg
145          150          155          160
Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
          165          170          175

```

10

<210> 45

<211> 175

<212> PRT

<213> *Arabidopsis lyrata*

15

<220>

<223> histona H3 de berro de roca de hoja lira (HTR12A)

20

<400> 45

ES 2 774 722 T3

```

Met Ala Arg Thr Lys His Phe Ala Thr Arg Thr Gly Ser Gly Asn Arg
1      5      10      15
Thr Asp Ala Asn Ala Ser Ser Ser Ser Gln Ala Ala Gly Pro Thr Lys
      20      25      30
Thr Pro Thr Thr Arg Gly Thr Glu Gly Gly Asp Asn Thr Gln Gln Thr
      35      40      45
Thr Ser Pro Ala Thr Gly Gly Arg Arg Gly Pro Arg Arg Ala Arg Gln
      50      55      60
Ala Met Pro Arg Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro Gly
65      70      75      80
Thr Val Ala Leu Arg Glu Ile Arg His Phe Gln Lys Gln Thr Asn Leu
      85      90      95
Leu Ile Pro Ala Ala Ser Phe Ile Arg Gln Ala Arg Ser Ile Thr His
      100     105     110
Ala Leu Ala Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu Ala Leu Val
      115     120     125
Ala Leu Gln Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser Asp
      130     135     140
Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Arg Lys
145     150     155     160
Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
      165     170     175

```

<210> 46
 <211> 172
 <212> PRT
 <213> *Arabidopsis lyrata*

5

<220>
 <223> histona H3 de berro de roca de hoja lira (HTR12B)

10

<400> 46

```

Met Ala Arg Thr Lys His Phe Ala Thr Lys Ser Arg Thr Asp Ala Asn
1      5      10      15
Ala Ser Ser Ser Gln Ala Ala Gly Pro Thr Thr Thr Pro Thr Thr Arg
      20      25      30
Gly Thr Glu Gly Gly Asp Asn Thr Gln Gln Thr Asn Pro Thr Thr Ser
      35      40      45
Pro Ala Thr Gly Gly Arg Arg Pro Arg Arg Ala Arg Gln Ala Met Pro
      50      55      60
Arg Gly Ser Gln Lys Lys Pro Tyr Arg Tyr Lys Pro Gly Thr Val Ala
65      70      75      80
Leu Arg Glu Ile Arg His Phe Gln Lys Gln Thr Asn Leu Leu Ile Pro
      85      90      95
Ala Ala Ser Phe Ile Arg Gln Val Arg Ser Ile Thr His Ala Leu Ala
      100     105     110
Pro Pro Gln Ile Asn Arg Trp Thr Ala Glu Ala Leu Val Ala Leu Gln
      115     120     125

Glu Ala Ala Glu Asp Tyr Leu Val Gly Leu Phe Ser Asp Ser Met Leu
      130     135     140
Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Arg Lys Asp Phe Glu
145     150     155     160
Leu Ala Arg Arg Leu Gly Gly Lys Gly Arg Pro Trp
      165     170

```

15 <210> 47
 <211> 163
 <212> PRT
 <213> *Saccharum officinalis*

ES 2 774 722 T3

<220>
 <223> CENH3 de caña de azúcar

5 <400> 47

Met Ala Arg Thr Lys His Gln Ala Val Arg Arg Pro Thr Gln Lys Pro
 1 5 10 15
 Lys Lys Lys Leu Gln Phe Glu Arg Ala Gly Gly Ala Ser Thr Ser Ala
 20 25 30
 Thr Pro Glu Arg Asn Ala Gly Thr Gly Gly Ala Ala Ala Arg Val
 35 40 45
 Thr Arg Gly Arg Val Glu Lys Lys His Arg Trp Arg Val Gly Thr Val
 50 55 60
 Ala Leu Arg Glu Ile Arg Lys Tyr Gln Lys Ser Thr Glu Pro Leu Ile
 65 70 75 80
 Pro Phe Ala Pro Phe Val Arg Val Val Lys Glu Leu Thr Gly Phe Ile
 85 90 95
 Thr Asp Trp Arg Ile Gly Arg Tyr Thr Pro Glu Ala Leu Leu Ala Leu
 100 105 110
 Gln Glu Ala Ala Glu Phe His Leu Ile Glu Leu Phe Gln Val Ala Asn
 115 120 125
 Leu Cys Ala Ile His Ala Lys Arg Val Thr Val Met Gln Lys Asp Ile
 130 135 140
 Gln Leu Ala Arg Arg Ile Gly Gly Lys Arg Trp Ala Tyr Pro Phe Phe
 145 150 155 160
 Leu Pro Tyr

10 <210> 48
 <211> 181
 <212> PRT
 <213> *Brassica napa*

15 <220>
 <223> CENH3 de nabo

<400> 48

ES 2 774 722 T3

```

Met Ala Arg Thr Lys His Phe Ala Ser Arg Ala Arg Asp Arg Asn Pro
 1          5          10          15
Thr Asn Ala Thr Ala Ser Ser Ser Ala Ala Ala Ala Ala Gly Pro Ser
      20          25          30
Ala Thr Pro Thr Arg Arg Gly Gly Ser Gln Gly Gly Glu Ala Gln Gln
      35          40          45
Thr Thr Pro Pro Ala Thr Thr Thr Ala Gly Arg Lys Lys Gly Gly Thr
      50          55          60
Lys Arg Thr Lys Gln Ala Met Pro Lys Ser Ser Asn Lys Lys Lys Thr
 65          70          75          80
Phe Arg Tyr Lys Pro Gly Thr Val Ala Leu Arg Glu Ile Arg His Phe
      85          90          95
Gln Lys Thr Thr Lys Leu Leu Ile Pro Ala Ala Ser Phe Ile Arg Glu
      100          105          110
Val Arg Ser Val Thr Gln Ile Phe Ala Pro Pro Asp Val Thr Arg Trp
      115          120          125

Thr Ala Glu Ala Leu Met Ala Ile Gln Glu Ala Ala Glu Asp Phe Leu
 130          135          140
Val Gly Leu Phe Ser Asp Ala Met Leu Cys Ala Ile His Ala Arg Arg
 145          150          155          160
Val Thr Leu Met Arg Lys Asp Phe Glu Leu Ala Arg Arg Leu Gly Gly
      165          170          175
Lys Gly Arg Pro Leu
      180

```

5 <210> 49
 <211> 91
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> dominio de histona CENH3 de musgo *Physcomitrella patens*

<400> 49

```

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

```

15 <210> 50
 <211> 98
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> dominio de histona CENH3 de pino taeda *Pinus taeda*

<400> 50

ES 2 774 722 T3

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

<210> 51
 <211> 97
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de berro de roca de Holboell *Boecheera holboellii*

<400> 51

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

<210> 52
 <211> 97
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de berro de roca de Drummond *Boecheera stricta*

<400> 52

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

ES 2 774 722 T3

5 <210> 53
 <211> 97
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de mastuerzo de Virginia *Lepidum virginicum*

10 <400> 53

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

15 <210> 54
 <211> 97
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de berro amargo del bosque *Cardamine flexuosa*

20 <400> 54

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

25 <210> 55
 <211> 102
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de cebada *Hordeum vulgare*

30 <400> 55

35

ES 2 774 722 T3

Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Lys Tyr Arg Lys Ser Thr
 1 5 10 15
 Asn Met Leu Ile Pro Phe Ala Pro Phe Val Arg Leu Val Arg Asp Ile
 20 25 30
 Ala Asp Asn Leu Thr Pro Leu Ser Asn Lys Lys Glu Ser Lys Pro Thr
 35 40 45
 Pro Trp Thr Pro Leu Ala Leu Leu Ser Leu Gln Glu Ser Ala Glu Tyr
 50 55 60
 His Leu Val Asp Leu Phe Gly Lys Ala Asn Leu Cys Ala Ile His Ser
 65 70 75 80
 His Arg Val Thr Ile Met Leu Lys Asp Met Gln Leu Ala Arg Arg Ile
 85 90 95
 Gly Thr Arg Ser Leu Trp
 100

- <210> 56
- <211> 97
- 5 <212> PRT
- <213> Secuencia artificial
- <220>
- 10 <223> dominio de histona CENH3 de arabidopsis *Arabidopsis thaliana*
- <400> 56

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

- 15 <210> 57
- <211> 96
- <212> PRT
- 20 <213> Secuencia artificial
- <220>
- <223> dominio de histona CENH3 de álamo negro *Populus trichocarpa*
- 25 <400> 57

ES 2 774 722 T3

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

- <210> 58
- <211> 95
- 5 <212> PRT
- <213> Secuencia artificial
- <220>
- 10 <223> dominio de histona CENH3 de trigo *Triticum aestivum*
- <400> 58

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

- 15 <210> 59
- <211> 101
- <212> PRT
- <213> Secuencia artificial
- 20 <220>
- <223> dominio de histona CENH3 de arroz *Oryza sativa*
- <400> 59

Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Lys Phe Gln Lys Thr Thr
 1 5 10 15
 Glu Leu Leu Ile Pro Phe Ala Pro Phe Ser Arg Leu Val Arg Glu Ile
 20 25 30
 Thr Asp Phe Tyr Ser Lys Asp Val Ser Arg Trp Thr Leu Glu Ala Leu
 35 40 45
 Leu Ala Leu Gln Glu Ala Ala Glu Tyr His Leu Val Asp Ile Phe Glu
 50 55 60
 Val Ser Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile Met Gln
 65 70 75 80
 Lys Asp Met Gln Leu Ala Arg Arg Ile Gly Gly Arg Arg Pro Trp Asn
 85 90 95
 Leu Asn Ser Leu Arg
 100

25

ES 2 774 722 T3

<210> 60
 <211> 91
 <212> PRT
 <213> Secuencia artificial

5

<220>
 <223> dominio de histona CENH3 de luzula *Luzula nivea*

<400> 60

10

```

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

<210> 61
 <211> 97
 <212> PRT
 <213> Secuencia artificial

15

<220>
 <223> dominio de histona CENH3 de berro de arena *Arabidopsis arenosa*

20

<400> 61

```

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

25 <210> 62
 <211> 96
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> dominio de histona CENH3 de maíz *Zea mays*

<400> 62

ES 2 774 722 T3

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

<210> 63
 <211> 96
 5 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de sorgo *Sorghum bicolor*

<400> 63

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

15 <210> 64
 <211> 96
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> dominio de histona CENH3 de achicoria *Cichorium intybus*

<400> 64

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

25

ES 2 774 722 T3

<210> 65
 <211> 87
 <212> PRT
 <213> Secuencia artificial

5

<220>
 <223> dominio de histona CENH3 de sagú reina *Cycas rumphii*

<400> 65

10

```

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

<210> 66
 <211> 96
 <212> PRT
 <213> Secuencia artificial

15

<220>
 <223> dominio de histona CENH3 de cebolla *Allium cepa*

20

<400> 66

```

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

<210> 67
 <211> 89
 <212> PRT
 <213> Secuencia artificial

25

<220>
 <223> dominio de histona CENH3 de manzana *Malus domestica*

30

<400> 67

```

Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Tyr Tyr Gln Lys Thr Trp
 1          5          10          15
Asn Leu Ile Ile Pro Ala Ala Pro Phe Ile Arg Thr Val Arg Glu Ile
      20          25          30
    
```

35

ES 2 774 722 T3

Ser Ile Asn Met Ser Lys Asp Pro Val Arg Trp Thr Pro Glu Ala Leu
 35 40 45
 Gln Ala Ile Gln Glu Ala Ala Glu Asp Phe Leu Val Arg Leu Phe Glu
 50 55 60
 Asp Ser Met Leu Cys Ala Ile His Ala Arg Arg Val Thr Leu Met Lys
 65 70 75 80
 Lys Asp Leu Glu Leu Ala Arg Arg Ile
 85

5 <210> 68
 <211> 96
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> dominio de histona CENH3 de lechuga *Lactuca sativa*
 <400> 68

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

15 <210> 69
 <211> 96
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> dominio de histona CENH3 de cártamo *Carthamus tinctorius*
 <400> 69

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

25 <210> 70
 <211> 96
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> dominio de histona CENH3 de girasol serpentina *Helianthus exilis*

ES 2 774 722 T3

<400> 70

```

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

```

5

<210> 71
 <211> 95
 <212> PRT
 <213> Secuencia artificial

10

<220>
 <223> dominio de histona CENH3 de algodón americano *Gossypium hirsutum*

15

<400> 71

```

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

```

20

<210> 72
 <211> 93
 <212> PRT
 <213> Secuencia artificial

25

<220>
 <223> dominio de histona CENH3 de soja *Glycine max*
 <400> 72

ES 2 774 722 T3

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

<210> 73
 <211> 96
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de melón *Cucumis melo*

<400> 73

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

<210> 74
 <211> 96
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de patata chaco *Solanum chacoense*

<400> 74

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

ES 2 774 722 T3

<210> 75
 <211> 96
 <212> PRT
 <213> Secuencia artificial

5

<220>
 <223> dominio de histona CENH3 de tomate *Solanum lycopersicum*

<400> 75

10

Pro	Gly	Thr	Val	Ala	Leu	Arg	Glu	Ile	Arg	His	Phe	Gln	Lys	Thr	Trp
1				5					10					15	
Asp	Leu	Leu	Ile	Pro	Ala	Ala	Pro	Phe	Ile	Arg	Leu	Val	Arg	Glu	Ile
			20					25					30		
Ser	His	Phe	Tyr	Ala	Pro	Gly	Val	Thr	Arg	Trp	Gln	Ala	Glu	Ala	Leu
		35					40					45			
Ile	Ala	Ile	Gln	Glu	Ala	Ala	Glu	Asp	Phe	Leu	Val	His	Leu	Phe	Glu
	50					55					60				
Asp	Ala	Met	Leu	Cys	Ala	Ile	His	Ala	Lys	Arg	Val	Thr	Leu	Met	Lys
65					70					75					80
Lys	Asp	Phe	Glu	Leu	Ala	Arg	Arg	Leu	Gly	Gly	Lys	Gly	Gln	Pro	Trp

85

90

95

<210> 76
 <211> 96
 <212> PRT
 <213> Secuencia artificial

15

<220>
 <223> dominio de histona CENH3-1 de tabaco alotetraploide *Nicotiana tabacum*

20

<400> 76

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

25

<210> 77
 <211> 96
 <212> PRT
 <213> Secuencia artificial

30

<220>
 <223> dominio de histona CENH3-2 de tabaco alotetraploide *Nicotiana tabacum*

<400> 77

ES 2 774 722 T3

```

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

```

<210> 78

<211> 96

5 <212> PRT

<213> Secuencia artificial

<220>

10 <223> dominio de histona de CENH3 de tabaco diploide *Nicotiana tomentosiformis*

<400> 78

```

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

```

15 <210> 79

<211> 96

<212> PRT

<213> Secuencia artificial

20 <220>

<223> dominio de histona de CENH3 de uva de vino europea *Vitis vinifera*

<400> 79

```

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

```

25

ES 2 774 722 T3

<210> 80
 <211> 96
 <212> PRT
 <213> Secuencia artificial

5

<220>
 <223> dominio de histona de CENH3 de tabaco del bosque *Nicotiana sylvestris*

<400> 80

10

```

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

<210> 81
 <211> 97
 <212> PRT
 <213> Secuencia artificial

15

<220>
 <223> dominio de histona CENH3 de berro de roca del Himalaya *Crucihimalaya himalaica*

20

<400> 81

```

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

<210> 82
 <211> 97
 <212> PRT
 <213> Secuencia artificial

25

<220>
 <223> dominio de histona CENH3 de berro de roca de hoja de lira *Arabidopsis lyrata*

30

<400> 82

ES 2 774 722 T3

```

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

```

5 <210> 83
 <211> 90
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> dominio de histona CENH3 de bolsa de pastor *Capsella bursapastoris*

<400> 83

```

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

```

15 <210> 84
 <211> 97
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> dominio de histona CENH3 de rábano *Raphanus sativus*

<400> 84

```

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

```

25

ES 2 774 722 T3

<210> 85
 <211> 97
 <212> PRT
 5 <213> Secuencia artificial

<220>
 <223> dominio de histona CENH3 de rúcula *Eruca sativa*

10 <400> 85

```

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

15 <210> 86
 <211> 97
 <212> PRT
 <213> Secuencia artificial

<220>
 20 <223> dominio de histona CENH3-1 de cohete enano *Olimarabidopsis pumila*

<400> 86

```

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

25 <210> 87
 <211> 97
 <212> PRT
 <213> Secuencia artificial

<220>
 30 <223> dominio de histona CENH3-2 de cohete enano *Olimarabidopsis pumila*

<400> 87

35

ES 2 774 722 T3

```

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

```

5 <210> 88
 <211> 97
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> dominio de histona CENH3 de mostaza torre *Turritis glabra*
 <400> 88

```

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

```

15 <210> 89
 <211> 97
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> dominio de histona CENH3-1 de *Arabidopsis halleri*
 <400> 89

ES 2 774 722 T3

```

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

```

- 5 <210> 90
- <211> 97
- <212> PRT
- <213> Secuencia artificial
- 10 <220>
- <223> dominio de histona CENH3-2 de *Arabidopsis halleri*
- <400> 90

```

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

```

- 15 <210> 91
- <211> 97
- <212> PRT
- <213> Secuencia artificial
- 20 <220>
- <223> dominio de histona CENH3 HTR12A de *Arabidopsis lyrata*
- <400> 91

ES 2 774 722 T3

```

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

```

5 <210> 92
 <211> 97
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> dominio de histona CENH3 HTR12B de *Arabidopsis lyrata*

<400> 92

```

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

```

15 <210> 93
 <211> 103
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> dominio de histona CENH3 de caña de azúcar *Saccharum officinalis*

<400> 93

ES 2 774 722 T3

```

Val Gly Thr Val Ala Leu Arg Glu Ile Arg Lys Tyr Gln Lys Ser Thr
 1          5          10          15
Glu Pro Leu Ile Pro Phe Ala Pro Phe Val Arg Val Val Lys Glu Leu
 20          25          30
Thr Gly Phe Ile Thr Asp Trp Arg Ile Gly Arg Tyr Thr Pro Glu Ala
 35          40          45
Leu Leu Ala Leu Gln Glu Ala Ala Glu Phe His Leu Ile Glu Leu Phe
 50          55          60
Gln Val Ala Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Val Met
 65          70          75          80
Gln Lys Asp Ile Gln Leu Ala Arg Arg Ile Gly Gly Lys Arg Trp Ala
 85          90          95
Tyr Pro Phe Phe Leu Pro Tyr
          100

```

5 <210> 94
 <211> 97
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> dominio de histona CENH3 de nabo *Brassica napa*
 <400> 94

```

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

```

15 <210> 95
 <211> 43
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> dominio de cola vegetal conservado de proteína de histona H3, cola N-terminal de H3, cola H3
 <400> 95

```

Met Ala Arg Thr Lys Gln Ser Ala Arg Lys Ser His Gly Gly Lys Ala
 1          5          10          15
Pro Thr Lys Gln Leu Ala Thr Lys Ala Ala Arg Lys Ser Ala Pro Thr
 20          25          30
Thr Gly Gly Val Lys Lys Pro His Arg Phe Arg
 35          40

```

25 <210> 96
 <211> 705
 <212> PRT

ES 2 774 722 T3

<213> Arabidopsis thaliana

<220>

<223> proteína centromérica C de arabidopsis (CENPC)

5

<400> 96

```

Met Ala Asp Val Ser Arg Ser Ser Ser Leu Tyr Thr Glu Glu Asp Pro
 1          5          10          15
Leu Gln Ala Tyr Ser Gly Leu Ser Leu Phe Pro Arg Thr Leu Lys Ser
 20          25          30
Leu Ser Asn Pro Leu Pro Pro Ser Tyr Gln Ser Glu Asp Leu Gln Gln
 35          40          45
Thr His Thr Leu Leu Gln Ser Met Pro Phe Glu Ile Gln Ser Glu His
 50          55          60
Gln Glu Gln Ala Lys Ala Ile Leu Glu Asp Val Asp Val Asp Val Gln
 65          70          75          80
Leu Asn Pro Ile Pro Asn Lys Arg Glu Arg Arg Pro Gly Leu Asp Arg
 85          90          95
Lys Arg Lys Ser Phe Ser Leu His Leu Thr Thr Ser Gln Pro Pro Pro
 100         105         110
Val Ala Pro Ser Phe Asp Pro Ser Lys Tyr Pro Arg Ser Glu Asp Phe
 115         120         125
Phe Ala Ala Tyr Asp Lys Phe Glu Leu Ala Asn Arg Glu Trp Gln Lys
 130         135         140
Gln Thr Gly Ser Ser Val Ile Asp Ile Gln Glu Asn Pro Pro Ser Arg
 145         150         155         160
Arg Pro Arg Arg Pro Gly Ile Pro Gly Arg Lys Arg Arg Pro Phe Lys
 165         170         175
Glu Ser Phe Thr Asp Ser Tyr Phe Thr Asp Val Ile Asn Leu Glu Ala
 180         185         190
Ser Glu Lys Glu Ile Pro Ile Ala Ser Glu Gln Ser Leu Glu Ser Ala
 195         200         205
Thr Ala Ala His Val Thr Thr Val Asp Arg Glu Val Asp Asp Ser Thr
 210         215         220

```

ES 2 774 722 T3

Val Asp Thr Asp Lys Asp Leu Asn Asn Val Leu Lys Asp Leu Leu Ala
 225 230 235 240
 Cys Ser Arg Glu Glu Leu Glu Gly Asp Gly Ala Ile Lys Leu Leu Glu
 245 250 255
 Glu Arg Leu Gln Ile Lys Ser Phe Asn Ile Glu Lys Phe Ser Ile Pro
 260 265 270
 Glu Phe Gln Asp Val Arg Lys Met Asn Leu Lys Ala Ser Gly Ser Asn
 275 280 285
 Pro Pro Asn Arg Lys Ser Leu Ser Asp Ile Gln Asn Ile Leu Lys Gly
 290 295 300
 Thr Asn Arg Val Ala Val Arg Lys Asn Ser His Ser Pro Ser Pro Gln
 305 310 315 320
 Thr Ile Lys His Phe Ser Ser Pro Asn Pro Pro Val Asp Gln Phe Ser
 325 330 335
 Phe Pro Asp Ile His Asn Leu Leu Pro Gly Asp Gln Gln Pro Ser Glu
 340 345 350
 Val Asn Val Gln Pro Ile Ala Lys Asp Ile Pro Asn Thr Ser Pro Thr
 355 360 365
 Asn Val Gly Thr Val Asp Val Ala Ser Pro Phe Asn Asp Ser Val Val
 370 375 380
 Lys Arg Ser Gly Glu Asp Asp Ser His Ile His Ser Gly Ile His Arg
 385 390 395 400
 Ser His Leu Ser Arg Asp Gly Asn Pro Asp Ile Cys Val Met Asp Ser
 405 410 415
 Ile Ser Asn Arg Ser Ser Ala Met Leu Gln Lys Asn Val Asp Met Arg
 420 425 430
 Thr Lys Gly Lys Glu Val Asp Val Pro Met Ser Glu Ser Gly Ala Asn
 435 440 445
 Arg Asn Thr Gly Asp Arg Glu Asn Asp Ala Glu Ile Asn Glu Glu Thr
 450 455 460
 Asp Asn Leu Glu Arg Leu Ala Glu Cys Ala Ser Lys Glu Val Thr Arg
 465 470 475 480
 Pro Phe Thr Val Glu Glu Asp Ser Ile Pro Tyr Gln Gln Gly Ala Ser
 485 490 495
 Ser Lys Ser Pro Asn Arg Ala Pro Glu Gln Tyr Asn Thr Met Gly Gly
 500 505 510
 Ser Leu Glu His Ala Glu His Asn Gln Gly Leu His Glu Glu Glu Asn
 515 520 525
 Val Asn Thr Gly Ser Ala Ser Gly Leu Gln Val Glu Asn Ala Pro Glu
 530 535 540
 Val His Lys Tyr Ser His Lys Gln Thr Asn Lys Arg Arg Lys Arg Gly
 545 550 555 560
 Ser Ser Asp Ser Asn Val Lys Lys Arg Ser Lys Thr Val His Gly Glu
 565 570 575
 Thr Gly Gly Asp Lys Gln Met Lys Thr Leu Pro His Glu Ser Arg Ala
 580 585 590
 Lys Lys Gln Thr Lys Gly Lys Ser Asn Glu Arg Glu Glu Lys Lys Pro
 595 600 605
 Lys Lys Thr Leu Thr His Glu Gly Lys Leu Phe Ser Cys Arg Lys Ser
 610 615 620
 Leu Ala Ala Ala Gly Thr Lys Ile Glu Gly Gly Val Arg Arg Ser Thr
 625 630 635 640
 Arg Ile Lys Ser Arg Pro Leu Glu Tyr Trp Arg Gly Glu Arg Phe Leu
 645 650 655
 Tyr Gly Arg Ile His Glu Ser Leu Thr Thr Val Ile Gly Ile Lys Tyr
 660 665 670
 Ala Ser Pro Gly Glu Gly Lys Arg Asp Ser Arg Ala Ser Lys Val Lys
 675 680 685
 Ser Phe Val Ser Asp Glu Tyr Lys Lys Leu Val Asp Phe Ala Ala Leu
 690 695 700
 His
 705

ES 2 774 722 T3

<210> 97
 <211> 312
 <212> PRT
 <213> Arabidopsis thaliana

5

<220>
 <223> MCM21 de arabidopsis

<400> 97

10

```

Met Gly Glu Met Ile Val Ser Met Asp Gln Asp Ile Arg Leu Asp Thr
 1          5          10          15
Thr Arg Ala Arg Leu Ser Asn Leu Leu Lys Arg His Arg Glu Leu Ser
 20          25          30
Asp Arg Leu Thr Arg Asp Ser Asp Lys Thr Met Leu Asp Arg Leu Asn
 35          40          45
Lys Glu Phe Glu Ala Ala Arg Arg Ser Gln Ser Gln Glu Val Phe Leu
 50          55          60
Asp Gly Glu Glu Trp Asn Asp Gly Leu Leu Ala Thr Leu Arg Glu Arg
 65          70          75          80
Val His Met Glu Ala Asp Arg Lys Ala Asp Asn Gly Asn Ala Gly Phe
 85          90          95
Ser Leu Val Cys His Pro Glu Glu Arg Ile Thr Tyr Arg Val Gly Asn
 100         105         110
Lys Val Ile Cys Cys Leu Asp Gly Ser Arg Ile Gly Ile Gln Phe Glu
 115         120         125
Thr Ser Thr Ala Gly Glu Thr Tyr Glu Val Tyr His Cys Val Leu Glu
 130         135         140
Ser Lys Ser Phe Leu Glu Lys Met Ile Val Leu Glu His Thr Ile Pro
 145         150         155         160
Phe Phe Leu Pro Leu Ser Asp Leu Glu Asn Asp Leu Leu Phe Ser Asn
 165         170         175
Ala Lys Lys Phe Ile Asp Asn Val Gly Asp Leu Leu Gln Ala Tyr Val
 180         185         190
Asp Arg Lys Glu Gln Val Arg Leu Ile Lys Glu Leu Phe Gly His Gln
 195         200         205
Ile Ser Glu Ile Tyr His Ser Leu Pro Tyr His Met Ile Glu Phe Ser
 210         215         220
Met Asp Asp Cys Asp Cys Lys Phe Val Val Ser Leu Arg Tyr Gly Asp
 225         230         235         240
Leu Leu Cys Glu Leu Pro Thr Lys Val Arg Ile Leu Val Trp Pro Met
 245         250         255
His His Leu Ser Lys Lys Gln Cys Thr Ser Pro Gly Ser Pro Ala Ile
 260         265         270
Pro Val Arg Leu Pro Phe Ala Glu Asp Ala Phe Arg Ile Gln Ser Leu
 275         280         285
Pro Glu Ala Tyr Ala Glu Ile Met Pro Asn Met Pro Asn Glu Ile Arg
 290         295         300
Gln Leu Phe Gln Thr Ser Pro Ser
 305         310
    
```

<210> 98
 <211> 285
 <212> PRT
 <213> Arabidopsis thaliana

15

<220>
 <223> inestabilidad del minicromosoma 12 de arabidopsis (MIS12)

20

<400> 98

ES 2 774 722 T3

Met Glu Gly Ser Lys Ser Glu Ala Val Phe Asp Ser Met Asn Leu Asn
 1 5 10 15
 Pro Gln Ile Phe Ile Asn Glu Ala Ile Asn Ser Val Glu Asp Tyr Val
 20 25 30
 Asp Gln Ala Phe Asp Phe Tyr Ala Arg Asp Ala Ser Lys Ser Leu Lys
 35 40 45

Ile Lys Gly Ser Asp Lys Gln Lys Ser Gln Ala Leu Ser Asn Gly Ile
 50 55 60
 Ala Arg Val Arg Gly Leu Leu Ser Val Ile Asp Asn Arg Leu Lys
 65 70 75 80
 Leu Trp Glu Ser Tyr Ser Leu Arg Phe Cys Phe Ala Val Pro Asp Gly
 85 90 95
 Phe Val Leu Pro Lys Ser Glu Glu Ser Ser Ser Val His Gln Asp Gly
 100 105 110
 Leu Tyr Asp Leu Glu Leu Asp Ala Glu Leu Asp Ser Leu Arg Asp Lys
 115 120 125
 Leu Asn Val Val Gly Lys Arg Ser Val Glu Leu Asp Ser Glu Leu Gln
 130 135 140
 Ala Leu Glu Arg Ser Ser Val Ser Arg Glu Arg Ser Leu Arg Leu Val
 145 150 155 160
 Asn Glu Ala Leu Glu Leu Tyr Asp Glu Ser Ser Met Asp Glu Ile Phe
 165 170 175
 Lys Glu Met Thr Lys Met Ala Ser Glu Leu Arg Ala Ser Val Glu Arg
 180 185 190
 Leu Lys Thr Arg Arg Met Lys Ala Ser Glu Ser Ala Lys Val Lys Arg
 195 200 205
 Leu Lys Asn His Gly Lys Glu Phe Ser Ala Met Thr Phe Asp Tyr Val
 210 215 220
 Val Ser Gly Leu Pro Asn Gly Gly Ser Arg Gln Lys Pro Val Ile Pro
 225 230 235 240
 Pro Asp Gln Lys Pro Gln His Ile Ile Leu Val Ser Ser Gln Val Ser
 245 250 255
 Leu Glu Leu Ala Thr Ser Lys Leu Arg Leu Gly Ser Leu Pro Ile Glu
 260 265 270
 Leu Thr Thr Leu Asp Tyr Phe Cys Met Cys Phe Ile Met
 275 280 285

<210> 99
 <211> 568
 5 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> NDC80 de arabidopsis

10 <400> 99

ES 2 774 722 T3

Met	Arg	Gly	Gly	Ala	Ala	Gly	Lys	Arg	Arg	Thr	Thr	Val	Gly	Phe	Gly
1				5					10					15	
Gly	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Ser	Ile	Glu	Gln	Gln	Arg	His	Leu
			20					25					30		
Phe	Asn	Ser	Arg	Asp	Ser	Asp	Ala	Ser	Phe	Ala	Ser	Ser	Arg	Pro	Ser
		35					40					45			
Ser	Ile	Gly	Leu	Gly	Gly	Arg	Gly	Ala	Ser	Asp	Asp	Arg	Ser	Ser	Met
		50				55					60				
Ile	Arg	Phe	Ile	Asn	Ala	Phe	Leu	Ser	Thr	His	Asn	Phe	Pro	Ile	Ser
65					70					75					80
Ile	Arg	Gly	Asn	Pro	Val	Pro	Ser	Val	Lys	Asp	Ile	Ser	Glu	Thr	Leu
			85						90					95	
Lys	Phe	Leu	Leu	Ser	Ala	Leu	Asp	Tyr	Pro	Cys	Asp	Ser	Ile	Lys	Trp
			100					105					110		
Asp	Glu	Asp	Leu	Val	Phe	Phe	Leu	Lys	Ser	Gln	Lys	Cys	Pro	Phe	Lys
		115					120					125			
Ile	Thr	Lys	Ser	Ser	Leu	Lys	Ala	Pro	Asn	Thr	Pro	His	Asn	Trp	Pro
	130					135					140				
Thr	Val	Leu	Ala	Val	Val	His	Trp	Leu	Ala	Glu	Leu	Ala	Arg	Phe	His
145					150					155					160
Gln	His	Leu	Val	Ser	Asn	Ser	Thr	Ser	Val	Pro	Glu	Asp	Asn	Ser	Met
				165					170				175		
Asn	Phe	Phe	Ala	Ile	Gln	Ser	Phe	Gly	His	Phe	Ile	Arg	Gly	Glu	Asp
			180					185					190		

ES 2 774 722 T3

Asp Lys Val Asn Asp Leu Asp Ser Gln Phe Leu Gly Lys Leu Glu Ala
 195 200 205
 Glu Lys Thr Ser Val Ala Glu Thr Ile Ser Gly Cys Glu Lys Ile Ser
 210 215 220
 Gly Glu Leu Glu Ala Lys Leu Glu Ser Leu Arg Lys Gly Pro Ser Lys
 225 230 235 240
 Lys Glu Ser Leu Glu Lys Val Lys Ala Asp Leu Glu Asn Asp Val Asn
 245 250 255
 Lys Phe Arg Thr Ile Val Val Glu Tyr Thr Asp Arg Asn Pro Ala Met
 260 265 270
 Glu Lys Val Val Glu Glu Lys Ala Lys Glu Leu Lys Ala Lys Glu Glu
 275 280 285
 Glu Arg Glu Arg Ile Ser Val Glu Asn Lys Glu Leu Lys Lys Ser Val
 290 295 300
 Glu Leu Gln Asn Phe Ser Ala Ala Asp Val Asn Arg Met Arg Arg Glu
 305 310 315 320
 Leu Gln Ala Val Glu Arg Asp Val Ala Asp Ala Glu Val Ala Arg Asp
 325 330 335
 Gly Trp Asp Gln Lys Ala Trp Glu Leu Asn Ser Gln Ile Arg Asn Gln
 340 345 350
 Phe His Gln Ile Gln Thr Leu Ala Ile Asp Cys Asn Gln Ala Leu Arg
 355 360 365
 Arg Leu Lys Leu Asp Ile Gln Phe Ala Val Asn Glu Arg Gly Glu Thr
 370 375 380
 Pro Ala Ala Val Met Gly Val Asp Tyr Lys Ser Val Val Lys Pro Ala
 385 390 395 400
 Leu Cys Ser Leu Cys Asp Gly Ile Lys Gly Ser Ser Ala Glu Lys Val
 405 410 415
 Glu Glu Leu Val Thr Leu Gln His His Lys Ser Glu Met Ala Ser Lys
 420 425 430
 Ile Glu Ser Lys Arg Ser Leu Leu Gly Ser Ile Gln Leu Gln Ile Asn
 435 440 445
 Asp Leu Glu Glu Lys Met Lys Leu Val Lys Lys Glu Thr Gln Glu Leu
 450 455 460
 Ser Thr Lys Cys Asp Leu Glu Ala Lys Thr Leu Val Glu Ser Val Lys
 465 470 475 480
 Ala Glu Ala Leu Asn Leu Glu Val Val Glu Lys Glu Ala Ala Glu Phe
 485 490 495
 Val Lys Ala Ser Glu Leu Arg Leu Gln Glu Ala Val Lys Glu Ser Glu
 500 505 510
 Glu Glu Val Gln Ala Cys Ala Ala Gln Leu Phe Ala Leu Ile Asp Ser
 515 520 525
 Ile Ser Lys Gln Lys Glu Tyr Met Asp Ser Lys Ile Ser Glu Ile Lys
 530 535 540
 Thr Gly Val Ala Asp Thr Ala Ser Ala Val Ser Glu Ile Tyr Lys Ala
 545 550 555 560
 Asn Phe Lys Lys Asn Leu Gly Ile
 565

<210> 100
 <211> 440
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> NUF2 de arabidopsis

<400> 100

5

10

ES 2 774 722 T3

Met Ser Ala Tyr Glu Tyr Pro Arg Leu Ser Arg Ser Asp Ile Ile Thr
1 5 10 15
Ala Leu Lys Asp Ala Gln Ile Ala Ser Val Thr Glu Thr Asp Leu Lys
20 25 30
Thr Pro Thr Ser Asp Phe Val Ser Glu Leu Tyr Thr Arg Ile Leu Ile
35 40 45

Tyr Leu Asp Ala Leu Asp Glu Glu Glu Lys Gly Gln Val Asp Phe Glu
50 55 60
Ala Leu Glu Gln Leu Glu Asn Pro Asp His His Ala Thr Ser Met Gln
65 70 75 80
Ala Met Lys Leu Tyr Cys Lys Val Lys Asp Met Leu Glu Met Leu Asp
85 90 95
Cys Pro Leu Pro Ile Ser Phe Lys Asp Leu Leu Arg Pro Glu Ser Ser
100 105 110
Arg Thr Glu Phe Phe Ile Ser Ala Leu Leu Asn Tyr Gly Leu Tyr Lys
115 120 125
Asp Ser Lys Met Asp Leu Ile Arg Pro Lys Ala Glu Glu Leu Gly Leu
130 135 140
Leu Asp Glu Gln Arg Lys Gln Cys Glu Ala Lys Val Ala Gln Leu Asn
145 150 155 160
Ala Glu Ile Gly Glu Phe Asp Glu Ala Val Glu Arg Asp Leu Pro Phe
165 170 175
Val Gln Glu Leu Glu Ala Asn Ile Glu Gln Leu Asn Lys Lys Ile Leu
180 185 190
Glu Leu Asn Asn Gln Gln Met Ser Leu Arg Ala Thr Phe Gln Lys Met
195 200 205
Arg Glu Lys Ser Thr Gln Met Asp Asn Glu Ile Ser Lys Ala Glu Phe
210 215 220
Asp Leu Val Glu Thr Val Gln Glu Asn Ala Asn Leu Arg Ser Gln Ile
225 230 235 240
Val Gln Ser Pro Asp Lys Leu Gln Gly Ala Leu Glu Glu Lys Lys Leu
245 250 255
Val Leu Gly Glu Thr Lys Lys Ala Glu Gln Ser Ala Met Val Thr Phe
260 265 270
Gln Glu Lys Ala Ala Ile Leu Glu Val Phe Glu Lys Ala Leu Lys Lys
275 280 285
Ile Leu Lys Ser Ser Ser Gln Leu Gln Leu Ile Asn Glu Gln Val Thr
290 295 300
Asn Ala Lys Thr Val Glu Lys Glu Phe Lys Ala Leu Lys Asp Lys Leu
305 310 315 320
Ser Glu Asp Gly Val Ala Tyr Lys Ser Leu Glu Ala Lys Val Val Glu
325 330 335
Arg Glu Arg Ile Val Glu Gln Leu Asn Glu Ser Leu Lys Gln Leu Glu
340 345 350
Lys Glu Lys Ala Val Met Phe Asp Asp Trp Thr Lys Gln Leu Asn Glu
355 360 365
Leu Lys Val Glu Val Glu Ser Arg Arg Arg Glu Leu Glu Thr Arg Gln
370 375 380
Thr Asn Val Glu Ser Val Val Ala Met Val Asp Asp Asn Thr Ala Lys
385 390 395 400
Thr Asn Gln Val Arg Gln Ser Gly Glu Ala Lys Val Lys Lys Leu Ala
405 410 415
Ala Lys Tyr Glu Glu Ile Val Lys Gln Phe His Glu Tyr Thr Val Ser
420 425 430
Phe Asp Ala Phe Leu Pro Ser Leu
435 440

<210> 101
<211> 1109
<212> ADN

ES 2 774 722 T3

<213> Secuencia artificial

<220>

<223> cola de CENH3 de maíz sintética, dominio de plegamiento de histona de arabidopsis

5

<400> 101

atggctcgaa ccaagcacca ggccgtgagg aagacggcgg agaagcccaa gaagaagctc 60
 cagttcgagc gctcaggtgg tgcgagtacc tcggcgacgc cggaaagggc tgctgggacc 120
 gggggaagag cggcgtctgg aggtgactca gttaagaaga cgaaccacg ccaccgctgg 180

aggccaggaa ccgttgctct aaaagagatt cgccatttcc agaagcagac aaaccttctt 240
 attccggctg ccagtttcat aagagaagtt agttactctt tttcttacca gccataataa 300
 gtttcacagc ttaacaatat tcatatatac taacagaggc acaagccttt tgggtgttaa 360
 tgtggctagt tttaggattt gcacacccca cacatatctg agcatcaatg cagtgtacat 420
 agtgagtgat atagcaattt aactaaaatt cagagtaatc gtgaggccaa ccctccttgt 480
 ttaaggagtg tgtaactctag tttgtctttg aggttatgag ctcatagatt cagaaccata 540
 tgattcctgt agctacaaaa ctcaacatga atcgtcagtg atgtggaaat gctgatttgt 600
 gttacaaaca aactatttta cattgttttt ccaggtgaga agtataacc atagttggc 660
 ccctcccaa atcaatcgtt ggacagctga agctcttgtt gctcttcaag aggtaccaat 720
 ccttcaactt tttctttata cgaatgtatg aatatagata tagagatagt cacacatttc 780
 aactaatgtc attccccttg atgaccaatc aacctaatac cacaattctt ttgtggtagg 840
 cggcagaaga ttacttggtt ggtttgttct cagattcaat gctctgtgct atccatgcaa 900
 gacgtgttac tctaagtaag tactctaaaa gaagacattt ttcagtctca acttaggaat 960
 cacaagcata cattttatat ccctttgaat cattagttac ttgaatatca tatataaaaa 1020
 tgcttatcta tatctgtttt ttgttcatat cagtgagaaa agactttgaa cttgcacgcc 1080
 ggcttgaggg aaaaggcaga ccatgggtga 1109

10 <210> 102

<211> 477

<212> ADN

<213> Secuencia artificial

15

<220>

<223> ADNc de cola de CENH3 maíz sintético

<400> 102

atggctcgaa ccaagcacca ggccgtgagg aagacggcgg agaagcccaa gaagaagctc 60
 cagttcgagc gctcaggtgg tgcgagtacc tcggcgacgc cggaaagggc tgctgggacc 120
 gggggaagag cggcgtctgg aggtgactca gttaagaaga cgaaccacg ccaccgctgg 180
 aggccaggaa ccgttgctct aaaagagatt cgccatttcc agaagcagac aaaccttctt 240
 attccggctg ccagtttcat aagagaagtg agaagtataa cccatatggt ggcccctccc 300
 caaatcaatc gttggacagc tgaagctctt gttgctcttc aagaggcggc agaagattac 360
 ttggttggtt tgttctcaga ttcaatgctc tgtgctatcc atgcaagacg tgttactcta 420
 atgagaaaag actttgaact tgcacgcggc cttggaggaa aaggcagacc atggtgga 477

20

<210> 103

<211> 158

<212> PRT

25

<213> Secuencia artificial

<220>

<223> proteína tailswap de maíz sintética

30

<400> 103

ES 2 774 722 T3

Met	Ala	Arg	Thr	Lys	His	Gln	Ala	Val	Arg	Lys	Thr	Ala	Glu	Lys	Pro
1				5					10					15	
Lys	Lys	Lys	Leu	Gln	Phe	Glu	Arg	Ser	Gly	Gly	Ala	Ser	Thr	Ser	Ala
			20					25					30		
Thr	Pro	Glu	Arg	Ala	Ala	Gly	Thr	Gly	Gly	Arg	Ala	Ala	Ser	Gly	Gly
		35					40					45			
Asp	Ser	Val	Lys	Lys	Thr	Lys	Pro	Arg	His	Arg	Trp	Arg	Pro	Gly	Thr
	50					55					60				
Val	Ala	Leu	Lys	Glu	Ile	Arg	His	Phe	Gln	Lys	Gln	Thr	Asn	Leu	Leu
65					70					75					80
Ile	Pro	Ala	Ala	Ser	Phe	Ile	Arg	Glu	Val	Arg	Ser	Ile	Thr	His	Met
				85					90					95	
Leu	Ala	Pro	Pro	Gln	Ile	Asn	Arg	Trp	Thr	Ala	Glu	Ala	Leu	Val	Ala
			100					105						110	
Leu	Gln	Glu	Ala	Ala	Glu	Asp	Tyr	Leu	Val	Gly	Leu	Phe	Ser	Asp	Ser
		115					120						125		
Met	Leu	Cys	Ala	Ile	His	Ala	Arg	Arg	Val	Thr	Leu	Met	Arg	Lys	Asp
	130					135						140			
Phe	Glu	Leu	Ala	Arg	Arg	Leu	Gly	Gly	Lys	Gly	Arg	Pro	Trp		
145					150					155					

<210> 104
 <211> 708
 <212> ADN
 <213> Secuencia artificial

5

<220>
 <223> variante DsRed de proteína roja fluorescente mCherry sintética

10

<400> 104

```

atggtgagca agggcgagga ggataacatg gccatcatca aggagtcat gcgcttcaag 60
gtgcacatgg agggctccgt gaacggccac gagttcgaga tcgagggcga gggcgagggc 120
cgcccctacg agggcaccca gaccgccaag ctgaaggtga ccaaggggtg ccccctgccc 180
ttcgctggg acatcctgtc ccctcagttc atgtacggct ccaaggccta cgtgaagcac 240
cccgccgaca tccccgacta cttgaagctg tccttccccg agggcttcaa gtgggagcgc 300
gtgatgaact tcgaggacgg cggcgtggtg accgtgaccc aggactcctc cctgcaggac 360
ggcgagtca tctacaaggt gaagctgcgc ggcaccaact tcccctccga cggccccgta 420
atgcagaaga agaccatggg ctgggaggcc tcctccgagc ggatgtaccc cgaggacggc 480
gccctgaagg gcgagatcaa gcagaggctg aagctgaagg acggcggcca ctacgacgct 540
gaggtcaaga ccacctaaa ggccaagaag cccgtgcagc tgcccggcgc ctacaacgctc 600
aacatcaagt tggacatcac ctcccacaac gaggactaca ccatcgtgga acagtacgaa 660
cgcgccgagg gccgccactc caccggcggc atggacgagc tgtacaag 708
    
```

<210> 105
 <211> 236
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> variante DsRed de proteína roja fluorescente mCherry sintética

20

<400> 105

ES 2 774 722 T3

Met	Val	Ser	Lys	Gly	Glu	Glu	Asp	Asn	Met	Ala	Ile	Ile	Lys	Glu	Phe
1				5					10					15	
Met	Arg	Phe	Lys	Val	His	Met	Glu	Gly	Ser	Val	Asn	Gly	His	Glu	Phe
			20					25					30		
Glu	Ile	Glu	Gly	Glu	Gly	Glu	Gly	Arg	Pro	Tyr	Glu	Gly	Thr	Gln	Thr
		35					40					45			
Ala	Lys	Leu	Lys	Val	Thr	Lys	Gly	Gly	Pro	Leu	Pro	Phe	Ala	Trp	Asp
		50				55					60				
Ile	Leu	Ser	Pro	Gln	Phe	Met	Tyr	Gly	Ser	Lys	Ala	Tyr	Val	Lys	His
65					70					75					80
Pro	Ala	Asp	Ile	Pro	Asp	Tyr	Leu	Lys	Leu	Ser	Phe	Pro	Glu	Gly	Phe
				85					90					95	
Lys	Trp	Glu	Arg	Val	Met	Asn	Phe	Glu	Asp	Gly	Gly	Val	Val	Thr	Val
			100					105					110		
Thr	Gln	Asp	Ser	Ser	Leu	Gln	Asp	Gly	Glu	Phe	Ile	Tyr	Lys	Val	Lys
		115					120					125			
Leu	Arg	Gly	Thr	Asn	Phe	Pro	Ser	Asp	Gly	Pro	Val	Met	Gln	Lys	Lys
		130				135					140				
Thr	Met	Gly	Trp	Glu	Ala	Ser	Ser	Glu	Arg	Met	Tyr	Pro	Glu	Asp	Gly
145					150					155					160
Ala	Leu	Lys	Gly	Glu	Ile	Lys	Gln	Arg	Leu	Lys	Leu	Lys	Asp	Gly	Gly
			165						170				175		
His	Tyr	Asp	Ala	Glu	Val	Lys	Thr	Thr	Tyr	Lys	Ala	Lys	Lys	Pro	Val
			180					185					190		
Gln	Leu	Pro	Gly	Ala	Tyr	Asn	Val	Asn	Ile	Lys	Leu	Asp	Ile	Thr	Ser
		195					200					205			
His	Asn	Glu	Asp	Tyr	Thr	Ile	Val	Glu	Gln	Tyr	Glu	Arg	Ala	Glu	Gly
	210					215					220				
Arg	His	Ser	Thr	Gly	Gly	Met	Asp	Glu	Leu	Tyr	Lys				
225					230						235				

<210> 106

<211> 211

5 <212> PRT

<213> *Candida albicans*

<220>

10 <223> CENH3 de levadura

<400> 106

ES 2 774 722 T3

Met Ala Arg Leu Ser Gly Gln Ser Ser Gly Arg Gln Thr Gly Gln Gly
 1 5 10 15
 Thr Ser Ala Glu Ala Ile Arg Gln Gln Arg Glu Glu Leu Arg Arg Gln
 20 25 30
 Arg Glu Leu Arg Leu Gln Gln Gln Gln Ala Glu Arg Gln Gln Gln
 35 40 45
 Arg Gln Gln Tyr Arg Thr Glu Gln Ser Pro Ile Val Pro Ala Ala Thr
 50 55 60
 Ser Ser Ser Arg Tyr Ser Gln Phe Gly Ile Tyr Arg Asn Gln Pro Gly
 65 70 75 80
 Asp Val Val Asp Thr Leu Ala Ser Ser Leu Pro Arg Arg Thr Thr Thr
 85 90 95
 Thr Arg Pro Glu Val Asn Arg Thr Val Pro Arg Val Lys Lys Arg Tyr
 100 105 110
 Arg Pro Gly Thr Lys Ala Leu Arg Glu Ile Arg Gln Tyr Gln Lys Ser
 115 120 125
 Thr Asp Leu Leu Ile Arg Lys Leu Pro Phe Ala Arg Leu Val Arg Glu
 130 135 140
 Ile Ser Leu Asp Phe Val Gly Pro Ser Tyr Gly Leu Arg Trp Gln Ser
 145 150 155 160
 Asn Ala Ile Leu Ala Leu Gln Glu Ala Ser Glu Ser Phe Leu Ile His
 165 170 175
 Leu Leu Glu Asp Thr Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr
 180 185 190
 Ile Met Gln Lys Asp Ile Gln Leu Ala Arg Arg Ile Arg Gly Gln Ser
 195 200 205
 Trp Ile Leu
 210

<210> 107
 <211> 140
 <212> PRT
 <213> *Homo sapiens*

5

<220>
 <223> CENH3 humana

10

<400> 107

Met Gly Pro Arg Arg Arg Ser Arg Lys Pro Glu Ala Pro Arg Arg Arg
 1 5 10 15
 Ser Pro Ser Pro Thr Pro Thr Pro Gly Pro Ser Arg Arg Gly Pro Ser
 20 25 30
 Leu Gly Ala Ser Ser His Gln His Ser Arg Arg Arg Gln Gly Trp Leu
 35 40 45
 Lys Glu Ile Arg Lys Leu Gln Lys Ser Thr His Leu Leu Ile Arg Lys
 50 55 60
 Leu Pro Phe Ser Arg Leu Ala Arg Glu Ile Cys Val Lys Phe Thr Arg
 65 70 75 80
 Gly Val Asp Phe Asn Trp Gln Ala Gln Ala Leu Leu Ala Leu Gln Glu
 85 90 95
 Ala Ala Glu Ala Phe Leu Val His Leu Phe Glu Asp Ala Tyr Leu Leu
 100 105 110
 Thr Leu His Ala Gly Arg Val Thr Leu Phe Pro Lys Asp Val Gln Leu
 115 120 125
 Ala Arg Arg Ile Arg Gly Leu Glu Glu Gly Leu Gly
 130 135 140

15 <210> 108
 <211> 164

ES 2 774 722 T3

<212> PRT
<213> *Oryza sativa*

5 <220>
<223> histona 3 centromérica de arroz (CENH3)

<400> 108

```

Met Ala Arg Thr Lys His Pro Ala Val Arg Lys Ser Lys Ala Glu Pro
 1          5          10          15
Lys Lys Lys Leu Gln Phe Glu Arg Ser Pro Arg Pro Ser Lys Ala Gln
 20          25          30
Arg Ala Gly Gly Gly Thr Gly Thr Ser Ala Thr Thr Arg Ser Ala Ala
 35          40          45
Gly Thr Ser Ala Ser Gly Thr Pro Arg Gln Gln Thr Lys Gln Arg Lys
 50          55          60
Pro His Arg Phe Arg Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Lys
 65          70          75          80
Phe Gln Lys Thr Thr Glu Leu Leu Ile Pro Phe Ala Pro Phe Ser Arg
 85          90          95
Leu Val Arg Glu Ile Thr Asp Phe Tyr Ser Lys Asp Val Ser Arg Trp
 100         105         110
Thr Leu Glu Ala Leu Leu Ala Leu Gln Glu Ala Ala Glu Tyr His Leu
 115         120         125
Val Asp Ile Phe Glu Val Ser Asn Leu Cys Ala Ile His Ala Lys Arg
 130         135         140
Val Thr Ile Met Gln Lys Asp Met Gln Leu Ala Arg Arg Ile Gly Gly
 145         150         155         160
Arg Arg Pro Trp
    
```

10 <210> 109
<211> 97
<212> PRT
<213> Secuencia artificial

15 <220>
<223> dominio de plegamiento de histona CENH3 sintético

20 <400> 109

```

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

25 <210> 110
<211> 30
<212> ADN
<213> Secuencia artificial

30 <220>
<223> cebador sintético de amplificación por PCR CP 384

<220>

ES 2 774 722 T3

<221> base_modificada
 <222> (1)...(4)
 <223> n = g, a, c o t

5 <400> 110
 nnnngtcgac atggctcgaa ccaagcacca 30

<210> 111
 <211> 27
 10 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador sintético de amplificación por PCR CP 572

15 <400> 111
 caacggttcc tggcctccag cggtggc 27

<210> 112
 <211> 27
 20 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador sintético de amplificación por PCR CP 571

25 <400> 112
 gccaccgctg gaggccagga accgttg 27

<210> 113
 <211> 34
 30 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador sintético de amplificación por PCR CP 375

<220>
 <221> base_modificada
 <222> (1)...(4)
 <223> n = g, a, c o t

40 <400> 113
 nnnntctaga tcaccatggt ctgcctttc ctcc 34

<210> 114
 <211> 6
 <212> PRT
 <213> Secuencia artificial

50 <220>
 <223> secuencia conservada sintética en el borde entre el dominio de cola y el dominio de plegamiento de histona de proteínas CENH3

55 <400> 114

Pro Gly Thr Val Ala Leu
1 5

REIVINDICACIONES

- 5 1. Una planta transgénica que comprende un casete de expresión transgénica heterólogo, comprendiendo el casete de expresión un promotor unido operativamente a un polinucleótido que codifica un polipéptido que comprende una secuencia de aminoácidos heteróloga de al menos 5 aminoácidos unidos a un extremo amino terminal de una proteína que comprende un dominio de plegamiento de histona CENH3, en donde la secuencia de aminoácidos heteróloga es heteróloga con respecto al dominio de plegamiento de histona CENH3 y en donde todos los alelos de la secuencia codificante genómica de CENH3 endógena de la planta están inactivados o desactivados.
- 10 2. La planta transgénica de la reivindicación 1, en donde la secuencia de aminoácidos heteróloga se une al dominio de plegamiento de histona CENH3 a través de una secuencia de proteína interpuesta.
- 15 3. La planta transgénica de la reivindicación 2, en donde la secuencia de proteína interpuesta comprende un dominio de cola de histona H3 no CENH3.
- 20 4. La planta transgénica de la reivindicación 2, en donde la secuencia de proteína interpuesta comprende un dominio de cola de CENH3.
- 25 5. La planta transgénica de la reivindicación 4, en donde el dominio de cola de CENH3 es heterólogo con respecto al dominio de plegamiento de histona CENH3.
- 30 6. La planta transgénica de la reivindicación 2, en donde la secuencia de aminoácidos heteróloga tiene al menos 10 aminoácidos de longitud.
- 35 7. La planta transgénica de la reivindicación 1, en donde la secuencia de aminoácidos heteróloga comprende un dominio de cola de histona H3 no CENH3 unido a un extremo amino de un dominio de plegamiento de histona CENH3.
- 40 8. Un ácido nucleico aislado que comprende un polinucleótido que codifica el polipéptido de una cualquiera de las reivindicaciones 3, 5 y 7.
- 45 9. Un método para generar una planta de progenie F1 que tiene la mitad de la ploidía de una planta parental que expresa una proteína CENH3 endógena, comprendiendo el método cruzar la planta parental con la planta de una cualquiera de las reivindicaciones 1 a 7; y seleccionar la progenie F1 generada a partir de la etapa de cruce que tiene la mitad de la ploidía de la planta parental.
- 50 10. El método de la reivindicación 9, en donde la planta parental es el progenitor de polen.
- 55 11. El método de la reivindicación 9, en donde la planta parental es el progenitor de óvulo.
12. El método de la reivindicación 9, que comprende adicionalmente convertir al menos una planta haploide seleccionada en una planta doblemente haploide.
13. Un método para hacer la planta transgénica de una cualquiera de las reivindicaciones 1 a 7, que comprende transformar células vegetales con un ácido nucleico que comprende el casete de expresión; y seleccionar transformantes que comprenden el ácido nucleico, haciendo de este modo la planta transgénica.
14. El método de la reivindicación 12, en donde la planta parental es un diploide y la progenie F1 seleccionada es haploide.
15. Una planta que comprende una o dos copias de un alelo de un gen de CENH3 endógena mutado, en donde la CENH3 endógena mutada se muta para reducir la función CENH3 y en donde dicha planta no se obtiene exclusivamente mediante un proceso esencialmente biológico.
16. Un método para generar una planta de progenie F1 que tiene la mitad de la ploidía de una planta parental que expresa una proteína CENH3 endógena, comprendiendo el método cruzar la planta parental con la planta de la reivindicación 15 y seleccionar la progenie F1 generada a partir de la etapa de cruce que tiene la mitad de la ploidía de la planta parental.

