

MÉTODOS PARA PRODUCIR PLANTAS CON DESARROLLO DE FRUTAS ALTERADAS Y PLANTAS DERIVAS DE LAS MISMAS

CAMPO.

La presente descripción se refiere al campo de la agricultura, particularmente a los cultivos de consumo, lo que incluye métodos para generar plantas que producen frutos con un fenotipo de desarrollo de fruta alterado, tal como un fenotipo sin semilla o de reducida simiente.

DECLARACIÓN DE PRIORIDAD

Esta solicitud reivindica el beneficio, conforme al Título 35 U. S. C. § 119 (e), de la Solicitud Provisional de los Estados Unidos Nro. 62/859.992, presentada el 11 de junio de 2019, cuyo contenido completo se incorpora aquí como referencia.

DESCRIPCIÓN DEL ARCHIVO DE TEXTO PRESENTADO ELECTRÓNICAMENTE.

Un Listado de Secuencias en formato de texto ASCII, presentado conforme al Título 37 C. F. R. § 1.821, titulado 1499-38WO_ST25.txt, de 380.154 bytes de tamaño, generado el 9 de junio de 2020 y presentado a través de EFS-Web, se proporciona en lugar de una copia en papel. Este Listado de Secuencias se incorpora aquí como referencia en la memoria descriptiva para sus divulgaciones.

ANTECEDENTES.

El desarrollo del fruto normalmente depende de la fertilización y la posterior formación de la semilla. La fruta puede tener semillas duras, tales como uvas o sandías, o puede ser drupas, tales como cerezas, duraznos o frambuesas negras. Una drupa (o fruta de hueso) es una fruta indehiscente en la cual una parte carnosa externa (exocarpio, o piel; y mesocarpio, o carne) rodea un único caparazón (hoyo, hueso o pireno) de endocarpio endurecido con una semilla (grano) dentro. El endocarpio está compuesto de lignina para proteger la semilla. El endocarpio se desarrolla desde la capa más interna del ovario.

Las frutas sin semillas pueden desarrollarse de dos maneras: partenocarpia y estenospermocarpia. En la partenocarpia, el fruto se desarrolla sin polinización ni fertilización, y es deseable en cultivos frutales que pueden ser difíciles de polinizar o fertilizar, como el higo, el tomate

y la calabaza de verano. En la estenospermocarpia, la polinización o la fertilización desencadenan el desarrollo del fruto, pero los óvulos o embriones abortan sin producir semillas maduras.

La presente invención proporciona nuevos enfoques para alterar el desarrollo de frutos de plantas y la producción de semillas.

SÍNTESIS.

La presente descripción proporciona un método de mejoramiento de plantas de próxima generación, para producir plantas con un fenotipo de desarrollo de fruta alterado. El método de mejoramiento vegetal es capaz de producir plantas que exhiben un desarrollo alterado del fruto (por ejemplo, un fenotipo sin semillas, y/o plantas con un nivel reducido de formación de endocarpio). Un aspecto del método de mejoramiento de plantas de próxima generación es el descubrimiento de los inventores de un objetivo genómico que es altamente conservado a través de una cantidad de taxones de plantas, y está correlacionado con la producción de fruta; en consecuencia, cuando dicho objetivo genómico es editado de la manera enseñada, es posible producir una planta con, por ejemplo, un fenotipo sin semillas, y/o una planta con un nivel reducido de formación de endocarpio.

En un aspecto, la invención proporciona una planta o parte de planta de la misma que comprende por lo menos una mutación no natural en un gen endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS.

En otro aspecto, se proporciona una célula vegetal, donde la célula vegetal comprende un sistema de edición de bases que comprende: (a) una proteína efectora asociada a CRISPR; (b) una citidina desaminasa o adenosina desaminasa; y (c) un ácido nucleico guía (ARNg) que tiene una secuencia espaciadora con complementariedad a un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AG.

En un aspecto adicional, se proporciona una planta o parte de la misma (por ejemplo, una célula vegetal), que comprende por lo menos una mutación no natural en un gen de factor de transcripción de caja MADS clado AG endógeno que previene o reduce la formación de semillas, en donde por lo menos una mutación es una sustitución, inserción o delección que se introduce utilizando un sistema de edición que comprende un dominio de unión de ácido nucleico que se une a un sitio objetivo en el gen del factor de transcripción de la caja MADS clado AG, en donde el factor de transcripción de la caja MADS clado AG gen es un gen SEEDSTICK (STK), donde el gen STK (a) comprende por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o una región que tiene por lo

menos 90% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**; o (b) codifica un factor de transcripción de caja MADS clado AG que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164 o 166–171** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**.

En otro aspecto, una planta o parte de la misma que comprende un gen de factor de transcripción de caja MADS clado AG endógeno mutado, en donde el gen de factor de transcripción de caja MADS clado AG endógeno mutado es un gen SEEDSTICK (STK) que (a) comprende una secuencia que tiene por lo menos 80% de identidad con la secuencia de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o secuencia que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**; o (b) codifica una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164 o 166–171** o una secuencia que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**.

Además se proporciona un método para producir/fitomejorar una planta editada con bases libre de transgén, que comprende: (a) cruzar una planta de la invención con una planta libre de transgén, para introducir así la por lo menos una mutación, donde la mutación o la modificación en la planta está libre de transgenes; y (b) seleccionar una planta de progenie que comprende la por lo menos una sustitución de nucleótido simple pero que esté libre de transgén, de manera de producir así una planta editada con bases libre de transgenes.

En otro aspecto, se proporciona un método para proporcionar una pluralidad de plantas que exhiben un fenotipo de desarrollo de fruto alterado (por ejemplo, producción de semillas reducida o nula), en donde el método comprende plantar dos o más plantas de cualquiera de las reivindicaciones 1–24 o 30–49 en un área de cultivo (p. ej., un campo (p. ej., un campo cultivado, un campo agrícola), una cámara de crecimiento, un invernadero, un área recreativa, un césped y/o una carretera y similares), de manera de proporcionar una pluralidad de plantas que exhiben un desarrollo alterado de frutos en comparación con una pluralidad de plantas de control que no comprenden la mutación.

En un aspecto adicional, se proporciona un método para editar un sitio específico en el genoma de una célula vegetal, donde el método comprende: escindir, de una manera específica de sitio, un

sitio objetivo dentro de un gen de factor de transcripción de caja MADS clado AG endógeno en una célula vegetal, donde el gen de factor de transcripción de caja MADS clado AG endógeno es un gen SEEDSTICK (STK) (a) que comprende por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**; o (b) codifica un factor de transcripción de caja MADS clado AG que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164 o 166–171** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, de modo de generar una edición en el gen de factor de transcripción de caja MADS clado AG endógeno de la célula vegetal.

Un aspecto adicional proporciona un método para producir una planta, que comprende: (a) poner en contacto una población de células vegetales que comprenden una secuencia de ADN que codifica un gen SEEDSTICK (STK) endógeno de tipo salvaje con una nucleasa unida a un dominio de unión de ácido nucleico (por ejemplo, sistema de edición) que se une a una secuencia: (a) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**; o (b) que codifica un factor de transcripción SEEDSTICK (STK) (i) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171** o (ii) que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**; (b) seleccionar una célula vegetal de dicha población en donde por lo menos una secuencia de ADN que codifica el factor de transcripción STK se ha mutado, en donde la mutación comprende una sustitución de por lo menos un nucleótido en la por lo menos una secuencia de ADN; y (c) cultivar la célula vegetal seleccionada hasta una planta.

Además, se proporciona un método para producir una planta o una parte de la misma que comprende por lo menos una célula en la que un gen del factor de transcripción de caja MADS clado AG endógeno está mutado, donde el método comprende poner en contacto un sitio objetivo en el gen del factor de transcripción de caja MADS clado AG en la planta o parte de la planta, con una nucleasa unida a un dominio de unión de ácido nucleico (por ejemplo, sistema de edición) que se une al sitio objetivo, en donde el gen del factor de transcripción de caja MADS clado AG es un gen SEEDSTICK

(STK) (a) que comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, o (b) que codifica un factor de transcripción SEEDSTICK (STK) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o que comprende una región que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, de modo de producir una planta o parte de la misma que comprende por lo menos una célula que tiene una mutación en el gen de factor de transcripción de caja MADS clado AG endógeno.

En un aspecto adicional, se proporciona un método para producir una planta que tiene un fenotipo de desarrollo de fruto alterado, que comprende (a) poner en contacto una célula vegetal que comprende un gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje con una nucleasa que se dirige al gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje, en donde la nucleasa se une a un dominio de unión de ADN que se une a un sitio objetivo en el gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje, en donde el gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje es un gen SEEDSTICK (STK) que (i) comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174** y/o (ii) codifica una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o una secuencia que comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**; (b) cultivar la célula vegetal hasta una planta, de manera de producir una planta que tiene un fenotipo de fruto alterado.

En un aspecto adicional, un método para producir una planta o parte de la misma que comprende por lo menos una célula que tiene un gen de factor de transcripción de caja MADS clado AG endógeno mutado, donde el método comprende poner en contacto un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG endógeno en la planta o parte de planta con una

nucleasa que comprende un dominio de escisión y un dominio de unión de ADN, en donde el dominio de unión de ADN se une a un sitio objetivo en el gen de factor de transcripción de caja MADS clado AG endógeno, en donde el gen de factor de transcripción de caja MADS clado AG endógeno es un gen SEEDSTICK (STK) que (a) codifica (i) una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171** o (ii) una secuencia que comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**; y/o (b) comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** or o una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, de modo de producir la planta o parte de la misma que comprende por lo menos una célula que tiene un gen de factor de transcripción de caja MADS clado AG endógeno mutado.

Se proporciona adicionalmente un método de producción de una planta o parte de la misma que comprende un gen de factor de transcripción de caja MADS clado AG endógeno mutado y que tiene desarrollo de fruto alterado, donde el método comprende poner en contacto un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG endógeno, con una nucleasa que comprende un dominio de escisión y un dominio de unión de ADN (por ejemplo, sistema de edición) que comprende un dominio de unión de ácido nucleico que se une al sitio objetivo en el gen del factor de transcripción de caja MADS clado AG, en donde el gen del factor de transcripción de caja MADS clado AG es un gen SEEDSTICK (STK) (a) que comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, o (b) que codifica un factor de transcripción SEEDSTICK (STK) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o que comprende una región que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, de modo de producir una planta o parte de la misma que comprende un gen de factor de transcripción de caja MADS clado AG endógeno mutado y que tiene desarrollo de fruto alterado.

En un aspecto adicional, la invención proporciona un ácido nucleico guía (por ejemplo, ARNg, ADNg, ARNcr, ADNcr) que se une a un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG, en donde el gen de factor de transcripción de caja MADS clado AG endógeno es un gen SEEDSTICK (STK) (a) que comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, o (b) que codifica un factor de transcripción SEEDSTICK (STK) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o una región que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**.

Adicionalmente se proporciona un sistema que comprende un ácido nucleico guía de la invención y una proteína efectora CRISPR–Cas que se asocia con el ácido nucleico guía.

Además se proporciona un sistema de edición de genes que comprende una proteína efectora CRISPR–Cas en asociación con un ácido nucleico guía, en donde el ácido nucleico guía comprende una secuencia espaciadora que se une a un gen de factor de transcripción de caja MADS clado AG.

En un aspecto adicional, se proporciona un complejo que comprende una proteína efectora CRISPR–Cas que comprende un dominio de escisión y un ácido nucleico guía, en donde el ácido nucleico guía se une a un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG, en donde el gen del factor de transcripción de caja MADS clado AG es un gen SEEDSTICK (STK) (a) que comprende por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, y/o (b) que codifica (i) una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128 o 150**, o (ii) una secuencia que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o una región que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, en donde el dominio de escisión escinde una hebra objetivo en el gen del factor de transcripción de caja MADS clado AG.

En un aspecto adicional, se proporciona un casete de expresión que comprende (a) polinucleótido que codifica proteína efectora CRISPR–Cas que comprende un dominio de escisión y (b) un ácido nucleico guía que se une a un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG, en donde el gen de factor de transcripción de caja MADS clado AG es un gen SEEDSTICK, y el ácido nucleico guía comprende una secuencia espaciadora que es complementaria y se une a una secuencia que tiene por lo menos 80% de identidad de secuencia con por lo menos una porción de una secuencia que tiene por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o una porción de una secuencia que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**; o una secuencia que codifica (i) un polipéptido que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o (ii) un polipéptido que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, opcionalmente, en donde una porción es de alrededor de 2 a alrededor de 22 nucleótidos consecutivos de longitud.

Se proporciona adicionalmente un ácido nucleico que codifica un alelo dominante–negativo, alelo semidominante, alelo de pérdida de función débil, o una mutación hipomórfica de un factor de transcripción de caja MADS clado AG, opcionalmente, en donde el factor de transcripción de caja MADS clado AG es AGAMOUS (AG), SHATTERPROOF 1 (SHP1), SHATTERPROOF 2 (SHP2), y/o SEEDSTICK (STK).

Además, se proporcionan plantas que comprenden en su genoma uno o más genes del factor de transcripción de caja MADS clado AG mutado producidas por los métodos de la invención, así como polipéptidos, polinucleótidos, construcciones de ácido nucleico, casetes de expresión y vectores para hacer una planta de esta invención.

Estos y otros aspectos de la invención se exponen con más detalle en la descripción de la invención a continuación.

BREVE DESCRIPCIÓN DE LAS FIGURAS.

FIG. 1 ilustra una forma de realización de una métrica que se puede utilizar para definir un fenotipo de desarrollo de fruto alterado, por ejemplo, reducción de simiente. Es decir, en aspectos de la divulgación, la simiente reducida puede describirse como contenido reducido de lignina en

comparación con el tipo salvaje (lignina/mg de pared celular libre de proteína en endocarpio y/o cubierta de semilla). La Figura 3 representa una comparación gráfica de la "simiente" de la soja, la frambuesa roja, la frambuesa negra y la mora, sobre la base de μg de lignina/mg de pared celular libre de proteína.

FIG. 2 proporciona una alineación entre las secuencias de polipéptidos SEEDSTICK. De arriba a abajo, **SEQ ID NO: 53, SEQ ID NO: 66, SEQ ID NO: 160, SEQ ID NO: 158, SEQ ID NO: 156, SEQ ID NO: 154, SEQ ID NO: 152, SEQ ID NO: 97, SEQ ID NO: 166, SEQ ID NO: 150, SEQ ID NO: 78, SEQ ID NO: 29, SEQ ID NO: 128.**

FIG. 3 proporciona una alineación entre secuencias de polipéptido SEEDSTICK. De arriba a abajo, **SEQ ID NO: 171, SEQ ID NO: 166, SEQ ID NO: 170, SEQ ID NO: 169, SEQ ID NO: 78, SEQ ID NO: 168, SEQ ID NO: 167.**

FIG. 4 ilustra modificaciones ejemplares de una secuencia *SEEDSTICK* de mora de tipo salvaje (porción de secuencia de aminoácido **SEQ ID NO: 204**; porción de secuencia de nucleótido **SEQ ID NO: 205**). Ejemplos de secuencias de ácido nucleico mutadas y correspondientes secuencias de aminoácidos, de arriba a abajo: **SEQ ID NO: 219, SEQ ID NO: 220, SEQ ID NO: 221, SEQ ID NO: 222, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 225, SEQ ID NO: 226, SEQ ID NO: 227, SEQ ID NO: 228, SEQ ID NO: 229, SEQ ID NO: 230, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 233, SEQ ID NO: 234, SEQ ID NO: 235, SEQ ID NO: 236, SEQ ID NO: 237, SEQ ID NO: 238, SEQ ID NO: 239, SEQ ID NO: 20, SEQ ID NO: 241, SEQ ID NO: 242, SEQ ID NO: 243, SEQ ID NO: 244, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 247, SEQ ID NO: 248, SEQ ID NO: 249, SEQ ID NO: 250.**

FIG. 5 ilustra una comparación de las secuencias de aminoácidos putativas de los ortólogos STK de *Arabidopsis* (**SEQ ID NO: 13**), frutilla silvestre (**SEQ ID NO: 29**) y frambuesa negra (**SEQ ID NO: 97**). Las secuencias de aminoácidos alineadas demuestran que las secuencias de proteínas de tres ortólogos STK tienen la estructura de dominio típica de tipo MIKC que comprende un dominio de tipo caja MADS/MADS MEF2 (M), un dominio Intermedio (I), un dominio de tipo queratina (caja K) y un dominio C-terminal (C). Además, los motivos específicos de Ag, que distinguen los ortólogos del factor de transcripción del clado AG de otros genes de caja MADS, se ilustran mediante una caja rectangular (por ejemplo, residuos múltiples) o cuadrada (por ejemplo, residuo único) debajo de las residencias de aminoácidos correspondientes.

FIG. 6 ilustra tres ejemplos de ARN guía diseñados para apuntar al codón de Arginina (R) conservado en el dominio C-terminal de la proteína STK. Los ARN guía S1 y S2 pueden

desencadenar, por ejemplo, la sustitución de aminoácidos Arginina (R) a Histidina (H) por cambio de nucleótido C a T (o G a A) por el editor de bases de citosina. El ARN guía S3 puede inducir, por ejemplo, la sustitución de Arginina (R) a Cisteína (C) por cambio de nucleótido C a T (o G a A) por el editor de bases de citosina.

FIG. 7 proporciona los resultados de un ensayo para determinar la actividad negativa dominante de tres mutaciones diferentes en genes de *Arabidopsis thaliana* SEEDSTICK (STK). El ensayo se lleva a cabo como se describe en el Ejemplo 3. R>H, R>L es edición del R conservado a un H o L; dC es una delección de C-terminal.

FIG. 8 proporciona los resultados de un ensayo para determinar la actividad negativa dominante de una mutación en un gen STK de uva (*Vitis vinifera*). El ensayo se lleva a cabo como se describe en el Ejemplo 3. R>L es la edición del R conservado a un L.

FIG. 9 proporciona los resultados de un ensayo para determinar la actividad negativa dominante de una mutación en un gen STK de frutilla silvestre (*Fragaria vesca*). El ensayo se lleva a cabo como se describe en el Ejemplo 3.

FIG. 10 proporciona los resultados de un ensayo para la actividad negativa dominante (dom) de dos mutaciones en un gen STK de mora (*Rubus occidentalis*). El ensayo se lleva a cabo como se describe en el Ejemplo 3. ELNAdel = delección de residuos consecutivos de ELNA; sp661 se refiere a una mutación generada por el espaciador 661 (sp661) como se muestra en la FIG. 4).

FIG. 11 proporciona los resultados de un ensayo para determinar la actividad negativa dominante de una mutación en un gen STK de palma aceitera (*Elaeis guineensis*). El ensayo se lleva a cabo como se describe en el Ejemplo 3.

FIGS. 12A–12B. **FIG. 12A** representa dos factores de transcripción SEPALLATA de tipo salvaje y dos factores de transcripción STK de tipo salvaje que interactúan en el ensayo de luciferasa, y **FIG. 12B** representa dos factores de transcripción SEPALLATA de tipo salvaje, un factor de transcripción STK mutado (tono más oscuro) y un factor de transcripción STK de tipo salvaje interactuando en el ensayo de luciferasa.

FIG. 13 muestra una alineación de STK de frutilla (*Fragaria vesca*) de tipo salvaje con STK mutante generado mediante el uso de un vector de corte como se describe en el Ejemplo 4.

DESCRIPCIÓN DETALLADA

La presente divulgación proporciona métodos para generar plantas que producen frutos con un fenotipo de desarrollo de fruto alterado. En algunos aspectos, los métodos proporcionan la

producción de frutas sin semillas o frutas con una simiente reducida. "Sin semillas" y/o "simiente reducida" se definen como un fruto que no tiene una semilla o un hoyo/drupa (sin drupa), o frutos con semillas en donde la semilla está alterada en comparación con una fruta de tipo silvestre de la misma especie de planta, debido a una reducción del tamaño de semilla, o al contenido reducido de lignina de semilla, o la formación reducida de endocarpio. La simiente reducida se define de la siguiente manera: contenido de lignina reducido en comparación con el tipo salvaje (lignina/mg de pared celular libre de proteína en endocarpio y/o cubierta de semilla; tamaño de semilla reducido en comparación con el tipo salvaje (volumen de semilla y/o drupa); y/o superficie de semilla/drupa más lisa en comparación con el tipo silvestre (área de superficie/volumen de semilla/drupa).

La divulgación proporciona un método para crear frutas sin semillas (donde una fruta sin semillas se define de modo de no tener una semilla o un hoyo) o frutas con semillas donde la semilla se puede comer debido al ablandamiento de la semilla o el hoyo, mediante una combinación de: análisis bioinformático, análisis evolutivo, edición de genes y cribado fenotípico. Primero, el análisis bioinformático se utiliza para determinar qué genes se expresan en las semillas en desarrollo mediante el examen de los datos de expresión génica (RNASeq, Microarray). Los genes que solo se expresan en la semilla en desarrollo y no en otras partes de la planta se seleccionan para el segundo paso, el análisis evolutivo. En este paso, las alineaciones de aminoácidos se usan para diagnosticar residuos bajo una selección de purificación positiva, una fuerte indicación de un papel crítico en el desarrollo de la semilla. En el tercer paso, los codones que codifican estos residuos seleccionados se editan para afectar el cambio no conservador en el aminoácido, lo que resulta en una variación en la función de la proteína. Estas posibles ediciones candidatas luego se criban fenotípicamente para seleccionar un fenotipo sin semillas.

Los consumidores valoran los cultivos sin semillas debido a la conveniencia y facilidad de preparación y consumo. Por ejemplo, la sandía sin semillas permite comer la sandía sin escupir las semillas, al igual que las uvas, naranjas, bananas, etc. También hay una fácil preparación; las bananas requerirían la eliminación de semillas antes de comerlas. Incluso donde las semillas son actualmente comestibles, como en las moras, el contenido sustancial de semillas es inaceptable para algunos consumidores. Estos beneficios se disfrutan en un número muy limitado de cultivos, porque hasta la fecha, el fenotipo sin semillas se ha sustentado en la captura de la variación que surge en las poblaciones silvestres y de reproducción, o en la complicada producción triploide que solo funciona en algunos cultivos. Los métodos de la presente invención permiten el desarrollo del fenotipo sin semillas en muchas, si no en todas, las especies de plantas.

Además de producir plantas con un fenotipo sin semillas, también es deseable reducir la "simiente" a través del direccionamiento al tamaño de la semilla o el grosor del endocarpio. La razón del bajo nivel de simiente en algunos cultivares de moras se debe a una percepción, aparentemente debido a la forma de la semilla y al grosor del endocarpio. El rasgo del grosor del endocarpio es un factor importante que contribuye a la simiente, según lo experimentado por el consumidor.

Los mecanismos genéticos subyacentes y la arquitectura genómica que dan lugar al fenotipo sin semilla, o de reducida simiente, no son bien comprendidos. Por lo tanto, existe una gran necesidad en la técnica de un método para crear un fenotipo sin semillas, en muchas, si no todas, las especies de plantas. Tal método permitiría a los consumidores tener la opción de consumir frutas o verduras sin semillas, o que tienen una percepción reducida de las semillas cuando se comen.

La presente divulgación proporciona composiciones, métodos y productos relacionados con la modificación de células vegetales para inducir la falta de semilla y/o tamaño reducido de semilla en el contexto de rasgos agronómicos mejorados. En algunas formas de realización, las células vegetales se modifican mediante los métodos de la presente divulgación para editar un solo nucleótido de uno o más genes objetivo endógenos, o para desencadenar una mutación puntual en la que cambia un único nucleótido, tal como C a T (o G a A), que da como resultado un codón que codifica un aminoácido diferente. En algunas formas de realización, las células vegetales se modifican por los métodos de la presente divulgación para proporcionar un polipéptido truncado, por ejemplo, un truncamiento C-terminal en un polipéptido de factor de transcripción de caja MADS clado AG. En algunas formas de realización, las células vegetales se modifican usando técnicas de edición de genes, que incluyen, sin limitación, sistemas CRISPR-Cas y/o editores de base asociados con un complejo CRISPR-Cas. En algunas formas de realización, la presente divulgación proporciona composiciones, métodos y productos (es decir, plantas editadas genéticamente) para modificar células vegetales usando técnicas de edición de bases dirigidas con precisión de la presente divulgación. En algunas formas de realización, la presente descripción proporciona métodos para realizar la edición dirigida de gen/genoma dentro de una célula vegetal y producir una planta, parte y célula editada genéticamente, que comprende componentes para la ingeniería dirigida de gen/genoma descrita en este documento. En otras formas de realización, la planta editada genéticamente está libre de transgenes. En la descripción, cuando se usa la frase "sustitución de un solo nucleótido", se entiende que pueden realizarse múltiples sustituciones de un solo nucleótido utilizando los métodos descritos en este documento. Los métodos enseñados en el presente documento pueden realizar sustituciones o delecciones dirigidas de uno o más nucleótidos.

La presente invención se describirá ahora a continuación con referencia a los dibujos y ejemplos adjuntos, en los que se muestran formas de realización de la invención. Esta descripción no pretende ser un catálogo detallado de todas las diferentes formas en que se puede implementar la invención, o de todas las características que se pueden agregar a la presente invención. Por ejemplo, las características ilustradas con respecto a una forma de realización pueden incorporarse en otras formas de realización, y las características ilustradas con respecto a una forma de realización particular pueden eliminarse de esa forma de realización. Por lo tanto, la invención contempla que en algunas formas de realización de la invención, cualquier característica o combinación de características expuestas en el presente documento puede excluirse u omitirse. Además, numerosas variaciones y adiciones a las diversas formas de realización sugeridas aquí serán evidentes para los expertos en la técnica en vista de la presente divulgación, que no se apartan de la presente invención. Por lo tanto, las siguientes descripciones pretenden ilustrar algunas formas de realización particulares de la invención, y no especificar exhaustivamente todas sus permutaciones, combinaciones y variaciones.

A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en este documento tienen el mismo significado que comúnmente entiende un experto en la técnica a la que pertenece esta invención. La terminología utilizada en la descripción de la invención en este documento tiene el propósito de describir formas de realización particulares solamente, y no pretende ser limitativa de la invención.

Todas las publicaciones, solicitudes de patentes, patentes y otras referencias citadas en este documento se incorporan por referencia en su totalidad para las enseñanzas pertinentes a la oración y/o el párrafo en donde se presenta la referencia.

A menos que el contexto indique lo contrario, se pretende específicamente que las diversas características de la invención descritas en el presente documento se puedan usar en cualquier combinación. Además, la presente invención también contempla que en algunas formas de realización de la invención, cualquier característica o combinación de características expuestas en el presente documento pueda excluirse u omitirse. Con fines de ilustración, si la memoria descriptiva establece que una composición comprende los componentes A, B y C, se pretende específicamente que cualquiera de A, B o C, o una combinación de los mismos, pueda omitirse y rechazarse en forma individual o en cualquier combinación.

El término "un" o "una" se refiere a una o más de esa entidad, es decir, puede referirse a un referente plural. De este modo, los términos "un" o "una", "uno o más" y "por lo menos uno" se usan

indistintamente en el presente documento. Además, la referencia a "un elemento" por el artículo indefinido "un" o "una" no excluye la posibilidad de que esté presente más de uno de los elementos, a menos que el contexto requiera claramente que haya uno y solo uno de los elementos .

Como se usa en esta memoria descriptiva, el término "y/o" se usa en esta divulgación para significar "y" u "o", a menos que se indique lo contrario, y abarca todas y cada una de las combinaciones posibles de uno o más de los elementos enumerados asociados, así como la falta de combinaciones cuando se interpreta en la alternativa ("o").

A lo largo de esta memoria descriptiva, a menos que el contexto requiera lo contrario, se entenderá que las palabras "comprender" o variaciones tales como "comprende" o "que comprende" implican la inclusión de un elemento o número entero o grupo de elementos o números enteros establecidos, pero no la exclusión de cualquier otro elemento o número entero o grupo de elementos o números enteros.

Como se usa en este documento, la frase de transición "que consiste esencialmente en" significa que el alcance de una reivindicación debe interpretarse para abarcar los materiales o pasos especificados en la reivindicación y aquellos que no afectan materialmente las características básicas y novedosas de la invención reivindicada. Por lo tanto, el término "que consiste esencialmente en" cuando se usa en una reivindicación de esta invención no pretende interpretarse como equivalente a "que comprende".

Como se usan en esta solicitud, los términos "alrededor de" y "aproximadamente" se usan como equivalentes. Todos los números utilizados en esta solicitud con o sin alrededor de/ aproximadamente están destinados a cubrir cualquier fluctuación normal apreciada por un experto en la técnica pertinente. En ciertas formas de realización, el término "aproximadamente" o "alrededor de" se refiere a un rango de valores que se encuentran dentro del 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1 %, o menos en cualquier dirección (mayor o menor que) del valor de referencia establecido, a menos que se indique lo contrario o sea evidente por el contexto (excepto donde dicho número excedería el 100% de un valor posible). Por lo tanto, el término "alrededor de", como se usa en el presente documento cuando se refiere a un valor medible tal como una cantidad o concentración y similares, pretende abarcar variaciones de $\pm 10\%$, $\pm 5\%$, $\pm 1\%$, $\pm 0,5\%$, o incluso $\pm 0,1\%$ del valor especificado, así como el valor especificado. Por ejemplo, "alrededor de X", donde X es el valor medible, debe incluir X, así como variaciones de $\pm 10\%$, $\pm 5\%$, $\pm 1\%$, $\pm 0,5\%$, o incluso $\pm 0,1\%$ de X. Un rango proporcionado aquí para un valor medible puede incluir cualquier otro rango y/o valor individual dentro de este.

Como se usa en este documento, las frases como "entre X e Y" y "entre alrededor de X e Y" deben interpretarse para incluir X e Y. Como se usa en este documento, frases como "entre alrededor de X e Y" significan "entre alrededor de X y alrededor de Y", y frases como "de alrededor de X a Y" significa "de alrededor de X a alrededor de Y".

La cita de los rangos de valores en el presente documento solo pretende servir como un método abreviado para referirse individualmente a cada valor separado que cae dentro del rango, a menos que se indique lo contrario en el presente documento, y cada valor por separado se incorpora a la memoria como si se mencionara individualmente en el presente documento. Por ejemplo, si se divulga el rango de 10 a 15, entonces también se divulgan 11, 12, 13 y 14.

Como se usa en este documento, los términos "aumentar", "aumento", "aumentado", "mejorar", "mejorado", "mejoramiento" y "mayor" (y sus variaciones gramaticales) describen una elevación de por lo menos alrededor de 5%, 10%, 15%, 20%, 25%, 50%, 75%, 100%, 150%, 200%, 300%, 400%, 500% o más en comparación con un control.

Como se usa en el presente documento, los términos "reducir", "reducido", "reducción", "menor", "disminuir" y "disminuido" (y sus variaciones gramaticales), describen, por ejemplo, una disminución de por lo menos alrededor de 5%, 10%, 15%, 20%, 25%, 35%, 50%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99,5%, 99,6%, 99,7%, 99,8%, 99,9%, o 100% en comparación con un control. En formas de realización particulares, la reducción puede producir ninguna o esencialmente ninguna (es decir, una cantidad insignificante, por ejemplo, inferior a alrededor de 10% o incluso 5%) de cantidad o actividad detectable. Por ejemplo, una planta que comprende una mutación en un gen de factor de transcripción de caja MADS clado AG como se describe aquí puede exhibir reducida producción de semillas (simiente) o reducida simiente, que se reduce en alrededor de 5% a alrededor de 75%, alrededor de 5% a alrededor de 80%, alrededor de 5% a alrededor de 85%, alrededor de 5% a alrededor de 90%, alrededor de 5% a alrededor de 95%, alrededor de 15% a alrededor de 80%, alrededor de 15% a alrededor de 85%, alrededor de 25% a alrededor de 75%, alrededor de 25% a alrededor de 80%, alrededor de 25% a alrededor de 85%, alrededor de 25% a alrededor de 90%, alrededor de 25% a alrededor de 95%, alrededor de 50% a alrededor de 80%, alrededor de 50% a alrededor de 85%, alrededor de 75% a alrededor de 80%, alrededor de 75% a alrededor de 85%, alrededor de 75% a alrededor de 90%, alrededor de 75% a alrededor de 95%, alrededor de 80% a alrededor de 85%, alrededor de 80% a alrededor de 95%, o cualquier rango o valor entre ellos, en comparación con una planta de control que no comprende la misma mutación.

Como se usa en el presente documento, los términos "expresar", "expresa", "expresado" o "expresión" y similares, con respecto a una molécula de ácido nucleico y/o una secuencia de nucleótidos (por ejemplo, ARN o ADN) indica que la molécula de ácido nucleico y/o una secuencia de nucleótidos se transcribe y, opcionalmente, se traduce. Por lo tanto, una molécula de ácido nucleico y/o una secuencia de nucleótidos pueden expresar un polipéptido de interés o, por ejemplo, un ARN no traducido funcional.

Los términos "célula huésped genéticamente diseñada", "célula huésped recombinante" y "cepa recombinante" se usan indistintamente en el presente documento, y se refieren a células huéspedes que han sido genéticamente diseñadas por los métodos de la presente divulgación. Por lo tanto, los términos incluyen una célula huésped (p. ej., bacterias, células de levadura, células fúngicas, CHO, células humanas, células vegetales, protoplastos derivados de plantas, callos, etc.) que ha sido genéticamente alterada, modificada o diseñada, de modo que exhibe un genotipo y/o fenotipo alterado, modificado o diferente (p. ej., cuando la modificación genética afecta secuencias de ácido nucleico codificantes), en comparación con la célula huésped natural de la que se deriva. Se entiende que los términos se refieren no solo a la célula huésped recombinante particular en cuestión, sino también a la progenie o progenie potencial de dicha célula huésped.

El término "genéticamente diseñado" puede referirse a cualquier manipulación del genoma de una célula huésped (por ejemplo, mediante la inserción o eliminación de ácidos nucleicos).

El término "fitomejoramiento de próxima generación" se refiere a una gran cantidad de herramientas y metodologías de fitomejoramiento que están disponibles para el productor agrícola de hoy. Una característica distintiva clave del fitomejoramiento de próxima generación es que el productor agrícola ya no se limita a depender de la variación fenotípica observada, para inferir las causas genéticas subyacentes de un rasgo determinado. En cambio, el fitomejoramiento de próxima generación puede incluir la utilización de marcadores moleculares y selección asistida por marcadores (MAS, conforme a sus siglas en inglés), de modo que el productor agrícola pueda observar directamente el movimiento de alelos y elementos genéticos de interés de una planta en la población reproductora a otra, y no se limita simplemente a observar el fenotipo. Además, los métodos de fitomejoramiento de próxima generación no se limitan a utilizar la variación genética natural que se encuentra dentro de una población de plantas. Por el contrario, el productor agrícola que utiliza la metodología de fitomejoramiento de próxima generación puede acceder a una serie de herramientas modernas de ingeniería genética que alteran/cambian/editan directamente la arquitectura genética subyacente de la planta de una manera dirigida, a fin de generar un rasgo fenotípico de interés. En algunos aspectos, las plantas criadas con una metodología de fitomejoramiento de próxima generación son indistinguibles de una planta que fue criada

de manera tradicional, ya que la planta de producto final resultante podría desarrollarse teóricamente por cualquiera de estos métodos. En aspectos particulares, una metodología de fitomejoramiento de próxima generación puede dar como resultado una planta que comprende: una modificación genética que es una eliminación de cualquier tamaño (por ejemplo, que resulta en un truncamiento); una modificación genética que es una sustitución de un solo par de bases; una modificación genética que es una introducción de secuencias de ácido nucleico del interior del conjunto de genes naturales de la planta (por ejemplo, cualquier planta que pueda cruzarse o reproducirse con una planta de interés) o de la edición de secuencias de ácido nucleico en una planta para que corresponden a una secuencia que se sabe que ocurre en el acervo genético natural de la planta; y descendencia de dichas plantas.

El término "fitomejoramiento tradicional" se refiere a la utilización de la variación natural que se encuentra dentro de una población de plantas como fuente de alelos y variantes genéticas que imparten un rasgo de interés a una planta determinada. Los métodos de reproducción tradicionales hacen uso de procedimientos de cruzamiento que dependen en gran medida de la variación fenotípica observada para inferir la asociación alelo causal. Es decir, el fitomejoramiento tradicional se basa en observaciones del fenotipo expresado de una planta determinada para inferir la causa genética subyacente. Estas observaciones se utilizan para informar el procedimiento de fitomejoramiento a fin de trasladar la variación alélica al germoplasma de interés. Además, el fitomejoramiento tradicional también se ha caracterizado por comprender técnicas de mutagénesis al azar, que pueden usarse para introducir la variación genética en un germoplasma dado. Estas técnicas de mutagénesis al azar pueden incluir procedimientos de mutagénesis sobre la base de radiación y/o químicos. En consecuencia, una característica clave del fitomejoramiento tradicional es que el productor agrícola no utiliza una herramienta de ingeniería genética que altere/modifique/edite directamente la arquitectura genética subyacente de la planta de manera específica, a fin de introducir la diversidad genética y generar un rasgo fenotípico de interés.

Los términos "polinucleótido", "ácido nucleico" y "secuencia de nucleótidos", usados indistintamente en el presente documento, se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sea ribonucleótidos o desoxirribonucleótidos, o análogos de los mismos. Este término se refiere a la estructura primaria de la molécula y, por lo tanto, incluye ADN bicatenario y monocatenario, así como ARN bicatenario y monocatenario. Este término incluye, sin limitación, ADN o ARN monocatenario, bicatenario o multicatenario, ADN genómico, ADNc, híbridos de ADN-ARN o un polímero que comprende bases de purina y pirimidina u otras bases de nucleótidos naturales, química o bioquímicamente modificadas, no naturales o derivadas. También incluye ácidos

nucleicos modificados tales como ácidos nucleicos metilados y/o tapados, ácidos nucleicos que contienen bases modificadas, modificaciones del esqueleto y similares. "Oligonucleótido" generalmente se refiere a polinucleótidos de entre alrededor de 5 y alrededor de 100 nucleótidos de ADN monocatenario o bicatenario. Sin embargo, a los fines de esta divulgación, no existe un límite superior para la longitud de un oligonucleótido. Los oligonucleótidos también se conocen como "oligómeros" u "oligos", y pueden aislarse de genes o sintetizarse químicamente mediante métodos conocidos en la técnica. Debe entenderse que los términos "polinucleótido", "ácido nucleico" y "secuencia de nucleótidos" incluyen, según corresponda a las formas de realización que se describen, polinucleótidos monocatenarios (tal como de sentido o antisentido) y bicatenarios.

Los términos "péptido", "polipéptido" y "proteína" se usan indistintamente en el presente documento, y se refieren a una forma polimérica de aminoácidos de cualquier longitud, que puede incluir aminoácidos codificados y no codificados, aminoácidos modificados química o bioquímicamente o derivados, y polipéptidos que tienen esqueletos de péptidos modificados.

Una secuencia de nucleótidos "heteróloga" o "recombinante" es una secuencia de nucleótidos que no está asociada de forma natural con una célula huésped en la que se introduce, incluidas múltiples copias no naturales de una secuencia de nucleótidos natural.

Una secuencia de ácido nucleico, secuencia de nucleótidos, polipéptido o aminoácidos "nativa" o de "tipo salvaje" se refiere a una secuencia de ácido nucleico, secuencia de nucleótidos, polipéptido o secuencia de aminoácidos natural o endógena. Así, por ejemplo, un "gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje" es un gen del factor de transcripción de caja MADS clado AG que se produce naturalmente o es endógeno en el organismo de referencia, por ejemplo, una planta *Rubus*, una planta *Prunus* y similares.

Como se usa en el presente documento, el término "heterocigoto" se refiere a un estado genético en el que diferentes alelos residen en los loci (lugares) correspondientes en cromosomas homólogos.

Como se usa en el presente documento, el término "homocigoto" se refiere a un estado genético en el que alelos idénticos residen en los loci correspondientes en cromosomas homólogos.

Como se usa en el presente documento, el término "alelo" se refiere a uno de dos o más nucleótidos o secuencias de nucleótidos diferentes que se producen en un locus específico.

Un "alelo nulo" es un alelo no funcional causado por una mutación genética que resulta en una falta total de producción de la proteína correspondiente o produce una proteína que no es funcional.

Una "mutación negativa dominante" es una mutación que produce un producto genético alterado (por ejemplo, que tiene una función aberrante en relación con el tipo salvaje), cuyo producto génico afecta negativamente la función del producto génico o alelo de tipo salvaje. Por ejemplo, una "mutación negativa dominante" puede bloquear una función del producto del gen de tipo salvaje. Una mutación negativa dominante también puede denominarse "mutación antimórfica".

Una "mutación semidominante" se refiere a una mutación en la cual la penetración del fenotipo en un organismo heterocigoto es menor que la observada para un organismo homocigoto.

Una "mutación de pérdida de función débil" es una mutación que da como resultado un producto génico que tiene una función parcial o una función reducida (parcialmente inactivada) en comparación con el producto génico de tipo salvaje.

Una "mutación hipomórfica" es una mutación que resulta en una pérdida parcial de la función del gen, que puede ocurrir a través de una expresión reducida (p. ej., proteína reducida y/o ARN reducido) o un rendimiento funcional reducido (p. ej., actividad reducida), pero no la completa pérdida de función/actividad. Un alelo "hipomórfico" es un alelo semifuncional causado por una mutación genética que resulta en la producción de la proteína correspondiente que funciona en cualquier lugar entre 1% y 99% de la eficiencia normal.

Como se usa en el presente documento, los términos "alelo deseado", "alelo objetivo" y/o "alelo de interés" se usan indistintamente para referirse a un alelo asociado con un rasgo deseado. En algunas formas de realización, un alelo deseado puede estar asociado con un aumento o una disminución (en relación con un control) en un rasgo dado, de acuerdo con la naturaleza del fenotipo deseado.

Un marcador está "asociado con" un rasgo cuando dicho rasgo está unido a él, y cuando la presencia del marcador es un indicador de si y/o en qué medida el rasgo o la forma del rasgo deseado ocurrirá en una planta/germoplasma que comprende el marcador. De manera similar, un marcador está "asociado con" un intervalo de cromosoma o alelo cuando está unido a él, y cuando la presencia del marcador es un indicador de si el intervalo de cromosoma o alelo está presente en una planta/germoplasma que comprende el marcador.

Como se usan en este documento, los términos "retrocruzamiento/retrocruce" y "retrocruzar" se refieren al proceso mediante el cual una planta de progenie se cruza de nuevo con uno de sus progenitores una o más veces (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, etc.). En un esquema de retrocruzamiento, el progenitor "donante" se refiere a la planta parental con el gen o locus deseado para ser introgresado. El progenitor "receptor" (usado una o más veces) o el progenitor "recurrente"

(usado dos o más veces) se refiere a la planta parental en la que el gen o locus está siendo introgresado. Por ejemplo, ver Ragot, M. et al. *Marker-assisted Backcrossing: A Practical Example*, in *TECHNIQUES ET UTILISATIONS DES MARQUEURS MOLECULAIRES LES COLLOQUES*, Vol. 72, pp. 45–56 (1995); and Openshaw et al., *Marker-assisted Selection in Backcross Breeding*, in *PROCEEDINGS OF THE SYMPOSIUM "ANALYSIS OF MOLECULAR MARKER DATA,"* pp. 41–43 (1994). El cruce inicial da lugar a la generación F1. El término "BC1" se refiere al segundo uso del progenitor recurrente, "BC2" se refiere al tercer uso del progenitor recurrente, y así sucesivamente.

Como se usa en el presente documento, los términos "cruzado" o "cruce" se refieren a la fusión de gametos mediante polinización para producir progenie (por ejemplo, células, semillas o plantas). El término abarca tanto los cruces sexuales (la polinización de una planta por otra) como la autofertilización (autopolinización, por ejemplo, cuando el polen y el óvulo son de la misma planta). El término "cruce" se refiere al acto de fusionar gametos a través de la polinización para producir progenie.

Como se usa en el presente documento, los términos "introgresión", "introgresar" e "introgresado" se refieren tanto a la transmisión natural como artificial de un alelo deseado o combinación de alelos deseados de un locus genético o loci genéticos, de un fondo genético a otro. Por ejemplo, un alelo deseado en un locus específico se puede transmitir a por lo menos una progenie a través de un cruce sexual entre dos progenitores de la misma especie, donde por lo menos uno de los progenitores tiene el alelo deseado en su genoma. Alternativamente, por ejemplo, la transmisión de un alelo puede ocurrir por recombinación entre dos genomas donantes, por ejemplo, en un protoplasto fusionado, donde por lo menos uno de los protoplastos donantes tiene el alelo deseado en su genoma. El alelo deseado puede ser un alelo seleccionado de un marcador, un QTL, un transgén o similar. La descendencia que comprende el alelo deseado puede retrocruzarse una o más veces (por ejemplo, 1, 2, 3, 4 o más veces) con una línea que tiene un fondo genético deseado, seleccionando el alelo deseado, con el resultado de que el alelo deseado se fija en el fondo genético deseado. Por ejemplo, un marcador asociado con un aumento del rendimiento en condiciones sin estrés hídrico puede ser introgresado de un donante a un progenitor recurrente que no comprende el marcador y no exhibe un aumento del rendimiento en condiciones sin estrés hídrico. La descendencia resultante podría ser retrocruzada una o más veces y seleccionada hasta que la progenie posea los marcadores genéticos asociados con un mayor rendimiento en condiciones sin estrés hídrico en el fondo parental recurrente.

Como se usa en el presente documento, el término "genotipo" se refiere a la constitución genética de un individuo (o grupo de individuos) en uno o más loci genéticos, en contraste con el rasgo observable y/o detectable y/o manifestado (el fenotipo). El genotipo se define por el alelo(s) de uno o más loci conocidos que el individuo ha heredado de sus progenitores. El término genotipo puede usarse para referirse a la constitución genética de un individuo en un solo locus, en múltiples loci, o más generalmente, el término genotipo puede usarse para referirse a la composición genética de un individuo para todos los genes en su genoma. Los genotipos se pueden caracterizar indirectamente, por ejemplo, usando marcadores, y/o se pueden caracterizar directamente por secuenciación de ácido nucleico.

Como se usa en el presente documento, el término "germoplasma" se refiere al material genético de un individuo (por ejemplo, una planta), un grupo de individuos (por ejemplo, una línea de planta, variedad o familia) o un clon derivado de una línea, variedad, especie o cultivo. El germoplasma puede ser parte de un organismo o de una célula, o puede estar separado del organismo o la célula. En general, el germoplasma proporciona material genético con una composición genética específica que proporciona una base para algunas o todas las cualidades hereditarias de un organismo o cultivo celular. Como se usa en el presente documento, el germoplasma incluye células, semillas o tejidos a partir de los cuales se pueden cultivar nuevas plantas, así como partes de plantas que se pueden cultivar hasta una planta completa (por ejemplo, hojas, tallos, brotes, raíces, polen, células, etc.).

Como se usa en este documento, los términos "cultivar" y "variedad" se refieren a un grupo de plantas similares, que por características estructurales o genéticas y/o de rendimiento pueden distinguirse de otras variedades dentro de la misma especie.

Como se usa en el presente documento, el término "por lo menos una porción" o "fragmento" de un ácido nucleico o polipéptido significa una porción que tiene las características de tamaño mínimo de tales secuencias, o cualquier fragmento más grande de la molécula de longitud completa, hasta e incluyendo la molécula de longitud completa. Un fragmento de un polinucleótido de la descripción puede codificar una porción biológicamente activa de un elemento regulador genético. Se puede preparar una porción biológicamente activa de un elemento regulador genético aislando una porción de uno de los polinucleótidos de la divulgación que comprende el elemento regulador genético y evaluando la actividad como se describe en este documento. En algunas formas de realización, una "porción" puede codificar una región objetivo para producir un fenotipo deseado (p. ej., desarrollo alterado del fruto; p. ej., falta de semillas o reducida simiente). Por lo tanto, con

respecto a los ácidos nucleicos, el término "fragmento" o "porción" se refiere a un ácido nucleico que es reducido en longitud relativo (por ejemplo, reducido en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 400, 450, 500, 550, 600, 650, 660, 670, 680, 690 o 700 o más nucleótidos, o cualquier rango o valor de ellos) a un ácido nucleico de referencia, y que comprende, consiste esencialmente en y/o consiste en una secuencia de nucleótidos contiguos idéntica o casi idéntica (por ejemplo, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% idéntica) a una porción correspondiente del ácido nucleico de referencia. Tal fragmento de ácido nucleico puede, cuando sea apropiado, incluirse en un polinucleótido más grande del cual es un constituyente. Como ejemplo, una secuencia de repetición de ácido nucleico guía de esta invención puede comprender una "porción" de una secuencia de repetición CRISPR–Cas de tipo salvaje (por ejemplo, una repetición CRISPR–Cas de tipo salvaje; por ejemplo, una repetición del sistema CRISPR Cas de, por ejemplo, un Cas9, Cas12a (Cpf1), Cas12b, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12g, Cas12h, Cas12i, C2c4, C2c5, C2c8, C2c9, C2c10, Cas14a, Cas14b, y/o a Cas14c, y similares). En algunas formas de realización, un fragmento de ácido nucleico puede comprender, consistir esencialmente en, o consistir en alrededor de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, o 650 o más nucleótidos consecutivos o cualquier rango o valor entre ellos, de un ácido nucleico que codifica un factor de transcripción SEEDSTICK (STK), opcionalmente, un fragmento de ácido nucleico STK puede ser de alrededor de 50 nucleótidos a alrededor de 300 nucleótidos de longitud, alrededor de 50 nucleótidos a alrededor de 350 nucleótidos de longitud, alrededor de 50 nucleótidos a alrededor de 400 nucleótidos de longitud, alrededor de 50 nucleótidos a alrededor de 450 nucleótidos de longitud, alrededor de 50 nucleótidos a alrededor de 500 nucleótidos de longitud, alrededor de 50 nucleótidos a alrededor de 600 nucleótidos de longitud, alrededor de 100 nucleótidos a alrededor de 300 nucleótidos de longitud, alrededor de 100 nucleótidos a alrededor de 350 nucleótidos de longitud, alrededor de 100 nucleótidos a alrededor de 400 nucleótidos de longitud, alrededor de 100 nucleótidos a alrededor de 450 nucleótidos de longitud, alrededor de 100 nucleótidos a alrededor de 500 nucleótidos de longitud, o alrededor de 100 nucleótidos a alrededor de 600 nucleótidos de longitud, o cualquier rango o valor entre ellos.

En algunas formas de realización, un fragmento de ácido nucleico de un gen *STK* puede ser el resultado de una delección de nucleótidos del extremo 3', el extremo 5', y/o del interior de un gen que codifica una proteína *STK*. En algunas formas de realización, una delección de una porción de un gen que codifica una proteína *STK* puede comprender una delección de una porción de nucleótidos consecutivos del extremo 5', el extremo 3' o del interior, por ejemplo, de una secuencia de nucleótidos que tiene por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**. En algunas formas de realización, una delección de una porción de un gen *STK* puede comprender delección de una porción de nucleótidos consecutivos del extremo 3' de una secuencia de nucleótidos que tiene por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**. En algunas formas de realización, una delección de una porción de un gen *STK* puede comprender una delección de una porción de nucleótidos consecutivos del extremo 3' de una secuencia de nucleótidos que tiene por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** de alrededor de 1 nucleótido a alrededor de 300 nucleótidos consecutivos o más (por ejemplo, alrededor de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 130, 140, 150, 175, 200, 225, 250, o 300 o más nucleótidos consecutivos, o cualquier rango o valor entre ellos, opcionalmente, alrededor de 4 nucleótidos consecutivos a alrededor de 150 nucleótidos consecutivos). En algunas formas de realización, dicha delección puede producir un alelo nulo, que cuando está en una planta, puede producir un fenotipo de desarrollo alterado del fruto. En algunas formas de realización, dicha delección puede ser un alelo dominante negativo, un alelo semidominante, un alelo de pérdida de función débil, un alelo nulo o una mutación hipomórfica, que cuando se encuentra en una planta, puede producir un fenotipo de desarrollo de fruto alterado en la planta, en donde el desarrollo alterado del fruto produce, por ejemplo, falta de semillas o reducción de simiente.

De manera similar, una porción de un polipéptido puede ser 4 aminoácidos, 5 aminoácidos, 6 aminoácidos, 7 aminoácidos, y así sucesivamente, subiendo al polipéptido de longitud completa. La longitud de la porción que se utilizará dependerá de la aplicación particular. Una porción de un polipéptido útil como epítopo puede ser tan corta como 4 aminoácidos. Una porción de un polipéptido

que realiza la función del polipéptido de longitud completa generalmente será más larga que 4 aminoácidos. En algunas formas de realización, un fragmento de un polipéptido o polinucleótido comprende por lo menos 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, o 99% de la longitud total del polipéptido o polinucleótido de referencia.

Además, con respecto a polipéptidos, el término "fragmento" o "porción" se refiere a un polipéptido que tiene reducida longitud en relación con un polipéptido de referencia, y que comprende, consiste esencialmente en y/o consiste en una secuencia de aminoácidos de aminoácidos contiguos idéntica o casi idéntica (por ej., 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% idéntica) a una porción correspondiente del polipéptido de referencia. Dicho fragmento de polipéptido puede incluirse cuando sea apropiado en un polipéptido más grande del cual es constituyente. En algunas formas de realización, el fragmento de polipéptido comprende, consiste esencialmente en o consiste en por lo menos alrededor de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400 o más aminoácidos consecutivos de un polipéptido de referencia. En algunas formas de realización, un fragmento de polipéptido puede comprender, consistir esencialmente en o consistir en alrededor de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 166, 167, 168, 169, 170, 171, 172 o más residuos de aminoácidos consecutivos (o cualquier rango o valor entre ellos) de, por ejemplo, un polipéptido STK (por ej., un fragmento o una porción de un polipéptido que tiene por lo menos 80% de identidad de secuencia con cualquiera de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**).

En algunas formas de realización, una "porción" puede estar relacionada con el número de aminoácidos que se deletan de un polipéptido. Así, por ejemplo, una "porción" deletada de, por ejemplo, un polipéptido STK puede comprender por lo menos un residuo de aminoácido (por ejemplo, por lo menos 1, o por lo menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 o más residuos de aminoácidos consecutivos) deletados de un polipéptido que tiene por lo menos 80% de identidad de secuencia con cualquiera de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, opcionalmente, por lo menos un residuo de aminoácido a alrededor de 50 residuos de aminoácidos consecutivos. En algunas formas de realización, una

deleción de una porción de una proteína STK puede comprender una deleción de una porción de residuos de aminoácidos consecutivos del término N o C de, o dentro de un polipéptido que tiene por lo menos 80% de identidad de secuencia con cualquiera de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, opcionalmente, una deleción de por lo menos residuo de aminoácido a alrededor de 50 residuos de aminoácidos consecutivos (por ejemplo, alrededor de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, o 50 nucleótidos consecutivos). En algunas formas de realización, dicha deleción en un polipéptido puede dar como resultado un alelo nulo, que cuando está en una planta, la planta exhibe un desarrollo alterado del fruto como se describe en el presente documento en comparación con una planta que no comprende dicha deleción. En algunas formas de realización, dicha deleción puede ser un alelo dominante negativo, un alelo semidominante, un alelo de pérdida de función débil, un alelo nulo o una mutación hipomórfica, que cuando se encuentra en una planta, la planta exhibe un desarrollo alterado del fruto en comparación con una planta que no comprende la mutación.

Como se usa en el presente documento, el término "gen" se refiere a cualquier segmento de ADN asociado con una función biológica. Por lo tanto, los genes incluyen, sin limitación, secuencias de codificación y/o las secuencias reguladoras requeridas para su expresión. Los genes también pueden incluir segmentos de ADN no expresados que, por ejemplo, forman secuencias de reconocimiento para otras proteínas. Los genes pueden obtenerse de una diversidad de fuentes, incluida la clonación de una fuente de interés, o la síntesis a partir de información de secuencia conocida o predicha, y pueden incluir secuencias diseñadas para tener los parámetros deseados.

Como se usa en el presente documento, el término "endógeno" o "gen endógeno" se refiere al gen natural, en la ubicación en la que se encuentra naturalmente dentro del genoma de la célula huésped.

Como se usa en el presente documento, el término "exógeno" se refiere a una sustancia que proviene de alguna fuente distinta de su fuente nativa. Por ejemplo, los términos "proteína exógena" o "gen exógeno" se refieren a una proteína o un gen de una fuente no nativa, y que se ha suministrado artificialmente a un sistema biológico. Como se usa en el presente documento, el término "exógeno" se usa indistintamente con el término "heterólogo", y se refiere a una sustancia que proviene de alguna fuente distinta de su fuente nativa.

Como se usa en este documento, el término "heterólogo" se refiere a una sustancia que proviene de alguna fuente o ubicación distinta de su fuente o ubicación nativa. En algunas formas de

realización, el término "ácido nucleico heterólogo" se refiere a una secuencia de ácido nucleico que no se encuentra naturalmente en el organismo particular. Por ejemplo, el término "promotor heterólogo" puede referirse a un promotor que ha sido tomado de un organismo fuente y utilizado en otro organismo, en el que el promotor no se encuentra naturalmente. Sin embargo, el término "promotor heterólogo" también puede referirse a un promotor que se encuentra dentro del mismo organismo fuente, pero que simplemente se ha trasladado a una nueva ubicación, en la que dicho promotor normalmente no se encuentra.

Las secuencias de genes heterólogos pueden introducirse en una célula objetivo utilizando un "vector de expresión", que puede ser un vector de expresión eucariota, por ejemplo, un vector de expresión vegetal. Los métodos utilizados para construir vectores son bien conocidos por el experto en la técnica, y se describen en diversas publicaciones. En particular, las técnicas para construir vectores adecuados, incluida una descripción de los componentes funcionales tales como promotores, mejoradores, señales de terminación y poliadenilación, marcadores de selección, orígenes de replicación y señales de empalme, se revisan en la técnica anterior. Los vectores pueden incluir, entre otros, vectores plasmídicos, fagémidos, cósmidos, cromosomas artificiales/minicromosomas (p. ej., ACE) o vectores virales tales como baculovirus, retrovirus, adenovirus, virus asociado a adeno, virus herpes simple, retrovirus, bacteriófagos. Los vectores de expresión eucarióticos típicamente contendrán también secuencias procariotas que facilitan la propagación del vector en bacterias tales como un origen de replicación y genes de resistencia a antibióticos para la selección en bacterias. Una diversidad de vectores de expresión eucariotas, que contienen un sitio de clonación en el que un polinucleótido puede unirse operativamente, son bien conocidos en la técnica, y algunos están disponibles comercialmente en compañías tales como Stratagene, La Jolla, California; Invitrogen, Carlsbad, California; Promega, Madison, Wisconsin o BD Biosciences Clontech, Palo Alto, California. En una forma de realización, el vector de expresión comprende por lo menos una secuencia de ácido nucleico que es una secuencia reguladora necesaria para la transcripción y traducción de secuencias de nucleótidos que codifican un péptido/polipéptido/proteína de interés.

Como se usa en el presente documento, el término "homólogo" u "ortólogo" se conoce en la técnica y se refiere a secuencias relacionadas que comparten un ancestro común o un miembro de la familia, y se determinan en función del grado de identidad de secuencia. Las secuencias homólogas son ortólogas cuando se infiere que descenden de la misma secuencia ancestral separadas por un evento de especiación: cuando una especie diverge en dos especies separadas, se dice que las copias de un solo gen en las dos especies resultantes son ortólogas. Los "ortólogos", o "genes ortólogos",

son genes de diferentes especies que se originaron por el descenso vertical de un solo gen del último ancestro común. Los términos "homología", "homólogo", "sustancialmente similar" y "correspondiente sustancialmente" se usan indistintamente en el presente documento. Se refieren a fragmentos de ácido nucleico en los que los cambios en una o más bases de nucleótidos no afectan la capacidad del fragmento de ácido nucleico para mediar en la expresión génica o producir un cierto fenotipo. Estos términos también se refieren a modificaciones de los fragmentos de ácido nucleico de la divulgación presente, tales como la delección o inserción de uno o más nucleótidos que no alteran sustancialmente las propiedades funcionales del fragmento de ácido nucleico resultante en relación con el fragmento inicial no modificado. Por lo tanto, se entiende, como apreciarán los expertos en la técnica, que la divulgación abarca más que las secuencias ejemplares específicas. Estos términos describen la relación entre un gen encontrado en una especie, subespecie, variedad, cultivar o cepa y el gen correspondiente o equivalente en otra especie, subespecie, variedad, cultivar o cepa. Para los fines de esta descripción, se comparan secuencias homólogas. Se cree o se sabe que las "secuencias homólogas" o los "homólogos" u "ortólogos" están relacionados funcionalmente. Una relación funcional puede indicarse de cualquiera de varias maneras, que incluyen, sin limitación: (a) el grado de identidad de secuencia y/o (b) la misma o similar función biológica. Preferentemente, se indican tanto (a) como (b). La homología se puede determinar usando programas de soporte lógico fácilmente disponibles en la técnica, tales como los discutidos en *Current Protocols in Molecular Biology* (F. M. Ausubel et al., Eds., 1987). Suplemento 30, sección 7.718, Tabla 7.71. Algunos programas de alineación son MacVector (Oxford Molecular Ltd, Oxford, Reino Unido), ALIGN Plus (Scientific and Educational Software, Pennsylvania) y AlignX (Vector NTI, Invitrogen, Carlsbad, CA). Otro programa de alineación es Sequencher (Gene Codes, Ann Arbor, Michigan), que utiliza parámetros predeterminados, y MUSCLE (Comparación de secuencia múltiple por Log-Expectation; un soporte lógico de computadora con licencia como dominio público).

Como se usa en el presente documento, el término "cambio de nucleótido" se refiere, por ejemplo, a la sustitución, delección y/o inserción de nucleótido, como se entiende bien en la técnica. Por ejemplo, las mutaciones contienen alteraciones que producen sustituciones de nucleótidos individuales, sustituciones silenciosas, adiciones o delecciones, pero no alteran las propiedades o actividades de la proteína codificada o cómo se hacen las proteínas. Alternativamente, las mutaciones que dan como resultado sustituciones, adiciones o delecciones de nucleótidos (p. ej., una delección en marco, delección fuera del marco) pueden alterar las propiedades o actividades de la proteína

codificada, por ejemplo, dar como resultado un codón de parada no natural o prematuro, un truncamiento C-terminal y/o N-terminal.

Como se usa en el presente documento, el término "modificación de proteína" se refiere, por ejemplo, a la modificación, sustitución, delección y/o inserción de aminoácidos, como se entiende bien en la técnica, que puede dar como resultado, por ejemplo, un codón de parada prematuro y/o un truncamiento C-terminal, un truncamiento N-terminal u otra alteración que afecta la función de un polipéptido (por ejemplo, una modificación STK que afecta la capacidad del polipéptido STK para interactuar con otros polipéptidos STK y/o con uno o más polipéptidos SEPALLATA (SEP1 , SEP2, SEP3, SEP4)).

Como se usa en el presente documento, el término "optimización de codones" se refiere al uso de codones de un ADN o ARN que se adapta al de una célula u organismo de interés para mejorar la tasa de transcripción de dicho ácido nucleico recombinante en la célula o el organismo de interés. El experto tiene conocimiento del hecho de que un ácido nucleico objetivo puede modificarse en una posición debido a la degeneración del codón, mientras que esta modificación aún conducirá a la misma secuencia de aminoácidos en esa posición después de la traducción, que se logra mediante la optimización de codón para tener en cuenta el uso de codones específicos de especie de una célula u organismo objetivo. En algunas formas de realización, los polinucleótidos, construcciones de ácido nucleico, casetes de expresión y/o vectores de los sistemas de edición de la invención (por ejemplo, que comprenden/codifican un dominio de unión de ADN específico de secuencia (p. ej., un dominio de unión de ADN específico de secuencia de una endonucleasa guiada por polinucleótido, una nucleasa de dedo de zinc, una nucleasa efectora activadora de la transcripción (TALEN, conforme a sus siglas en inglés), una proteína argonauta y/o una endonucleasa CRISPR-Cas (por ejemplo, proteína efectora CRISPR-Cas) (por ejemplo, una proteína efectora CRISPR-Cas tipo I, una proteína efectora CRISPR-Cas tipo II, una proteína efectora CRISPR-Cas Tipo III, una proteína efectora CRISPR-Cas tipo IV, una proteína efectora CRISPR-Cas tipo V o una proteína efectora CRISPR-Cas tipo VI)), una nucleasa (p. ej., una endonucleasa (p. ej., Fok1), una endonucleasa guiada por polinucleótido, una endonucleasa CRISPR-Cas (p. ej., proteína efectora CRISPR-Cas), una nucleasa de dedo de zinc y/o una nucleasa efectora de tipo activador de la transcripción (TALEN), proteínas/dominios de desaminasas (p. ej., adenina desaminasa, citosina desaminasa), un polinucleótido que codifica una proteína o dominio de transcriptasa inversa, un polinucleótido que codifica un polipéptido de exonucleasa 5'-3' y/o polipéptidos de afinidad, etiquetas peptídicas, etc.) pueden ser codones optimizados para la expresión en una planta. En algunas formas de realización,

los ácidos nucleicos, polinucleótidos, casetes de expresión y/o vectores optimizados por codón de la invención tienen alrededor de 70% a alrededor de 99,9% (por ejemplo, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5%, 99,9% o 100%) de identidad o más a los ácidos nucleicos, polinucleótidos, casetes de expresión y/o vectores de referencia que no han sido optimizados por codón.

Como se usa en el presente documento, el término "de origen natural/natural", tal como se aplica a un ácido nucleico, un polipéptido, una célula o un organismo, se refiere a un ácido nucleico, polipéptido, una célula o un organismo que se encuentra en la naturaleza. El término "natural" puede referirse a un gen o secuencia derivado de una fuente natural. Por lo tanto, para los fines de esta divulgación, una secuencia "no natural" es una secuencia que se ha sintetizado, mutado, diseñado, editado o modificado de otro modo para tener una secuencia diferente de las secuencias naturales conocidas. En algunas formas de realización, la modificación puede estar en el nivel de la proteína (por ejemplo, modificación de aminoácidos, por ejemplo, sustituciones, adiciones, deleciones). En otras formas de realización, la modificación puede ser en el nivel de ADN (por ejemplo, modificaciones de nucleótidos, por ejemplo, sustituciones, adiciones, deleciones). "No natural" se refiere a un ácido nucleico, un polipéptido, una célula o un organismo que no se encuentra en la naturaleza (por ejemplo, una mutación no natural) o que no se encuentra en la naturaleza en la combinación específica descripta.

El término "mutación" se refiere a mutaciones puntuales (p. ej., sentido erróneo o sin sentido, o inserciones o deleciones de pares de bases individuales que dan como resultado desplazamientos de marco), inserciones, deleciones y/o truncamientos. Cuando la mutación es una sustitución de un residuo dentro de una secuencia de aminoácidos con otro residuo, o una delección o inserción de uno o más residuos dentro de una secuencia, las mutaciones se describen típicamente identificando el residuo original seguido de la posición del residuo dentro la secuencia y por la identidad del residuo recientemente sustituido. Un truncamiento puede incluir un truncamiento en el extremo C-terminal de un polipéptido o en el extremo N-terminal de un polipéptido. Un truncamiento de un polipéptido puede ser el resultado de una delección del extremo 5' o 3' correspondiente del gen que codifica el polipéptido. En algunas formas de realización, un truncamiento C-terminal es el resultado de una mutación tal como una delección, adición o sustitución de base que produce un codón de parada prematuro.

Como se usa en el presente documento, las frases "construcción recombinante", "construcción de expresión", "construcción quimérica", "construcción" y "construcción de ADN recombinante" se usan indistintamente en el presente documento. Una construcción recombinante comprende una combinación artificial de fragmentos de ácido nucleico, por ejemplo, secuencias reguladoras y de codificación que no se encuentran juntas en la naturaleza. Por ejemplo, una construcción quimérica puede comprender secuencias reguladoras y secuencias de codificación que se derivan de diferentes fuentes, o secuencias reguladoras y secuencias de codificación derivadas de la misma fuente, pero dispuestas de manera diferente de la encontrada en la naturaleza. Tal construcción puede usarse por sí misma o puede usarse junto con un vector. Si se usa un vector, entonces la elección del vector depende del método que se usará para transformar las células huéspedes, como es bien conocido por los expertos en la técnica. Por ejemplo, se puede usar un vector plasmídico. El experto en la técnica conoce bien los elementos genéticos que deben estar presentes en el vector para transformar, seleccionar y propagar con éxito las células huéspedes que comprenden cualquiera de los fragmentos de ácido nucleico aislados de la divulgación. El experto en la técnica también reconocerá que diferentes eventos de transformación independientes darán como resultado diferentes niveles y patrones de expresión (Jones et al., (1985) EMBO J. 4: 2411–2418; De Almeida et al., (1989) Mol. Gen. Genetics 218: 78–86) y, por lo tanto, deben cribarse múltiples eventos para obtener líneas que muestren el nivel de expresión y el patrón deseados. Dicho cribado puede realizarse mediante análisis Southern de ADN, análisis Northern de expresión de ARNm, análisis de inmunotransferencia de expresión de proteína, o análisis fenotípico, entre otros. Los vectores pueden ser plásmidos, de virus, bacteriófagos, provirus, fagémidos, transposones, cromosomas artificiales y similares, que se replican de forma autónoma o pueden integrarse en un cromosoma de una célula huésped. Un vector también puede ser un polinucleótido de ARN desnudo, un polinucleótido de ADN desnudo, un polinucleótido compuesto de ADN y ARN dentro de la misma hebra, un ADN o ARN conjugado con polilisina, un ADN o ARN conjugado con péptidos, un ADN conjugado con liposomas o similar, que no se replica de forma autónoma. Como se usa en el presente documento, el término "expresión" se refiere a la producción de un producto final funcional, por ejemplo, un ARNm o una proteína (precursora o madura).

Por "unido operativamente" o "asociado operativamente", como se usa en este documento con referencia a polinucleótidos, se entiende que los elementos indicados están funcionalmente relacionados entre sí, y también están generalmente relacionados físicamente. Por lo tanto, el término "operativamente unido" o "operativamente asociado", como se usa en el presente documento, se

refiere a secuencias de nucleótidos en una sola molécula de ácido nucleico que están funcionalmente asociadas. Por lo tanto, que una primera secuencia de nucleótidos que está operativamente unida a una segunda secuencia de nucleótidos significa una situación en la que la primera secuencia de nucleótidos se coloca en una relación funcional con la segunda secuencia de nucleótidos. Por ejemplo, un promotor está asociado operativamente con una secuencia de nucleótidos si el promotor efectúa la transcripción o expresión de dicha secuencia de nucleótidos. Los expertos en la técnica apreciarán que las secuencias de control (por ejemplo, el promotor) no necesitan estar contiguas a la secuencia de nucleótidos a la que está operativamente asociada, siempre que las secuencias de control funcionen para dirigir su expresión. Así, por ejemplo, las secuencias de ácido nucleico no traducidas, pero transcriptas intermedias pueden estar presentes entre un promotor y la secuencia de nucleótidos, y el promotor todavía puede considerarse "unido operativamente" a la secuencia de nucleótidos.

Como se usa en el presente documento, el término "unido", con referencia a los polipéptidos, se refiere a la unión de un polipéptido a otro. Un polipéptido puede estar unido a otro polipéptido (en el extremo N-terminal o el extremo C-terminal) directamente (por ejemplo, a través de un enlace peptídico) o mediante un enlazador.

El término "enlazador" es reconocido en la técnica, y se refiere a un grupo químico, o una molécula que une dos moléculas o restos, por ejemplo, dos dominios de una proteína de fusión, tal como, por ejemplo, un polipéptido o dominio de unión de ADN y una etiqueta peptídica y/o una transcriptasa inversa y un polipéptido de afinidad que se une a la etiqueta peptídica; o un dominio o polipéptido de endonucleasa de ADN y etiqueta de péptido y/o una transcriptasa inversa y un polipéptido de afinidad que se une a la etiqueta de péptido. Un enlazador puede estar compuesto por una sola molécula de unión, o puede comprender más de una molécula de unión. En algunas formas de realización, el enlazador puede ser una molécula orgánica, grupo, polímero o resto químico tal como un resto orgánico bivalente. En algunas formas de realización, el enlazador puede ser un aminoácido, o puede ser un péptido. En algunas formas de realización, el enlazador es un péptido.

En algunas formas de realización, un enlazador peptídico útil con esta invención puede ser de alrededor de 2 a alrededor de 100 o más aminoácidos de longitud, por ejemplo, alrededor de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 o más aminoácidos de longitud (por ejemplo, alrededor de 2 a alrededor de 40, alrededor de 2 a alrededor de 50, alrededor de 2 a alrededor de 60,

alrededor de 4 a alrededor de 40, alrededor de 4 a alrededor de 50, alrededor de 4 a alrededor de 60, alrededor de 5 a alrededor de 40, alrededor de 5 a alrededor de 50, alrededor de 5 a alrededor de 60, alrededor de 9 a alrededor de 40, alrededor de 9 a alrededor de 50, alrededor de 9 a alrededor de 60, alrededor de 10 a alrededor de 40, alrededor de 10 a alrededor de 50, alrededor de 10 a alrededor de 60, o alrededor de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 aminoácidos a alrededor de 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 o más aminoácidos de longitud (por ejemplo, alrededor de 105, 110, 115, 120, 130, 140 150 o más aminoácidos de longitud). En algunas formas de realización, un enlazador peptídico puede ser un enlazador GS.

Como se usa en el presente documento, el término "unido" o "fusionado" con referencia a polinucleótidos, se refiere a la unión de un polinucleótido a otro. En algunas formas de realización, dos o más moléculas polinucleotídicas pueden estar unidas por un enlazador que puede ser una molécula orgánica, grupo, polímero o resto químico tal como un resto orgánico bivalente. Un polinucleótido puede unirse o fusionarse con otro polinucleótido (en el extremo 5' o el extremo 3') a través de un enlace o unión covalente o no covalente, que incluye, por ejemplo, el emparejamiento de bases de Watson–Crick, o mediante uno o más nucleótidos de unión. En algunas formas de realización, se puede insertar un motivo polinucleotídico de una determinada estructura dentro de otra secuencia polinucleotídica (por ejemplo, extensión de la estructura de horquilla en el ARN guía). En algunas formas de realización, los nucleótidos de unión pueden ser nucleótidos naturales. En algunas formas de realización, los nucleótidos de unión pueden ser nucleótidos de origen no natural.

Por lo tanto, en el contexto de la disposición secuencial de un promotor, el polinucleótido de acuerdo con la descripción con un oligo o polinucleótido adicional (por ejemplo, unido operativamente) produce la transcripción de dicho polinucleótido adicional. En algunas formas de realización, las secuencias promotoras de la presente descripción se insertan justo antes de 5'UTR de un ácido nucleico o marco de lectura abierto por expresar. En algunas formas de realización, las secuencias promotoras unidas operativamente y las secuencias de ácido nucleico de la presente descripción pueden estar separadas por uno o más nucleótidos enlazadores.

Como se usa en este documento, una "proteína efectora CRISPR–Cas" o "efector asociado a CRISPR" es una proteína o un polipéptido o su dominio que escinde o corta un ácido nucleico, se une a un ácido nucleico (por ejemplo, un ácido nucleico objetivo y/o un ácido nucleico guía), y/o que

identifica, reconoce o une un ácido nucleico guía como se define en el presente documento. En algunas formas de realización, una proteína efectora CRISPR–Cas puede ser una enzima (por ejemplo, una nucleasa, endonucleasa, nickasa, etc.) o una porción de esta, y/o puede funcionar como una enzima. En algunas formas de realización, una proteína efectora CRISPR–Cas se refiere a un polipéptido de nucleasa CRISPR–Cas o su dominio que comprende actividad de nucleasa o en el que la actividad de nucleasa se ha reducido o eliminado, y/o comprende actividad de nickasa o en el que la nickasa ha sido reducida o eliminada, y/o comprende actividad de escisión de ADN monocatenario (actividad de ADNasa ss) o en donde la actividad ADNasa ss se ha reducido o eliminado, y/o comprende actividad de ARNasa de autoprosesamiento o en donde la actividad de ARNasa de autoprosesamiento se ha reducido o eliminado. Una proteína efectora CRISPR–Cas puede unirse a un ácido nucleico objetivo. Por lo tanto, un "efector asociado a CRISPR" o "proteína efectora CRISPR–Cas" como se usa en el presente documento puede definirse como cualquier nucleasa, nickasa o recombinasa asociada con CRISPR (repeticiones palindrómicas cortas agrupadas regularmente intercaladas, conforme a sus siglas en inglés), que tiene la capacidad de introducir una escisión de hebra simple o doble en un sitio objetivo genómico, o que tiene la capacidad de introducir una modificación dirigida, que incluye una mutación puntual, una inserción o una delección, en un sitio objetivo genómico de interés. El por lo menos un efector asociado a CRISPR puede actuar por sí solo o en combinación con otras moléculas como parte de un complejo molecular. El efector asociado a CRISPR puede estar presente como molécula de fusión, o como moléculas individuales que se asocian o están asociadas por por lo menos una de una interacción covalente o no covalente con el ARNg y/o el sitio objetivo, para que los componentes del complejo asociado con CRISPR estén en proximidad física cercana.

En algunas formas de realización, un dominio de unión de ADN específico de secuencia puede ser una proteína efectora CRISPR–Cas. En algunas formas de realización, una proteína efectora CRISPR–Cas puede ser de un sistema CRISPR–Cas Tipo I, un sistema CRISPR–Cas Tipo II, un sistema CRISPR–Cas Tipo III, un sistema CRISPR–Cas Tipo IV, un sistema CRISPR–Cas Tipo V o un sistema CRISPR–Cas Tipo VI. En algunas formas de realización, una proteína efectora CRISPR–Cas de la invención puede ser de un sistema CRISPR–Cas Tipo II o un sistema CRISPR–Cas Tipo V. En algunas formas de realización, una proteína efectora CRISPR–Cas puede ser proteína efectora CRISPR–Cas tipo II, por ejemplo, una proteína efectora Cas9. En algunas formas de realización, una proteína efectora CRISPR–Cas puede ser proteína efectora CRISPR–Cas tipo V, por ejemplo, una proteína efectora Cas12.

En algunas formas de realización, una proteína efectora CRISPR–Cas puede incluir, sin limitación, una nucleasa Cas9, C2c1, C2c3, Cas12a (también referida como Cpf1), Cas12b, Cas12c, Cas12d, Cas12e, Cas13a, Cas13b, Cas13c, Cas13d, Cas1, Cas1B, Cas2, Cas3, Cas3', Cas3'', Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también denominado Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Cse3, Cse4, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 (dinG), y/o Csf5, opcionalmente en donde la proteína efectora CRISPR–Cas puede ser una proteína efectora Cas9, Cas12a (Cpf1), Cas12b, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12g, Cas12h, Cas12i, C2c4, C2c5, C2c8, C2c9, C2c10, Cas14a, Cas14b, y/o Cas14c.

En algunas formas de realización, una proteína efectora CRISPR–Cas útil con la invención puede comprender una mutación en su sitio activo de nucleasa (por ejemplo, RuvC, HNH, por ejemplo, el sitio RuvC de un dominio de nucleasa Cas12a; por ejemplo, el sitio RuvC y/o el sitio HNH de un dominio de nucleasa Cas9). Una proteína efectora CRISPR–Cas que tiene una mutación en su sitio activo de nucleasa y, por lo tanto, ya no comprende actividad de nucleasa, se denomina comúnmente "muerta", por ejemplo, dCas. En algunas formas de realización, un dominio o polipéptido de proteína efectora CRISPR–Cas que tiene una mutación en su sitio activo de nucleasa puede tener actividad deteriorada o actividad reducida en comparación con la misma proteína efectora CRISPR–Cas sin la mutación, por ejemplo, una nickasa, por ejemplo, nickasa Cas9, nickasa Cas12a.

Una proteína efectora CRISPR Cas9 o dominio efector CRISPR Cas9 útil con esta invención puede ser cualquier nucleasa Cas9 conocida o identificada más adelante. En algunas formas de realización, un polipéptido CRISPR Cas9 puede ser un polipéptido Cas9 de, por ejemplo, el género *Streptococcus* (por ejemplo, *S. pyogenes*, *S. thermophilus*), el género *Lactobacillus*, el género *Bifidobacterium*, el género *Kandleria*, el género *Leuconostoc*, el género *Oenococcus*, el género *Pediococcus*, el género *Weissella* y/o el género *Olsenella*.

En algunas formas de realización, la proteína efectora CRISPR–Cas puede ser un polipéptido Cas9 derivado de *Streptococcus pyogenes*, y reconoce el motivo de secuencia PAM NGG, NAG, NGA (Mali et al, Science 2013; 339(6121): 823–826). En algunas formas de realización, la proteína efectora CRISPR–Cas puede ser un polipéptido Cas9 derivado de *Streptococcus thermophiles* y reconoce el motivo de secuencia PAM NGGNG y/o NNAGAAW (W = A o T) (ver, por ejemplo, Horvath et al, Science, 2010; 327(5962): 167–170, y Deveau et al, J Bacteriol 2008; 190(4): 1390–1400). En algunas formas de realización, la proteína efectora CRISPR–Cas puede ser un polipéptido Cas9 derivado de *Streptococcus mutans* y reconoce el motivo de secuencia PAM NGG y/o NAAR

(R = A o G) (ver, por ejemplo, Deveau et al, J BACTERIOL 2008; 190(4): 1390–1400). En algunas formas de realización, la proteína efectora CRISPR–Cas puede ser un polipéptido Cas9 derivado de *Streptococcus aureus* y reconoce el motivo de secuencia PAM NNGRR (R = A o G). En algunas formas de realización, la proteína efectora CRISPR–Cas puede ser una proteína Cas9 derivada de *S. aureus*, que reconoce el motivo de secuencia PAM N GRRT (R = A o G). En algunas formas de realización, la proteína efectora CRISPR–Cas puede ser un polipéptido Cas9 derivado de *S. aureus*, que reconoce el motivo de secuencia PAM N GRRV (R = A o G). En algunas formas de realización, la proteína efectora CRISPR–Cas puede ser un polipéptido Cas9 derivado de *Neisseria meningitidis* y reconoce el motivo de secuencia PAM N GATT o N GCTT (R = A o G, V = A, G o C) (ver, por ejemplo, Hou et al, PNAS 2013, 1–6). En las formas de realización arriba mencionadas, N puede ser cualquier residuo de nucleótido, por ejemplo, cualquiera de A, G, C o T. En algunas formas de realización, la proteína efectora CRISPR–Cas puede ser una proteína Cas13a derivada de *Leptotrichia shahii*, que reconoce un motivo de secuencia de secuencia flanqueante de protoespaciador (PFS, conforme a sus siglas en inglés) (o ARN PAM (rPAM)) de un solo 3' A, U, o C, que puede ubicarse dentro del ácido nucleico objetivo.

En algunas formas de realización, la proteína efectora CRISPR–Cas se puede derivar de Cas12a, que es una nucleasa de repetición palindrómica corta agrupada regularmente intercalada de tipo V (CRISPR)–Cas. Cas12a difiere en varios aspectos de la nucleasa Cas9 CRISPR Tipo II más conocida. Por ejemplo, Cas9 reconoce un motivo adyacente de protoespaciador (PAM, conforme a sus siglas en inglés) rico en G que está 3' de su sitio de unión de ARN guía (ARNg, ARNsg, ARNcr, ADNcr, matriz CRISPR) (protoespaciador, ácido nucleico objetivo, ADN objetivo) (3'–NGG), mientras que Cas12a reconoce un PAM rico en T que se encuentra 5' del ácido nucleico objetivo (5'–TTN, 5'–TTTN). De hecho, las orientaciones en las que Cas9 y Cas12a se unen a sus ARN guía son revertidas muy cercanamente en relación con sus extremos N y C. Además, las enzimas Cas12a usan un ARN guía único (ARNg, matriz CRISPR, ARNcr) en lugar del ARN guía doble (ARNsg (p. ej., ARNcr y ARNtracr)) que se encuentra en los sistemas naturales Cas9, y Cas12a procesa sus propios ARNg. Además, la actividad de nucleasa de Cas12a produce roturas de doble hebra de ADN escalonadas en lugar de extremos romos producidos por la actividad de nucleasa Cas9, y Cas12a se basa en un solo dominio RuvC para escindir ambas hebras de ADN, mientras que Cas9 utiliza un dominio HNH y un dominio RuvC para la escisión.

Una proteína/dominio efector CRISPR Cas12a útil con esta invención puede ser cualquier polipéptido Cas12a conocido o identificado posteriormente (anteriormente conocido como Cpf1)

(véase, por ejemplo, la Patente de los Estados Unidos Núm. 9.790.490, que se incorpora como referencia para sus divulgaciones de secuencias Cpf1 (Cas12a)). El término "Cas12a", "polipéptido Cas12a" o "dominio Cas12a" se refiere a una nucleasa guiada por ARN que comprende un polipéptido Cas12a, o un fragmento del mismo, que comprende el dominio de unión de ácido nucleico guía de Cas12a y/o un dominio de escisión de ADN activo, inactivo, o parcialmente activo de Cas12a. En algunas formas de realización, un Cas12a útil con la invención puede comprender una mutación en el sitio activo de nucleasa (por ejemplo, el sitio RuvC del dominio Cas12a). Un dominio Cas12a o polipéptido Cas12a que tiene una mutación en su sitio activo de nucleasa y, por lo tanto, ya no comprende actividad de nucleasa, se denomina comúnmente Cas12a muerto (por ejemplo, dCas12a). En algunas formas de realización, un dominio Cas12a o un polipéptido Cas12a que tiene una mutación en su sitio activo de nucleasa puede tener una actividad deteriorada, por ejemplo, puede tener actividad de nickasa.

Un "editor de bases", como se usa en el presente documento, se refiere a una proteína o un fragmento de la misma que tiene la misma actividad catalítica que la proteína de la que deriva, cuya proteína o fragmento de la misma, sola o cuando se proporciona como complejo molecular, denominado complejo de edición de bases en el presente documento, tiene la capacidad de mediar una modificación de base dirigida, es decir, la conversión de una base de interés que produce una mutación puntual de interés, que a su vez, puede producir una mutación dirigida, si la conversión de base no causa una mutación silenciosa, sino una conversión de un aminoácido codificado por el codón que comprende la posición por convertir con el editor de base. Por lo menos un editor de bases de acuerdo con la presente divulgación es unido temporal o permanentemente a por lo menos un efector asociado a CRISPR, u opcionalmente a un componente de por lo menos un complejo efector asociado a CRISPR.

El término "nucleasa Cas9" y "Cas9" se pueden usar indistintamente en el presente documento, y se refieren a una enzima endonucleasa de ADN guiada por ARN asociada con el CRISPR (repeticiones palindrómicas cortas agrupadas regularmente intercaladas), incluida la proteína Cas9 o fragmentos de la misma (tal como una proteína que comprende un dominio de escisión de ADN activo de Cas9 y/o un dominio de unión de ARNg de Cas9). Cas9 es un componente del sistema de edición del genoma CRISPR/Cas, que apunta y corta una secuencia objetivo de ADN para formar rupturas de doble hebra (DSB, conforme a sus siglas en inglés) de ADN bajo la guía de un ARN guía.

Los términos "ácido nucleico guía", "ARN guía", "ARNg", "CRISPR", "ARN CRISPR" o "ARNcr" se refieren a la molécula de ARN responsable de la hibridación con las secuencias de ADN objetivo y el reclutamiento de endonucleasas CRISPR y/o efectores asociados con CRISPR. Los ARNcr pueden ser de origen natural, o pueden sintetizarse de acuerdo con cualquier método conocido para producir ARN. Los CRISPR se pueden introducir en forma de ARN o ADN. Los CRISPR comprenden por lo menos una secuencia espaciadora (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más), que es complementaria a (e hibrida con) un ADN objetivo (por ejemplo, protoespaciador), y por lo menos una secuencia de repetición (por ejemplo, una repetición de un sistema CRISPR–Cas Tipo V Cas12a, o su fragmento o porción; una repetición de un sistema CRISPR–Cas Tipo II Cas9, o su fragmento; una repetición de un sistema CRISPR Cas Tipo V C2c1 o su fragmento; una repetición de un sistema CRISPR–Cas de, por ejemplo, C2c3, Cas12a (también denominado Cpf1), Cas12b, Cas12c, Cas12d, Cas12e, Cas13a, Cas13b, Cas13c, Cas13d, Cas1, Cas1B, Cas2, Cas3, Cas3', Cas3'', Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también denominado Csn1 y Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Cse3, Cse4, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 (dinG), y/o Csf5, o su fragmento), en donde la secuencia de repetición puede estar unida al extremo 5' y/o el extremo 3' de la secuencia espaciadora. El diseño de un ARNg de esta invención puede ser sobre la base de un sistema CRISPR–Cas Tipo I, Tipo II, Tipo III, Tipo IV, Tipo V, o Tipo VI.

Un CRISPR o ácido nucleico guía puede comprender más de una secuencia espaciadora de secuencia de repetición (p. ej., 2, 3, 4, 5, 6, 7, 8, 9, 10 o más secuencias espaciadoras de repetición) (p. ej., repetición–espaciador–repetición, por ej., repetición–espaciador–repetición–espaciador–repetición–espaciador–repetición–espaciador–repetición–espaciador, y similares). Los ácidos nucleicos guía de esta invención son sintéticos, hechos por el hombre, y no se encuentran en la naturaleza. Un ARNg puede ser bastante largo y puede usarse como un aptámero (como en la estrategia de reclutamiento de MS2) u otras estructuras de ARN que cuelgan del espaciador.

Una "secuencia de repetición", como se usa en el presente documento, se refiere, por ejemplo, a cualquier secuencia de repetición de un locus CRISPR Cas de tipo salvaje (por ejemplo, un locus Cas9, un locus Cas12a, un locus C2c1, etc.) o una secuencia repetida de un ARNc sintético que es funcional con la proteína efectora CRISPR–Cas codificada por las construcciones de ácido nucleico de la invención. Una secuencia de repetición útil con esta invención puede ser cualquier secuencia de repetición conocida o identificada posteriormente de un locus CRISPR–Cas (por

ejemplo, Tipo I, Tipo II, Tipo III, Tipo IV, Tipo V o Tipo VI) o puede ser una repetición sintética diseñado para funcionar en un sistema CRISPR–Cas tipo I, II, III, IV, V o VI. Una secuencia repetida (o de repetición) puede comprender una estructura de horquilla y/o una estructura de bucle de tallo. En algunas formas de realización, una secuencia repetida puede formar una estructura similar a un pseudonudo en su extremo 5' (es decir, "asa"). Por lo tanto, en algunas formas de realización, una secuencia repetida puede ser idéntica o sustancialmente idéntica a una secuencia repetida de loci de CRISPR–Cas Tipo I, loci de CRISPR–Cas Tipo II, loci de CRISPR–Cas Tipo III, loci de CRISPR–Cas Tipo IV, loci de CRISPR–Cas Tipo V y/o loci de CRISPR–Cas Tipo VI de tipo salvaje. Una secuencia repetida de un locus CRISPR–Cas de tipo salvaje puede determinarse mediante algoritmos establecidos, como el uso del CRISPRfinder ofrecido a través de CRISPRdb (ver, Grissa et al. *Nucleic Acids Res.* 35 (publicación del servidor web): W52–7). En algunas formas de realización, una secuencia repetida o una porción de la misma está unida en su extremo 3' al extremo 5' de una secuencia espaciadora, formando así una secuencia espaciadora repetida (por ejemplo, ácido nucleico guía, ARN/ADN guía, ARNcr, ADNcr).

En algunas formas de realización, una secuencia de repetición comprende, consiste esencialmente en, o consiste en por lo menos 10 nucleótidos de acuerdo con la repetición particular y si el ácido nucleico guía que comprende la repetición es procesado o no procesado (por ejemplo, alrededor de 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 a 100 o más nucleótidos, o cualquier rango o valor entre ellos). En algunas formas de realización, una secuencia de repetición comprende, consiste esencialmente en, o consiste en alrededor de 10 a alrededor de 20, alrededor de 10 a alrededor de 30, alrededor de 10 a alrededor de 45, alrededor de 10 a alrededor de 50, alrededor de 15 a alrededor de 30, alrededor de 15 a alrededor de 40, alrededor de 15 a alrededor de 45, alrededor de 15 a alrededor de 50, alrededor de 20 a alrededor de 30, alrededor de 20 a alrededor de 40, alrededor de 20 a alrededor de 50, alrededor de 30 a alrededor de 40, alrededor de 40 a alrededor de 80, alrededor de 50 a alrededor de 100 o más nucleótidos.

Una secuencia de repetición unida al extremo 5' de una secuencia espaciadora puede comprender una porción de una secuencia de repetición (por ejemplo, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 o más nucleótidos contiguos de una secuencia de repetición de tipo salvaje). En algunas formas de realización, una porción de una secuencia de repetición unida al extremo 5' de una secuencia espaciadora puede ser de alrededor de cinco a alrededor de diez nucleótidos consecutivos de longitud (por ejemplo,

alrededor de 5, 6, 7, 8, 9, 10 nucleótidos) y tener por lo menos 90% de identidad de secuencia (por ejemplo, por lo menos alrededor de 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más) con la misma región (por ejemplo, extremo 5') de una secuencia de nucleótidos de repetición de tipo salvaje CRISPR Cas. En algunas formas de realización, una porción de una secuencia de repetición puede comprender una estructura de tipo pseudonudo en su extremo 5' (por ejemplo, "asa").

El término "ARNtracr" se refiere a un ARN transcodificado pequeño. ARNtracr es complementario y se empareja en bases con ARNcr para formar un híbrido ARNcr/ARNtracr, capaz de reclutar endonucleasas CRISPR y/o efectores asociados a CRISPR para secuencias objetivo.

En algunas formas de realización, un ARNg está compuesto de moléculas de ARNcr y ARNtracr que forman complejos a través del complemento parcial, en donde el ARNc comprende una secuencia que es suficientemente complementaria a una secuencia objetivo para la hibridación y dirige el complejo CRISPR (es decir, el híbrido Cas9–ARNcr/ARNtracr) para específicamente unirse a la secuencia objetivo. Además, se puede diseñar un ARN de guía único (ARNsg), que comprende las características tanto de ARNcr como de ARNtracr. Por lo tanto, como se usa en el presente documento, un ARN guía puede ser un ARNcr natural o sintético (por ejemplo, para Cpf1), un híbrido de ARNcr/ARNtracr natural o sintético (por ejemplo, para Cas9), o un ARN de guía único (ARNsg).

El término "secuencia guía" o "secuencia espaciadora" se refiere a la porción de un ARNcr o ARN guía (ARNg) que es responsable de la hibridación con el ADN objetivo.

Una "secuencia espaciadora" como se usa en el presente documento es una secuencia de nucleótidos que es complementaria y se une a un ácido nucleico diana (p. ej., ADN objetivo) (p. ej., protoespaciador) (p. ej., nucleótidos consecutivos de una secuencia (i) que codifica una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID Nos: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**; (ii) que codifica a región o una porción de una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**; (iii) que comprende una porción de una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**; y/o (iv) que comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**. En algunas formas de realización, una secuencia espaciadora puede incluir, sin limitación, cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 175, 176 o 184–186**. Una secuencia espaciadora puede ser completamente complementaria o sustancialmente complementaria (por ejemplo, por lo menos

alrededor de 70% complementaria (por ejemplo, alrededor de 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más)) a un ácido nucleico objetivo. Así, en algunas formas de realización, la secuencia espaciadora puede tener una, dos, tres, cuatro, cinco no coincidencias, en comparación con el ácido nucleico objetivo, cuyas no coincidencias pueden ser contiguas o no contiguas. En algunas formas de realización, la secuencia espaciadora puede tener 70% de complementariedad con un ácido nucleico objetivo. En otras formas de realización, la secuencia de nucleótidos espaciadora puede tener 80% de complementariedad con un ácido nucleico objetivo. En aún otras formas de realización, la secuencia de nucleótidos espaciadora puede tener 85%, 90%, 95%, 96%, 97%, 98%, 99% o 99,5% de complementariedad, y similares, con el ácido nucleico objetivo (protoespaciador). En algunas formas de realización, la secuencia espaciadora es 100% complementaria al ácido nucleico objetivo. Una secuencia espaciadora puede tener una longitud de alrededor de 15 nucleótidos a alrededor de 30 nucleótidos (por ejemplo, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, o 30 nucleótidos, o cualquier rango o valor entre ellos). Por lo tanto, en algunas formas de realización, una secuencia espaciadora puede tener completa complementariedad o sustancial complementariedad sobre una región de un ácido nucleico objetivo (por ejemplo, protoespaciador) que es por lo menos alrededor de 15 nucleótidos a alrededor de 30 nucleótidos de longitud. En algunas formas de realización, el espaciador es de alrededor de 20 nucleótidos de longitud. En algunas formas de realización, el espaciador es de alrededor de 21, 22, o 23 nucleótidos de longitud.

En algunas formas de realización, la región 5' de una secuencia espaciadora de un ácido nucleico guía puede ser idéntica a un ADN objetivo, mientras que la región 3' del espaciador puede ser sustancialmente complementaria al ADN objetivo (p. ej., Tipo V CRISPR–Cas), o la región 3' de una secuencia espaciadora de un ácido nucleico guía puede ser idéntica a un ADN objetivo, mientras que la región 5' del espaciador puede ser sustancialmente complementaria al ADN objetivo (p. ej., Tipo II CRISPR–Cas), y por lo tanto, la complementariedad general de la secuencia espaciadora con el ADN objetivo puede ser inferior al 100%. Así, por ejemplo, en una guía para un sistema CRISPR–Cas Tipo V, los primeros 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 nucleótidos en la región 5' (es decir, región de semilla) de, por ejemplo, una secuencia espaciadora de 20 nucleótidos pueden ser 100% complementarios al ADN objetivo, mientras que los nucleótidos restantes en la región 3' de la secuencia espaciadora son sustancialmente complementarios (por ejemplo, por lo menos alrededor de 70% complementario) al ADN objetivo. En algunas formas de realización, los primeros 1 a 8

nucleótidos (p. ej., los primeros 1, 2, 3, 4, 5, 6, 7, 8, nucleótidos y cualquier rango en el medio) del extremo 5' de la secuencia espaciadora pueden ser 100 % complementarios al ADN objetivo, mientras que los nucleótidos restantes en la región 3' de la secuencia espaciadora son sustancialmente complementarios (por ejemplo, por lo menos alrededor de 50% complementarios (por ejemplo, 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más)) al ADN objetivo.

Como otro ejemplo, en una guía para un sistema CRISPR–Cas Tipo II, los primeros 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 nucleótidos en la región 3' (es decir, región de semilla) de, por ejemplo, una secuencia espaciadora de 20 nucleótidos pueden ser 100% complementarios al ADN objetivo, mientras que los nucleótidos restantes en la región 5' de la secuencia espaciadora son sustancialmente complementarios (p. ej., por lo menos alrededor de 70% complementario) al ADN objetivo . En algunas formas de realización, los primeros 1 a 10 nucleótidos (p. ej., los primeros 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 nucleótidos, y cualquier rango en el medio) del extremo 3' de la secuencia espaciadora pueden ser 100% complementarios al ADN objetivo, mientras que los nucleótidos restantes en la región 5' de la secuencia espaciadora son sustancialmente complementarios (por ejemplo, por lo menos alrededor de 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más o cualquier rango o valor en el medio)) al ADN objetivo

En algunas formas de realización, una región de semilla de un espaciador puede ser de alrededor de 8 a alrededor de 10 nucleótidos de longitud, alrededor de 5 a alrededor de 6 nucleótidos de longitud, o alrededor de 6 nucleótidos de longitud.

El término "protoespaciador" se refiere al ADN bicatenario dirigido por una secuencia guía de un CRISPR (por ejemplo, ARNcr, ARNg). En algunas formas de realización, la secuencia de protoespaciador se hibrida (por ejemplo, es completa o sustancialmente complementaria como se describe anteriormente) con la secuencia espaciadora de un CRISPR.

El término "región semilla" se refiere a la porción de una secuencia de guía de un ARNcr o ARN guía que es más susceptible a la falla de funcionar como guía debido a no coincidencias con sus objetivos. En algunas formas de realización, una sola falta de coincidencia en la región de la semilla de un ARNcr/ARNg puede hacer que un complejo CRISPR sea inactivo en ese sitio de unión. En algunas formas de realización, las regiones de semillas para las endonucleasas Cas9 están ubicadas a lo largo de alrededor de los últimos 12 nts de la porción 3' de la secuencia de guía, que corresponden

(hibridan) a la porción de la secuencia diana protoespaciadora que está adyacente al PAM. En algunas formas de realización, las regiones de semillas para las endonucleasas Cpf1 se ubican a lo largo de aproximadamente los primeros 5 nts de la porción 5' de la secuencia guía, que corresponden (hibridan) a la porción de la secuencia diana protoespaciadora adyacente al PAM.

El término "sitio de aterrizaje CRISPR", "ácido nucleico diana", "ADN diana", "secuencia de nucleótidos diana", "región diana" o una "región diana en el genoma" como se usa en el presente documento, se refiere a una secuencia de ADN capaz de ser objetivo de un complejo CRISPR–Cas, por ejemplo, una región del genoma de una planta que es completamente complementaria (100% complementaria) o sustancialmente complementaria (por ejemplo, por lo menos 70% complementaria (por ejemplo, 70%, 71%, 72%, 73 %, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más)) a una secuencia espaciadora en un ácido nucleico guía de esta invención. En algunas formas de realización, un sitio de aterrizaje CRISPR comprende una secuencia combinada de protoespaciador/motivo adyacente a protoespaciador próximamente ubicada, que puede ser escindida por un complejo CRISPR. En algunas formas de realización, una región objetivo útil para un sistema CRISPR–Cas puede ubicarse inmediatamente 3' (por ejemplo, Tipo V sistema CRISPR–Cas) o inmediatamente 5' (por ejemplo, tipo II sistema CRISPR–Cas) a una secuencia PAM en el genoma del organismo (por ejemplo, un genoma de planta). Se puede seleccionar una región objetivo de cualquier región de por lo menos 15 nucleótidos consecutivos (por ejemplo, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 nucleótidos, y similar) ubicada inmediatamente adyacente a una secuencia PAM.

En el caso de los sistemas Tipo II CRISPR–Cas (por ejemplo, Cas9), el PAM está ubicado inmediatamente a 3' de la región objetivo. El PAM para los sistemas Tipo I CRISPR–Cas está ubicado 5' de la cadena objetivo. No se conoce PAM para los sistemas CRISPR–Cas Tipo III. Makarova et al. describen la nomenclatura para todas las clases, tipos y subtipos de sistemas CRISPR (Nature Reviews Microbiology 13: 722–736 (2015)). Las estructuras de guía y los PAM se describen en R. Barrangou (Genome Biol. 16: 247 (2015)).

Los PAM canónicos Cas12a son ricos en T. En algunas formas de realización, una secuencia canónica PAM Cas12a puede ser 5'–TTN, 5'–TTTN o 5'–TTTV. En algunas formas de realización, las PAM canónicas Cas9 (por ejemplo, *S. pyogenes*) pueden ser 5'–NGG–3'. En algunas formas de realización, se pueden usar PAM no canónicos pero pueden ser menos eficientes.

Los expertos en la técnica pueden determinar secuencias PAM adicionales a través de enfoques experimentales y computacionales establecidos. Por lo tanto, por ejemplo, los enfoques

experimentales incluyen el direccionamiento de una secuencia flanqueada por todas las secuencias de nucleótidos posibles y la identificación de miembros de secuencia que no experimentan el direccionamiento, como a través de la transformación del ADN plasmídico objetivo (Esvelt et al. 2013. *Nat. Methods* 10: 1116– 1121; Jiang et al. 2013. *Nat. Biotechnol.* 31: 233–239). En algunos aspectos, un enfoque computacional puede incluir realizar búsquedas BLAST de espaciadores naturales para identificar las secuencias de ADN objetivo originales en bacteriófagos o plásmidos y alinear estas secuencias para determinar secuencias conservadas adyacentes a la secuencia objetivo (Briner y Barrangou. 2014. *Appl. Environ. Microbiol.* 80: 994–1001; Mojica et al. 2009. *Microbiology* 155: 733–740).

El término "complejo CRISPR", "complejo de endonucleasa CRISPR", "complejo CRISPR Cas" o "complejo CRISPR–ARNg" se usan indistintamente en el presente documento. "Complejo CRISPR" se refiere, por ejemplo, a una nucleasa Cas9 y/u otro efector asociado a CRISPR (por ejemplo, Cas12a/Cpf1) acompañado con un ARN guía (ARNg). El término "complejo CRISPR" se refiere así a una combinación de endonucleasa CRISPR y ARN guía (y cualquier otra secuencia de nucleótidos CRISPR/polipéptidos) capaz de inducir una ruptura bicatenaria en un sitio de aterrizaje CRISPR. En algunas formas de realización, el "complejo CRISPR" de la presente descripción se refiere a una combinación de proteína Cas9 catalíticamente muerta y ARN guía capaz de apuntar a una secuencia objetivo, pero no capaz de inducir una ruptura bicatenaria en un sitio de aterrizaje CRISPR porque pierde una nucleasa actividad. En otras formas de realización, el "complejo CRISPR" de la presente divulgación se refiere a una combinación de nickasa Cas9 y ARN guía capaz de introducir rupturas de hebra única dirigidas a ARNg en el ADN en lugar de las rupturas bicatenarias creadas por las enzimas Cas de tipo salvaje.

Como se usa en el presente documento, el término "dirigir la unión específica de secuencia" en el contexto de los complejos CRISPR se refiere a la capacidad de un ARN guía para reclutar una endonucleasa CRISPR y/o efectores asociados a CRISPR en un sitio de aterrizaje CRISPR o ácido nucleico objetivo.

Como se usa en el presente documento, el término "desaminasa" se refiere a una enzima que cataliza la reacción de desaminación. En algunas formas de realización de la presente descripción, la desaminasa se refiere a una citidina desaminasa, que cataliza la desaminación de una citidina o una desoxicitidina a un uracilo o una desoxiuridina, respectivamente. En otras formas de realización de la presente descripción, la desaminasa se refiere a una adenosina desaminasa, que cataliza la

desaminación de una adenina para formar hipoxantina (en forma de su nucleósido inosina), que es leída como guanina por la ADN polimerasa.

Como se usa en el presente documento, el término "glicosilasa" se refiere a una familia de enzimas implicadas en la reparación de escisión de bases, clasificada bajo el número EC EC 3.2.2. La reparación de escisión de bases es el mecanismo por el cual las bases dañadas en el ADN se eliminan y reemplazan. Las glicosilasas de ADN catalizan el primer paso de este proceso. Quitan la base nitrogenosa dañada mientras dejan intacto el esqueleto de azúcar y fosfato, de modo de crear un sitio apurínico/apirimidínico, comúnmente conocido como un sitio AP. Esto se logra volteando la base dañada fuera de la doble hélice, seguido de la escisión del enlace N–glicosídico. En algunas formas de realización de la presente descripción, con la expectativa de proporcionar una tendencia de introducción de mutación diferente de la de desaminasa y similares, se usa una reacción de escisión de bases por hidrólisis del enlace N–glicosídico de ADN, y luego la inducción de la introducción de mutación en un proceso de reparación de células. En aspectos, se usa una enzima que tiene actividad de citosina–ADN glicosilasa (CDG) o actividad de timina–ADN glicosilasa (TDG). En algunos aspectos, se usa un mutante de uracilo–ADN glicosilasa mitocondrial de levadura (UNG 1), como una enzima que realiza dicha reacción de escisión de bases. Nishida et al., US 2017/0321210 A1, publicado el 9 de noviembre de 2017, se incorpora aquí como referencia.

Como se usa en el presente documento, el término "dirigido" se refiere a la expectativa de que un elemento o una molécula interactuará con otro elemento u otra molécula con un grado de especificidad, a fin de excluir elementos o moléculas no dirigidos. Por ejemplo, un primer polinucleótido que está dirigido a un segundo polinucleótido, de acuerdo con la presente descripción, se ha diseñado para hibridarse con el segundo polinucleótido de una manera específica de secuencia (por ejemplo, mediante emparejamiento de bases Watson–Crick). En algunas formas de realización, la región seleccionada de hibridación está diseñada para hacer que la hibridación sea única para la una o más regiones objetivo. Un segundo polinucleótido puede dejar de ser un objetivo de un primer polinucleótido de direccionamiento, si su secuencia de direccionamiento u objetivo (región de hibridación) está mutada, o si está de otro modo eliminada/separada del segundo polinucleótido. Además, "dirigido" se puede usar indistintamente con "específico de sitio" o "dirigido a sitio", que se refiere a una acción de la biología molecular que utiliza información sobre la secuencia de una región genómica de interés por ser modificada, y que se sustenta aún más en la información del mecanismo de acción de las herramientas moleculares, por ejemplo, nucleasas, incluidas las nucleasas CRISPR y sus variantes, TALEN, ZFN, meganucleasas o recombinasas, enzimas modificadoras de ADN,

incluidas las enzimas modificadoras de bases como las enzimas citidina desaminasas, las enzimas modificadoras de histonas y similares, proteínas de unión al ADN, ARN cr/tracr, ARN guía y similares.

El término "identidad de secuencia" se refiere al porcentaje de bases o aminoácidos entre dos secuencias de polinucleótidos o polipéptidos que son iguales y están en la misma posición relativa. De este modo, una secuencia de polinucleótidos o polipéptidos tiene un cierto porcentaje de identidad de secuencia en comparación con otra secuencia de polinucleótidos o polipéptidos. Para la comparación de secuencias, típicamente una secuencia actúa como una secuencia de referencia, con la cual se comparan las secuencias de prueba. El término "secuencia de referencia" se refiere a una molécula con la que se compara una secuencia de prueba. La "identidad" se puede calcular fácilmente mediante métodos conocidos que incluyen, entre otros, los descritos en: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, New York (1993); *Computer Analysis of SEquence Data, Part I* (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press (1987); y *Sequence Analysis Primer* (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991).

Como se usa en el presente documento, el término "porcentaje de identidad de secuencia" o "porcentaje de identidad" se refiere al porcentaje de nucleótidos idénticos en una secuencia de polinucleótido lineal de una molécula de polinucleótido de referencia ("consulta") (o su hebra complementaria) en comparación con una molécula de polinucleótido de prueba ("sujeto") (o su hebra complementaria) cuando las dos secuencias están alineadas de manera óptima. En algunas formas de realización, "porcentaje de identidad" puede referirse al porcentaje de aminoácidos idénticos en una secuencia de aminoácidos en comparación con un polipéptido de referencia.

Como se usa en el presente documento, la frase "sustancialmente idéntica" o "identidad sustancial" en el contexto de dos moléculas de ácido nucleico, secuencias de nucleótidos o secuencias de polipéptidos, se refiere a dos o más secuencias o subsecuencias que tienen por lo menos alrededor de 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% o 100% de identidad de residuos de aminoácidos o nucleótidos, cuando se comparan y se alinean para la máxima correspondencia, medido usando uno de los siguientes algoritmos de comparación de secuencias, o por inspección visual. En algunas formas de realización de la invención, existe identidad sustancial sobre una región de nucleótidos consecutivos de una secuencia de nucleótidos de la invención que es

de alrededor de 10 nucleótidos a alrededor de 20 nucleótidos, alrededor de 10 nucleótidos a alrededor de 25 nucleótidos, alrededor de 10 nucleótidos a alrededor de 30 nucleótidos, alrededor de 15 nucleótidos a alrededor de 25 nucleótidos, alrededor de 30 nucleótidos a alrededor de 40 nucleótidos, alrededor de 50 nucleótidos a alrededor de 60 nucleótidos, alrededor de 70 nucleótidos a alrededor de 80 nucleótidos, alrededor de 90 nucleótidos a alrededor de 100 nucleótidos, alrededor de 100 nucleótidos a alrededor de 200 nucleótidos, alrededor de 100 nucleótidos a alrededor de 300 nucleótidos, alrededor de 100 nucleótidos a alrededor de 400 nucleótidos, alrededor de 100 nucleótidos a alrededor de 500 nucleótidos, alrededor de 100 nucleótidos a alrededor de 600 nucleótidos, alrededor de 100 nucleótidos a alrededor de 800 nucleótidos, alrededor de 100 nucleótidos a alrededor de 900 nucleótidos, o más de longitud, o cualquier rango dentro de ellos, hasta la longitud total de la secuencia. En algunas formas de realización, las secuencias de nucleótidos pueden ser sustancialmente idénticas sobre por lo menos alrededor de 20 nucleótidos (por ejemplo, alrededor de 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70 u 80 nucleótidos o más).

En algunas formas de realización de la invención, existe identidad sustancial sobre una región de residuos de aminoácidos consecutivos de un polipéptido de la invención que es de alrededor de 3 residuos de aminoácidos a alrededor de 20 residuos de aminoácidos, alrededor de 5 residuos de aminoácidos a alrededor de 25 residuos de aminoácidos, alrededor de 7 residuos de aminoácidos a alrededor de 30 residuos de aminoácidos, alrededor de 10 residuos de aminoácidos a alrededor de 25 residuos de aminoácidos, alrededor de 15 residuos de aminoácidos a alrededor de 30 residuos de aminoácidos, alrededor de 20 residuos de aminoácidos a alrededor de 40 residuos de aminoácidos, alrededor de 25 residuos de aminoácidos a alrededor de 40 residuos de aminoácidos, alrededor de 25 residuos de aminoácidos a alrededor de 50 residuos de aminoácidos, alrededor de 30 residuos de aminoácidos a alrededor de 50 residuos de aminoácidos, alrededor de 40 residuos de aminoácidos a alrededor de 50 residuos de aminoácidos, alrededor de 40 residuos de aminoácidos a alrededor de 70 residuos de aminoácidos, alrededor de 50 residuos de aminoácidos a alrededor de 70 residuos de aminoácidos, alrededor de 60 residuos de aminoácidos a alrededor de 80 residuos de aminoácidos, alrededor de 70 residuos de aminoácidos a alrededor de 80 residuos de aminoácidos, alrededor de 90 residuos de aminoácidos a alrededor de 100 residuos de aminoácidos, o más residuos de aminoácidos de longitud, y cualquier rango entre ellos, hasta la longitud total de la secuencia. En algunas formas de realización, las secuencias de polipéptidos pueden ser sustancialmente idénticas entre sí sobre por lo menos alrededor de 8 residuos de aminoácidos consecutivos (por ejemplo, alrededor de 8, 9, 10,

11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 130, 140, 150, 175, 200, 225, 250, 300, 350 o más aminoácidos de longitud o más residuos de aminoácidos consecutivos). En algunas formas de realización, dos o más polipéptidos STL pueden ser idénticos (100%) o sustancialmente idénticos (por ejemplo, por lo menos 70% a 99,9% idénticos; por ejemplo, alrededor de 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5%. 99,9% idénticos o cualquier rango o valor en el medio) entre sí. En algunas formas de realización, dos o más proteínas STL pueden ser sustancialmente idénticas a través de residuos de aminoácidos consecutivos 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 a alrededor de 40, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 660, 665 o más de cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164 o 166–171**.

"Complementario" o "complementariedad", se refiere a la capacidad de emparejamiento, mediante apilamiento de bases y enlaces de hidrógeno específicos, entre dos secuencias que comprenden bases naturales o no naturales o sus análogos. Por ejemplo, si una base en una posición de un ácido nucleico es capaz de enlace de hidrógeno con una base en la posición correspondiente de un objetivo, entonces las bases se consideran complementarias entre sí en esa posición. Los ácidos nucleicos pueden comprender bases universales, o separadores abásicos inertes que no proporcionan una contribución positiva o negativa al enlace de hidrógeno. Los emparejamientos de bases pueden incluir emparejamientos de bases canónicas Watson–Crick y emparejamientos de bases no Watson–Crick (por ejemplo, emparejamientos de bases Wobble y emparejamientos de bases Hoogsteen). Se entiende que para las parejas de bases complementarias, las bases de tipo adenosina (A) son complementarias de las bases de tipo timidina (T) o las bases de tipo uracilo (U), que las bases de tipo citosina (C) son complementarias de las bases de tipo guanosina

(G), y que las bases universales tales como 3-nitropirrol o 5-nitroindol pueden hibridarse y se consideran complementarias con cualquier A, C, U o T. Nichols et al., *Nature*, 1994; 369: 492– 493, y Loakes et al., *Nucleic Acids Res.*, 1994; 22: 4039–4043. La inosina (I) también se ha considerado en la técnica como una base universal, y se considera complementaria a cualquier A, C, U o T. Ver Watkins y Santa Lucia, *Nucl. Acids Research*, 2005; 33 (19): 6258–6267.

Como se menciona aquí, una "secuencia de ácido nucleico complementaria" es una secuencia de ácido nucleico que comprende una secuencia de nucleótidos que le permite unirse de manera no covalente a otro ácido nucleico de una manera antiparalela específica de secuencia (es decir, un ácido nucleico específicamente se une a un ácido nucleico complementario) bajo las condiciones apropiadas *in vitro* y/o *in vivo* de temperatura y concentración iónica de solución. La complementariedad entre dos moléculas monocatenarias puede ser "parcial", en la que solo se unen algunos de los nucleótidos, o puede ser completa, cuando existe una complementariedad total entre las moléculas monocatenarias. El grado de complementariedad entre las hebras de ácido nucleico tiene efectos significativos sobre la eficiencia y la resistencia de la hibridación entre hebras de ácido nucleico.

"Complemento", como se usa en el presente documento, puede significar 100% de complementariedad con la secuencia de nucleótidos del comparador, o puede significar menos del 100% de complementariedad (por ejemplo, alrededor de 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, y similares, de complementariedad) con la secuencia de nucleótidos del comparador.

Los métodos de alineación de secuencias para la comparación y determinación del porcentaje de identidad de secuencia y el porcentaje de complementariedad son bien conocidos en la técnica. La alineación óptima de secuencias para la comparación puede llevarse a cabo, por ejemplo, por el algoritmo de alineación de homología de Needleman y Wunsch, (1970) *J. Mol. Biol.* 48: 443, por el método de búsqueda de similitud de Pearson y Lipman, (1988) *Proc. Nat'l. Acad. Sci. USA* 85: 2444, por implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA, y TFASTA en el paquete de soporte lógico Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), por alineación manual e inspección visual (ver, por ejemplo, Brent *et al.*, (2003) *Current Protocols in Molecular Biology*), mediante el uso de algoritmos conocidos en la técnica, que incluyen los algoritmos BLAST y BLAST 2.0, que se describen en las referencias de Altschul *et al.*, (1977) *Nuc. Acids Res.* 25: 3389–3402; y Altschul *et al.*, (1990) *J. Mol. Biol.* 215:403–410, respectivamente. El soporte lógico para realizar análisis BLAST está disponible

públicamente a través del Centro Nacional de Información Biotecnológica. Algunos programas de alineación son MacVector (Oxford Molecular Ltd, Oxford, Reino Unido), ALIGN Plus (Scientific and Educational Software, Pennsylvania) y AlignX (Vector NTI, Invitrogen, Carlsbad, CA). Otro programa de alineación es Sequencher (Gene Codes, Ann Arbor, Michigan), que utiliza parámetros predeterminados, y MUSCLE (Comparación de secuencia múltiple por Log-Expectation; un soporte lógico de computadora con licencia como dominio público).

Aquí, el término "hibridar" se refiere al emparejamiento entre bases de nucleótidos complementarias (por ejemplo, adenina (A) forma un par de bases con timina (T) en una molécula de ADN, y con uracilo (U), en una molécula de ARN, y guanina (G) forma un par de bases con citosina (C) en las moléculas de ADN y ARN) para formar una molécula de ácido nucleico bicatenario. (Ver, por ejemplo, Wahl y Berger (1987) *Methods Enzymol.* 152: 399; Kimmel, (1987) *Methods Enzymol.* 152: 507). Además, también se sabe en la técnica que para la hibridación entre dos moléculas de ARN (p. ej., ARNds), guanina (G) empareja en base con uracilo (U). Por ejemplo, el emparejamiento de bases G/U es parcialmente responsable de la degeneración (es decir, la redundancia) del código genético en el contexto del emparejamiento de bases anticodones de ARNt con codones en ARNm. En el contexto de esta descripción, una guanina (G) de un segmento de unión de proteína (dupla ARNds) de una molécula de ARN guía se considera complementaria a un uracilo (U), y viceversa. De este modo, cuando se puede hacer un par de bases G / U en una posición de nucleótido dada, un segmento de unión de proteína (dupla ARNds) de una molécula de ARN guía, la posición no se considera no complementaria, sino que se considera complementaria. Se entiende en la técnica que la secuencia del polinucleótido no necesita ser 100% complementaria a la de su ácido nucleico diana para ser específicamente hibridable. Además, un polinucleótido puede hibridarse sobre uno o más segmentos de modo que los segmentos intermedios o adyacentes no estén involucrados en el evento de hibridación (por ejemplo, una estructura de bucle o estructura de horquilla). Un polinucleótido puede comprender por lo menos 70%, por lo menos 80%, por lo menos 90%, por lo menos 95%, por lo menos 99% o 100% de complementariedad de secuencia a una región objetivo dentro de la secuencia de ácido nucleico objetivo a la que se dirigen.

El término "modificado" se refiere a una sustancia o un compuesto (por ejemplo, una célula, una secuencia de polinucleótidos y/o una secuencia de polipéptidos) que se ha alterado o cambiado en comparación con la sustancia o compuesto no modificado correspondiente.

"Aislado" se refiere a un material que está libre en diversos grados de componentes que normalmente lo acompañan como se encuentra en su estado nativo.

El término "planta, parte o célula editada por genes" como se usa en el presente documento se refiere a una planta, parte o célula que comprende uno o más genes endógenos que son editados por un sistema de edición de genes. El sistema de edición de genes de la presente descripción comprende un elemento de direccionamiento y/o un elemento de edición. El elemento de direccionamiento es capaz de reconocer una secuencia genómica diana. El elemento de edición es capaz de modificar la secuencia genómica diana, por ejemplo, mediante sustitución o inserción de uno o más nucleótidos en la secuencia genómica, delección de uno o más nucleótidos en la secuencia genómica, alteración de secuencias genómicas para incluir secuencias reguladoras, inserción de transgenes en un sitio genómico de puerto seguro u otra ubicación específica en el genoma, o cualquier combinación de los anteriores. El elemento de direccionamiento y el elemento de edición pueden estar en la misma molécula de ácido nucleico o en diferentes moléculas de ácido nucleico. En algunas formas de realización, el elemento de edición es capaz de editar con precisión el genoma mediante la sustitución de un solo nucleótido usando un editor de base, tal como editor de base de citosina (CBE) y/o editor de base de adenina (ABE), que se fusiona directa o indirectamente a una proteína efectora asociada a CRISPR.

Como se usa en el presente documento, el término "planta" se refiere a plantas enteras. El término "parte de la planta" incluye tejidos diferenciados y no diferenciados que incluyen, entre otros: órganos vegetales, tejidos vegetales, raíces, tallos, brotes, portainjertos, injertos, estípulas, pétalos, hojas, flores, óvulos, pólenes, brácteas, pecíolos, entrenudos, corteza, pubescencia, vástagos, rizomas, frondas, láminas, estambres, frutos, semillas, tejido tumoral y células vegetales (p. ej., células individuales, protoplastos, embriones y tejido calloso). Las células vegetales incluyen, sin limitación, células de semillas, cultivos en suspensión, embriones, regiones meristemáticas, tejido calloso, hojas, raíces, brotes, gametofitos, esporofitos, polen y microsporas. El tejido vegetal puede estar en una planta o en un órgano vegetal, tejido o cultivo celular. Por lo tanto, el término "parte de la planta", como se usa en el presente documento, incluye pero no se limita a tejidos reproductivos (p. ej., pétalos, sépalos, estambres, pistilos, receptáculos, anteras, polen, flores, frutos, capullos, óvulos, semillas y embriones); tejidos vegetativos (p. ej., pecíolos, tallos, raíces, pelos radiculares, puntas de raíz, médula, coleóptilos, tallos, brotes, ramas, corteza, meristemo apical, yema axilar, cotiledón, hipocotilos y hojas); tejidos vasculares (por ejemplo, floema y xilema); células especializadas tales como células epidérmicas, células de parénquima, células de colénquima, células de esclerénquima, estomas, células protectoras, cutícula, células mesofílicas; tejido calloso; y estacas. El término "parte de planta" también incluye células vegetales, incluidas las células vegetales que están intactas en

plantas y/o partes de plantas, protoplastos vegetales, tejidos vegetales, órganos vegetales, cultivos de tejidos celulares vegetales, callos vegetales, agrupamientos vegetales y similares. Como se usa en el presente documento, "brote" se refiere a las partes por encima del suelo, incluidas las hojas y los tallos. Como se usa en el presente documento, el término "cultivo de tejido" abarca cultivos de tejidos, células, protoplastos y callosidades. El término "tallo", como se usa en el presente documento, se refiere al eje estructural de la planta sobre el suelo, que consiste tanto en nudos (por ejemplo, hojas y flores) como entrenudos (por ejemplo, material de conexión entre nodos).

Como se usa en el presente documento, "célula vegetal" se refiere a una unidad estructural y fisiológica de la planta, que típicamente comprende una pared celular pero también incluye protoplastos. Una célula vegetal de la presente invención puede estar en forma de una célula individual aislada, o puede ser una célula cultivada, o puede ser parte de una unidad organizada superior como, por ejemplo, un tejido vegetal (incluyendo callos) o un órgano vegetal. Un "protoplasto" es una célula vegetal aislada sin una pared celular o con solo partes de la pared celular. Por lo tanto, en algunas formas de realización de la invención, una célula transgénica que comprende una molécula de ácido nucleico y/o una secuencia de nucleótidos de la invención es una célula de cualquier planta o parte de la planta que incluye, sin limitación, una célula de raíz, una célula de hoja, una célula de cultivo de tejidos, una célula de semilla, una célula de flor, una célula de fruta, una célula de polen y similares. En algunos aspectos de la invención, la parte de la planta puede ser un germoplasma de plantas. En algunos aspectos, una célula vegetal puede ser una célula vegetal no propagante que no se regenera en una planta.

"Cultivo de células vegetales" significa cultivos de unidades vegetales tales como, por ejemplo, protoplastos, células de cultivo celular, células en tejidos vegetales, polen, tubos de polen, óvulos, sacos embrionarios, cigotos y embriones en diversas etapas de desarrollo.

El término "órgano vegetal" se refiere al tejido vegetal o un grupo de tejidos que constituyen una parte morfológica y funcionalmente distinta de una planta, como una raíz, tallo, hoja, botón floral o embrión.

"Tejido vegetal" como se usa en el presente documento significa un grupo de células vegetales organizadas en una unidad estructural y funcional. Se incluye cualquier tejido de una planta, en planta o en cultivo. Este término incluye, sin limitación, plantas enteras, órganos vegetales, semillas vegetales, cultivo de tejidos y cualquier grupo de células vegetales organizadas en unidades estructurales y/o funcionales. El uso de este término en conjunción con cualquier tipo específico de

tejido vegetal, o en ausencia de este, como se menciona anteriormente o de otra manera abarcado por esta definición no pretende ser exclusivo de ningún otro tipo de tejido vegetal.

Como se usa en el presente documento, el término "cultivo de tejidos" indica una composición que comprende células aisladas del mismo tipo o de un tipo diferente o una colección de tales células organizadas en partes de una planta. Ejemplos de cultivos de tejidos son protoplastos, callos, grupos de plantas y células vegetales que pueden generar cultivos de tejidos intactos en plantas o partes de plantas, como embriones, polen, flores, semillas, hojas, tallos, raíces, puntas de raíces, anteras, pistilos, células meristemáticas, yemas axilares, ovarios, cubierta de semillas, endospermo, hipocotilos, cotiledones y similares. En algunas formas de realización de la invención, se proporciona un cultivo de tejido transgénico o cultivo de células vegetales transgénicas, en donde el cultivo de tejido o célula transgénico comprende una molécula de ácido nucleico/secuencia de nucleótidos de la invención. En algunas formas de realización, los transgenes pueden eliminarse de una planta desarrollada a partir del tejido o célula transgénicos mediante la reproducción de la planta transgénica con una planta no transgénica, y seleccionando de la progenie las plantas que comprenden la edición del gen deseado y no los transgenes utilizados en la producción de la edición.

"Progenie" comprende cualquier generación posterior de una planta.

Los términos "transgén" o "transgénico", como se usan en el presente documento, se refieren a por lo menos una secuencia de ácido nucleico que se toma del genoma de un organismo, o se produce sintéticamente, y que luego se introduce en una célula, u organismo o tejido huésped de interés y que posteriormente se integra en el genoma del huésped mediante enfoques de transformación o transfección "estable". En contraste, el término transformación o transfección o introducción "transitoria" se refiere a una forma de introducir herramientas moleculares que incluyen por lo menos un ácido nucleico (ADN, ARN, monocatenario o bicatenario o una mezcla de los mismos) y/o por lo menos una secuencia de aminoácidos, que opcionalmente comprende agentes químicos o biológicos adecuados, para lograr una transferencia en por lo menos un compartimento de interés de una célula, que incluye, entre otros, el citoplasma, un orgánulo, que incluye el núcleo, una mitocondria, una vacuola, un cloroplasto, o dentro de una membrana, que produce la transcripción y/o traducción y/o asociación y/o actividad de la por lo menos una molécula introducida sin lograr una integración o incorporación estable y, por lo tanto, la herencia de la respectiva por lo menos una molécula introducida en el genoma de una célula. Los términos "libre de transgén" se refieren a una condición en la que el transgén no está presente o no se encuentra en el genoma de una célula o tejido u organismo huésped de interés.

El "rasgo agronómico" es un parámetro medible que incluye, entre otros, falta de semillas (p. ej., producción de semillas reducida o nula), tamaño de semilla reducido, grosor reducido del endocarpio de la fruta, lignificación reducida del endocarpio de la fruta, disminución de la simiente, verdor de hojas, rendimiento, tasa de crecimiento, biomasa, peso fresco en la maduración, peso seco en la maduración, rendimiento de fruta, rendimiento de semilla, contenido total de nitrógeno de la planta, contenido de nitrógeno de la fruta, contenido de nitrógeno de la semilla, contenido de nitrógeno en un tejido vegetativo, contenido total de aminoácidos libres de plantas, contenido de aminoácidos libres de fruto, contenido de aminoácidos libres de semillas, contenido de aminoácidos libres en un tejido vegetativo, contenido total de proteína vegetal, contenido de proteína frutal, contenido de proteína de semillas, contenido de proteína en un tejido vegetativo, tolerancia a la sequía, absorción de nitrógeno, alojamiento de raíces, índice de cosecha, alojamiento del tallo, altura de la planta, altura de la mazorca, longitud de la mazorca, resistencia a enfermedades, resistencia al frío, tolerancia a la sal y número de vástagos, etc.

Los métodos generales en bioquímica molecular y celular se pueden encontrar en libros de texto estándares tales como *Molecular Cloning: A Laboratory Manual*, 3rd Ed. (Sambrook *et al.*, HaRBor Laboratory Press 2001); *Short Protocols in Molecular Biology*, 4th Ed. (Ausubel *et al.* eds., John Wiley & Sons 1999); *Protein Methods* (Bollag *et al.*, John Wiley & Sons 1996); *Nonviral Vectors for Gene Therapy* (Wagner *et al.* eds., Academic Press 1999); *Viral Vectors* (Kaplift & Loewy eds., Academic Press 1995); *Immunology Methods Manual* (I. Lefkovits ed., Academic Press 1997); y *Cell and Tissue Culture: Laboratory Procedures in Biotechnology* (Doyle & Griffiths, John Wiley & Sons 1998), cuyas divulgaciones se incorporan en este documento por referencia.

Como se usa en el presente documento, "contacto", "poner en contacto", "contactado" y sus variaciones gramaticales se refieren a colocar los componentes de una reacción deseada juntos en condiciones adecuadas para llevar a cabo la reacción deseada (por ejemplo, transformación, control transcripcional, edición del genoma, mellado y/o escisión). Como ejemplo, se puede poner en contacto un ácido nucleico diana con una proteína de unión de ADN específica de secuencia (por ejemplo, endonucleasa guiada por polinucleótido, una endonucleasa CRISPR–Cas (por ejemplo, proteína efectora CRISPR–Cas), una nucleasa de dedo de zinc, una nucleasa efectora de tipo activador de la transcripción (TALEN) y/o una proteína Argonauta)) y una desaminasa o una construcción de ácido nucleico que la codifica, en condiciones en las que se expresan la proteína de unión de ADN específica de secuencia, la transcriptasa inversa y/o la desaminasa, y la proteína de unión de ADN específica de secuencia se une al ácido nucleico diana, y la transcriptasa inversa y/o

la desaminasa pueden fusionarse a la proteína de unión de ADN específica de secuencia o reclutarse en la proteína de unión de ADN específica de secuencia (mediante, por ejemplo, un etiqueta peptídica fusionada a la proteína de unión de ADN específica de secuencia y una etiqueta de afinidad fusionada a la transcriptasa inversa y/o desaminasa) y, por lo tanto, la desaminasa y/o transcriptasa inversa se colocan cerca del ácido nucleico objetivo, de manera de modificar así el ácido nucleico objetivo. Pueden usarse otros métodos para reclutar transcriptasa inversa y/o desaminasa que aprovechan otras interacciones proteína-proteína, y también pueden usarse interacciones ARN-proteína e interacciones químicas para el reclutamiento proteína-proteína y proteína-ácido nucleico.

Como se usa en el presente documento, "modificar" o "modificación" con referencia a un ácido nucleico diana incluye editar (por ejemplo, mutar), modificación covalente, intercambiar / sustituir ácidos nucleicos / bases de nucleótidos, deletear, escindir, mellar y/o alterar el control transcripcional de un ácido nucleico diana. En algunas formas de realización, una modificación puede incluir uno o más cambios de base única (SNP) de cualquier tipo.

"Introducir", "introducción", "introducido" (y sus variaciones gramaticales) en el contexto de un polinucleótido de interés significa presentar una secuencia de nucleótidos de interés (por ejemplo, polinucleótido, patrón RT, una construcción de ácido nucleico y/o un ácido nucleico guía) a una planta, parte de la planta o célula de la misma, de tal manera que la secuencia de nucleótidos tenga acceso al interior de una célula.

Los términos "transformación" o "transfección" se pueden usar indistintamente y como se usa en el presente documento se refieren a la introducción de un ácido nucleico heterólogo en una célula. La transformación de una célula puede ser estable o transitoria. Por lo tanto, en algunas formas de realización, una célula huésped u organismo huésped (por ejemplo, una planta) puede transformarse de manera estable con una molécula de polinucleótido/ácido nucleico de la invención. En algunas formas de realización, una célula huésped u organismo huésped puede transformarse transitoriamente con una molécula de polinucleótido/ácido nucleico de la invención.

La "transformación transitoria" en el contexto de un polinucleótido significa que se introduce un polinucleótido en la célula y no se integra en el genoma de la célula.

Por "introducción estable" o "introducido de manera estable" en el contexto de un polinucleótido introducido en una célula se pretende que el polinucleótido introducido se incorpore de forma estable en el genoma de la célula, y por lo tanto la célula se transforma de manera estable con el polinucleótido.

"Transformación estable" o "transformada establemente" como se usa en el presente documento significa que una molécula de ácido nucleico se introduce en una célula y se integra en el genoma de la célula. De este modo, la molécula de ácido nucleico integrada es capaz de ser heredada por su progenie, más particularmente, por la progenie de múltiples generaciones sucesivas. "Genoma", como se usa en el presente documento, incluye el genoma nuclear y el plastídico, y por lo tanto incluye la integración del ácido nucleico en, por ejemplo, el genoma mitocondrial y cloroplasto. La transformación estable como se usa en el presente documento también puede referirse a un transgén que se mantiene extracromosómicamente, por ejemplo, como un minicromosoma o un plásmido.

La transformación transitoria puede detectarse, por ejemplo, mediante un ensayo inmunoabsorbente ligado a enzimas (ELISA, conforme a sus siglas en inglés) o transferencia Western, que puede detectar la presencia de un péptido o polipéptido codificado por uno o más transgenes introducidos en un organismo. La transformación estable de una célula puede detectarse, por ejemplo, mediante un ensayo de hibridación Southern blot de ADN genómico de la célula con secuencias de ácido nucleico que hibridan específicamente con una secuencia de nucleótidos de un transgén introducido en un organismo (por ejemplo, una planta). La transformación estable de una célula se puede detectar, por ejemplo, mediante un ensayo de hibridación de transferencia Northern de ARN de la célula con secuencias de ácido nucleico que hibridan específicamente con una secuencia de nucleótidos de un transgén introducido en un organismo huésped. La transformación estable de una célula también se puede detectar, por ejemplo, mediante una reacción en cadena de la polimerasa (PCR, conforme a sus siglas en inglés) u otras reacciones de amplificación como se conocen bien en la técnica, empleando secuencias cebadoras específicas que se hibridan con las secuencias diana de un transgén, lo que produce la amplificación de la secuencia transgénica, que puede detectarse de acuerdo con métodos estándares. La transformación también puede detectarse mediante protocolos de secuenciación directa y/o hibridación bien conocidos en la técnica.

Por consiguiente, en algunas formas de realización, las secuencias de nucleótidos, los polinucleótidos, las construcciones de ácido nucleico y/o los casetes de expresión de la invención pueden expresarse de manera transitoria y/o pueden incorporarse de manera estable en el genoma del organismo huésped. Por lo tanto, en algunas formas de realización, una construcción de ácido nucleico de la invención (por ejemplo, uno o más casetes de expresión que comprenden polinucleótidos para edición como se describe en el presente documento) puede introducirse

transitoriamente en una célula con un ácido nucleico guía y, de este modo, no se mantiene ADN en la célula.

Una construcción de ácido nucleico de la invención puede introducirse en una célula vegetal mediante cualquier método conocido por los expertos en la técnica. Los ejemplos no limitativos de métodos de transformación incluyen la transformación a través del suministro de ácido nucleico mediado por bacterias (por ejemplo, a través de agrobacterias), el suministro de ácido nucleico mediado por virus, suministro de ácido nucleico mediado por carburo de silicio o virutas de ácido nucleico, suministro de ácido nucleico mediado por liposomas, microinyección, bombardeo de micropartículas, transformación mediada por fosfato de calcio, transformación mediada por ciclodextrina, electroporación, transformación mediada por nanopartículas, sonicación, infiltración, absorción de ácido nucleico mediada por PEG, así como cualquier otro mecanismo eléctrico, químico, físico (mecánico) y/o biológico que logra la introducción de ácido nucleico en la célula vegetal, incluida cualquiera de sus combinaciones. Los procedimientos para transformar los organismos eucariotas y procariotas son bien conocidos y rutinarios en la técnica y se describen a lo largo de la literatura (véase, por ejemplo, *Jiang et al. 2013. Nat. Biotechnol.* 31:233–239; Ran et al. *Nature Protocols* 8:2281–2308 (2013)). Las guías generales para diversos métodos de transformación de plantas conocidos en la técnica incluyen aquellas en las referencias de Miki et al. ("Procedures for Introducing Foreign DNA into Plants" en *Methods in Plant Molecular Biology and Biotechnology*, Glick, B. R. y Thompson, J. E., Eds. (CRC Press, Inc., Boca Raton, 1993), pp. 67–88) y Rakowoczy–Trojanowska (*Cell. Mol. Biol. Lett.* 7:849–858 (2002)).

En algunas formas de realización de la invención, la transformación de una célula puede comprender transformación nuclear. En otras formas de realización, la transformación de una célula puede comprender la transformación de plástidos (por ejemplo, transformación de cloroplasto). En otras formas de realización adicionales, los ácidos nucleicos de la invención pueden introducirse en una célula mediante técnicas de reproducción convencionales. En algunas formas de realización, uno o más de los polinucleótidos, casetes de expresión y/o vectores pueden introducirse en una célula vegetal mediante la transformación de *Agrobacterium*.

Por lo tanto, un polinucleótido puede introducirse en una planta, parte de planta, célula de planta en cualquier número de formas que son bien conocidas en la técnica. Los métodos de la invención no dependen de un método particular para introducir una o más secuencias de nucleótidos en una planta, solo de que obtengan acceso al interior de la célula. Cuando se va a introducir más de polinucleótido, se pueden ensamblar como parte de una construcción de ácido nucleico única, o

como construcciones de ácido nucleico separadas, y se pueden ubicar en las mismas o diferentes construcciones de ácido nucleico. Por consiguiente, el polinucleótido se puede introducir en la célula de interés en un solo evento de transformación, o en eventos de transformación separados, o, alternativamente, se puede incorporar un polinucleótido en una planta como parte de un protocolo de reproducción.

Como se usa en el presente documento, el término "factor de transcripción clado AGAMOUS" o "factor de transcripción clado AG" es un miembro de la subfamilia AGAMOUS (AG) de genes de caja MADS tipo MIKC. Las proteínas "tipo MIKC" representan una clase de factores de transcripción del dominio MADS y están definidas por una estructura de dominio única: (1) 'M' – un dominio MADS de unión de ADN altamente conservado, (2) 'I' – un dominio intermedio, (3) 'K' – un dominio K de tipo queratina, y (4) 'C' – un dominio C-terminal. En algunas formas de realización, el "Factor de transcripción clado AGAMOUS" o el "factor de transcripción clado AG" comprenden además una región N-terminal. En formas de realización adicionales, el "Factor de transcripción clado AGAMOUS" o el "factor de transcripción clado AG" comprende genes AG, SHP1, SHP2 y STK en plantas de la presente descripción, cada uno de los cuales tiene un motivo NN en el dominio M, un motivo YQQ en el dominio K, y/o un R / Q (R o Q) en el dominio C.

Como se usa en el presente documento, el término "una sustitución de nucleótidos única estable y dirigida" se refiere a una sustitución de nucleótidos no natural mediada por una tecnología de edición de bases (que incluye CBE y/o ABE) que permite la conversión directa e irreversible de un par de bases a otro en un locus genómico objetivo sin requerir roturas de ADN bicatenario (DSB), procesos de reparación dirigida por homología (HDR) o patrones de ADN donante. La tecnología de edición de bases que se enseña en la presente descripción también se combina con un sistema CRISPR-Cas para la focalización precisa, que permite una sustitución de nucleótidos única estable y dirigida. Como se mencionó anteriormente, se puede hacer más de una sustitución de nucleótido única con los sistemas de edición enseñados aquí, en uno o más genes de interés. La sustitución de nucleótido único estable y dirigida es, por lo tanto, una sustitución de nucleótidos no natural que resulta de los esfuerzos del productor agrícola que pone en práctica los métodos de fitomejoramiento de próxima generación de la divulgación.

Las plantas útiles para modificar como se describe en el presente documento incluyen cualquier planta para la que sea deseable un desarrollo alterado del fruto, opcionalmente falta de semillas, o reducción de la simiente. Dichas plantas pueden incluir, entre otras, aquellas de la familia

de plantas *Rosaceae* (rosáceas). Como se usa en el presente documento, el término "familia de plantas *Rosaceae*", "familia *Rosaceae*" o "*Rosaceae*" se refiere a una familia de plantas con flores, que incluye 4828 especies conocidas en 91 géneros. La familia *Rosaceae* incluye hierbas, arbustos y árboles. Muchos productos económicamente importantes provienen de las rosáceas, incluidas muchas frutas comestibles (como manzanas, peras, membrillos, damascos, ciruelas, cerezas, duraznos, moras, frambuesas, frambuesas negras, nísperos, dátiles y frutillas), almendras y arbustos y árboles ornamentales (como rosas, espireas, fotinias, espinos de fuego, cenizos de la montaña y majuelos). Los ejemplos no limitantes de plantas de *Rosaceae* que pueden modificarse como se describe en el presente documento incluyen, sin limitación, el género *Rubus* (por ejemplo, moras, frambuesa negra o frambuesa, y similares), el género *Prunus*, el género *Fragaria* y/o el género *Malus*. Ejemplos de plantas *Rubus* útiles con la invención pueden incluir, sin limitación, *Rubus occidentalis* L., *Rubus pergratus* Blanch., *Rubus oklahomus* L. H. Bailey *Rubus originalis* L. H. Bailey, *Rubus ortivus* (L. H. Bailey) L. H. Bailey, *Rubus parcifrondifer* L. H. Bailey, *Rubus odoratus* L., *Rubus parvifolius* L., *Rubus pedatus* Sm., y *Rubus phoenicolasius* Maxim. Ejemplos de plantas del género *Prunus* útiles con la invención pueden incluir, sin limitación, *P. persica*, *P. pyrifolia*, *P. serotina*, *P. armeniaca*, *P. spinosa*, *P. avium*, o *P. dulcis* (por ejemplo, ciruela, damasco, cereza, nectarina, durazno, almendra, cerezo de Virginia, laurel cerezo y endrino). Ejemplos de plantas del género *Fragaria* útiles con la invención pueden incluir, sin limitación, *F. vesca*, *Fragaria x ananassa* Duchesne, o *F. chiloensis*. Ejemplos de plantas del género *Malus* útiles con la invención pueden incluir, sin limitación, *M. domesticus*, *Pyrus communis*, *Cydonia oblonga*, género *Crataegus*, género *Chaenomeles*, o género *Amelanchier*. En algunas formas de realización, una planta *Rosaceae* o parte de la misma es una del género *Rubus* o una drupa o fruto de hueso. En algunas formas de realización, la planta *Rosaceae* o parte de la misma es una mora, frambuesa negra, cereza, ciruela o durazno.

En algunas formas de realización, una planta útil con esta invención puede incluir, pero sin limitación, plantas angiospermas y gimnospermas tales como acacia, alfalfa, amaranto, manzana, damasco, alcaucil, fresno, espárragos, palta, banana, cebada, porotos, remolacha, abedul, haya, mora, frambuesa negra, arándano, brócoli, coles de Bruselas, repollo, canola, melón, zanahoria, mandioca, coliflor, cedro, un cereal, apio, castaño, cereza, col china, cítricos, clementina, trébol, café, maíz, algodón, caupí, pepino, ciprés, berenjena, olmo, endibia, eucalipto, hinojo, higos, abeto, geranio, uva, pomelo, maní, alquequenje, cicuta, nuez dura, col rizada, kiwi, colinabo, alerce, lechuga, puerro, limón, lima, falsa acacia, pino, nogal de Japón, maíz, mango, arce, melón, mijo, champiñón, mostaza, nueces, roble, avena, palma aceitera (*Elaeis guineensis*), okra, cebolla, naranja, una planta ornamental

o flor o árbol, papaya, palma, perejil, chirivía, arveja, durazno, maní, pera, turba, pimienta, caqui, guandú, durazno, pino, piña, plátano, ciruela, granada, papa, calabaza, achicoria, rábano, colza, frambuesa, arroz, centeno, sorgo, cártamo, sauce llorón, soja, espinaca, picea, calabaza, frutilla, remolacha azucarera, caña de azúcar, girasol, batata, maíz dulce, mandarina, té, tabaco, tomate, árboles, triticale, césped, nabos, vid, nuez, berro, sandía, trigo, frutilla silvestre, ñame, tejo y zucchini, opcionalmente en donde la planta es mora, frambuesa, frambuesa negra, frutilla, cereza, durazno, ciruela, damasco, manzana, pera, membrillo, níspero, dátiles y almendras.

Metodología de fitomejoramiento de próxima generación utilizando mutaciones en factores de transcripción de caja MADS clado AGAMOUS para el desarrollo de fruto alterado.

A. Identificación de factores de transcripción de caja MADS clado AGAMOUS.

Un factor de transcripción de caja MADS clado AGAMOUS comprende por lo menos los siguientes tres dominios de aminoácidos conservados: (1) motivo Y{A,S}NN en las posiciones 83–86 con respecto a

Arabidopsis thaliana STK (AT4G09960.3); (2) motivo YQQE{A,S}{A,S,N,K,T}KL{R,H}{Q,H,A,N,R}QI en las posiciones 116–127 con respecto a *Arabidopsis thaliana* STK (AT4G09960.3); y (3) R en la posición 225 con respecto a *Arabidopsis thaliana* STK (AT4G09960.3).

En algunas formas de realización alternativas, el factor de transcripción de caja MADS clado AGAMOUS comprende los dominios anteriormente mencionados en (1) y (2), pero tiene uno expandido (3), que incluye el R en la posición 225 con respecto a *Arabidopsis thaliana* STK (AT4G09960.3), excepto el gen Grape VIT_12s0142g00360.0, que contiene un Q en esta posición. Este “factor de transcripción de caja MADS clado AGAMOUS expandido” por lo tanto incluye R y Q correspondientes a la posición 225 de *Arabidopsis thaliana* STK (AT4G09960.3).

En algunas formas de realización, la presente divulgación enseña métodos desarrollados por los inventores, que utilizan bioinformática y análisis evolutivos de diversos factores de transcripción de caja MADS para identificar factores de transcripción de caja MADS clado AGAMOUS. En algunas formas de realización, las secuencias de gen de caja MADS de *Arabidopsis thaliana* se usan para la bioinformática, tal como búsquedas Blast de nucleótidos y/o proteínas para identificar y recoger información de secuencia en homólogos, ortólogos y/o parálogos de los genes de caja MADS de *Arabidopsis* de interés en otras plantas. En algunas formas de realización, las secuencias de aminoácidos recogidas correspondientes a los factores de transcripción de caja MADS se analizan

adicionalmente para la caracterización de similitud estructural y características funcionales entre los factores de transcripción de caja MADS usando alineación de secuencia múltiple (MSA, conforme a sus siglas en inglés) y análisis filogenéticos.

En algunas formas de realización, la presente divulgación enseña especies de los factores de transcripción de caja MADS clado AGAMOUS, que pueden comprender: STK, AG y SHP1/2 a través de plantas de interés, que incluyen: *Arabidopsis*, frambuesa negra, frutilla silvestre, durazno, manzana, tomate y uva de árboles filogenéticos como se describe en la Solicitud Provisional de los EE. UU. Nro. 62/859.992, presentada el 11 de junio de 2019. En otras formas de realización, la presente divulgación enseña especies de factores de transcripción de la caja MADS del clado AGAMOUS, que pueden comprender: STK, AG y SHP1/2 a través de cultivos de interés incluidos, entre otros, frambuesa, frambuesa negra, mora, frutilla silvestre, cereza, durazno, ciruela, palta, manzana, tomate, dátiles y uva. En formas de realización adicionales, los cultivos de interés enseñados en la presente descripción son frambuesa negra, mora, cereza, durazno y palta.

En algunas formas de realización, los residuos conservados, motivos y/o regiones de factores de transcripción clado AG identificados a partir de los análisis MSA y filogenético se analizan adicionalmente para identificar posibles sitios objetivo para introducir modificaciones genéticas que incluyen, sin limitación, edición de bases, tecnologías de edición del genoma para alterar el desarrollo del fruto en plantas de interés. En formas de realización adicionales, los codones que codifican estos residuos seleccionados en las regiones conservadas también se analizan en busca de ediciones potenciales que podrían afectar un cambio no conservador en el aminoácido, dando como resultado una variación en la función de la proteína.

En algunas formas de realización, los sitios diana para la modificación genética de los codones que codifican los residuos seleccionados están en un péptido de señal nuclear del factor de transcripción clado AG. En algunas formas de realización, los sitios objetivo para la modificación genética de los codones que codifican los residuos seleccionados están en un dominio N-terminal del factor de transcripción clado AG. En algunas formas de realización, los sitios diana para la modificación genética de los codones que codifican los residuos seleccionados están en un dominio de caja MADS (M) del factor de transcripción clado AG. En algunas formas de realización, los sitios objetivo para la modificación genética de los codones que codifican los residuos seleccionados están en un dominio Intermedio (I) del factor de transcripción clado AG. En algunas formas de realización, los sitios diana para la modificación genética de los codones que codifican los residuos seleccionados están en un dominio de queratina (K) del factor de transcripción clado AG. En algunas formas de

realización, los sitios objetivo para la modificación genética de los codones que codifican los residuos seleccionados están en un dominio C-terminal (C) del factor de transcripción clado AG, ver, como ejemplo, la **FIG. 4**. En otras formas de realización, los sitios objetivo para la modificación genética de los codones que codifican los residuos seleccionados están en un motivo NN en el dominio M del factor de transcripción clado AG. En otras formas de realización, los sitios objetivo para la modificación genética de los codones que codifican los residuos seleccionados están en un motivo YQQ en el dominio M del factor de transcripción clado AG. En otras formas de realización, los sitios objetivo para la modificación genética de los codones que codifican los residuos seleccionados están en un residuo R/Q en el dominio C del factor de transcripción clado AG. En formas de realización adicionales, el motivo NN comprende un motivo extendido Y[A/S]NN como sitios objetivo para la modificación genética. En algunas formas de realización, el motivo puede ser YQQE (**SEQ ID NO: 177**). En algunas formas de realización, el motivo YQQ o YQQE puede comprender un motivo extendido, YQQE[A/S][A/S/N/K/T]KL[R/H][Q/H/A/N/R]QI como sitios objetivo para la modificación genética (ver, por ejemplo, **SEQ ID NO: 178, 179, 180, 181, o 182**). En algunas formas de realización, un sitio objetivo para la modificación de un factor de transcripción STK es una arginina conservada (R conservada) ubicada en la posición 225 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 13**; posición 196 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 66**; posición 197 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 53, 78, 150, 166–171**; posición 198 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 29**; posición 199 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 128**; o posición 201 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 97, 152, 154, 156, 158, 160, 162, 164**). En algunas formas de realización, un dominio en un polipéptido STK que puede ser apuntado para la modificación puede ser un dominio que comprende un R conservado que incluye, sin limitación, NVLVRAKIADLERLQHADMVSGDQDLELNAIQALVSRNFFTS (**SEQ ID NO: 183**) o un dominio que tiene cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 187, 188, o 189**, en donde la modificación puede ser una delección de uno o más aminoácidos, opcionalmente en donde la delección produce un truncamiento C-terminal del polipéptido STK. En algunas formas de realización, como mínimo, un truncamiento C-terminal útil para esta invención elimina por lo menos la arginina conservada ubicada en la posición 225 de **SEQ ID NO: 13** (AtSTK), 196 de **SEQ ID NO: 66**; posición 197 de **SEQ ID NOs: 53, 78, 150, 166–171**; 198 de **SEQ ID NO: 29**; 199 de **SEQ ID NO: 128** o 201 de **SEQ ID NO: 97** (RoSTK)) o **SEQ ID NOs: 152, 154, 156, 158, 160, 162, 164**).

Ejemplos de mutaciones que producen un truncamiento C-terminal se proporcionan en la **Fig. 5**. La presente divulgación enseña que **SEQ ID NO: 13** es AT4G09960.3, también denominado AGAMOUS-LIKE 11 (AGL11) o SEEDSTICK (STK), que es un factor de transcripción de caja MADS expresado en el carpelo y los óvulos. AtSTK (a.k.a. AtAGL11) y sus ortólogos en otras plantas (ver, por ejemplo, **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164 o 166–171**) cumplen un rol maternal en la fertilización y el desarrollo de semillas, y controlan la estructura y las propiedades mecánicas de la cubierta de la semilla.

En algunas formas de realización, **SEQ ID NO: 13** es AT4G09960.3 (Proteína STK de *Arabidopsis*) que incluye una región N-terminal (1–26 aminoácidos de AT4G09960.3 Secuencia de proteína) como una secuencia de referencia para identificar homólogos, ortólogos y/o parálogos en otras especies de plantas. En formas de realización adicionales, la proteína STK de *Arabidopsis* STK comprende: un motivo 'NN' en las posiciones 85–86 de **SEQ ID NO: 13**, un motivo 'YQQ' en las posiciones 116–118 de **SEQ ID NO: 13** y un "R (Arginina)" en la posición 225 de **SEQ ID NO: 13**, como se presenta en la **Tabla 1** y la **Tabla 1B** (los motivos están en negritas y subrayados).

La presente divulgación enseña que la **SEQ ID NO: 97** es una proteína STK de frambuesa negra (Ro05_G22454) sin una región N-terminal. Cuando se compara con la secuencia STK de *Arabidopsis* (**SEQ ID NO: 13**), la secuencia STK de frambuesa negra no contiene una región N-terminal correspondiente a 1–26 de **SEQ ID NO: 13** (26 aminoácidos del codón de inicio). En otras formas de realización, la proteína STK de frambuesa negra comprende: un motivo 'NN' en las posiciones 59–60 correspondientes a 85–86 de **SEQ ID NO: 13**, un motivo 'YQQ' en las posiciones 91–93 correspondientes a 116–118 de **SEQ ID NO: 13** y una "R (Arginina)" en la posición 201 correspondiente a 225 de **SEQ ID NO: 13** como se presenta en la **Tabla 1** y la **Tabla 1B** (los motivos están en negrita y subrayados).

La presente descripción enseña que la **SEQ ID NO: 29** es proteína STK de bosque o frutilla silvestre (FvH4_5g32540.1) sin una región N-terminal. Cuando se compara con la secuencia STK de *Arabidopsis* (**SEQ ID NO: 13**), la secuencia STK de frambuesa negra no contiene una región N-terminal correspondiente a 1–26 de **SEQ ID NO: 13** (26 aminoácidos desde el codón de inicio). En otras formas de realización, la proteína STK de frambuesa negra comprende: un motivo 'NN' en las posiciones 59–60 correspondientes a 8586 de **SEQ ID NO: 13**, un motivo 'YQQ' en las posiciones 91–93 correspondientes a 116–118 de **SEQ ID NO: 13** y una "R (Arginina)" en la posición 198 correspondiente a 225 de la **SEQ ID NO: 13** como se presenta en la **Tabla 1** y la **Tabla 1B** (los motivos están en negrita y subrayados).

Los proteomas de las especies se consultaron con la secuencia de codificación de proteínas para STK de *Arabidopsis* del genoma de *A. thaliana* (ID de proteína AT4G09960.3 en la versión de ensamblaje TAIR10) usando Blast (blastp Ver. 2.5.0; parámetros: '-seg no -max_hsps 1 -max_objetivo_seqs 40 -uso_sw_tretro ') y aciertos Blast de secuencia de proteína extraída al pasar el valor $E < 1e-10$ de corte. Se realizó una segunda iteración de la búsqueda de Blast utilizando todos los aciertos de Blast que pasaban como secuencias de consulta, para generar puntajes de valor E recíprocos, que se usaron en la agrupación. La agrupación se realizó con el programa hcluster_sg (Li 2006; Ver. 0.5.1, parámetros: "-m 750 -w 0 -s 0.34"). Las secuencias se alinearon utilizando el programa de alineación de secuencias múltiples T-Coffee con ajustes de parámetros predeterminados (Notredame et al. 2000; Version_11.00.8cbe486). Para la reconstrucción de árboles filogenéticos, utilizamos ModelFinder para la selección del modelo de sustitución (Kalyaanamoorthy et al. 2017) e IQ-Tree para la inferencia de árboles y la estimación de soporte de ramas con aproximación de arranque ultrarrápido (Hoang et al. 2018, Nguyen et al. 2015; IQ-Tree Ver. 1.5.5; parámetros: "-bb 1000") usando secuencias de codificación de ADN después de la traducción inversa con la herramienta de soporte lógico tranalign del paquete EMBOSS (Rice et al. 2000. Ver. 6.6.0.0). Se encontró evidencia de ortología al examinar la topología de los árboles para agrupar genes de múltiples especies para las cuales el nodo ancestro más común es un evento de especiación. La alineación de secuencia múltiple para el grupo que contiene STK de *Arabidopsis* se investigó en busca de motivos de aminoácidos conservados en la subfamilia de genes de caja MADS Agamous (* Referencias agregadas debajo de la sección de referencia existente).

Tabla 1. Información de secuencia de SEQ ID NO: 13, SEQ ID NO: 97 y SEQ ID NO: 29.

SEQ ID NO	Secuencia de aminoácidos
SEQ ID NO: 13 (AT4G09960.3)	MLFPHERKKEKERSQGFYLVTRLRIRMGRGKIEIKRIENSTNRQ VTFCKRRNGLLKAYELSVLCDAEVALIVFSTRGRLYEYANN IRSTIERYKKACSDSTNTSTVQEINAAYYQQESAKLRQQIQTIQ NSNRNLMGDSLSSLVKELKQVENRLEKAISRIRSKKHELLLVE IENAQKREIELDNENIYLRTKVAEVERYQQHHHQMVSQSEINAI EALASRNYFAHSIMTAGSGSGNGGSYSDPKKILHLG

SEQ ID NO: 97 (Ro05_G22454)	MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSILCEAEVALIVFSSRGRLYEYS <u>NN</u> NSIRNTIERYKKACSDSSGATTITEINAQY <u>YQQ</u> ESAKLRHQIQMLQNSNRHLMGDSLSNLTVKELKQLENRLERGLTRIRSKKHEMLLAEIEYLQKREVELENENVLVRAKIAELERLQHADMVSGDQDLELNAIQALVS <u>R</u> NFFASTMIEGEASYSQPEKKFLNLGAGKGLVKQGKTSSSGFYVL
SEQ ID NO: 29 (FvH4_5g32540.1)	MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSILCEAEVALIVFSSRGRLYEYS <u>NN</u> NSIRNTIERYKKASSDMSGATTITEINAQY <u>YQQ</u> ESTKLRHQIQMLQNSNRHLMGDSLSNLTVKELKQLENRLERGLTRIRSKKHEMLLAEIEYLQKREIELENENVLIRAKIAEVE RLQQADLVSGAELNAIQALAS <u>R</u> NFFESTMMEGETSYSQPEKKL LHLG

Tabla 1B. Tabla de anotación para SEQ ID NO: 13, SEQ ID NO: 97, y SEQ ID NO: 29.

ID Secuencia	Descripción	AA Inicio	AA Fin	AA Longitud
SEQ ID NO: 13 (AT4G09960)	Dominio específico de Ag extendido	83	86	4
SEQ ID NO: 13 (AT4G09960)	Dominio específico de Ag extendido	116	127	12
SEQ ID NO: 13 (AT4G09960)	Dominio específico de Ag extendido	225	225	1
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	85	86	2
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	98	98	1
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	101	101	1
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	116	118	3
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	124	124	1
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	126	126	1
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	131	131	1
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	182	182	1
SEQ ID NO: 13 (AT4G09960)	Motivo específico de Ag	225	225	1
SEQ ID NO: 13 (AT4G09960)	C-terminal	197	256	60
SEQ ID NO: 13 (AT4G09960)	Dominio intermedio	103	109	7

SEQ ID NO: 13 (AT4G09960)	Caja K	110	196	87
SEQ ID NO: 13 (AT4G09960)	tipo MADS MEF2	27	102	76
SEQ ID NO: 13 (AT4G09960)	N-terminal	1	26	26
SEQ ID NO: 97 (Ro05_G22454)	Dominio específico de Ag extendido	57	60	4
SEQ ID NO: 97 (Ro05_G22454)	Dominio específico de Ag extendido	91	102	12
SEQ ID NO: 97 (Ro05_G22454)	Dominio específico de Ag extendido	201	201	1
SEQ ID NO: 97 (Ro05_G22454)	Motivo específico de Ag	73	73	1
SEQ ID NO: 97 (Ro05_G22454)	Motivo específico de Ag	76	76	1
SEQ ID NO: 97 (Ro05_G22454)	Motivo específico de Ag	91	93	3
SEQ ID NO: 97 (Ro05_G22454)	Motivo específico de Ag	99	99	1
SEQ ID NO: 97 (Ro05_G22454)	Motivo específico de Ag	101	101	1
SEQ ID NO: 97 (Ro05_G22454)	Motivo específico de Ag	106	106	1
SEQ ID NO: 97 (Ro05_G22454)	Motivo específico de Ag	157	157	1
SEQ ID NO: 97 (Ro05_G22454)	Motivo específico de Ag	201	201	1
SEQ ID NO: 29 (FvH4_5g32540.1)	Dominio específico de Ag extendido	57	60	4
SEQ ID NO: 29 (FvH4_5g32540.1)	Dominio específico de Ag extendido	91	102	12
SEQ ID NO: 29 (FvH4_5g32540.1)	Dominio específico de Ag extendido	198	198	1
SEQ ID NO: 29 (FvH4_5g32540.1)	Motivo específico de Ag	73	73	1
SEQ ID NO: 29 (FvH4_5g32540.1)	Motivo específico de Ag	76	76	1
SEQ ID NO: 29 (FvH4_5g32540.1)	Motivo específico de Ag	91	93	3
SEQ ID NO: 29 (FvH4_5g32540.1)	Motivo específico de Ag	99	99	1
SEQ ID NO: 29 (FvH4_5g32540.1)	Motivo específico de Ag	101	101	1
SEQ ID NO: 29 (FvH4_5g32540.1)	Motivo específico de Ag	106	106	1
SEQ ID NO: 29 (FvH4_5g32540.1)	Motivo específico de Ag	157	157	1
SEQ ID NO: 29 (FvH4_5g32540.1)	Motivo específico de Ag	198	198	1

B. Generación de alelos mutantes en factores de transcripción caja MADS clado AGAMOUS para crear fenotipos de desarrollo de fruto alterado.

Aquí se proporcionan métodos para generar modificaciones en los factores de transcripción de la caja MADS del clado AGAMOUS, que son clave para el desarrollo de semillas y frutos. Dicha modificación incluye, sin limitación, aquellas que generan alelos negativos dominantes, alelos semidominantes, alelos de pérdida de función débil o mutaciones hipomórficas. En la presente descripción, el producto genético mutado todavía puede interactuar con otros elementos dentro de la célula como lo haría el producto génico de tipo salvaje, pero algunas funciones del producto génico se han cambiado o reducido o bloqueado por completo. Por ejemplo, una mutación en un factor de transcripción, tal como un factor de transcripción de la caja MADS, podría eliminar el dominio de activación pero mantener el dominio de unión de ADN. Este producto alterado todavía puede unirse al sitio de ADN, de manera de competir con el factor de transcripción de tipo salvaje y, por lo tanto, conducir a niveles reducidos de activación del gen objetivo. En consecuencia, un fenotipo alterado del desarrollo de la fruta resulta de la actividad reducida de los genes diana debido a los alelos negativos dominantes, los alelos semidominantes, los alelos de pérdida de función débil o las mutaciones hipomórficas generadas en los factores de transcripción de la caja MADS del clado AGAMOUS.

C. Cribado de HTP *in vivo* para identificar alelos modificados en factores de transcripción de caja MADS de clado AGAMOUS candidatos.

En algunas formas de realización, los sitios diana candidatos se mutan y luego se criban mediante un ensayo de luciferasa *in vivo* en un sistema heterólogo tal como tabaco, levadura o células humanas para seleccionar ediciones efectivas que afecten la transcripción. Luego, las ediciones seleccionadas se realizan en plantas para detectar fenotípicamente un fenotipo de desarrollo de fruta alterado, particularmente un fenotipo sin semillas o reducida simiente. En algunas formas de realización, se usa un sistema de ensayo *in vivo* para analizar la expresión transitoria de luciferasa en células HEK293T transfectadas con constructores de expresión de proteína de factor de transcripción de caja de MADS y reportero, para cribar sitios diana adecuados para mutar factores de transcripción de caja de MADS clado AGAMOUS, con el fin de eventualmente producir ninguna semilla o de tamaño reducido de semillas en frutos de las plantas de interés. En algunas formas de realización, la presente divulgación enseña métodos para generar construcciones de expresión de luciferasa duales insertando siete repeticiones de la secuencia de unión de proteína MADS (secuencia 7x CArG)

corriente ascendente de un promotor mínimo (promotor de la caja CArG 7x), que se usa para permitir que los factores de transcripción de caja MADS se unan a múltiples cajas de CArG. En otras formas de realización, el sistema de luciferasa dual 7xCArG::LUC permite la expresión de (i) el promotor de la caja 7xCArG que controla la expresión del gen reportero de luciferasa de luciérnaga (LUC) y (ii) un promotor de control (CMV) que regula la expresión de la luciferasa de *Renilla* (REN) reportero genético. En otras formas de realización, el cofactor de clado AG SEPALLATA3 se expresa en la misma célula que el sistema dual de luciferasa 7xCArG::LUC. La proteína SEPALLATA es un cofactor requerido para que el factor de transcripción de clado AG forme un complejo para unir la caja 7xCArG. El gen reportero LUC se transcribe y se expresa cuando un complejo de factor de transcripción de caja MADS apunta y se une a la región promotora de caja 7xCArG. El nivel de expresión del gen reportero LUC depende de la activación transcripcional por el complejo de factor de transcripción de caja MADS de interés. En consecuencia, la efectividad relativa de las diferentes mutaciones del factor de transcripción del clado AG en diversos sitios (p. ej., dominio N-terminal, M, I, K o C, o tres motivos descritos en el Ejemplo 2) se determina al cambiar la expresión de una secuencia del gen reportero LUC controlada por el promotor de la caja 7xCArG en comparación con la expresión constitutiva del gen reportero REN.

En algunas formas de realización, la construcción TF clado AG no mutada se cotransfecta con las construcciones de luciferasa duales fusionadas en caja 7x CArG. En algunas formas de realización, la construcción TF clado AG mutada y la construcción TF clado AG no mutada se cotransfectan con las construcciones de luciferasa duales fusionadas en caja 7xCArG. En formas de realización adicionales, la relación LUC:REN con la cotransfección de la versión mutada y la versión no mutada de los TF clado AG se compara con la relación LUC:REN con la cotransfección de la versión no mutada. Si la relación LUC:REN en la versión mutada y la versión no mutada es menor que la versión no mutada sola de los TF de clado AG de interés, el sitio de mutación será un objetivo de interés para el estudio *in vivo* al crear una línea de planta estable con una mutación específica introducida por una técnica de edición de base única estable pero dirigida que se enseña en la presente descripción.

La presente descripción enseña que hay una reducción en la transactivación cuando la versión mutada del TF clado AG se cotransfecta con el sistema de ensayo dual de luciferasa, porque la función de los TF clado AG mutados se verá afectada negativamente por las mutaciones. Por lo tanto, la transactivación se reducirá en la versión mutada del TF clado AG en comparación con la versión no mutada. En algunas formas de realización, las mutaciones que inducen la reducción de la relación

LUC:REN en comparación con la versión de tipo salvaje no mutada se seleccionan para modificar genéticamente las células vegetales de interés.

Por lo tanto, la metodología anteriormente mencionada proporciona un rápido cribado *in vivo* HTP en un organismo modelo, para mutaciones candidatas en factores de transcripción de caja MADS clado AGAMOUS. Si hay una reducción en la activación mediada por el factor de transcripción evidente en el cribado, es decir una indicación de que una mutación podría servir como un alelo negativo dominante, un alelo semidominante, un alelo de pérdida de función débil o una mutación hipomórfica, entonces dicha mutación se direcciona en una planta de interés, que finalmente se fenotipifica para confirmar el "acierto" positivo del cribado e identificar una mutación de interés asociada con el desarrollo alterado de la fruta.

Células vegetales modificadas.

En algunas formas de realización, la presente divulgación proporciona células vegetales modificadas. Aquí, el término "células vegetales modificadas" abarca células vegetales que comprenden una o más modificaciones genómicas que dan como resultado la función alterada de uno o más genes diana endógenos, así como células vegetales que comprenden un sistema de edición de bases dirigido capaz de modificar una base de ADN dentro de una secuencia de ácido nucleico de uno o más genes diana endógenos y alterar la función de una proteína codificada por uno o más genes diana editados en base. En el presente documento, una "célula vegetal no modificada" o "célula vegetal de control" se refiere a una célula o población de células derivadas de una planta en la que los genomas no se han modificado y que no comprende un sistema de edición de genes o comprende un sistema de edición de genes de control (p. ej., un control de vector vacío, un ARNg no dirigido, un ARNsi de mezcla aleatoria, etc.). En algunas formas de realización, una célula vegetal modificada es una célula vegetal editada genéticamente dirigida por las técnicas de edición de la presente descripción.

La presente descripción proporciona células vegetales, que en algunos aspectos se refieren a células derivadas de plantas de angiospermas, que incluyen, sin limitación, la familia de plantas *Rosaceae*. La familia de plantas *Rosaceae* comprende una planta fructífera como: moras, frambuesas negras, frambuesas, frutillas, cerezas, duraznos, ciruelas, damascos, manzanas, peras, membrillos, dátiles, nísperos y almendras. En algunas formas de realización, la presente descripción proporciona células vegetales de frambuesa negra, mora, frambuesa, cereza, damasco, manzana, pera, membrillo, dátiles, durazno, palta, frutilla silvestre, níspero y ciruela para la edición de bases dirigida. En algunas

formas de realización, la presente descripción proporciona células vegetales de frambuesa negra, mora, frambuesa, cereza, damasco, manzana, pera, membrillo, dátiles, durazno, palta, frutilla silvestre, níspero y ciruela, para técnicas de edición de genes.

En algunas formas de realización, la célula vegetal es una célula de frambuesa negra que se ha aislado de una planta de frambuesa negra o parte de ella (*Rubus occidentalis*). En algunas formas de realización, la célula vegetal es una célula de frutilla silvestre que se ha aislado de una planta de frutilla silvestre o parte de ella (*Fragaria vesca*). En algunas formas de realización, la célula vegetal es una célula de ciruela que se ha aislado de una planta de ciruela o parte de ella (*Prunus domestica*). En algunas formas de realización, la célula vegetal es una célula de durazno que se ha aislado de una planta de durazno o parte de ella (*Prunus persica*). En algunas formas de realización, la célula vegetal es una célula de manzana que se ha aislado de una manzana (*Malus domestica*). En algunas formas de realización, la célula vegetal es una célula de tomate que se ha aislado de una planta de tomate o parte de ella (*Solanum lycopersicum*). En algunas formas de realización, la célula vegetal es una célula de ciruela china que se ha aislado de una planta de ciruela china o parte de ella (*Prunus mume*). En algunas formas de realización, la célula vegetal es una célula de almendra que se ha aislado de una planta de almendra o parte de ella (*Prunus dulcis*). En algunas formas de realización, la célula vegetal es una célula de cereza japonesa que se ha aislado de una planta de cereza japonesa o parte de ella (*Prunus serrulata*). En algunas formas de realización, la célula vegetal es una célula de cereza que se ha aislado de una planta de cereza o parte de ella (incluidas muchas especies del género *Prunus*; por ejemplo, *Prinis*). En algunas formas de realización, la célula vegetal es una célula de cereza o célula de cereza dulce que se ha aislado de una planta de cereza o de cereza dulce, o parte de ella (*Prunus avium*). En algunas formas de realización, la célula vegetal es una célula de cereza agria que se ha aislado de una planta de cereza agria o parte de ella (*Prunus cerasus*). En algunas formas de realización, la célula vegetal es una célula de frambuesa roja que se ha aislado de una planta de frambuesa roja o parte de ella (*Rubus idaeus*). En algunas formas de realización, la célula vegetal es una célula de frambuesa silvestre que se ha aislado de una planta de frambuesa silvestre o parte de ella (*Rubus ulmifolius*). En algunas formas de realización, la célula vegetal es una célula de mora que se ha aislado de una planta de mora o parte de ella (incluidas muchas especies en el género *Rubus*, por ejemplo, especies *Rubus*). En algunas formas de realización, la célula vegetal es una célula de palta que se ha aislado de una planta de palta o parte de ella (*Persea americana*). En algunas formas de realización, la célula vegetal es una célula de manzana que se ha aislado de una manzana (*Malus pumila*). En algunas formas de realización, la célula vegetal es una célula de uva que se ha aislado de

una planta de uva o parte de ella (*Vitis vinifera*). En algunas formas de realización, la célula vegetal es una célula de tomate que se ha aislado de un tomate (*Solanum lycopersicum*). En algunas formas de realización, la célula vegetal es una célula de *Arabidopsis* que se ha aislado de una planta de *Arabidopsis* (*Arabidopsis thaliana*).

En algunas formas de realización, las células vegetales modificadas comprenden una o más modificaciones (por ejemplo, inserciones, deleciones o mutaciones de uno o más ácidos nucleicos) en la secuencia de ADN genómico de un gen diana endógeno, que logran la función alterada del gen endógeno, de manera de inducir así falta de semillas (p. ej., producción de semillas nula o reducida), simiente reducida, grosor reducido del endocarpio del fruto y/o lignificación reducida del endocarpio del fruto. En tales formas de realización, las células vegetales modificadas comprenden un "gen diana endógeno modificado". En algunas formas de realización, las modificaciones en la secuencia de ADN genómico causan una mutación de sentido erróneo y/o una mutación sin sentido, de modo de alterar así la función de la proteína, tal como la mutación negativa dominante, una mutación semidominante, una mutación débil por pérdida de función o una mutación hipomórfica. En algunas formas de realización, las modificaciones en la secuencia de ADN genómico dan como resultado sustituciones de aminoácidos, de manera de alterar así la función normal de la proteína codificada. En algunas formas de realización, las modificaciones en la secuencia de ADN genómico codifican una proteína endógena modificada con función reducida, regulada en forma ascendente o alterada en comparación con la versión no modificada (es decir, de tipo salvaje) de la proteína endógena (por ejemplo, un mutante negativo dominante, un mutante semidominante, mutante de pérdida de función débil o mutante hipomórfico).

En algunas formas de realización, las células vegetales modificadas descritas en el presente documento comprenden uno o más genes diana endógenos modificados, en donde la una o más modificaciones dan como resultado una función alterada de un producto génico (es decir, una proteína) codificado por el gen diana endógeno, en comparación con una célula vegetal no modificada. Por ejemplo, en algunas formas de realización, una célula vegetal modificada demuestra una expresión regulada en forma descendente de una proteína, o una expresión regulada en forma ascendente de una proteína. En algunas formas de realización, la expresión del producto génico en una célula vegetal modificada se reduce por lo menos en 5% en comparación con la expresión del producto génico en una célula vegetal no modificada. En otras formas de realización, la expresión del producto génico en una célula vegetal modificada se reduce en por lo menos 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% o más en comparación a la expresión del producto génico en una célula

vegetal no modificada. En algunas formas de realización, las células vegetales modificadas descritas en el presente documento demuestran una expresión y/o función reducida de productos génicos codificados por una pluralidad (por ejemplo, dos o más) de genes diana endógenos, en comparación con la expresión de los productos génicos en una célula vegetal no modificada. Por ejemplo, en algunas formas de realización, una célula vegetal modificada demuestra una expresión y/o función reducida de productos génicos de 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos en comparación con la expresión de los productos génicos en una célula vegetal no modificada.

En otras formas de realización, la expresión del producto génico en una célula vegetal modificada está regulada en forma ascendente en por lo menos 5% en comparación con la expresión del producto génico en una célula vegetal no modificada. En otras formas de realización, la expresión del producto génico en una célula vegetal modificada se regula en forma ascendente en por lo menos 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% o más en comparación con la expresión del producto génico en una célula vegetal no modificada. En algunas formas de realización, las células vegetales modificadas descritas en el presente documento demuestran la función y/o expresión regulada en forma ascendente de productos génicos codificados por una pluralidad (por ejemplo, dos o más) de genes diana endógenos en comparación con la expresión de los productos génicos en una célula vegetal no modificada. Por ejemplo, en algunas formas de realización, una célula vegetal modificada demuestra la función y/o expresión regulada en forma ascendente de productos génicos de 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos en comparación con la expresión de los productos genéticos en una célula vegetal no modificada.

En algunas formas de realización, la presente descripción proporciona una célula vegetal modificada en la que un nucleótido único de uno o más genes diana endógenos, o una porción de los mismos, está sustituido (por ejemplo, "C>T (sustitución de C a T)" o "G>A (Sustitución de G a A)" en la hebra de ADN complementaria), o "A>G (sustitución de A a G)" o "T>1C (sustitución de T a C)" en la hebra de ADN complementaria), de modo que las células vegetales modificadas expresan la proteína en el que un aminoácido está sustituido, lo que puede desencadenar una mutación sin sentido o una mutación de sentido erróneo. En algunas formas de realización, una célula vegetal modificada comprende la sustitución de nucleótidos de una pluralidad de genes diana endógenos, o sus porciones. En algunas formas de realización, una célula vegetal modificada comprende la sustitución de nucleótidos de 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos.

En algunas formas de realización, las células vegetales modificadas descritas en el presente documento comprenden uno o más genes diana endógenos modificados, en donde la una o más

modificaciones a la secuencia de ADN diana dan como resultado la expresión de una proteína con función reducida o alterada (por ejemplo, una "proteína endógena modificada") en comparación con la función de la proteína correspondiente expresada en una célula vegetal no modificada (por ejemplo, una "proteína endógena no modificada"). En algunas formas de realización, las células vegetales modificadas descritas en el presente documento comprenden 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos modificados que codifican 2, 3, 4, 5, 6, 7, 8, 9, 10 o más proteínas endógenas modificadas. En algunas formas de realización, la proteína endógena modificada demuestra una afinidad de unión reducida o alterada para otra proteína expresada por la célula vegetal modificada o expresada por otra célula; capacidad de señalización reducida o alterada; actividad enzimática reducida o alterada; actividad de unión a ADN reducida o alterada; o capacidad reducida o alterada para funcionar como una proteína de andamiaje.

En algunas formas de realización, el gen objetivo endógeno modificado comprende una o más mutaciones negativas dominantes. Como se usa en el presente documento, una "mutación negativa dominante" se refiere a una sustitución, delección o inserción de uno o más nucleótidos de un gen diana de manera que la proteína codificada actúa de forma antagónica a la proteína codificada por el gen diana no modificado. La mutación es dominante negativa porque el alelo produce un fenotipo a pesar de la presencia del gen correspondiente no modificado. Un gen que comprende una o más mutaciones negativas dominantes y la proteína codificada de ese modo se denominan "mutantes negativos dominantes", p. ej., genes negativos dominantes y proteínas negativas dominantes. En algunas formas de realización, la proteína mutante negativa dominante está codificada en una célula vegetal modificada en la que uno o más nucleótidos de genes diana endógenos están editados por un sistema de edición de genes de la presente descripción.

Se conocen diversos mecanismos para la negatividad dominante. Típicamente, el producto génico de un mutante negativo dominante retiene algunas funciones del producto génico no modificado pero carece de una o más funciones cruciales adicionales del producto génico no modificado. Esto hace que el mutante negativo dominante antagonice el producto génico no modificado. Por ejemplo, como una forma de realización ilustrativa, un mutante negativo dominante de un factor de transcripción puede carecer de un dominio de activación funcional, pero puede retener un dominio de unión a ADN funcional. En este ejemplo, el factor de transcripción negativo dominante no puede activar la transcripción del ADN como lo hace el factor de transcripción no modificado, pero el factor de transcripción negativo dominante puede inhibir indirectamente la expresión génica al evitar que el factor de transcripción no modificado se una al sitio de unión del factor de

transcripción. Como otra forma de realización ilustrativa, se presentan mutaciones dominantes negativas de proteínas que funcionan como dímeros o multímeros. Por ejemplo, los mutantes negativos dominantes de tales proteínas diméricas pueden retener la capacidad de dimerizarse con proteínas no modificadas, pero no pueden funcionar de otra manera. Los monómeros negativos dominantes, al dimerizar con monómeros no modificados para formar heterodímeros, evitan la formación de homodímeros funcionales de los monómeros no modificados. Las mutaciones negativas dominantes del gen AG de *Arabidopsis*, el gen 1 estéril de cáscara de hoja de arroz y el gen LMADS1 de lirio se conocen en la técnica (véase, por ejemplo, Mizukami et al., 1996, Jeon et al, 2000, y Tzeng y Yang, 2001). Además de la mutación negativa dominante, también se puede generar una planta de esta invención que comprende además o en lugar de una mutación negativa dominante, una o más mutaciones semidominantes, mutaciones de pérdida de función débil o mutaciones hipomórficas.

En algunas formas de realización, las células vegetales modificadas comprenden un sistema de edición de bases dirigido capaz de modificar una base de ADN dentro de una secuencia de ácido nucleico de uno o más genes diana endógenos. El sistema de edición de bases dirigido puede alterar la expresión y/o la función de las modificaciones de los genes diana endógenos mediante una diversidad de mecanismos, incluida la modificación de la secuencia de ADN genómico del gen diana endógeno (por ejemplo, mediante la inserción, delección o mutación de uno o más ácidos nucleicos en la secuencia de ADN genómico). En alguna forma de realización, la presente descripción enseña una modificación de la secuencia de ADN del gen diana endógeno mediante la sustitución de nucleótidos de uno o más ácidos nucleicos en la secuencia de ADN genómico.

La presente descripción enseña que las células vegetales modificadas descritas en el presente documento comprenden un sistema de edición de bases dirigido (por ejemplo, una combinación de sistema regulador de genes sobre la base de proteína/ácido nucleico). En algunas formas de realización, el sistema de edición de bases dirigido comprendido en la célula vegetal modificada es capaz de modificar uno o más genes diana endógenos. En algunas formas de realización, las células vegetales modificadas descritas en el presente documento comprenden un sistema de edición de bases dirigido que comprende: (1) uno o más ARN guía (ARNg) capaces de unirse a una secuencia de ADN objetivo en un gen endógeno; (2) uno o más polinucleótidos que codifican uno o más ARNg capaces de unirse a una secuencia de ADN diana en un gen endógeno; (3) uno o más polipéptidos modificadores dirigidos a sitio capaces de interactuar con un ARNg y modificar una secuencia de ADN diana en un gen endógeno; (4) uno o más polinucleótidos que codifican un polipéptido modificador dirigido a sitio capaz de interactuar con un ARNg y modificar una secuencia de ADN

diana en un gen endógeno; (5) uno o más polinucleótidos que codifican un polipéptido de edición de bases capaz de sustituir C con T (o G con A); y/o (6) uno o más polinucleótidos que codifican un polipéptido de edición de bases capaz de sustituir A con G (o T con C).

En algunas formas de realización, uno o más polinucleótidos que codifican el sistema de edición de bases dirigido se insertan en el genoma de la célula vegetal. En algunas formas de realización, uno o más polinucleótidos que codifican el sistema de edición de bases dirigido se expresan episomalmente y no se insertan en el genoma de la célula vegetal.

Regulación de genes diana endógenos.

En algunas formas de realización, las células vegetales modificadas descritas en el presente documento demuestran una expresión o función alterada de uno o más genes diana endógenos. En algunas formas de realización, el uno o más genes diana endógenos están relacionados o asociados con la determinación, diferenciación o lignificación del endocarpio. En algunas formas de realización, el uno o más genes diana endógenos son genes de caja MADS. En algunas formas de realización, el gen objetivo es un factor de transcripción clado AGAMOUS (AG). En otras formas de realización, el gen diana es un gen de la caja MADS seleccionado de un clado AGAMOUS enseñado en la presente descripción. En formas de realización adicionales, el gen objetivo es AG, SHP1, SHP2 o STK en plantas de la presente descripción. En algunas formas de realización, el gen diana es un factor de transcripción de caja MADS clado AGAMOUS.

En algunas formas de realización, la expresión de un gen diana endógeno en una vía particular se reduce en las células vegetales modificadas. En algunas formas de realización, la expresión de una pluralidad (por ejemplo, dos o más) de genes diana endógenos en una vía particular se reduce en las células vegetales modificadas. Por ejemplo, la expresión de 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos en una vía particular puede reducirse. En algunas formas de realización, la expresión de un gen diana endógeno en una vía y la expresión de genes diana endógenos en otra vía se reducen en las células vegetales modificadas. En algunas formas de realización, la expresión de una pluralidad de genes diana endógenos en una vía y la expresión de una pluralidad de genes diana endógenos en otra vía se reducen en las células vegetales modificadas. Por ejemplo, la expresión de 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos en una vía puede reducirse, y la expresión de 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos en otra vía particular puede reducirse. En algunas formas de realización, se reduce la expresión de una pluralidad de genes diana endógenos en una pluralidad de vías. Por ejemplo, la expresión de un gen endógeno de cada una de una pluralidad de vías (por

ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más vías) puede reducirse. En aspectos adicionales, la expresión de una pluralidad de genes endógenos (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes) de cada una de una pluralidad de vías (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más vías) puede reducirse.

En algunas formas de realización, la función de una proteína codificada por un gen diana endógeno en una vía particular se altera en las células vegetales modificadas. En algunas formas de realización, las funciones de las proteínas codificadas por una pluralidad (por ejemplo, dos o más) de genes diana endógenos en una vía particular se alteran en las células vegetales modificadas. Por ejemplo, la función de las proteínas codificadas por 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos en una vía particular puede alterarse. En algunas formas de realización, la función de una proteína codificada por un gen objetivo endógeno en una vía y la función de los genes objetivo endógenos en otra vía se alteran en las células vegetales modificadas. En algunas formas de realización, las funciones de las proteínas codificadas por una pluralidad de genes diana endógenos en una vía y la función de las proteínas codificadas por una pluralidad de genes diana endógenos en otra vía están alteradas en las células vegetales modificadas. Por ejemplo, la función de las proteínas codificadas por 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos en una vía puede alterarse, y la función de las proteínas codificadas por 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes diana endógenos en otra vía particular puede ser alterada. En algunas formas de realización, las funciones de las proteínas codificadas por una pluralidad de genes diana endógenos en una pluralidad de vías están alteradas. Por ejemplo, la función de una proteína codificada por un gen endógeno de cada una de una pluralidad de vías (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más vías) puede verse alterada. En aspectos adicionales, la función de las proteínas codificadas por una pluralidad de genes endógenos (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes) de cada una de una pluralidad de vías (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más vías) puede alterarse.

En algunas formas de realización, las células vegetales modificadas descritas en el presente documento comprenden un gen mutado en un clado AGAMOUS. En algunas formas de realización, las células vegetales modificadas descritas en el presente documento comprenden la función alterada de un factor de transcripción codificado por el gen de caja MADS de un clado AGAMOUS que comprende AG, SHP1, SHP2 y STK en plantas de la presente descripción. En algunas formas de realización, las células vegetales modificadas descritas en el presente documento comprenden un factor de transcripción clado AGAMOUS (AG) mutado. En otras formas de realización, las células vegetales modificadas descritas en el presente documento comprenden un gen de caja MADS mutado de un clado AGAMOUS como se describe en la presente descripción. En formas de

realización adicionales, las células vegetales modificadas descritas en el presente documento comprenden un AG, SHP1, SHP2 y/o STK mutado en plantas de la presente descripción.

Función molecular de los genes de la caja MADS relacionados con la formación de semillas.

La familia de factores de transcripción de caja MADS está presente en todos los genomas eucariotas analizados hasta ahora, aunque con un mayor número de miembros de gen en genomas de plantas que en otros reinos. Esta familia de factores de transcripción se define por la presencia de un dominio conservado, la caja MADS, en la región N-terminal, implicada en la unión del ADN y la dimerización con otras proteínas de la caja MADS. La duplicación de genes de la caja MADS ancestral anterior a la divergencia de plantas y animales separó los dos linajes principales, tipo I y tipo II (Alvarez-Buylla et al., 2000 y Bodt et al., 2003). Los genes del grupo tipo II incluyen genes de animales y levaduras de tipo MEF2 y genes de tipo MIKC que solo se encuentran en plantas. Los genes de tipo MIKC recibieron este nombre porque, además del dominio MADS (M), contienen tres dominios conservados adicionales: el dominio intermedio débilmente conservado (I), el dominio conservado de tipo queratina (K) y el dominio C-terminal altamente variable (C), donde este último generalmente contiene motivos conservados de secuencia específica de subfamilia (Theissen et al, 1996 y Kaufmann et al., 2005). El dominio I es responsable de la especificidad en la formación de dímeros de unión al ADN, el dominio K media en la dimerización, y el dominio C funciona en la activación transcripcional y en la formación de complejos de proteínas de orden superior. Por otro lado, los genes del grupo Tipo I muestran una estructura genética más simple. Son más cortos, generalmente codifican un solo exón y carecen del dominio K. Los factores de transcripción MADS en plantas codifican reguladores clave del desarrollo del desarrollo vegetativo y reproductivo.

La mayoría de las proteínas MADS de planta comparten una estructura estereotípica de MIKC. Esta comprende (de N a C-terminal) un dominio N-terminal, que, sin embargo, está presente solo en una minoría de proteínas; un dominio MADS (PDOC00302, IPR002100), que es el principal determinante de la unión al ADN pero que también realiza funciones de dimerización y de unión del factor accesorio; un dominio intermedio (I) débilmente conservado, que constituye un determinante molecular clave para la formación selectiva de dímeros de unión a ADN; un dominio de tipo queratina (K), que promueve la dimerización de proteínas; y un dominio C-terminal (C), que está implicado en la activación transcripcional o en la formación de complejos proteicos ternarios o cuaternarios.

Los genes de tipo MIKC se identificaron inicialmente como genes de identidad de órganos florales en *Antirrhinum majus* y *Arabidopsis thaliana*. Otros análisis genéticos y moleculares agruparon sus funciones biológicas en la organogénesis de las flores en cinco clases: A, B, C, D y E, que se requieren, en diferentes combinaciones, para especificar la identidad de los sépalos (A + E), pétalos (A + B + E), estambres (B + C + E), carpelos (C + E) y óvulos (D + E). En *Arabidopsis*, los genes que pertenecen a estas clases funcionales fueron APETALA1 (AP1) en la clase A, PISTILATA (PI) y APETALA3 (AP3) en la clase B, AGAMOUS (AG) en la clase C, SEEDSTICK / AGAMOUS–LIKE 11 (STK/AGL11) y SHATTERPROOF (SHP) en los genes de clase D y SEPALLATA (SEP1, SEP2, SEP3, SEP4) en la clase E (Bowman et al, 1991, Colombo et al., 1995, y Theissen et al, 2001). Los genes MIKC en las subfamilias AG y APETALA1/FRUITFULL (AP1/FUL) también participan en el desarrollo de frutos y semillas (Immink et al., 2003, Rijpkema et al., 2006, y Smaczniak et al., 2012).

SEEDSTICK (STK, también conocido como AGAMOUS–LIKE 11) – AT4G09960.3, un factor de transcripción de la caja MADS, es conocido por controlar la identidad del óvulo y el desarrollo de semillas en *Arabidopsis*. STK forma un clado estrechamente relacionado con otros tres factores de transcripción en *Arabidopsis*: AGAMOUS (AG) – AT4G18960.1, SHATTERPROOF1 (SHP1) – AT3G58780.4 y SHATTERPROOF2 (SHP2) – AT2G42830.2. Los investigadores han revelado que AG controla la identidad del órgano y la determinación del meristemo floral. AG especifica la identidad del carpelo y la antera dentro de la flor en desarrollo. La proteína AG es un factor de transcripción de la caja MADS. Además, otro factor de transcripción de la caja MADS, STK, en el mismo clado actúa de forma redundante con otros dos genes de caja MADS, SHP1 y SHP2, en el control de la identidad del óvulo. En el tomate, los homólogos de SHP controlan la expansión de la fruta carnosa, mientras que en los duraznos, que son drupas, los homólogos de ambos genes han sido implicados en la formación de fosas divididas lignificadas. AG, SHP1, SHP2 y STK están involucrados con la identidad del órgano, como la determinación de óvulos y meristemas florales para el desarrollo de semillas, así como también, están asociados con la determinación, diferenciación o lignificación del endocarpio. En algunas formas de realización, los genes AG, SHP1, SHP2 y STK y sus ortólogos en plantas son genes de la caja MADS objetivo para la edición de genes, por ejemplo, edición de bases precisa. En otras formas de realización, los genes AG, SHP1, SHP2 y STK y sus ortólogos en plantas de la presente divulgación comparten una estructura MIKC estereotípica, que proporciona similitud funcional implicada con la formación de semillas y/o endocarpios y la lignificación del endocarpio del fruto.

La expresión ectópica de ADNc de AG con estructura de MIKC intacta que incluye regiones C–terminales produce la transformación de los pétalos en estambres, que también se conoce como un fenotipo "apetala". Sin embargo, la eliminación de la región C–terminal de la proteína AG, que es AG con solo dominios MIK (AG–MIK), no permite que las plantas transgénicas AG–MIK produzcan un fenotipo "apetala", lo que indica que esta región es necesaria para la función AG en la especificación de las identidades de estambre y carpelo en las espirales florales exteriores. Además, las plantas transgénicas que llevan una construcción para expresar constitutivamente AG–MIK (por construcción 35S:AG–MIK) producen flores similares a las flores de mutantes *agamous* (*ag*), con indeterminación y conversión de órganos en las espirales internas, lo que sugiere que la proteína AG que carece de su región C–terminal inhibe la función de la proteína AG endógena y funciona como una mutación negativa dominante de la proteína AG endógena. Por lo tanto, el fracaso de la sobreexpresión de AG–MIK para convertir los órganos de espiral externos en órganos reproductivos, así como la inhibición de AG–MIK en funciones normales de AG indican que la región C–terminal de la proteína AG puede tener un rol esencial para identidad del órgano y determinación del meristemo floral. La presente divulgación enseña que se espera que la proteína STK que comparte una estructura MIKC similar con la proteína AG tenga la misma función que la AG.

Como se describió anteriormente, la eliminación del dominio C–terminal en plantas transgénicas 35S::AG–MIK muestra el mutante *ag* de pérdida de función más fuerte. El extremo C–terminal se puede eliminar por lo menos de dos maneras como se enseña en este documento. Se puede introducir un codón de parada a través de la edición de bases. Alternativamente, se podría cortar el extremo C–terminal a través de una nucleasa dirigida CRISPR. Por lo tanto, el extremo C–terminal se puede deletarse por lo menos mediante la metodología de edición de bases o enzimas de corte.

Como se mencionó anteriormente, la presente divulgación enseña la delección del dominio C–terminal de la proteína STK en plantas usando un sistema de edición de base dirigido con complejo CRISPR/Cas fusionado con el dominio de desaminasa base, lo que produce un alelo STK negativo dominante fuerte, o alternativamente, una CRISPR/Cas endonucleasa. En algunas formas de realización, se puede introducir una delección C–terminal de la invención mediante la nucleasa dirigida por CRISPR que corta el polinucleótido STK. En algunas formas de realización, los alelos mutantes con fuerte pérdida de función *stk* dan como resultado la infertilidad. En algunas formas de realización, las células vegetales de la presente descripción en las que el gen STK objetivo se edita en bases usando el sistema de edición de bases dirigido descrito en el presente documento inducen la ausencia de semillas al reducir la función STK (por ejemplo, mutación de sentido erróneo por sustitución de

aminoácidos). En algunas formas de realización, las células vegetales de la presente descripción en las que el gen STK objetivo se edita en bases utilizando el sistema de edición de bases dirigido descrito en el presente documento inducen la ausencia de semillas al abolir la función STK (es decir, mutación de sentido erróneo por sustitución de aminoácidos).

Una mutación en STK es el gen causante del locus INHIBIDOR DE DESARROLLO DE SEMILLAS (SDI) sin semillas en uvas comerciales (Royo et al, 2018, incorporado aquí como referencia). Esta mutación natural se asocia con un rasgo estenospermocárpico, en el que se produce la fertilización pero las semillas no se desarrollan correctamente y son esencialmente imperceptibles. La mutación stk^{sdi} en la uva es una mutación de sentido erróneo que implica una conversión de G> T en la posición. Esto conduce a una sustitución no conservadora de aminoácidos de arginina a leucina en la 'región C-terminal' de STK. La mutación stk^{sdi} causa un efecto negativo dominante ya que los heterocigotos tienen un tamaño de semilla reducido. En contraste con el fuerte efecto fenotípico negativo dominante de 35S::AG-MIK en *Arabidopsis*, la mutación stk^{sdi} en uva tiene un efecto negativo hipomórfico dominante. En algunas formas de realización, la mutación de residuos específicos del dominio 'C-terminal' usando un sistema de edición de bases preciso puede hacer alelos negativos dominantes hipomórficos en plantas tales como la familia de plantas *Rosaceae*, que comprenden una planta fructífera, por ejemplo, moras, frambuesas negras, frambuesas, frutillas, cerezas, duraznos, ciruelas, damascos, manzanas, peras, membrillos, nísperos, dátiles y almendras.

STK también está involucrada en un evento clave en la domesticación y el mejoramiento de la palma aceitera, así como en la pérdida de la gruesa cáscara de coco que rodea el grano de la palma aceitera (*Elaeis guineensis*). La palma aceitera moderna tiene tres formas de fruta, dura (con cáscara gruesa; DeliDura), pisífera (sin cáscara: AVROS y MPOB) y tenera (con cáscara delgada), un híbrido entre dura y pisífera. La palma pisífera suele ser femenina. La tenera produce mucho más aceite que la dura, y es la base para la producción comercial de aceite de palma en todo el sudeste asiático. Una sustitución de nucleótidos en el 'dominio de la caja MADS' de STK causa dos alelos que conducen al fenotipo pisífera: i) una sustitución de nucleótidos en stk^{MPOB} da como resultado un cambio de aminoácidos de leucina a prolina en el dominio conservado de unión y dimerización del ADN, mientras que ii) una sustitución en stk^{AVROS} da como resultado un cambio de aminoácido lisina a asparagina en el mismo dominio, solo dos aminoácidos carboxi-terminales al stk^{MPOB} . En proteínas relacionadas, este residuo lisina altamente conservado está involucrado en la localización nuclear y la unión directa al ADN, mientras que la sustitución por una prolina de solo dos residuos de aminoácidos amino-terminales a esta posición interrumpiría la hélice alfa que está involucrada en la

dimerización de MADS y unión de ADN En algunas formas de realización, ambas mutaciones de stk^{MPOB} y stk^{AVROS} se encuentran en el "dominio de la caja MADS" de STK que está implicado en la localización nuclear, la dimerización de MADS y la unión al ADN.

La presente divulgación enseña que los alelos mutantes de un gen objetivo pueden inducir la falta de semillas en las plantas, que incluyen, entre otras, frambuesa negra, mora, frambuesa, cereza, palta, frutilla silvestre, manzana, ciruela, uva, tomate, dátiles, níspero y durazno. En algunas formas de realización, el gen objetivo es un gen de caja MADS de un clado AGAMOUS que comprende AG, SHP1, SHP2 y STK. En algunas formas de realización, el gen objetivo es un factor de transcripción clado AGAMOUS (AG). En otras formas de realización, el gen objetivo es un gen de caja MADS de un clado AGAMOUS enseñado en la presente descripción. En formas de realización adicionales, el gen objetivo es AG, SHP1, SHP2 o STK en plantas de la presente descripción.

En algunas formas de realización, los ARNg están diseñados para dirigirse a la NLS (secuencia de localización nuclear) del gen objetivo. La mutación de NLS usando el sistema de edición de bases dirigido de la presente descripción puede evitar que el factor de transcripción sea transportado al núcleo, lo que resulta en falta de semillas, tamaño reducido de semillas y/o endocarpio de fruto reducido en una planta editada genéticamente. En algunas formas de realización, los ARNg están diseñados para dirigirse al dominio de la caja MADS del gen diana. La mutación del dominio de la caja MADS utilizando el sistema de edición de bases dirigido de la presente divulgación puede evitar que el factor de transcripción se una al ADN y/o dimerice, lo que resulta en falta de semilla, reducción de la simiente, tamaño reducido de la semilla y/o reducción del endocarpio de la fruta en una planta con edición de genes.

En algunas formas de realización, los ARNg están diseñados para apuntar a un motivo NN en el dominio de la caja MADS del gen diana. La mutación del motivo NN del dominio de la caja MADS utilizando el sistema de edición de bases dirigido de la presente divulgación puede evitar la unión de ADN y/o la dimerización del factor de transcripción, lo que resulta en falta de semillas, reducción de la simiente, reducción del tamaño de la semilla y/o reducción del endocarpio de la fruta en una planta editada genéticamente. En otras formas de realización, el motivo NN comprende un motivo extendido Y [A/S]NN.

En algunas formas de realización, los ARNg están diseñados para apuntar al dominio K del gen objetivo que está involucrado en la dimerización de proteínas. La mutación del dominio K usando el sistema de edición de bases dirigido de la presente divulgación puede dar lugar a un mal funcionamiento de la transcripción de la caja MADS en la dimerización, lo que resulta en falta de

semilla, tamaño reducido de la semilla y/o reducido endocarpio de la fruta en una planta editada genéticamente.

En algunas formas de realización, los ARNg están diseñados para apuntar a un motivo YQQ (o motivo YQQE) del dominio K del gen objetivo que está implicado en la dimerización de proteínas. La mutación del motivo YQQ del dominio K usando el sistema de edición de bases dirigido de la presente divulgación puede dar lugar a un mal funcionamiento de la transcripción de la caja MADS en la dimerización, lo que da como resultado falta de semillas, tamaño reducido de semillas y/o endocarpio de fruta reducido en una planta editada genéticamente. En otras formas de realización, el motivo YQQ comprende un motivo extendido, YQQE[A/S][A/S/N/K/T]KL[R/H][Q/H/A/N/R]QI (véase, por ejemplo, **SEQ ID NOs.** 178–182).

En algunas formas de realización, los ARNg están diseñados para dirigirse al dominio C-terminal del gen objetivo. La mutación del dominio C-terminal usando el sistema de edición de bases dirigido de la presente divulgación puede afectar negativamente la activación transcripcional o la formación de complejos de proteínas ternarios o cuaternarios, lo que resulta en falta de semillas, reducción de la simiente, tamaño reducido de las semillas y/o reducción del endocarpio de la fruta en una planta editada genéticamente. El dominio C-terminal se puede truncar usando diversos métodos de edición de genes.

En algunas formas de realización, los ARNg están diseñados para dirigirse al R en la posición correspondiente a 225 de la **SEQ ID NO:** 13 o 201 de la **SEQ ID NO:** 97 en el gen ortólogo de plantas de la presente descripción. En algunas formas de realización, la arginina conservada es el residuo 196 con la numeración de posición de aminoácidos de referencia de **SEQ ID NO:** 66; residuo 197 con numeración de posición de aminoácidos de referencia de **SEQ ID NOs:** 53, 78, 150, 166–171; residuo 198 con la numeración de la posición del aminoácido de referencia de **SEQ ID NO:** 29; residuo 199 con numeración de posición de aminoácido de referencia de **SEQ ID NO:** 128; o el residuo 201 con la numeración de la posición del aminoácido de referencia de **SEQ ID NO:** 152, 154, 156, 158, 160, 162, 164. La mutación de la posición R usando el sistema de edición de la base objetivo de la presente divulgación puede afectar negativamente la activación transcripcional o la formación de complejos proteicos ternarios o cuaternarios, lo que resulta en falta de semillas, reducción del tamaño de la semilla y/o reducción del endocarpio de la fruta en una planta editada genéticamente.

En algunas formas de realización, la presente divulgación enseña un sistema de edición de bases dirigido que usa un sistema CRSIPR/CAS y un editor de bases, tal como el editor de bases de citosina (CBE) o el editor de bases de adenina (ABE). En algunas formas de realización, el sistema

de edición de bases dirigido comprende un dominio efector asociado a CRISPR y un dominio de citidina desaminasa. Este sistema de edición de bases con un ARNg tiene como objetivo un cambio de par de bases de nucleótidos C>T (o G> T en la hebra complementaria), lo que conduciría a una sustitución de aminoácidos en una secuencia dirigida.

En otras formas de realización, el sistema de edición de bases dirigido comprende un dominio efector asociado a CRISPR y un dominio de adenosina desaminasa. Este sistema de edición de bases con un ARNg apunta a un cambio de par de bases de nucleótidos A>G (o T> C en la hebra complementaria), lo que conduciría a una sustitución de aminoácidos en una secuencia dirigida.

La presente divulgación enseña un sistema de edición de bases dirigido para lograr mutaciones dominante negativa, semidominante o de pérdida de función débil o mutaciones hipomórficas al dirigirse a aminoácidos conservados en diferentes dominios MIKC del gen objetivo (es decir, STK, SHP1/SHP2 o AG) . En algunas formas de realización, hay secuencias de aminoácidos que están bien conservadas entre los ortólogos de STK entre plantas que incluyen, sin limitación, la familia de plantas *Rosaceae*. Estas plantas pueden comprender una planta fructífera que incluye, entre otras, moras, frambuesas negras, frambuesas, frutillas, cerezas, duraznos, ciruelas, damascos, manzanas, peras, membrillos, dátiles, nísperos y almendras.

En algunas formas de realización, la secuencia de aminoácidos bien conservada que se encuentra en los dominios MIKC de la proteína diana (es decir, STK) está sustituida y/o mutada en las plantas, partes y células de la presente descripción por el sistema de edición de bases dirigido. En otras formas de realización, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más secuencias de aminoácidos bien conservadas que se encuentran en los dominios MIKC de la proteína diana (es decir, STK, SHP1/SHP2 o AG) pueden ser sustituidas y/o mutadas en las plantas, partes y células de la presente divulgación por el sistema de edición de bases dirigido.

Hay otras secuencias de aminoácidos bien conservadas, tales como NN, A, D, YYQ, R, Q, N, R, R, que son conservadas a más del 95% en los factores de transcripción del clado AG (por ejemplo, STK) a través de todo el reino vegetal, como se muestra, por ejemplo, en las Tablas 1, 1B y 3. Dichas sustituciones de aminoácidos pueden generar mutaciones dominante negativa, semidominante, de pérdida de función débil o mutaciones hipomórficas en las células vegetales, partes y plantas modificadas, lo que puede dar lugar a un aumento de la falta de semillas, reducción de la simiente, una reducción en el tamaño de la semilla o una reducción en el tejido de endocarpio en las frutas. La presente divulgación enseña que la introducción de la mutación STK para un cambio específico en el nucleótido en una ubicación objetivo puede alterar la función molecular de STK para obtener el

fenotipo deseado, como falta de semillas, disminución de la simiente, una reducción en el tamaño de la semilla o una reducción en el tejido del endocarpio en frutas. En otras formas de realización, la presente divulgación proporciona un método para generar un alelo específico con un solo cambio de par de bases *in planta*, y realizar una edición de bases dirigida para inducir la falta de semillas en frutos de muchas especies de plantas diferentes.

La presente divulgación enseña que un motivo NN, un motivo YQQ (por ejemplo, un motivo YQQE en la posición 91–94 de SEQ ID NOs: 29, 78, 97, 53, 66, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164, 166–171; o la posición 116–119 de SEQ ID NO: 13), o una posición R (225 de SEQ ID NO: 13; posición 196 de SEQ ID NO: 66; posición 197 de SEQ ID NOs: 53, 78, 150, 166–171; posición 198 de SEQ ID NO: 29; posición 199 de SEQ ID NO: 128; o la posición 201 de SEQ ID NOs: 97, 152, 154, 156, 158, 160, 162, 164) tienen por lo menos 90%, 91%, 92%, 93%, 94%, 95% o más identidad de secuencia en dominios MIKC entre homólogos y/u ortólogos de STK a través del reino vegetal (ver, por ejemplo, FIGS. 2–5). En algunas formas de realización, un motivo NN, un motivo YQQ o una posición R son dirigidos para una o más sustituciones de aminoácidos mediante la edición de bases de nucleótidos que codifican para el aminoácido correspondiente. En algunas formas de realización, un motivo NN es dirigido para una o más sustituciones de aminoácidos en el dominio de la caja MADS de la proteína objetivo (es decir, STK, SHP1/SHP2 o AG). En otras formas de realización, un motivo Y[A/S]NN es dirigido para una o más sustituciones de aminoácidos en el dominio de la caja MADS de la proteína objetivo (es decir, STK, SHP1/SHP2 o AG) (por ejemplo, posición 57 de la SEQ ID NO: 152, 154, 156, 158, 160, 162, 164 o 166–171).

En algunas formas de realización, un motivo YQQ es dirigido para una o más sustituciones de aminoácidos en el dominio K de la proteína objetivo (es decir, STK, SHP1/SHP2 o AG). En otras formas de realización, un motivo YQQE[A/S][A/S/N/K/T]KL[R/H][Q/H/A/N/R]QI es dirigido para una o más sustituciones de aminoácidos en el dominio K de la proteína objetivo (es decir, STK, SHP1/SHP2 o AG).

En algunas formas de realización, una posición R (225 de SEQ ID NO: 13; posición 196 de SEQ ID NO: 66; posición 197 de SEQ ID NOs: 53, 78, 150, 166–171; posición 198 de SEQ ID NO: 29; posición 199 de SEQ ID NO: 128; o la posición 201 de SEQ ID NOs: 97, 152, 154, 156, 158, 160, 162, 164) es dirigida para la sustitución de aminoácidos en el dominio C de la proteína objetivo (es decir, STK, SHP1/SHP2, o AG). En algunas formas de realización, una sustitución de aminoácido puede ser una R>H o R>L. En algunas formas de realización, una posición Q que corresponde a 225 de SEQ ID NO: 13 o 201 de SEQ ID NO: 97 se encuentra en el factor de transcripción SHP de la

uva (VIT_12s0142g00360.0) y de manera similar, puede ser dirigida para la sustitución de aminoácido en el dominio C de la proteína SHP objetivo.

Dichas sustituciones de aminoácidos producen mutaciones negativas dominantes en las células vegetales, partes y plantas modificadas, lo que puede dar lugar a falta de semillas, reducción de la simiente, una reducción en el tamaño de la semilla o una reducción en el tejido del endocarpio en las frutas. La presente divulgación enseña que la introducción de una mutación en una STK para un cambio específico en el nucleótido en una ubicación objetivo puede alterar la función molecular de STK para obtener el fenotipo deseado, tal como falta de semillas, reducción de la simiente, una reducción en el tamaño de la semilla o una reducción en el tejido de endocarpio en frutas. En otras formas de realización, la presente divulgación proporciona un método para generar un alelo específico con un solo cambio de par de bases *in planta* y realizar una edición de bases dirigida para inducir la falta de semillas en frutos de muchas especies de plantas diferentes.

La conservación de secuencias de aminoácidos entre plantas distantemente relacionadas sugiere una presión selectiva para mantener la función. En algunas formas de realización, las mutaciones que causan un tamaño de semilla reducido o inducen falta de semilla serían una fuerte presión selectiva. En algunas formas de realización, tales mutaciones son mutación de sentido erróneo y/o mutación sin sentido causada por el sistema de edición de bases dirigido de la presente descripción. Como los aminoácidos bien conservados están en el dominio C de la proteína diana, la presente descripción proporciona candidatos para la mutación para imitar la función molecular del alelo mutante stk^{sdi} descrito anteriormente. En algunas formas de realización, una mutación introducida por el sistema de edición de bases dirigido en plantas de la presente descripción proporciona fenotipos similares al alelo stk^{sdi} en el dominio C-terminal del gen STK de uva. Este tipo de mutación puede no eliminar completamente la función STK y no producir infertilidad.

Sistemas de edición de genes.

La edición de genes generalmente se refiere al proceso de modificación de la secuencia de nucleótidos de un genoma, preferentemente de manera precisa o predeterminada. Los ejemplos de métodos de edición de genes descritos en el presente documento incluyen métodos de uso de nucleasas dirigidas al sitio para cortar el ácido desoxirribonucleico (ADN) en ubicaciones objetivo precisas en el genoma, creando así roturas de ADN monocatenarias o bicatenarias en ubicaciones particulares dentro del genoma. Tales roturas pueden repararse y se reparan regularmente mediante procesos celulares endógenos naturales, como la reparación dirigida por homología (HDR) y la unión

de extremo no homólogo (NHEJ, conforme a sus siglas en inglés). Estos dos procesos principales de reparación del ADN consisten en una familia de vías alternativas. NHEJ une directamente los extremos del ADN resultantes de una ruptura bicatenaria, a veces, con la pérdida o la adición de la secuencia de nucleótidos, que puede interrumpir o mejorar la expresión génica. HDR utiliza una secuencia homóloga, o secuencia donante, como patrón para insertar una secuencia de ADN definida en el punto de ruptura. La secuencia homóloga puede estar en el genoma endógeno, tal como una cromátida hermana. Alternativamente, el donante puede ser un ácido nucleico exógeno, tal como un plásmido, un oligonucleótido monocatenario, un oligonucleótido bicatenario, un oligonucleótido dupla o un virus, que tiene regiones de alta homología con el locus escindido por nucleasa, pero que también puede contener una secuencia adicional o cambios en la secuencia, incluidas las deleciones que se pueden incorporar al locus diana escindido. Un tercer mecanismo de reparación puede ser la unión de extremos mediada por microhomología (MMEJ, conforme a sus siglas en inglés), también conocida como "NHEJ alternativa", en la que el resultado genético es similar a NHEJ, en términos de que pueden producirse pequeñas deleciones e inserciones en el sitio de escisión. MMEJ puede hacer uso de secuencias homólogas de algunos pares de bases que flanquean el sitio de ruptura de ADN para impulsar un resultado de reparación de unión de extremos de ADN más favorecido, y los informes recientes han aclarado aún más el mecanismo molecular de este proceso; ver, por ejemplo, Cho y Greenberg, *Nature* 518, 174–76 (2015); Kent et al., *Nature Structural and Molecular Biology*, Adv. Online doi: 10.1038 / nsmb.2961 (2015); Mateos–Gomez et al., *Nature* 518, 254–57 (2015); Ceccaldi et al., *Nature* 528, 258–62 (2015). En algunos casos, puede ser posible predecir resultados probables de reparación sobre la base del análisis de microhomologías potenciales en el sitio de la ruptura del ADN.

"Recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos, que incluye, sin limitación, captura de donante por unión de extremos no homólogos (NHEJ) y recombinación homóloga. Para los propósitos de esta divulgación, "recombinación homóloga (HR)" se refiere a la forma especializada de dicho intercambio que tiene lugar, por ejemplo, durante la reparación de roturas bicatenarias en las células a través de mecanismos de reparación dirigida por homología (HDR). Este proceso requiere homología de secuencia de nucleótidos, utiliza una molécula "donante" como patrón para reparar una molécula "objetivo" (es decir, la que experimentó la ruptura bicatenaria), y se conoce de diversas maneras como "conversión génica no cruzada" o "conversión génica del tracto corto", porque conduce a la transferencia de información genética del donante al objetivo. Sin desear limitarse a ninguna teoría en particular, dicha

transferencia puede implicar una corrección de no coincidencia del ADN heterodupla que se forma entre el objetivo roto y el donante, y/o el recocido de hebra dependiente de síntesis, en donde el donante se utiliza para volver a sintetizar información genética que formará parte del objetivo y/o procesos relacionados. Dicha HR especializada a menudo da como resultado una alteración de la secuencia de la molécula diana de tal manera que parte o toda la secuencia del polinucleótido donante se incorpora al polinucleótido diana.

Los métodos de edición de genes contemplados en diversas formas de realización comprenden nucleasas diseñadas, diseñadas para unir y escindir una secuencia de ADN diana en un gen de interés, tal como los genes del factor de transcripción clado AG descritos aquí. Las nucleasas diseñadas por ingeniería contempladas en formas de realización particulares pueden usarse para introducir una ruptura de doble hebra en una secuencia de polinucleótidos diana, que puede repararse (i) mediante unión de extremo no homólogo (NHEJ) en ausencia de un patrón de polinucleótidos, por ejemplo, un patrón de reparación donante, o (ii) por reparación dirigida por homología (HDR), es decir, recombinación homóloga, en presencia de un patrón de reparación donante. Las nucleasas de ingeniería contempladas en ciertas formas de realización también pueden diseñarse como nickasas, que generan roturas de ADN monocatenarias que pueden repararse usando la maquinaria de reparación de escisión de bases (BER, conforme a sus siglas en inglés) de la célula o recombinación homóloga en presencia de un patrón de reparación donante.

La edición de genes mediante nucleasas específicas de secuencia se conoce en la técnica. Ver referencias (1) Carroll, D. (2011) Genome engineering with zinc–finger nucleases. *Genetics*, 188, 773–82; (2) Wood, A. J. et al. (2011) Targeted gene editing across species using ZFNs and TALENs. *Science* (New York, N.Y.), 333, 307; (3) Perez–Pinera, P. et al. (2012) Advances in targeted gene editing. *Current opinion in chemical biology*, 16, 268–77, cada una de las cuales se incorpora aquí como referencia en su totalidad.

Una ruptura de ADN de doble hebra mediada por nucleasas (ADNds) en el genoma puede repararse mediante dos mecanismos principales: unión de extremo no homólogo (NHEJ), que con frecuencia da como resultado la introducción de inserciones y deleciones no específicas (indels), o la reparación dirigida por homología (HDR), que incorpora una hebra homóloga como patrón de reparación. Ver Symington, L. S. y Gautier, J. (2011) Double–strand break end resection and repair pathway choice. *Annual review of genetics*, 45, 247–71, que se incorpora por referencia en su totalidad.

Cuando se suministra una nucleasa específica de secuencia junto con una construcción de ADN donante homólogo que contiene las mutaciones deseadas, las eficiencias de direccionamiento de genes aumentan en 1000 veces en comparación con solo la construcción del donante. Ver Urnov et al. (2005) Highly efficient endogenous human gene correction using designed zinc–finger nucleases. *Nature*, 435, 646–51, que se incorpora por referencia en su totalidad.

En algunas formas de realización, las técnicas de edición de genes de la presente divulgación se usan para plantas que se modifican usando cualquier herramienta de edición de genes, incluidas, entre otras: tecnologías ZFN, TALENS, CRISPR y Mega nucleasa. En algunas formas de realización, las herramientas de edición de genes de la presente descripción comprenden proteínas o polinucleótidos que se han diseñado a medida para apuntar y cortar secuencias de ácido desoxirribonucleico (ADN) específicas. En algunas formas de realización, las proteínas de edición de genes son capaces de reconocer y unirse directamente a secuencias de ADN seleccionadas. En otras formas de realización, las herramientas de edición de genes de la presente divulgación forman complejos, en los que los componentes de nucleasa se basan en moléculas de ácido nucleico para unirse y reclutar el complejo a la secuencia de ADN diana.

En algunas formas de realización, las herramientas de edición de genes de un solo componente comprenden un dominio de unión capaz de reconocer secuencias de ADN específicas en el genoma de la planta y una nucleasa que corta el ADN bicatenario. El fundamento de la tecnología de edición de genes que se enseña en la presente divulgación es el uso de una herramienta que permite la introducción de mutaciones específicas del sitio en el genoma de la planta o la integración de genes específica de sitio.

Hay muchos métodos disponibles para suministrar genes en las células vegetales, por ejemplo, transfección, electroporación, vectores virales y transferencia mediada por *Agrobacterium*. Los genes pueden expresarse transitoriamente a partir de un vector plasmídico. Una vez expresados, los genes generan la mutación dirigida que se heredará de manera estable, incluso después de la degradación del plásmido que contiene el gen.

Las nucleasas a medida se pueden usar para realizar rupturas de doble hebra (DSB) dirigidas en células vivas, cuya reparación se puede explotar para inducir los cambios de secuencia deseados. Dos vías competidoras efectúan reparaciones en la mayoría de las células, incluidas las células vegetales. La reparación de una DSB inducida por nucleasas, mediante la unión de extremo no homólogo (NHEJ) conduce a la introducción de mutaciones de inserción/delección (indeles) con altas frecuencias. Por el contrario, la reparación de DSB por reparación dirigida por homología (HDR) con

un ADN de "patrón donante" suministrado por el usuario puede conducir a la introducción de alteraciones específicas (por ejemplo, mutaciones puntuales e inserciones) o la corrección de secuencias mutantes nuevamente al tipo salvaje.

Nucleasas.

En algunas formas de realización, se genera una célula vegetal de interés mediante la edición de genes realizada con nucleasas diseñadas que se dirigen a uno o más loci que contribuyen al factor de transcripción de la caja MADS. Sin desear limitarse a ninguna teoría en particular, se contempla que las nucleasas diseñadas estén diseñadas para alterar con precisión uno o más factores de transcripción de caja MADS a través de la edición de genes y, una vez que se validan la actividad y la especificidad de nucleasas, conducen a una interrupción predecible de la expresión y/o función del factor de transcripción de caja MADS, de modo de ofrecer un desarrollo alterado del fruto, tal como un fenotipo sin semillas, de reducida simiente, y/o un endocarpio reducido en los frutos de las plantas.

Las nucleasas diseñadas genéticamente descritas en el presente documento generan cortes de ADN de hebra única o roturas de ADN de hebra doble (DSB) en una secuencia diana. Además, se puede lograr una DSB en el ADN objetivo mediante el uso de dos nucleasas que generan mellas monocatenarias (nickasas). Cada nickasa escinde una hebra del ADN, y el uso de dos o más nickasas puede crear una ruptura de doble hebra (o bicatenaria) (por ejemplo, una ruptura escalonada de doble hebra) en una secuencia de ADN objetivo. En otras formas de realización, las nucleasas se usan en combinación con un patrón de reparación de donante, que se introduce en la secuencia diana en el sitio de ruptura de ADN mediante recombinación homóloga en una DSB.

Las nucleasas diseñadas descritas en el presente documento que son adecuadas para la edición de genes comprenden uno o más dominios de unión al ADN y uno o más dominios de escisión del ADN (por ejemplo, uno o más dominios de endonucleasa y/o exonucleasa), y opcionalmente, uno o más enlazadores contemplados en el presente documento. Una "nucleasa diseñada por ingeniería genética" se refiere a una nucleasa que comprende uno o más dominios de unión a ADN y uno o más dominios de escisión de ADN, en donde la nucleasa se ha diseñado y/o modificado para unir una secuencia objetivo de unión a ADN adyacente a una secuencia objetivo de escisión de ADN. La nucleasa diseñada por ingeniería puede diseñarse y/o modificarse a partir de una nucleasa natural o de una nucleasa diseñada previamente.

Los ejemplos ilustrativos de nucleasas que se pueden diseñar para unir y escindir una secuencia diana incluyen, sin limitación, endonucleasas orientadoras (*homíng*) (meganucleasas), megaTAL, nucleasas efectoras de tipo activadoras de la transcripción (TALEN), nucleasas de dedos

de zinc (ZFN) y repeticiones palindrómicas cortas agrupadas regularmente intercaladas (CRISPR)/Cas. En algunas formas de realización, las nucleasas contempladas en el presente documento comprenden uno o más dominios de unión y escisión de ADN heterólogos (por ejemplo, ZFN, TALEN, megaTAL) (Boissel et al., 2014; Christian et al., 2010). En otras formas de realización, el dominio de unión de ADN de una nucleasa natural puede alterarse para unirse a un sitio objetivo seleccionado (por ejemplo, una meganucleasa que ha sido diseñada para unirse a un sitio diferente al sitio de unión afín). Por ejemplo, las meganucleasas se han diseñado para unir sitios objetivo diferentes de sus sitios de unión afines (Boissel et al., 2014). En formas de realización particulares, una nucleasa requiere una secuencia de ácido nucleico para dirigir la nucleasa a un sitio objetivo (por ejemplo, CRISPR/Cas).

(i) TALEN.

Las nucleasas efectoras de tipo activador de la transcripción (TALEN) comprenden una nucleasa de escisión de ADN no específica (p. ej., un dominio de escisión de Fok I) fusionada a un dominio de unión de ADN que puede diseñarse fácilmente para que las TALEN puedan dirigirse esencialmente a cualquier secuencia (ver, por ejemplo, Joung y Sander, *Nature Reviews Molecular Cell Biology* 14: 49–55 (2013)). Los métodos para generar TALEN de ingeniería son conocidos en la técnica; ver, por ejemplo, el sistema de alto rendimiento de fase sólida automatizable sobre la base de ligadura rápida (FLASH) descrito en el documento de los Estados Unidos. Acta Nro. 61/610.212, y la referencia de Reyon et al., *Nature Biotechnology* 30, 460–465 (2012); al igual que los métodos descritos en las referencias de Bogdanove & Voytas, *Science* 333, 1843–1846 (2011); Bogdanove et al., *Curr Opin Plant Biol* 13, 394–401 (2010); Scholze & Boch, *J. Curr Opin Microbiol* (2011); Boch et al., *Science* 326, 1509–1512 (2009); Moscou & Bogdanove, *Science* 326, 1501 (2009); Miller et al., *Nat Biotechnol* 29, 143–148 (2011); Morbitzer et al., *T. Proc Natl Acad Sci USA* 107, 21617–21622 (2010); Morbitzer et al., *Nucleic Acids Res* 39, 5790–5799 (2011); Zhang et al., *Nat Biotechnol* 29, 149–153 (2011); Geissler et al., *PLoS ONE* 6, e19509 (2011); Weber et al., *PLoS ONE* 6, e19722 (2011); Christian et al., *Genetics* 186, 757–761 (2010); Li et al., *Nucleic Acids Res* 39, 359–372 (2011); Mahfouz et al., *Proc Natl Acad Sci USA* 108, 2623–2628 (2011); Mussolino et al., *Nucleic Acids Res* (2011); Li et al., *Nucleic Acids Res* 39, 6315–6325 (2011); Cermak et al., *Nucleic Acids Res* 39, e82 (2011); Wood et al., *Science* 333, 307 (2011); Hockemeyer et al. *Nat Biotechnol* 29, 731–734 (2011); Tesson et al., *Nat Biotechnol* 29, 695–696 (2011); Sander et al., *Nat Biotechnol* 29, 697–698 (2011); Huang et al., *Nat Biotechnol* 29, 699–700 (2011); y Zhang et al., *Nat Biotechnol* 29, 149–153 (2011); todos los cuales se incorporan en este documento por referencia en su totalidad.

En algunas formas de realización, una TALEN se une y escinde una región objetivo de un locus que contribuye al factor de transcripción de la caja MADS. Una "TALEN" se refiere a una nucleasa diseñada que comprende un dominio de unión al ADN TALE diseñado contemplado en otra parte del presente documento y un dominio de endonucleasa (o su medio dominio de endonucleasa), y opcionalmente comprende uno o más enlazadores y/o dominios funcionales adicionales, por ejemplo, un dominio enzimático de procesamiento de extremo una enzima de procesamiento de extremo que exhibe actividad de 5–3' exonucleasa, 5–3' exonucleasa alcalina, 3–5'–exonucleasa (p. ej., Trex2), 5' aleta endonucleasa, helicasa o polimerasas de ADN independiente de patrón.

En algunas formas de realización, las plantas de interés se modifican a través de nucleasas efectoras de tipo activador de transcripción (TAL) (TALEN). Las TALEN son polipéptidos con brazos de polipéptidos repetidos capaces de reconocer y unirse a regiones específicas de ácido nucleico. Al diseñar los brazos de polipéptidos para reconocer secuencias diana seleccionadas, las nucleasas TAL se pueden usar para dirigir rupturas de ADN de doble hebra a regiones genómicas específicas. Estas rupturas pueden repararse luego mediante recombinación para editar, deletear, insertar o de otro modo modificar el ADN de un organismo huésped. En algunas formas de realización, las TALEN se usan solas para la edición de genes (por ejemplo, para la delección o la interrupción de un gen). En otras formas de realización, las TAL se usan junto con secuencias de donantes y/u otras proteínas de factor de recombinación que ayudarán en el proceso de unión de extremo no homólogo (NHEJ) para reemplazar la región de ADN objetivo. Para obtener más información sobre las composiciones de edición de genes mediadas por TAL y los métodos de la presente divulgación, ver las Patentes de EE.UU. 8.440.432; 8.440.432; US 8.450.471; US 8.586.526; US 8.586.363; US 8.592.645; US 8.697.853; 8.704.041; 8.921.112; y 8.912.138, cada uno de las cuales se incorpora por la presente en su totalidad para todos los propósitos.

(ii) MegaTAL.

Diversas formas de realización ilustrativas contemplan una nucleasa megaTAL que se une y escinde una región objetivo de un locus que contribuye al factor de transcripción de la caja MADS. Un "megaTAL" se refiere a una nucleasa diseñada por ingeniería que comprende un dominio de unión de ADN TALE diseñado y una meganucleasa diseñada, y opcionalmente comprende uno o más enlazadores y/o dominios funcionales adicionales, por ejemplo, un dominio enzimático de procesamiento de extremo de una enzima de procesamiento de extremo que presenta una actividad de exonucleasa 5–3', una exonucleasa alcalina 5–3', una exonucleasa 3–5', una endonucleasa de aleta 5', helicasa o polimerasas de ADN independiente de patrón.

Un "dominio de unión de ADN TALE" es la porción de unión de ADN de los efectores de tipo activadores de la transcripción (efectores TALE o TAL), que imita a los activadores transcripcionales de la planta para manipular el transcriptoma de la planta (véase, por ejemplo, Kay et al., 2007. Science 318: 648– 651). Los dominios de unión de ADN de TALE contemplados en formas de realización particulares están diseñados *de novo* o a partir de TALE de origen natural. Ejemplos ilustrativos de proteínas TALE para derivar y diseñar dominios de unión de ADN se describen en la Patente de los Estados Unidos Nro. 9.017.967, y referencias allí citadas, todas las cuales se incorporan aquí como referencia en su totalidad.

En algunas formas de realización, las plantas de interés se modifican a través de megaTAL. En algunas formas de realización, las megaTAL son endonucleasas diseñadas capaces de dirigirse a secuencias de ADN seleccionadas e inducir roturas de ADN.

(iii) Meganucleasas/Endonucleasas de orientación (*homing*) (HE).

Las meganucleasas son endonucleasas específicas de secuencia que se originan a partir de una diversidad de organismos tales como bacterias, levaduras, algas y orgánulos vegetales. Se conocen varias meganucleasas en la técnica; ver, por ejemplo, el documento WO 2012010976 (variantes de meganucleasas que escinden secuencias diana de ADN del gen TERT); Pat. U. S.. Nros. 8.021.867; 8.119.361 y 8.119.381 (meganucleasas I–CreI); Pat. U. S. Nro. 7.897.372 (variantes de meganucleasa I–CreI con modificada especificidad).

En algunas formas de realización, una endonucleasa orientadora o meganucleasa está diseñada para unirse, e introducir mellas monocatenarias o roturas bicatenarias (DSB) a uno o más loci que contribuyen al factor de transcripción de la caja MADS. La "endonucleasa de orientación" y la "meganucleasa" se usan indistintamente y se refieren a nucleasas de origen natural o meganucleasas diseñadas genéticamente que reconocen sitios de escisión de 12–45 pares de bases y se agrupan comúnmente en cinco familias según los motivos de secuencia y estructura: LAGLIDADG, GIY–YIG, HNH, caja His–Cys y PD–(D/E)XK.

Las HE de ingeniería no existen en la naturaleza y pueden obtenerse mediante tecnología de ADN recombinante o mediante mutagénesis aleatoria. Las HE diseñadas pueden obtenerse haciendo una o más alteraciones de aminoácidos, por ejemplo, mutando, sustituyendo, agregando o deleteando uno o más aminoácidos, en una HE natural o HE previamente diseñada. En formas de realización particulares, una HE diseñada comprende una o más alteraciones de aminoácidos en la interfaz de reconocimiento de ADN. Las HE de ingeniería contempladas en formas de realización particulares pueden comprender además uno o más enlazadores y/o dominios funcionales adicionales, por

ejemplo, un dominio enzimático de procesamiento de extremo de una enzima de procesamiento de extremo que exhibe actividad de 5–3' exonucleasa, 5–3' exonucleasa alcalina, 3–5' exonucleasa, 5' aleta endonucleasa, helicasa o ADN polimerasas independiente de patrón.

En algunas formas de realización, las plantas de interés se modifican a través de meganucleasas. En algunas formas de realización, las meganucleasas son endonucleasas diseñadas genéticamente capaces de dirigirse a secuencias de ADN seleccionadas e inducir roturas de ADN. En algunas formas de realización, se desarrollan nuevas meganucleasas dirigidas a regiones específicas a través de técnicas recombinantes que combinan los motivos de unión al ADN de diversas otras nucleasas identificadas. En otras formas de realización, se crean nuevas meganucleasas mediante análisis mutacional semirracional, que intenta modificar la estructura de los dominios de unión existentes para obtener especificidad para secuencias adicionales. Para obtener más información sobre el uso de meganucleasas para la edición del genoma, ver Silva et al., 2011 *Current Gene Therapy* 11 pg 11–27; y Stoddard et al., 2014 *Mobile DNA* 5 pg 7, cada uno de los cuales se incorpora en su totalidad para todos los propósitos.

(iv) ZFN.

Las nucleasas de dedos de zinc (ZFN) están compuestas de módulos de unión de ADN de dedos de zinc programables y específicos de secuencia (ver arriba) unidos a un dominio de escisión de ADN no específico, por ejemplo, un dominio de escisión Fok I. Los métodos para fabricar y usar ZFN son conocidos en la técnica, véase, por ejemplo, (Maeder et al., 2008, *Mol. Cell*, 31: 294–301; Joung et al., 2010, *Nat. Methods*, 7: 91–92 ; Isalan et al., 2001, *Nat. Biotechnol.*, 19: 656–660; Sander et al., *Nat Methods*. 8 (1): 67–9, 2011; Bhakta et al., *Genome Res*. 23 (3) : 530–8, 2013). En algunas formas de realización, las ZFN se describen o se generan como se describe en los documentos WO 2011/017293 o WO 2004/099366. ZFN adicionales adecuadas se describen en las Pat. de EE. UU. Nros. 6.511.808, 6.013.453, 6.007.988 y 6.503.717 y Solicitud de Patente de EE. UU. 2002/0160940.

En algunas formas de realización, una nucleasa de dedo de zinc (ZFN) se une y escinde una región objetivo de un locus que contribuye al factor de transcripción de la caja MADS. Un "ZFN" se refiere a una nucleasa diseñada por ingeniería que comprende uno o más dominios de unión de ADN de dedo de zinc y un dominio de endonucleasa (o su medio dominio de endonucleasa), y opcionalmente comprende uno o más enlazadores y/o dominios funcionales adicionales, por ejemplo, un dominio enzimático de procesamiento de extremo de una enzima de procesamiento de extremo que exhibe actividad de exonucleasa 5–3', exonucleasa alcalina 5–3', exonucleasa 3–5', endonucleasa de aleta 5', helicasa o actividad de ADN polimerasas independiente de patrón.

En una forma de realización, la escisión bicatenaria dirigida se logra usando dos ZFN, cada uno de los cuales comprende un medio dominio de endonucleasa que puede usarse para reconstituir un dominio de escisión catalíticamente activo. En otra forma de realización, la escisión bicatenaria dirigida se consigue con un polipéptido único que comprende uno o más dominios de unión de ADN de dedo de zinc y dos medios dominios de endonucleasa.

En una forma de realización, una ZFN comprende un dominio de unión de ADN de TALE contemplado en otro lugar del presente documento, un dominio de unión de ADN del dedo de zinc y un dominio de endonucleasa (o medio dominio de endonucleasa) contemplado en otro lugar del presente documento.

En una forma de realización, una ZFN comprende un dominio de unión de ADN de dedo de zinc y una meganucleasa contemplada en otra parte del presente documento.

En formas de realización particulares, el ZFN comprende un dominio de unión de ADN de dedo de zinc que tiene uno, dos, tres, cuatro, cinco, seis, siete u ocho o más motivos de dedo de zinc y un dominio de endonucleasa (o medio dominio de endonucleasa). Típicamente, un único motivo de dedo de zinc tiene aproximadamente 30 aminoácidos de longitud. Los motivos de dedos de zinc incluyen tanto dedos de zinc canónicos C_2H_2 como dedos de zinc no canónicos tales como, por ejemplo, dedos de zinc C_3H y dedos de zinc C_4 .

Los dominios de unión de dedos de zinc se pueden diseñar para unir cualquier secuencia de ADN. Se han identificado dominios de unión a ADN de dedo de zinc candidatos para una secuencia objetivo de ADN de 3 pb dada, y se han ideado estrategias de ensamblaje modulares para unir una pluralidad de los dominios en un péptido multidedos dirigido a la secuencia objetivo de ADN compuesta correspondiente. Otros métodos adecuados conocidos en la técnica también se pueden usar para diseñar y construir ácidos nucleicos que codifican dominios de unión de ADN de dedos de zinc, por ejemplo, visualización de fagos, mutagénesis aleatoria, genotecas combinatorias, diseño informático/racional, selección por afinidad, PCR, clonación a partir de genotecas genómicas o de ADNc, construcción sintética y similares. (Ver, por ejemplo, la Patente de Estados Unidos Nro. 5.786.538; Wu et al., *PNAS* 92: 344348 (1995); Jamieson et al., *Biochemistry* 33: 5689–5695 (1994); Rebar & Pabo, *Science* 263: 671673 (1994); Choo y Klug, *PNAS* 91: 11163–11167 (1994); Choo y Klug, *PNAS* 91: 11168–11172 (1994); Desjarlais & Berg, *PNAS* 90:2256–2260 (1993); Desjarlais & Berg, *PNAS* 89:7345–7349 (1992); Pomerantz et al., *Science* 267:93–96 (1995); Pomerantz et al., *PNAS* 92:9752–9756 (1995); Liu et al., *PNAS* 94:5525–5530 (1997); Griesman & Pabo, *Science* 275:657–661 (1997); Desjarlais & Berg, *PNAS* 91:11–99–11103 (1994)).

Los motivos individuales de dedos de zinc se unen a una secuencia de tres o cuatro nucleótidos. La longitud de una secuencia a la que un dominio de unión de dedos de zinc está diseñado para unirse (por ejemplo, una secuencia objetivo) determinará el número de motivos de dedos de zinc en un dominio de unión de dedos de zinc diseñado. Por ejemplo, para los ZFN en los que los motivos de los dedos de zinc no se unen a subsitios superpuestos, una secuencia objetivo de seis nucleótidos está unida por un dominio de unión de dos dedos; una secuencia diana de nueve nucleótidos está unida por un dominio de unión de tres dedos, etc. En formas de realización particulares, los sitios de unión de ADN para motivos de dedos de zinc individuales en un sitio objetivo no necesitan ser contiguos, sino que pueden estar separados por uno o varios nucleótidos, de acuerdo con la longitud y la naturaleza de las secuencias de enlazador entre los motivos de dedos de zinc en un dominio de unión de multiedos.

En algunas formas de realización, las plantas de interés se modifican a través de nucleasas de dedos de zinc. Se reconocen tres variantes de la tecnología ZFN en la ingeniería del genoma de la planta (con aplicaciones que van desde la producción de mutaciones únicas o pequeñas deleciones/inserciones en el caso de las técnicas ZFN-1 y -2, hasta la introducción dirigida de nuevos genes, en el caso de la técnica ZFN-3):

ZFN-1: los genes que codifican ZFN se suministran a las células de la planta sin un patrón de reparación. Las ZFN se unen al ADN de la planta y generan roturas de doble hebra (DSB) específicas de sitio. El proceso natural de reparación del ADN (que ocurre a través de la unión de extremo no homólogo, NHEJ) conduce a mutaciones específicas de sitio, en uno o solo algunos pares de bases, o a deleciones o inserciones cortas.

ZFN-2: los genes que codifican ZFN se suministran a las células de la planta junto con un patrón de reparación homólogo al área objetivo, que abarca algunos pares de kilobases. Las ZFN se unen al ADN de la planta y generan DSB específicas de sitio. Los mecanismos naturales de reparación de genes generan mutaciones puntuales específicas de sitio, p. ej., cambios a uno o algunos pares de bases mediante recombinación homóloga y la copia del patrón de reparación.

ZFN-3: los genes que codifican las ZFN se suministran a las células vegetales junto con un tramo de ADN que puede tener varios pares de kilobases de largo y cuyos extremos son homólogos a las secuencias de ADN que flanquean el sitio de escisión. Como resultado, el tramo de ADN se inserta en el genoma de la planta de una manera específica de sitio.

(v) FokI.

FokI es una endonucleasa de restricción de tipo IIs que incluye un dominio de reconocimiento de ADN y un dominio catalítico (endonucleasa). Las proteínas de fusión descritas en el presente documento pueden incluir todo FokI o solo el dominio de endonucleasa catalítica, por ejemplo, los aminoácidos 388–583 o 408–583 de GenBank Acc. Nro. AAA24927.1, por ejemplo, como se describe en WO95/09233, Li et al., *Nucleic Acids Res.* 39 (1): 359–372 (2011); Cathomen y Joung, *Mol. Ther.* 16: 1200–1207 (2008), o una forma mutada de FokI como se describe en Miller et al. *Nat Biotechnol* 25: 778–785 (2007); Szczepek y col., *Nat Biotechnol* 25: 786–793 (2007); o Bitinaite y col., *Proc. Natl. Acad. Sci. USA.* 95: 10570–10575 (1998). Ver también Tsai et al., *Nat Biotechnol.* Junio de 2014; 32 (6): 569–76.

En algunas formas de realización, las plantas de interés se modifican a través de endonucleasas FokI.

Sistemas CRISPR–Cas dirigidos.

En este documento, el término "sistema de edición de bases dirigido" se refiere a una proteína, ácido nucleico o combinación de los mismos que es capaz de sustituir un solo nucleótido en un sitio objetivo y modificar una secuencia de ADN objetivo endógeno cuando se introduce en una célula, de modo de causar así un o más sustituciones de aminoácidos. Los numerosos sistemas de edición de genes adecuados para el uso en los métodos de la presente descripción incluyen, sin limitación, sistemas de nucleasas de dedos de zinc, sistemas TALEN y sistemas CRISPR/Cas.

En algunas formas de realización, se utiliza un sistema CRISPR/Cas inactivado por nucleasas que tiene una actividad de base desaminasa para una edición de bases dirigida. En otros aspectos, se usa una nickasa.

En algunas formas de realización, el sistema de edición de bases dirigido puede mediar un cambio en la secuencia del gen diana endógeno, por ejemplo, mediante la introducción de una o más mutaciones puntuales en la secuencia diana endógena, tal como sustituyendo C con T (o G con A) o A con G (o T con C) en la secuencia diana endógena.

En algunas formas de realización, el sistema de edición de bases dirigido puede mediar un cambio en la expresión de la proteína codificada por el gen objetivo endógeno. En tales formas de realización, el sistema de edición de bases dirigido puede regular la expresión de la proteína codificada mediante modificaciones de la secuencia de ADN diana endógeno, o actuando sobre el producto de ARNm codificado por la secuencia de ADN. En algunas formas de realización, el sistema de edición de bases dirigido puede dar como resultado la expresión de una proteína endógena

modificada. En algunas formas de realización, las modificaciones a la secuencia de ADN endógeno mediada por el sistema de edición de bases dirigido dan como resultado una función alterada de la proteína endógena modificada en comparación con la proteína endógena correspondiente en una célula vegetal no modificada. En tales formas de realización, el nivel de expresión de la proteína endógena modificada puede aumentarse, disminuirse o puede ser el mismo, o sustancialmente similar al nivel de expresión de la proteína endógena correspondiente en una célula vegetal no modificada.

La presente descripción proporciona un sistema de edición de bases dirigido para editar una secuencia de nucleótidos objetivo en el genoma de una planta, que comprende por lo menos uno de los siguientes; i) una proteína de fusión de edición de bases y un ARN guía; ii) una construcción de expresión que comprende una secuencia de nucleótidos que codifica una proteína de fusión de edición de bases y un ARN guía; iii) una proteína de fusión de edición de bases, y una construcción de expresión que comprende una secuencia de nucleótidos que codifica un ARN guía; iv) una construcción de expresión que comprende una secuencia de nucleótidos que codifica una proteína de fusión de edición de bases, y una construcción de expresión que comprende una secuencia de nucleótidos que codifica un ARN guía; v) una construcción de expresión que comprende una secuencia de nucleótidos que codifica la proteína de fusión de edición de bases y una secuencia de nucleótidos que codifica el ARN guía; en donde dicha proteína de fusión de edición de base contiene un dominio efector asociado a CRISPR y un dominio de desaminasa, dicho ARN guía puede dirigir dicha proteína de fusión de edición de base a la secuencia objetivo en el genoma de la planta. En algunas formas de realización, una proteína de fusión de edición de base dirigida comprende un dominio efector asociado a CRISPR inactivado por nucleasa y un dominio citidina desaminasa. En algunas formas de realización, una proteína de fusión de edición de base dirigida comprende un dominio efector asociado a CRISPR inactivado por nucleasa y un dominio de adenosina desaminasa.

Las proteínas de fusión de la invención pueden comprender dominios de unión de ADN específicos de secuencia, polipéptidos CRISPR–Cas y/o dominios de desaminasa fusionados a etiquetas peptídicas o polipéptidos de afinidad que interactúan con las etiquetas peptídicas, como se conoce en la técnica, para usar en el reclutamiento de la desaminasa al ácido nucleico objetivo. Los métodos de reclutamiento también pueden comprender ácidos nucleicos guía unidos a motivos de reclutamiento de ARN y desaminasas fusionadas a polipéptidos de afinidad capaces de interactuar con motivos de reclutamiento de ARN, de modo de reclutar así la desaminasa al ácido nucleico diana. Alternativamente, las interacciones químicas pueden usarse para reclutar polipéptidos (por ejemplo, desaminasas) a un ácido nucleico diana.

Una etiqueta peptídica (p. ej., epítipo) útil con esta invención puede incluir, sin limitación, una etiqueta peptídica GCN4 (p. ej., etiqueta Sun–Tag), una etiqueta de afinidad c–Myc, una etiqueta de afinidad HA, una etiqueta de afinidad His, una etiqueta de afinidad S, una etiqueta de afinidad de metionina–His, una etiqueta de afinidad de RGD–His, un octapéptido FLAG, una etiqueta de estrep o una etiqueta de estrep II, una etiqueta de V5 y/o un epítipo VSV–G. Cualquier epítipo que pueda unirse a un polipéptido y para el que haya un polipéptido de afinidad correspondiente que pueda unirse a otro polipéptido puede usarse con esta invención como una etiqueta peptídica. En algunas formas de realización, una etiqueta peptídica puede comprender 1 o 2 o más copias de una etiqueta peptídica (por ejemplo, unidad de repetición, epítipo multimerizado (por ejemplo, repeticiones en tándem)) (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 o más unidades repetidas). En algunas formas de realización, un polipéptido de afinidad que interactúa o se une con una etiqueta peptídica puede ser un anticuerpo. En algunas formas de realización, el anticuerpo puede ser un anticuerpo scFv. En algunas formas de realización, un polipéptido de afinidad que se une a una etiqueta peptídica puede ser sintético (por ejemplo, evolucionado para interacción de afinidad) que incluye, sin limitación, un aficuerpo, una anticalina, un monocuerpo y/o un DARPin (véase, por ejemplo, Sha et al., *Protein Sci.* 26 (5): 910–924 (2017)); Gilbreth (*Curr Opin Struc Biol* 22 (4): 413–420 (2013)), Patente de los Estados Unidos Núm. 9.982.053, cada una de las cuales se incorpora por referencia en su totalidad para las enseñanzas pertinentes a aficuerpos, anticalinas, monocuerpos y/o DARPins.

En algunas formas de realización, un ácido nucleico guía puede unirse a un motivo de reclutamiento de ARN, y un polipéptido por reclutar (por ejemplo, una desaminasa) puede fusionarse a un polipéptido de afinidad que se une al motivo de reclutamiento de ARN, en donde la guía se une al ácido nucleico diana, y el motivo de reclutamiento de ARN se une al polipéptido de afinidad, de manera de reclutar el polipéptido a la guía y contactar el ácido nucleico diana con el polipéptido (por ejemplo, desaminasa). En algunas formas de realización, se pueden reclutar dos o más polipéptidos para un ácido nucleico guía, poniendo en contacto el ácido nucleico diana con dos o más polipéptidos (por ejemplo, desaminasas).

En algunas formas de realización, un polipéptido fusionado con un polipéptido de afinidad puede ser una transcriptasa inversa, y el ácido nucleico guía puede ser un ácido nucleico guía extendido unido a un motivo de reclutamiento de ARN. En algunas formas de realización, un motivo de reclutamiento de ARN puede ubicarse en el extremo 3' de la porción extendida de un ácido nucleico guía extendido (por ejemplo, porción extendida de espaciador–repetición 5'–3' (sitio de unión de

patrón–cebador RT) – motivo de reclutamiento de ARN). En algunas formas de realización, un motivo de reclutamiento de ARN puede incrustarse en la porción extendida.

En algunas formas de realización de la invención, un ARN guía extendido y/o ARN guía puede estar unido a uno o dos o más motivos de reclutamiento de ARN (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más motivos; por ejemplo, por lo menos 10 a alrededor de 25 motivos), opcionalmente en donde los dos o más motivos de reclutamiento de ARN pueden ser el mismo motivo de reclutamiento de ARN o diferentes motivos de reclutamiento de ARN. En algunas formas de realización, un motivo de reclutamiento de ARN y el polipéptido de afinidad correspondiente pueden incluir, pero sin limitación, un motivo de unión de telomerasa Ku (por ejemplo, horquilla de unión a Ku) y el polipéptido de afinidad correspondiente Ku (por ejemplo, heterodímero de Ku), un motivo de unión Sm7 de telomerasa motivo y el correspondiente polipéptido de afinidad Sm7, un bucle de tallo de operador de fago MS2 y el correspondiente polipéptido de afinidad de proteína de cubierta MS2 (MCP), un bucle de tallo de operador de fago PP7 y el correspondiente polipéptido de afinidad de proteína de cubierta PP7 (PCP), un bucle de tallo Com de fago SfMu y el correspondiente polipéptido de afinidad de proteína de unión a ARN Com, un sitio de unión PUF (PBS) y el polipéptido de afinidad de factor de unión a ARNm de Pumilio/fem–3 (PUF), y/o un aptámero de ARN sintético y el ligando de aptámero como polipéptido de afinidad correspondiente. En algunas formas de realización, el motivo de reclutamiento de ARN y el polipéptido de afinidad correspondiente pueden ser un bucle de tallo de operador de fago MS2 y el polipéptido de afinidad de proteína de cubierta MS2 (MCP). En algunas formas de realización, el motivo de reclutamiento de ARN y el polipéptido de afinidad correspondiente pueden ser un sitio de unión a PUF (PBS) y el polipéptido de afinidad de factor de unión a ARNm de Pumilio/fem–3 (PUF).

En algunas formas de realización, los componentes para reclutar polipéptidos y ácidos nucleicos pueden ser aquellos que funcionan a través de interacciones químicas que pueden incluir, sin limitación, dimerización inducible por rapamicina de FRB–FKBP; Biotina–estreptavidina; etiqueta SNAP; etiqueta de halo; etiqueta CLIP; heterodímero DmrA–DmrC inducido por un compuesto; ligando bifuncional (por ejemplo, fusión de dos productos químicos de unión a proteínas juntos; por ejemplo, dihirofolato reductasa (DHFR).

En algunas formas de realización, el sistema de edición de bases dirigido descrito en este documento contiene un sistema de nucleasa CRISPR (repeticiones palindrómicas cortas agrupadas regularmente intercaladas)/Cas (asociado a CRISPR) como un dominio efector asociado a CRISPR. En algunas formas de realización, el sistema CRISPR/Cas es un sistema de Clase 2. Los sistemas

CRISPR/Cas de clase 2 se dividen en tres tipos: sistemas Tipo II, Tipo V y Tipo VI. En algunas formas de realización, el sistema CRISPR/Cas es un sistema de Clase 2 Tipo II, que utiliza la proteína Cas9. En tales formas de realización, el polipéptido modificador dirigido al sitio es una endonucleasa de ADN Cas9 (o una variante de la misma), y la molécula guía de ácido nucleico es un ARN guía (ARNg). En algunas formas de realización, el sistema CRISPR/Cas es un sistema Clase 2 Tipo V, que utiliza las proteínas Cas12 (por ejemplo, Cas12a (también conocido como Cpf1), Cas12b (también conocido como C2c1), Cas12c (también conocido como C2c3), Cas12d (también conocido como CasY) y Cas12e (también conocido como CasX)). En algunas formas de realización, el polipéptido modificador dirigido al sitio es una endonucleasa de ADN Cas12 (o una variante de la misma), y la molécula guía de ácido nucleico es un ARNg. En algunas formas de realización, el sistema CRISPR/Cas es un sistema de Clase 2 y Tipo VI, que utiliza las proteínas Cas13 (por ejemplo, Cas13a (también conocido como C2c2), Cas13b y Cas13c). (Ver, Pyzocha et al., ACS Chemical Biology, 13 (2), 347–356). En algunas formas de realización, el polipéptido modificador dirigido al sitio es una riboendonucleasa de ARN Cas13, y la molécula guía de ácido nucleico es un ARNg.

Un polipéptido Cas se refiere a un polipéptido que puede interactuar con una molécula de ARNg y, junto con la molécula de ARNg, orientarse o localizarse en una secuencia de ADN o ARN objetivo. Los polipéptidos Cas incluyen proteínas Cas de origen natural y proteínas Cas diseñadas mediante ingeniería, alteradas o modificadas de otro modo que difieren en uno o más residuos de aminoácidos de una secuencia Cas de origen natural.

Un ARN guía (ARNg) comprende dos segmentos: un segmento de unión a ADN y un segmento de unión a proteína. En algunas formas de realización, el segmento de unión a proteína de un ARNg está comprendido en una molécula de ARN, y el segmento de unión a ADN está comprendido en otra molécula de ARN separada. Dichas formas de realización se denominan en el presente documento "ARNm de doble molécula" o "ARNg de dos moléculas" o "ARNg dual". En algunas formas de realización, el ARNg es una molécula de ARN única y se denomina en el presente documento "ARN guía única" o "ARNsg". El término "ARN guía" o "ARNg" es inclusivo, refiriéndose tanto a ARN guía de dos moléculas como a ARNsg.

El segmento de unión a proteína de un ARNg comprende, en parte, dos tramos complementarios de nucleótidos que se hibridan entre sí para formar una dupla de ARN bicatenario (dupla de ARNd), que facilita la unión a la proteína Cas. El segmento de unión a ácido nucleico (o "secuencia de unión a ácido nucleico") de un ARNg comprende una secuencia de nucleótidos que es complementaria y capaz de unirse a una secuencia de ácido nucleico diana específica. El segmento

de unión a proteína del ARNg interactúa con un polipéptido Cas, y la interacción de la molécula de ARNg y el polipéptido modificador dirigido al sitio produce la unión de Cas a la secuencia de ácido nucleico endógeno y produce una o más modificaciones dentro o alrededor de la secuencia de ácido nucleico diana. La ubicación precisa del sitio de modificación objetivo está determinada por (i) la complementariedad de emparejamiento de bases entre el ARNg y la secuencia de ácido nucleico objetivo; y (ii) la ubicación de un motivo corto, denominado motivo adyacente de protoespaciador (PAM), en la secuencia de ADN objetivo (denominada secuencia flanqueante de protoespaciador (PFS) en secuencias de ARN objetivo). La secuencia PAM/PFS se requiere para la unión de Cas a la secuencia de ácido nucleico diana. Se conoce una diversidad de secuencias PAM/PFS en la técnica, que son adecuadas para usar con una endonucleasa Cas particular (por ejemplo, una endonucleasa Cas9) (ver, por ejemplo, Nat Methods. 2013 Nov; 10 (11): 1116–1121 y Sci Rep 2014; 4: 5405). En algunas formas de realización, la secuencia PAM se encuentra dentro de 50 pares de bases del sitio de modificación objetivo en una secuencia de ADN objetivo. En algunas formas de realización, la secuencia PAM se encuentra dentro de 10 pares de bases del sitio de modificación objetivo en una secuencia de ADN objetivo. Las secuencias de ADN que pueden ser dirigidas por este método están limitadas solo por la distancia relativa de la secuencia PAM al sitio de modificación objetivo y la presencia de una secuencia única de 20 pares de bases para mediar la unión de Cas mediada por ARNg específica de secuencia. En algunas formas de realización, la secuencia PFS se encuentra en el extremo 3' de la secuencia de ARN objetivo. En algunas formas de realización, el sitio de modificación del objetivo está ubicado en el extremo 5' del locus del objetivo. En algunas formas de realización, el sitio de modificación objetivo está ubicado en el extremo 3' del locus objetivo. En algunas formas de realización, el sitio de modificación objetivo está ubicado dentro de un intrón o un exón del locus objetivo.

En algunas formas de realización, la presente divulgación proporciona un polinucleótido que codifica un ARNg. En algunas formas de realización, un ácido nucleico que codifica ARNg está comprendido en un vector de expresión, por ejemplo, un vector de expresión recombinante. En algunas formas de realización, la presente descripción proporciona un polinucleótido que codifica un polipéptido modificador dirigido al sitio. En algunas formas de realización, el polinucleótido que codifica un polipéptido modificador dirigido al sitio está comprendido en un vector de expresión, por ejemplo, un vector de expresión recombinante.

1. Proteínas Cas.

En algunas formas de realización, el sistema de edición de bases dirigido comprende una proteína Cas como un dominio efector asociado a CRISPR. Las moléculas Cas de una diversidad de especies pueden usarse en los métodos y las composiciones descritos en este documento, incluidas las moléculas Cas derivadas de *S. pyogenes*, *S. aureus*, *N. meningitidis*, *S. thermophiles*, *Acidovorax avenae*, *Actinobacillus pleuropneumoniae*, *Actinobacillus succinogenes*, *Actinobacillus suis*, *Actinomyces* sp., *Cycliphilus denitrificans*, *Aminomonas paucivorans*, *Bacillus cereus*, *Bacillus smithii*, *Bacillus thuringiensis*, género *Bacteroides*, *Blastopirellula marina*, género *Bradyrhizobium*, *Brevibacillus laterosporus*, *Campylobacter coli*, *Campylobacter jejuni*, *Campylobacter lari*, *Candidatus puniceispirillum*, *Clostridium cellulolyticum*, *Clostridium perfringens*, *Corynebacterium accolens*, *Corynebacterium diphtheria*, *Corynebacterium matruchotii*, *Dinoroseobacter shibae*, *Eubacterium dolichum*, *Gammaproteobacterium*, *Gluconacetobacter diazotrophicus*, *Haemophilus parainfluenzae*, *Haemophilus sputum*, *Helicobacter canadensis*, *Helicobacter cinaedi*, *Helicobacter mustelae*, *Ilyobacter polytropus*, *Kingella kingae*, *Lactobacillus crispatus*, *Listeria ivanovii*, *Listeria monocytogenes*, *Listeriaceae bacterium*, género *Methylocystis*, *Methylosinus trichosporium*, *Mobiluncus mulieris*, *Neisseria bacilliformis*, *Neisseria cinerea*, *Neisseria flavescens*, *Neisseria lactamica*, *Neisseria meningitidis*, género *Neisseria*., *Neisseria wadsworthii*, género *Nitrosomonas*, *Parvibaculum lavamentivorans*, *Pasteurella multocida*, *Phascolarctobacterium succinatutens*, *Ralstonia syzygii*, *Rhodopseudomonas palustris*, género *Rhodovulum*, *Simonsiella muelleri*, género *Sphingomonas*, *Sporolactobacillus vineae*, *Staphylococcus aureus*, *Staphylococcus lugdunensis*, género *Streptococcus*., género *Subdoligranulum*, *Tistrella mobilis*, género *Treponema*, o *Verminephrobacter eiseniae*.

En algunas formas de realización, la proteína Cas es una proteína Cas natural. En otras formas de realización, la proteína Cas es una proteína Cas diseñada mediante ingeniería. En algunas formas de realización, la endonucleasa Cas incluye, sin limitación, C2C1, C2C3, Cpf1 (también denominada Cas12a), Cas12b, Cas12c, Cas12d, Cas12e, Cas13a, Cas13b, Cas13c, Cas13d, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también denominada Csn1 y Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Cse1, Cse2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, y Csf4.

En algunas formas de realización, la proteína Cas es una endorribonucleasa tal como una proteína Cas13. En algunas formas de realización, la proteína Cas13 es una proteína Cas13a (Abudayyeh *et al.*, Nature 550 (2017), 280284), Cas13b (Cox *et al.*, Science (2017) 358:6336, 1019–

1027), Cas13c (Cox *et al.*, Science (2017) 358:6336, 1019–1027), o Cas13d (Zhang *et al.*, Cell 175 (2018), 212–223).

En algunas formas de realización, la proteína Cas es una proteína Cas9 de tipo salvaje o natural o un ortólogo Cas9. La Cas9 de tipo salvaje es una enzima multidominio que utiliza un dominio de nucleasa HNH para escindir la hebra diana de ADN y un dominio de tipo RuvC para escindir la hebra no diana. La unión de Cas9 WT al ADN en función de la especificidad de ARNg produce roturas de ADN bicatenario que pueden repararse mediante unión de extremo no homólogo (NHEJ) o reparación dirigida por homología (HDR).

En algunas formas de realización, un polipéptido Cas9 natural puede incluir, sin limitación, SpCas9, SpCas9–HF1, SpCas9–HF2, SpCas9–HF3, SpCas9–HF4, SaCas9, FnCpf, FnCas9, eSpCas9, y NmeCas9. En algunas formas de realización, la proteína Cas9 comprende una secuencia de aminoácidos que tiene por lo menos 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con una secuencia de aminoácidos Cas9 descrita en las referencias de Chylinski *et al.*, RNA Biology 2013 10:5, 727–737; Hou *et al.*, PNAS Early Edition 2013, 1–6).

En algunas formas de realización, un polipéptido Cas comprende una o más de las siguientes actividades: (a) una actividad de nickasa, es decir, la capacidad de escindir una hebra sencilla, por ejemplo, la hebra no complementaria o la hebra complementaria, de una molécula de ácido nucleico; (b) una actividad de nucleasa de doble hebra, es decir, la capacidad de escindir ambas hebras de un ácido nucleico de doble hebra y crear una ruptura de doble hebra, que en una forma de realización, es la presencia de dos actividades de nickasa; (c) una actividad endonucleasa; (d) una actividad exonucleasa; y/o (e) una actividad helicasa, es decir, la capacidad de desenrollar la estructura helicoidal de un ácido nucleico bicatenario. En otras formas de realización, la proteína Cas puede estar muerta o inactiva (por ejemplo, dCas).

En algunas formas de realización, el polipéptido Cas se fusiona con un polipéptido/proteína heterólogo que tiene, por ejemplo, actividad de desaminasa base.

En algunas formas de realización, diferentes proteínas Cas (es decir, proteínas Cas9 de diversas especies) pueden ser ventajosas para usar en los diversos métodos proporcionados con el fin de capitalizar diversas características enzimáticas de las diferentes proteínas Cas (por ejemplo, para diferentes preferencias de secuencia PAM; para aumentar o disminuir la actividad enzimática; para un nivel aumentado o disminuido de toxicidad celular; para cambiar el equilibrio entre NHEJ, reparación dirigida por homología, roturas de hebra sencilla, roturas de hebra doble, etc.).

En algunas formas de realización, la proteína Cas es una proteína Cas9 derivada de *S. pyogenes* y reconoce el motivo de secuencia PAM NGG, NAG, NGA (Mali *et al*, Science 2013; 339(6121): 823826). En algunas formas de realización, la proteína Cas es una proteína Cas9 derivada de *S. thermophiles* y reconoce el motivo de secuencia PAM NGGNG y/o NNAGAAW (W = A o T) (ver, por ejemplo, Horvath *et al*, Science, 2010; 327(5962): 167–170, y Deveau *et al*, J Bacteriol 2008; 190(4): 1390–1400). En algunas formas de realización, la proteína Cas es una proteína Cas9 derivada de *S. mutans* y reconoce el motivo de secuencia PAM NGG y/o NAAR (R = A o G) (ver, por ejemplo, Deveau *et al*, J BACTERIOL 2008; 190(4): 1390–1400). En algunas formas de realización, la proteína Cas es una proteína Cas9 derivada de *S. aureus* y reconoce el motivo de secuencia PAM NNGRR (R = A o G). En algunas formas de realización, la proteína Cas es una proteína Cas9 derivada de *S. aureus* y reconoce el motivo de secuencia PAM N GRRT (R = A o G). En algunas formas de realización, la proteína Cas es una proteína Cas9 derivada de *S. aureus* y reconoce el motivo de secuencia PAM N GRRV (R = A o G). En algunas formas de realización, la proteína Cas es una proteína Cas9 derivada de *N. meningitidis* y reconoce el motivo de secuencia PAM N GATT o N GCTT (R = A o G, V = A, G o C) (ver, por ejemplo, Hou *et al*, PNAS 2013, 1–6). En las formas de realización mencionadas, N puede ser cualquier residuo de nucleótido, por ejemplo, cualquiera de A, G, C o T. En algunas formas de realización, la proteína Cas es una proteína Cas13a derivada de *Leptotrichia shahii* y reconoce el motivo de secuencia PFS de un único 3' A, U, o C.

En algunas formas de realización, una proteína Cas como un dominio efector asociado a CRISPR es optimizada en codones sobre la base del genoma de la planta por ser apunado.

En algunas formas de realización, el por lo menos un efector asociado a CRISPR es una nucleasa, por ejemplo, una nucleasa CRISPR que incluye, sin limitación, nucleasas Cas9 o Cpf1 (Cas12a).

En algunas formas de realización, el por lo menos un efector asociado a CRISPR es un polipéptido Cas, en donde el polipéptido Cas comprende un dominio de unión a ADN específico del sitio unido a por lo menos un editor de base. El efector asociado a CRISPR o la secuencia de ácido nucleico que codifica el mismo incluye, sin limitación, (i) Cas9, que incluye SpCas9, SaCas9, SaKKH–Cas9, VQRCas9, StlCas9, (ii) Cpf1, que incluye AsCpf1, LbCpf1, FnCpf1, (iii) CasX, o (iv) CasY, o cualquier variante o derivado del efector asociado a CRISPR mencionado anteriormente, preferentemente en donde el por lo menos un efector asociado a CRISPR comprende una mutación en comparación con la secuencia de tipo salvaje respectiva, de modo que el efector asociado a

CRISPR resultante se convierte en una ADN nickasa específica de hebra sencilla, o en un efector de unión a ADN que carece de toda capacidad de escisión de ADN, como se describe a continuación. Por lo tanto, de acuerdo con la presente divulgación, se prevén nucleasas CRISPR modificadas artificialmente, que de hecho podrían no ser "nucleasas" en el sentido de enzimas de escisión de doble hebra, pero que son variantes muertas de nucleasas o nickasas, que todavía tienen reconocimiento de ADN inherente y, por lo tanto, capacidad de unión.

2. Mutantes/variantes Cas.

En algunas formas de realización, la proteína Cas descrita anteriormente se puede diseñar para alterar una o más propiedades del polipéptido Cas. Por ejemplo, en algunas formas de realización, el polipéptido Cas comprende propiedades enzimáticas alteradas, por ejemplo, actividad de nucleasa alterada (en comparación con una molécula Cas natural u otra de referencia) o actividad de helicasa alterada. En algunas formas de realización, un polipéptido Cas diseñado puede tener una alteración que altera su tamaño, por ejemplo, una delección de la secuencia de aminoácidos que reduce su tamaño sin un efecto significativo sobre otra propiedad del polipéptido Cas. En algunas formas de realización, un polipéptido Cas diseñado comprende una alteración que afecta el reconocimiento de PAM. Por ejemplo, un polipéptido Cas diseñado puede alterarse para reconocer una secuencia PAM distinta de la secuencia PAM reconocida por la proteína Cas de tipo salvaje correspondiente. En algunas formas de realización, el sistema de edición de bases dirigido comprende una proteína Cas como un dominio efector asociado a CRISPR.

Los polipéptidos Cas con propiedades deseadas pueden prepararse de varias maneras, que incluyen la alteración de un polipéptido Cas natural o polipéptido Cas parental, para proporcionar un polipéptido Cas mutante o alterado que tenga una propiedad deseada. Por ejemplo, se pueden introducir una o más mutaciones en la secuencia de un polipéptido Cas parental (por ejemplo, un polipéptido Cas de origen natural o diseñado por ingeniería genética). Dichas mutaciones y diferencias pueden comprender sustituciones (por ejemplo, sustituciones conservadoras o sustituciones de aminoácidos no esenciales); inserciones; o delecciones. En algunas formas de realización, un polipéptido Cas mutante comprende una o más mutaciones (por ejemplo, por lo menos 1, 2, 3, 4, 5, 10, 15, 20, 30, 40 o 50 mutaciones) en relación con un polipéptido Cas parental.

En una forma de realización, un polipéptido Cas mutante comprende una propiedad de escisión que difiere de un polipéptido Cas de origen natural. En algunas formas de realización, el Cas es un mutante Cas desactivado (dCas), que está catalíticamente muerto. En tales formas de realización, el polipéptido Cas no comprende ninguna actividad enzimática intrínseca y es incapaz de

mediar en la escisión del ácido nucleico diana. En tales formas de realización, el dCas se fusiona con una proteína heteróloga que es capaz de modificar el ácido nucleico diana de una manera no sustentada en la escisión. En algunas formas de realización, el sistema de edición de bases dirigido comprende una proteína Cas como un dominio efector asociado a CRISPR.

En algunas formas de realización, el dCas es un mutante dCas9. En algunas formas de realización, una proteína dCas se fusiona con dominios de desaminasa base (por ejemplo, citidina desaminasa o adenosina desaminasa). En algunos de estos casos, la proteína de fusión dCas es apuntada por el ARNg a una ubicación específica (es decir, secuencia) en el ácido nucleico objetivo y ejerce una modificación específica del locus, tal como el reemplazo de C con T (o G con A) si la proteína de fusión tiene actividad citidina desaminasa) o el reemplazado de A con G (o T con C), si la proteína de fusión tiene actividad adenosina desaminasa.

En algunas formas de realización, el dCas es un mutante dCas13 (Konermann et al., Cell 173 (2018), 665–676). Estos mutantes dCas13 pueden entonces fusionarse con enzimas que modifican el ARN, incluidas las adenosina desaminasas (por ejemplo, ADAR1 y ADAR2). Las adenosina desaminasas convierten la adenina en inosina, que la maquinaria de traducción trata como guanina, de modo de crear así un cambio funcional de A a G en la secuencia diana.

En algunas formas de realización, la proteína efectora asociada a CRISPR es la endonucleasa Cas9. En algunas formas de realización, la proteína efectora asociada a CRISPR es una variante de CRISPR–Cas, que es mutante dCas9 o mutante nCas9 nickasa. La endonucleasa Cas9 tiene un dominio de escisión de ADN que contiene dos subdominios: i) el subdominio RuvC que escinde la hebra monocatenaria no complementaria y ii) el subdominio de nucleasa HNH que escinde la hebra que es complementaria al ARNg. Las mutaciones en estos subdominios pueden inactivar la endonucleasa Cas9 para formar Cas9 desactivado (dCas9), que se usa indistintamente con "Cas9 catalíticamente muerto". El Cas9 inactivado por nucleasa retiene la capacidad de unión al ADN dirigida por ARNg. Por lo tanto, en principio, cuando se fusiona con una proteína adicional, el dCas9 puede simplemente dirigir dicha proteína adicional a casi cualquier secuencia de ADN mediante la coexpresión con el ARN guía apropiado. Por ejemplo, Cas9 catalíticamente muerto (dCas9), que contiene las mutaciones Asp10Ala (D10A) y His840Ala (H840A) que inactivan su actividad nucleasa, conserva su capacidad de unirse al ADN de una manera programada por el ARN guía, pero no escinde el esqueleto del ADN (Komor et al., Nature (2016), Vol 533: 420–424). En algunas formas de realización, la conjugación de dCas9 con un catalizador enzimático o químico que media la

conversión directa de una base a otra podría permitir la edición de bases de ADN programada con ARN.

En algunas formas de realización, el mutante Cas9 es un mutante Cas9 nickasa (nCas9). Los mutantes de la nickasa Cas9 comprenden solo un dominio catalíticamente activo (ya sea el dominio RuvC (D10A) o el dominio HNH (H840A)). Los mutantes de la Cas9 nickasa retienen la unión al ADN en función de la especificidad del ARNg, pero son capaces de cortar solo una hebra de ADN, lo que produce una rotura de cadena sencilla (por ejemplo, una "mella"). En algunas formas de realización, dos mutantes complementarios de la nickasa Cas9 (por ejemplo, un mutante de la nickasa Cas9 con un dominio RuvC inactivado y un mutante de la nickasa Cas9 con un dominio HNH inactivado) se expresan en la misma célula con dos ARNg correspondientes a dos secuencias diana respectivas; una secuencia objetivo en la hebra de ADN sentido, y otra en la hebra de ADN antisentido. Este sistema de doble nickasa produce roturas escalonadas de doble hebra y puede aumentar la especificidad de objetivo, ya que es poco probable que se generen dos mellas fuera del objetivo lo suficientemente cerca como para generar una ruptura de doble hebra. En algunas formas de realización, un mutante de Cas9 nickasa se coexpresa con un patrón de reparación de ácido nucleico para facilitar la incorporación de una secuencia exógena de ácido nucleico mediante reparación dirigida por homología.

El dCas9 de la presente descripción puede derivarse de Cas9 de diferentes especies, por ejemplo, derivado de Cas9 de *S. pyogenes* (SpCas9). Las mutaciones tanto en el subdominio RuvC como en el subdominio nucleasa HNH de SpCas9 (incluye, por ejemplo, las mutaciones D10A y H840A) inactivan la nucleasa Cas9 de *S. pyogenes*, lo que produce una Cas9 muerta catalíticamente/muerta por nucleasa (dCas9). En algunas formas de realización de la presente descripción, la Cas9 inactivada por nucleasa comprende el dCas9. En algunas formas de realización preferidas, la Cas9 inactivada por nucleasa comprende.

La inactivación de uno de los subdominios por mutación permite a Cas9 ganar actividad de nickasa, es decir, para producir una nickasa Cas9 (nCas9), por ejemplo, nCas9 son solo una mutación D10A.

En algunas formas de realización, el Cas9 inactivado por nucleasa comprende sustituciones de aminoácidos D10A y/o H840A con respecto al Cas9 de tipo salvaje. En algunas formas de realización, el Cas9 inactivado por nucleasas de la presente descripción pierde completamente la actividad de nucleasas, que está catalíticamente muerta. En tales formas de realización, la Cas9

inactivada por nucleasa es dCas9 con D10A y H840A. Por lo tanto, el término "Cas9 inactivada por nucleasas" se refiere a dCas9 y/o nCas9.

En algunas formas de realización, la Cas9 inactivada por nucleasas de la presente descripción tiene actividad de nickasa. En algunas formas de realización, la Cas9 inactivada por nucleasas es una Cas9 nickasa que retiene la actividad de escisión del subdominio HNH de Cas9, mientras que la actividad de escisión del subdominio RuvC está inactivada. Por ejemplo, la Cas9 inactivada por nucleasa contiene una sustitución de aminoácidos D10A en relación con la Cas9 de tipo salvaje. En tales formas de realización, la Cas9 inactivada por nucleasa es la nCas9 con D10A solamente. En algunas formas de realización de la presente descripción, la Cas9 inactivada por nucleasa comprende la nCas9.

En algunas formas de realización, los polipéptidos Cas descritos en el presente documento pueden diseñarse para modificar la especificidad PAM/PFS del polipéptido Cas. En algunas formas de realización, un polipéptido Cas mutante tiene una especificidad PAM/PFS que es diferente de la especificidad PAM / PFS del polipéptido Cas parental. Por ejemplo, una proteína Cas de origen natural se puede modificar para alterar la secuencia PAM/PFS que el polipéptido Cas mutante reconoce para disminuir los sitios fuera de objetivo, mejorar la especificidad o eliminar un requisito de reconocimiento de PAM/PFS. En algunas formas de realización, una proteína Cas puede modificarse para aumentar la longitud de la secuencia de reconocimiento PAM/PFS. En algunas formas de realización, la longitud de la secuencia de reconocimiento de PAM es de por lo menos 4, 5, 6, 7, 8, 9, 10 o 15 aminoácidos de longitud. Los polipéptidos Cas que reconocen diferentes secuencias de PAM/PFS y/o tienen una actividad reducida fuera del objetivo pueden generarse usando la evolución dirigida. Se describen métodos y sistemas ejemplares que pueden usarse para la evolución dirigida de polipéptidos Cas, por ejemplo, en Esvelt et al. Nature 2011, 472 (7344): 499–503. Los mutantes Cas ejemplares se describen en la publicación internacional PCT Nro. WO 2015/161276 y Konermann et al., Cell 173 (2018), 665–676, que se incorporan aquí como referencia en su totalidad.

3. Editores de bases.

En algunas formas de realización, el dominio de desaminasa está fusionado con el extremo N-terminal del dominio Cas9 inactivado por nucleasa. En algunas formas de realización, el dominio de desaminasa está fusionado con el extremo C-terminal del dominio Cas9 inactivado por nucleasa. En algunas formas de realización, el dominio de desaminasa y el dominio Cas9 inactivado por nucleasa se fusionan a través de un enlazador. El enlazador puede ser una secuencia de aminoácidos

no funcional que no tiene estructura secundaria o superior, extremo N-terminal y uno o más NLS en el extremo C-terminal. Cuando hay más de una NLS, cada NLS puede seleccionarse como independiente de otras NLS. En algunas formas de realización, la proteína de fusión de edición de base dirigida comprende dos NLS, por ejemplo, las dos NLS están ubicadas en el extremo N-terminal y el extremo C-terminal, respectivamente.

En alguna forma de realización, una modificación de base dirigida es una conversión de cualquier nucleótido C, A, T o G, a cualquier otro nucleótido. Cualquiera de los nucleótidos C, A, T o G puede intercambiarse de forma dirigida al sitio como mediado por un editor de base, o un fragmento catalíticamente activo del mismo, a otro nucleótido. Un complejo de edición de base puede comprender cualquier editor de base, o un dominio de editor de base o un fragmento catalíticamente activo del mismo, que puede convertir un nucleótido de interés en cualquier otro nucleótido de interés de una manera dirigida.

Un dominio de edición base de acuerdo con la presente descripción puede comprender por lo menos una citidina desaminasa, o un fragmento catalíticamente activo de la misma. El por lo menos un complejo de edición de base puede comprender la citidina desaminasa, o un dominio de la misma en forma de un fragmento catalíticamente activo, como editor de base.

(i) Citidina desaminasa.

El término "citidina desaminasa" o "proteína citidina desaminasa", como se usa en el presente documento, se refiere a una proteína, un polipéptido o uno o más dominios funcionales de una proteína o un polipéptido que es capaz de catalizar una reacción de desaminación hidrolítica que convierte un citosina (o un resto de citosina de una molécula) a un uracilo (o un resto de uracilo de una molécula), como se muestra a continuación. En algunas formas de realización, la molécula que contiene citosina es una citidina (C), y la molécula que contiene uracilo es una uridina (U). La molécula que contiene citosina puede ser ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN).

De acuerdo con la presente descripción, las citidina desaminasas que se pueden usar en relación con la presente divulgación incluyen, sin limitación, miembros de la familia de enzimas conocida como familia de desaminasas del complejo de edición de ARNm de apolipoproteína B (APOBEC), una desaminasa inducida por activación (AID), o una citidina desaminasa 1 (CDA1). En formas de realización particulares, la desaminasa es una APOBEC1 desaminasa, una APOBEC2 desaminasa, una APOBEC3A desaminasa, una APOBEC3B desaminasa, una APOBEC3C desaminasa, y APOBEC3D desaminasa, una APOBEC3E desaminasa, una APOBEC3F desaminasa, un APOBEC3G desaminasa, una APOBEC3H desaminasa, o una APOBEC4 desaminasa.

La citidina desaminasa es capaz de dirigir la citosina en una hebra sencilla de ADN. En ciertas formas de realización de ejemplo, una citodina desaminasa puede editarse en una hebra única presente fuera del componente de unión, p. ej., Cas9 y/o Cas13 enlazados. En otras formas de realización de ejemplo, una citodina desaminasa puede editarse en una burbuja localizada, tal como una burbuja localizada formada por una no coincidencia en el sitio de edición objetivo, excepto la secuencia guía.

En algunas formas de realización, la proteína citidina desaminasa reconoce y convierte uno o más residuos de citosina diana en una burbuja monocatenaria de una heterodupla de ADN-ARN, en residuos de uracilo. En algunas formas de realización, la proteína citidina desaminasa reconoce una ventana de unión en la burbuja monocatenaria de una heterodupla de ADN-ARN. En algunas formas de realización, la ventana de unión contiene por lo menos un residuo de citosina objetivo. En algunas formas de realización, la ventana de unión está en el rango de alrededor de 3 pb a alrededor de 100 pb. En algunas formas de realización, la ventana de unión está en el rango de alrededor de 5 pb a alrededor de 50 pb. En algunas formas de realización, la ventana de unión está en el rango de alrededor de 10 pb a alrededor de 30 pb. En algunas formas de realización, la ventana de unión es de alrededor de 1 pb, 2 pb, 3 pb, 5 pb, 7 pb, 10 pb, 15 pb, 20 pb, 25 pb, 30 pb, 40 pb, 45 pb, 50 pb, 55 pb, 60 pb, 65 pb, 70 pb, 75 pb, 80 pb, 85 pb, 90 pb, 95 pb, o 100 pb.

En algunas formas de realización, la proteína citidina desaminasa comprende uno o más dominios de desaminasa. Sin pretender estar limitado por la teoría, se contempla que el dominio de desaminasa funcione para reconocer y convertir uno o más residuos de citosina (C) objetivo contenidos en una burbuja monocatenaria de una heterodupla de ADN-ARN en (un) residuo uracilo (U). En algunas formas de realización, el dominio de desaminasa comprende un centro activo. En algunas formas de realización, el centro activo comprende un ion zinc. En algunas formas de realización, los residuos de aminoácidos en o cerca del centro activo interactúan con uno o más nucleótidos 5' a un residuo citosina diana. En algunas formas de realización, los residuos de aminoácidos en o cerca del centro activo interactúan con uno o más nucleótidos 3' a un residuo citosina diana.

En algunas formas de realización, una citosina desaminasa útil con la invención puede ser una desaminasa de la familia del complejo de edición de ARNm de apolipoproteína B (APOBEC). En algunas formas de realización, la citosina desaminasa puede ser una APOBEC1 desaminasa, una APOBEC2 desaminasa, una APOBEC3A desaminasa, una APOBEC3B desaminasa, una APOBEC3C desaminasa, una APOBEC3D desaminasa, una APOBEC3F desaminasa, una APOBEC3G desaminasa, una APOBEC3H desaminasa, una APOBEC4 desaminasa, una desaminasa

inducida por activación humana (hAID), una rAPOBEC1, FERNY y/o una CDA1, opcionalmente una pmCDA1, una atCDA1 (p. ej., At2g19570) y sus versiones evolucionadas. En algunas formas de realización, la citidina desaminasa comprende la proteína APOBEC1 humana completa (hAPOBEC1) o el dominio de desaminasa de la misma (hAPOBEC1-D) o una versión truncada en extremo C-terminal de ella (hAPOBEC-T). En algunas formas de realización, la citidina desaminasa es un miembro de la familia APOBEC que es homólogo a hAPOBEC1, hAPOBEC-D o hAPOBEC-T. En algunas formas de realización, la citidina desaminasa comprende la proteína completa AID1 humana (hAID) o el dominio de desaminasa de la misma (hAID-D) o una versión truncada en extremo C-terminal (hAID-T). En algunas formas de realización, la citidina desaminasa es un miembro de la familia de AID que es homólogo a hAID, hAID-D o hAID-T. En algunas formas de realización, el hAIDT es un hAID que está truncado en el extremo C-terminal en aproximadamente 20 aminoácidos. En algunas formas de realización, la citosina desaminasa puede ser una APOBEC1 desaminasa. En algunas formas de realización, la citosina desaminasa puede ser una APOBEC3A desaminasa. En algunas formas de realización, la citosina desaminasa puede ser una desaminasa CDA1. En algunas formas de realización, la citosina desaminasa puede ser una FERNY desaminasa.

En algunas formas de realización, la citidina desaminasa comprende la secuencia de aminoácidos de tipo salvaje de una citosina desaminasa. En algunas formas de realización, la citidina desaminasa comprende una o más mutaciones en la secuencia de citosina desaminasa, de modo que la eficacia de edición y/o la preferencia de edición de sustrato de la citosina desaminasa se cambia de acuerdo con las necesidades específicas.

Se han descrito ciertas mutaciones de las proteínas APOBEC 1 y APOBEC3 en Kim et al., *Nature Biotechnology* (2017) 35 (4): 371–377 y Harris et al. *Mol. Cell* (2002) 10: 1247–1253, cada uno de los cuales se incorpora aquí como referencia en su totalidad. Se describen formas de realización adicionales de la citidina desaminasa en los documentos WO2017/070632 y WO2018/213726, cada uno de los cuales se incorpora aquí como referencia en su totalidad.

En algunas formas de realización, por lo menos un efector asociado a CRISPR está unido temporal o permanentemente a por lo menos un editor de bases para formar un complejo de edición de bases dirigido, que es una proteína de fusión de edición de bases, en donde el complejo de edición de bases media por lo menos una modificación de base dirigida. El por lo menos un efector asociado a CRISPR puede estar unido de forma no covalente (temporal) o covalente (permanente) a por lo menos un editor de base. Cualquier componente del por lo menos un editor de base se puede unir temporal o permanentemente al por lo menos un efector asociado a CRISPR.

(ii) adenosina desaminasa.

El término "adenosina desaminasa" o "proteína adenosina desaminasa" como se usa en el presente documento se refiere a una proteína, un polipéptido o uno o más dominios funcionales de una proteína o un polipéptido que es capaz de catalizar una reacción de desaminación hidrolítica que convierte un adenina (o un resto adenina de una molécula) en una hipoxantina (o un resto hipoxantina de una molécula), como se muestra a continuación. En algunas formas de realización, la molécula que contiene adenina es una adenosina (A), y la molécula que contiene hipoxantina es una inosina (I). La molécula que contiene adenina puede ser ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN).

Las adenosina desaminasas (p. ej., adenosina desaminasas diseñadas, adenosina desaminasas evolucionadas) pueden provenir de cualquier organismo, tal como una bacteria. En algunas formas de realización, el dominio desaminasa o la desaminasa es una variante de una desaminasa natural de un organismo. En algunas formas de realización, el dominio desaminasa o la desaminasa no se produce en la naturaleza. Por ejemplo, en algunas formas de realización, el dominio desaminasa o la desaminasa es por lo menos 50%, por lo menos 55%, por lo menos 60%, por lo menos 65%, por lo menos 70%, por lo menos 75% al menos 80%, por lo menos 85 %, por lo menos 90%, por lo menos 95%, por lo menos 96%, por lo menos 97%, por lo menos 98%, por lo menos 99% o por lo menos 99,5% idéntico a una desaminasa natural. En algunas formas de realización, la adenosina desaminasa es de una bacteria, tal como *E. coli*, *S. aureus*, *S. typhi*, *S. putrefaciens*, *H. influenzae* o *C. crescentus*.

En algunas formas de realización, un dominio de adenina desaminasa puede ser un dominio de adenosina desaminasa específico de ARNt de tipo salvaje, por ejemplo, una adenosina desaminasa específica de ARNt (TadA) y/o un dominio de adenosina desaminasa mutado/evolucionado, por ejemplo, dominio de adenosina desaminasa específico de ARNt mutado/evolucionado (TadA*). En algunas formas de realización, un dominio TadA puede ser de *E. coli*. En algunas formas de realización, el TadA puede modificarse, por ejemplo, truncarse, perder uno o más aminoácidos N-terminales y/o C-terminales en relación con un TadA de longitud completa (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19 o 20 residuos de aminoácidos N-terminales y/o C terminales pueden faltar en relación con un TadA de longitud completa). En algunas formas de realización, un polipéptido TadA o dominio TadA no comprende una metionina N-terminal. En algunas formas de realización, la adenosina desaminasa es una desaminasa TadA. En algunas formas de realización, la desaminasa TadA es una desaminasa TadA de *E. coli* (ecTadA). En algunas formas de realización, la TadA desaminasa es una TadA desaminasa truncada de *E. coli*. Por ejemplo, la

ecTadA truncada puede carecer de uno o más aminoácidos N–terminales en relación con una ecTadA de longitud completa. En algunas formas de realización, una adenosina desaminasa es como se describe en Gaudelli et al. (Evolución dirigida de editores de base de adenina con mayor actividad y aplicación terapéutica. Nat Biotechnol (2020) (doi.org/10.1038/s41587–020–0491–6)) y Richter et al. (Evolución asistida por fagos de un editor de base de adenina con compatibilidad y actividad mejoradas del dominio Cas. Nat Biotechnol (2020) (doi.org/10.1038/s41587–020–0453–z)). En algunas formas de realización, la ecTadA truncada puede carecer de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19 o 20 residuos de aminoácidos N–terminales en relación con la ecTadA de longitud completa. En algunas formas de realización, la ecTadA truncada puede carecer de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19 o 20 residuos de aminoácidos C–terminales en relación con la ecTadA de longitud completa. En algunas formas de realización, la desaminasa de ecTadA no comprende una metionina N–terminal. En algunas formas de realización, un polinucleótido que codifica una TadA/TadA* puede ser optimizada por codones para la expresión en una planta.

Algunos aspectos de la divulgación utilizan adenosina desaminasas. En algunas formas de realización, las adenosina desaminasas proporcionadas en el presente documento son capaces de desaminar la adenina. En algunas formas de realización, las adenosina desaminasas proporcionadas en el presente documento son capaces de desaminar adenina en un residuo de ADN de desoxiadenosina. La adenosina desaminasa puede derivarse de cualquier organismo adecuado (por ejemplo, *E. coli*). En algunas formas de realización, la adenina desaminasa es una adenosina desaminasa natural que incluye una o más mutaciones correspondientes a cualquiera de las mutaciones proporcionadas en el presente documento (por ejemplo, mutaciones en ecTadA). Un experto en la técnica podrá identificar el residuo correspondiente en cualquier proteína homóloga y en el ácido nucleico codificador respectivo mediante métodos bien conocidos en la técnica, por ejemplo, mediante alineamiento de secuencia y determinación de residuos homólogos. En consecuencia, un experto en la técnica podría generar mutaciones en cualquier adenosina desaminasa natural (por ejemplo, que tenga homología con ecTadA) que corresponda a cualquiera de las mutaciones descritas en el presente documento, por ejemplo, cualquiera de las mutaciones identificadas en ecTadA. En algunas formas de realización, la adenosina desaminasa es de un procariota. En algunas formas de realización, la adenosina desaminasa es de una bacteria. En algunas formas de realización, la adenosina desaminasa es de *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Shewanella putrefaciens*, *Haemophilus influenzae*, *Caulobacter crescentus*, or *Bacillus subtilis*. En algunas formas de realización, la adenosina desaminasa es de *E. coli*.

En otras formas de realización, las adenosina desaminasas que se pueden usar incluyen, sin limitación, miembros de la familia de enzimas conocidas como adenosina desaminasas que actúan sobre el ARN (ADAR), miembros de la familia de enzimas conocidas como adenosina desaminasas que actúan sobre el ARNt (ADAT) y otros miembros de la familia que contienen el dominio de adenosina desaminasa (ADAD). De acuerdo con la presente divulgación, la adenosina desaminasa es capaz de apuntar a la adenina en una heterodupla de ARN/ADN. De hecho, Zheng et al. (Nucleic Acids Res. 2017, 45 (6): 3369–3377) ha demostrado que las ADAR pueden llevar a cabo reacciones de edición de adenosina a inosina en heteroduplas de ARN/ADN. En formas de realización particulares, la adenosina desaminasa se ha modificado para aumentar su capacidad de editar ADN en una heterodupla de ARN/ADN como se detalla a continuación en el presente documento.

En algunas formas de realización, la adenosina desaminasa se deriva de una o más especies de metazoos, incluidos, entre otros, mamíferos, aves, ranas, calamares, peces, moscas y gusanos. En algunas formas de realización, la adenosina desaminasa es una adenosina desaminasa humana, de calamar o *Drosophila*. En algunas formas de realización, la adenosina desaminasa es un ADAR humano, que incluye hADAR1, hADAR2, hADAR3. En algunas formas de realización, la adenosina desaminasa es una proteína ADAR de *Caenorhabditis elegans*, que incluye ADR-1 y ADR-2. En algunas formas de realización, la adenosina desaminasa es una proteína ADAR de *Drosophila*, que incluye dAdar. En algunas formas de realización, la adenosina desaminasa es una proteína ADAR de calamar *Loligo pealeii*, que incluye sqADAR2a y sqADAR2b. En algunas formas de realización, la adenosina desaminasa es una proteína ADAT humana. En algunas formas de realización, la adenosina desaminasa es una proteína ADAT de *Drosophila*. En algunas formas de realización, la adenosina desaminasa es una proteína ADAD humana, que incluye TE R (hADAD1) y TE RL (hADAD2).

En algunas formas de realización, la adenosina desaminasa es una proteína TadA tal como TadA de *E. coli*. Ver Kim et al., Biochemistry 45: 6407–6416 (2006); Wolf et al., EMBO J. 21: 3841–3851 (2002), cada uno de los cuales se incorpora aquí como referencia en su totalidad. En algunas formas de realización, la adenosina desaminasa es ADA de ratón (Ver Grunebaum et al., Curr. Opin. Allergy Clin. Immunol. 13: 630–638 (2013)) o ADAT2 humano (Ver Fukui et al., J. Nucleic Acids 2010 : 260512 (2010)), cada uno de los cuales se incorpora aquí como referencia en su totalidad. Formas de realización adicionales de la adenosina desaminasa se describen en la Patente de los Estados Unidos Núm. US10.113.163 y WO2018/213708, cada una de las cuales se incorpora aquí como referencia en su totalidad.

En algunas formas de realización, por lo menos un efector asociado a CRISPR–Cas está unido temporal o permanentemente a por lo menos un editor de bases para formar un complejo de edición de bases dirigido, que es una proteína de fusión de edición de bases, en donde el complejo de edición de bases media por lo menos una modificación de base dirigida. El por lo menos un efector asociado a CRISPR puede estar unido de forma no covalente (temporal) o covalente (permanente) a por lo menos un editor de base. Cualquier componente del por lo menos un editor de base se puede unir temporal o permanentemente al por lo menos un efector asociado a CRISPR–Cas.

En un aspecto, la presente descripción proporciona métodos para la desaminación dirigida de adenina en un ADN, más particularmente en un locus de interés. La divulgación enseña que la proteína adenosina desaminasa (AD) es reclutada específicamente en la adenina pertinente en el locus de interés por un complejo Cas CRISPR que puede unirse específicamente a una secuencia diana. Para lograr esto, la proteína adenosina desaminasa puede estar unida covalentemente a la enzima CRISPR–Cas o puede proporcionarse como una proteína separada, pero adaptada para asegurar su reclutamiento de la misma en el complejo CRISPR–Cas.

En algunas formas de realización, el reclutamiento de la adenosina desaminasa en el locus diana se asegura fusionando la adenosina desaminasa o el dominio catalítico de la misma a la proteína CRISPR–Cas, que es una proteína Cas o Cpf1. Los métodos para generar una proteína de fusión a partir de dos proteínas separadas se conocen en la técnica y típicamente implican el uso de espaciadores o enlazadores. La proteína CRISPR–Cas puede fusionarse con la proteína adenosina desaminasa o su dominio catalítico en su extremo N– o C–terminal. En formas de realización particulares, la proteína CRISPR–Cas es una proteína Cas o Cpf1 y está unida al extremo N–terminal de la proteína desaminasa o su dominio catalítico.

4. Enlazador.

En la presente descripción, el enlazador se refiere a una molécula que se une a las proteínas para formar una proteína de fusión. Generalmente, tales moléculas no tienen actividad biológica específica más que unir o preservar alguna distancia mínima u otra relación espacial entre las proteínas. Sin embargo, en algunas formas de realización, el enlazador puede seleccionarse para influir en alguna propiedad del enlazador y/o la proteína de fusión tal como el plegamiento, la carga neta o la hidrofobicidad del enlazador.

Los enlazadores adecuados para usar en los métodos de la presente descripción son bien conocidos por los expertos en la técnica e incluyen, sin limitación, enlazadores de carbono de cadena lineal o ramificada, enlazadores de carbono heterocíclicos o enlazadores de péptidos. Sin embargo,

como se usa en el presente documento, el enlazador también puede ser un enlace covalente (enlace carbono-carbono o enlace carbono-heteroátomo). En formas de realización particulares, el enlazador se usa para separar la proteína CRISPR-Cas y la citidina desaminasa en una distancia suficiente para asegurar que cada proteína conserve su propiedad funcional requerida. Las secuencias enlazadoras de péptidos preferidas adoptan una conformación extendida flexible y no exhiben una propensión a desarrollar una estructura secundaria ordenada.

En algunas formas de realización, el enlazador puede ser un resto químico que puede ser monomérico, dimérico, multimérico o polimérico. Ejemplos de enlazadores se describen en las referencias de Maratea et al. (1985), Gene 40: 39-46; Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62; Pat. U. S. Nro. 4.935.233; y Pat. U. S. Nro. 4.751.180. Por ejemplo, pueden usarse los enlazadores GlySer GGS, GGGS o GSG. Los enlazadores GGS, GSG, GGGS o GGGGS pueden usarse en repeticiones de 3, tal como (GGS)₃. En otras formas de realización, se usan 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o más de (GGGS), para proporcionar longitudes adecuadas. En algunas formas de realización, pueden usarse enlazadores tales como (GGGS)₁, (GGGS)₂, (GGGS)₃, (GGGS)₄, (GGGS)₅, (GGGS)₆, (GGGS)₇, (GGGS)₈, (GGGS)₉, (GGGS)₁₀, (GGGS)₁₁, o (GGGS)₁₂. En otras formas de realización, el enlazador es el enlazador XTEN. En formas de realización particulares, la proteína efectora asociada a CRISPR inactivada por nucleasa tal como dCas9 o nCas9 se une a la proteína desaminasa o su dominio catalítico por medio de un enlazador XTEN. En algunas formas de realización, el mutante Cas inactivado por nucleasa se une C-terminalmente al extremo N-terminal de una proteína desaminasa o su dominio catalítico por medio de un enlazador XTEN. Además, los NLS N y C terminales también pueden funcionar como enlazadores.

5. ARNg.

La presente divulgación proporciona ARN guía (ARNg) que dirigen un polipéptido modificador dirigido al sitio a una secuencia de ácido nucleico diana específica. Un ARNg comprende un segmento dirigido al ácido nucleico y un segmento de unión a proteínas. El segmento dirigido al ácido nucleico de un ARNg comprende una secuencia de nucleótidos que es complementaria a una secuencia en la secuencia de ácido nucleico diana. De este modo, el segmento dirigido a ácido nucleico de un ARNg interactúa con un ácido nucleico objetivo de una manera específica de secuencia mediante hibridación (es decir, emparejamiento de bases), y la secuencia de nucleótidos del segmento dirigido a ácido nucleico determina la ubicación dentro del ácido nucleico objetivo al que se unirá el ARNg. El segmento dirigido al ácido nucleico de un ARNg puede modificarse (por ejemplo, mediante

ingeniería genética) para hibridarse con cualquier secuencia deseada dentro de una secuencia de ácido nucleico diana. El segmento de unión a proteína de un ARN guía interactúa con un polipéptido modificador dirigido al sitio (por ejemplo, una proteína Cas) para formar un complejo. El ARN guía guía el polipéptido unido a una secuencia de nucleótidos específica dentro del ácido nucleico diana a través del segmento de direccionamiento de ácido nucleico descrito anteriormente. El segmento de unión de proteína de un ARN guía comprende dos tramos de nucleótidos que son complementarios entre sí y que forman una dupla de ARN bicatenario.

En algunas formas de realización, un ARNg comprende dos moléculas de ARN separadas. En tales formas de realización, cada una de las dos moléculas de ARN comprende un tramo de nucleótidos que son complementarios entre sí, de modo que los nucleótidos complementarios de las dos moléculas de ARN hibridan para formar la dupla de ARN bicatenario del segmento de unión a proteína. En algunas formas de realización, un ARNg comprende una única molécula de ARN guía (ARNg).

La especificidad de un ARNg para un loci diana está mediada por la secuencia del segmento de unión a ácido nucleico (por ejemplo, el espaciador), que comprende alrededor de 20 nucleótidos que son complementarios a una secuencia de ácido nucleico diana dentro del locus diana. En algunas formas de realización, la secuencia de ácido nucleico diana correspondiente tiene aproximadamente 20 nucleótidos de longitud, opcionalmente aproximadamente 20, 21, 22, 23, 24 o 25 nucleótidos de longitud. En algunas formas de realización, los segmentos de unión a ácido nucleico de las secuencias de ARNg de la presente descripción son por lo menos 90% complementarios a una secuencia de ácido nucleico diana dentro de un locus diana. En algunas formas de realización, los segmentos de unión a ácido nucleico de las secuencias de ARNg de la presente descripción son por lo menos 95%, 96%, 97%, 98% o 99% complementarios a una secuencia de ácido nucleico diana dentro de un locus diana. En algunas formas de realización, los segmentos de unión a ácido nucleico de las secuencias de ARNg de la presente descripción son 100% complementarios a una secuencia de ácido nucleico diana dentro de un locus diana.

En algunas formas de realización, la secuencia de ácido nucleico diana es una secuencia diana de ARN. En algunas formas de realización, la secuencia de ácido nucleico diana es una secuencia diana de ADN.

En algunas formas de realización, un sistema de edición dirigido (por ejemplo, sistema de edición de base) comprende una o dos o más moléculas de ARNg, cada una de las cuales comprende un segmento de unión al ADN, en donde por lo menos uno de los segmentos de unión al ácido nucleico

se une a una secuencia de ADN objetivo de un gen objetivo de un clado AGAMOUS que comprende AG, SHP1, SHP2 y STK en plantas de la presente descripción. En algunas formas de realización, el ARN guía es un ARN guía único (ARNsg). Los métodos de construcción de ARNsg adecuados de acuerdo con una secuencia diana dada son conocidos en la técnica. Ver, por ejemplo, Wang, Y. et al. La edición simultánea de tres homóalelos en trigo de grano entero hexaploide confiere resistencia hereditaria al mildiu polvoriento. *Nat. Biotecnología* 32, 947–951 (2014); Shan, Q. et al. Modificación dirigida del genoma de plantas de cultivo utilizando un sistema CRISPR–Cas. *Nat. Biotechnol.* 31, 686–688 (2013); Liang, Z. et al. Mutagénesis dirigida en *Zea mays* utilizando TALEN y el sistema CRISPR/Cas. *J Genet Genomics.* 41, 63–68 (2014).

6. Inhibidor de reparación de escisión de bases.

La adición de un dominio de uracilo ADN glicosilasa (UGI) aumentó aún más la eficiencia de edición de bases. En algunas formas de realización, el sistema de edición de bases dirigido comprende además un inhibidor de reparación de escisión de bases (BER, conforme a sus siglas en inglés). La respuesta de reparación de ADN celular a la presencia de un emparejamiento U:G en el ADN puede ser responsable de una disminución en la eficiencia de edición de nucleobase en las células vegetales. La uracilo ADN glicosilasa cataliza la eliminación de uracilo del ADN en las células vegetales, lo que puede iniciar la reparación de la escisión de la base, de modo que el par U:G se invierte a C:G. En algunas formas de realización, el inhibidor de BER es un inhibidor de uracilo glicosilasa o un dominio activo del mismo.

En algunas formas de realización, el inhibidor de BER es un inhibidor de uracilo ADN glicosilasa (UDG). En algunas formas de realización, el inhibidor de BER es un inhibidor de UDG. En algunas formas de realización, el inhibidor de BER es un inhibidor de polipéptido. En algunas formas de realización, el inhibidor de BER es una proteína que se une al ADN monocatenario. Por ejemplo, el inhibidor de BER puede ser una proteína de unión monocatenaria de *Erwinia tasmaniensis*. En algunas formas de realización, el inhibidor de BER es una proteína que se une al uracilo. En algunas formas de realización, el inhibidor de BER es una proteína que se une al uracilo en el ADN. En algunas formas de realización, el inhibidor de BER es un UDG catalíticamente inactivo o un dominio de unión del mismo. En algunas formas de realización, el inhibidor de BER es un UDG catalíticamente inactivo o un dominio de unión del mismo que no elimina el uracilo del ADN. Otras proteínas que son capaces de inhibir (por ejemplo, bloquear estéricamente) UDG están dentro del alcance de esta descripción. Además, cualquier proteína que bloquee o inhiba la reparación de escisión de bases también está dentro del alcance de esta descripción.

La reparación de la escisión de bases puede ser inhibida por moléculas que se unen a la hebra editada, bloquean la base editada, inhiben la uracilo ADN glicosilasa, inhiben la reparación de la escisión de la base, protegen la base editada y/o promueven la fijación de la hebra no dirigida. Por consiguiente, el uso del inhibidor de BER descrito en este documento puede aumentar la eficacia de edición de una citidina desaminasa que es capaz de catalizar un cambio de C a U.

En algunas formas de realización, el inhibidor de uracilo glicosilasa (UGI) es el inhibidor de uracilo ADN glicosilasa del bacteriófago PBS1 de *Bacillus subtilis* o un fragmento activo del mismo, tal como una proteína de 83 residuos del bacteriófago PBS1 de *Bacillus subtilis*.

La proteína UGI adecuada y las secuencias de nucleótidos se proporcionan en el presente documento, y las secuencias UGI adecuadas adicionales son conocidas por los expertos en la técnica, e incluyen, por ejemplo, las publicadas en Wang et al., El gen inhibidor de la uracilo ADN glicosilasa del bacteriófago PBS2 codifica una proteína de unión específica para uracilo-ADN glicosilasa. *J. Biol. Chem* 264: 1 163–1 171 (1989); Lundquist et al., Mutagénesis dirigida al sitio y caracterización de la proteína inhibidora de la uracilo ADN glicosilasa. Papel de aminoácidos carboxílicos específicos en la formación de complejos con uracilo ADN glicosilasa de *Escherichia coli*. *J. Biol. Chem* 272: 21408–21419 (1997); Ravishankar et al., Análisis de rayos X de un complejo de uracilo ADN glicosilasa de *Escherichia coli* (EcUDG) con un inhibidor proteínico. Elucidación de la estructura de un UDG procariota. *Nucleic Acids Res.* 26: 4880–4887 (1998); y Putnam et al., Mimetismo proteico del ADN de las estructuras cristalinas de la proteína inhibidora de la uracilo ADN glicosilasa y su complejo con la uracilo ADN glicosilasa de *Escherichia coli*. *J. Mol. Biol.* 287: 331–346 (1999), cada uno de los cuales se incorpora aquí como referencia. Se describen formas de realización adicionales del inhibidor de uracilo glicosilasa (UGI) en los documentos WO2018/086623, WO2018/205995, WO2017/70632, y WO2018/213726, cada uno de los cuales se incorpora aquí como referencia en su totalidad.

En algunas formas de realización, el dominio UGI comprende un UGI de tipo salvaje o un UGI. En algunas formas de realización, las proteínas UGI proporcionadas en el presente documento incluyen fragmentos de UGI y proteínas homólogas a un fragmento UGI o UGI.

Las proteínas adicionales pueden ser inhibidores de uracilo glicosilasa. Por ejemplo, otras proteínas que son capaces de inhibir (por ejemplo, bloquear estéricamente) una enzima de reparación de la escisión de base de uracilo-ADN glicosilasa están dentro del alcance de esta descripción. Además, cualquier proteína que bloquee o inhiba la reparación de escisión de bases también está dentro del alcance de esta descripción. En algunas formas de realización, se usa una proteína que se

une al ADN. En otra forma de realización, se usa un sustituto de UGI. En algunas formas de realización, un inhibidor de uracilo glicosilasa es una proteína que se une al ADN monocatenario. Por ejemplo, un inhibidor de la uracilo glicosilasa puede ser una proteína de unión monocatenaria de *Erwinia tasmaniensis*. En algunas formas de realización, un inhibidor de la glicosilasa de uracilo es una proteína que se une al uracilo. En algunas formas de realización, un inhibidor de la glicosilasa de uracilo es una proteína que se une al uracilo en el ADN. En algunas formas de realización, un inhibidor de uracilo glicosilasa es una proteína de uracilo ADN–glicosilasa catalíticamente inactiva. En algunas formas de realización, un inhibidor de la glicosilasa de uracilo es una proteína de uracilo ADN glicosilasa catalíticamente inactiva que no elimina el uracilo del ADN. Como otro ejemplo, un inhibidor de uracilo glicosilasa es una UDG catalíticamente inactivo.

En algunas formas de realización, el sistema de edición base comprende los siguientes dominios; i) la proteína CRISPR–Cas (dCas9 o nCas9) y ii) la citidina desaminasa, que puede fusionarse o unirse a un inhibidor de BER (p. ej., un inhibidor de uracilo ADN glicosilasa).

7. Uracilo ADN glicosilasas para la diversificación.

La uracilo ADN glicosilasa (UDG) es una enzima que revierte las mutaciones en el ADN. La mutación más común es la desaminación de la citosina a uracilo. UDG repara estas mutaciones, y es crucial en la reparación del ADN. Se presentan diversas uracilo ADN glicosilasas y ADN glicosilasas relacionadas (CE), como la uracilo–ADN glicosilasa, la uracilo ADN glicosilasa termofílica, la ADN glicosilasa específica de no coincidencia G:T/U (Mug) y la uracilo ADN glicosilasa monofuncional selectiva monocatenaria (SMUG1).

Las uracilo ADN glicosilasas eliminan el uracilo del ADN, lo que puede surgir o bien por desaminación espontánea de citosina o por la incorporación incorrecta de dU opuesto a dA durante la replicación de ADN. El miembro prototípico de esta familia es UDG de *E. coli*, que fue una de las primeras glicosilasas descubiertas. Se han identificado cuatro actividades diferentes de uracilo–ADN glicosilasa en células de mamíferos, incluidas UNG, SMUG1, TDG y MBD4, que varían en especificidad de sustrato y localización subcelular. SMUG1 prefiere el ADN monocatenario como sustrato, pero también elimina U del ADN bicatenario. Además del uracilo no modificado, SMUG1 puede escindir 5–hidroxiuracilo, 5–hidroximetiluracilo y 5–formiluracilo que portan un grupo oxidado en el anillo C5.[13] TDG y MBD4 son estrictamente específicas para el ADN bicatenario. TDG puede eliminar el timina glicol cuando está presente opuesto a la guanina, así como los derivados de U con modificaciones en el carbono 5.

TDG y SMUG1 son las principales enzimas responsables de la reparación de los pares incorrectos U:G causados por la desaminación de citosina espontánea, mientras que uracilo que surge en ADN a través de la incorporación errónea de dU es principalmente tratado por UNG. Se cree que MBD4 corrige las no coincidencias o desajustes T:G que surgen de la desaminación de 5-metilcitosina a timina en sitios CpG.

El uracilo que surge en el ADN, o bien por incorporación incorrecta de dUMP o por desaminación de citosina, se elimina activamente a través de la vía de reparación de escisión de bases (BER) de múltiples pasos. La BER de uracilo se inicia por una actividad de uracilo ADN glicosilasa (UDG) que escinde el enlace N-glicosídico y elimina el uracilo como base libre, de manera de generar un sitio abásico (apurínico/apirimidínico, AP) en el ADN. La reparación se completa a través de pasos posteriores que incluyen incisión en el sitio AP, adaptación de espacios, síntesis de reparación y ligadura. En algunas formas de realización, la adición de una uracilo ADN glicosilasa (UDG) tal como la uracilo N-glicosilasa (UNG) puede inducir diversas mutaciones en la base objetivo. En algunas formas de realización, el sistema de edición de bases dirigido comprende además una uracilo ADN glicosilasa (UDG). La respuesta de reparación de ADN celular a la presencia de un emparejamiento U:G en ADN puede ser responsable de una disminución en la eficiencia de edición de nucleobases en células vegetales. La glicosilasa de ADN de uracilo cataliza la eliminación de uracilo del ADN en las células vegetales, lo que puede iniciar la reparación de la escisión de la base, de modo que el par U:G se invierta en C:G. En otras formas de realización, la eliminación de uracilo del ADN en las células vegetales no siempre se invierte a C para el emparejamiento C:G, sino que se aleatoriza a otras bases como T, A y G.

En algunas formas de realización, una uracilo ADN glicosilasa (UDG) se fusiona con el sistema de edición de bases dirigido enseñado en la presente descripción para introducir una sustitución de nucleótido único estable y dirigida, pero aleatorizada, en el gen del factor de transcripción de la caja MADS.

El uso de la UDG descrita en el presente documento puede aumentar la aleatorización de bases en un nucleótido único dirigido de un gen diana tal como un miembro de la familia de genes de caja MADS que se enseña en la presente descripción.

En algunas formas de realización, se proporciona una UDG en cis. En algunas formas de realización, se proporciona una UDG en trans. En algunas formas de realización, una UDG se fusiona a un editor de base (o un sistema de edición de base) descrito en la presente divulgación. En otras formas de realización, una UDG desencadena una replicación de ADN de casilla para la aleatorización

de bases. En otras formas de realización, una UDG desencadena una reparación de ADN a través de la replicación de ADN, lo que incluye la aleatorización de bases. En formas de realización adicionales, las variantes UDG de origen natural pueden usarse como un dominio UDG. En formas de realización adicionales, las variantes UDG que no se producen naturalmente pueden usarse como un dominio UDG. En formas de realización adicionales, una UDG puede diseñarse genéticamente para mejorar una actividad UDG funcional.

8. Secuencias de localización nuclear para direccionamiento al núcleo.

Una señal de localización nuclear (NLS, conforme a sus siglas en inglés), o cualquier otra señal de direccionamiento de orgánulos, puede ser necesaria adicionalmente para garantizar el direccionamiento adecuado del complejo. La presente descripción se refiere a la modificación de una citosina en un locus objetivo de interés, por lo que el locus o lugar objetivo está dentro de una célula vegetal. Para mejorar el direccionamiento de la proteína CRISPR–Cas y/o la proteína citidina desaminasa o su dominio catalítico utilizada en los métodos de la presente divulgación al núcleo, puede ser ventajoso proporcionar uno o ambos de estos componentes con una o más secuencias de localización nuclear (NLS).

En algunas formas de realización, los NLS pueden ser heterólogos a las proteínas. En general, NLS consiste en una o más secuencias cortas de lisina o arginina con carga positiva expuestas en la superficie de una proteína, pero también se conocen otros tipos de NLS en la técnica.

En algunas formas de realización, el extremo N–terminal de la proteína de fusión de edición de bases comprende una NLS con una secuencia de aminoácidos. En algunas formas de realización, el extremo C–terminal de la proteína de fusión de edición de bases comprende una NLS.

Una proteína de fusión de edición de base también puede incluir otras secuencias de localización, tales como secuencias de localización citoplasmática, secuencias de localización de cloroplastos, secuencias de localización mitocondrial y similares, de acuerdo con la ubicación del ADN por editar. Con el fin de obtener una expresión eficiente en plantas, en algunas formas de realización, la secuencia de nucleótidos que codifica la proteína de fusión de edición de bases está optimizada por codones para que la planta pueda editarse en bases.

La optimización de codones se refiere a un proceso de modificación de una secuencia de ácido nucleico para mejorar la expresión en las células huéspedes de interés mediante el reemplazo de por lo menos un codón (por ejemplo, alrededor de o más de alrededor de 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 o más codones) de la secuencia nativa con codones que se usan con más o mayor frecuencia en los genes de esa célula huésped, mientras se mantiene la secuencia de aminoácidos nativa. Diversas

especies exhiben un sesgo particular para ciertos codones de un aminoácido particular. El sesgo de codón (diferencias en el uso de codón entre organismos) a menudo se correlaciona con la eficiencia de la traducción del ARN mensajero (ARNm), que a su vez se cree que depende, entre otras cosas, de las propiedades de los codones que se están traduciendo y la disponibilidad de las moléculas de ARN de transferencia particulares (ARNt). El predominio de ARNt seleccionados en una célula es generalmente un reflejo de los codones utilizados con mayor frecuencia en la síntesis de péptidos. En consecuencia, los genes se pueden adaptar para una expresión génica óptima en un organismo determinado sobre la base de la optimización de codones. Las tablas de uso de codones están fácilmente disponibles, por ejemplo, en la "Base de datos de uso de codones" disponible en www.kazusa.or.jp/codon/, y estas tablas se pueden adaptar de varias maneras. Ver Nakamura, Y., et al. "Uso de codones tabulado de las bases de datos internacionales de secuencias de ADN: estado para el año 2000" Nucl. Acids Res. 28: 292 (2000). En algunas formas de realización, la secuencia de nucleótidos con optimización de codones que codifica la proteína de fusión de edición de bases se proporciona en este documento. En algunas formas de realización, el ARN guía es un ARN guía único (ARNsg). Los métodos de construcción de ARNsg adecuados de acuerdo con una secuencia diana dada son conocidos en la técnica. Ver, por ejemplo, Wang, Y. et al. La edición simultánea de tres homóalelos en trigo de grano entero hexaploide confiere resistencia hereditaria al mildiú polvoriento. Nat. Biotechnol. 32, 947-951 (2014); Shan, Q. et al. Modificación dirigida del genoma de plantas de cultivo utilizando un sistema CRISPR-Cas. Nat. Biotechnol. 31, 686-688 (2013); Liang, Z. et al. Mutagénesis dirigida en *Zea mays* utilizando TALEN y el sistema CRISPR/Cas. J Genet Genomics. 41, 63-68 (2014).

9. Promotor.

Con el fin de garantizar una expresión adecuada en una célula vegetal, los componentes del sistema de edición de bases dirigido descrito en el presente documento se colocan típicamente bajo el control de un promotor de planta, es decir, un promotor operable en células vegetales. Se prevé el uso de diferentes tipos de promotores.

Un promotor constitutivo de plantas es un promotor que puede expresar el marco de lectura abierto (ORF, conforme a sus siglas en inglés) que controla en todos o casi todos los tejidos vegetales durante todas o casi todas las etapas de desarrollo de la planta (denominado "expresión constitutiva"). Un ejemplo no limitante de un promotor constitutivo es el promotor 35S del virus del mosaico de la coliflor. "Promotor regulado" se refiere a promotores que dirigen la expresión génica no constitutivamente, sino de una manera regulada temporal y/o espacialmente, e incluye promotores

específicos de tejido, preferidos de tejido e inducibles. Diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos de células, o en diferentes etapas de desarrollo, o en respuesta a diferentes condiciones ambientales. En formas de realización particulares, uno o más de los componentes del sistema de edición de bases dirigido descrito aquí se expresan bajo el control de un promotor constitutivo, tal como el promotor 35S del virus del mosaico de la coliflor. Los promotores preferidos de tejido se pueden utilizar para dirigir la expresión mejorada en ciertos tipos de células dentro de un tejido vegetal particular, por ejemplo, células vasculares en hojas o raíces o en células específicas de la semilla. Se pueden encontrar ejemplos de promotores particulares para el uso de la presente divulgación en las referencias de Kawamata et al., (1997) *Plant Cell Physiol* 38: 792–803; Yamamoto et al., (1997) *Plant J* 12: 25565; Hire et al., (1992) *Plant Mol Biol* 20: 207–18, Kuster et al., (1995) *Plant Mol Biol* 29: 759–72, y Capana et al. (1994) *Plant Mol Biol* 25: 681–91.

Los promotores inducibles pueden ser de interés para expresar uno o más de los componentes del sistema de edición de bases dirigido descrito aquí en circunstancias limitadas para evitar la actividad no específica de la desaminasa. En formas de realización particulares, uno o más elementos del sistema de edición de bases dirigido descrito aquí se expresan bajo el control de un promotor inducible. Los ejemplos de promotores que son inducibles y que permiten el control espacio-temporal de la edición génica o la expresión génica pueden usar una forma de energía. La forma de energía puede incluir, entre otros, energía sonora, radiación electromagnética, energía química y/o energía térmica. Los ejemplos de sistemas inducibles incluyen promotores inducibles por tetraciclina (Tet-On o Tet-Off), sistemas de activación de transcripción de dos híbridos de molécula pequeña (FKBP, ABA, etc.) o sistemas inducibles por luz (Fitocromo, dominios de LOV o criptocromo), tal como un efector transcripcional inducible por luz (LITE, conforme a sus siglas en inglés) que dirige los cambios en la actividad transcripcional de una manera específica de secuencia. Los componentes de un sistema inducible por luz pueden incluir una proteína de fusión del sistema de edición de bases dirigido y un heterodímero de citocromo sensible a la luz (por ejemplo, de *Arabidopsis thaliana*). Se proporcionan ejemplos adicionales de proteínas de unión a ADN inducibles y métodos para su uso en los documentos US 61/736465 y US 61/721.283, que se incorporan aquí como referencia en su totalidad.

En algunas formas de realización, la expresión transitoria o inducible se puede lograr usando, por ejemplo, promotores regulados por productos químicos, es decir, mediante lo cual la aplicación de un producto químico exógeno induce la expresión génica. La modulación de la expresión génica también puede obtenerse mediante un promotor químico reprimible, donde la aplicación del químico

reprime la expresión génica. Los promotores inducibles químicamente incluyen, sin limitación, el promotor In2-2 de maíz, activado por aseguradores de herbicidas de bencenosulfonamida (De Veylder et al., (1997) *Plant Cell Physiol* 38: 568-77), el promotor GST de maíz (GST- 11-27, WO93/01294), activado por compuestos electrofílicos hidrófobos utilizados como herbicidas preemergentes, y el promotor de tabaco PR-1 a (Ono et al., (2004) *Biosci Biotechnol Biochem* 68: 803-7) activado por ácido salicílico. Los promotores que están regulados por antibióticos, tales como los promotores inducibles por tetraciclina y los represibles por tetraciclina (Gatz et al., (1991) *Mol Gen Genet* 227: 229-37; las Patentes de EE. UU. Números 5.814.618 y 5.789.156) también se pueden usar en este documento.

En algunas formas de realización, la secuencia de nucleótidos que codifica la proteína de fusión editada con bases y/o la secuencia de nucleótidos que codifica el ARN guía está operativamente unida a un elemento regulador de la expresión de plantas, tal como un promotor. Los ejemplos de promotores que pueden usarse en la presente descripción incluyen, sin limitación, el promotor 35S del virus del mosaico de la coliflor (Odell et al. (1985) *Nature* 313: 810-812), un promotor Ubi-1 de maíz, un promotor U6 de trigo, un promotor de arroz U3, un promotor de maíz U3, un promotor actina de arroz, un promotor TrpPro5 (Solicitud de Pat. U S. Nro. 10/377.318; presentada el 16 de marzo de 2005), un promotor pEMU (Last et al. *Theor. Appl. Genet.* 8 1: 581588), un promotor MAS (Velten et al. (1984) *EMBO J.* 3 : 2723-2730), un promotor H3 histona de maíz (Lepetit et al. *Mol. Gen. Genet.* 231: 276-285 y Atanassova et al. (1992) *Plant J.* 2 (3): 291300), y un promotor de *Brassica napus* ALS3 (Solicitud PCT WO 97/41228). Los promotores que pueden usarse en la presente divulgación además incluyen los comúnmente usados promotores específicos de tejido como se revisa en la referencia de Moore et al. (2006) *Plant J.* 45 (4): 651-683.

Generación de una planta con edición génica con mejorados rasgos agronómicos.

La presente divulgación proporciona métodos para la edición dirigida en una célula, tejido, órgano vegetal o planta. En un aspecto, la presente descripción proporciona métodos para producir una planta editada genéticamente, que comprende introducir un sistema para realizar la edición de bases a una secuencia objetivo en el genoma de una planta, y de ese modo, dicha proteína de fusión de edición de bases es dirigida a la secuencia objetivo en dicho genoma de planta por el ARN guía, y produce una o más sustituciones de nucleótidos en dicha secuencia diana.

En algunas formas de realización, el sistema de edición de bases dirigido como se describe en el presente documento se usa para introducir mutaciones dirigidas (por ejemplo, C> T o G> A), que

pueden causar una mutación sin sentido (por ejemplo, codón de parada prematuro) o una mutación de sentido erróneo (por ejemplo, que codifica diferente residuo de aminoácido). Esto es de interés cuando las mutaciones de un solo nucleótido en ciertos genes endógenos pueden conferir o contribuir a un rasgo deseado, tal como falta de semillas, disminución de la simiente, tamaño reducido de las semillas, tamaño reducido del endocarpio y/o lignificación reducida en el tejido del endocarpio. En algunas formas de realización, el gen objetivo es un gen de caja MADS de un clado AGAMOUS.

Los métodos descritos en el presente documento logran la generación de plantas editadas genéticamente que tienen uno o más rasgos deseables en comparación con la planta de tipo silvestre.

En algunas formas de realización, se obtienen plantas, partes de plantas o células no transgénicas pero editadas genéticamente, ya que no se incorpora una secuencia de ADN exógeno en el genoma de ninguna de las células de la planta. En tales formas de realización, las plantas editadas genéticamente no son transgénicas. Cuando solo se garantiza la modificación de un gen endógeno y no se introducen ni mantienen genes extraños en el genoma de la planta, los cultivos genéticamente modificados resultantes no contienen genes extraños y, por lo tanto, pueden considerarse básicamente no transgénicos.

En algunas formas de realización, la modificación de la secuencia diana se puede lograr simplemente introduciendo o produciendo la proteína de fusión de edición de base y el ARN guía en células vegetales, y la modificación se puede heredar de manera estable sin la necesidad de una transformación estable de plantas con el sistema de edición de base.

En otras formas de realización, los polinucleótidos se suministran a la célula mediante un virus de ADN (por ejemplo, un geminivirus) o un virus de ARN (por ejemplo, un tobnavirus). En otras formas de realización, los pasos de introducción incluyen suministrar a la célula de la planta un ADN-T que contiene una o más secuencias de polinucleótidos que codifican la proteína CRISPR-Cas, la citidina desaminasa y el ARN guía, donde el suministro se realiza a través de *Agrobacterium*. La secuencia polinucleotídica que codifica los componentes del sistema CRISPR/Cas acoplado a citidina desaminasa puede estar operativamente unida a un promotor, tal como un promotor constitutivo (por ejemplo, un promotor 35S del virus del mosaico de la coliflor), o un promotor inducible o específico de célula en una célula vegetal de interés descrita aquí. En otras formas de realización, el polinucleótido se introduce mediante bombardeo con microproyectiles.

En algunas formas de realización, el sistema de edición de bases puede introducirse en plantas mediante cualquier método conocido en la técnica o desarrollado posteriormente, incluidos, sin limitación, bombardeo de partículas, transformación de protoplastos mediada por PEG,

transformación mediada por *Agrobacterium*, transformación mediada por virus vegetal, inyección de ovario y tubo de polen.

En algunas formas de realización, la introducción de un ácido nucleico puede realizarse en ausencia de una presión selectiva, de manera de evitar así la integración de secuencias de nucleótidos exógenas en el genoma de la planta. En algunas formas de realización, la introducción comprende transformar el sistema de edición de bases en células o tejidos vegetales aislados, y luego regenerar las células o los tejidos vegetales transformados como una planta intacta. Preferentemente, la regeneración se realiza en ausencia de una presión selectiva, es decir, no se usa ningún agente selectivo contra el gen selectivo portado en el vector de expresión durante el cultivo de tejidos. Sin el uso de un agente selectivo, la eficiencia de regeneración de la planta se puede aumentar para obtener una planta modificada que no contenga secuencias de nucleótidos exógenas.

Un casete de expresión de la invención también puede incluir un polinucleótido que codifica un marcador seleccionable, que puede usarse para seleccionar una célula huésped transformada. Como se usa en el presente documento, "marcador seleccionable" significa una secuencia de polinucleótidos que cuando se expresa imparte un fenotipo distinto a la célula huésped que expresa el marcador y, por lo tanto, permite que dichas células transformadas se distingan de las que no tienen el marcador. Dicha secuencia de polinucleótidos puede codificar un marcador seleccionable o detectable, según si el marcador confiere un rasgo que puede seleccionarse por medios químicos, tal como mediante el uso de un agente selectivo (por ejemplo, un antibiótico y similares), o si el marcador es simplemente un rasgo que se puede identificar a través de la observación o de pruebas, tal como mediante cribado (por ejemplo, fluorescencia). Muchos ejemplos de marcadores seleccionables adecuados son conocidos en la técnica y pueden usarse en los casetes de expresión descritos en este documento.

En otras formas de realización, el sistema de edición base de la presente descripción se puede transformar en un sitio particular en una planta intacta, tal como hoja, punta de brote, tubo de polen, mazorca joven o hipocotilo. Esto es particularmente adecuado para la transformación de plantas que son difíciles de regenerar mediante cultivo de tejidos. En algunas formas de realización, las proteínas expresadas *in vitro* y/o las moléculas de ARN transcriptas *in vitro* se transforman directamente en la planta. Las proteínas y/o moléculas de ARN son capaces de lograr la edición de bases en las células vegetales, y posteriormente son degradadas por las células para evitar la integración de secuencias de nucleótidos exógenas en el genoma de la planta. Las plantas que pueden editarse con bases mediante los métodos incluyen monocotiledóneas y dicotiledóneas. Por ejemplo, la planta puede ser una planta

de cultivo tal como trigo, arroz, maíz, soja, girasol, sorgo, colza, alfalfa, algodón, cebada, mijo, caña de azúcar, tomate, tabaco, mandioca o papa. En otro ejemplo, la planta puede ser un cultivo frutal, tal como tomate, almendra, dátiles, níspero, manzana, durazno, pera, ciruela, frambuesa, frambuesa negra, mora, cereza, palta, frutilla, frutilla silvestre, uva y naranja.

En algunas formas de realización, la secuencia objetivo se asocia con rasgos de plantas tales como rasgos agronómicos, y en consecuencia, la edición de bases produce una planta que tiene alterados rasgos en relación con una planta de tipo silvestre. En la presente divulgación, la secuencia objetivo por ser modificada puede ubicarse en cualquier lugar en el genoma, por ejemplo, dentro de un gen funcional tal como un gen codificador de proteína, o por ejemplo, puede ubicarse en una región reguladora de la expresión génica, tal como una región de promotor o una región de mejorador, y de este modo, lograr la modificación funcional de dicho gen o lograr la modificación de la expresión de un gen.

En algunas formas de realización, el método además comprende obtener progenie de la planta editada genéticamente. En un aspecto adicional, la divulgación además proporciona una planta editada genéticamente o su progenie, o sus partes, en donde la planta se obtiene por el método descripto anteriormente.

En otro aspecto, la presente divulgación proporciona un método de fitomejoramiento, que comprende cruzar una primera planta editada genéticamente por el método anteriormente mencionado de la presente divulgación, con una segunda planta que no contiene dicha modificación genética, de modo de introducir dicha modificación genética en dicha segunda planta.

En formas de realización particulares, el método además incluye cribar la célula vegetal después de los pasos de introducción, a fin de determinar si la expresión del gen de interés se ha modificado. En formas de realización particulares, los métodos incluyen el paso de regenerar una planta a partir de la célula vegetal. En formas de realización adicionales, los métodos incluyen la reproducción cruzada de la planta para obtener un linaje de planta genéticamente deseado.

Suministro transitorio de componentes del sistema de edición dirigido a la célula vegetal.

En algunas formas de realización, la presente divulgación proporciona un suministro de uno o más componentes del sistema de edición de bases dirigido directamente a la célula vegetal. Esto es de interés, *inter alia*, para la generación de plantas no transgénicas. En algunas formas de realización, uno o más de los compuestos del sistema de edición de bases dirigido se preparan fuera de la planta o célula vegetal y se suministran a la célula. En algunas formas de realización, la proteína de fusión de edición de bases dirigida que comprende un dominio CRISPR–Cas inactivado por nucleasas y el

dominio de citosina desaminasa se prepara *in vitro* antes de la introducción a la célula vegetal. La proteína de fusión de edición de bases puede prepararse por diversos métodos conocidos por el experto en la técnica, que incluyen la producción recombinante. Después de la expresión, la proteína CRISPR–Cas es aislada, replegada si es necesario, purificada y opcionalmente tratada para eliminar cualquier etiqueta de purificación si se presenta una etiqueta tal como una etiqueta His en la proteína de fusión. Una vez que se obtiene proteína CRISPR–Cas cruda, parcialmente purificada o más completamente purificada, la proteína puede introducirse en la célula vegetal.

En algunas formas de realización, la proteína de fusión de edición de bases se mezcla con ARN guía que se dirige al gen de interés para formar una ribonucleoproteína preensamblada. Los componentes individuales o la ribonucleoproteína preensamblada pueden introducirse en la célula vegetal mediante electroporación, por bombardeo con partículas recubiertas con sistema de edición de bases dirigido, mediante transfección química o por cualquier algún otro medio de transporte a través de una membrana celular. Por ejemplo, se ha demostrado que la transfección de un protoplasto vegetal con una ribonucleoproteína CRISPR–Cas9 preensamblada asegura la modificación dirigida del genoma de la planta (Woo et al. *Nature Biotechnology*, 2015; 33(11): 1162–1164).

En algunas formas de realización, los componentes del sistema de edición de bases dirigido que se describe en la presente memoria se introducen en la célula vegetal usando nanopartículas. Los componentes, o bien como proteína o ácido nucleico, o en una de sus combinaciones, pueden cargarse o empaquetarse en nanopartículas y aplicarse a las plantas (tal como se describe, por ejemplo, en los documentos WO2008/042156 y US2013/0185823). En particular, la divulgación enseña nanopartículas cargadas o empaquetadas con moléculas de ADN que codifica la proteína efectora CRISPR–Cas, moléculas de ADN que codifica citosina desaminasa (que pueden fusionarse con la proteína CRISPR–Cas o un enlazador), y moléculas de ADN que codifica el ARN guía y/o ARN guía aislado, como se describe en el documento WO2015/089419.

Otros medios para introducir uno o más componentes del sistema de edición de bases dirigido a la célula vegetal son mediante el uso de péptidos de penetración celular (CPP, conforme a sus siglas en inglés). Por lo tanto, en formas de realización particulares, las formas de realización de la divulgación comprenden composiciones que comprenden un péptido de penetración celular unido a la proteína de fusión de edición de bases. En formas de realización particulares, la proteína de fusión de edición de bases y/o ARN guía se acopla con uno o más CPP para transportarlos de manera eficaz dentro de los protoplastos de las plantas. Ramakrishna (*Genome Res.* 2014 Jun; 24(6): 1020–7, para Cas9 en células humanas). En otras formas de realización, la proteína de fusión de edición de bases

y/o ARN guía son codificadas por una o más moléculas de ADN circular o no circular que se acoplan a uno o más CPP para el suministro al protoplasto de la planta. Los protoplastos vegetales luego son regenerados como células vegetales, y adicionalmente, en plantas. Los CPP se describen generalmente como péptidos cortos de menos de 35 aminoácidos, o bien derivados de proteínas, o de secuencias quiméricas que son capaces de transportar biomoléculas a través de la membrana celular de una manera independiente de receptor. Los CPP pueden ser péptidos catiónicos, péptidos que tienen secuencias hidrófobas, péptidos anfipáticos, péptidos que tienen secuencia rica en prolina y antimicrobiana, y péptidos quiméricos o bipartitos (Pooga y Langel 2005). Los CPP pueden penetrar las membranas biológicas y, de este modo, desencadenar el movimiento de diversas biomoléculas a través de las membranas celulares hacia el citoplasma y mejorar su ruta intracelular, y en consecuencia, facilitar la interacción de la biomolécula con el objetivo.

En algunas formas de realización, los métodos que se describen en esta memoria se usan para modificar genes endógenos, o para modificar su expresión, sin la introducción permanente de ningún gen extraño que incluye aquellos que codifican componentes de CRISPR, en el genoma de la planta, de modo de evitar la presencia de ADN extraño en el genoma de la planta. En formas de realización particulares, esto se garantiza mediante la expresión transitoria de los componentes del sistema de edición de bases dirigido. En formas de realización particulares, uno o más de los componentes se expresan en uno o más vectores virales que producen suficiente proteína CRISPR–Cas, citosina desaminasa, y ARN guía para garantizar de manera uniforme y constante la modificación de un gen de interés de acuerdo con un método descrito en esta memoria. En formas de realización particulares, la expresión transitoria de las construcciones de sistemas de edición de bases dirigidos se asegura en los protoplastos vegetales, y en consecuencia, no se integran en el genoma. La limitada ventana de expresión puede ser suficiente para permitir que el sistema de edición de bases dirigido asegure la modificación de un gen objetivo como se describe aquí.

En algunas formas de realización, los diferentes componentes del sistema de edición de bases dirigido se introducen en una célula vegetal, protoplasto o tejido vegetal, o bien en forma separada, o en mezcla, con la ayuda de moléculas de suministro tales como nanopartículas o moléculas de CPP como se describe aquí anteriormente. La expresión de los componentes del sistema de edición de bases dirigido puede inducir la modificación dirigida del genoma, mediante la actividad de la citosina desaminasa. Las diferentes estrategias descritas aquí permiten la edición del genoma dirigida mediada por CRISPR, sin requerir la introducción de los componentes del sistema de edición de bases

dirigido en el genoma de la planta. Los componentes que se introducen en forma transitoria en la célula vegetal típicamente se eliminan con el cruce.

En algunas formas de realización, las células vegetales que tienen un genoma modificado y que son producidas u obtenidas por cualquiera de los métodos que se describen en esta memoria pueden cultivarse para regenerar una planta entera que posee el genotipo modificado o transformado y, en consecuencia, el fenotipo deseado. Las técnicas de regeneración convencionales son bien conocidas por los expertos en la técnica. Ejemplos particulares de dichas técnicas de regeneración se sustentan en la manipulación de ciertas fitohormonas en un medio de crecimiento de cultivo de tejido, y típicamente, se sustentan en un marcador de biocida y/o herbicida que se ha introducido junto con las secuencias de nucleótidos deseadas. En formas de realización particulares adicionales, la regeneración de plantas se obtiene a partir de protoplastos cultivados, callo vegetal, explantaciones, órganos, polen, embriones o sus partes (ver, por ejemplo, la referencia de Evans et al. (1983), *Handbook of Plant Cell Culture*, Klee et al (1987) *Ann. Rev. of Plant Phys.*).

En algunas formas de realización, las plantas mejoradas o transformadas como se describen aquí pueden ser autopolinizadas para proporcionar semillas para plantas homocigotas mejoradas en rasgos de la divulgación, que tienen un rasgo deseado tal como falta de semillas, reducido tamaño de semilla, reducido tejido de endocarpio o endocarpio menos lignificado (homocigotas para la modificación de ADN), o pueden cruzarse con plantas no transgénicas o plantas mejoradas en rasgos diferentes, a fin de proporcionar semillas para plantas heterocigotas. Cuando se introduce un ADN recombinante en la célula vegetal, la planta resultante de dicho cruce es una planta que es heterocigota para la molécula de ADN recombinante. Estas plantas homocigotas y heterocigotas obtenidas mediante el cruce a partir de plantas con rasgos mejorados y que comprenden la modificación genética (que puede ser un ADN recombinante) se refieren aquí como “progenie”. Alternativamente, las plantas editadas genéticamente pueden obtenerse por cualquiera de los métodos que se describen en esta memoria usando el sistema de edición de bases dirigido, por lo cual no se incorpora ningún ADN extraño en el genoma usando la expresión y el suministro transitorios, o por lo cual se incorpora ADN extraño en el genoma usando la transformación estable, pero es eliminado/segregado con el cruce. La progenie de dichas plantas, obtenida mediante el fitomejoramiento adicional, también puede contener la modificación genética tal como sustituciones de nucleótidos. Los fitomejoramientos se realizan mediante cualquier método de fitomejoramiento comúnmente usado para diferentes cultivos (por ejemplo, Allard, *Principles of Plant Breeding*, John Wiley & Sons, NY, U. of CA, Davis, CA, 50–98 (1960).

Plantas y partes de plantas que tienen un factor de transcripción de caja MADS clado AG modificado y métodos para su producción.

En vista de lo anterior, se describen aquí formas de realización adicionales de la invención. Por lo tanto, en algunas formas de realización, se proporciona una planta o parte de planta de esta, que comprende por lo menos una mutación no natural en un gen endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS. Un factor de transcripción de caja MADS clado AG útil con esta invención incluye, sin limitación, AGAMOUS (AG), SHATTERPROOF 1 (SHP1), SHATTERPROOF 2 (SHP2), y/o SEEDSTICK (STK). En algunas formas de realización, el gen endógeno que codifica un factor de transcripción de caja MADS clado AG es un gen endógeno gen SEEDSTICK (STK).

Una mutación no natural como se describe en esta memoria incluye, sin limitación, un alelo dominante–negativo, alelo semidominante, alelo de pérdida de función débil, o una mutación hipomórfica. En algunas formas de realización, la mutación es un alelo dominante–negativo o un alelo de pérdida de función débil. En algunas formas de realización, la mutación es un alelo de pérdida de función débil, o una mutación hipomórfica.

En algunas formas de realización, el gen endógeno que codifica un factor de transcripción de caja MADS clado AG es capaz de regular la producción de semillas. En algunas formas de realización, una planta que comprende la por lo menos una mutación no natural en un gen endógeno que codifica un factor de transcripción de caja MADS clado AG exhibe desarrollo de fruto alterado, opcionalmente en donde el gen endógeno que codifica un factor de transcripción de caja MADS clado AG es un gen de factor de transcripción SEEDSTICK (STK). En algunas formas de realización, el desarrollo de fruto alterado comprende un fenotipo que incluye, sin limitación, falta de semillas (por ejemplo, sin semillas o un número reducido de semillas), reducida simiente, reducido contenido de lignina de semillas, reducida formación de endocarpio de semilla, o superficie de semilla más lisa.

En algunas formas de realización, el gen endógeno SEEDSTICK (STK) codifica un polipéptido que tiene por lo menos 80% de identidad (por ejemplo, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, o 100%) con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o codifica una región de un polipéptido que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**; o

comprende una secuencia de nucleótidos que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**. Por lo tanto, en algunas formas de realización, los motivos/dominios que se describen aquí pueden usarse para identificar un gen STK para modificar una planta o parte de la misma de acuerdo con los métodos de esta invención, a fin de producir una planta sin semillas o parte de la misma o una planta o parte de la misma que tiene reducida simiente, reducido contenido de lignina de semillas, reducida formación de endocarpio de semillas, o superficie de semilla más lisa.

En algunas formas de realización, la por lo menos una mutación no natural se produce en una región de un factor de transcripción de caja MADS clado AG que comprende cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174** o una de sus porciones. En algunas formas de realización, la por lo menos una mutación no natural produce una sustitución en un residuo de aminoácido ubicado en cualquiera de una o más de las posiciones 83–86 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 13** o ubicado en la posición 57–60 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164, o 166–171**. En algunas formas de realización, la por lo menos una mutación no natural produce una sustitución en un residuo de aminoácido ubicado en cualquiera de una o más de las posiciones 116–128 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 13**, ubicado en la posición 91–103 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 29, SEQ ID NO: 78 SEQ ID NO: 97, SEQ ID NO: 152, SEQ ID NO: 154, SEQ ID NO: 156, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 162, SEQ ID NO: 164 o SEQ ID NOs: 166–171**, o ubicado en la posición 90–102 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 53, SEQ ID NO: 66, SEQ ID NO: 128, o SEQ ID NO: 150**, opcionalmente ubicado en la posición 116–119 (YQQE) con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 13**, posición 91–94 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 29, 53, 66, 78, 97, 152, 154, 156, 158, 159, 160, 162, 164, o 166–171**, o posición 90–93 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 53, 66, 128 o 150**.

En algunas formas de realización, la por lo menos una mutación no natural es una sustitución de base, una delección de base y/o una inserción de base. En algunas formas de realización, una sustitución de base a una A, una T, una G o una C. En algunas formas de

realización, la por lo menos una mutación no natural es una sustitución de por lo menos un par de bases (por ejemplo, 1, 2, 3, 4 o 5 pares de bases). En algunas formas de realización, la por lo menos una mutación no natural da como resultado una delección del extremo 3' del gen y un polipéptido que tiene un truncamiento C-terminal, o da como resultado una delección del extremo 5' del gen y un polipéptido que tiene un truncamiento N-terminal.

En algunas formas de realización, la por lo menos una mutación no natural es una sustitución de aminoácido en una arginina conservada en un gen STK, en donde el residuo arginina conservado (arg, R) se ubica: en el residuo 196 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 66**; en el residuo 197 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 53, 78, 150, 166–171**; en el residuo 198 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 29**; residuo 199 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 128**; en el residuo 201 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 97, 152, 154, 156, 158, 160, 162, 164**; o en el residuo 225 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 13**. En algunas formas de realización, la sustitución es R>L o R>H.

En algunas formas de realización, la por lo menos una mutación no natural está en el dominio C-terminal del polipéptido codificado por el gen endógeno que codifica un factor de transcripción de caja MADS clado AG, opcionalmente, en el dominio C-terminal de un polipéptido codificado por un gen STK. En algunas formas de realización, por lo menos una mutación no natural en un factor de transcripción de caja MADS clado AG produce una delección de por lo menos un aminoácido o por lo menos dos o más aminoácidos consecutivos.

En algunas formas de realización, una delección es una delección en marco o una delección fuera de marco, opcionalmente en donde la delección en marco o la delección fuera de marco produce un codón de parada no natural/prematuro y un truncamiento C-terminal del residuo de aminoácido o dos o más aminoácidos consecutivos. En algunas formas de realización, dicha delección elimina el residuo arginina conservado (arg, R) ubicado en el residuo 196 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 66**; en el residuo 197 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 53, 78, 150, 166–171**; en el residuo 198 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 29**; residuo 199 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 128**; en el residuo 201 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 97, 152, 154, 156, 158, 160, 162, 164**; o

en el residuo 225 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 13**, opcionalmente en donde los residuos que rodean a la arginina conservada son eliminados.

En algunas formas de realización, la invención proporciona una célula vegetal, que comprende un sistema de edición de bases que comprende: (a) una proteína efectora asociada a CRISPR; (b) una citidina desaminasa o adenosina desaminasa; y (c) un ácido nucleico guía (ARNg) que tiene una secuencia espaciadora con complementariedad a un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AG. En algunas formas de realización, el factor de transcripción de caja MADS clado AG es AGAMOUS (AG), SHATTERPROOF 1 (SHP1), SHATTERPROOF 2 (SHP2), y/o SEEDSTICK (STK), opcionalmente el factor de transcripción de caja MADS clado AG es un gen SEEDSTICK (STK). En algunas formas de realización, el gen de factor de transcripción de caja MADS clado AG endógeno codifica un factor de transcripción SEEDSTICK (STK) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164 o 166–171**, o comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**. En algunas formas de realización, el factor de transcripción SEEDSTICK (STK) es codificado por una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o es codificado por una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**. En algunas formas de realización, el ARNg útil con esta invención comprende una secuencia espaciadora que tiene por lo menos 80% de identidad con cualquiera de **SEQ ID NOs: 175, 176, 184, 185, o 186**.

En algunas formas de realización, se proporciona una planta o parte de la misma (por ejemplo, una célula vegetal) comprende por lo menos una mutación no natural en un gen de factor de transcripción de caja MADS clado AG endógeno que tiene desarrollo de fruto alterado, en donde la por lo menos una mutación no natural es una sustitución, inserción o delección que se introduce usando un sistema de edición que comprende un dominio de unión de ácido nucleico que se une a un sitio objetivo en el gen del factor de transcripción de caja MADS clado AG, en donde el gen del factor de transcripción de caja MADS clado AG es un gen SEEDSTICK (STK), el gen STK (a) comprende por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID**

NOs: 172–174; o (b) codifica un factor de transcripción de caja MADS clado AG que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164 o 166–171** o comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189.** En algunas formas de realización, el dominio de unión de ácido nucleico se une a un sitio objetivo que tiene la secuencia de nucleótidos de cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172, 173 o 174.** El desarrollo de fruto alterado incluye, sin limitación, un fenotipo de falta de semillas (sin semillas o reducida producción de semillas) o reducida simiente, como se describe en esta memoria.

En algunas formas de realización, la por lo menos una mutación no natural es un alelo dominante–negativo, alelo semidominante, alelo de pérdida de función débil, o una mutación hipomórfica. En algunas formas de realización, la por lo menos una mutación no natural es por lo menos una (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, o más) sustitución de base, deleción de base, o adición de base. En algunas formas de realización, la por lo menos una mutación no natural es una mutación puntual. En algunas formas de realización, la por lo menos una sustitución de base comprende una sustitución a un A, un T, un G, o un C. En algunas formas de realización, la por lo menos una mutación no natural es una mutación de desplazamiento de marco (por ejemplo, una deleción en marco; una deleción fuera de marco). Dichas deleciones pueden alterar el marco de lectura, para producir una terminación prematura de la traducción, por ejemplo, un codón de parada prematuro, un truncamiento del polipéptido. En algunas formas de realización, la por lo menos una mutación no natural en un gen de factor de transcripción de caja MADS clado AG endógeno produce una deleción de por lo menos un aminoácido o dos o más aminoácidos consecutivos (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20 o más), opcionalmente, donde los dos o más aminoácidos son aminoácidos consecutivos.

En algunas formas de realización, por lo menos una mutación no natural está en el dominio N–terminal del polipéptido codificado por el gen endógeno que codifica un factor de transcripción de caja MADS clado AG; opcionalmente, un polipéptido STK. En algunas formas de realización, por lo menos una mutación no natural está en el dominio C–terminal del polipéptido codificado por el gen endógeno que codifica un factor de transcripción de caja MADS clado AG; opcionalmente, un polipéptido STK.

En algunas formas de realización, la por lo menos una mutación no natural en el dominio C–terminal es un truncamiento C–terminal, opcionalmente en donde por lo menos los últimos 25, 26,

27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 aminoácidos consecutivos del extremo C-terminal de un polipéptido STK están truncados. Dicho truncamiento produce la eliminación de una arginina conservada en el extremo C-terminal del polipéptido STK, en donde la arginina conservada es el residuo 196 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 66**; residuo 197 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 53, 78, 150, 166–171**; residuo 198 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 29**; residuo 199 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 128**; residuo 201 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 97, 152, 154, 156, 158, 160, 162, 164**; o residuo 225 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 13**.

En algunas formas de realización, la por lo menos una mutación no natural es en un gen endógeno SEEDSTICK (STK) y produce un residuo de aminoácido sustituido ubicado en la posición 196 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 66**; en la posición 197 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 53, 78, 150, 166–171**; en la posición 198 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 29**; en la posición 199 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 128**; en la posición 201 con referencia a la numeración de posición de aminoácido de **SEQ ID NOs: 97, 152, 154, 156, 158, 160, 162, 164**; o en la posición 225 con referencia a la numeración de posición de aminoácido de **SEQ ID NO: 13**. En algunas formas de realización, la sustitución es arginina a leucina (R>L) o arginina a histidina (R>H).

La invención además proporciona una planta o parte de la misma que comprende un gen de factor de transcripción de caja MADS clado AG endógeno mutado, en donde el gen de factor de transcripción de caja MADS clado AG endógeno mutado es un gen SEEDSTICK (STK) que (a) comprende una secuencia que tiene por lo menos 80% de identidad con la secuencia de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o secuencia que comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**; o (b) codifica una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164 o 166–171** o una secuencia que comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de

aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, en donde la planta que comprende un gen STK mutado exhibe desarrollo de fruto alterado.

Se proporcionan métodos para producir una planta o parte de planta que tiene un desarrollo alterado del fruto. En algunas formas de realización, se proporciona un método para producir/reproducir una planta editada con bases libre de transgén, donde el método comprende: (a) cruzar la planta de la invención con una planta libre de transgén, de modo de introducir así la por lo menos una mutación, la mutación o la modificación en la planta que está libre de transgenes; y (b) seleccionar una planta de progenie que comprenda por lo menos una sustitución de nucleótido único, pero que esté libre de transgén, a fin de producir así una planta editada con bases libre de transgenes que tiene un desarrollo alterado del fruto.

En algunas formas de realización, se proporciona un método para producir una pluralidad de plantas que exhiben desarrollo alterado del fruto, en donde el método comprende plantar dos o más plantas de la invención en un área de cultivo (por ejemplo, un campo (por ejemplo, un campo cultivado, un campo agrícola), una cámara de crecimiento, un invernadero, un área recreativa, un césped y/o una carretera y similares), para proporcionar una pluralidad de plantas que exhiben un desarrollo alterado del fruto en comparación con una pluralidad de plantas de control que no comprenden la mutación.

En algunas formas de realización, se proporciona un método para editar un sitio específico en el genoma de una célula vegetal, en donde el método comprende: escindir, de una manera específica del sitio, un sitio objetivo dentro de un gen de factor de transcripción de caja MADS clado AG endógeno en la célula vegetal, en donde el gen del factor de transcripción de caja MADS clado AG endógeno es un gen SEEDSTICK (STK) (a) que comprende por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**; o (b) que codifica un factor de transcripción de caja MADS clado AG que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164 o 166–171** o comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, de modo de generar una edición en el gen de factor de transcripción de caja MADS clado AG endógeno de la célula vegetal. En algunas formas de realización, el método además comprende regenerar una planta a partir de la célula vegetal que comprende una edición en

el gen STK endógeno para producir una planta que comprende la edición en su gen STK endógeno, en el que la edición da como resultado una mutación no natural. En algunas formas de realización, una planta que comprende la edición en su gen STK endógeno exhibe un desarrollo alterado del fruto, opcionalmente en el que el desarrollo alterado del fruto es un fenotipo de falta de semilla y/o disminución de la simiente. En algunas formas de realización, la mutación no natural produce un alelo dominante negativo, un alelo semidominante, un alelo de pérdida de función débil o una mutación hipomórfica. En algunas formas de realización, la mutación no natural es una sustitución, una adición o una delección. En algunas formas de realización, la mutación no natural es una delección de por lo menos 4 pares de bases consecutivos a alrededor de 150 pares de bases consecutivos desde el extremo 3' del gen STK, en donde la delección da como resultado un truncamiento C-terminal que comprende un truncamiento de por lo menos 1 residuo de aminoácido a alrededor de 50 residuos de aminoácido consecutivos.

Adicionalmente se proporciona un método para producir una planta, que comprende: (a) poner en contacto una población de células vegetales que comprenden una secuencia de ADN que codifica un gen SEEDSTICK (STK) endógeno de tipo salvaje con una nucleasa unida a un dominio de unión de ácido nucleico (por ejemplo, sistema de edición) que se une a una secuencia: (a) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165**, o que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**; o (b) que codifica un factor de transcripción SEEDSTICK (STK) (i) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171** o (ii) que comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**; (b) seleccionar una célula vegetal de dicha población en donde por lo menos una secuencia de ADN que codifica el factor de transcripción STK se ha mutado, en donde la mutación comprende una sustitución de por lo menos un nucleótido en la por lo menos una secuencia de ADN; y (c) cultivar la célula vegetal seleccionada hasta una planta.

En algunas formas de realización, se proporciona un método para producir una planta o parte de la misma que comprende por lo menos una célula en la que un gen de factor de transcripción de caja MADS clado AG endógeno está mutado, donde el método comprende poner en contacto un sitio objetivo en el gen del factor de transcripción de caja MADS clado AG en la planta o parte de

planta con una nucleasa unida a un dominio de unión de ácido nucleico (por ejemplo, sistema de edición) que se une al sitio objetivo, en donde el gen del factor de transcripción de caja MADS clado AG es un gen SEEDSTICK (STK) (a) que comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, o (b) que codifica un factor de transcripción SEEDSTICK (STK) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o que comprende una región que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, de modo de producir una planta o parte de la misma que comprende por lo menos una célula que tiene una mutación en el gen de factor de transcripción de caja MADS clado AG endógeno.

En algunas formas de realización, se proporciona un método para producir una planta que tiene un fenotipo de desarrollo de fruto alterado, que comprende (a) poner en contacto una célula vegetal que comprende un gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje con una nucleasa que se dirige al gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje, en donde la nucleasa se une a un dominio de unión de ADN que se une a un sitio objetivo en el gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje, en donde el gen de factor de transcripción de caja MADS clado AG endógeno de tipo salvaje es un gen SEEDSTICK (STK) que (i) comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174** y/o (ii) codifica una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o una secuencia que comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**; (b) cultivar la célula vegetal hasta una planta, de manera de producir una planta que tiene un fenotipo de fruto alterado.

En algunas formas de realización, se proporciona un método para producir una planta o parte de la misma que comprende por lo menos una célula que tiene un gen de factor de transcripción de caja MADS clado AG endógeno mutado, donde el método comprende poner en contacto un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG endógeno en la planta o parte de planta con una nucleasa que comprende un dominio de escisión y un dominio de unión de ADN, en donde el dominio de unión de ADN se une a un sitio objetivo en el gen de factor de transcripción de caja MADS clado AG endógeno, en donde el gen de factor de transcripción de caja MADS clado AG endógeno es un gen SEEDSTICK (STK) que (a) codifica (i) una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171** o (ii) una secuencia que comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**; y/o (b) comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o o una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, de modo de producir la planta o parte de la misma que comprende por lo menos una célula que tiene un gen de factor de transcripción de caja MADS clado AG endógeno mutado.

En algunas formas de realización, se proporciona un método de producción de una planta o parte de la misma que comprende un gen de factor de transcripción de caja MADS clado AG endógeno mutado y que tiene desarrollo de fruto alterado, donde el método comprende poner en contacto un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG endógeno con una nucleasa que comprende un dominio de escisión y un dominio de unión de ADN (por ejemplo, sistema de edición) que comprende un dominio de unión de ácido nucleico que se une al sitio objetivo en el gen del factor de transcripción de caja MADS clado AG, en donde el gen del factor de transcripción de caja MADS clado AG es un gen SEEDSTICK (STK) (a) que comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, o (b) que codifica un factor de transcripción SEEDSTICK (STK) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o**

166–171, o que comprende una región que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, de modo de producir una planta o parte de la misma que comprende un gen de factor de transcripción de caja MADS clado AG endógeno mutado y que tiene desarrollo de fruto alterado.

En algunas formas de realización, la nucleasa (por ejemplo, de un sistema de edición) escinde el gen STK endógeno, y se introduce una mutación en el gen STK endógeno. Como se describe en el presente documento, una nucleasa incluye una nucleasa de dedo de zinc, una nucleasa efectora de tipo activadora de la transcripción (TALEN), una endonucleasa (por ejemplo, Fok1) o una proteína efectora CRISPR–Cas. En algunas formas de realización, un dominio de unión a ADN es un dedo de zinc, un dominio de unión a ADN de tipo activador de transcripción (TAL), un dominio de unión de ADN de efector CRISPR–Cas o argonauta.

En algunas formas de realización, la mutación que se introduce es una mutación no natural, en la que la mutación no natural da como resultado un desarrollo alterado del fruto en la planta. Como se describe en el presente documento, un fenotipo de desarrollo de fruta alterado incluye, sin limitación, falta de semilla o reducción de la simiente.

Una mutación no natural útil para los métodos de esta invención incluye, sin limitación, una sustitución, una inserción y/o una deleción de por lo menos una base (nucleótido) (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, o más) o por lo menos un aminoácido (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, o más). En algunas formas de realización, la por lo menos una mutación no natural es una mutación puntual. En algunas formas de realización, la por lo menos una sustitución de base comprende una sustitución a un A, un T, un G, o un C. En algunas formas de realización, la por lo menos una mutación no natural es una mutación de desplazamiento de marco (por ejemplo, una deleción en marco; una deleción fuera de marco). Dichas deleciones pueden alterar el marco de lectura para producir la terminación prematura de la traducción, por ejemplo, un codón de parada prematuro o un truncamiento del polipéptido.

En algunas formas de realización, una deleción puede ser una deleción de por lo menos 4 pares de bases consecutivos a alrededor de 150 pares de bases consecutivos del extremo 3' del gen STK, cuya deleción puede dar como resultado un truncamiento C–terminal que comprende un truncamiento de por lo menos 1 residuo de aminoácido a alrededor de 50 residuos de aminoácidos consecutivos desde el extremo C–terminal de un polipéptido codificado por el gen STK.

También se proporcionan aquí ejemplos de ácidos nucleicos guía para llevar a cabo los métodos de la invención. En función de la guía de la presente invención, se pueden desarrollar guías

adicionales más allá de las específicamente proporcionadas. En algunas formas de realización, se proporciona un ácido nucleico guía (por ejemplo, ARNg, ADNg, ARNcr, ADNcr) que se une a un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG, en donde el gen de factor de transcripción de caja MADS clado AG endógeno es un gen SEEDSTICK (STK) (a) que comprende una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, o

(b) que codifica un factor de transcripción SEEDSTICK (STK) que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o una región que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**. En algunas formas de realización, el ácido nucleico guía comprende un espaciador que tiene la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 175, 176, 184, 185, o 186**.

En algunas formas de realización, se proporciona un sistema que comprende un ácido nucleico guía de la invención y una proteína efectora CRISPR–Cas que se asocia con el ácido nucleico guía. En algunas formas de realización, el sistema comprende además un ácido nucleico tracr que se asocia con el ácido nucleico guía y una proteína efectora CRISPR–Cas, opcionalmente en donde el ácido nucleico tracr y el ácido nucleico guía están unidos covalentemente.

En algunas formas de realización, se proporciona un sistema de edición de genes que comprende una proteína efectora CRISPR–Cas en asociación con un ácido nucleico guía, en donde el ácido nucleico guía comprende una secuencia espaciadora que se une a un gen de factor de transcripción de caja MADS clado AG. En algunas formas de realización, el gen del factor de transcripción de caja MADS clado AG es un gen SEEDSTICK (STK) (a) que comprende por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, y/o

(b) que codifica (i) una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o (ii) una secuencia que comprende una región que tiene

por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**. En algunas formas de realización, el ácido nucleico guía del sistema de edición génica comprende una secuencia espaciadora que tiene la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 175, 176, 184, 185, o 186**. En algunas formas de realización, el sistema de edición génica puede comprender además un ácido nucleico tracr que se asocia con el ácido nucleico guía y una proteína efectora CRISPR–Cas, opcionalmente en donde el ácido nucleico tracr y el ácido nucleico guía están unidos covalentemente.

En algunas formas de realización, se proporciona un complejo que comprende una proteína efectora CRISPR–Cas que comprende un dominio de escisión y un ácido nucleico guía, en donde el ácido nucleico guía se une a un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG, en donde el gen del factor de transcripción de caja MADS clado AG es un gen SEEDSTICK (STK) (a) que comprende por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de **SEQ ID NOs: 172–174**, y/o (b) que codifica (i) una secuencia que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128 o 150**, o (ii) una secuencia que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164, o 166–171**, o una región que tiene por lo menos 80% de identidad con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188, o 189**, en donde el dominio de escisión escinde una hebra objetivo en el gen del factor de transcripción de caja MADS clado AG.

En algunas formas de realización, se proporciona un casete de expresión, donde el casete de expresión comprende (a) un polinucleótido que codifica proteína efectora CRISPR–Cas que comprende un dominio de escisión y (b) un ácido nucleico guía que se une a un sitio objetivo en un gen de factor de transcripción de caja MADS clado AG, en donde el gen de factor de transcripción de caja MADS clado AG es un gen SEEDSTICK, y el ácido nucleico guía comprende una secuencia espaciadora que es complementaria y se une a una secuencia que tiene por lo menos 80% de identidad de secuencia con por lo menos una porción de una secuencia que tiene por lo menos 80% de identidad de secuencia con la secuencia de nucleótidos de cualquiera de **SEQ ID NOs: 153, 155, 157, 159, 161, 163, o 165** o una porción de una secuencia que comprende una región que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de nucleótidos de

SEQ ID NOs: 172–174; o una secuencia que codifica (i) un polipéptido que tiene por lo menos 80% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 13, 29, 53, 66, 78, 97, 128, 150, 152, 154, 156, 158, 160, 162, 164,** o **166–171,** o (ii) un polipéptido que comprende una región que tiene por lo menos 90% de identidad de secuencia con cualquiera de las secuencias de aminoácidos de **SEQ ID NOs: 177–183, 187, 188,** o **189,** opcionalmente, en donde una porción es de alrededor de 2 a alrededor de 22 nucleótidos consecutivos de longitud.

En algunas formas de realización, se proporciona un ácido nucleico, donde el ácido nucleico codifica un alelo dominante–negativo, alelo semidominante, alelo de pérdida de función débil, o una mutación hipomórfica de un factor de transcripción de caja MADS clado AG, opcionalmente en donde el factor de transcripción de caja MADS clado AG es AGAMOUS (AG), SHATTERPROOF 1 (SHP1), SHATTERPROOF 2 (SHP2), y/o SEEDSTICK (STK). En algunas formas de realización, el gen endógeno que codifica un factor de transcripción de caja MADS clado AG es un gen endógeno SEEDSTICK (STK) como se describe aquí. Además se proporciona una planta o parte de la misma que comprende el ácido nucleico de la invención, en donde la planta o parte de la misma exhibe desarrollo de fruto alterado.

Las plantas o partes de las mismas son como se describen en el presente documento y pueden incluir cualquier planta o parte de las mismas para las cuales es deseable la falta de semillas, la disminución de la simiente, el contenido reducido de lignina de la semilla, la formación reducida de endocarpio de la semilla o una superficie de la semilla más lisa. Por lo tanto, una planta o parte de la misma puede ser una dicotiledónea o una monocotiledónea, por ejemplo, y puede incluir, entre otras, frambuesa, frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, manzana, tomate, uva, durazno, ciruela, damasco, pera, membrillo, níspero, dátiles o almendras. Cuando se modifica un factor de transcripción de la caja MADS clado AG de una parte de la planta, incluida una célula de la planta, la parte de la planta puede regenerarse hasta una planta. Se conocen métodos para regenerar plantas, que se pueden usar fácilmente con las partes de plantas de esta invención.

EJEMPLOS.

Los siguientes ejemplos se proporcionan con el fin de ilustrar diversas formas de realización de la divulgación, y no pretenden limitar la presente divulgación de ninguna manera. Los expertos en la técnica podrán idear cambios de la presente divulgación y otros usos que están abarcados dentro del espíritu de la divulgación, tal como se define por el alcance de las reivindicaciones.

Ejemplo 1: Identificación de factores de transcripción de caja MADS clado AGAMOUS.

Los factores de transcripción de caja MADS clado AGAMOUS fueron definidos y descriptos en § II “Metodología de fitomejoramiento de próxima generación que utiliza mutaciones en factores de transcripción de caja MADS clado AGAMOUS para el desarrollo alterado del fruto”. Los contornos del género y la definición de los mismos se proporcionan en esa sección. El **Ejemplo 1** presente describe cómo se descubrió este clado como un posible gen diana para la mutación, a fin de evaluar los fenotipos del desarrollo de fruto alterado.

Se realizaron análisis bioinformáticos y evolutivos moleculares en la base de datos del genoma de siete especies de plantas, que incluyen: *Arabidopsis* (*Arabidopsis thaliana*), Manzana (*Malus pumila*), Frambuesa negra (*Rubus occidentalis*), Uva (*Vitis vinifera*), Durazno (*Prunus persica*), Tomate (*Solanum lycopersicum*) y Frutilla silvestre (*Fragaria vesca*). Utilizando técnicas analíticas derivadas de la biología evolutiva molecular, se buscaron secuencias de genes de caja MADS de *Arabidopsis thaliana* en la base de datos del genoma de otras especies de plantas de interés para obtener homólogos, ortólogos y/o parálogos de los genes de interés de caja MADS de *Arabidopsis*. En consecuencia, cada secuencia de aminoácidos de los factores de transcripción de la caja MADS se recolectó de la base de datos del genoma de cada especie (**Tabla 2**) y luego se realizaron análisis de alineación de secuencias múltiples (MSA).

[0001] Tabla 2.

SEQ ID NO:	Tipo de secuencia	Factor de transcripción caja MADS clado AGAMOUS (Sí/No)	Nombre común	Nombre científico	ID Gen
1	Aminoácido	Sí (STK)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT4G09960.3
2	Aminoácido	Sí (STK)	Frambuesa negra	<i>Rubus occidentalis</i>	Ro05_G22454
3	Aminoácido	Sí (STK)	Frutilla (silvestre)	<i>Fragaria vesca</i>	FvH4_5g32540.1
4	Aminoácido	Sí (STK)	Durazno	<i>Prunus persica</i>	Prupe.1G549600.1
5	Aminoácido	Sí (STK)	Manzana	<i>Malus pumila</i>	MD08G1216500
6	Aminoácido	Sí (STK)	Manzana	<i>Malus pumila</i>	MD15G1403600
7	Aminoácido	Sí (STK)	Uva	<i>Vitis vinifera</i>	VIT_18s0041g01880.0
8	Aminoácido	Sí (STK)	Tomate	<i>Solanum lycopersicum</i>	Solyc06g064840.3.1
9	Aminoácido	Sí (STK)	Tomate	<i>Solanum lycopersicum</i>	Solyc11g028020.2.1
10	Aminoácido	Sí (SHP2)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT2G42830.2
11	Aminoácido	Sí (SHP1)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT3G58780.4
12	Aminoácido	Sí (SHP1/2)	Frambuesa negra	<i>Rubus occidentalis</i>	Ro06_G09924
13	Aminoácido	Sí (SHP1/2)	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_6g37880.1
14	Aminoácido	Sí (SHP1/2)	Durazno	<i>Prunus persica</i>	Prupe.3G170600.1
15	Aminoácido	Sí (SHP1/2)	Manzana	<i>Malus pumila</i>	MD09G1155200
16	Aminoácido	Sí (SHP1/2)	Uva	<i>Vitis vinifera</i>	VIT_12s0142g00360.0
17	Aminoácido	Sí (SHP1/2)	Tomate	<i>Solanum lycopersicum</i>	Solyc07g055920.3.1

18	Aminoácido	Sí (AG)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT4G18960.1
19	Aminoácido	Sí (AG)	Frambuesa negra	<i>Rubus occidentalis</i>	Ro03_G05331
20	Aminoácido	Sí (AG)	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_3g06720.1
21	Aminoácido	Sí (AG)	Durazno	<i>Prunus persica</i>	Prupe.4G070500.1
22	Aminoácido	Sí (AG)	Manzana	<i>Malus pumila</i>	MD05G1293700
23	Aminoácido	Sí (AG)	Manzana	<i>Malus pumila</i>	MD10G1271000
24	Aminoácido	Sí (AG)	Uva	<i>Vitis vinifera</i>	VIT_10s0003g02070.0
25	Aminoácido	Sí (AG)	Tomate	<i>Solanum lycopersicum</i>	Solyc02g071730.3.1
26	Aminoácido	No (AGL42/FYF)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT5G62165.2
27	Aminoácido	No (AGL71)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT5G51870.3
28	Aminoácido	No (AGL72)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT5G51860.1
29	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_16s0022g02400.0
30	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro07_G09508
31	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_7g28740.1
32	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.2G287500.1
33	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD00G1004500'
34	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD01G1193300'
35	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD01G1193500'
36	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD01G1193600'
37	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc03g006830.3.1
38	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc06g035560.2.1
39	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_02s0025g04650.0
40	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc08g080100.3.1

41	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc12g056460.2.1
42	Aminoácido	No (AGL19)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT4G22950.1
43	Aminoácido	No (AGL14/XAL2)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT4G11880.4
44	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro05_G14876
45	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_5g01240.1
46	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.5G087500.1
47	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD04G1070200
48	Aminoácido	No (AGL20/SOC1)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT2G45660.1
49	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_15s0048g01240.0
50	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc00g179240.2.1
51	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc10g017640.1.1
52	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro07_G21930
53	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_7g12700.1
54	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.2G151200.1
55	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD02G1197400
56	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD07G1123600
57	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro05_G31625
58	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc03g019710.3.1
59	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_17s0000g01230.0
60	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro05_G01909
61	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.5G149400.1
62	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD06G1140100

63	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD14G1155500
65	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD10G1055900
65	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD05G1049400
66	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_03s0097g00160.0
67	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_00s0211g00150.0
68	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.1G538700.1
69	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc12g088080.1.1
70	Aminoácido	No (AGL12/XAL1)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT1G71692.1
71	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc11g032100.2.1
72	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_18s0041g02140.0
73	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD08G1216000
74	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD15G1400400
75	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.1G548000.1
76	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_5g32850.1
77	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro05_G26898
78	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro05_G26897
79	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_16s0022g02330.0
80	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc01g093960.3.1
81	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_15s0048g01270.0
82	Aminoácido	No (AGL13)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT3G61120.1
83	Aminoácido	No (AGL6/RSB1)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT2G45650.1
84	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro07_G21927
85	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_7g12670.1
86	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.2G151000.1

87	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD02G1197600
88	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro06_G03786
89	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro04_G12651
90	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD10G1041100
91	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD05G1037100
92	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD17G1001300
93	Aminoácido	No (AGL79)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT3G30260.1
94	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_01s0011g00100.0
95	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc05g056620.2.1
96	Aminoácido	No (AGL7/AP1)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT1G69120.1
97	Aminoácido	No (AGL10/CAL1)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT1G26310.1
98	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_4g29600.1
99	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.1G290600.1
100	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD13G1059200
101	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD16G1058500
102	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_14s0083g01030.0
103	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc02g065730.2.1
104	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc02g089210.3.1
105	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro06_G28654
106	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_6g46410.1
107	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.3G249300.1
108	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD17G1065500
109	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_17s0000g04990.0

110	Aminoácido	No (AGL8/FUL)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT5G60910.1
111	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc03g114830.3.1
112	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc06g069430.3.1
113	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_5g13500.1
114	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.5G208500.1
115	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD06G1204400
116	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD14G1215700
117	Aminoácido	No (AGL3/SEP4)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT2G03710.1
118	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc04g005320.3.1
119	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc05g012020.3.1
120	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_01s0011g00110.0
121	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro04_G02528
122	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_4g29620.1
123	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.1G290500.1
124	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD13G1059300
125	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD16G1058600
126	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc12g038510.2.1
127	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc03g114840.3.1
128	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD16G1121900
129	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_17s0000g05000.0
130	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro05_G03365
131	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_5g13510.1
132	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.5G208400.1

133	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD14G1215600
134	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD06G1204100
135	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD06G1204300
136	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_14s0083g01050.0
137	Aminoácido	No (AGL4/SEP2)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT3G02310.1
138	Aminoácido	No (AGL2/SEP1)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT5G15800.2
139	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD09G1073900
140	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD17G1065400
141	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.3G249400.1
142	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro06_G04893
143	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_6g46420.1
144	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc02g089200.3.1
145	Aminoácido	No (AGL9/SEP3)	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AT1G24260.2
146	Aminoácido	No	Tomate	<i>Solanum lycopersicum</i>	Solyc05g015750.3.1
147	Aminoácido	No	Uva	<i>Vitis vinifera</i>	VIT_01s0010g03900.0
148	Aminoácido	No	Frambuesa negra	<i>Rubus occidentalis</i>	Ro04_G07112
149	Aminoácido	No	Frutilla (Silvestre)	<i>Fragaria vesca</i>	FvH4_4g23530.1
150	Aminoácido	No	Durazno	<i>Prunus persica</i>	Prupe.1G223600.1
151	Aminoácido	No	Manzana	<i>Malus pumila</i>	MD13G1121500

Como un ejemplo para identificar homólogos, ortólogos y/o parálogos de los genes de interés de caja MADS de *Arabidopsis*, la secuencia codificante de proteína para *AtSTK/AtAGL11* se obtuvo del genoma de *A. thaliana* (ID de gen AT4G09960 en la versión de ensamblaje TAIR10), y se realizaron búsquedas Blast utilizando la secuencia de codificación de proteína de *Arabidopsis* STK/AGL11 como una consulta contra genomas de las plantas florecientes. A partir de las búsquedas Blast, las secuencias de codificación subyacentes a cada acierto de Blast se extrajeron para el posterior alineamiento de múltiples secuencias y análisis filogenético. La misma metodología se aplicó a otros genes de la caja MADS, con el fin de identificar y recopilar información de secuencia sobre homólogos, ortólogos y/o parálogos de los genes de interés de la caja MADS de *Arabidopsis*.

Para identificar un número ampliado de ortólogos de genes de la caja MADS de especies dispares, se utilizó el portal Plant Comparative Genomics (Genómica Comparativa de Plantas) del Instituto Joint Genome Institute del Departamento de Energía, también conocido como Phytozome, que alberga 93 genomas ensamblados y anotados de 82 especies de plantas. Las secuencias de aminoácidos de los ortólogos de genes de caja MADS identificados de 7 especies de plantas se alinearon con MUSCLE. A partir de este análisis, se identificaron los residuos de aminoácidos de los factores de transcripción de la caja MADS que se conservan en todo el reino vegetal. Como se mencionó anteriormente, los factores de transcripción de la caja MADS del clado AGAMOUS tal como se definen en el presente documento comprenden por lo menos los siguientes tres dominios conservados: **(1)** motivo Y{A,S}NN en las posiciones 83–86 con respecto a STK de *Arabidopsis thaliana* (AT4G09960.3); **(2)** motivo YQQE{A,S}{A,S,N,K,T}KL{R,H}{Q,H,A,N,R}QI en las posiciones 116–127 con respecto a STK de *Arabidopsis thaliana* (AT4G09960.3); y **(3)** R en la posición 225 con respecto a STK de *Arabidopsis thaliana* (AT4G09960.3).

Para la reconstrucción de árboles filogenéticos, ModelFinder se usó para la selección del modelo de sustitución, e IQ-Tree, para la inferencia de árboles y la estimación de soporte de ramas con aproximación de arranque ultrarrápido. Se encontró evidencia de ortología al examinar la topología de los árboles para la agrupación de genes de múltiples especies hasta la exclusión de parálogos de las especies focales.

Al principio, se analizaron las proteínas que poseían los tres motivos caracterizados anteriormente de *Arabidopsis*, frambuesa negra y frutilla silvestre. Los factores de transcripción

del clado AG se agruparon en el mismo clado, denominado en este documento factores de transcripción de caja MADS "clado AGAMOUS" o "clado AG". Este clado AG comprende por lo menos cuatro especies, que incluyen: genes de caja MADS, Agamous (AG), Seedstick (STK) y Shatterproof 1 y 2 (SHP1/2). Se realizó otro análisis filogenético con las proteínas que comparten los tres motivos descritos anteriormente de siete especies de plantas: *Arabidopsis* (*Arabidopsis thaliana*), manzana (*Malus pumila*), frambuesa negra (*Rubus occidentalis*), uva (*Vitis vinifera*), durazno (*Prunus persica*), tomate (*Solanum lycopersicum*) y frutilla silvestre (*Fragaria vesca*). A partir de los análisis MSA y filogenéticos, se identificaron los factores de transcripción de la caja MADS clado AG como objetivos potenciales para crear fenotipos de desarrollo del fruto alterado. Estos fenotipos se crearán mediante la introducción de modificaciones genéticas en los objetivos.

Además, un grupo bien respaldado con genes que provengan de especies separadas sería indicativo de un conjunto de genes ortólogos. El resultado de los análisis filogenéticos muestra que el gen *STK* de frambuesa negra (es decir, Ro05_G22454), el gen *STK* de frutilla silvestre (es decir, FvH4_5g32540), y el gen *STK* de durazno (es decir, Prupe.1G549600) están en el mismo grupo con el gen *AtSTK/AtAGL11* (es decir AT4G09960). Además, los genes de *Arabidopsis* AG, *SHP1* y *SHP2* se agrupan cerca del grupo de genes *STK*, lo que indica que los ortólogos *STK* son un clado estrechamente relacionado con otros tres factores de transcripción AG, *SHP1* y *SHP2*, todos los cuales se clasifican en un clado superior llamado factor de transcripción de la caja MADS clado AG enseñado en la presente descripción. Además, se realizó un análisis synteny entre los ortólogos pronosticados utilizando la herramienta GEvo en la plataforma web de genómica comparativa CoGe proporcionada en el sitio web disponible públicamente, genomevolution.org/coge. Se analizaron 400 kb de región cromosómica flanqueante que rodea los genes focales, y se evaluó la presencia de alineamientos de secuencia en la vecindad del gen. El análisis de synteny entre el gen *STK* de *Arabidopsis* (AT4G09960) (cromosoma 4: 6236473 – 6240681 pb) y el gen *STK* de frambuesa negra (Ro05_G22454) (cromosoma 5: 34530529 – 34523413, que incluye regiones cromosómicas flanqueantes de cada gen, mostró que era probable que los genes ortólogos sean parte del mismo vecindario de genes, en orden, en los respectivos genomas.

Ejemplo 2: Identificación de sitios objetivo para mutaciones en factores de transcripción de caja MADS clado AGAMOUS para el desarrollo de fruto alterado.

A fin de identificar sitios objetivo para mutaciones en cada uno de los factores de transcripción de caja MADS clado AG caracterizados en el **Ejemplo 1**, se analizaron adicionalmente el MSA y los resultados filogenéticos. Enre los factores de transcripción de caja MADS clado AG, SEEDSTICK(STK) está estrechamente agrupado con otros factores de transcripción en *Arabidopsis*: AGAMOUS (AG), SHATTERPROOF1 (SHP1) y SHATTERPROOF2 (SHP2). Estos cuatro genes son reconocidos en la técnica por compartir funciones redundantes en el control del crecimiento de óvulos, la determinación de meristemos florales y la abscisión de semillas en *Arabidopsis thaliana*, y son necesarios para especificar la identidad de los óvulos. Estos cuatro objetivos son especies dentro del género definido más grande. Por lo tanto, los inventores propusieron que la introducción de mutaciones en los factores de transcripción del clado AG (por ejemplo, las cuatro especies objetivo mencionadas dentro del género) perturba sus funciones normales, lo que puede dar lugar a un desarrollo alterado de la fruta.

Como un ejemplo, las secuencias de aminoácidos de los ortólogos STK recolectados de 4 especies de plantas, *Arabidopsis thaliana*, *Rubus occidentalis* (mora) y *Fragaria vesca* (frutilla silvestre) se alinearon para identificar los residuos conservados entre los dominios de caja MADS como se muestra en la **FIG. 5** con cuatro dominios MIKC; un MADS (M), un dominio Intermedio (I), uno de Queratina (K) y un dominio C-terminal (C). En este ejemplo, el dominio M, el dominio I, el dominio K y el dominio C son objetivos de interés para modificaciones genéticas tales como sustitución, adición o delección por técnicas de ingeniería genética que se enseñan en la presente divulgación, incluido un sistema de edición de bases descrito a continuación. Además, el factor de transcripción clado AG tiene la región N-terminal frente a los dominios MIKC, que es otro objetivo para las modificaciones genéticas descritas en este documento. Asimismo, entre los motivos altamente conservados de los factores de transcripción del clado AG seleccionados, el motivo NN (por ejemplo, 85–86 de SEQ ID NO: 13 (AtSTK) o 59–60 de SEQ ID NO: 97 (RoSTK), motivo YQQ (por ejemplo, 116–118 de SEQ ID NO: 13 (AtSTK) o 91–93 de SEQ ID NO: 97 (RoSTK)), o SEQ ID NOs: 29, 78, 97, 53, 66, 128, 150, 152, 154, 156, 158, 159, 160, 162, 164, 166–171), o una posición R conservada (por ejemplo, 225 de SEQ ID NO: 13; posición 196 de SEQ ID NO: 66; posición 197 de SEQ ID NOs: 53, 78, 150, 166–171; posición 198 de SEQ ID NO: 29; posición 199 de SEQ ID NO: 128; o la posición 201 de SEQ ID NOs: 97, 152,

154, 156, 158, 160, 162, 164) se seleccionan como objetivos de interés para la introducción de mutaciones enseñadas en la presente divulgación usando las técnicas de ingeniería genética de la presente descripción.

En algunas formas de realización, la presente invención proporciona la introducción de un truncamiento C-terminal de un polipéptido STK. Se puede introducir un truncamiento C-terminal, por ejemplo, mediante el direccionamiento del exón final de un polinucleótido STK como se describe en este documento. Una mutación dirigida, que incluye, por ejemplo, una o más deleciones o sustituciones, puede producir un codón de parada prematuro que produce un truncamiento C-terminal. En algunos aspectos, como mínimo, un truncamiento C-terminal útil para esta invención elimina por lo menos la arginina conservada ubicada en la posición 225 de SEQ ID NO: 13 (AtSTK), 196 de SEQ ID NO: 66; posición 197 de SEQ ID NOs: 53, 78, 150, 166–171; 198 de SEQ ID NO: 29; 199 de SEQ ID NO: 128 o 201 de SEQ ID NO: 97 (RoSTK)) o SEQ ID NOs: 152, 154, 156, 158, 160, 162, 164). Ejemplos de mutaciones que producen un truncamiento C-terminal se proporcionan en la **Fig. 5**.

Los codones que codifican los residuos seleccionados para el direccionamiento se analizaron para posibles ediciones, incluidos los truncamientos C-terminales, que podrían afectar un cambio no conservador en el aminoácido, lo que produciría una variación en la función de la proteína. El motivo NN extendido, Y[A/S]NN, y otro motivo YQQ extendido, YQQE[A/S][A/S/N/K/T]KL[R/H][Q/H/A/N/R]QI, también se identificaron como objetivos potenciales para editar uno o más aminoácidos para alterar el desarrollo del fruto en plantas de interés. Hay otras secuencias de aminoácidos bien conservadas, como NN, A, D, YYQ, R, Q, N, R, R, que se conservan a más del 95% en los factores de transcripción del clado AG (por ejemplo, STK) a través de todo el reino vegetal, como se muestra en la **FIG. 5** y las **Tablas 1, 1B, 3**. La conservación de secuencia entre plantas relacionadas en forma distante sugiere una presión selectiva para mantener la función. Las mutaciones que causan un tamaño de semilla reducido o inducen falta de semilla serían una fuerte presión selectiva. Por ejemplo, como estos aminoácidos están en el dominio C-terminal y están conservados, estas secuencias son candidatas para la mutación a fin de investigar la función molecular del alelo STK en plantas de interés. Mediante el uso de herramientas genéticas para inducir mutaciones que afecten a estos residuos, los ARNg se diseñarán, como se describe a continuación, para introducir una variación de aminoácido único, que provoca un cambio conformacional de la proteína STK inducido por la sustitución y/o el

truncamiento de aminoácidos únicos. La **Tabla 3** presenta la ubicación de caja MADS conservada y específico de AG de **SEQ ID NO: 97**.

Tabla 3. Posición de los motivos identificados en la SEQ ID 97.

Nombre	AA inicio	AA fin	Longitud
Específico de Ag	115	117	3
Específico de Ag	85	86	2
Específico de Ag	98	98	1
Específico de Ag	124	124	1
Específico de Ag	126	126	1
Específico de Ag	132	132	1
Específico de Ag	182	182	1
Específico de Ag	225	225	1
Específico de Ag	101	101	1
Caja MADS conservada	43	47	5
Caja MADS conservada	49	51	3
Caja MADS conservada	56	58	3
Caja MADS conservada	29	30	2
Caja MADS conservada	53	54	2
Caja MADS conservada	64	65	2
Caja MADS conservada	74	75	2
Caja MADS conservada	37	37	1
Caja MADS conservada	39	39	1
Caja MADS conservada	60	60	1
Caja MADS conservada	62	62	1
Caja MADS conservada	67	67	1
Caja MADS conservada	78	78	1
Caja MADS conservada	150	150	1

Estas ediciones/mutaciones candidatas identificadas se criban luego en un ensayo HTP *in vivo* descrito a continuación para seleccionar las ediciones que afectan la transcripción. Las ediciones seleccionadas, que se documentan como "aciertos" en el cribado, luego se introducen en las plantas de interés, con el fin de cribar fenotípicamente para hallar un fenotipo de desarrollo de la fruta alterado, particularmente un fenotipo sin semillas o con reducida simiente.

Ejemplo 3: Cribado HTP y selección de mutaciones efectivas en factores de transcripción de caja MADS clado AGAMOUS P.

Se usa un sistema de ensayo *in vivo* para analizar la expresión transitoria de luciferasa en células HEK293T para cribar sitios objetivo adecuados para mutar los factores de transcripción del clado AG, con el fin de producir eventualmente un tamaño de semilla reducido o un rasgo sin semillas en frutos de las plantas de interés.

En general, se sabe que los factores de transcripción de la caja MADS se dirigen a un sitio de unión de promotor particular, que es la caja CArG (Honma y Goto, 2001, Nature 409: 525–529). Por ejemplo, los factores de transcripción de la caja MADS pueden unirse a un promotor sintético de la caja CArG (7xCaRG::LUC) que tiene siete repeticiones de secuencia de consenso de unión a proteína MADS ((5WGGGGTGGCTTTCCTTTTTGGTAAATTTGGATCC–3W, la caja CArG está subrayada), corriente ascendente del promotor mínimo. La construcción que tenía siete repeticiones de la secuencia consenso de unión a la proteína MADS se usó para permitir que el tetrámero de factores de transcripción de la caja MADS se uniera a múltiples cajas CArG.

Al principio, el promotor de la caja 7xCaRG se inserta corriente ascendente del gen reportero de luciferasa de luciérnaga (LUC) en el vector utilizado para un sistema dual de luciferasa. El sistema dual de luciferasa 7xCaRG::LUC permite la expresión de (i) el promotor de la caja 7xCaRG que controla la expresión del gen reportero luciferasa de luciérnaga (LUC) y (ii) un promotor de control (CMV) que regula la expresión del gen reportero de luciferasa de *Renilla* (REN). A partir de este experimento, el gen reportero LUC se transcribe y se expresa cuando un factor de transcripción de la caja MADS apunta y se une a la región promotora de la caja 7xCaRG. El nivel de expresión del gen reportero LUC depende del grado de unión y la capacidad del factor de transcripción de la caja MADS de interés para modular la transcripción. Por lo tanto, la efectividad relativa de diferentes mutaciones del factor de transcripción clado AG en diversos sitios (p. ej., dominio N-terminal, M, I, K o C, o tres motivos descritos en el **Ejemplo 2**) se determina en la estimulación de la expresión de una secuencia del gen reportero LUC controlada por el promotor de la caja 7xCaRG (Hellens et al. 2005, Plant Methods 1:13) en comparación con la expresión constitutiva del gen reportero REN.

En ausencia de factores de transcripción de la caja MADS (TF), la relación LUC a REN es baja. Es probable que este nivel de actividad de fondo represente niveles iniciales de TF de caja MADS presentes en las células HEK293T. La adición de un TF clado AG tal como STK, AG o SHP1/2 no mutado a una mezcla de transfección para la expresión de luciferasa en células

HEK293T aumenta el nivel relativo de actividad de LUC en comparación con la actividad de promotor de fondo en ausencia de plásmido que contenga el TF 1 de clado AG en la mezcla de infiltración. Se predice que diferentes mutaciones introducidas en los TF clado AG (por ejemplo, STK, AG y/o SHP1 / 2) dan un nivel relativo diferente de actividad de LUC en comparación con la actividad del promotor de REN de fondo.

Luego, el efecto de las mutaciones en la proteína TF clado AG se cuantifica por la relación LUC:REN. Este ensayo se realiza en células HEK293T. Una vez que los vectores de expresión capaces de sobreexpresar los genes TF clado AG mutados candidatos (por ejemplo, mutaciones en la región N-terminal, dominio M, dominio I, dominio K, dominio C, el motivo NN extendido, el motivo YQQ extendido o el residuo R/Q en C-terminal) se construyen junto con el sistema de luciferasa dual 7xCARg::LUC preparado anteriormente, las células HEK293T se cotransfectan con los vectores; (1) un vector de expresión que posee los genes TF clado AG mutados candidatos (por ejemplo, mutación negativa dominante) bajo el control del promotor CMV, junto con (2) el vector para el sistema de ensayo dual de luciferasa 7xCARg::LUC. Las secuencias de promotor de la caja 7xCARg activadas por la presencia del factor de transcripción del clado AG se fusionan con LUCIFERASA de luciérnaga (LUC), y la transactivación del gen LUC se mide en relación con CMV:luciferasa de RENILLA (REN) mediante la medición de la luminiscencia después de la expresión en células HEK293T. La luminiscencia se analiza dos días después de la transfección. Los ensayos sobre la base de placas se realizan utilizando la función de luminómetro de un lector de microplacas multimodo SpectraMax iD3 (Molecular Devices) de acuerdo con las especificaciones del fabricante para el ensayo de luciferasa dual, utilizando los reactivos de ensayo Dual Glow (Promega) para luciferasa de luciérnaga y luciferasa de *Renilla*. La luminiscencia se calculó utilizando el soporte lógico SoftMax® Pro GxP (Molecular Devices). Los resultados se esperan de la siguiente manera, a la luz de los datos de Espley et al. 2009 Plant Cell 21: 168–183. El promotor de la caja 7xCARg mostrará poca diferencia en la actividad, según lo determinado por la relación de luminiscencia producida por el promotor de la caja 7xCARg–LUC a CMV–REN sin cotransfección del TF clado AG. Cuando el TF clado AG se cotransfecta con las construcciones de promotor de caja 7xCARg–LUC, se transactivarán los promotores. Cuando la construcción de TF clado AG mutada se cotransfecta con las construcciones de luciferasa dual fusionada a caja 7xCARg, habrá una reducción en la transactivación, porque la afinidad de unión o la transactivación del complejo de caja MADS con el (los) TF (s) clado AG mutados se verán

afectadas negativamente por las mutaciones hasta cierto grado. Por lo tanto, la transactivación se reducirá en la versión mutada del TF clado AG en comparación con la versión no mutada. La relación LUC:REN con las cotransfecciones de la versión mutada de los TF del clado AG será menor que la versión no mutada. Si la relación LUC:REN en la versión mutada es inferior a la versión no mutada de los TF del clado AG de interés, el sitio de mutación será un objetivo de interés (es decir, un "acierto") para el estudio *in vivo*, de manera de crear una línea de planta estable con una mutación específica introducida por una técnica de edición de base única estable pero dirigida que se enseña en la presente descripción.

Por lo tanto, la metodología anteriormente mencionada proporciona un cribado rápido y HTP *in vivo*, en un organismo modelo, para mutaciones candidatas en factores de transcripción de caja MADS clado AGAMOUS. Si hay una reducción en la unión del factor de transcripción evidente en el cribado, es decir, una indicación de que una mutación podría servir como un alelo negativo dominante, entonces dicha mutación es dirigida luego en una planta de interés, que finalmente se fenotipifica para confirmar el "acierto" positivo a partir del cribado, e identificar un alelo negativo dominante asociado con el desarrollo alterado de la fruta.

El ensayo anterior se realizó para SEEDSTICK como se describe en el Ejemplo 8 y las Figs. 7–11.

Ejemplo 4: Construcción de vectores de edición de bases.

Una vez que los sitios objetivo para las mutaciones específicas de sitio en los TF del clado AG candidatos (por ejemplo, STK, AG y SHP1/2) se criban y verifican mediante el ensayo de infiltración de *Agrobacterium* sobre la base de luciferasa *in vivo* descrito en el **Ejemplo 3**, se construyen los vectores de edición de bases para las mutaciones dirigidas en los genes candidatos (por ejemplo, STK, AG y SHP1/2) encontrados en plantas de interés, para la transformación de los vectores de edición de bases en la planta de interés, incluidas, entre otras: frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, *Arabidopsis*, manzana, tomate, uva y durazno.

Los vectores de expresión del editor de base de citosina (CBE) y los vectores de expresión del editor de base de adenina (ABE) se generarán para introducir en los genes candidatos (genes del factor de transcripción clado AG y sus ortólogos en plantas) mutaciones que causan un desarrollo alterado de la fruta, tal como reducción de tamaño de semilla o disminución de la simiente, sobre la base de la información proporcionada en el **Ejemplo 3**. Al usar herramientas

genéticas para introducir mutaciones para efectuar estos residuos como se describe en el **Ejemplo 3**, los ARNg se diseñarán en consecuencia para introducir una variación de un solo aminoácido, que causa un cambio conformacional de la proteína TF clado AG inducido por sustitución y/o truncamiento de un solo aminoácido. La modificación genética introducida por las técnicas de edición de bases dará como resultado la sustitución, adición, delección y/o truncamiento de aminoácidos. El truncamiento tendrá lugar en la región C-terminal una vez que la secuencia del codón se convierta en un codón de parada. Entonces, toda la secuencia peptídica corriente descendente del codón de parada recién introducido se deleteará/truncará en consecuencia.

Direccionamiento de vectores de corte para realizar un truncamiento/delección C-terminal.

Los vectores de expresión de Cas12a se generarán para introducir en los genes candidatos (p. ej., genes del factor de transcripción clado AG y sus ortólogos en plantas) mutaciones que causan un desarrollo alterado del fruto, como un tamaño de semilla reducido o una simiente reducida, según la información proporcionada en el **Ejemplo 3**. Mediante el uso de herramientas genéticas para introducir mutaciones para efectuar estos residuos como se describe en el **Ejemplo 3**, los ARNg se diseñarán en consecuencia para introducir delecciones de pares de bases, que causan un cambio conformacional de la proteína TF clado AG inducido por truncamiento o delecciones en marco. La modificación genética introducida por las delecciones de par de bases dará como resultado la delección y/o el truncamiento de aminoácidos. El truncamiento tendrá lugar en la región C-terminal una vez que la delección provoque un desplazamiento de marco en la secuencia de codificación, y esto haga que el codón se convierta en un codón de parada. Luego, toda la secuencia peptídica corriente descendente del codón de parada recién introducido se deleteará/truncará en consecuencia.

1. Construcción del vector de expresión del editor de base citosina (CBE).

Las citosina desaminasas (CD) típicamente desaminan las citosinas en sitios específicos en el ADN monocatenario, de modo que la desaminación de la citosina (C) es catalizada por las citidina desaminasas y da como resultado el uracilo (U), que tiene las propiedades de emparejamiento de bases de la timina (T). En este ejemplo, CD se fusiona con CRISPR/Cas9 de tipo II deficiente en nucleasas para lograr la desaminación de citosina guiada por ARN en ADN genómico, y CD edita una hebra de ADN no dirigida desplazada por la unión de un complejo de

ARN guía Cas9 a una hebra de ADN dirigida. Para inducir la sustitución de un solo nucleótido, se usa Cas9 catalíticamente muerta/deficiente (dCas9), cuya actividad de nucleasa se pierde, o se usa la variante Cas9 que tiene actividad de nickasa (nCas9) para la proteína de fusión Cas9–CD. Mientras que dCas9 o nCas9 se unen a su objetivo a través del ARN guía y forman la burbuja entre el ARN guía y su ADN complementario, la enzima CD actuará entonces sobre la citosina en el ADN monocatenario accesible en la burbuja no coincidente. Cuando se utiliza nCas9, la nickasa Cas9 cortará la hebra de ADN no dirigida y facilitará la conversión posterior de la guanina correspondiente a adenina en la hebra de ADN no dirigida.

En una forma de realización, se puede usar otro sistema CRISPR deficiente en nucleasas, como la nickasa Cpf1, C2c1 deficiente en nucleasas o Cms1 deficiente en nucleasas para mellar la hebra de ADN no dirigida para su necesidad. En este ejemplo, Cas9, CasX, CasY, Cpf1, C2c1, C2c2, C2c3 o Cms1, cada uno de los cuales es deficiente en nucleasas, pueden usarse para la sustitución de una base única junto con las enzimas de edición de bases citidina desaminasa y/o adenosina desaminasa. Para la construcción del vector de expresión CBE, las secuencias APOBEC1, XTEN, nCas9(D10A) o dCas9 (D10A y H840A) y UGI se optimizarán con codones para las plantas de interés. Se puede fusionar un inhibidor de uracilo glicosilasa (UGI) con nCas9 o dCas9 para inhibir la enzima de reparación de escisión de base ADN uracilo–glicosilasa y aumentar la eficiencia de la edición de bases.

Diseños para el vector de expresión CBE:

- nCas9, dCas9, Cpf1 nickasa, C2c1 con deficiencia de nucleasa o Cms1 con deficiencia de nucleasa se fusionan al CD en el extremo N o C–terminal.
- La secuencia de localización nuclear (NLS) se puede etiquetar a nCas9, nCas9, dCas9, Cpf1 nickasa, C2c1 deficiente en nucleasa o Cms1 deficiente en nucleasa.
- Se utiliza una diversidad de enlazadores, incluidos enlazadores flexibles como XTEN o enlazadores menos flexibles.
- Un inhibidor de uracilo glicosilasa (UGI) se puede unir a la nCas9, dCas9, Cpf1 nickasa, C2c1 deficiente en nucleasa o Cms1 deficiente en nucleasa.
- El andamio de ARN guía puede insertarse en el vector de expresión de CBE o puede expresarse en un vector de expresión de ARN guía separado.

2. Construcción del vector de expresión de editor de base Adenina (ABE).

Las adenosina desaminasas (AD) típicamente desaminan la adenina en un residuo de ADN de desoxiadenosina en sitios específicos en el ADN monocatenario, de modo que la desaminación de la adenosina (A) es catalizada por las adenosina desaminasas y da como resultado la inosina, que se empareja en base como guanina (G) en el contexto de ADN y tiene las propiedades de emparejamiento de bases de Guanina (G). En este ejemplo, AD se fusiona con CRISPR/Cas9 tipo II deficiente en nucleasas para lograr la desaminación de citosina guiada por ARN en ADN genómico, y AD edita una hebra de ADN no dirigida desplazada por la unión de un complejo de ARN guía Cas9 a una hebra de ADN dirigida. Para inducir la sustitución de un solo nucleótido, se usa Cas9 catalíticamente muerta/deficiente (dCas9) cuya actividad de nucleasa se pierde, o se usa la variante Cas9 que tiene actividad de nickasa (nCas9), para la proteína de fusión Cas9–AD. Mientras que dCas9 o nCas9 se unen a su objetivo a través del ARN guía y forman la burbuja entre el ARN guía y su ADN complementario, la enzima AD actuará sobre la adenosina en el ADN monocatenario accesible en la burbuja no coincidente. Cuando se utiliza nCas9, la nickasa Cas9 cortará la hebra de ADN no dirigida y facilitará la conversión posterior de la timina correspondiente a citosina en la hebra de ADN no dirigida.

En algunas formas de realización, pueden usarse otros sistemas deficientes en nucleasas CRISPR tales como Cpf1 nickasa, C2c1 deficiente en nucleasas o Cms1 deficiente en nucleasas para cortar la hebra de ADN no dirigida. En este ejemplo, Cas9, CasX, CasY, Cpf1, C2c1, C2c2, C2c3 o Cms1, cada uno de los cuales es deficiente en nucleasas, pueden usarse para la sustitución de una base única junto con las enzimas de edición de bases citidina desaminasa y/o adenosina desaminasa.

Para la construcción del vector de expresión ABE, el TadA de *E. coli* (ecTadA) que incluye truncamientos/mutaciones de adenosina desaminasa, ADAR humano que incluye hADAR1, hADAR2, hADAR3 y sus versiones mutadas, XTEN, nCas9 (D10A) o dCas9 (D10A y H840A), y las secuencias UGI serán optimizadas por codones para plantas de interés. Se puede fusionar un inhibidor de uracilo glicosilasa (UGI) con nCas9 o dCas9 para inhibir la enzima de reparación de escisión de base de uracilo–ADN glicosilasa y aumentar la eficiencia de la edición de bases.

Diseños para el vector de expresión ABE:

– nCas9, dCas9, Cpf1 nickasa, C2c1 deficiente en nucleasa, C2c2, C2c3 o Cms1 deficiente en nucleasa se fusiona a CD en el extremo N o C–terminal.

- La secuencia de localización nuclear (NLS) se puede etiquetar a nCas9, nCas9, dCas9, Cpf1 nickasa, C2c1 deficiente en nucleasa o Cms1 deficiente en nucleasa.
- Se utiliza una diversidad de enlazadores, incluidos enlazadores flexibles como XTEN o enlazadores menos flexibles.
- Un inhibidor de uracilo glicosilasa (UGI) se puede unir a nCas9, dCas9, Cpf1 nickasa, C2c1 deficiente en nucleasa, C2c1, C2c2, C2c3 o Cms1 deficiente en nucleasa.
- El andamio de ARN guía puede insertarse en el vector de expresión de CBE o expresarse en un vector de expresión de ARN guía separado.

3. Diseños para andamios de ARNg.

El sistema de guía dependerá del sistema CRISPR utilizado. Si es Cpf1, es una guía natural "única", pero con Cas9 puede ser única o dual/nativa con un ARNcr separado ARNtracr. Los vectores de edición de bases que se dirigen al gen TF clado AG se construirán adicionalmente insertando fragmentos de ARN de guía único que contienen un sitio único objetivo para la secuencia de Señal de Localización Nuclear (NLS) del dominio de caja MADS, dominio de caja MADS, dominio Intermedio, dominio de caja K, Dominio C-terminal y/o tres motivos (por ejemplo, el motivo NN extendido, el motivo YQQ extendido, el residuo R/Q en el extremo C-terminal). Una vez que el ARN guía conduce la proteína de fusión Cas/CD a los sitios objetivo en el dominio de la caja MADS del TF clado AG, la citosina ubicada en la ventana de edición de bases, que es 4–8 nt distal a PAM, se desaminará y se convertirá en Uracilo, es decir, las propiedades de emparejamiento de base de Timina en el sistema CBE. Por otro lado, una vez que el ARN guía conduce la proteína de fusión Cas9/AD a los sitios objetivo en el dominio de la caja MADS del TF clado AG, Adenina que se encuentra en la ventana de edición de base, que es 4–8 nt distal a PAM, será desaminada y convertida en inosina que está lista como guanina, por polimerasa en el sistema ABE.

(i) El ARNg que se dirige a la región N-terminal que incluye NLS guiará la proteína de fusión Cas/CD al sitio NLS para una sustitución de nucleótido único C> T o G> A. El ARNg que se dirige a la región N-terminal que incluye NLS guiará la proteína de fusión Cas9/AD al sitio NLS para una sustitución de nucleótido único A> G o T> C. La mutación del NLS evitará que el factor de transcripción de la caja MADS sea transportado al núcleo, lo que da como resultado una regulación errónea de genes de identidad de otros órganos en la corriente descendente y defectos en la especificación de la identidad del meristemo floral y la formación de órganos.

(ii) Los ARNg que se dirigen al dominio de la caja MADS guiarán la proteína de fusión Cas/CD al sitio del dominio M para una sustitución de nucleótidos único C> T o G> A. Los ARNg que se dirigen al dominio de la caja MADS guiarán la proteína de fusión Cas/AD al sitio del dominio M para una sustitución de un solo nucleótido A> G o T> C. El dominio de la caja MADS está involucrado en la unión del ADN, y también puede estar involucrado en la dimerización, cuya función se verá afectada por estas modificaciones.

(iii) Los ARNg que se dirigen al dominio I guiarán la proteína de fusión Cas/CD al sitio del dominio I para una sustitución de nucleótido único C> T o G> A. Los ARNg que se dirigen al dominio I guiarán la proteína de fusión Cas/AD al sitio del dominio I para una sustitución de un solo nucleótido A> G o T> C.

(iv) Los ARNg que se dirigen al dominio de la caja K guiarán la proteína de fusión Cas/CD al sitio del dominio K para una sustitución de nucleótido simple C> T o G> A. Los ARNg que se dirigen al dominio de la caja K guiarán la proteína de fusión Cas/AD al sitio del dominio K para una sustitución de un solo nucleótido A> G o T> C. El dominio de la caja K está involucrado en la dimerización de proteínas, cuya función se verá afectada por estas modificaciones.

(v) Los ARNg que se dirigen al dominio C-terminal guiarán la proteína de fusión Cas/CD al sitio del dominio C para una sustitución de nucleótido único C> T o G> A. Los ARNg que se dirigen al dominio C terminal guiarán la proteína de fusión Cas/AD al sitio del dominio C para una sustitución de un solo nucleótido A> G o T> C. El dominio C-terminal funciona como un dominio de activación transcripcional, cuya función se verá afectada por estas modificaciones.

La **FIG. 6** ilustra tres ejemplos de ARN guía (S1 [CCGCAATCATAGAGCGTTAA **SEQ ID NO: 184**], S2 [CGCAATCATAGAGCGTTAAAG **SEQ ID NO: 185**], y S3 [CGCAATTTCTTCACATCCAC **SEQ ID NO: 186**]) diseñados para dirigirse al codón de Arginina (R) en el dominio C-terminal de la proteína STK de frambuesa negra (RoSTK). Los ARN guía S1 y S2, por ejemplo, pueden desencadenar la sustitución de aminoácido de Arginina (R) a Histidina (H) por cambio de nucleótido C a T (o G a A) por el editor de base de citosina. El ARN guía S3 puede inducir la sustitución de Arginina (R) a Cisteína (C) por sustitución de nucleótido C a T (o G a A) por el editor de base de citosina. (1) Arginina a Histidina usando ARNg S1/S2 por un cambio de par de bases de nucleótidos G a A o (2) Arginina a Cisteína usando ARNg S3 por un cambio de par de bases de nucleótidos C a A. Espaciadores adicionales para STK de frambuesa negra incluyen **SEQ ID NOs: 175–176 o 184–186**.

Se usa la secuencia PAM apropiada para el diseño de ARNg, que depende del sistema Cas utilizado. La fuente/tipo de proteína Cas9 como las variantes de Cas9 tienen diferentes requisitos de PAM para el reconocimiento del objetivo. Se pueden determinar los sitios PAM en un gen de interés. Como ejemplo, el sitio PAM reconocido por Cas9 es NGG, Cpf1 (Cas12a) reconoce TTTN, y C2c1/Cms1 reconoce TTN.

En algunas formas de realización, la longitud de la guía de ARN (17–24 nt) está diseñada para apuntar al locus genómico de interés. En este ejemplo, RoSTK es un gen objetivo del complejo de proteína de fusión Cas–CD o Cas–AD/ARNg. La guía de ARN con una longitud más larga que la canónica puede usarse para formar heteroduplas fuera del complejo de proteína–ARN guía–ADN objetivo.

Ejemplo 5: Transformación con vectores de edición de bases en plantas de interés.

Los vectores de edición de bases descritos en el Ejemplo 4 se transformarán en plantas de interés tales como: frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, *Arabidopsis*, manzana, tomate, uva y durazno. La divulgación enseña todos los tipos de métodos de transformación, incluido el uso de protocolos mediados por agrobacterias que son conocidos en la técnica y/o desarrollados por los inventores, así como métodos de transformación biolística. El cultivo de tejidos y la regeneración de plantas transformadas se realizarán en consecuencia.

Ejemplo 6: Análisis molecular para plantas transformadas.

En este ejemplo, se examina la actividad de los vectores de edición de bases descritos en el **Ejemplo 4** en las plantas de interés transformadas que incluyen *Arabidopsis*, frambuesa negra, mora, cereza, palta, frutilla, frutilla silvestre, durazno, uva, manzana y/o tomate. Como control para la inducción de indel, en este experimento también se usa una construcción que expresa Cas9 de tipo salvaje. La edición de base en un sitio objetivo se examina en todas las plantas transformadas. La ventana de desaminación abarca las posiciones 4–8 nt de la secuencia de protoespaciador distal a PAM. La frecuencia de una sustitución de base única se estudia mediante amplificación por PCR y secuenciación posterior de Sanger y/o secuenciación profunda de alto rendimiento. Además, las sustituciones de aminoácidos causadas por la técnica de edición de base única introducida en la región del genoma objetivo se analizan mediante secuenciación de proteínas con espectrometría de masas.

Ejemplo 7: Observación/puntuación de rasgos fenotípicos de plantas transformadas.

Las plantas transformadas obtenidas del Ejemplo 6 se cultivan en el invernadero controlado y/o en condiciones de campo. Las plantas transformadas, verificadas con la sustitución de aminoácido de interés, se observan para el desarrollo de frutos alterado, tal como un fenotipo sin semillas o con reducida simiente y/o un endocarpio reducido en los frutos de las plantas. Por ejemplo, el ensayo/métrica descrito anteriormente se utiliza para definir un fenotipo de desarrollo de fruta alterado.

Ejemplo 8: Ensayos de actividad negativa dominante de genes *SEEDSTICK* (STK) mutados.

Usando la técnica descrita en el Ejemplo 3, se llevaron a cabo ensayos para evaluar la actividad negativa dominante de los genes *SEEDSTICK* (STK) en varias especies de plantas diferentes. Los genes fueron optimizados por codones para la expresión en las células humanas (HO). Los resultados son visibles en las **FIGS. 7–11**. En cada caso, el control comprende dos genes *SEPALLATA* de tipo salvaje y dos genes STK de tipo salvaje (véase, por ejemplo, (FIG. 12A). Las mutaciones evaluadas fueron truncamientos y mutaciones C–terminales negativos dominantes (etiquetados "dom" en las **FIGS. 7– 11**) en combinación con los genes *SEPALLATA* de tipo salvaje y un gen STK de tipo salvaje (véase, por ejemplo, la **Figura 12B**) Como se muestra en las **Figuras 7–11**, la introducción del gen STK mutado reduce la actividad del complejo SEP3–STK y reduce la expresión del reportero. Estos datos confirman que los alelos STK mutados todavía pueden unirse y formar un complejo con SEP3 y STK de tipo salvaje, pero que la expresión causada por el complejo se reduce.

Ejemplo 9. Modificación de genes STK en frutilla silvestre.

El gen STK de frutilla silvestre (*Fragaria vesca*) (que codifica el polipéptido de **SEQ ID NO: 29**) se modificó mediante el uso de un vector de corte como se describe en el Ejemplo 4. La **FIG. 13** muestra una alineación de STK de frutilla silvestre (*Fragaria vesca*) con la STK mutante que se generó.

Se diseccionaron manualmente los aquenios de los frutos de tipo silvestre y de los mutantes FvSTKdelta4bp heterocigotos a los 4-5 días después de la polinización. La semilla se extrajo del aquenio y se comparó con un microscopio de disección. La semilla de mutantes FvSTKdelta4bp heterocigotos presentó un aspecto más oscuro en comparación con la semilla de tipo salvaje. El aspecto más oscuro de la semilla puede corresponder al aborto del embrión debido a la presencia de la mutación FvSTKdelta4bp heterocigota. La fecundación y el aborto de la semilla es una característica de la mutación SEEDSTICK en la uva. Además, el aspecto más oscuro se produce en el estado heterocigota, lo que sugiere que la mutación FvSTKdelta4bp es una mutación dominante.

Otras formas de realización numeradas de la divulgación.

Otro objeto del asunto contemplado por la presente divulgación se expone en las siguientes formas de realización numeradas:

1. Una célula vegetal, que comprende: a. un sistema de edición de bases capaz de modificar un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS e introducir una sustitución de nucleótido única estable y dirigida en dicho gen, donde el sistema de edición de bases comprende: (i) una proteína de fusión de edición de bases, que comprende: una proteína efectora asociada a CRISPR y una citidina desaminasa o adenosina desaminasa; y (ii) un ARN guía (ARNg) capaz de dirigir la proteína de fusión al gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS.
2. La célula vegetal de la forma de realización 1, en donde el sistema de edición de bases introduce una sustitución de nucleótido única estable y dirigida C>T o G>A o A>G o T>C en el gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS.
3. La célula vegetal de cualquiera de las formas de realización 1–2, en donde el gen objetivo endógeno codifica un factor de transcripción de caja MADS clado AGAMOUS involucrado en la formación del fruto.
4. La célula vegetal de cualquiera de las formas de realización 1–3, en donde el gen objetivo endógeno codifica un factor de transcripción de caja MADS clado AGAMOUS involucrado en la formación del endocarpio.

5. La célula vegetal de cualquiera de las formas de realización 1–4, en donde el factor de transcripción de caja MADS clado AGAMOUS se selecciona del grupo que consiste en: AG, SHP1, SHP2, y STK.
6. La célula vegetal de cualquiera de las formas de realización 1–5, en donde el factor de transcripción de caja MADS clado AGAMOUS es STK.
7. La célula vegetal de cualquiera de las formas de realización 1–6, en donde el ARNg comprende una secuencia guía que se une a un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS que comparte por lo menos 90% de identidad de secuencia con **SEQ ID NO: 13**.
8. La célula vegetal de cualquiera de las formas de realización 1–7, en donde el sistema de edición de bases se adapta para crear una sustitución de nucleótido único estable y dirigida C>T o G>A o A>G o T>C en la región del gen objetivo endógeno que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS.
9. La célula vegetal de cualquiera de las formas de realización 1–8, en donde el sistema de edición de bases se adapta para crear una sustitución de nucleótido único estable y dirigida C>T o G>A o A>G o T>C en la región del gen objetivo endógeno que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS, en donde, con la expresión del gen, el factor de transcripción de caja MADS clado AGAMOUS resultante tiene alterada expresión y/o función.
10. La célula vegetal de cualquiera de las formas de realización 1–9, en donde una célula vegetal es de la familia *Rosaceae*.
11. La célula vegetal de cualquiera de las formas de realización 1–10, en donde una célula vegetal es de una planta seleccionada del grupo que consiste en: frambuesa, frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, manzana, tomate, uva, durazno, ciruela, damasco, pera, membrillo, níspero y almendra.
12. La célula vegetal de cualquiera de las formas de realización 1–11, en donde una célula vegetal es de *Rubus occidentalis*.
13. Una célula vegetal con un gen objetivo endógeno modificado que codifica un factor de transcripción de caja MADS clado AGAMOUS, que comprende: una sustitución de nucleótido único estable y dirigida en un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS.

14. La célula vegetal de la forma de realización 13, que comprende una sustitución de un solo nucleótido estable y dirigida C>T o G>A o A>G o T>C en el gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS.
15. La célula vegetal de cualquiera de las formas de realización 13–14, en donde el gen objetivo endógeno codifica un factor de transcripción de caja MADS clado AGAMOUS involucrado en la formación del fruto.
16. La célula vegetal de cualquiera de las formas de realización 13–15, en donde el gen objetivo endógeno codifica un factor de transcripción de caja MADS clado AGAMOUS involucrado en la formación del endocarpio.
17. La célula vegetal de cualquiera de las formas de realización 13–16, en donde el factor de transcripción de caja MADS clado AGAMOUS se selecciona del grupo que consiste en: AG, SHP1, SHP2, y STK.
18. La célula vegetal de cualquiera de las formas de realización 13–17, en donde el factor de transcripción de caja MADS clado AGAMOUS es STK.
19. La célula vegetal de cualquiera de las formas de realización 13–18, que comprende una sustitución de un solo nucleótido estable y dirigida C>T o G>A o A>G o T>C en la región del gen objetivo endógeno que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS.
20. La célula vegetal de cualquiera de las formas de realización 13–19, que comprende una sustitución de un solo nucleótido estable y dirigida C>T o G>A o A>G o T>C en la región del gen objetivo endógeno que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS, en donde, con la expresión del gen, el factor de transcripción de caja MADS clado AGAMOUS resultante tiene alterada expresión y/o función.
21. La célula vegetal de cualquiera de las formas de realización 13–20, en donde una célula vegetal es de la familia *Rosaceae*.
22. La célula vegetal de cualquiera de las formas de realización 13–21, en donde una célula vegetal es de una planta seleccionada del grupo que consiste en: frambuesa, frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, manzana, tomate, uva, durazno, ciruela, damasco, pera, membrillo, níspero y almendra.
23. La célula vegetal de cualquiera de las formas de realización 13–22, en donde una célula vegetal es de *Rubus occidentalis*.

24. Un método para la producción de una planta editada con bases que expresa un factor de transcripción caja MADS clado AGAMOUS alterado, que comprende: (a) introducir en la planta un sistema de edición de bases capaz de modificar un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS, e introducir una sustitución de nucleótido único estable y dirigida en dicho gen, donde el sistema de edición de bases comprende: (i) una proteína de fusión de edición de bases, que comprende: una proteína efectora asociada a CRISPR y una citidina desaminasa o adenosina desaminasa; y (ii) un ARN guía (ARNg) capaz de dirigir la proteína de fusión a un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS; y (b) expresar el sistema de edición de bases en la planta, de modo que el ARNg se una al gen objetivo endógeno y la proteína de fusión de edición de bases contacte el gen objetivo endógeno y desamine una base de nucleótido para producir una sustitución de nucleótido único estable y dirigida en el gen objetivo endógeno.
25. El método de la forma de realización 24, en donde el sistema de edición de bases introduce una sustitución de un solo nucleótido estable y dirigida C>T o G>A o A>G o T>C en el gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS.
26. El método de cualquiera de las formas de realización 24–25, en donde el gen objetivo endógeno codifica un factor de transcripción de caja MADS clado AGAMOUS involucrado en la formación del fruto.
27. El método de cualquiera de las formas de realización 24–26, en donde el gen objetivo endógeno codifica un factor de transcripción de caja MADS clado AGAMOUS involucrado en la formación del endocarpio.
28. El método de cualquiera de las formas de realización 24–27, en donde el factor de transcripción de caja MADS clado AGAMOUS se selecciona del grupo que consiste en: AG, SHP1, SHP2, y STK.
29. El método de cualquiera de las formas de realización 24–28, en donde el factor de transcripción de caja MADS clado AGAMOUS es STK.
30. El método de cualquiera de las formas de realización 24–29, en donde el ARNg comprende una secuencia guía que se une a un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS que comparte por lo menos 90% de identidad de secuencia con **SEQ ID NO: 13**.

31. El método de cualquiera de las formas de realización 24–30, en donde el sistema de edición de bases se adapta para crear una sustitución de un solo nucleótido estable y dirigida C>T o G>A o A>G o T>C en la región del gen objetivo endógeno que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS.
32. El método de cualquiera de las formas de realización 24–31, en donde el sistema de edición de bases se adapta para crear una sustitución de un solo nucleótido estable y dirigida C>T o G>A o A>G o T>C en la región del gen objetivo endógeno que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS, en donde con la expresión del gen, el factor de transcripción de caja MADS clado AGAMOUS resultante tiene alterada expresión y/o función.
33. El método de cualquiera de las formas de realización 24–32, en donde la planta es una *Rosaceae*.
34. El método de cualquiera de las formas de realización 24–33, en donde la planta se selecciona del grupo que consiste en: frambuesa, frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, manzana, tomate, uva, durazno, ciruela, damasco, pera, membrillo, níspero y almendra.
35. El método de cualquiera de las formas de realización 24–34, en donde la planta es *Rubus occidentalis*.
36. Un método para producir un alelo negativo dominante en un factor de transcripción de caja MADS clado AGAMOUS para crear un fenotipo de desarrollo de fruto alterado en una planta, que comprende: introducir una sustitución de nucleótido único estable y dirigida en un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS, en donde la planta tiene un fenotipo de desarrollo de fruto alterado cuando se compara con una planta de tipo salvaje sin dicha sustitución de un solo nucleótido.
37. El método de la forma de realización 36, en donde la sustitución de un solo nucleótido estable y dirigida es una sustitución de un solo nucleótido C>T o G>A o A>G o T>C en el gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS.
38. El método de cualquiera de las formas de realización 36–37, en donde el gen objetivo endógeno codifica un factor de transcripción de caja MADS clado AGAMOUS involucrado en la formación del fruto.

39. El método de cualquiera de las formas de realización 36–38, en donde el gen objetivo endógeno codifica un factor de transcripción de caja MADS clado AGAMOUS involucrado en la formación del endocarpio.
40. El método de cualquiera de las formas de realización 36–39, en donde el factor de transcripción de caja MADS clado AGAMOUS se selecciona del grupo que consiste en: AG, SHP1, SHP2, y STK.
41. El método de cualquiera de las formas de realización 36–40, en donde el factor de transcripción de caja MADS clado AGAMOUS es STK.
42. El método de cualquiera de las formas de realización 36–41, en donde la sustitución de un solo nucleótido estable y dirigida se introduce en la región del gen objetivo endógeno que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS.
43. El método de cualquiera de las formas de realización 36–42, en donde la sustitución de un solo nucleótido estable y dirigida se introduce en la región del gen objetivo endógeno que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS, en donde, con la expresión del gen, el factor de transcripción de caja MADS clado AGAMOUS resultante tiene alterada expresión y/o función.
44. El método de cualquiera de las formas de realización 36–43, en donde la planta es una *Rosaceae*.
45. El método de cualquiera de las formas de realización 36–44, en donde la planta se selecciona del grupo que consiste en: frambuesa, frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, manzana, tomate, uva, durazno, ciruela, damasco, pera, membrillo, níspero y almendra.
46. El método de cualquiera de las formas de realización 36–45, en donde la planta es *Rubus occidentalis*.
47. El método de cualquiera de las formas de realización 36–46, en donde un fenotipo de desarrollo de fruto alterado es un fruto sin semillas y/o reducida simiente.
48. El método de cualquiera de las formas de realización 36–47, en donde un fenotipo de desarrollo de fruto alterado es por lo menos uno de: reducido tamaño de semilla, reducido contenido de lignina en semilla, reducida formación de endocarpio de semilla o superficie de semilla más lisa.

49. Una planta editada genéticamente, que comprende: una o más modificaciones de base de un solo nucleótido en un gen de factor de transcripción de caja MADS clado AGAMOUS, en donde la planta editada genéticamente tiene un fenotipo de desarrollo de fruto alterado cuando se compara con una planta de tipo salvaje sin dichas modificaciones de base de un solo nucleótido.
50. La planta editada genéticamente de la forma de realización 49, en donde las una o más modificaciones de base de un solo nucleótido comprenden una sustitución C>T o G>A o A>G o T>C en el gen del factor de transcripción de caja MADS clado AGAMOUS.
51. La planta editada genéticamente de cualquiera de las formas de realización 49–50, en donde el gen de factor de transcripción de caja MADS clado AGAMOUS está involucrado en la formación del fruto.
52. La planta editada genéticamente de cualquiera de las formas de realización 49–51, en donde el gen de factor de transcripción de caja MADS clado AGAMOUS está involucrado en la formación del endocarpio.
53. La planta editada genéticamente de cualquiera de las formas de realización 49–52, en donde el factor de transcripción de caja MADS clado AGAMOUS se selecciona del grupo que consiste en: AG, SHP1, SHP2, y STK.
54. La planta editada genéticamente de cualquiera de las formas de realización 49–53, en donde el factor de transcripción de caja MADS clado AGAMOUS es STK.
55. La planta editada genéticamente de cualquiera de las formas de realización 49–54, en donde las una o más modificaciones de base de un solo nucleótido están en la región del gen que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS.
56. La planta editada genéticamente de cualquiera de las formas de realización 49–55, en donde las una o más modificaciones de base de un solo nucleótido están en la región del gen que codifica para la región C-terminal del factor de transcripción de caja MADS clado AGAMOUS, en donde, con la expresión del gen, el factor de transcripción de caja MADS clado AGAMOUS resultante tiene alterada expresión y/o función.
57. La planta editada genéticamente de cualquiera de las formas de realización 49–56, en donde la planta es una *Rosaceae*.

58. La planta editada genéticamente de cualquiera de las formas de realización 49–57, en donde la planta se selecciona del grupo que consiste en: frambuesa, frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, manzana, tomate, uva, durazno, ciruela, damasco, pera, membrillo, níspero y almendra.
59. La planta editada genéticamente de cualquiera de las formas de realización 49–58, en donde la planta es *Rubus occidentalis*.
60. La planta editada genéticamente de cualquiera de las formas de realización 49–59, en donde un fenotipo de desarrollo de fruto alterado es un fruto sin semillas y/o de reducida simiente.
61. La planta editada genéticamente de cualquiera de las formas de realización 49–60, en donde un fenotipo de desarrollo de fruto alterado es por lo menos uno de: reducido tamaño de semilla, reducido contenido de lignina en semilla, reducida formación de endocarpio de semilla o superficie de semilla más lisa.
62. Una planta con un proteoma modificado de factor de transcripción de caja MADS clado AGAMOUS, que comprende: una sustitución de aminoácido introducida en una secuencia de aminoácidos de factor de transcripción de caja MADS clado AGAMOUS, en donde la planta tiene un fenotipo de desarrollo de fruto alterado cuando se compara con una planta de tipo salvaje sin dicha introducción de aminoácido introducida.
63. La planta de la forma de realización 62, en donde la sustitución de aminoácido introducida es en la región de secuencia de aminoácidos N-terminal, de caja MADS, caja K, y/o C-terminal de dicho factor de transcripción de caja MADS clado AGAMOUS.
64. La planta de cualquiera de las formas de realización 62–63, en donde la sustitución de aminoácido introducida causa la alterada expresión y/o función del factor de transcripción de caja MADS clado AGAMOUS.
65. La planta de cualquiera de las formas de realización 62–64, en donde el factor de transcripción de caja MADS clado AGAMOUS está involucrado en la formación del fruto.
66. La planta de cualquiera de las formas de realización 62–65, en donde el factor de transcripción de caja MADS clado AGAMOUS está involucrado en la formación del endocarpio.
67. La planta de cualquiera de las formas de realización 62–66, en donde el factor de transcripción de caja MADS clado AGAMOUS se selecciona del grupo que consiste en: AG, SHP1, SHP2, y STK.

68. La planta de cualquiera de las formas de realización 62–67, en donde el factor de transcripción de caja MADS clado AGAMOUS es STK.
69. La planta de cualquiera de las formas de realización 62–68, en donde la planta es una *Rosaceae*.
70. La planta de cualquiera de las formas de realización 62–69, en donde la planta se selecciona del grupo que consiste en: frambuesa, frambuesa negra, mora, cereza, durazno, palta, frutilla, frutilla silvestre, manzana, tomate, uva, durazno, ciruela, damasco, pera, membrillo, níspero y almendra.
71. La planta de cualquiera de las formas de realización 62–70, en donde la planta es *Rubus occidentalis*.
72. La planta de cualquiera de las formas de realización 62–71, en donde un fenotipo de desarrollo de fruto alterado es un fruto sin semillas y/o de reducida simiente.
73. La planta de cualquiera de las formas de realización 62–72, en donde un fenotipo de desarrollo de fruto alterado es por lo menos uno de: reducido tamaño de semilla, reducido contenido de lignina en semilla, reducida formación de endocarpio de semilla o superficie de semilla más lisa.
74. Una planta *Rubus occidentalis* con un fenotipo de desarrollo de fruto alterado, que comprende: una o más sustituciones de un solo nucleótido dirigidas en un gen que codifica un factor de transcripción de caja MADS clado AGAMOUS, en donde la planta *Rubus occidentalis* tiene un fenotipo de desarrollo de fruto alterado en comparación con una planta de tipo salvaje *Rubus occidentalis* sin dicha sustitución de un solo nucleótido.
75. Un método para generar una planta editada con bases con un fenotipo de desarrollo de fruto alterado, que comprende: (a) proporcionar una célula o un tejido de la planta para la transformación; (b) transformar dicha célula o tejido con un sistema de edición de bases capaz de modificar un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS e introducir una sustitución de nucleótido único estable y dirigida en dicho gen; (c) expresar el sistema de edición de bases en la célula o el tejido transformado, en donde el sistema de edición de bases desamina una base de nucleótido para obtener una sustitución de nucleótido único estable y dirigida en el gen objetivo endógeno; (d) regenerar plantas a partir de dicha célula o tejido transformado; y (e) identificar una planta que tiene un fenotipo de desarrollo de fruto alterado.

76. Un método de fitomejoramiento de una planta editada con bases para producir una planta libre de transgenes, que comprende: (a) hacer un cruce entre una primera planta editada con bases producida a partir de la forma de realización 75 con una segunda planta, para introducir la sustitución de un solo nucleótido en la segunda planta; y (b) seleccionar una planta de progenie que tiene la sustitución de un solo nucleótido, pero que está libre de transgenes.
77. Una célula vegetal, que comprende: un sistema de edición génica capaz de modificar un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS e introducir una mutación en una región dirigida de dicho gen por medio de una rotura de ADN bicatenaria mediada por nucleasas.
78. La célula vegetal de la forma de realización 77, en donde la rotura de ADN bicatenaria mediada por nucleasas es reparada por unión de extremo no homólogo (NHEJ) o reparación dirigida por homología (HDR).
79. La célula vegetal de cualquiera de las formas de realización 77–78, en donde la nucleasa capaz de introducir la mutación por dicha rotura de ADN bicatenaria se selecciona del grupo que consiste en: una TALEN, una MegaTAL, una endonucleasa de orientación (homing), una ZFN, una endonucleasa FokI y una endonucleasa CRISPR/Cas.
80. La célula vegetal de cualquiera de las formas de realización 77–79, en donde el sistema de edición génica se introduce en la célula vegetal para la expresión y función de una molécula de edición génica.
81. La célula vegetal de cualquiera de las formas de realización 77–80, en donde la célula vegetal comprende una delección o inserción de uno o más nucleótidos en el gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS; en donde la delección o inserción se introduce por medio de la molécula de edición génica; y en donde el sistema de edición génica no está presente en la célula vegetal.
82. Un método para producir una planta que expresa un factor de transcripción de caja MADS clado AGAMOUS alterado, que comprende: (a) introducir en la planta un sistema de edición génica capaz de modificar un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS e introducir una mutación en una región dirigida de dicho gen por medio de una rotura de ADN bicatenaria mediada por nucleasas; y (b) expresar el sistema de edición génica en la planta, de modo que la molécula de edición génica expresada se una al

gen objetivo endógeno e introduzca una delección o inserción de uno o más nucleótidos en el gen objetivo endógeno.

83. El método de la forma de realización 82, en donde la rotura de ADN bicatenaria mediada por nucleasas es reparada por unión de extremo no homólogo (NHEJ) o reparación dirigida por homología (HDR).
84. El método de cualquiera de las formas de realización 82–83, en donde la nucleasa capaz de introducir la mutación por medio de dicha rotura de ADN bicatenaria se selecciona del grupo que consiste en una TALEN, una MegaTAL, una endonucleasa de orientación (homing), una ZFN, una endonucleasa FokI y una endonucleasa CRISPR/Cas.
85. El método de cualquiera de las formas de realización 82–84, en donde el sistema de edición génica se introduce en la célula vegetal para la expresión y función de una molécula de edición génica.
86. El método de cualquiera de las formas de realización 82–85, en donde la célula vegetal comprende una delección o inserción de uno o más nucleótidos en el gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS; en donde la delección o inserción se introduce por medio de la molécula de edición génica; y en donde el sistema de edición génica no está presente en la célula vegetal.
87. Un método para producir un alelo negativo dominante en un factor de transcripción de caja MADS clado AGAMOUS para crear un fenotipo de desarrollo de fruto alterado en una planta, que comprende: introducir una delección o inserción de uno o más nucleótidos en un gen objetivo endógeno que codifica un factor de transcripción de caja MADS clado AGAMOUS, en donde la planta tiene un fenotipo de desarrollo de fruto alterado cuando se compara con una planta de tipo salvaje sin dicha delección o inserción.

REFERENCIAS.

Patente de los Estados Unidos Nro. 6.603.061.

Patente de los Estados Unidos Nro. 7.868.149.

Publicación de Solicitud de Patente de los Estados Unidos 2009/0100536 A1.

Belhaj, K., Chaparro–Garcia, A., Kamoun, S., Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods*, 9(1):39.

Blast reference: <https://blast.ncbi.nlm.nih.gov/>

Brooks, C., Nekrasov, V., Lippman, Z. B., Van Eck, J. (2014). Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR–Associated9 System. *Plant Physiology*, 166(3), 1292–1297.

Caliando, B. J., Voigt, C. A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nature Communications*, 6:6989:1–10.

Carolina Royo, Rafael Torres–Pérez, Nuria Mauri, Nieves Diestro, José Antonio Cabezas, Cécile Marchal, Thierry Lacombe, Javier Ibáñez, Manuel Tornel, Juan Carreño, José Miguel Martínez–Zapater, Pablo Carbonell–Bejerano (2018). The Major Origin of Seedless Grapes Is Associated with a Missense Mutation in the MADS–Box Gene VviAGL11. *Plant Physiology*, 177(3), 1234–1253.

Espley, R. V., Brendolise, C., Chagne, D., Kutty–Amm, S., Green, S., Volz, R., ... Allan, A. C. (2009). Multiple Repeats of a Promoter Segment Causes Transcription Factor Autoregulation in Red Apples. *THE PLANT CELL ONLINE*, 21(1), 168–183. doi:10.1105/tpc.108.059329.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.–L., Wei, P., Cao F., Zhu S., Zhang F., Mao Y., Zhu, J.–K. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, 23(10):1229–1232.

Hellens, R., Allan, A., Friel, E., Bolitho, K., Grafton, K., Templeton, M., Laing, W. (2005). *Plant Methods*, 1(1), 13. doi:10.1186/1746–4811–1–13.

Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., and L.S. Vinh (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, 35:518–522.

Honma, T., & Goto, K. (2001). Complexes of MADS–box proteins are sufficient to convert leaves into floral organs. *Nature*, 409(6819), 525–529.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F, von Haeseler A., and L.S. Jermiin (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods*, 14:587589.

Li H. Constructing the TreeFam database. PhD thesis, Chinese Academy of Sciences Beijing. 2006.

Morrell, P. L., Buckler, E. S., Ross–Ibarra, J. (2011). Crop genomics: advances and applications. *Nature Reviews Genetics*, 13(2):85–96.

- Nguyen L.-T., Schmidt H.A., von Haeseler A., and B.Q. Minh (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies.. *Mol. Biol. Evol.*, 32:268–274.
- Notredame, Higgins, Heringa. 2000. T-Coffee: A novel method for multiple sequence alignments. *JMB*,302(205–217) DOI: 10.1006/jmbi.2000.4042.
- Rice,P. Longden,I. and Bleasby,A. EMBOSS: The European Molecular Biology Open Software Suite. (2000). *Trends in Genetics* 16, (6) pp276—277.
- Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Liu J., Xi J. J., Qiu J. L., Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR–Cas system. *Nature Biotechnology*, 31(8):686–688.
- Xie, K., Yang, Y. (2013). RNA–Guided Genome Editing in Plants Using a CRISPR–Cas System. *Molecular Plant*, 6(6):1975–1983.
- Xu, R., Li, H., Qin, R., Wang, L., Li, L., Wei, P., Yang, J. (2014). Gene targeting using the *Agrobacterium tumefaciens*–mediated CRISPR–Cas system in rice. *Rice*, 7(1):5.
- Zhou, X., Jacobs, T. B., Xue, L.–J., Harding, S. A., Tsai, C.–J. (2015). Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4coumarate:CoA ligase specificity and redundancy. *New Phytologist*, 208(2):298–301.