

EFFECTORES TIPO ACTIVADOR DE TRANSCRIPCIÓN FUSIONADOS A INTEÍNAS

INCORPORACIÓN POR REFERENCIA DE MATERIAL PRESENTADO ELECTRÓNICAMENTE

[0001] Se incorpora por referencia en su totalidad un listado de secuencias de nucleótidos/aminoácidos legible por computadora, un archivo de texto ASCII de 115 kb de tamaño, presentado simultáneamente con el presente, e identificado como sigue: "C1633112111_SequenceListing" y creado el 7 de julio de 2022.

ANTECEDENTES

[0002] Las tecnologías de edición del genoma mediante nucleasas de corte raro, tales como las nucleasas efectoras tipo activador de transcripción (TALEN), nucleasas de dedos de zinc (ZFN), repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) y los sistemas relacionados con la proteína 9 asociada a CRISPR (Cas9) o Cpf1, han acelerado la investigación en biología básica, biotecnología, mejoramiento genético y terapia génica. Las nucleasas de corte raro se pueden utilizar para generar supresiones, inserciones e iniciar la recombinación homóloga. Las TALEN incluyen una nucleasa de corte raro y un TALE que se pueden dirigir a una secuencia específica y generar una ruptura precisa en el ácido desoxirribonucleico (ADN). Un diseño de TALEN incluye una mitad izquierda de las TALEN que incluye una TALE izquierda fusionada a la nucleasa de corte raro que reconoce un primer sitio de unión seguido de una región espaciadora y una mitad derecha de las TALEN que incluye una TALE derecha fusionada a la nucleasa de corte raro que reconoce un segundo sitio de unión.

SÍNTESIS

[0003] La presente descripción está dirigida a superar los retos y necesidades mencionados anteriormente relacionados con los TALE.

[0004] Varios aspectos se refieren a una pluralidad de secuencias de nucleótidos, que comprenden una primera secuencia de nucleótidos que codifica una primera inteína

fusionada a al menos una porción de un primer efector tipo activador de transcripción (TALE), una segunda secuencia de nucleótidos que codifica la primera inteína fusionada a al menos una porción de un segundo TALE, y una tercera secuencia de nucleótidos que codifica una segunda inteína fusionada a al menos una porción de una nucleasa de corte raro.

[0005] En algunos aspectos, la primera secuencia de nucleótidos codifica la primera inteína fusionada al primer TALE y la segunda secuencia de nucleótidos codifica la primera inteína fusionada al segundo TALE. En algunos aspectos adicionales, la tercera secuencia de nucleótidos codifica la segunda inteína fusionada a la nucleasa de corte raro.

[0006] En algunos aspectos, la primera secuencia de nucleótidos codifica la primera inteína fusionada a una primera porción de la nucleasa de corte raro y el primer TALE, y la segunda secuencia de nucleótidos codifica la primera inteína fusionada a la primera porción de la nucleasa de corte raro y el segundo TALE. En algunos aspectos adicionales, la tercera secuencia de nucleótidos codifica la segunda inteína fusionada a una segunda porción de la nucleasa de corte raro, en la que la primera porción y la segunda porción de la nucleasa de corte raro forman la nucleasa de corte raro.

[0007] En algunos aspectos, la pluralidad de secuencias de nucleótidos, cada una incluye un vector separado y/o están en un constructo de expresión único.

[0008] En algunos aspectos, la primera inteína y la segunda inteína están configuradas para auto-empalmar cuando están en contacto y, en respuesta, para formar una primera mitad de las nucleasas efectoras tipo activador de transcripción (TALEN) que incluye el primer TALE unido a la nucleasa de corte raro. En algunos aspectos adicionales, la primera inteína y la segunda inteína están configuradas para auto-empalmar cuando están en contacto y, en respuesta, para formar una segunda mitad de las TALENmitad de las TALEN que incluye el segundo TALE unido a la nucleasa de corte raro.

[0009] En algunos aspectos, la primera inteína y la segunda inteína están configuradas para auto-empalmar cuando están en contacto y formar una proteína empalmada que incluye la primera inteína unida a la segunda inteína.

[0010] En algunos aspectos, cada una de la pluralidad de secuencias de nucleótidos también codifica un promotor y un terminador.

[0011] Algunos aspectos se refieren a un procedimiento que comprende poner en contacto una célula con: una primera secuencia de nucleótidos que codifica una primera

inteína fusionada a al menos una porción de un primer efector tipo activador de transcripción (TALE), una segunda secuencia de nucleótidos que codifica la primera inteína fusionada a al menos una porción de un segundo TALE, y una tercera secuencia de nucleótidos que codifica una segunda inteína fusionada a al menos una porción de una nucleasa de corte raro. El procedimiento también comprende, en respuesta al contacto con la célula, empalmar el primer TALE, el segundo TALE, y la nucleasa de corte raro mediante la primera inteína y la segunda inteína para formar: una primera mitad de las nucleasas efectoras tipo activador de transcripción (TALEN) que incluye el primer TALE y la nucleasa de corte raro, y una segunda mitad de las TALEN que incluye el segundo TALE y la nucleasa de corte raro.

[0012] En algunos aspectos, el procedimiento además incluye traducir la primera secuencia de nucleótidos, la segunda secuencia de nucleótidos, y la tercera secuencia de nucleótidos para formar la primera inteína fusionada al primer TALE, la primera inteína fusionada a el segundo TALE, y la segunda inteína fusionada a la nucleasa de corte raro.

[0013] En algunos aspectos, el procedimiento además incluye transformar la célula usando la primera mitad de las TALEN y la segunda mitad de las TALEN.

[0014] En algunos aspectos, la primera secuencia de nucleótidos codifica una inteína N-terminal fusionada a un C-terminal del primer TALE, la segunda secuencia de nucleótidos codifica la inteína N-terminal fusionada a un C-terminal del segundo TALE, y la tercera secuencia de nucleótidos codifica una inteína C-terminal fusionada a un N-terminal de la nucleasa de corte raro o la inteína C-terminal fusionada entre las porciones de la nucleasa de corte raro.

[0015] En algunos aspectos, la primera secuencia de nucleótidos codifica una inteína C-terminal fusionada a un N-terminal del primer TALE, la segunda secuencia de nucleótidos codifica la inteína C-terminal fusionada a un N-terminal del segundo TALE, y la tercera secuencia de nucleótidos codifica una inteína N-terminal fusionada a un C-terminal de la nucleasa de corte raro o la inteína C-terminal fusionada entre las porciones de la nucleasa de corte raro.

[0016] En algunos aspectos, el empalme incluye la unión de la primera inteína a la segunda inteína para formar: un primer intermediario que incluye la primera inteína unida a la segunda inteína, en el que la primera inteína está fusionada al primer TALE y la segunda inteína está fusionada a la nucleasa de corte raro, y un segundo intermediario que incluye la primera inteína unida a la segunda inteína, en el que la primera inteína está

fusionada al segundo TALE y la segunda inteína está fusionada a la nucleasa de corte raro.

[0017] En algunos aspectos, el empalme incluye: unir la primera inteína a la segunda inteína, cortar sitios de empalme asociados con la primera inteína y la segunda inteína, y unir el primer TALE a la nucleasa de corte raro y unir el segundo TALE a la nucleasa de corte raro para formar la primera mitad de las TALEN y la segunda mitad de las TALEN.

[0018] En algunos aspectos, los sitios de empalme están entre la primera inteína y el primer TALE, la primera inteína y el segundo TALE, y la segunda inteína y la nucleasa de corte raro o porciones de estas.

[0019] Algunos aspectos se refieren a un constructo de expresión, que comprende: una primera secuencia de nucleótidos que codifica una primera inteína fusionada a al menos un primer efector tipo activador de transcripción (TALE), una segunda secuencia de nucleótidos que codifica la primera inteína fusionada a al menos un segundo TALE, y una tercera secuencia de nucleótidos que codifica una segunda inteína fusionada a al menos una porción de una nucleasa de corte raro.

[0020] En algunos aspectos, la primera secuencia de nucleótidos codifica una inteína N-terminal fusionada a un C-terminal del primer TALE, la segunda secuencia de nucleótidos codifica la inteína N-terminal fusionada a un C-terminal del segundo TALE, y la tercera secuencia de nucleótidos codifica una inteína C-terminal fusionada a un N-terminal de la nucleasa de corte raro o fusionada entre las porciones de la nucleasa de corte raro.

[0021] En algunos aspectos, la primera secuencia de nucleótidos codifica una inteína C-terminal fusionada a un N-terminal del primer TALE, la segunda secuencia de nucleótidos codifica la inteína C-terminal fusionada a un N-terminal del segundo TALE, y la tercera secuencia de nucleótidos codifica una inteína N-terminal fusionada a un C-terminal de la nucleasa de corte raro o fusionada entre las porciones de la nucleasa de corte raro.

[0022] En algunos aspectos, en respuesta a la traducción de la primera secuencia de nucleótidos, la segunda secuencia de nucleótidos, y la tercera secuencia de nucleótidos mediante una célula, la primera inteína y segunda inteína están configuradas para unirse entre sí y autoempalmarse para formar: una primera mitad de las nucleasas efectoras tipo activador de transcripción (TALEN) que incluye el primer TALE unido a la nucleasa de corte raro, una segunda mitad de las TALEN que incluye el segundo TALE unido a la

nucleasa de corte raro, y a proteína empalmada que incluye la primera inteína unida a la segunda inteína.

[0023] En algunos aspectos, la primera secuencia de nucleótidos, la segunda secuencia de nucleótidos, y la tercera secuencia de nucleótidos incluyen vectores separados o están en un constructo de expresión único.

[0024] En algunos aspectos, el primer TALE que incluye una primera pluralidad de secuencias de repetición de TALE que, en combinación, se unen a una primera secuencia de nucleótidos en una secuencia de ADN diana, y el segundo TALE que incluye a segunda pluralidad de secuencias de repetición de TALE que, en combinación, se unen a una segunda secuencia de nucleótidos en la secuencia de ADN diana.

[0025] En algunos aspectos, cada una de la primera secuencia de nucleótidos, la segunda secuencia de nucleótidos, y la tercera secuencia de nucleótidos además codifican un promotor y un terminador.

[0026] Varios aspectos se refieren a una planta, parte de planta o células de plantas transformadas por una pluralidad de secuencias de nucleótidos, la pluralidad de secuencias de nucleótidos, que comprende: una primera secuencia de nucleótidos que codifica una primera inteína fusionada a al menos una porción de un primer efector tipo activador de transcripción (TALE), una segunda secuencia de nucleótidos que codifica la primera inteína fusionada a al menos una porción de un segundo TALE, y una tercera secuencia de nucleótidos que codifica una segunda inteína fusionada a al menos una porción de una nucleasa de corte raro. Y, donde la planta, parte de planta o células de plantas transformadas expresan la primera inteína fusionada al primer TALE, la primera inteína fusionada al segundo TALE, y la segunda inteína fusionada a la nucleasa de corte raro.

[0027] En algunos aspectos, la primera inteína y segunda inteína expresadas, de la primera inteína fusionada al primer TALE, la primera inteína fusionada al segundo TALE, y la segunda inteína fusionada a la nucleasa de corte raro, se empalman para formar: una primera mitad de las nucleasas efectoras tipo activador de transcripción (TALEN) que incluye el primer TALE unido a la nucleasa de corte raro, una segunda mitad de las TALEN que incluye el segundo TALE unido a la nucleasa de corte raro, y a proteína empalmada que incluye la primera inteína unida a la segunda inteína.

[0028] En algunos aspectos, la planta, parte de planta o células de plantas transformadas exhiben: una primera mitad de las nucleasas efectoras tipo activador de transcripción

(TALEN) que incluye el primer TALE unido a la nucleasa de corte raro, una segunda mitad de las TALEN que incluye el segundo TALE unido a la nucleasa de corte raro, y una proteína empalmada que incluye la primera inteína unida a la segunda inteína.

[0029] Varias realizaciones están dirigidas a la célula y/u organismo huésped transformado por los procedimientos, vectores, constructos de expresión, secuencias de nucleótidos y/o sistemas descritos en la presente.

[0030] Varias realizaciones están dirigidas a un procedimiento de formación de cualquiera de las secuencias de nucleótidos y/o sistemas reivindicados en la presente.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0031] Varios ejemplos de realizaciones pueden ser comprendidos más completamente en consideración de la siguiente descripción detallada en conexión con los dibujos adjuntos, en los cuales:

[0032] Las FIG. 1A-1D son diagramas que ilustran ejemplos de secuencias de nucleótidos que codifican una primera inteína con un TALE y codifican una segunda inteína con una nucleasa de corte raro, conforme a la presente descripción.

[0033] La FIG. 2 es un diagrama de flujo que ilustra un ejemplo de procedimiento para generar TALENs usando ejemplos de secuencias de nucleótidos, conforme a la presente descripción.

[0034] Las FIG. 3A-3B son diagramas que ilustran ejemplos de constructos de expresión, conforme a la presente descripción.

[0035] La FIG. 4 es un diagrama de flujo que ilustra un ejemplo de empalme mediante inteínas para formar una TALEN, conforme a la presente descripción.

[0036] La FIG. 5 ilustra imágenes que muestran la genotipificación resultante de editar una célula usando secuencias de nucleótidos, consistente con la presente descripción.

[0037] Las FIG. 6A-6C ilustran imágenes completas de las regiones genómicas de un control negativo YFP, un control positivo TALEN, e inteínas, conforme a la presente descripción.

[0038] La FIG. 7 ilustra ejemplos de resultados de transformación de explantes de soja utilizando los diferentes vectores plasmídicos, conforme a la presente descripción.

[0039] Las FIGs. 8A-8I ilustran resultados de transformación de explantes de cáñamo usando los diferentes vectores plasmídicos, conforme a la presente descripción.

DESCRIPCIÓN DETALLADA

[0040] Los aspectos de la presente descripción se refieren a una variedad de procedimientos, secuencias de nucleótidos, sistemas, constructos de expresión y células huésped y/u organismos transformados usando las secuencias de nucleótidos. Aunque la presente invención no se limita necesariamente a tales aplicaciones, se pueden apreciar diversos aspectos de la invención a través de una discusión de diversas realizaciones utilizando este contexto.

[0041] En consecuencia, en la siguiente descripción se exponen diversos detalles específicos para describir realizaciones concretas presentadas en la presente. Sin embargo, debería ser evidente para un experto en la materia que uno o más ejemplos y/o variaciones de estas realizaciones se pueden practicar sin todos los detalles específicos que se dan a continuación. En otros casos, características bien conocidas no se han descrito en detalle para no oscurecer la descripción de las realizaciones expuestas en la presente. Para facilitar la ilustración, los mismos números de referencia se pueden utilizar en diferentes diagramas para referirse a los mismos elementos o instancias adicionales del mismo elemento.

[0042] Las nucleasas de corte raro, tales como las endonucleasas de corte raro, se pueden utilizar para dirigirse a genes específicos o a múltiples genes diana. Los ejemplos no limitantes incluyen TALEN, endonucleasas homing manipuladas genéticamente, ZFNs, meganucleasas y CRISPR. Las nucleasas de corte raro pueden ser proteínas naturales o manipuladas genéticamente con actividad endonucleasa dirigida a una secuencia de nucleótidos con un sitio de reconocimiento, a veces denominado secuencia diana, de 12-40 pares de bases de longitud o más. Típicamente, las nucleasas de corte raro causan la escisión dentro del sitio de reconocimiento. En algunos casos, la nucleasa de corte raro es una proteína de fusión que contiene un dominio de unión fusionado a un dominio nucleasa con actividad de escisión. El dominio de unión se puede configurar para unirse a una secuencia diana. El dominio de la endonucleasa de corte raro se puede configurar para inducir una mutación en un locus genómico diana asociado a la localización diana.

Las TALEN y los ZFN son ejemplos de proteínas de fusión de los dominios de unión con el dominio de nucleasa, tal como una endonucleasa de FokI, a veces denominada Fok1. Las TALEN se pueden manipular genéticamente para que sean específicas de una secuencia diana mediante la fusión de los dominios de unión con el dominio nucleasa, lo que produce nucleasas quiméricas dirigidas a secuencias de ADN específicas y

seleccionadas, y que producen el corte del ADN en la secuencia diana o cerca de ella. Estos cortes de ADN (roturas de doble cadena) pueden inducir mutaciones, tal como la inactivación o alteración de la función génica con precisión y eficacia.

[0043] Las realizaciones de acuerdo con la presente descripción se refieren al uso de inteínas en secuencias de nucleótidos que contienen TALEs y nucleasas. En algunas realizaciones, las medias TALEs y nucleasas están separadas entre sí, y las inteínas pueden empalmar y ligar las respectivas medias de TALEs a la nucleasa después de la traducción. Mediante el uso de inteínas complementarias para las secuencias de nucleótidos que contienen las medias TALEN y las nucleasas, los vectores suministrados y utilizados para transformar las células de un organismo pueden tener un tamaño de plásmido menor, a veces denominado "tamaño de carga", en comparación con las secuencias de nucleótidos que contienen las medias TALEN completas. La disminución del tamaño del plásmido puede aumentar la frecuencia de expresión de las secuencias de nucleótidos. Además, el uso de inteínas en secuencias de nucleótidos que contienen TALENs y nucleasas puede añadir flexibilidad a la administración del material de edición genética a la célula.

[0044] En algunas realizaciones, las diferentes secuencias de nucleótidos se pueden administrar en un constructo de expresión común, tal como un plásmido. Aunque se administren juntas, cada secuencia de nucleótidos codifica un complejo molecular diferente que se transcribirá y traducirá por separado de otros complejos moleculares. Un complejo molecular es un compuesto o complejo de compuestos transcritos y traducidos juntos para formar una proteína o proteína de fusión. Por ejemplo, la primera secuencia de nucleótidos codifica un complejo molecular del primer TALE fusionado a una primera inteína que se traducen y transcriben juntas para formar una proteína de fusión que incluye el primer TALE fusionado y la primera inteína. Los complejos moleculares resultantes tienen un tamaño de carga menor en comparación con la traducción y transcripción de una nucleasa fusionada al primer TALE. Una proteína de fusión, tal como se utiliza en la presente, incluye y/o se refiere a una proteína o complejo proteico que incluye al menos dos dominios codificados por genes separados que se unen o fijan juntos de manera que los genes se transcriben y/o traducen juntos como una unidad única, lo que produce un polipéptido único.

[0045] En cualquiera de las realizaciones descritas anteriormente, los complejos moleculares se transcriben y traducen, y luego se empalman para formar TALENs para

transformar las células. La disminución del tamaño de la carga de los complejos moleculares que se administran dentro de una célula puede aumentar la frecuencia de expresión de las formas transcritas de los complejos moleculares. Además, el uso de inteínas en los complejos moleculares que contienen TALEs y nucleasas puede añadir flexibilidad a la administración del material de edición génica a las células. Por ejemplo, la secuencia de nucleótidos que codifica la nucleasa puede seguir siendo la misma, y solo se revisan las porciones variables de los TALE para diferentes dianas (por ejemplo, en la primera y segunda secuencias de nucleótidos).

[0046] Volviendo ahora a las figuras, las FIG. 1A-1D son diagramas que ilustran ejemplos de secuencias de nucleótidos que codifican una primera inteína con un TALE y codifican una segunda inteína con una nucleasa de corte raro, coherente con la presente descripción.

[0047] Como se muestra en la FIG. 1A, la pluralidad de secuencias de nucleótidos 100 codifica la primera inteína 104 fusionada a al menos una porción del primer TALE 102 y codifica la segunda inteína 106 fusionada a al menos una porción de la nucleasa de corte raro 108. Como también se ilustra en la FIG. 1B, el primer TALE 102 puede incluir un N-terminal de TALE 103, una primera pluralidad de secuencias de repetición de TALE 105 (ilustrado como un dominio de unión), y un C-terminal TALE 107. La primera pluralidad de secuencias de repetición de TALE 105 puede, en combinación, unirse a una primera secuencia de nucleótidos en una secuencia de ADN diana, y se puede denominar como dominio de unión. Las secuencias de repetición de TALE pueden ser una matriz de 33 a 35 repeticiones de aminoácidos de longitud que difieren en dos posiciones (posiciones 12 y 13), a veces denominadas "diresiduos variables repetidos (RVD)". Cada RVD reconoce un único nucleótido, por lo que la matriz de repeticiones múltiples identifica una secuencia de nucleótidos única de longitud correspondiente al número de repeticiones de la matriz, a veces denominada "la secuencia diana". Los RVD de la matriz forman juntos el dominio de unión que se une a una secuencia diana.

[0048] Como se muestra en la FIG. 1B, otra pluralidad de secuencias de nucleótidos 101 puede codificar la primera inteína 104 fusionada a al menos una porción del primer TALE 102, la segunda inteína 106 fusionada a al menos una porción de la nucleasa de corte raro 108, y una primera inteína 112 fusionada a al menos una porción de un segundo TALE 110. Las primeras inteínas 104, 112 pueden incluir secuencias de nucleótidos que codifican la misma inteína (por ejemplo, la secuencia de inteína aparece

dos veces). El segundo TALE 110 puede incluir un N-terminal de TALE 109, una segunda pluralidad de secuencias de repetición de TALE 111 (ilustrado como un dominio de unión), y un C-terminal TALE 113. La segunda pluralidad de secuencias de repetición de TALE 111, en combinación, se puede unir a una segunda secuencia de nucleótidos en una secuencia de ADN diana, y se puede denominar como un dominio de unión.

[0049] Como se muestra en la FIG. 1B, la pluralidad de secuencias de nucleótidos 101 puede incluir una primera secuencia de nucleótidos que codifica la primera inteína 104 fusionada al primer TALE 102, una segunda secuencia de nucleótidos que codifica la primera inteína 112 fusionada al segundo TALE 110, y una tercera secuencia de nucleótidos que codifica la segunda inteína 106 fusionada a la nucleasa de corte raro 108. Sin embargo, los ejemplos no son tan limitados y la primera y/o segunda inteínas se pueden fusionar a porciones del primer TALE 102, segundo TALE 110, y/o nucleasa de corte raro 108, como también se ilustra en la FIG. 1C.

[0050] En algunas realizaciones, la pluralidad de secuencias de nucleótidos 100, 101 puede formar un vector separado. En algunas realizaciones, la pluralidad de secuencias de nucleótidos 100, 101 se puede formar en un constructo de expresión único, tal como se ilustra en las FIG. 3A-3B.

[0051] Las primeras inteínas 104, 112 y la segunda inteína 106 se pueden configurar para autoempalmarse cuando están en contacto. En respuesta al autoempalme, se puede formar una primera mitad de las TALEN que incluye el primer TALE 102 unido a la nucleasa de corte raro 108. Además, se puede formar una segunda mitad de las TALEN que incluye el segundo TALE 110 unido a la nucleasa de corte raro 108. Adicionalmente, se puede formar una proteína empalmada que incluye la primera inteína 104, 112 unida a la segunda inteína 106. En algunos ejemplos, la proteína empalmada es una proteína trans-empalmada. En algunos ejemplos, la proteína cortada es una proteína cis-empalmada.

[0052] Tal como se utilizan en la presente, las inteínas son fragmentos o elementos proteicos internos que se autoextinguen de otras proteínas a las que las inteínas están unidas y las inteínas catalizan la ligación de componentes flanqueantes, a veces denominados exteínas, con un enlace peptídico. La escisión de inteínas es un proceso postraducciona que puede no requerir enzimas auxiliares o cofactores. Este proceso de autoexcisión se denomina "empalme de proteínas" por analogía con el empalme de intrones de ARN a partir de ARNpre-m (Perler F et al., Nucl Acids Res. 22:1125-1127 (1994)). Las primeras inteínas 104, 112 y la segunda inteína 106 pueden incluir

respectivamente una inteína N-terminal y una inteína C-terminal que tengan afinidad entre sí, y que se unan cuando estén en contacto. Las primeras inteínas 104, 112 y la segunda inteína 106 se pueden denominar inteínas de trans-empalme. Con el trans-empalme, una inteína es una inteína N-terminal (por ejemplo, un fragmento) que se une a una N-exteína y la otra inteína es una inteína C-terminal que se une a una C-exteína. La inteína N-terminal y la inteína C-terminal se unen entre sí, se auto-empalman y catalizan la ligación de la N-exteína y la C-exteína. En algunos ejemplos, las secuencias de inteína derivan de *Synechocystis* sp, *Saccharomyces cerevisiae*, *Pyrococcus horikoshii*, *Mycobacterium xenopi*, *Thermococcus kodakarensis*, *Methanocaldococcus jannaschii* y *Nostoc punctiforme*, entre otros. Las inteínas de ejemplo incluyen inteína Tfu pol-1, inteínas de ADN polimerasa (DnaE) (por ejemplo, Ssp DnaE, Npu DnaE), Gp41-1, Mxe GyrA, Mru RecA, MTU RecA, Tli Pol-2, Sce VMA y Ssp ADN helicasa (Dna B), entre otras.

[0053] En algunas realizaciones, la primera y segunda inteínas 104, 112, 106 pueden ser inteínas ortogonales. Por ejemplo, una inteína N-terminal de *Synechocystis* se puede unir a una inteína C-terminal de *Nostoc*, y/o dos inteínas N-terminal diferentes se pueden unir a la misma inteína C-terminal.

[0054] Como también se ilustra en la presente, tal como mediante la FIG. 1D, la pluralidad de secuencias de nucleótidos 100, 101 también puede cada una codificar elementos adicionales, tales como un promotor y/o un terminador.

[0055] La FIG. 1C ilustra un ejemplo de una pluralidad de secuencias de nucleótidos 115 que codifican una inteína fusionada a una porción de la nucleasa de corte raro y un TALE. Por ejemplo, la primera secuencia de nucleótidos codifica la primera inteína 104 fusionada a una primera porción de la nucleasa de corte raro 108-1 fusionada al primer TALE 102. La segunda secuencia de nucleótidos codifica la segunda inteína 106 fusionada a una segunda porción de la nucleasa de corte raro 108-2. Como se muestra, la primera porción de la nucleasa de corte raro 108-1 se fusiona entre el primer TALE 102 y la primera inteína 104. De modo similar a la pluralidad de secuencias de nucleótidos 100, 101, la primera inteína 104 y la segunda inteína 106 se pueden configurar para autoempalmar cuando están en contacto. En respuesta al empalme, la primera y segunda porciones de la nucleasa de corte raro 108-1, 108-2 se pueden unir entre sí para formar una primera mitad de las TALEN con el primer TALE 102. Aunque no se ilustra en la FIG. 1C, en algunas realizaciones, una tercera secuencia de nucleótidos puede codificar la

primera inteína 104 fusionada a una primera porción de la nucleasa de corte raro 108-1 fusionada al segundo TALE, y, en respuesta al empalme, la primera y segunda porciones de la nucleasa de corte raro 108-1, 108-2 pueden unir entre sí para formar una segunda mitad de las TALEN con el segundo TALE.

[0056] La FIG. 1D ilustra un ejemplo de una pluralidad de secuencias de nucleótidos 117 que incluye además cada una un promotor y un terminador. Por ejemplo, la primera secuencia de nucleótidos codifica un primer promotor 116, el primer TALE 102, la primera inteína 104, un primer terminador 118, y opcionalmente, una porción de la nucleasa de corte raro 108-2. La segunda secuencia de nucleótidos codifica un segundo promotor 124, el segundo TALE 110, la primera inteína 112, un segundo terminador 126, y opcionalmente, una porción de la nucleasa de corte raro 108-3. La tercera secuencia de nucleótidos codifica un tercer promotor 120, una segunda inteína 106, al menos una porción de la nucleasa de corte raro 108-2, y un tercer terminador 122. En algunas realizaciones, la segunda inteína 106 está fusionada a una segunda porción de la nucleasa de corte raro 108-2, y las primeras inteínas 104, 112 están respectivamente fusionadas a una primera porción de la nucleasa de corte raro 108-1, 108-3 y el primer TALE 102 o el segundo TALE 110.

[0057] Sin embargo, las realizaciones no son tan limitadas y en varias realizaciones, la primera inteína 104, 112 se puede fusionar a una primera porción del primer TALE 102 y/o una primera porción del segundo TALE 110, y/o la segunda inteína 106 está fusionada a la segunda porción del primer TALE 102 y/o la segunda porción del segundo TALE 110 que está fusionada a la nucleasa de corte raro como una secuencia de nucleótidos única. En otras realizaciones, la segunda inteína 106 está fusionada a la nucleasa de corte raro como una secuencia de nucleótidos única.

[0058] Como ejemplos no limitantes, los promotores pueden incluir un promotor de nopalina sintasa (NosPro) o un promotor T7, entre otros. Otros ejemplos de promotores pueden incluir el promotor Sp6, un promotor T3, un promotor Ubi, un promotor 35S del virus del mosaico de coliflor (CaMV), un promotor ADHI, y un promotor ADH1, un promotor GDS, un promotor TEF1, un promotor Gal1, un promotor CaMKIIa, un promotor T7lac, un promotor araBAD, un promotor trp, un promotor lac, un promotor Ptac, entre otros.

[0059] Como ejemplos no limitantes, los terminadores pueden incluir terminador Nos (NosTerm), terminador CaMV, t7S, tE9, tmas, tocs, tTr9, tpinIII, tORF25, ttml, entre otros.

[0060] La FIG. 2 es un diagrama de flujo que ilustra un ejemplo del procedimiento para generar las TALEN usando ejemplos de secuencias de nucleótidos, conforme a la presente descripción. Las secuencias de nucleótidos usadas en el procedimiento 230 pueden incluir las secuencias de nucleótidos 100, 101, 115, 117 de cualquiera de las FIG. 1A-1D.

[0061] En 232, el procedimiento 230 incluye poner en contacto una célula con una primera secuencia de nucleótidos, una segunda secuencia de nucleótidos, y una tercera secuencia de nucleótidos. La primera secuencia de nucleótidos codifica una primera inteína fusionada a al menos una porción de un primer TALE. La segunda secuencia de nucleótidos codifica la primera inteína fusionada a al menos una porción de un segundo TALE. La tercera secuencia de nucleótidos codifica una segunda inteína fusionada a al menos una porción de una nucleasa de corte raro.

[0062] En 234, en respuesta al contacto con la célula, el procedimiento 230 incluye el empalme del primer TALE, el segundo TALE, y la nucleasa de corte raro mediante la primera inteína y la segunda inteína para formar una primera mitad de las TALEN que incluye el primer TALE y la nucleasa de corte raro, y una segunda mitad de las TALEN que incluye el segundo TALE y la nucleasa de corte raro. La primera y segunda mitad de las TALEN pueden incluir las medias izquierda y derecha TALEN.

[0063] Por ejemplo y en respuesta al contacto con la célula con las secuencias de nucleótidos, el procedimiento 230 puede incluir transcribir y/o traducir la primera secuencia de nucleótidos, la segunda secuencia de nucleótidos, y la tercera secuencia de nucleótidos mediante la célula. En respuesta a la transcripción y/o traducción, se pueden formar o expresar la primera inteína fusionada al primer TALE, la primera inteína fusionada al segundo TALE, y la segunda inteína fusionada a la nucleasa de corte raro. En varias realizaciones, una pluralidad de copias de cada una de la primera secuencia de nucleótidos, la segunda secuencia de nucleótidos, y la tercera secuencia de nucleótidos se puede transcribir y/o traducir, lo que produce una pluralidad de copias de cada una de la primera inteína fusionada al primer TALE, la primera inteína fusionada al segundo TALE, y la segunda inteína fusionada a la nucleasa de corte raro.

[0064] La primera inteína y segunda inteína se pueden autoempalmar cuando están en contacto. El proceso de empalme puede producir intermediarios se forman antes de formar la primera y segunda mitad de las TALEN. Por ejemplo, la primera inteína y la segunda inteína se pueden unir para formar un primer intermediario y un segundo intermediario, como también se ilustra en la FIG. 4. El primer intermediario puede incluir la primera inteína unida a la segunda inteína, donde la primera inteína está fusionada al primer TALE y la segunda inteína fusionada a la nucleasa de corte raro. El segundo intermediario puede incluir la primera inteína unida a la segunda inteína, donde la primera inteína está fusionada al segundo TALE y la segunda inteína está fusionada a la nucleasa de corte raro.

[0065] Como se describió anteriormente, el empalme incluye la unión entre las inteínas, el corte en los sitios de empalme y la unión entre componentes. Por ejemplo, el empalme comprende la unión de la primera inteína a la segunda inteína, el corte en los sitios de empalme asociados con la primera inteína y la segunda inteína, y la unión del primer TALE a la nucleasa de corte raro y el segundo TALE a la nucleasa de corte raro para formar la primera mitad de las TALEN y la segunda mitad de las TALEN. En algunos ejemplos, el empalme puede incluir la unión de la primera y segunda porciones de la nucleasa de corte raro entre sí para formar la nucleasa de corte raro completa.

[0066] Los sitios de empalme pueden ser entre componentes de la pluralidad de secuencias de nucleótidos. Por ejemplo, los sitios de empalme pueden ser entre la primera inteína y el primer TALE, entre la primera inteína y el segundo TALE, y entre la segunda inteína y la nucleasa de corte raro o porciones de estas. En algunas realizaciones, los sitios de empalme pueden estar entre la segunda inteína y cada una de la primera y segunda porciones de la nucleasa de corte raro.

[0067] En algunas realizaciones, el procedimiento 230 también puede incluir la transformación de la célula usando la primera mitad de las TALEN y la segunda mitad de las TALEN. Por ejemplo, la primera mitad de las TALEN y la segunda mitad de las TALEN se pueden unir a una secuencia de ADN diana por medio de los dominios de unión y, en respuesta, las endonucleasas de la primera mitad de las TALEN y la segunda mitad de las TALEN pueden causar una ruptura de cadena doble en o cerca de la secuencia diana.

[0068] Varias realizaciones se refieren a sistemas que incluyen la pluralidad de secuencias de nucleótidos. En algunas realizaciones, un constructo de expresión único

puede incluir cada una de la pluralidad de secuencias de nucleótidos. En otras realizaciones y/o además, cada secuencia de nucleótidos puede formar un vector individual.

[0069] Las FIG. 3A-3B son diagramas que ilustran ejemplos de constructos de expresión, tales como constructos de expresión que incluyen las secuencias de nucleótidos ilustradas por las FIG. 1A-1D, conforme a la presente descripción. Aunque las FIG. 3A-3B ilustran un constructo de expresión único 340, 360, las realizaciones no son tan limitadas y las respectivas secuencias de nucleótidos pueden estar en vectores separados que forman el sistema.

[0070] El sistema o constructo de expresión 340, 360 comprenden la primera secuencia de nucleótidos 341-A, 341-B, segunda secuencia de nucleótidos 343-A, 343-B, y tercera secuencia de nucleótidos 345-A, 345-B descritas anteriormente. La primera secuencia de nucleótidos 341-A, 341-B codifica una primera inteína 304, 362 fusionada a al menos una porción de n primer TALE 303, 305, 307. El primer TALE 303, 305, 307 puede incluir un N-terminal de TALE 303, una primera pluralidad de secuencias de repetición de TALE, por ejemplo, el primer dominio de unión (BD1) 305, y un C-terminal TALE 307. La primera pluralidad de secuencias de repetición de TALE se unen a una primera secuencia de nucleótidos en una secuencia de ADN diana. La primera secuencia de nucleótidos 341-A, 341-B además incluye un primer promotor 342 y un primer terminador 344. El primer promotor 342 puede estar corriente arriba del primer TALE 303, 305, 307 y la primera inteína 304, 362 y el primer terminador 344 pueden estar corriente abajo del primer TALE 303, 305, 307 y la primera inteína 304, 362.

[0071] La segunda secuencia de nucleótidos 343-A, 343-B codifica la primera inteína 312, 364 fusionada a al menos una porción de un segundo TALE 309, 311, 313. El segundo TALE 309, 311, 313 puede incluir un N-terminal de TALE 309, una segunda pluralidad de secuencias de repetición de TALE, por ejemplo, el segundo dominio de unión (BD2) 311, y un C-terminal TALE 313. La segunda pluralidad de secuencias de repetición de TALE se unen a una segunda secuencia de nucleótidos en la secuencia de ADN diana. La segunda secuencia de nucleótidos 343-A, 343-B además incluye un segundo promotor 346 y un segundo terminador 348. El segundo promotor 346 puede estar corriente arriba del segundo TALE 309, 311, 313 y la primera inteína 312, 364 y el segundo terminador 348 pueden estar corriente abajo del segundo TALE 309, 311, 313 y la primera inteína 312, 364.

[0072] La tercera secuencia de nucleótidos 345-A, 345-B codifica una segunda inteína 306, 366 fusionada a al menos una porción de una nucleasa de corte raro 308. La tercera secuencia de nucleótidos 345-A, 345-B además incluye un tercer promotor 350 y un tercer terminador 352. El tercer promotor 350 puede estar corriente arriba de la nucleasa de corte raro 308 y la segunda inteína 306, 366 y el tercer terminador 352 pueden estar corriente abajo de la nucleasa de corte raro 308 y la segunda inteína 306, 366.

[0073] Como se muestra en el constructo de expresión 340 de la FIG. 3A, en algunas realizaciones la primera secuencia de nucleótidos 341-A codifica una inteína N-terminal 304 fusionada a un C-terminal 307 del primer TALE 303, 305, 307. La segunda secuencia de nucleótidos 343-A codifica la inteína N-terminal 312 fusionada a un C-terminal 313 del segundo TALE 309, 311, 313. Y, la tercera secuencia de nucleótidos 345-A codifica una inteína C-terminal 306 fusionada a un N-terminal de la nucleasa de corte raro 308 o fusionada entre las porciones de la nucleasa de corte raro 308. En tales realizaciones, el primer TALE 303, 305, 307 y segundo TALE 309, 311, 313 están respectivamente corriente arriba de las primeras inteínas 304, 312. La segunda inteína 306 está corriente arriba de al menos una porción de la nucleasa de corte raro 308.

[0074] Como se muestra en el constructo de expresión 360 de la FIG. 3B, en algunas realizaciones la primera secuencia de nucleótidos 341-B codifica una inteína C-terminal 362 fusionada a un N-terminal 303 del primer TALE 303, 305, 307. La segunda secuencia de nucleótidos 343-B codifica la inteína C-terminal 364 fusionada a un N-terminal 309 del segundo TALE 309, 311, 313. Y, la tercera secuencia de nucleótidos 345-B codifica una inteína N-terminal 366 fusionada a un C-terminal de la nucleasa de corte raro 308 o fusionada entre las porciones de la nucleasa de corte raro 308. En tales realizaciones, el primer TALE 303, 305, 307 y segundo TALE 309, 311, 313 están respectivamente corriente abajo de las primeras inteínas 362, 364. La segunda inteína 366 está corriente abajo de al menos una porción de la nucleasa de corte raro 308.

[0075] Como se usa en la presente, un constructo de expresión incluye y/o se refiere a una secuencia de nucleótidos (por ejemplo, una secuencia de ácido nucleico o secuencia de ADN) que incluye uno o más vectores o vectores binarios portadores de reactivos de edición del genoma. Los reactivos de edición del genoma pueden incluir o codificar una nucleasa y/o una TALE. En algunas realizaciones, el constructo de expresión puede incluir una variedad de secuencias de nucleótidos, seleccionadas y dispuestas para facilitar el transporte de los reactivos de edición del genoma en las células. Por ejemplo,

el constructo de expresión puede incluir la primera secuencia de nucleótidos, la segunda secuencia de nucleótidos y la tercera secuencia de nucleótidos descritas anteriormente. La nucleasa de corte raro puede incluir una proteína FokI, entre otras nucleasas. En algunas realizaciones, el constructo y/o los vectores de expresión pueden incluir otros componentes, tal como una marca detectable, un promotor y un terminador. La marca detectable puede incluir una proteína fluorescente, un fluoróforo o un nucleótido unido a un fluoróforo, entre otros tipos de marcas.

[0076] Un vector o vector binario incluye o se refiere a una secuencia de ácido nucleico que incluye uno o más transgenes, a veces denominados "insertos", y un esqueleto. El vector binario puede incluir un casete de expresión que incluye el transgén y una secuencia reguladora para ser expresada por una célula transformada.

[0077] Como se usa en la presente, un dominio incluye y/o se refiere a una parte conservada de una secuencia proteica y estructura terciaria de la proteína que puede formar una estructura tridimensional. Los dominios pueden ser codificados por los constructos de expresión.

[0078] Como también se ilustra en la presente, en respuesta a la transcripción y/o traducción de los constructos de expresión 340, 360 respectivamente que comprende la primera secuencia de nucleótidos 341-A, 341-B, la segunda secuencia de nucleótidos 343-A, 343-B, y la tercera secuencia de nucleótidos 345-A, 345-B, se puede formar una pluralidad de copias de cada una de la primera inteína fusionada al primer TALE, la primera inteína fusionada al segundo TALE, y la segunda inteína fusionada a la nucleasa de corte raro. Además, las respectivas inteínas 304, 312, 306, 362, 364, 366 se pueden unir entre sí y autoempalmar para formar las primeras medias TALEN y segundas medias TALEN.

[0079] En algunas realizaciones, el primer promotor 342, el segundo promotor 346, y el tercer promotor 350 puede incluir el mismo promotor. En otras realizaciones, el primer promotor 342, el segundo promotor 346, y el tercer promotor 350 pueden incluir cada uno diferentes promotores. En realizaciones adicionales, el primer promotor 342 y el segundo promotor 346 pueden ser el mismo promotor, y el tercer promotor 350 puede ser un promotor diferente del primer promotor 342 y el segundo promotor 346. Por ejemplo, el tercer promotor 350 puede ser un promotor más fuerte que el primer promotor 342 y el segundo promotor 346, de modo que se forman copias adicionales de la segunda inteína fusionada a la nucleasa de corte raro en comparación con el número de copias de la

primera inteína fusionada al primer TALE y la primera inteína fusionada al segundo TALE después de la transcripción y/o traducción.

[0080] La FIG. 4 es un diagrama de flujo que ilustra un ejemplo del proceso de empalme mediante las inteínas para formar las TALEN, conforme a la presente descripción.

[0081] Después del contacto de una población de células con la pluralidad de secuencias de nucleótidos, la pluralidad de secuencias de nucleótidos se transcriben y/o traducen mediante las células para formar los componentes 471, 473, y 475 que incluye la primera inteína 478-A fusionada al primer TALE 474, la primera inteína 478-B fusionada al segundo TALE 482, y la segunda inteína 481 fusionada a la nucleasa de corte raro 483. Si bien se ilustra una copia única de los componentes 471, 473, y 475 mediante la FIG. 4, se puede formar una pluralidad de copias de cada uno de los componentes 471, 473, y 475.

[0082] El componente 471, 473, y 475 puede entrar en contacto con otro y en respuesta, las primeras inteínas 478-A, 478-B se unen respectivamente a copias de la segunda inteína 481. En respuesta a la unión, se forman un primer intermediario 477 y un segundo intermediario 479. El primer intermediario 477 incluye la primera inteína 478-A unida a la segunda inteína 481, donde la primera inteína 478-A está fusionada al primer TALE 474 y la segunda inteína 481 está fusionada a la nucleasa de corte raro 483. El segundo intermediario 479 incluye la primera inteína 478-B unida a la segunda inteína 481, donde la primera inteína 478-B está fusionada al segundo TALE 482 y la segunda inteína 481 está fusionada a la nucleasa de corte raro 483.

[0083] Después de que las primeras inteínas 478-A, 478-B se unen respectivamente a copias de la segunda inteína 481, las inteínas 478-A, 478-B, 481 se autoempalman para formar una primera mitad de las TALEN 485, una segunda mitad de las TALEN 489, y una proteína empalmada 487. Por ejemplo, las inteínas 478-A, 478-B, 481 pueden cortar en sitios de empalme asociados con las primeras inteínas 478-A, 478-B y la segunda inteína 481, unir el primer TALE 474 a la nucleasa de corte raro 483, y unir el segundo TALE 482 a la nucleasa de corte raro 483. La primera mitad de las TALEN 485 puede incluir el primer TALE 474 unido a la nucleasa de corte raro 483 próximo al C-terminal del primer TALE 474. La segunda mitad de las TALEN 489 puede incluir el segundo TALE 482 unido a la nucleasa de corte raro 483 proximal al C-terminal del segundo TALE 482. La proteína empalmada 487 puede incluir la primera inteína 478 (por ejemplo, 478-A o 478-B) unida a la segunda inteína 481. Tal como se utiliza en la presente, una proteína empalmada incluye y/o se refiere a una proteína o complejo

proteico que incluye al menos dos dominios codificados por genes separados que se transcriben y/o traducen por separado y que se empalman para formar una única unidad, por ejemplo, un único polipéptido.

[0084] Como se describió anteriormente, la primera mitad de las TALEN 485 y la segunda mitad de las TALEN 489 pueden transformar las células. Por ejemplo, varias realizaciones se refieren a una célula y/u organismo huésped transformado por los procedimientos, vectores, constructos de expresión, secuencias de nucleótidos, y/o sistemas descritos en la presente. En algunas realizaciones, las células se pueden transformar y/o transformar o regenerar el organismo como se describe en la Patente US 8440431, emitida el 14 de mayo de 2013, titulada "TAL effector-mediated ADN medication", que se incorpora en la presente en su totalidad por su enseñanza.

[0085] Como se usa en la presente, el contacto de la población de células con la pluralidad de secuencias de nucleótidos puede incluir la administración de un constructo de expresión a la población de células. El constructo de expresión se puede introducir en las células a través de diferentes enfoques que incluyen, pero no se limitan a, transformación mediada por PEG, infección por *Agrobacterium*, electroporación, bombardeo de partículas o transformación de protoplastos mediada por microinyección, así como combinaciones de los mismos.

[0086] En varias realizaciones, antes de poner en contacto una población de células con la pluralidad de secuencias de nucleótidos, tal como un constructo de expresión que comprende la pluralidad de secuencias de nucleótidos. La pluralidad de secuencias de nucleótidos se puede generar usando técnicas moleculares estándares.

[0087] En algunos ejemplos, la población de células se puede analizar para identificar células diana que estén genéticamente transformadas por la pluralidad de secuencias de nucleótidos y/o el constructo de expresión. Las células diana, tal como se utilizan en la presente, incluyen y/o se refieren a células que expresan la pluralidad de secuencias de nucleótidos y/o que de otro modo exhiben o expresan la modificación genética. Las células diana pueden incluir la mutación prevista en el locus genómico diana. En algunas realizaciones, la población de células se puede analizar y las células diana se pueden seleccionar para la expresión del constructo de expresión mediante una marca detectable. Las pruebas para la población de células para la marca detectable puede incluir el aislamiento de las células diana que tienen la marca detectable de un resto de la población

de células. Varias realizaciones incluyen la selección de células transformadas basada en la clasificación celular activada por fluorescencia (FACS).

[0088] En consecuencia, una serie de realizaciones se refieren a la combinación de edición génica de células mediada por ADN, junto con la selección de células diana que reciben ambas medias TALEN utilizando FACS y proteínas fluorescentes o marcación de fluoróforo de las dos TALEN. Los organismos regenerados a partir de células seleccionadas por FACS se pueden enriquecer para las ediciones génicas pretendidas, de este modo se reducen los esfuerzos de pruebas típicamente requeridos con la expresión génica transitoria.

[0089] Varias realizaciones de la presente descripción se dirigen a una célula huésped no natural y/u organismos generados por el procedimiento 230 descrito por la FIG. 2 y/o usando la pluralidad de secuencias de nucleótidos o componentes ilustrados por las FIG. 1A-1D, 3A-3B, y 4. Por ejemplo, el procedimiento 230 puede incluir además cultivar las células diana identificadas que se transforman con la pluralidad de secuencias de nucleótidos, y regenerar un organismo a partir de las células diana cultivadas, donde los organismos regenerados expresan la modificación diana. La pluralidad de secuencias de nucleótidos y/o los componentes resultantes (por ejemplo, medias TALEN y proteínas empalmadas) se pueden eliminar (por ejemplo, separar) del organismo regenerado.

[0090] En algunas realizaciones y en consonancia con el procedimiento 230, se puede generar un organismo no natural mediante una técnica de edición genómica que incluye el uso de la pluralidad de secuencias de nucleótidos. La pluralidad de secuencias de nucleótidos pueden ser vectores separados y/o formados en un constructo de expresión único. La técnica de edición genómica puede incluir el contacto de una población de células con la pluralidad de secuencias de nucleótidos, el análisis de la población de células para identificar células diana que se transforman con la pluralidad de secuencias de nucleótidos y, opcionalmente, la regeneración de un organismo no natural a partir de las células diana identificadas.

[0091] La célula y/o población celular, como se usa en la presente, puede ser de una variedad de diferentes tipos de organismos. Ejemplos de células pueden ser de mamíferos, aves, reptiles, anfibios, peces, insectos, crustáceos, arácnidos, equinodermos, gusanos, moluscos, esponjas, plantas, hongos, algas, bacterias, entre otros.

[0092] Como se usa en la presente, corriente arriba puede incluir una localización proximal y/o más cercana al extremo 5' de la secuencia de nucleótidos en comparación

con la secuencia de referencia. Por el contrario, corriente abajo puede incluir una localización proximal y/o más cercana al extremo 3' de la secuencia de nucleótidos en comparación con la secuencia de referencia. Como se usa en la presente, una secuencia con adjetivos enumerados delante, tal como la secuencia de nucleasa de corte raro, la secuencia de inteína o la secuencia TALE, incluye o se refiere a una secuencia de nucleótidos que codifica o es los adjetivos (por ejemplo, codifica o es la nucleasa).

[0093] A continuación se describen diferentes enfoques de ejemplo para el enriquecimiento y/o selección de las células para las ediciones génicas deseadas. El enriquecimiento y/o selección de las células puede aumentar la representación de células que probablemente contengan la edición genómica deseada.

[0094] La pluralidad de secuencias de nucleótidos se puede administrar en células u otros tejidos utilizando una variedad de procedimientos conocidos, tales como transformación mediada por PEG, electroporación, bombardeo o transformación mediada por microinyección. En el caso de tejidos más grandes con paredes celulares, tales como embriones, se pueden utilizar como procedimientos de administración el bombardeo (o la biolística) con partículas de oro recubiertas de ADN. Después de la administración de las secuencias de nucleótidos, se puede utilizar FACS para seleccionar células positivas de color fluorescente.

[0095] Para la transformación por bombardeo de partículas, los constructos de expresión se pueden recubrir sobre partículas, tales como partículas de oro. Para recubrir el ácido nucleico sobre las partículas de oro, se mezclan diferentes volúmenes de solución de ácido nucleico con una cantidad fija de suspensión de oro mediante pipeteo.

[0096] Sin embargo las realizaciones no son tan limitadas, y se pueden utilizar diversos protocolos de transformación por bombardeo de partículas.

[0097] Por conveniencia, en la presente se proporcionan ciertos términos empleados en la memoria descriptiva, ejemplos y reivindicaciones anexas. Las definiciones se proporcionan para ayudar a describir realizaciones particulares y no pretenden limitar la invención reivindicada, ya que el alcance de la invención está limitado únicamente por las reivindicaciones.

[0098] El uso del término "o" en las reivindicaciones y la memoria descriptiva se utiliza para significar "y/o" a menos que se indique explícitamente que se refiere solo a alternativas o que las alternativas son mutuamente excluyentes, aunque la descripción apoya una definición que se refiere solo a alternativas y "y/o".

[0099] A menos que el contexto requiera claramente lo contrario, a lo largo de la descripción y las reivindicaciones, las palabras "incluye", "que incluye", "comprende", "que comprende", y similares, se deben interpretar en un sentido abierto e inclusivo en contraposición a un sentido cerrado, exclusivo o exhaustivo. Por ejemplo, el término "que comprende" se puede interpretar que indica "que incluye, pero no se limita a". Las palabras que utilizan el número singular o plural también incluyen el número plural y singular, respectivamente. Las palabras "un" y "una", cuando se utilizan junto con la palabra "que comprende" o "que incluye" en las reivindicaciones o la memoria descriptiva, indican uno o más, a menos que se indique específicamente.

[0100] Como se usa en la presente, el término "polipéptido" o "proteína" incluye y/o se refiere a un polímero en el que los monómeros son residuos de aminoácidos que se unen mediante enlaces amida. Cuando los aminoácidos son alfa-aminoácidos, se puede utilizar el isómero L-óptico o el isómero D-óptico, el típico es el L-isómero. El término polipéptido o proteína, como se usa en la presente, abarca cualquier secuencia de aminoácidos e incluye secuencias modificadas, tales como glicoproteínas. El término polipéptido, a menos que se indique lo contrario, se refiere específicamente a las proteínas naturales, así como a las producidas de forma recombinante o sintética.

[0101] El término "secuencia de nucleótidos" incluye y/o se refiere a una pluralidad de nucleótidos en una cadena o secuencia. Un nucleótido incluye y/o se refiere a un compuesto que incluye un nucleósido (por ejemplo, una nucleobase con un azúcar de carbono) unido a un grupo fosfato. La secuencia de nucleótidos se puede denominar a veces ácido nucleico, los nucleótidos de la secuencia forman los bloques de construcción del ácido nucleico. El término "ácido nucleico" incluye y/o se refiere al ácido nucleico de ADN o ARN y una secuencias de ácidos nucleicos en forma de cadena simple o doble, a menos que se indique lo contrario, abarca análogos conocidos de nucleótidos naturales que hibridan con ácidos nucleicos de forma similar a los nucleótidos naturales. A menos que se indique lo contrario, una secuencia particular de nucleótidos o de ácidos nucleicos incluye su secuencia complementaria.

[0102] Se describen materiales, composiciones y componentes que se pueden usar para, se pueden usar en conjunción con, se pueden usar en preparación para, o son productos de los procedimientos y composiciones divulgados. Se entiende que, cuando se describen combinaciones, subconjuntos, interacciones, grupos, etc., de estos materiales, se contempla específicamente cada una de las diversas combinaciones individuales y

colectivas, aunque no se describa explícitamente la referencia específica a todas y cada una de las combinaciones y permutaciones de estos compuestos. Este concepto se aplica a todos los aspectos de la presente descripción, que incluye, pero sin limitación, las etapas de los procedimientos descritos. Los elementos específicos de cualquiera de las realizaciones anteriores se pueden combinar o sustituir por elementos de otras realizaciones. Por ejemplo, si hay una variedad de etapas adicionales que se pueden realizar, cada uno de estas etapas adicionales se puede realizar con cualquier etapa específica del procedimiento o combinación de etapas del procedimiento de los procedimientos descritos, y cada tal combinación o subconjunto de combinaciones se contempla y se describe específicamente. Además, se entiende que las realizaciones descritas en la presente pueden implementarse utilizando cualquier material adecuado, como los descritos en la presente o los conocidos en la técnica.

[0103] Varias realizaciones se implementan de acuerdo con la solicitud provisional subyacente, Solicitud Provisional US N° 63/219,291, presentada el 7 de julio de 2021, y titulada " Transcription Activator-Like Effectors Fused to Inteins" a la cual se reivindica beneficio y se incorpora completamente en la presente por referencia. Por ejemplo, las realizaciones del presente documento y/o de la solicitud provisional se pueden combinar en diversos grados (incluso totalmente). Las realizaciones descritas en las solicitudes provisionales no pretenden, en modo alguno, limitar la descripción técnica general, ni ninguna parte de la invención reivindicada, a menos que se indique específicamente.

[0104] Aunque se han ilustrado y descrito realizaciones ilustrativas, se apreciará que se pueden introducir diversos cambios en las mismas sin apartarse del alcance de la invención.

REALIZACIONES EXPERIMENTALES

[0105] Varias realizaciones experimentales se dirigieron a diseñar diferentes vectores de ácido nucleico, a veces denominados en la presente vectores para facilitar la referencia y que pueden incluir los constructos de expresión descritos anteriormente o una parte de las mismas, tal como un constructo de ADN o ARNm. Los vectores codifican una primera inteína fusionada a al menos una porción de un segundo TALE, y una segunda inteína fusionada a al menos una porción de una nucleasa de corte raro. Se diseñaron experimentos específicos para demostrar la edición genética mediante los vectores descritos anteriormente. En la presente se describen algunos de los experimentos realizados.

menos una porción de un segundo TALE, y una segunda inteína fusionada a al menos una porción de una nucleasa de corte.

[0106] En varias realizaciones experimentales, se diseñaron y construyeron varios vectores. Los vectores incluyeron un efector TAL fusionado a una secuencia peptídica int-N de Ssp DnaE, y una endonucleasa fusionada a una secuencia peptídica int-C de Ssp DnaE. Juntas, las secuencias peptídicas int-N e int-C de Ssp DnaE constituyen una trans-inteína, denominada aquí inteína. Se realizaron experimentos específicos para demostrar la edición genética a partir de la actividad conjunta de estos vectores inteína. Los ejemplos de constructos y secuencias utilizadas para las realizaciones experimentales incluyen las secuencias de nucleótidos expuestas en SEQ ID NOs: 1-159. Los SEQ ID NOs: 1-164 son cada uno ADN sintético.

[0107] Los diferentes vectores se muestran a continuación en la Tabla 1. Los vectores de ácido nucleico de la Tabla 1 incluyen constructos de ADN.

Tabla 1

Nombre	Composición	Descripción
Vector plasmídico 1	NosPro-TALE-BnFAD2 (T03-L)-Ssp DnaE int-N-NosTerm	Efector TAL izquierdo con dominio de unión de ADN BnFAD2 fusionado a una secuencia Ssp DnaE int-N
Vector plasmídico 9	NosPro-TALE-BnFAD2 (T03-R)-Ssp DnaE int-N-NosTerm	Efector TAL derecho con dominio de unión de ADN BnFAD2 fusionada a una secuencia Ssp DnaE int-N
Vector plasmídico 13	NosPro-TALE-Ssp DnaE int-N-NosTerm	Efector TAL izquierdo fusionado a una secuencia Ssp DnaE int-N, incluye sitios BsaI para la clonación de GG TALE
Vector plasmídico 15	NosPro-TALE-Ssp DnaE int-N-NosTerm	Efector TAL derecho fusionado a una secuencia Ssp DnaE int-N, incluye sitios BsaI para la clonación de GG TALE

Vector plasmídico 17	NosPro- Ssp DnaE int-C-FokI- NosTerm	FokI endonucleasa fusionada a una secuencia peptídica Ssp DnaE int-C
Vector plasmídico Control 1	NosPro-TALE-BnFAD2 (T03-L)- N-NosTerm	Control LHT con dominio de unión de ADN BnFAD2
Vector plasmídico Control 2	NosPro-TALE-BnFAD2 (T03-R)- N-NosTerm	Control RHT con dominio de unión de ADN BnFAD2

[0108] Los constructos en la Tabla 1 se generaron en las realizaciones experimentales se describen en detalle a continuación. Los vectores plasmídicos 1 y 9 codifican un efector TAL dirigido al gen BnFAD2 fusionado a una secuencia peptídica int-N Ssp DnaE. El vector plasmídico 1 (SEQ ID NO: 1) codifica un promotor NosPro, un efector TAL dirigido al gen BnFAD2, una secuencia peptídica Ssp DnaE int-N, un terminador NosTerm y un esqueleto de mitad de las TALEN (HT) izquierda. El vector plasmídico 9 (SEQ ID NO: 9) codifica un promotor NosPro, un efector TAL dirigido al gen BnFAD2, una secuencia peptídica Ssp DnaE int-N, un terminador NosTerm, y un esqueleto HT derecho. El vector plasmídico 17 (SEQ ID NO: 17) codifica una secuencia peptídica Ssp DnaE int-C fusionada a la FokI endonucleasa. El vector plasmídico 17 codifica un promotor NosPro, una secuencia peptídica int-C de Ssp DnaE, una FokI endonucleasa y un terminador NosTerm. En las realizaciones experimentales, los vectores plasmídicos 1, 9 y 17 se utilizaron conjuntamente para demostrar la actividad de edición génica TALEN.

[0109] Una secuencia del vector plasmídico 1 se expone en la SEQ ID NO: 1, que codifica un promotor Nos (SEQ ID NO: 2), N-terminal de TALE izquierdo (SEQ ID NO: 3), dominio de unión BnFAD2 (T03-L) (SEQ ID NO: 4), C-terminal de TALE izquierdo (SEQ ID NO: 5), un ligador (SEQ ID NO: 6), Ssp DnaE int-N (SEQ ID NO: 7), y un terminador Nos (SEQ ID NO: 8). Una secuencia del vector plasmídico 9 se expone en la SEQ ID NO: 9, que codifica un promotor Nos (SEQ ID NO: 2), N-terminal de TALE derecho (SEQ ID NO: 10), dominio de unión BnFAD2 (T03-R) (SEQ ID NO: 11), C-terminal TALE derecho (SEQ ID NO: 12), un ligador (SEQ ID NO: 6), Ssp DnaE int-N (SEQ ID NO: 7), y un terminador Nos (SEQ ID NO: 8). Una secuencia del vector plasmídico 17 se expone en la SEQ ID NO: 17, que codifica un promotor Nos (SEQ ID

NO: 2), Ssp DnaE int-C (SEQ ID NO: 18), un ligador (AGCCGTTCC), FokI (SEQ ID NO: 19), y un terminador Nos (SEQ ID NO: 8).

[0110] La actividad de edición génica de los vectores plasmídicos 1, 9, y 17 se comparó contra los vectores plasmídicos control 1 y control 2. El vector plasmídico control 1 consiste en un promotor NosPro, un TALEN dirigido a BnFAD2 (T03-L), y un terminador NosTerm en un esqueleto de HT izquierdo. El vector plasmídico control 2 consiste en un promotor NosPro, un TALEN dirigido a BnFAD2 (T03-R), y un terminador NosTerm en un esqueleto de HT derecho.

[0111] El promotor NosPro y el terminador NosTerm se basan en secuencias de *Agrobacterium tumefaciens*, el efector TAL se basa en una secuencia de *Xanthomonas*, el BnFAD2 (T03) se dirige a una secuencia encontrada en *Brassica napus*, la FokI endonucleasa se basa en una secuencia de *Flavobacterium okeanoikoites*, y las secuencias peptídicas int-C e int-N de Ssp DnaE son de *Synechocystis* sp. PCC6803.

[0112] Los ejemplos de constructos restantes de la Tabla 1 se describen a continuación. El vector plasmídico 13 (expuesto en la SEQ ID NO: 13) y el vector plasmídico 15 (expuesto en la SEQ ID NO: 15) son vectores de entrada que codifican un efector TAL fusionado a una Secuencia peptídica Ssp DnaE int-N. El vector plasmídico 13 codifica un promotor NosPro, un efector TAL, un casete lacZ flanqueado por los sitios BsaI para la clonación Golden Gate, una secuencia Ssp DnaE int-N, y terminador NosTerm en un esqueleto de HT izquierdo. El Vector plasmídico 15 codifica un promotor NosPro, un efector TAL, un casete lacZ flanqueado por los sitios BsaI para la clonación Golden Gate, una secuencia Ssp DnaE int-N, y terminador NosTerm en un esqueleto de HT derecho.

[0113] Una secuencia del vector plasmídico 13 se expone en la SEQ ID NO: 13, que codifica un promotor Nos (SEQ ID NO: 2), N-terminal de TALE izquierdo (SEQ ID NO: 3), casete LacZ (SEQ ID NO: 14), C-terminal de TALE izquierdo (SEQ ID NO: 5), un ligador (SEQ ID NO: 6), Ssp DnaE int-N (SEQ ID NO: 7), y un terminador Nos (SEQ ID NO: 8). Una secuencia del vector plasmídico 15 se expone en la SEQ ID NO: 15, que codifica un promotor Nos (SEQ ID NO: 2), N-terminal de TALE derecho (SEQ ID NO: 10), casete LacZ (SEQ ID NO: 16), extremo c-terminal TALE derecho (SEQ ID NO: 12), un ligador (SEQ ID NO: 6), Ssp DnaE int-N (SEQ ID NO: 7), y un terminador Nos (SEQ ID NO: 8).

[0114] Se realizaron experimentos adicionales para ilustrar la transformación de células con los vectores de la Tabla 1, como se muestra en la Tabla 2. Más específicamente, se

transformaron protoplastos de canola utilizando los vectores ilustrados en la Tabla 1. Estos incluyen los vectores SEQ ID NO: 1, SEQ ID NO: 9, SEQ ID NO: 17, el vector plasmídico control 1 y el vector plasmídico control 2 descritos previamente. Todas las muestras se utilizaron para transformar 200,000 protoplastos cada una, y todas las muestras se ensayaron como replicados biológicos. Las muestras A y D incluían 30 ug cada una de un plásmido de control del proceso, que funcionó como control negativo durante el análisis de datos. Las muestras B y E incluyeron 30 ug cada una de vectores de control para la edición genética de la diana BnFAD2 (T03), el vector plasmídico de control 1 y el vector plasmídico de control. Las muestras C y F incluyeron 30 ug de los tres plásmidos de inteína SEQ ID NO: 1, SEQ ID NO: 9 y SEQ ID NO: 17. Las muestras de inteína C y F se compararon con las muestras de control positivo B y E para evaluar la eficacia de la edición genética. Todas las muestras se prepararon utilizando la misma secuencia Illumina para el análisis. Los vectores se utilizaron para transformar protoplastos de canola para comparar la eficiencia de edición génica de los vectores TALEN de inteína con los vectores TALEN sin las secuencias peptídicas de inteína. La Tabla 2 ilustra los experimentos realizados.

Tabla 2

Muestra	A	B	C	D	E	F
Plásmido 1	Control negativo	Vector plasmídico control 1	Vector plasmídico 1	Control negativo	Vector plasmídico control 1	Vector plasmídico 1
Descripción	pVaUbi3_YFP_NosTerm (control negativo)	NosPro-TALE-BnFAD2 (T03-L)-N-NosTerm	NosPro-TALE-BnFAD2 (T03-L)- Ssp DnaE int-N-NosTerm	pVaUbi3_YFP_NosTerm (control negativo)	NosPro-TALE-BnFAD2 (T03-L)-N-NosTerm	NosPro-TALE-BnFAD2 (T03-L)- Ssp DnaE int-N-NosTerm
Tipo	ADN	ADN	ADN	ADN	ADN	ADN
Por 100k células (ug)	15	15	15	15	15	15
Cantidad total(ug)	30	30	30	30	30	30
Conc. (ug/ul)	2,604	1,691	5,2	0,300	0,300	0,300

Vol. (ul)	11,52	17,74	5,77	100,00	100,00	100,00
Protoplastos #	200K	200K	200K	200K	200K	200K
Plásmido 2		Vector plasmídico control 2	Vector plasmídico 9		Vector plasmídico control 2	Vector plasmídico 9
Descripción		NosPro-TALE-BnFAD2 (T03-R)-N-NosTerm	NosPro-TALE-BnFAD2 (T03-R)- Ssp DnaE int-N-NosTerm		NosPro-TALE-BnFAD2 (T03-R)-N-NosTerm	NosPro-TALE-BnFAD2 (T03-R)-Ssp DnaE int-N-NosTerm
Tipo		ADN	ADN		ADN	ADN
Por 100k células (ug)		15	15		15	15
Cantidad total (ug)		30	30		30	30
Conc. (ug/ul)		1,838	6,3		0,300	0,300
Vol. (ul)		16,32	4,76		100,00	100,00
Plásmido 3			Vector plasmídico 17			Vector plasmídico 17
Descripción						
Tipo			ADN			ADN
Por 100k células (ug)			15			15
Cantidad total (ug)			30			30
Conc. (ug/ul)			3,1			0,300
Vol. (ul)			9,68			100,00
Nota	Control negativo	Control positivo TALEN	Muestra de TALEN inteína	Control negativo	Control positivo TALEN	Muestra de TALEN inteína

[0115] Para realizar la transformación de canola, 30-60 días antes del experimento, se lavaron 10 semillas de canola con 1,5 ml de etanol 70%, y luego 1,8 ml de agua estéril.

Para esterilizar las semillas, se utilizó 1,5 ml de una solución de hipoclorito de sodio 1% para lavar las semillas, y después se lavaron las semillas otras cinco veces con 1,8 ml de agua estéril. Después de la imbibición, seis de las semillas recién esterilizadas se plantaron en medio 8P-MS-G en una PlantCon. Las semillas se incubaron a 25C en una relación de 16/8 horas de luz/oscuridad.

[0116] Después de 30-60 días de incubación, se digirieron las plántulas de canola germinadas. Se utilizaron tijeras estériles para extirpar 4-6 hojas jóvenes de canola y las hojas se colocaron en una placa de Petri que contiene 50-100 ul de CPDS. Las hojas se cortaron en trozos de 0,5-1 mm utilizando una cuchilla estéril de borde recto. Se añadieron otros 4 ml de CPDS a la placa y ésta se colocó dentro de una placa petri más grande de 100 mm para asegurar la esterilidad. Las placas se trasladaron a una cámara de vacío y se aspiraron a 30 inHg durante 10 minutos. Después de 10 minutos, la placa se incubó a 25 rpm durante 16 horas en la oscuridad a 25C.

[0117] Después de 16 horas, la digestión de protoplastos se lavó con 4 ml de solución W-5 más Carb100, y la solución de protoplastos se pipeteó suavemente a través de un filtro de células Falcon de 40 um a un tubo de 50 ml. Esta etapa se repitió una vez más. A continuación, se centrifugó el tubo durante 5 minutos a 100xg. Después de la centrifugación, se desechó el sobrenadante y el pellet se resuspendió con 4 ml de buffer de lavado W-5, se centrifugó una vez más y se desechó el sobrenadante restante. A continuación, se resuspendió el pellet lavado con 2 ml de buffer de lavado W-5 y se cargaron 20 ul de la suspensión en un hemocitómetro para el recuento de células. Se utilizaron los recuentos celulares entre cuatro rejillas del hemocitómetro para obtener un número medio de protoplastos por rejilla. El # total de protoplastos de la muestra se calculó de la siguiente manera: $((\bar{x}_{\{a,b,c,d\}})/0,2) \times 1000 \times 2(\text{ml}) = \text{número total de protoplastos}$, donde $\bar{x}$ es la media de las cuatro rejillas {a,b,c,d} que se contaron previamente. Para realizar la transformación, el tubo de 50 ml que contiene 2 ml de buffer W-5 y la suspensión de protoplastos se centrifugó durante 5 minutos a 100xg. Para las muestras A-C, se eliminó el sobrenadante y se añadió 1 ml de MMG 1X a temperatura ambiente por 1×10^6 protoplastos, y se añadió un volumen correspondiente a 200,000 protoplastos a un tubo de microcentrífuga de 1,5 ml que contiene la cantidad especificada de vectores para cada transformación. Para las muestras D-F, se añadieron 500 ul de 2X MMG a temperatura ambiente por 1×10^6 protoplastos junto con la cantidad especificada de vectores de ADN a los 200,000 protoplastos en un tubo de microcentrífuga de 1,5 ml.

Para todas las muestras, la suspensión de protoplastos se mezcló con los vectores pipeteando lentamente el líquido arriba y abajo. Los protoplastos y los vectores plasmídicos se incubaron a temperatura ambiente durante 5 minutos. Después de 5 minutos, se añadió un volumen 1X de PEG a temperatura ambiente a cada tubo de microcentrífuga, y se pipeteó arriba y abajo hasta que se mezcló completamente. A continuación, los tubos se incubaron a temperatura ambiente durante 20 minutos. Después de 20 minutos, se añadieron 1,5 ml de buffer de lavado W-5 para resuspender los protoplastos. Los tubos se centrifugaron a 200xg durante 5 minutos, se eliminó el sobrenadante y se añadieron 800 μ l más de buffer de lavado W-5 para resuspender las células. Se transfirieron 0,5-2 ml de los protoplastos lavados a una placa de 6-24 pocillos y se incubaron durante 24-48 horas.

[0118] Para evaluar la actividad de edición génica, se recolectaron los protoplastos mediante su transferencia a un tubo de microcentrífuga de 1,5 ml. La suspensión se centrifugó a 200xg durante 5 minutos, el sobrenadante se desechó y los tubos se colocaron en nitrógeno líquido durante dos minutos. A continuación, los tubos se almacenaron a -80C hasta que se extrajo el ADN del protoplasto y se analizó con Illumina.

[0119] La Tabla 3 ilustra las supresiones detectadas en los protoplastos transformados con los vectores descritos en la Tabla 1. Las eficiencias de edición génica, mostradas en la Tabla 2, son las siguientes. Las eficiencias de edición génica, mostradas aquí como porcentaje de eventos, se compararon a través de las muestras A-F. Las muestras A y D incluyeron protoplastos de colza transformados con vectores que expresan YFP y sirvieron como control negativo, en el que no se esperaba ninguna edición. Las muestras B y E incluyeron protoplastos de canola transformados con TALENs dirigidos al gen BnFAD2. Las muestras C y F incluyeron protoplastos de canola transformados con efectores TAL y una FokI endonucleasa, cada una fusionada a una secuencia int-N o int-C de Ssp DnaE, también dirigida a BnFAD2.

[0120] La Tabla 3 muestra los resultados de un ensayo de mutación NHEJ que detecta el número de supresiones, o eventos, en la población de células protoplasto que se transformaron con los vectores anteriores. El ensayo amplifica tres regiones genómicas que contienen la diana BnFAD2, representadas en la Tabla 3 como Illumina 1, 2, y 3. Como se muestra, las muestras de TALEN inteína produjeron un número

significativamente mayor de supresiones que el control negativo YFP, aunque no produjeron supresiones con la misma frecuencia que los vectores de control TALEN.

Tabla 3

Muestra	Replicado biológico	Descripción de muestra	Porcentaje de eventos (prom entre las copias 1-3 Illumina)
A	1	pVaUbi3_YFP_NosTerm control negativo	0,006074698
A	2	pVaUbi3_YFP_NosTerm control negativo	0,007731365
B	1	Control positivo TALEN	2,460016807
B	2	Control positivo TALEN	2,114974022
C	1	Muestra de TALEN inteína	0,208803999
C	2	Muestra de TALEN inteína	0,42680571
D	1	pVaUbi3_YFP_NosTerm control negativo	0,004661722
D	2	pVaUbi3_YFP_NosTerm control negativo	0,011750975
E	1	Control positivo TALEN	4,474206527
E	2	Control positivo TALEN	5,84799401
F	1	Muestra de TALEN inteína	0,130803747
F	2	Muestra de TALEN inteína	0,223925849

[0121] Los experimentos probaron el uso de inteínas de transempalme como un procedimiento para reducir el tamaño de la carga plasmídica de los vectores TALEN, y también agregar flexibilidad al administrar materiales de edición génica a la célula. Como se describió anteriormente, los experimentos se realizaron en protoplastos de canola (Bn-Westar). La diana con TALEN T03,01 en el gen FAD2, que tiene tres copias conocidas. Los protoplastos se genotipificaron utilizando secuenciación de amplicones para detectar

ediciones en las tres copias del gen FAD2. La Tabla 4 a continuación proporciona un resumen de la cobertura de secuencia resultante.

[0122] Para los experimentos, hubo una buena cobertura en las tres copias del gen para cada muestra de inteína (por ejemplo, promedio de más de 28,000 lecturas, con un rango de entre 24,000 y 41,000 lecturas). Los experimentos 1 y 2 se realizaron para probar diferentes concentraciones de entradas de transformación. Específicamente, en el experimento 2 (el procedimiento "Mod") se utilizó un mayor volumen de ADN a una concentración más baja.

[0123] La Tabla 4 y la Tabla 5 a continuación proporcionan el porcentaje de edición para los experimentos 1 y 2. En ambos experimentos, las muestras de inteína mostraron sistemáticamente un nivel de edición superior al de los controles negativos, pero inferior al de los controles positivos. En el experimento 2 (procedimiento Mod, menor cantidad de ADN), el porcentaje de edición fue mayor en el control positivo, pero menor en las muestras de inteína. Como se puede apreciar, el porcentaje de edición = (número de lecturas con ediciones / número total de lecturas analizadas) * 100.

Tabla 4. Experimento 1 Porcentaje de edición

	Muestra A: YFP (-) control	Muestra A: YFP (-) control	Muestra B: (+) BnFAD2 (T03)	Muestra B: (+) BnFAD2 (T03)	Muestra C: Inteínas	Muestra C: Inteínas
BnaA.F AD2.a	0,007	0,003	2,568	2,110	0,265	0,504
BnaC.F AD2.a	0,009	0,006	2,687	2,228	0,237	0,548
BnaC.F AD2.b	0,003	0,014	2,125	2,007	0,125	0,228
Prom.	0,006	0,008	2,460	2,115	0,209	0,427

Tabla 5 Experimento 2 Porcentaje de edición

	Muestra D: YFP (-) control	Muestra D: YFP (-) control	Muestra E: (+) BnFAD2 (T03)	Muestra E: (+) BnFAD2 (T03)	Muestra F: Inteínas	Muestra F: Inteínas
BnaA. FAD2. a	0,004	0,010	4,775	6,835	0,171	0,252
BnaC. FAD2. a	0,007	0,009	5,048	6,122	0,177	0,316
BnaC. FAD2. b	0,003	0,016	3,599	4,587	0,044	0,104
Prom.	0,005	0,012	4,474	5,848	0,131	0,224

[0124] La FIG. 5 ilustra imágenes que muestran el genotipificado resultante de la edición de una célula usando secuencias de nucleótidos, conforme a la presente descripción. Más en particular, la FIG. 5 ilustra imágenes alineadas de regiones genómicas del control negativo 590, el control positivo 591, y la inteína 592.

[0125] Las FIG. 6A-6C ilustran imágenes completas de las regiones genómicas de un control negativo YFP, un control positivo TALEN, e inteínas, conforme a la presente descripción. Por ejemplo, la FIG. 6A ilustra una imagen de regiones genómicas de la Muestra A control negativo YFP. La FIG. 6B ilustra una imagen de regiones genómicas de la Muestra B (+) BnFAD2 (T03) Control positivo TALEN. La FIG. 6C ilustra una imagen de regiones genómicas de la Muestra C (+) Inteínas.

[0126] Se realizaron varios experimentos utilizando vectores plasmídicos adicionales para transformar células vegetales, tales como células vegetales de canola, cannabis y/o soja. Aunque los ejemplos describen células vegetales particulares, las realizaciones no están tan limitadas y pueden incluir cualquier tipo de planta y/o células que no sean vegetales, tales como células de mamífero. Se utilizaron diferentes tipos de inteínas, que incluyen Ssp DNAE y Gp41-1, exteínas nativas y no nativas, y TALEs con dominios de

unión específicos para diferentes dianas. Los experimentos incluyeron además controles positivos que incluyeron TALEN y controles negativos sin TALE.

[0127] Algunos experimentos se llevaron a cabo utilizando vectores plasmídicos adicionales para transformar células vegetales de soja utilizando los TALE asociados con diferentes genes, tales como un transgén de sintasa (ALS), transgén de ácido graso desaturasa 3 (FAD3) y transgén de factor regulador del crecimiento (GRF). Los ejemplos de constructos y secuencias utilizados para las realizaciones experimentales incluyen las secuencias de nucleótidos expuestas en SEQ ID NOs: 92-159. Las siguientes Tablas 6-8 ilustran diferentes vectores plasmídicos.

Tabla 6

Nombre	Composición	Descripción
Vector plasmídico 154	NosPro-TALE-GmALS (T04-L)- Ssp DnaE int-n- NosTerm	Efector TAL izquierdo con GmALS_T04 ADN Dominio de unión fusionada a una secuencia Ssp DnaE int-N
Vector plasmídico 155	NosPro-TALE-GmALS (T04-R)- Ssp DnaE int-n- NosTerm	Efector TAL derecho con dominio de unión ADN GmALS_T04 fusionada a una secuencia Ssp DnaE int-N
Vector plasmídico 127	NosPro-Ssp DnaE int-c- FokI-NosTerm	FokI endonucleasa fusionada a una secuencia peptídica Ssp DnaE int-C
Vector plasmídico 132	NosPro-Ssp DnaE int-c- exteína-FokI-NosTerm	FokI endonucleasa fusionada a una secuencia peptídica Ssp DnaE int-C con exteína de CFN nativa
Vector plasmídico 149	NosPro-TALE-GmALS (T04-L)-gp41-1 int-n- NosTerm	Efector TAL izquierdo con dominio de unión ADN GmALS_T04 fusionada a una secuencia gp41-1 int-n

Vector plasmídico 150	NosPro-TALE-GmALS (T04-R)-gp41-1 int-n-NosTerm	Efector TAL derecho con dominio de unión ADN GmALS_T04 fusionada a una secuencia gp41-1 int-n
Vector plasmídico 131	NosPro-gp41-1 int-c-FokI-NosTerm	FokI endonucleasa fusionada a una secuencia peptídica gp41-1 int-c
Vector plasmídico 122	NosPro-TALE-GmALS (T04-L)-N-NosTerm	Control LHT con dominio de unión ADN GmALS_T04
Vector plasmídico 123	NosPro-TALE-GmALS (T04-R)-N-NosTerm	Control RHT con dominio de unión ADN GmALS_T04

Tabla 7

Muestra	AA	G	H	I	J
Plásmido 1	Control negativo	Vector plasmídico 154	Vector plasmídico 154	Vector plasmídico 149	Vector plasmídico 122
Descripción	pVaUbi3_YFP_NosTerm (control negativo)	NosPro-TALE-GmALS (T04-L)-Ssp DnaE int-n-NosTerm	NosPro-TALE-GmALS (T04-L)-Ssp DnaE int-n-NosTerm	NosPro-TALE-GmALS (T04-L)-gp41-1 int-n-NosTerm	NosPro-TALE-GmALS (T04-L)-N-NosTerm
Tipo	ADN	ADN	ADN	ADN	ADN
Por 100k células (ug)	15	15	15	15	15
Cantidad total (ug)	30	30	30	30	30
Conc. (ug/ul)	0,9	0,9	0,9	0,9	0,9
Vol. (ul)	100	100	100	100	100
Protoplasto #	200K	200K	200K	200K	200K
Plásmido 2		Vector plasmídico 155	Vector plasmídico 155	Vector plasmídico 150	Vector plasmídico 123
Descripción		NosPro-TALE-	NosPro-TALE-	NosPro-TALE-	NosPro-TALE-

		GmALS (T04-R)- Ssp DnaE int-n- NosTerm	GmALS (T04-R)- Ssp DnaE int-n- NosTerm	GmALS (T04-R)- gp41-1 int- n-NosTerm	GmALS (T04-R)-N- NosTerm
Tipo		ADN	ADN	ADN	ADN
Por 100k células (ug)		15	15	15	15
Cantidad total (ug)		30	30	30	30
Conc. (ug/ul)		0,9	0,9	0,9	0,9
Vol. (ul)		100	100	100	100,00
Plásmido 3		Vector plasmídico 127	Vector plasmídico 132	Vector plasmídico 131	
Descripción		NosPro- Ssp DnaE int-c-FokI- NosTerm	NosPro-Ssp DnaE int-c- exteína- FokI- NosTerm	NosPro- gp41-1 int- c-FokI- NosTerm	
Tipo		ADN	ADN	ADN	
Por 100k células (ug)		15	15	15	
Cantidad total (ug)		30	30	30	
Conc. (ug/ul)		0,9	0,9	0,9	
Vol. (ul)		100	100	100	
Nota	Control negativo	Muestra de inteína	Muestra de inteína	Muestra de inteína	Control positivo

Tabla 8

Muestra	Replicado biológico	Descripción de muestra	Porcentaje de eventos
AA	1	pVaUbi3_YFP_NosTerm control negativo	0,035018
AA	2	pVaUbi3_YFP_NosTerm control negativo	0,048517
G	1	Ssp DnaE Muestra de TALEN inteína	3,728738
G	2	Ssp DnaE Muestra de TALEN inteína	4,145849
H	1	Ssp DnaE Exteína nativa Muestra de TALEN inteína	8,979292
H	2	Ssp DnaE Exteína nativa Muestra de	7,736866

		TALEN inteína	
I	1	Gp41-1 Muestra de TALEN inteína	8,30517
I	2	Gp41-1 Muestra de TALEN inteína	8,629669
J	1	Control positivo TALEN	1,643263
J	2	Control positivo TALEN	2,329923

[0128] En varios experimentos, se transformaron protoplastos de soja usando los vectores plasmídicos descritos anteriormente y de acuerdo con el protocolo descrito en Xiong, L., et al., "A transient expression system in soybean mesophyll protoplasts reveals the formation of cytoplasmic GmCRY1 photobody-like structures", Science China Life Sciences, 2019, 62(8), 1070-1077, que se incorpora por la presente en su totalidad para su enseñanza, y además de otros vectores plasmídicos. En algunos ejemplos, se combinaron 2,4 millones de células en los tubos de replicación para 12 x 200,000 células por biorreplicación. Se identificó una media de 180 por cuadrado, lo que equivale a 3,6 M de células, ya que se utilizó un volumen de 4 ml. A continuación se procedió como se describe para el resto del protocolo.

[0129] Las muestras fueron las siguientes:

dos bioreplicaciones de cada una (por ejemplo para la muestra 1: 1A y 1B con 200K células para cada una; y

cada Muestra #1-6 tiene cada plásmido a una concentración final de 300 ng/ul.

La Tabla 9 proporciona las diferentes muestras que se probaron:

Tabla 9

Muestra	Diana	Inteína	Exteína	Plásmido
1	GmFAD3_T08	Ssp DnaE	No nativo	Vector plasmídico 110 - Ssp DnaE GmGRF_T03-L1
				Vector plasmídico 152 - Ssp DnaE GmGRF_T03-R1
				Vector plasmídico 127 - Ssp DnaE intC no nativo
2	GmFAD3_T08	Ssp DnaE	Nativo	Vector plasmídico 110- Ssp DnaE GmGRF_T03-L1
				Vector plasmídico 152 - Ssp DnaE GmGRF_T03-R1

				Vector plasmídico 132 - Ssp DnaE intC nativo
3	GmFAD3_T08	Gp41-1	No nativo	Vector plasmídico 137 - GmFAD3_T08-L1 gp41-1 no nativo
				Vector plasmídico 145 - GmFAD3_T08-R1 gp41-1 no nativo
				Vector plasmídico 131- gp41-1 intC no nativo
4	GmFAD3_T08	Gp41-1	Nativo	Vector plasmídico 134- GmFAD3_T08-L1 gp41-1 nativo
				Vector plasmídico 141 - GmFAD3_T08-R1 gp41-1 nativo
				Vector plasmídico 130 - gp41-1 intC nativo
5	GmFAD3_T08	N/A		pCLS control - GmFAD3_T08-L1
				Vector plasmídico 126 - GmFAD3_T08-R1
6	Control negativo			Control negativo de plásmido

** El vector plasmídico 152 y el vector plasmídico 110 utilizaron menos [adn] para dividirse entre los experimentos 1 y 2. Para el vector plasmídico 152 (295 ng/ul) y para el vector plasmídico 110 (244 ng/ul).

[0130] Cada muestra se preparó hasta un volumen final de 220 ul para compensar el error de pipeteo, a partir de las concentraciones y volúmenes del patrón que se indican a continuación. Por ejemplo, a continuación se describen las preparaciones de protoplastos de las Muestras 1-5 en 10 placas (por ejemplo, 2 por muestra). Los protoplastos se

sumaron para incluir 1 ul de solución x1000mlx4ml (volumen total). La Tabla 10 proporciona la suma para cada preparación de muestra.

Tabla 10

Muestra	Volumen
1	491
2	829
3	549
4	710
5	119

[0131] Lo anterior produjo un total de 3119 células (por ul) x1000 (convertir a ml) x 4 (4 ml de volumen total cada uno) = 12,5 millones de células. Las células se dividieron en 200,000 cada una (320 ul) sumergidas en tubos.

[0132] Para los ejemplos, las células se dividieron en 200,000 tubos biorreplicados con 2 por montaje abajo. Después de los lavados y la cuantificación: se sedimentaron las células, se eliminó W5, luego se utilizó 2xMMG (100 ul + 100 ul del ADN indicado a continuación). A continuación, se procedió como se describe en el resto del protocolo (el centrifugado final después de la adición de PEG se siguió realizando con 100 g durante 5 minutos y no con 200 g). Después de la transformación y los lavados, se transfiere en 1 ml de solución W5 a una placa de 24 pocillos que se mantiene en la oscuridad. Las muestras son las siguientes:

- a. Dos biorreplicados de cada uno (por ejemplo para la muestra 1: 1A y 1B con 200K células para cada uno; y
- b. Cada Muestra #1-6 tiene cada plásmido a una concentración final de 300 ng/ul.

[0133] La Tabla 11 a continuación proporciona vectores plasmídicos adicionales proporcionados en los diferentes bloques y para las diferentes muestras. Cada muestra se enumera como 1,1 para el bloque 1 muestra 1 y A y B indican las biorreplicaciones individuales.

Tabla 11

Block	Muestra	Diana	Inteína	Exteín a	Plásmidos
1	1	GmALS_T04	Ssp DnaE	No nativo	Vector plasmídico 154 – Ssp DnaE GmALS_T04-L1
					Vector plasmídico 155 - Ssp DnaE GmALS_T04-R1
					Vector plasmídico 127 - Ssp DnaE intC no nativo
	2	GmALS_T04	Ssp DnaE	Nativo	Vector plasmídico 154 - Ssp DnaE GmALS_T04-L1
					Vector plasmídico 155- Ssp DnaE GmALS_T04-R1
					Vector plasmídico 132 - Ssp DnaE intC nativo
	3	GmALS_T04	Gp41-1	No nativo	Vector plasmídico 149 - gp41-1 intN no nativo GmALS_T04-L1
					Vector plasmídico 150 - gp41-1 intN no nativo GmALS_T04-R1
					Vector plasmídico 131 - gp41-1 intC no nativo en pCLS30416
	4	GmALS_T04	Gp41-1	Nativo	Vector plasmídico 147 - gp41-1 intN nativo GmALS_T04-L1
					Vector plasmídico 148 - gp41-1 intN

					nativo GmALS_T04-R1
					Vector plasmídico 130 - gp41-1 intC nativo en pCLS30416
	5	GmALS_T04	Estándar de TALEN		Vector plasmídico 122 - GmALS_T04-L1
					Vector plasmídico 123 - GmALS_T04-R1
	6		Control negativo		Control negativo de plásmido
	7	AvYFP_T02	Control negativo		Vector plasmídico 128 - AvYFP_T02-L1
					Vector plasmídico 129 - AvYFP_T02-R1
2	1	GmALS_T07	Ssp DnaE	No nativo	Vector plasmídico 156 - Ssp DnaE GmALS_T07-L1
					Vector plasmídico 157 - Ssp DnaE GmALS_T07-R1
					Vector plasmídico 127 - Ssp DnaE intC no nativo
	2	GmALS	Ssp DnaE	Nativ	Vector plasmídico 156- Ssp DnaE

		_T07		o	GmALS_T07-L1
					Vector plasmídico 157- Ssp DnaE GmALS_T07-R1
					Vector plasmídico 132- Ssp DnaE intC nativo
	3	GmALS _T07	Gp41-1	No nativo	Vector plasmídico 138- GmALS_T07-L1 gp41-1 no nativo
					Vector plasmídico 146 - GmALS_T07-R1 gp41-1 no nativo
					Vector plasmídico 131- gp41-1 intC no nativo en pCLS30416
	4	GmALS _T07	Gp41-1	Nativ o	Vector plasmídico 135 - GmALS_T07-L1 gp41-1 nativo
					Vector plasmídico 142- GmALS_T07-R1 gp41-1 nativo
					Vector plasmídico 130 - gp41-1 intC nativo en pCLS30416
	5	GmALS _T07	Estándar de TALEN		Vector plasmídico 124 - GmALS_T07-L1
					Vector plasmídico 125 - GmALS_T07-R1

	6		Control negativo		Control negativo de plásmido
3*					
4	1	GmGRF _T03	Ssp DnaE	No nativo	Vector plasmídico 153 - Ssp DnaE GmGRF_T04-L1
					Vector plasmídico 151 - Ssp DnaE GmGRF_T04-R1
					Vector plasmídico 127 - Ssp DnaE intC no nativo
	2	GmGRF _T03	Ssp DnaE	Nativ o	Vector plasmídico 153- Ssp DnaE GmGRF_T04-L1
					Vector plasmídico 151 - Ssp DnaE GmGRF_T04-R1
					Vector plasmídico 132 - Ssp DnaE intC nativo
	3	GmGRF _T03	Gp41-1	No nativo	Vector plasmídico 136 - GmGRF3_T04-L1 gp41-1 no nativo
					Vector plasmídico 143- GmGRF3_T04-R1 gp41-1 no nativo
					Vector plasmídico 131- gp41-1 intC no nativo en pCLS30416
	4	GmGRF	Gp41-1	Nativ	Vector plasmídico 92 -

		_T03		o	GmGRF3_T03-L1 gp41-1 nativo
					Vector plasmídico 140 - GmGRF3_T03-R1 gp41-1 nativo
					Vector plasmídico 130 - gp41-1 intC nativo en pCLS30416
	5	GmGRF _T03	Estándar de TALEN		Vector plasmídico 120 - GmGRF_T03-L1
					Vector plasmídico 119 - GmGRF_T03-R1
	6		Control negativo		Control negativo
5	1	GmGRF _T04	Ssp DnaE	No nativo	Vector plasmídico 158 - Ssp DnaE GmFAD3_T08-L1
					Vector plasmídico 159 - Ssp DnaE GmFAD3_T08-R1
					Vector plasmídico 127- Ssp DnaE intC no nativo
	2	GmGRF _T04	Ssp DnaE	Nativ o	Vector plasmídico 158 - Ssp DnaE GmFAD3_T08-L1
					Vector plasmídico 159 - Ssp DnaE

					GmFAD3_T08-R1
					Vector plasmídico 132 - Ssp DnaE intC nativo
	3	GmGRF _T04	Gp41-1	No nativo	Vector plasmídico 102- GmGRF3_T03-L1 gp41-1 no nativo
					Vector plasmídico 144 - GmGRF3_T03-R1 gp41-1 no nativo
					Vector plasmídico 131- gp41-1 intC no nativo en pCLS30416
	4	GmGRF _T04	Gp41-1	Nativ o	Vector plasmídico 133 - GmGRF3_T04-L1 gp41-1 nativo
					Vector plasmídico 139 - GmGRF3_T04-R1 gp41-1 nativo
					Vector plasmídico 130 - gp41-1 intC nativo en pCLS30416
	5	GmGRF _T04	Estándar de TALEN		Vector plasmídico 121 - GmGRF_T04-L1
					Vector plasmídico 118 - GmGRF_T04-R1
	6		Control negativo		Control negativo de plásmido

En lo anterior, el bloque 3 de la Tabla 11 corresponde a la Tabla 9.

[0134] La FIG. 7 ilustra ejemplos de resultados de transformación de explantes de soja utilizando los diferentes vectores plasmídicos, conforme a la presente descripción. Más particularmente, se transformaron protoplastos de soja usando vectores coadministrados que codifica un TALE izquierdo fusionado a una primera inteína, un TALE derecho fusionado a la primera inteína, y una nucleasa fusionada a la segunda inteína. Las TALE incluyeron dominios de unión asociados a genes de ALS, FAD3 y GRF.

[0135] Los diferentes vectores plasmídicos incluyeron un primer conjunto que se dirige al gen ALS (por ejemplo, que incluye ALS-T04 y ALS-T07 y vectores plasmídicos 154, 155, 149, 150, 147, 148, 122, 123, 156, 157, 138, 146, 135, 142, 124, y 125), un segundo conjunto que se dirige al gen FAD3 (por ejemplo, que incluye FAD3-T08 y vectores plasmídicos 110, 152, 137, 145, 134, 141, y 126), y un tercer conjunto que se dirige al gen GRF (por ejemplo, que incluye GRF-T03 y GRF-T04 y vectores plasmídicos 153, 151, 136, 143, 92, 140, 120, 119, 158, 159, 102, 144, 133 139, 121, y 123). Dentro de cada uno del primer, segundo y tercer conjunto, los vectores respectivos no incluyeron inteínas (grupos de vector plasmídico de (vector plasmídico 122 y el vector plasmídico 123), (vector plasmídico 124 y el vector plasmídico 125), (vector plasmídico control y el vector plasmídico 126), (vector plasmídico 120 y el vector plasmídico 119), y (vector plasmídico 121 y el vector plasmídico 123), inteínas de SSP DnaE y exteínas no nativas (grupos de vector plasmídico de (vector plasmídico 154, vector plasmídico 155 y el vector plasmídico 127), (vector plasmídico 156, vector plasmídico 157, vector plasmídico 127), (vector plasmídico 110, vector plasmídico 152, vector plasmídico 127), y (vector plasmídico 153, vector plasmídico 151, vector plasmídico 127), inteínas de SSP DnaE y exteínas nativas (grupos de vectores plasmídicos de (vector plasmídico 154, vector plasmídico 155, y el vector plasmídico 132), (vector plasmídico 156, vector plasmídico 157, y el vector plasmídico 132), (vector plasmídico 158, vector plasmídico 159, y el vector plasmídico 132), (vector plasmídico 110, vector plasmídico 152, vector plasmídico 132), y (vector plasmídico 153, vector plasmídico 151, vector plasmídico 132), inteínas de GP41-1 y exteínas no nativas (grupos de vector plasmídico de (vector plasmídico 149, vector plasmídico 150, y el vector plasmídico 131), (vector plasmídico 138, vector plasmídico 146, y el vector plasmídico 131), (vector plasmídico 137, vector plasmídico 145, y el vector plasmídico 131), (vector plasmídico 102, vector plasmídico 144, y el vector plasmídico 131), y (vector plasmídico 136, vector plasmídico 143, y el vector

plasmídico 131), y inteínas de GP41-1 y exteínas nativas (grupos de vector plasmídico de (vector plasmídico 147, vector plasmídico 148, y el vector plasmídico 130), (vector plasmídico 135, vector plasmídico 142, y el vector plasmídico 130), (vector plasmídico 134, vector plasmídico 141, y el vector plasmídico 130), (vector plasmídico 92, vector plasmídico 140, y el vector plasmídico 130), y (vector plasmídico 144, vector plasmídico 139, y el vector plasmídico 130). Los ejemplos de vectores plasmídicos de control negativo incluyen el vector plasmídico 128 (expuesto en la SEQ ID NO: 128) y el vector plasmídico 129 (expuesto en la SEQ ID NO: 129).

[0136] Como se describió anteriormente, se pueden usar tres vectores plasmídicos en cada experimento para demostrar en conjunto actividad del gen de TALEN, con dos de tres vectores plasmídicos que incluye un efector TAL que dirige un gen (por ejemplo, efectores de medio TAL izquierdo y derecho) fusionado a una inteína (por ejemplo, int-N o int-C) y el tercer vector plasmídico que incluye una inteína (por ejemplo, int-C o int-N) fusionada a la FokI endonucleasa. Diferentes vectores pueden incluir un efector TAL que dirige el gen GRF3, FAD3, o ALS fusionado a una secuencia peptídica Ssp DnaE int-N o gp41-1 int-N. El vector plasmídico 92 (SEQ ID NO: 92) codifica un promotor NosPro, un efector TAL que dirige el gen GRF3, una secuencia peptídica Gp41-1 int-N nativa, un terminador NosTerm, y un esqueleto HT izquierdo. El vector plasmídico 102 (SEQ ID NO: 102) codifica un promotor NosPro, un efector TAL que dirige el gen GRF3, una secuencia peptídica Gp41-1 int-N no nativa, un terminador NosTerm, y un esqueleto HT izquierdo. El vector plasmídico 110 (SEQ ID NO: 110) codifica un promotor NosPro, un efector TAL que dirige el gen GRF3, y una secuencia peptídica Ssp DnaE int-N, un terminador NosTerm, y un esqueleto HT izquierdo.

[0137] Una secuencia del vector plasmídico 92 se expone en la SEQ ID NO: 92, que codifica un casete TALE medio izquierdo (SEQ ID NO: 93) que incluye un promotor Nos (SEQ ID NO: 94), una fusión del efector TAL inteína (SEQ ID NO: 95), y un terminador Nos (SEQ ID No: 33). La fusión del efector TAL inteína (SEQ ID NO: 95) codifica un N-terminal de TALE izquierdo (SEQ ID NO: 96), GmGRF3 (T03-L1) dominio de unión (SEQ ID NO: 97), C-terminal de TALE izquierdo (SEQ ID NO: 98), un ligador (SEQ ID NO: 99), un sitio de empalme nativo int-N (SEQ ID NO: 100), y secuencia peptídica Gp41-1 int-N (SEQ ID NO: 101).

[0138] Una secuencia del vector plasmídico 102 se expone en la SEQ ID NO: 102, que codifica un casete TALE medio izquierdo (SEQ ID NO: 103) que incluye un promotor

Nos (SEQ ID NO: 94), una fusión del efector TAL inteína (SEQ ID NO: 104), y un terminador Nos (SEQ ID No: 33). La fusión del efector TAL inteína (SEQ ID NO: 104) codifica un N-terminal de TALE izquierdo (SEQ ID NO: 105), GmGRF3 (T03-L1) dominio de unión (SEQ ID NO: 106), C-terminal de TALE izquierdo (SEQ ID NO: 107), un ligador (SEQ ID NO: 108), y secuencia peptídica Gp41-1 int-N (SEQ ID NO: 109).

[0139] Una secuencia del vector plasmídico 110 se expone en la SEQ ID NO: 110, que codifica un casete TALE medio izquierdo (SEQ ID NO: 111) que incluye un promotor Nos (SEQ ID NO: 94), una fusión del efector TAL inteína (SEQ ID NO: 112), y un terminador Nos (SEQ ID No: 33). La fusión del efector TAL inteína (SEQ ID NO: 112) codifica un N-terminal de TALE izquierdo (SEQ ID NO: 113), GmGRF3 (T03-L1) dominio de unión (SEQ ID NO: 114), C-terminal de TALE izquierdo (SEQ ID NO: 115), un ligador (SEQ ID NO: 116), y secuencia peptídica Ssp DnaE int-N (SEQ ID NO: 117).

[0140] Los ejemplos de constructos restantes se describen a continuación. El vector plasmídico 118 (expuesto en la SEQ ID NO: 118) es un vector TALEN de control que codifica una mitad derecha de las TALEN dirigida a GmGRF (T04) y el vector plasmídico 119 (expuesto en la SEQ ID NO: 119) es un vector TALEN de control que codifica una mitad derecha de las TALEN dirigida a GmGRF (T03). El vector plasmídico 120 (expuesto en la SEQ ID NO: 120) es un vector TALEN de control que codifica una mitad izquierda de las TALEN dirigida a GmGRF (T03) y el vector plasmídico 121 (expuesto en la SEQ ID NO: 121) es un vector TALEN de control que codifica una mitad izquierda de las TALEN dirigida a GmGRF (T04). El vector plasmídico 122 (expuesto en la SEQ ID NO: 122) es un vector TALEN de control que codifica una mitad izquierda de las TALEN dirigida a GmALS (T04) y el vector plasmídico 123 (expuesto en la SEQ ID NO: 123) es un vector TALEN de control que codifica una mitad derecha de las TALEN dirigida a GmALS (T04). El vector plasmídico 124 (expuesto en la SEQ ID NO: 124) es un vector TALEN de control que codifica una mitad izquierda de las TALEN dirigida a GmALS (T07) y el vector plasmídico 125 (expuesto en la SEQ ID NO: 125) es un vector TALEN de control que codifica una mitad derecha de las TALEN dirigida a GmALS (T07). El vector plasmídico 126 (expuesto en la SEQ ID NO: 126) es un vector TALEN de control que codifica una mitad derecha de las TALEN dirigida a GmFAD3 (T08).

[0141] El vector plasmídico 127 (expuesto en la SEQ ID NO: 127) codifica una secuencia peptídica Ssp DnaE int-C, con una exteína no nativa, y una FokI endonucleasa. Como se describió previamente, los vectores plasmídicos 128 y 129 (expuestos en las SEQ ID

NOs: 128-129) son controles de TALEN con YFP, con el vector plasmídico 128 que incluye una mitad izquierda de las TALEN y el vector plasmídico 129 que incluye una mitad derecha de las TALEN vector. El vector plasmídico 130 (expuesto en la SEQ ID NO: 130) codifica una secuencia peptídica gp41-1 int-C, con una exteína nativa, y una FokI endonucleasa. El vector plasmídico 131 (expuesto en la SEQ ID NO: 131) codifica una secuencia peptídica gp41-1 int-C, con una exteína no nativa, y una FokI endonucleasa. El vector plasmídico 132 (expuesto en la SEQ ID NO: 132) es una secuencia peptídica Ssp DnaE int-C, con una exteína nativa, y una FokI endonucleasa.

[0142] El vector plasmídico 133 (expuesto en la SEQ ID NO: 133) codifica un efector TAL medio izquierdo que dirige el gen GRF3 (T04) fusionado a una secuencia peptídica Gp41-1 int-N nativa. El vector plasmídico 134 (expuesto en la SEQ ID NO: 134) codifica un efector TAL medio izquierdo que dirige el gen FAD3 (T08) fusionado a una secuencia peptídica Gp41-1 int-N nativa. El vector plasmídico 135 (expuesto en la SEQ ID NO: 135) codifica un efector TAL medio izquierdo que dirige el gen ALS (T07) fusionado a una secuencia peptídica Gp41-1 int-N nativa.

[0143] El vector plasmídico 136 (expuesto en la SEQ ID NO: 136) codifica un efector TAL medio izquierdo que dirige el gen GRF3 (T04) fusionado a una secuencia peptídica Gp41-1 int-N no nativa. El vector plasmídico 137 (expuesto en la SEQ ID NO: 137) codifica un efector TAL medio izquierdo que dirige el gen FAD3 (T08) fusionado a una secuencia peptídica Gp41-1 int-N no nativa. El vector plasmídico 138 (expuesto en la SEQ ID NO: 138) codifica un efector TAL medio izquierdo que dirige el gen ALS (T07) fusionado a una secuencia peptídica Gp41-1 int-N no nativa.

[0144] El vector plasmídico 139 (expuesto en la SEQ ID NO: 139) codifica un efector TAL medio derecho que dirige el gen GRF3 (T04) fusionado a una secuencia peptídica Gp41-1 int-N nativa. El vector plasmídico 140 (expuesto en la SEQ ID NO: 140) codifica un efector TAL medio derecho que dirige el gen GRF3 (T03) fusionado a una secuencia peptídica Gp41-1 int-N nativa. El vector plasmídico 141 (expuesto en la SEQ ID NO: 141) codifica un efector TAL medio derecho que dirige el gen FAD3 (T08) fusionado a una secuencia peptídica Gp41-1 int-N nativa. El vector plasmídico 142 (expuesto en la SEQ ID NO: 142) codifica un efector TAL medio izquierdo que dirige el gen ALS (T07) fusionado a una secuencia peptídica Gp41-1 int-N nativa.

[0145] El vector plasmídico 143 (expuesto en la SEQ ID NO: 143) codifica un efector TAL medio derecho que dirige el gen GRF3 (T04) fusionado a una secuencia peptídica

Gp41-1 int-N no nativa. El vector plasmídico 144 (expuesto en la SEQ ID NO: 144) codifica un efector TAL medio derecho que dirige el gen GRF3 (T03) fusionado a una secuencia peptídica Gp41-1 int-N no nativa. El vector plasmídico 145 (expuesto en la SEQ ID NO: 145) codifica un efector TAL medio derecho que dirige el gen FAD3 (T08) fusionado a una secuencia peptídica Gp41-1 int-N no nativa. El vector plasmídico 146 (expuesto en la SEQ ID NO: 146) codifica un efector TAL medio derecho que dirige el gen ALS (T07) fusionado a una secuencia peptídica Gp41-1 int-N no nativa.

[0146] El vector plasmídico 147 (expuesto en la SEQ ID NO: 147) codifica un efector TAL medio izquierdo que dirige el gen ALS (T04) fusionado a una secuencia peptídica Gp41-1 int-N nativa. El vector plasmídico 148 (expuesto en la SEQ ID NO: 148) codifica un efector TAL medio derecho que dirige el gen ALS (T04) fusionado a una secuencia peptídica Gp41-1 int-N nativa.

[0147] El vector plasmídico 149 (expuesto en la SEQ ID NO: 149) codifica un efector TAL medio izquierdo que dirige el gen ALS (T04) fusionado a una secuencia peptídica Gp41-1 int-N, no nativa. El vector plasmídico 150 (expuesto en la SEQ ID NO: 150) codifica un efector TAL medio derecho que dirige el gen ALS (T04) fusionado a una secuencia peptídica Gp41-1 int-N, no nativa.

[0148] El vector plasmídico 151 (expuesto en la SEQ ID NO: 151) codifica un efector TAL medio derecho que dirige el gen GRF3 (T04) fusionado a una secuencia peptídica Ssp DnaE int-N. El vector plasmídico 152 (expuesto en la SEQ ID NO: 152) codifica un efector TAL medio derecho que dirige el gen GRF3 (T03) fusionado a una secuencia peptídica Ssp DnaE int-N. El vector plasmídico 153 (expuesto en la SEQ ID NO: 153) codifica un efector TAL medio izquierdo que dirige el gen GRF3 (T04) fusionado a una secuencia peptídica Ssp DnaE int-N.

[0149] El vector plasmídico 154 (expuesto en la SEQ ID NO: 154) codifica un efector TAL medio izquierdo que dirige el gen ALS (T04) fusionado a una secuencia peptídica Ssp DnaE int-N. El vector plasmídico 155 (expuesto en la SEQ ID NO: 155) codifica un efector TAL medio derecho que dirige el gen ALS (T04) fusionado a una secuencia peptídica Ssp DnaE int-N. El vector plasmídico 156 (expuesto en la SEQ ID NO: 156) codifica un efector TAL medio izquierdo que dirige el gen ALS (T07) fusionado a una secuencia peptídica Ssp DnaE int-N. El vector plasmídico 157 (expuesto en la SEQ ID NO: 157) codifica un efector TAL medio derecho que dirige el gen ALS (T07) fusionado a una secuencia peptídica Ssp DnaE int-N

[0150] Vector plasmídico 158 (expuesto en la SEQ ID NO: 158) codifica un efector TAL medio izquierdo que dirige el gen FAD3 (T08) fusionado a una secuencia peptídica Ssp DnaE int-N. El vector plasmídico 159 (expuesto en la SEQ ID NO: 159) codifica un efector TAL medio derecho que dirige el gen FAD3 (T08) fusionado a una secuencia peptídica Ssp DnaE int-N.

[0151] Por ejemplo, la FIG. 7 es un gráfico que compara las eficiencias de edición de diferentes vectores plasmídicos de 92-159 que ilustran las eficiencias de edición de no inteína, inteínas de SSP DnaE y exteínas no nativas, inteínas de SSP DnaE y exteínas nativas, inteínas de GP41-1 y exteínas no nativas, e inteínas de GP41-1 y exteínas nativas.

[0152] Se realizaron varios experimentos utilizando vectores plasmídicos adicionales para transformar células de plantas de cáñamo utilizando los TALEs asociados con diferentes genes, tales como un transgén de fitoeno desaturasa (PDS) y un transgén de ácido tetrahidrocannabinólico sintasa (THCAS).

[0153] Las células de la planta de cannabis se transformaron mediante una transformación mediada por agrobacterias de tejidos del eje embrionario (EA) del cannabis. Brevemente, las semillas de cannabis se esterilizaron mediante un lavado con peróxido de hidrógeno. Después de la esterilización, las semillas se embebieron durante la noche en una solución antibiótica líquida. Después de la imbibición durante la noche, los embriones de cannabis se separaron de la cubierta de la semilla, y los tejidos del EA se recolectaron y se eliminaron los cotiledones y las hojas primarias.

[0154] A continuación, los tejidos de EA de cannabis se colocaron en una placa de Petri que contiene una solución de infección líquida que consistió en medio más agrobacterium portador del vector binario de interés a una DO de 0,2. La placa de Petri de infección que contiene los EA en la solución de agrobacterium se selló y se sometió a sonicación durante 40 segundos. La placa petri de infección que contenía las EA en la solución de agrobacterias se selló y se sometió a sonicación durante 40 segundos. Después de la sonicación, los EA se mantuvieron en el medio de infección durante una hora. Después de una hora, se eliminaron los EA del medio de infección y se sembraron en nuevas placas de Petri de co-cultivo que contienen un papel de filtro húmedo. Las placas se sellaron y se colocaron en una incubadora a 16/8 h de luz y 23 °C durante cuatro días. Después del co-cultivo, los EA se sembraron en placas de Petri que contienen un medio de regeneración. Las placas de regeneración que contienen los EA se sellaron y se colocaron en una

incubadora a 16/8 h de luz y 23 °C durante 7 días. Las EA se pueden transformar utilizando cualquier técnica conocida en el campo.

[0155] Después de 7 días en el medio de regeneración, las EAs se retiraron del medio y se congelaron a -80C para la extracción de ADN utilizando cualquier técnica bien conocida. Por ejemplo, la extracción de ADN se puede implementar como se describe en la Publicación US 2021/0277411, publicada el 9 de septiembre de 2021, y titulada "Canola with High Oleic Acid", que se incorpora en la presente en su totalidad por su enseñanza.

[0156] En varios experimentos, las células de plantas de cannabis se transformaron con vectores plasmídicos como se expone en las SEQ ID NOs: 20-91. Los EA de cannabis se transformaron utilizando el protocolo de transformación de Cannabaceae descrito anteriormente y utilizando bacterias que contienen un vector binario respectivo. En varias realizaciones, se compararon las eficiencias de edición entre los diferentes vectores plasmídicos. Los diferentes vectores plasmídicos incluyeron un primer conjunto que se dirige al gen PDS (por ejemplo, vectores plasmídicos 91, 20, 46, y 87) y un segundo conjunto que se dirige al gen THCAS (por ejemplo, vectores plasmídicos 90, 88, 89, y 65). Dentro de cada uno del primer conjunto y el segundo conjunto, los respectivos vectores incluyeron no inteínas (por ejemplo, vectores plasmídicos 90 y 91), inteínas de SSP DnaE y exteínas nativas (por ejemplo, vectores plasmídicos 20, y 87), inteínas de GP41-1 y exteínas no nativas (por ejemplo, vectores plasmídicos 46 y 88), e inteínas de GP41-1 y exteínas nativas (por ejemplo, vectores plasmídicos 89 y 65).

[0157] Una secuencia del vector plasmídico 20 se expone en la SEQ ID NO: 20, que codifica un casete YFP (SEQ ID NO: 21), un casete TALE medio izquierdo (SEQ ID NO: 25), un casete TALE medio derecho (SEQ ID NO: 34), y un casete Fok1 (SEQ ID NO: 41). El casete YFP codifica un promotor FMV (SEQ ID NO: 22), una proteína de YFP (SEQ ID NO: 23), y un terminador Rbcs (SEQ ID NO: 24). El casete TALE medio izquierdo (SEQ ID NO: 25) codifica un promotor VaUbi3 (SEQ ID NO: 26), una fusión del efector TAL inteína (SEQ ID NO: 27), y un terminador Nos (SEQ ID NO: 33). La fusión del efector TAL inteína (SEQ ID NO: 27) codifica un N-terminal de TALE izquierdo (SEQ ID NO: 28), CsPDS (T02-L1) dominio de unión (SEQ ID NO: 29), C-terminal de TALE izquierdo (SEQ ID NO: 30), un ligador (SEQ ID NO: 31), y secuencia peptídica Ssp DnaE int-N (SEQ ID NO: 32). El casete TALE medio derecho (SEQ ID NO: 34) codifica un promotor VaUbi3 (SEQ ID NO: 26), una fusión del efector TAL inteína (SEQ ID NO: 35), y un terminador Nos (SEQ ID NO: 33). La fusión del efector

TAL inteína (SEQ ID NO: 35) codifica un N-terminal de TALE derecho (SEQ ID NO: 36), CsPDS (T02-R1) dominio de unión (SEQ ID NO: 37), C-terminal TALE derecho (SEQ ID NO: 38), un ligador (SEQ ID NO: 39), y secuencia peptídica Ssp DnaE int-N (SEQ ID NO: 40). El casete FokI (SEQ ID NO: 41) codifica un promotor de MtEF1a (SEQ ID NO: 42), una fusión de inteína y FokI (SEQ ID NO: 43), y un terminador Nos (SEQ ID NO: 33). La fusión de inteína y FokI (SEQ ID NO: 43) codifica una señal de localización nuclear (SEQ ID NO: 44), una secuencia peptídica Ssp DnaE int-C (SEQ ID NO: 45), un CFN (TGCTTCAAC), un ligador (AGCCGTTCC), y FokI (SEQ ID NO: 19).

[0158] Una secuencia del vector plasmídico 46 se expone en la SEQ ID NO: 46, que codifica un casete YFP (SEQ ID NO: 21), un casete TALE medio izquierdo (SEQ ID NO: 47), un casete TALE medio derecho (SEQ ID NO: 54), y un casete Fok1 (SEQ ID NO: 61). El casete YFP codifica un promotor FMV (SEQ ID NO: 22), una proteína de YFP (SEQ ID NO: 23), y un terminador Rbcs (SEQ ID NO: 24). El casete TALE medio izquierdo (SEQ ID NO: 47) codifica un promotor VaUbi3 (SEQ ID NO: 26), una fusión del efector TAL inteína (SEQ ID NO: 48), y un terminador Nos (SEQ ID NO: 33). La fusión del efector TAL inteína (SEQ ID NO: 48) codifica un N-terminal de TALE izquierdo (SEQ ID NO: 49), CsPDS (T02-L1) dominio de unión (SEQ ID NO: 50), C-terminal de TALE izquierdo (SEQ ID NO: 51), un ligador (SEQ ID NO: 52), y secuencia peptídica Gp41-1 int-N (SEQ ID NO: 53). El casete TALE medio derecho (SEQ ID NO: 54) codifica un promotor VaUbi3 (SEQ ID NO: 26), una fusión del efector TAL inteína (SEQ ID NO: 55), y un terminador Nos (SEQ ID NO: 33). La fusión del efector TAL inteína (SEQ ID NO: 55) codifica un N-terminal de TALE derecho (SEQ ID NO: 56), CsPDS (T02-R1) dominio de unión (SEQ ID NO: 57), C-terminal TALE derecho (SEQ ID NO: 58), un ligador (SEQ ID NO: 59), y secuencia peptídica Gp41-1 int-N (SEQ ID NO: 60). El casete FokI (SEQ ID NO: 61) codifica un promotor de MtEF1a (SEQ ID NO: 42), una fusión de inteína y FokI (SEQ ID NO: 62), y un terminador Nos (SEQ ID NO: 33). La fusión de inteína y FokI (SEQ ID NO: 62) codifica una señal de localización nuclear (SEQ ID NO: 63), una secuencia peptídica gp41-1 int-C (SEQ ID NO: 64), un ligador (AGCCGTTCC), y FokI (SEQ ID NO: 19).

[0159] Una secuencia del vector plasmídico 65 se expone en la SEQ ID NO: 65, que codifica un casete YFP (SEQ ID NO: 21), un casete TALE medio izquierdo (SEQ ID NO: 66), un casete TALE medio derecho (SEQ ID NO: 74), y un casete Fok1 (SEQ ID

NO: 82). El casete YFP codifica un promotor FMV (SEQ ID NO: 22), una proteína de YFP (SEQ ID NO: 23), y un terminador Rbcs (SEQ ID NO: 24). El casete TALE medio izquierdo (SEQ ID NO: 66) codifica un promotor VaUbi3 (SEQ ID NO: 26), una fusión del efector TAL inteína (SEQ ID NO: 67), y un terminador Nos (SEQ ID NO: 33). La fusión del efector TAL inteína (SEQ ID NO: 67) codifica un N-terminal de TALE izquierdo (SEQ ID NO: 68), CsTHCAS (T22-L1) dominio de unión (SEQ ID NO: 69), C-terminal de TALE izquierdo (SEQ ID NO: 70), un ligador (SEQ ID NO: 71), un sitio de empalme nativo int-N (SEQ ID NO: 72), y secuencia peptídica Gp41-1 int-N (SEQ ID NO: 73). El casete TALE medio derecho (SEQ ID NO: 74) codifica un promotor VaUbi3 (SEQ ID NO: 26), una fusión del efector TAL inteína (SEQ ID NO: 75), y un terminador Nos (SEQ ID NO: 33). La fusión del efector TAL inteína (SEQ ID NO: 75) codifica un N-terminal de TALE derecho (SEQ ID NO: 76), CsTHCAS (T22-R1) dominio de unión (SEQ ID NO: 77), C-terminal TALE derecho (SEQ ID NO: 78), un ligador (SEQ ID NO: 79), un sitio de empalme nativo int-N (SEQ ID NO: 80), y secuencia peptídica Gp41-1 int-N (SEQ ID NO: 81). El casete FokI (SEQ ID NO: 82) codifica un promotor de a MtEF1a (SEQ ID NO: 42), una fusión de inteína y FokI (SEQ ID NO: 83), y un terminador Nos (SEQ ID NO: 33). La fusión de inteína y FokI (SEQ ID NO: 83) codifica una señal de localización nuclear (SEQ ID NO: 84), una secuencia peptídica Gp41-1 int-C (SEQ ID NO: 85), un sitio de empalme nativo int-C (SEQ ID NO: 86), un ligador (AGCCGTTCC), y FokI (SEQ ID NO: 19).

[0160] Los ejemplos de constructos restantes se describen a continuación. El vector plasmídico 87 (expuesto en la SEQ ID NO: 87) codifica los efectores TALE izquierdo y derecho que se dirigen a THCAS, junto con un casete Fok1, cada uno de los cuales incluye inteínas Ssp DnaE, nativas. El vector plasmídico 88 (expuesto en la SEQ ID NO: 88) codifica los efectores TALE izquierdo y derecho que se dirigen a THCAS, junto con un casete Fok1 cada uno de los cuales incluye las inteínas Gp41-1, no nativas. El vector plasmídico 89 (expuesto en la SEQ ID NO: 89) codifica los efectores TALE izquierdo y derecho que se dirigen a THCAS, junto con un casete Fok1 cada uno de los cuales incluye las inteínas Gp41-1, nativas. El vector plasmídico 90 (expuesto en la SEQ ID NO: 90) es un vector TALEN de control que codifica una mitad de las TALEN izquierda y derecha dirigida al gen THCAS. El vector plasmídico 91 (expuesto en la SEQ ID NO: 91) es un vector TALEN de control que codifica las mitad izquierda de las TALEN y derecha dirigidas al gen PDS.

[0161] Las FIG. 8A-8I ilustran los resultados de la transformación de explantes de cáñamo utilizando los diferentes vectores plasmídicos, conforme a la presente descripción. La FIG. 8A es un gráfico que compara las eficiencias de edición de diferentes vectores plasmídicos del primer conjunto y del segundo conjunto, que ilustra las eficiencias de edición de no inteínas (por ejemplo, vectores plasmídicos 90 y 91), inteínas de SSP DnaE y exteínas nativas (por ejemplo, vectores plasmídicos 20 y 87), inteínas de GP41-1 y exteínas no nativas (por ejemplo, vectores plasmídicos 46 y 88), y inteínas de GP41-1 y exteínas nativas (por ejemplo, vectores plasmídicos 89 y 65).

[0162] Las FIG. 8B-8C son gráficos que comparan el porcentaje de eventos de edición de los diferentes vectores plasmídicos del primer conjunto y del segundo conjunto, como se ilustra en la FIG. 8A. Las FIG. 8D-8I son imágenes de explantes de experimentos que evalúan la transformación de los explantes con los diferentes vectores plasmídicos de del vector plasmídico 20 (FIG. 8D), vector plasmídico 46 (FIG. 8E), vector plasmídico 89 (FIG. 8F), vector plasmídico 87 (FIG. 8G), vector plasmídico 88 (FIG. 8H) y el vector plasmídico 65 (FIG. 8I), conforme a la presente descripción.