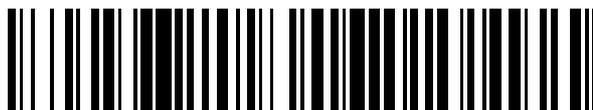


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 666 179**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/64 (2006.01)

C12N 15/66 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.06.2015 PCT/US2015/037199**

87 Fecha y número de publicación internacional: **30.12.2015 WO15200334**

96 Fecha de presentación y número de la solicitud europea: **23.06.2015 E 15735807 (8)**

97 Fecha y número de publicación de la concesión europea: **21.03.2018 EP 3155099**

54 Título: **Montaje de ADN mediado por nucleasas**

30 Prioridad:

23.06.2014 US 201462015809 P

24.06.2014 US 201462016400 P

13.08.2014 US 201462036983 P

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

03.05.2018

73 Titular/es:

REGENERON PHARMACEUTICALS, INC.
(100.0%)

777 Old Saw Mill River Road
Tarrytown, NY 10591, US

72 Inventor/es:

SCHOENHERR, CHRIS;
MCWHIRTER, JOHN;
MOMONT, COREY;
MACDONALD, LYNN;
MURPHY, ANDREW J.;
WARSHAW, GREGG S.;
ROJAS, JOSE F.;
LAI, KA-MAN, VENUS;
VALENZUELA, DAVID M. y
MONTAGNA, CAITLIN

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 666 179 T3

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

DESCRIPCIÓN

Montaje de ADN mediado por nucleasas

Como archivo de texto a través de la red EFS

5 La copia oficial del listado de secuencias se envía electrónicamente a través de la red EFS como un listado de secuencias con formato ASCII con un archivo llamado 461002SEQLIST.TXT, creado el 23 de junio de 2015, y que tiene un tamaño de 66 KB, y se presenta al mismo tiempo con la memoria descriptiva. El listado de secuencias contenido en este documento con formato ASCII es parte de la memoria descriptiva y se incorpora aquí como referencia en su totalidad.

Antecedentes

10 Históricamente, la extensión de solapamiento podría usarse como un medio para sintetizar moléculas de ADN de cadena doble más grandes, particularmente genes, a partir del solapamiento de oligonucleótidos sintéticos. Sin embargo, estos métodos no podrían combinar eficazmente grandes moléculas de ADN de manera rápida. Además, la combinación específica del sitio de ácidos nucleicos grandes que usan secuencias solapantes a menudo está limitada por la disponibilidad de secuencias solapantes en la posición deseada en los ácidos nucleicos a combinar. Las enzimas
15 de nucleasa modificadas diseñadas para secuencias de ADN específicas objetivo han atraído la atención como herramientas poderosas para la manipulación genética que permite la eliminación, el reemplazo y la reparación de genes objetivo, así como la inserción de secuencias exógenas. Sin embargo, las tecnologías existentes adolecen de una precisión limitada, lo que puede conducir a efectos imprevistos fuera del objetivo y a reacciones de varias etapas que consumen tiempo.

20 El documento US 2010/035768 A1 (Gibson Daniel G, 2010-02-11) discute métodos para el ensamblaje *in vitro* combinatorio y de unión de moléculas de ácido nucleico.

Sumario

Se proporcionan en la presente memoria métodos para ensamblar ácidos nucleicos que tienen secuencias solapantes. Tales métodos comprenden un método para ensamblar al menos dos ácidos nucleicos, que comprende: (a) poner en
25 contacto un primer ácido nucleico con un primer agente de nucleasa, donde el primer agente de nucleasa comprende una proteína Cas y un ARN guía (ARNg) (complejo de ARNg-Cas), una nucleasa de dedo de zinc, o una nucleasa efectora similar a un activador de transcripción (TALEN), donde el primer agente de nucleasa escinde el primer ácido nucleico en un primer sitio objetivo para producir un primer ácido nucleico digerido con una secuencia final solapante compartida por un segundo ácido nucleico; (b) poner en contacto el primer ácido nucleico digerido y el segundo ácido
30 nucleico con una exonucleasa para exponer secuencias complementarias entre el primer ácido nucleico digerido y el segundo ácido nucleico; y (c) ensamblar los dos fragmentos de ácido nucleico generados a partir de la etapa (b). En algunos de dichos métodos, la etapa (c) comprende, además: (i) hibridar las secuencias complementarias expuestas; (ii) extender los extremos 3' de las secuencias complementarias hibridadas; y (iii) ligar el primer y el segundo ácidos nucleicos.

35 En algunos de los métodos, la etapa (a) comprende además poner en contacto el segundo ácido nucleico con un segundo agente de nucleasa, donde el segundo agente de nucleasa escinde el segundo ácido nucleico en un segundo sitio objetivo para producir un segundo ácido nucleico digerido con la secuencia final solapante, y en donde el segundo ácido nucleico de la etapa (b) es el segundo ácido nucleico digerido. En algunos de los métodos, al menos dos ácidos nucleicos son de cadena doble. En algunos de los métodos, la secuencia final solapante varía de 20 pb a 200 pb de
40 longitud.

En algunos de los métodos, al menos uno del primer o segundo agente de nucleasa comprende una proteína Cas y un ARN guía (ARNg) (complejo de ARNg-Cas) que se dirige al primer o al segundo sitio objetivo. Por ejemplo, la proteína Cas puede ser una proteína Cas9. La proteína Cas9 puede comprender un dominio RuvC y un dominio HNH, al menos uno de los cuales carece de actividad endonucleasa. En algunas realizaciones, el ARNg comprende una secuencia de
45 ácido nucleico que codifica un ARN de repeticiones palindrómicas cortas agrupadas y regularmente intercaladas (CRISPR) (ARNcr) y un ARN CRISPR que se activa en forma trans (ARNcrtra). El primer sitio objetivo y/o el segundo sitio objetivo pueden estar flanqueados por una secuencia del motivo adyacente protoespaciador (PAM). En algunos de los métodos, el agente de nucleasa comprende una nucleasa de dedo de zinc o una nucleasa efectora similar al activador de la transcripción (TALEN).

50 En algunos de los métodos, el primer ácido nucleico, el segundo ácido nucleico o ambos ácidos nucleicos se derivan de un cromosoma artificial bacteriano. El cromosoma artificial bacteriano puede comprender un ADN humano, un ADN de roedor, un ADN sintético o una combinación de los mismos. El cromosoma artificial bacteriano puede comprender una secuencia humana.

55 Los métodos divulgados aquí incluyen un método para ensamblar dos o más ácidos nucleicos, que comprende: (a) poner en contacto un primer ácido nucleico con al menos un agente de nucleasa, donde al menos un agente de nucleasa comprende una proteína Cas y un ARN guía (ARNg) (complejo de ARNg-Cas), una nucleasa de dedo de zinc

o una nucleasa efectora similar a un activador de transcripción (TALEN), donde al menos un agente de nucleasa escinde el primer ácido nucleico en un primer sitio objetivo para generar un primer ácido nucleico digerido; (b) poner en contacto el primer ácido nucleico digerido con un segundo ácido nucleico, un oligo de unión y una exonucleasa, en el que el oligo de unión comprende: (i) una primera secuencia complementaria que es complementaria al primer ácido nucleico digerido; (ii) un espaciador; y (iii) una segunda secuencia complementaria que es complementaria al segundo ácido nucleico; en el que la exonucleasa expone la primera y la segunda secuencias complementarias; y (c) ensamblar el oligo de unión con el primer ácido nucleico digerido y el segundo ácido nucleico.

También se divulga un método para ensamblar al menos dos ácidos nucleicos, que comprende: (a) poner en contacto un primer ácido nucleico con un primer agente de nucleasa y un segundo agente de nucleasa para producir un primer ácido nucleico digerido, en el que el primer agente de nucleasa genera una muesca en una primera cadena del primer ácido nucleico en un primer sitio objetivo, y el segundo agente de nucleasa genera una muesca en una segunda cadena del primer ácido nucleico en un segundo sitio objetivo, para producir un primer ácido nucleico digerido que comprende una secuencia sobresaliente 5' o 3' en uno de sus extremos; (b) hibridar el primer ácido nucleico digerido y un segundo ácido nucleico que comprende una secuencia complementaria a la secuencia sobresaliente 5' o 3'; y (c) ligar el primer ácido nucleico digerido y el segundo ácido nucleico. En algunos de los métodos, la etapa (b) comprende adicionalmente extender el extremo 3' de la primera cadena usando la segunda cadena como una plantilla y extender el extremo 3' de la segunda cadena basándose en la primera cadena como plantilla. En algunos de los métodos, el primer sitio objetivo está separado por al menos 4 pb del segundo sitio objetivo.

En algunos de los métodos, al menos uno del primer o segundo agente de nucleasa comprende una proteína Cas9 y un ARN guía (ARNg) (complejo de ARNg-Cas) que elige como blanco al primer o al segundo sitio objetivo. El ARNg puede comprender una secuencia de ácido nucleico que codifica un ARN de repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) (ARNcr) y un ARN CRISPR que se activa en forma trans (ARNcrtra). En algunos de los métodos, al menos uno del primer sitio objetivo y el segundo sitio objetivo está flanqueado por una secuencia del motivo adyacente protoespaciador (PAM). La proteína Cas9 puede comprender un dominio RuvC y un dominio HNH, uno de los cuales carece de actividad endonucleasa.

En algunos de los métodos, el segundo ácido nucleico no comprende la secuencia complementaria a la secuencia sobresaliente 5' o 3' del primer ácido nucleico digerido, y la etapa (a) comprende además poner en contacto el primer ácido nucleico digerido y el segundo ácido nucleico digerido con un oligo de unión, en el que el oligo de unión comprende: (i) una primera secuencia complementaria a la secuencia sobresaliente 5' o 3' del primer ácido nucleico digerido; y (ii) una segunda secuencia complementaria a la secuencia sobresaliente 5' o 3' del segundo ácido nucleico digerido. En algunos métodos, el primero, el segundo o ambos ácidos nucleicos se derivan de un cromosoma artificial bacteriano. El cromosoma artificial bacteriano puede comprender un ADN humano, un ADN de roedor, un ADN sintético o una combinación de los mismos. El cromosoma artificial bacteriano puede comprender una secuencia de polinucleótidos humanos. En algunos métodos, el segundo ácido nucleico comprende un cromosoma artificial bacteriano.

Los métodos proporcionados también incluyen un método para ensamblar dos o más fragmentos de ácido nucleico, que comprende: (a) poner en contacto un primer ácido nucleico con al menos un agente de nucleasa para generar un primer ácido nucleico digerido; (b) poner en contacto el primer ácido nucleico digerido con un segundo ácido nucleico, un oligo de unión y una exonucleasa, en el que el oligo de unión comprende: (i) una primera secuencia complementaria que es complementaria al primer ácido nucleico digerido; (ii) un espaciador; y (iii) una segunda secuencia complementaria que es complementaria al segundo ácido nucleico; en el que la exonucleasa expone la primera y la segunda secuencias complementarias; y (c) ensamblar el oligo de unión con el primer ácido nucleico digerido y el segundo ácido nucleico. En algunos de tales métodos, el ensamblaje en la etapa (c) comprende: (i) hibridar la primera secuencia complementaria del oligo de unión con el primer ácido nucleico digerido y la segunda secuencia complementaria del oligo de unión con el segundo ácido nucleico; y (ii) ligar el oligo de unión al primer ácido nucleico digerido y al segundo ácido nucleico.

En algunos métodos, la primera secuencia complementaria y la segunda secuencia complementaria del oligo de unión comprenden entre 15 y 120 bases complementarias. En algunos métodos, el espaciador del oligo de unión comprende ácidos nucleicos no complementarios. En algunas realizaciones, el primer ácido nucleico digerido se ensambla en forma continua con el segundo ácido nucleico.

En algunos métodos, al menos un agente de nucleasa está diseñado para escindir un fragmento de al menos 20 pb del extremo del primer ácido nucleico en el que se producirá el ensamblaje sin problemas, en el que el espaciador del oligo de unión comprende una secuencia idéntica al fragmento de al menos 20 pb, en el que no están presentes bases de ácido nucleico entre la primera secuencia complementaria y el fragmento de al menos 20 pb, y no hay bases de ácido nucleico entre la segunda secuencia complementaria y el fragmento de al menos 20 pb, de manera que el ensamblaje del primer ácido nucleico digerido con el oligo de unión y el segundo ácido nucleico reconstituye el fragmento de al menos 20 pb y ensambla en forma continua el primer ácido nucleico y el segundo ácido nucleico. Opcionalmente, el fragmento de al menos 20 pb es de cadena doble. En algunos métodos, se realiza el mismo método con un fragmento de al menos 20 pb del segundo ácido nucleico como secuencia espaciadora. En algunos métodos, el oligo de unión es de aproximadamente 50 pb a aproximadamente 400 pb, opcionalmente en el que el oligo de unión es de aproximadamente 100 pb a aproximadamente 300 pb. En algunos métodos, el espaciador comprende de aproximadamente 20 pb a aproximadamente 120 pb. En algunos métodos, el segundo ácido nucleico se pone en

contacto con un segundo agente de nucleasa, donde el segundo agente de nucleasa escinde el segundo ácido nucleico para producir un segundo ácido nucleico digerido que comprende una secuencia de nucleótidos que es complementaria a la segunda secuencia complementaria del oligo de unión, donde el primer ácido nucleico digerido se ensambla al segundo ácido nucleico digerido. En algunos métodos, el segundo ácido nucleico se pone en contacto con una enzima de restricción o meganucleasa, en el que la enzima de restricción o meganucleasa escinde el segundo ácido nucleico para producir un segundo ácido nucleico digerido que comprende una secuencia de nucleótidos que es complementaria a la segunda secuencia complementaria en el oligo de unión, en donde el primer ácido nucleico digerido se ensambla al segundo ácido nucleico digerido. En algunos métodos, el extremo 3' del primer y/o el segundo ácidos nucleicos digeridos se extiende en la etapa (b). El oligo de unión se puede ensamblar a dicho primer ácido nucleico y dicho segundo ácido nucleico en la misma reacción o secuencialmente. En algunos métodos, el primer, el segundo, o ambos ácidos nucleicos se derivan de un cromosoma artificial bacteriano, al menos 10 kb, y/o comprenden un ADN humano, ADN de roedor, un ADN sintético o una combinación de los mismos.

En algunos de los métodos, al menos un agente de nucleasa o un segundo agente de nucleasa comprende una proteína Cas y un ARN guía (ARNg) (complejo de ARNg-Cas) que elige como blanco al primer o al segundo sitio objetivo. Por ejemplo, la proteína Cas puede ser una proteína Cas9. La proteína Cas9 puede comprender un dominio RuvC y un dominio HNH, al menos uno de los cuales carece de actividad endonucleasa. En algunas realizaciones, el ARNg comprende una secuencia de ácido nucleico que codifica un ARN de repeticiones palindrómicas cortas agrupadas y regularmente intercaladas (CRISPR) (ARNcr) y un ARN CRISPR que se activa en forma trans (ARNcrtra). El primer sitio objetivo y/o el segundo sitio objetivo pueden estar flanqueados por una secuencia del motivo adyacente protoespaciador (PAM). En algunos de los métodos, al menos un agente de nucleasa y/o el segundo agente de nucleasa comprende una nucleasa de dedo de zinc o una nucleasa efectora similar al activador de transcripción (TALEN).

En algunas realizaciones, el oligo de unión comprende un Bloque g. En algunos de tales métodos, el Bloque g no comprende un casete de selección.

Se proporcionan además métodos para ensamblar dos o más ácidos nucleicos, que comprenden: (a) poner en contacto un primer ácido nucleico con al menos un agente de nucleasa para generar un primer ácido nucleico digerido; (b) poner en contacto un segundo ácido nucleico con un segundo agente de nucleasa para generar un segundo ácido nucleico digerido; (c) poner en contacto el primer ácido nucleico digerido y el segundo ácido nucleico digerido con un oligo de unión y una exonucleasa, donde el oligo de unión comprende: (i) una primera secuencia complementaria que es complementaria al primer ácido nucleico digerido; (ii) un espaciador; y (iii) una segunda secuencia complementaria que es complementaria al segundo ácido nucleico digerido; en el que la exonucleasa expone la primera y la segunda secuencias complementarias; y (d) ensamblar el oligo de unión con el primer ácido nucleico digerido y el segundo ácido nucleico.

Se proporcionan en el presente documento métodos para ensamblar ácidos nucleicos que tienen secuencias solapantes. Dichos métodos comprenden un método para ensamblar al menos dos fragmentos de ácido nucleico, que comprende (a) poner en contacto un primer y un segundo ácidos nucleicos que comprenden secuencias solapantes con al menos un complejo de ARNg-Cas y una exonucleasa, generando de este modo dos fragmentos de ácido nucleico digeridos que comprenden secuencias complementarias en uno de sus extremos; (b) ensamblar los dos fragmentos de ácido nucleico generados a partir de la etapa (a). En algunos métodos, al menos un complejo de ARNg-Cas escinde el primer ácido nucleico en un primer sitio objetivo para producir un primer ácido nucleico digerido que comprende secuencias finales complementarias entre el primer ácido nucleico digerido y el segundo ácido nucleico. En ciertos métodos, la etapa (b) comprende, además: (i) hibridar las secuencias complementarias expuestas; (ii) extender los extremos 3' de las secuencias complementarias hibridadas; y (iii) ligar el primer y el segundo ácido nucleico. En algunos métodos, la etapa (a) comprende además poner en contacto el segundo ácido nucleico con un segundo complejo de ARNg-Cas, donde el segundo ácido nucleico no comprende la secuencia final solapante, y el segundo complejo de ARNg-Cas escinde el segundo ácido nucleico para producir un segundo ácido nucleico digerido que comprende las secuencias finales solapantes entre el primer ácido nucleico digerido y el segundo ácido nucleico digerido. Por ejemplo, el complejo de ARNg-Cas comprende una proteína Cas9. La proteína Cas9 puede comprender un dominio RuvC y un dominio HNH, al menos uno de los cuales carece de actividad endonucleasa. En algunos métodos, la secuencia de superposición varía de 20 pb a 200 pb de longitud. El método de una cualquiera de las reivindicaciones 1-7, en el que el primero, el segundo, o ambos ácidos nucleicos son de un cromosoma artificial bacteriano. En algunos métodos, el cromosoma artificial bacteriano comprende un ADN humano, un ADN de roedor, un ADN sintético o una combinación de los mismos. El cromosoma artificial bacteriano puede comprender una secuencia humana.

Los métodos proporcionados también incluyen un método para ensamblar dos o más fragmentos de ácido nucleico, que comprende: (a) exponer un primer y un segundo ácido nucleico a al menos un complejo de ARNg-Cas para generar un primer y un segundo ácidos nucleicos digeridos que comprenden una secuencia sobresaliente 5' o 3' en uno de sus extremos; (b) ensamblar los dos fragmentos de ácido nucleico generados a partir de la etapa (a). En algunos métodos, la etapa de ensamblaje (b) comprende: (i) hibridar las secuencias sobresalientes 5' y 3'; y (ii) ligar el primer ácido nucleico digerido y el segundo ácido nucleico digerido. En algunos métodos, las secuencias sobresalientes 5' y/o 3' comprenden al menos 4 bases complementarias. En algunos métodos, la etapa (b) comprende además extender el extremo 3' del primer y el segundo ácidos nucleicos digeridos. En algunos métodos, el segundo ácido nucleico no comprende una secuencia complementaria a la secuencia sobresaliente 5' o 3' del primer ácido nucleico digerido, y la

etapa (a) comprende además poner en contacto el primer ácido nucleico digerido y el segundo ácido nucleico digerido con un oligo de unión, en el que el oligo de unión comprende: (i) una primera secuencia complementaria a la secuencia sobresaliente 5' o 3' del primer ácido nucleico digerido; y (ii) una segunda secuencia complementaria a la secuencia sobresaliente 5' o 3' del segundo ácido nucleico digerido. En algunos métodos, el complejo de proteína ARNg-Cas comprende una proteína Cas9 que comprende un dominio RuvC y un dominio HNH, uno de los cuales carece de actividad endonucleasa. En algunos métodos, el complejo de ARNg-Cas se proporciona por separado como ARNcr, ARNcrtra y proteína Cas. En algunos métodos, el primer y el segundo ácidos nucleicos comprenden una secuencia del motivo adyacente protoespaciador (PAM). En algunos métodos, el primero, el segundo o ambos ácidos nucleicos se derivan de un cromosoma artificial bacteriano. En algunos métodos, el cromosoma artificial bacteriano comprende un ADN humano, un ADN de roedor, un ADN sintético o una combinación de los mismos. Por ejemplo, el cromosoma artificial bacteriano puede comprender una secuencia de polinucleótido humano.

Se proporcionan además métodos para ensamblar dos o más ácidos nucleicos, que comprenden: (a) poner en contacto un primer ácido nucleico con al menos un complejo de ARNg-Cas para generar un primer ácido nucleico digerido; y (b) poner en contacto el primer ácido nucleico digerido con un segundo ácido nucleico, un oligo de unión y una exonucleasa, en el que el oligo de unión comprende: (i) una primera secuencia complementaria que es complementaria al primer ácido nucleico digerido; (ii) un espaciador; y (iii) una segunda secuencia complementaria que es complementaria al segundo ácido nucleico; en el que la exonucleasa expone la primera y la segunda secuencias complementarias; y (c) ensamblar el oligo de unión con el primer ácido nucleico digerido y el segundo ácido nucleico. En algunos métodos, la etapa de ensamblaje (c) comprende (i) hibridar la primera secuencia complementaria del oligo de unión con el primer ácido nucleico digerido y la segunda secuencia complementaria del oligo de unión con el segundo ácido nucleico; y (ii) ligar el oligo de unión al primer ácido nucleico digerido y el segundo ácido nucleico. En algunos métodos, la primera secuencia complementaria y la segunda secuencia complementaria del oligo de unión comprenden entre 15 y 120 bases complementarias. En algunos métodos, el espaciador del oligo de unión comprende ácidos nucleicos no complementarios.

Usando el oligo de unión, el primer ácido nucleico digerido se puede ensamblar sin problemas con el segundo ácido nucleico. En algunos métodos, el complejo de ARNg-Cas está diseñado para escindir un fragmento de al menos 20 pb del extremo del primer ácido nucleico en el que se producirá el ensamblaje perfecto, en el que el espaciador del oligo de unión comprende una secuencia idéntica a dicho fragmento de al menos 20 pb, en el que no hay bases de ácido nucleico entre la primera secuencia complementaria y el fragmento de al menos 20 pb, y no hay bases de ácido nucleico entre la segunda secuencia complementaria y el fragmento de al menos 20 pb, de modo que el ensamblaje de dicho primer ácido nucleico con dicho oligo de unión y dicho segundo ácido nucleico reconstituye el fragmento de al menos 20 pb y ensambla sin problemas el primer y el segundo ácido nucleico. En algunos métodos, se realiza el mismo método con un fragmento de al menos 20 pb del segundo ácido nucleico como secuencia espaciadora. En algunos métodos, el espaciador comprende de aproximadamente 20 pb a aproximadamente 120 pb. En algunos métodos, el segundo ácido nucleico se pone en contacto con un segundo complejo de ARNg-Cas y una exonucleasa, en donde el segundo complejo de ARNg-Cas escinde el segundo ácido nucleico para producir un segundo ácido nucleico digerido que comprende una secuencia de nucleótidos que es complementaria a la segunda secuencia complementaria del oligo de unión, en donde el primer ácido nucleico digerido se ensambla al segundo ácido nucleico digerido. En algunos métodos, el segundo ácido nucleico se pone en contacto con una enzima de restricción o meganucleasa y una exonucleasa, donde la enzima de restricción o meganucleasa escinde el segundo ácido nucleico para producir un segundo ácido nucleico digerido que comprende una secuencia de nucleótidos que es complementaria a la segunda secuencia complementaria en el oligo de unión, en donde el primer ácido nucleico digerido se ensambla al segundo ácido nucleico digerido. En algunos métodos, el extremo 3' del primer y/o el segundo ácidos nucleicos digeridos se extiende en la etapa (b). El oligo de unión se puede ensamblar a dicho primer ácido nucleico y dicho segundo ácido nucleico en la misma reacción o secuencialmente. En algunos métodos, el complejo de ARNg-Cas comprende una proteína Cas9. En algunos métodos, el primer, el segundo, o ambos ácidos nucleicos se derivan de un cromosoma artificial bacteriano, de al menos 10 kb, y/o comprenden un ADN humano, ADN de roedor, un ADN sintético o una combinación de los mismos.

Breve descripción de los dibujos

La Fig. 1 muestra el ensamblaje de un BAC a un producto de PCR que tiene solapamientos diseñados para ser específicos para el BAC. Se agregaron superposiciones de 50 pb al casete HYG mediante PCR.

La Fig. 2 muestra el ensamblaje de dos BAC que tienen secuencias superpuestas usando dos sitios objetivo Cas9 en cada BAC. El proceso de ensamblaje usando el método divulgado aquí tomó 2 días.

La Fig. 3 muestra el ensamblaje de dos BAC con secuencias superpuestas usando métodos tradicionales. El proceso de ensamblaje usando métodos tradicionales tomó 4 semanas.

La Fig. 4 muestra las eficiencias de clonación del método de ensamblaje de Cas9/isotérmico y el tiempo requerido para las etapas de clonación de BAC.

La Fig. 5 muestra la construcción de un vector de direccionamiento grande (LTVEC) usando el sistema CRISPR/Cas9 y el ensamblaje isotérmico. Los fragmentos de ADN escindidos con CRISPR/Cas9 se ensamblaron en forma continua usando uno o más oligos de unión y ensamblaje isotérmico.

- La Fig. 6 muestra la estrategia para usar enlazadores (oligos de unión) para el ensamblaje sin problemas de ácidos nucleicos después de la escisión de Cas9. Un complejo de ARNg/Cas9 está diseñado para escindir un sitio objetivo localizado 5' secuencia arriba de un área de interés (flecha) para generar un primer fragmento de ADN digerido con Cas9 (ADN 5'). La porción eliminada del ADN 5' (caja cortada) se usa luego como un espaciador entre las secuencias superpuestas 5' y 3' en un oligo de unión. Se ensamblan tres componentes en la reacción de ensamblaje isotérmico: (a) un primer fragmento de ADN digerido con Cas9 (ADN 5'); (b) un oligo de unión; y (c) un segundo fragmento de ADN (ADN 3'). El oligo de unión comprende desde 5' hasta 3': (1) una secuencia de solapamiento con ADN 5', (2) un espaciador que contiene la porción suprimida del primer fragmento digerido y (3) una secuencia de solapamiento con ADN 3'. La porción eliminada del ADN 5' se reconstituye durante la etapa de ensamblaje.
- La Fig. 7 muestra la construcción de un vector de ADN usando el sistema CRISPR/Cas9 y el ensamblaje isotérmico.
- La Fig. 8 muestra la construcción de un vector de direccionamiento grande usando el sistema CRISPR/Cas9 y el ensamblaje isotérmico.
- La Fig. 9 muestra la construcción de un vector de direccionamiento para el reemplazo de una porción de un vector de BAC con un casete usando un ensamblaje isotérmico y dos enlazadores (oligos de unión). Los resultados de varias relaciones de mBAC con respecto a fragmentos o enlazadores se presentan en los paneles # 1, # 2, # 3 y # 4.
- La Fig. 10 muestra la confirmación de la secuencia de ensamblaje sin problemas a través de ambas uniones de la reacción de ensamblaje entre un mBAC (BAC ID: RP23-399M19) y un casete que usa dos enlazadores.
- La Fig. 11 muestra el ensamblaje de dos mBAC usando Cas9 y ensamblaje isotérmico. El ensamblaje entre el vector bMQ50f19 y el casete que comprende un promotor de ubiquitina del gen de resistencia a higromicina fue sin problemas.
- La Fig. 12 muestra la confirmación de secuencia del ensamblaje sin problemas en el enlazador 1, y la confirmación de secuencia del ensamblaje que fue intencionalmente sin problemas en el enlazador 2 y el enlazador 3.
- La Fig. 13 muestra la inserción de grandes fragmentos de genes humanos en un mBAC utilizando cuatro enlazadores y un ensamblaje isotérmico. Cas9 escindió el fragmento A de hGen de hBAC1, el fragmento B de hGen de hBAC2 y mBAC para eliminar los fragmentos de mGen.
- La Fig. 14 muestra la inserción de la secuencia humana en un vector BAC usando Cas9 y ensamblaje isotérmico.
- La Fig. 15 muestra la inserción de un Bloque g que comprende un sitio de meganucleasa usando Cas9 y ensamblaje isotérmico. La Fig. 15A muestra la inserción de un Bloque g que comprende un sitio PI-SceI; y la Fig. 15B muestra la inserción de un Bloque g que comprende un sitio MauBI.
- La Fig. 16 ilustra un ejemplo de humanización directa de un vector de direccionamiento usando tres oligos de unión, Cas9 y ensamblaje isotérmico.
- La Fig. 17 ilustra un ejemplo de humanización indirecta de un vector de direccionamiento usando un donante con oligos de unión ascendente y descendente, Cas9 y ensamblaje isotérmico.
- La Fig. 18 ilustra un ejemplo de introducción de una mutación puntual usando Cas9 y ensamblaje isotérmico.
- La Fig. 19 ilustra un ejemplo de recorte de BAC por Cas9 y ensamblaje isotérmico. En este ejemplo, el recorte elimina la secuencia de Ori. La secuencia de Ori se reinserta en el vector usando dos oligos de unión y ensamblaje isotérmico.

Descripción detallada

I. Definiciones

Los términos "proteína", "polipéptido" y "péptido", usados indistintamente en la presente memoria, incluyen formas poliméricas de aminoácidos de cualquier longitud, que incluyen aminoácidos codificados y no codificados y aminoácidos modificados química o bioquímicamente o que forman derivados. Los términos también incluyen polímeros que se han modificado, tales como polipéptidos que tienen cadenas principales de péptidos modificados.

Los términos "ácido nucleico" y "polinucleótido", usados indistintamente en la presente memoria, incluyen formas poliméricas de nucleótidos de cualquier longitud, que incluyen ribonucleótidos, desoxirribonucleótidos o análogos o versiones modificadas de los mismos. Incluyen ADN o ARN de cadena sencilla, doble y cadenas múltiples, ADN genómico, ADNc, híbridos de ADN-ARN y polímeros que comprenden bases de purina, bases de pirimidina u otras bases nucleotídicas naturales, modificadas químicamente, modificadas bioquímicamente, no naturales o que forman derivados.

La "optimización de codones" generalmente incluye un proceso de modificación de una secuencia de ácido nucleico para expresión mejorada en células huésped particulares mediante la sustitución de al menos un codón de la secuencia nativa por un codón que se usa con más frecuencia o más frecuentemente en los genes de la célula huésped, aunque se mantiene la secuencia de aminoácidos nativa. Por ejemplo, un ácido nucleico que codifica una proteína Cas puede

modificarse para sustituir codones que tienen una mayor frecuencia de uso en una célula procarionta o eucariótica dada, incluyendo una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster, o cualquier otra célula huésped, en comparación con la secuencia de ácido nucleico natural. Las tablas de uso de codones están disponibles, por ejemplo, en la "Codon Usage Database". Estas tablas se pueden adaptar de varias maneras. Véase Nakamura et al. (2000) *Nucleic Acids Research* 28: 292. Los algoritmos informáticos para la optimización de codones de una secuencia particular para la expresión en un huésped particular también están disponibles (véase, por ejemplo, Gene Forge).

"Enlace operable" o estar "operativamente enlazado" incluye yuxtaposición de dos o más componentes (por ejemplo, un promotor y otro elemento de secuencia) de modo que ambos componentes funcionan normalmente y permitan la posibilidad de que al menos uno de los componentes pueda mediar una función que se ejerce sobre al menos uno de los otros componentes. Por ejemplo, un promotor puede enlazarse operativamente a una secuencia codificante si el promotor controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o ausencia de uno o más factores reguladores transcripcionales.

"Complementariedad" de ácidos nucleicos significa que una secuencia de nucleótidos en una cadena de ácido nucleico, debido a la orientación de sus grupos nucleobases, forma enlaces de hidrógeno con otra secuencia en una cadena de ácido nucleico opuesta. Las bases complementarias en el ADN son típicamente A con T y C con G. En el ARN, son típicamente C con G y U con A. La complementariedad puede ser perfecta o sustancial/suficiente. La complementariedad perfecta entre dos ácidos nucleicos significa que los dos ácidos nucleicos pueden formar un dúplex en el que cada base en el dúplex se une a una base complementaria mediante emparejamiento Watson-Crick. Complementariedad "sustancial" o "suficiente" significa que una secuencia en una cadena no es completa y/o perfectamente complementaria a una secuencia en una cadena opuesta, pero que se produce una unión suficiente entre bases en las dos cadenas para formar un complejo híbrido estable en conjunto con las condiciones de hibridación (p. ej., concentración salina y temperatura). Tales condiciones pueden predecirse usando las secuencias y cálculos matemáticos estándar para predecir la T_m de cadenas hibridadas, o mediante la determinación empírica de T_m usando métodos rutinarios. La T_m incluye la temperatura a la que se desnaturaliza 50% de una población de complejos de hibridación formados entre dos cadenas de ácido nucleico. A una temperatura por debajo de la T_m , se favorece la formación de un complejo de hibridación, mientras que, a una temperatura superior a la T_m , se favorece la fusión o la separación de las cadenas en el complejo de hibridación. La T_m puede estimarse para un ácido nucleico que tiene un contenido de G + C conocido en una solución acuosa de NaCl 1 M utilizando, por ejemplo, $T_m = 81,5 + 0,41 (\% \text{ de G} + \text{C})$, aunque otros cálculos de T_m conocidos tienen en cuenta características estructurales del ácido nucleico.

La "condición de hibridación" incluye el entorno acumulativo en el que una cadena de ácido nucleico se une a una segunda cadena de ácido nucleico mediante interacciones de cadena complementaria y enlaces de hidrógeno para producir un complejo de hibridación. Tales condiciones incluyen los componentes químicos y sus concentraciones (por ejemplo, sales, agentes quelantes, formamida) de una solución acuosa u orgánica que contiene los ácidos nucleicos, y la temperatura de la mezcla. Otros factores, como la duración del tiempo de incubación o las dimensiones de la cámara de reacción pueden contribuir al entorno. Véase, por ejemplo, Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2a ed., páginas 1.90-1.91, 9.47-9.51, 11.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque son posibles faltas de coincidencia entre las bases. Las condiciones apropiadas para la hibridación entre dos ácidos nucleicos dependen de la longitud de los ácidos nucleicos y del grado de complementación, variables bien conocidas en la técnica. Cuanto mayor es el grado de complementación entre dos secuencias de nucleótidos, mayor es el valor de la temperatura de fusión (T_m) para los híbridos de ácidos nucleicos que tienen esas secuencias. Para hibridaciones entre ácidos nucleicos con tramos cortos de complementariedad (por ejemplo, complementariedad mayor a 35 o menos, 30 o menos, 25 o menos, 22 o menos, 20 o menos, o 18 o menos nucleótidos), la posición de las faltas de coincidencia se vuelve importante (véase Sambrook et al., citado anteriormente, 11.7-11.8). Típicamente, la longitud para un ácido nucleico hibridable es al menos aproximadamente de 10 nucleótidos. Las longitudes mínimas ilustrativas para un ácido nucleico hibridable incluyen al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 22 nucleótidos, al menos aproximadamente 25 nucleótidos y al menos aproximadamente 30 nucleótidos. Además, la temperatura y la concentración salina de la solución de lavado pueden ajustarse según sea necesario de acuerdo con factores tales como la longitud de la región de complementación y el grado de complementación.

La secuencia del polinucleótido no necesita ser 100% complementaria a la de su ácido nucleico objetivo para ser específicamente hibridable. Además, un polinucleótido puede hibridarse sobre uno o más segmentos de manera que los segmentos intermedios o adyacentes no están implicados en el evento de hibridación (por ejemplo, una estructura de bucle o estructura en horquilla). Un polinucleótido (por ejemplo, ARNg) puede comprender al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 99%, o 100% de complementariedad de secuencia con una región objetivo dentro de la secuencia de ácido nucleico objetivo a la cuál ellos apuntan. Por ejemplo, un ARNg en el que 18 de 20 nucleótidos son complementarios a una región objetivo, y por lo tanto hibridarían específicamente, representaría un 90% de complementariedad. En este ejemplo, los nucleótidos no complementarios restantes pueden estar agrupados o intercalados con nucleótidos complementarios y no necesitan ser contiguos entre sí o con nucleótidos complementarios.

El porcentaje de complementariedad entre tramos particulares de secuencias de ácidos nucleicos dentro de ácidos

nucleicos se puede determinar rutinariamente usando programas BLAST (herramientas de búsqueda de alineación local básica) y programas PowerBLAST conocidos en la técnica (AltschL et al. (1990) J. Mol. Biol. 215: 403-410; Zhang y Madden (1997) Genome Res. 7: 649-656) o usando el programa Gap (Wisconsin Sequence Analysis Package, Versión 8 para Unix, Genetics Computer Group, University Research Park, Madison, WI), usando la configuración predeterminada, que usa el algoritmo de Smith y Waterman (Adv. Appl. Math., 1981, 2, 482-489).

Los métodos y composiciones proporcionados aquí emplean una variedad de diferentes componentes. Se reconoce en toda la descripción que algunos componentes pueden tener variantes activas y fragmentos. Dichos componentes incluyen, por ejemplo, proteínas Cas, ARN de CRISPR, ARNcrtra y ARN guía. La actividad biológica para cada uno de estos componentes se describe en otra parte del presente documento.

"Identidad de secuencia" o "identidad" en el contexto de dos secuencias de polinucleótidos o polipéptidos hace referencia a los residuos en las dos secuencias que son iguales cuando se alinean para correspondencia máxima sobre una ventana de comparación especificada. Cuando se usa un porcentaje de identidad de secuencia en referencia a proteínas, se reconoce que las posiciones de los residuos que no son idénticas a menudo difieren por sustituciones conservadoras de aminoácidos, donde los residuos de aminoácidos se sustituyen por otros residuos de aminoácidos con propiedades químicas similares (por ejemplo, carga o hidrofobicidad) y por lo tanto no cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en las sustituciones conservadoras, el porcentaje de identidad de secuencia se puede ajustar hacia arriba para corregir la naturaleza conservadora de la sustitución. Se dice que las secuencias que difieren en tales sustituciones conservadoras tienen "similitud de secuencia" o "similitud". Los medios para realizar este ajuste son bien conocidos por los expertos en la técnica. Típicamente, esto implica calificar una sustitución conservadora como una falta de coincidencia parcial en lugar de completa, aumentando así el porcentaje de identidad de secuencia. Así, por ejemplo, cuando un aminoácido idéntico recibe una puntuación de 1 y una sustitución no conservadora recibe una puntuación de cero, una sustitución conservadora recibe una puntuación entre cero y 1. La puntuación de las sustituciones conservadoras se calcula, por ejemplo, como se implementó en el programa PC/GENE (Intelligenetics, Mountain View, California).

"Porcentaje de identidad de secuencia" incluye el valor determinado comparando dos secuencias óptimamente alineadas sobre una ventana de comparación, donde la porción de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o eliminaciones (es decir, espacios) en comparación con la secuencia de referencia (que no comprende adiciones o eliminaciones) para la alineación óptima de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que la base de ácido nucleico o residuo de aminoácido idéntico se produce en ambas secuencias para obtener el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación y multiplicar el resultado por 100 para obtener el porcentaje de identidad de secuencia.

A menos que se establezca lo contrario, los valores de identidad/similitud de secuencia incluyen el valor obtenido usando la Versión 10 de GAP usando los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos usando ponderación GAP de 50 y ponderación de longitud de 3, y la matriz de puntuación nwsgapdna.cmp; % de identidad y % de similitud para una secuencia de aminoácidos usando ponderación GAP de 8 y ponderación de longitud de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente del mismo. "Programa equivalente" incluye cualquier programa de comparación de secuencias que, para cualquiera de las dos secuencias en cuestión, genere una alineación que tenga coincidencias idénticas de residuos de nucleótidos o aminoácidos y una identidad de secuencia porcentual idéntica cuando se compara con la alineación correspondiente generada por la Versión 10 de GAP.

Las composiciones o métodos que "comprenden" o "incluyen" uno o más elementos mencionados pueden incluir otros elementos que no se mencionan específicamente. Por ejemplo, una composición que "comprende" o "incluye" una proteína puede contener la proteína sola o en combinación con otros ingredientes.

La designación de un intervalo de valores incluye todos los números enteros dentro o que definen el intervalo, y todos los subintervalos definidos por números enteros dentro del intervalo.

A menos que sea evidente a partir del contexto, el término "aproximadamente" abarca valores dentro de un margen de error estándar de medición (por ejemplo, SEM) de un valor establecido.

Las formas singulares de los artículos "un", "uno, una" y "el, la" incluyen referencias en plural a menos que el contexto indique claramente lo contrario. Por ejemplo, el término "una proteína Cas" o "al menos una proteína Cas" puede incluir una pluralidad de proteínas Cas, que incluyen mezclas de las mismas.

II. General

Los métodos tradicionales de ensamblaje de ácidos nucleicos emplean etapas que consumen mucho tiempo de digestión enzimática convencional con enzimas de restricción, clonación de los ácidos nucleicos y ácidos nucleicos ligantes juntos (véanse la Figura 3 y la Figura 4 para una ilustración de los métodos tradicionales y la línea de tiempo). Estos métodos se vuelven más difíciles cuando se ensamblan grandes fragmentos o vectores. Los métodos proporcionados en la presente memoria aprovechan la especificidad objetivo maleable de las nucleasas (por ejemplo, ARN guía y nucleasas Cas9) para convertir los ácidos nucleicos en una forma adecuada para uso en reacciones de

ensamblaje rápido.

Se proporcionan aquí métodos para ensamblar al menos dos ácidos nucleicos usando agentes de nucleasa dirigidos a sitios objetivo específicos, tales como por ARN guía (ARNg) (por ejemplo, proteína Cas dirigida a sitios objetivo específicos por ARN guía (ARNg)). Los agentes de nucleasa dirigidos al sitio, por ejemplo, proteínas Cas dirigidas por ARN guía, permiten una combinación rápida y eficiente de ácidos nucleicos seleccionando y manipulando las secuencias finales generadas por su actividad de endonucleasa. Los métodos proporcionados en la presente memoria combinan un primer polinucleótido con un agente de nucleasa (por ejemplo, un complejo de ARNg-Cas) específico para un sitio objetivo deseado y una exonucleasa. El sitio objetivo se puede elegir de manera que cuando la nucleasa escinde el ácido nucleico, los extremos resultantes creados por la escisión tengan regiones complementarias a los extremos del segundo ácido nucleico (por ejemplo, extremos superpuestos). Estos extremos complementarios se pueden ensamblar para producir un único ácido nucleico ensamblado. Debido a que el agente de nucleasa (por ejemplo, el complejo de ARNg-Cas) es específico para un sitio objetivo individual, el presente método permite la modificación de ácidos nucleicos de una manera precisa dirigida al sitio. El presente método aprovecha adicionalmente el agente de nucleasa, por ejemplo, un complejo de ARNg-Cas, especificidad utilizando métodos de ensamblaje rápidos y eficientes especialmente diseñados para combinar extremos de ácido nucleico solapantes generados por escisión de nucleasa o diseñados y sintetizados para la reacción de ensamblaje. Por ejemplo, seleccionando un agente de nucleasa (p. ej., un complejo de ARNg-Cas) específico para un sitio objetivo de forma que, en la escisión, se producen secuencias finales complementarias a las de un segundo ácido nucleico, se puede usar un ensamblaje isotérmico para ensamblar el ácido nucleico digerido resultante. Por lo tanto, seleccionando ácidos nucleicos y agentes de nucleasa (por ejemplo, complejos de ARNg-Cas) que dan como resultado superposición de secuencias finales, los ácidos nucleicos se pueden ensamblar por métodos combinatorios rápidos para producir el ácido nucleico ensamblado final de una manera rápida y eficiente. Alternativamente, los ácidos nucleicos que no tienen extremos complementarios se pueden ensamblar con oligos de unión diseñados para tener extremos complementarios para cada ácido nucleico. Usando los oligos de unión, dos o más ácidos nucleicos se pueden ensamblar en forma continua, reduciendo de este modo las secuencias innecesarias en el ácido nucleico ensamblado resultante.

III. Agente de nucleasa

Los presentes métodos emplean un agente de nucleasa para la escisión de polinucleótidos dirigida al sitio. Específicamente, la escisión por endonucleasa de polinucleótidos en un sitio objetivo identificado produce un polinucleótido digerido con extremos que luego se pueden unir a un segundo polinucleótido para ensamblar dos o más polinucleótidos de una manera específica del sitio.

"Agente de nucleasa" incluye moléculas que poseen actividad para la escisión del ADN. Los ejemplos particulares de agentes de nucleasa para uso en los métodos divulgados en este documento incluyen el sistema CRISPR-Cas9 guiado por ARN, proteínas de dedo de zinc, meganucleasas, dominios TAL, TALEN, ensamblaje de levadura, recombinasas, cremalleras de leucina, CRISPR/Cas, endonucleasas y otros agentes de nucleasa conocidos por los expertos en la materia. Los agentes de nucleasa pueden seleccionarse o diseñarse para especificidad en la escisión en un sitio objetivo dado. Por ejemplo, los agentes de nucleasa pueden seleccionarse para la escisión en un sitio objetivo que crea extremos solapantes entre el polinucleótido escindido y un polinucleótido diferente. Los agentes de nucleasa que tienen elementos tanto de proteína como de ARN como en CRISPR-Cas9 pueden suministrarse con los agentes ya complejados como un agente de nucleasa, o pueden suministrarse con los elementos de proteína y ARN separados, en cuyo caso se complejan para formar un agente de nucleasa en las mezclas de reacción descritas aquí.

El término "sitio de reconocimiento para un agente de nucleasa" incluye una secuencia de ADN en la que un agente de nucleasa induce una muesca o rompimiento de la doble cadena. El sitio de reconocimiento para un agente de nucleasa puede ser endógeno (o nativo) a la célula o el sitio de reconocimiento puede ser exógeno a la célula. En realizaciones específicas, el sitio de reconocimiento es exógeno a la célula y por lo tanto no se produce naturalmente en el genoma de la célula. En aún otras realizaciones, el sitio de reconocimiento es exógeno a la célula y a los polinucleótidos de interés que se desea colocar en el locus objetivo. En realizaciones adicionales, el sitio de reconocimiento exógeno o endógeno está presente solo una vez en el genoma de la célula huésped. En realizaciones específicas, se identifica un sitio endógeno o nativo que aparece solo una vez dentro del genoma. Tal sitio puede usarse luego para diseñar agentes de nucleasa que producirán una muesca o rompimiento de la doble cadena en el sitio de reconocimiento endógeno.

La longitud del sitio de reconocimiento puede variar, e incluye, por ejemplo, sitios de reconocimiento que son de aproximadamente 30-36 pb para un par de nucleasa de dedo de zinc (ZFN) (es decir, aproximadamente 15-18 pb para cada ZFN), aproximadamente 36 pb para una nucleasa efectora similar al activador de transcripción (TALEN), o aproximadamente 20 pb para un ARN guía de CRISPR/Cas9.

Se proporcionan también variantes activas y fragmentos de los ejemplos de sitios de reconocimiento. Tales variantes activas pueden comprender al menos 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de identidad de secuencia para el sitio de reconocimiento dado, donde las variantes activas retienen actividad biológica y, por lo tanto, son capaces de ser reconocidas y escindidas por un agente de nucleasa de una manera específica de la secuencia. Los ensayos para medir el rompimiento de cadena doble de un sitio de reconocimiento mediante un agente de nucleasa son conocidos en la técnica (por ejemplo, ensayo qPCR TaqMan®, Frenthewey D. et al., *Methods in Enzymology*, 2010, 476: 295-307).

En realizaciones específicas, el sitio de reconocimiento se coloca dentro del polinucleótido que codifica el marcador de selección. Tal posición puede ubicarse dentro de la región de codificación del marcador de selección o dentro de las regiones reguladoras, lo que influye en la expresión del marcador de selección. Por lo tanto, un sitio de reconocimiento del agente de nucleasa puede localizarse en un intrón del marcador de selección, un promotor, un potenciador, una región reguladora o cualquier región que no codifica proteína del polinucleótido que codifica el marcador de selección. En realizaciones específicas, una muesca o rotura de cadena doble en el sitio de reconocimiento interrumpe la actividad del marcador de selección. Se conocen métodos para ensayar la presencia o ausencia de un marcador de selección funcional.

Puede usarse cualquier agente de nucleasa que induzca una muesca o rompimiento de cadena doble en un sitio de reconocimiento deseado en los métodos y composiciones descritos en este documento. Puede emplearse un agente de nucleasa natural o nativo siempre que el agente de nucleasa induzca una muesca o rompimiento de cadena doble en un sitio de reconocimiento deseado. Alternativamente, se puede emplear un agente de nucleasa modificado o alterado por ingeniería genética. Un "agente de nucleasa modificado por ingeniería genética" comprende una nucleasa que está modificada por ingeniería genética (modificada o derivada) a partir de su forma nativa para reconocer e inducir específicamente una muesca o rompimiento de cadena doble en el sitio de reconocimiento deseado. De este modo, un agente de nucleasa modificado por ingeniería genética puede derivarse de un agente de nucleasa nativo que se produce de manera natural o puede crearse o sintetizarse artificialmente. La modificación del agente de nucleasa puede ser tan pequeña como un aminoácido en un agente de escisión de proteínas o un nucleótido en un agente de escisión de ácidos nucleicos. En algunas realizaciones, la nucleasa modificada induce una muesca o rompimiento de cadena doble en un sitio de reconocimiento, en el que el sitio de reconocimiento no era una secuencia que hubiera sido reconocida por un agente de nucleasa nativo (no alterado o no modificado). La producción de una muesca o rompimiento de cadena doble en un sitio de reconocimiento u otro ADN se puede denominar aquí "corte" o "escisión" del sitio de reconocimiento u otro ADN.

Estos rompimientos pueden ser reparados por la célula en una de dos formas: unión del extremo no homólogo y reparación dirigida por homología (recombinación homóloga). En la unión del extremo no homólogo (NHEJ), los rompimientos de cadena doble se reparan por ligadura directa de los extremos de rompimiento entre sí. Como tal, no se inserta ningún nuevo material de ácido nucleico en el sitio, aunque se puede perder algo de material de ácido nucleico, dando como resultado una eliminación. En la reparación dirigida por homología, puede usarse un polinucleótido donador con homología con la secuencia de ADN objetivo escindida como molde para la reparación de la secuencia de ADN objetivo escindida, dando como resultado la transferencia de información genética desde el polinucleótido donante al ADN objetivo. Por lo tanto, se puede insertar/copiar nuevo material de ácido nucleico en el sitio. Las modificaciones del ADN objetivo debido a NHEJ y/o la reparación dirigida por homología se pueden usar para la corrección de genes, reemplazo de genes, marcación de genes, inserción de transgenes, eliminación de nucleótidos, rompimiento de genes, mutación genética, etc.

En una realización, el agente de nucleasa es una nucleasa efectora similar al activador de la transcripción (TALEN). Las nucleasas efectoras TAL son una clase de nucleasas específicas de secuencia que se pueden usar para producir rompimientos de cadena doble en secuencias objetivo específicas en el genoma de un organismo procarionota o eucariota. Las nucleasas efectoras TAL se crean fusionando un efector similar al activador de transcripción (TAL) nativo o modificado, o parte funcional del mismo, al dominio catalítico de una endonucleasa, tal como, por ejemplo, *FokI*. El único dominio de unión de ADN efector TAL modular, permite el diseño de proteínas con potencialmente cualquier especificidad de reconocimiento de ADN dada. De este modo, los dominios de unión a ADN de las nucleasas efectoras TAL se pueden modificar por ingeniería genética para reconocer sitios objetivo específicos de ADN y, por lo tanto, se usan para hacer rompimientos de cadena doble en las secuencias objetivo deseadas. Véase, WO 2010/079430; Morbitzer et al. (2010) PNAS 10.1073/pnas.1013133107; Scholze y Boch (2010) Virulence 1: 428-432; Christian et al. Genetics (2010) 186: 757-761; Li et al. (2010) Nuc. Acids Res. (2010) doi: 10.1093/nar/gkq704; y Miller et al. (2011) Nature Biotechnology 29: 143-148.

Se describen ejemplos de nucleasas TAL adecuadas, y métodos para preparar nucleasas TAL adecuadas, por ejemplo, en las solicitudes de patente de Estados Unidos Nos. 2011/0239315 A1, 2011/0269234 A1, 2011/0145940 A1, 2003/0232410 A1, 2005/0208489 A1, 2005/0026157 A1, 2005/0064474 A1, 2006/0188987 A1 y 2006/0063231 A1. En diversas realizaciones, las nucleasas efectoras TAL se modifican genéticamente para cortar en o cerca de una secuencia de ácido nucleico objetivo, por ejemplo, en un locus genómico de interés, en el que la secuencia de ácido nucleico objetivo está en o cerca de una secuencia que se va a modificar mediante un vector de direccionamiento. Las nucleasas TAL adecuadas para su uso con los diversos métodos y composiciones proporcionados en la presente memoria incluyen aquellas que están diseñadas específicamente para unirse a o cerca de secuencias de ácido nucleico objetivo que se van a modificar por vectores de direccionamiento como se describe en la presente memoria.

En una realización, cada monómero de TALEN comprende 33-35 repeticiones de TAL que reconocen un único par de bases a través de dos residuos hipervariables. En una realización, el agente de nucleasa es una proteína quimérica que comprende un dominio de unión a ADN basado en la repetición de TAL operativamente unido a una nucleasa independiente. En una realización, la nucleasa independiente es una endonucleasa *FokI*. En una realización, el agente de nucleasa comprende un primer dominio de unión a ADN basado en la repetición de TAL y un segundo dominio de unión a ADN basado en la repetición de TAL, en el que cada uno del primer y segundo dominio de unión a ADN basado en repetición de TAL está operativamente unido a una subunidad de nucleasa plegada, donde el primer y el segundo

dominio de unión a ADN basado en repetición de TAL reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separadas por una secuencia espaciadora de longitud variable (12-20 pb), y en donde las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que produce un rompimiento de cadena doble en una secuencia objetivo.

5 El agente de nucleasa empleado en los diversos métodos y composiciones descritos en la presente memoria puede comprender además una nucleasa de dedo de zinc (ZFN). En una realización, cada monómero de la ZFN comprende 3 o más dominios de unión a ADN basados en dedos de zinc, en donde cada dominio de unión a ADN basado en dedos de zinc se une a un subsitio de 3 pb. En otras realizaciones, la ZFN es una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de zinc operativamente unido a una nucleasa independiente. En una
10 realización, la endonucleasa independiente es una endonucleasa FokI. En una realización, el agente de nucleasa comprende un primer ZFN y un segundo ZFN, en el que cada uno del primer ZFN y el segundo ZFN está operativamente unido a una subunidad de nucleasa Fold, en donde la primera y la segunda ZFN reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separada por un espaciador de aproximadamente 5-7 pb, y en el que las subunidades de nucleasa FokI dimerizan para crear una nucleasa activa que produce un rompimiento de cadena doble. Véase, por ejemplo, los documentos US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; WO/2011/017293A2; y Gaj et al. (2013) Trends in Biotechnology, 31 (7): 397-405.

En una realización de los métodos proporcionados en el presente documento, el agente de nucleasa comprende (a) una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de zinc fusionado a una endonucleasa Fold; o, (b) una proteína quimérica que comprende una nucleasa efectora similar al activador de la transcripción (TALEN) fusionada a una endonucleasa FokI.
20

En aún otra realización, el agente de nucleasa es una meganucleasa. Las meganucleasas se han clasificado en cuatro familias basadas en motivos de secuencia conservados, las familias son las familias de cajas LAGLIDADG (SEQ ID NO: 16), GIY-YIG, H-N-H y His-Cys. Estos motivos participan en la coordinación de iones metálicos y la hidrólisis de enlaces fosfodiéster. Las HEAsas son notables por sus sitios largos de reconocimiento, y por tolerar algunos polimorfismos de secuencia en sus sustratos de ADN. Los dominios, estructura y función de la meganucleasa son conocidos, véase, por ejemplo, Guhan y Muniyappa (2003) Crit Rev Biochem Mol Biol 38: 199-248; Lucas et al., (2001) Nucleic Acids Res 29: 960-9; Jurica y Stoddard, (1999) Cell Mol Life Sci 55: 1304-26; Stoddard, (2006) Q Rev Biophys 38: 49-95; y Moure et al., (2002) Nat Struct Biol 9: 764. En algunos ejemplos, se usa una meganucleasa variante natural y/o una derivada modificada. Se conocen métodos para modificar la cinética, las interacciones del cofactor, la expresión, las condiciones óptimas, y/o la especificidad del sitio de reconocimiento, y el cribado de la actividad, véase, por ejemplo, Epinat et al., (2003) Nucleic Acids Res 31: 2952-62; Chevalier et al., (2002) Mol Cell 10: 895-905; Gimble et al., (2003) Mol Biol 334: 993-1008; Seligman et al., (2002) Nucleic Acids Res 30: 3870-9; Sussman et al., (2004) J Mol Biol 342: 31-41; Rosen et al., (2006) Nucleic Acids Res 34: 4791-800; Chames et al., (2005) Nucleic Acids Res 33: e178; Smith et al., (2006) Nucleic Acids Res 34: e149; Gruen et al., (2002) Nucleic Acids Res 30: e29; Chen y Zhao, (2005) Nucleic Acids Res 33: e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; y WO2004031346.
25
30
35

Se puede usar cualquier meganucleasa en el presente documento, incluyendo, pero sin limitarse a, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeAIIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIIVP, I-TII, I-PpoI, PI-PspI, F-SceI, F-SceII, F-SuVI, F-TevI, F-TevII, I-Amal, I-Anil, I-Chul, I-Cmoel, I-Cpal, I-Cpall, I-Csml, I-CvuI, I-CvuAIP, I-DdII, I-DdIII, I-Dirl, I-Dmol, I-Hmul, I-Hmull, I-HsNIP, I-L1aI, I-Msol, I-Naal, I-NanI, I-NcIIP, I-NgrIP, I-NitI, I-Njal, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-Scal, I-SexIP, I-SnelIP, I-SpomI, I-SpomCP, I-SpomIIP, I-SpomIIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbilP, PI-MtuI, PI-MtuHIIP, PI-Pful, PI-Pfull, PI-Pkol, PI-Pkoll, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-Tful, PI-Tfull, PI-Thyl, PI-TIil, PI-TIiil, o cualquier variante activa o fragmento de las mismas.
40
45

En una realización, la meganucleasa reconoce secuencias de ADN de cadena doble de 12 a 40 pares de bases. En una realización, la meganucleasa reconoce una secuencia objetivo perfectamente coincidente en el genoma. En una realización, la meganucleasa es una nucleasa de inicio. En una realización, la nucleasa codificada por el intrón es una familia LAGLIDADG (SEQ ID NO: 16) de nucleasa codificada por el intrón. En una realización, la familia LAGLIDADG (SEQ ID NO: 16) de nucleasa codificada por el intrón se selecciona de I-SceI, I-CreI e I-Dmol.
50

Los agentes de nucleasa pueden comprender adicionalmente endonucleasas de restricción (enzimas de restricción), que incluyen endonucleasas de Tipo I, Tipo II, Tipo III y Tipo IV. Las endonucleasas de restricción Tipo I y Tipo III reconocen sitios de reconocimiento específicos, pero típicamente escinden en una posición variable del sitio de unión de la nucleasa, que puede estar a cientos de pares de bases del sitio de escisión (sitio de reconocimiento). En los sistemas de Tipo II, la actividad de restricción es independiente de cualquier actividad de metilasa, y la escisión típicamente ocurre en sitios específicos dentro o cerca del sitio de unión. La mayoría de las enzimas Tipo II cortan secuencias palindrómicas, sin embargo, las enzimas Tipo IIa reconocen sitios de reconocimiento no palindrómicos y se escinden fuera del sitio de reconocimiento, las enzimas Tipo IIb cortan secuencias dos veces con ambos sitios fuera del sitio de reconocimiento, y las enzimas Tipo II reconocen un sitio de reconocimiento asimétrico y escinde en un lado y a una distancia definida de aproximadamente 1-20 nucleótidos desde el sitio de reconocimiento. Las enzimas de restricción tipo IV se dirigen al ADN metilado. Las enzimas de restricción se describen y clasifican adicionalmente, por ejemplo, en
55
60

la base de datos REBASE (página web en rebase.neb.com; Roberts et al., (2003) *Nucleic Acids Res* 31: 418-20), Roberts et al., (2003) *Nucleic Acids Res* 31: 1805-12, y Belfort et al., (2002) en *Mobile ADN II*, págs. 761-783, Eds. Craigie y otros, (ASM Press, Washington, DC). En realizaciones específicas, se pueden seleccionar al menos dos enzimas endonucleasas como agentes de nucleasa en donde las enzimas crean extremos adhesivos compatibles o complementarios.

El agente de nucleasa empleado en los diversos métodos y composiciones también puede comprender un sistema CRISPR/Cas. Dichos sistemas pueden emplear una nucleasa Cas9, que, en algunos casos, es de codón optimizado para el tipo de célula deseado en el que se va a expresar. El sistema emplea además una construcción fusionada de ARNcr-ARNcrtra que funciona con la Cas9 de codón optimizado. Este único ARN a menudo se denomina ARN guía o ARNg. Dentro de un ARNg, la porción de ARNcr se identifica como la 'secuencia objetivo' para el sitio de reconocimiento dado y el ARNcrtra a menudo se denomina 'andamio'. Se ha demostrado que este sistema funciona en una variedad de células eucariotas y procariontas. Brevemente, se inserta un fragmento corto de ADN que contiene la secuencia objetivo en un plásmido de expresión de ARN guía. El plásmido de expresión de ARNg comprende la secuencia objetivo (en algunas realizaciones alrededor de 20 nucleótidos), una forma de la secuencia de ARNcrtra (el andamio) así como un promotor adecuado que es activo en la célula y elementos necesarios para el procesamiento apropiado en células eucariotas. Muchos de los sistemas se basan en oligos complementarios personalizados que se hibridan para formar un ADN de cadena doble y luego se clonan en el plásmido de expresión de ARNg. El casete de expresión de ARNg y el casete de expresión de Cas9 se introducen luego en la célula. Véase, por ejemplo, Mali P et al. (2013) *Science* 2013, 15 de febrero; 339 (6121): 823-6; Jinek M et al. *Science* 2012, 17 de agosto; 337 (6096): 816-21; Hwang WY et al. *Nat Biotechnol* 2013, marzo; 31 (3): 227-9; Jiang W et al. *Nat Biotechnol* 2013, marzo; 31 (3): 233-9; y, Cong L et al. *Science* 2013, 15 de febrero; 339 (6121): 819-23.

Los métodos y composiciones divulgados en la presente memoria pueden utilizar sistemas asociados a CRISPR (Cas) de repeticiones palindrómicas cortas agrupadas y regularmente intercaladas (CRISPR) o componentes de tales sistemas para modificar un genoma dentro de una célula. Los sistemas CRISPR/Cas incluyen transcritos y otros elementos implicados en la expresión de, o que dirigen la actividad de genes Cas. Un sistema CRISPR/Cas puede ser un sistema tipo I, un tipo II o un tipo III. Los métodos y composiciones divulgadas en la presente memoria emplean sistemas CRISPR/Cas utilizando complejos CRISPR (que comprenden un ARN guía (ARNg) complejado con una proteína Cas) para la escisión dirigida al sitio de ácidos nucleicos.

Algunos sistemas CRISPR/Cas utilizados en los métodos divulgados en este documento son de origen no natural. Un sistema "no natural" incluye cualquier cosa que indique la participación de la mano del hombre, tal como uno o más componentes del sistema que son alterados o mutados de su estado natural, al menos sustancialmente libres de al menos otro componente con que están naturalmente asociados en la naturaleza, o están asociados con al menos otro componente con el que no están asociados de forma natural. Por ejemplo, algunos sistemas CRISPR/Cas emplean complejos CRISPR no naturales que comprenden un ARNg y una proteína Cas que no se presentan de forma natural juntos.

Se proporcionan también variantes activas y fragmentos de agentes de nucleasa (es decir, un agente de nucleasa modificado por ingeniería genética). Tales variantes activas pueden comprender al menos 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 % o más de identidad de secuencia con el agente de nucleasa nativa, en las que las variantes activas retienen la capacidad de cortar en un sitio de reconocimiento deseado y, por lo tanto, retienen la actividad de inducción de muesca o rompimiento de cadena doble. Por ejemplo, cualquiera de los agentes de nucleasa descritos en este documento puede modificarse a partir de una secuencia de endonucleasa nativa y diseñarse para reconocer e inducir una muesca o rompimiento de la cadena doble en un sitio de reconocimiento que no fue reconocido por el agente de nucleasa nativo. Por lo tanto, en algunas realizaciones, la nucleasa modificada por ingeniería genética tiene una especificidad para inducir una muesca o rompimiento de la cadena doble en un sitio de reconocimiento que es diferente del sitio de reconocimiento del agente de nucleasa nativo correspondiente. Los ensayos para la actividad de inducción de una muesca o rompimiento de cadena doble son conocidos y generalmente miden la actividad global y la especificidad de la endonucleasa en sustratos de ADN que contienen el sitio de reconocimiento.

IV. Sistemas CRISPR/Cas (complejo de ARNg-Cas)

Los presentes métodos pueden emplear un sistema CRISPR/Cas (por ejemplo, un complejo de ARNg-Cas) para la escisión dirigida al sitio de ácidos nucleicos. Específicamente, la escisión de Cas de ácidos nucleicos dirigida por ARNg a un sitio objetivo identificado produce un ácido nucleico digerido con extremos que luego pueden unirse a un segundo ácido nucleico para ensamblar dos o más ácidos nucleicos de una manera específica del sitio.

Un "complejo de ARNg-Cas" incluye un complejo de una proteína Cas con un ARNg. El ARNg se puede diseñar o seleccionar para dirigir la escisión de Cas a un sitio objetivo que crea extremos solapantes entre el ácido nucleico escindido y un ácido nucleico diferente. El complejo de ARNg-Cas puede suministrarse con los agentes ya complejados, o puede suministrarse con la proteína y los elementos de ARN por separado, en cuyo caso se complejan para formar un complejo de ARNg-Cas en los métodos y mezclas de reacción descritos en este documento.

A. Endonucleasas guiadas por ARN de Cas

Las proteínas Cas generalmente comprenden al menos un dominio de unión o reconocimiento de ARN. Dichos dominios pueden interactuar con los ARN guía (ARNg, que se describen con más detalle a continuación). Las proteínas Cas también pueden comprender dominios de nucleasas (por ejemplo, dominios de DNasa o RNasa), dominios de unión a ADN, dominios de helicasa, dominios de interacción proteína-proteína, dominios de dimerización y otros dominios. Un dominio de nucleasa posee actividad catalítica para la escisión del ácido nucleico. La escisión incluye el rompimiento de los enlaces covalentes de una molécula de ácido nucleico. La escisión puede producir extremos romos o extremos escalonados, y puede ser de cadena sencilla o de cadena doble.

Ejemplos de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 o Csx12), Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 y Cu1966, y homólogos o versiones modificadas de las mismas.

Se puede usar cualquier proteína Cas que induzca una muesca o rompimiento de cadena doble en un sitio de reconocimiento deseado en los métodos y composiciones divulgados en este documento. Puede emplearse una proteína Cas nativa o natural siempre que la proteína Cas induzca el rompimiento de cadena doble en un sitio de reconocimiento deseado. Alternativamente, se puede emplear una proteína Cas modificada o alterada por ingeniería genética. Una "proteína Cas modificada por ingeniería genética" comprende una proteína Cas que está modificada por ingeniería genética (modificada o derivada) a partir de su forma nativa para reconocer específicamente e inducir una muesca o rompimiento de cadena doble en el sitio de reconocimiento deseado. Por lo tanto, una proteína Cas modificada por ingeniería genética puede derivarse de una proteína Cas nativa de origen natural o puede crearse o sintetizarse artificialmente.

En realizaciones particulares, la proteína Cas es Cas9. Las proteínas Cas9 generalmente comparten cuatro motivos clave con una arquitectura conservada. Los motivos 1, 2 y 4 son motivos similares a RuvC, y el motivo 3 es un motivo HNH. La actividad nucleasa de Cas9 escinde el ADN objetivo para producir rompimientos de cadena doble. Estos rompimientos pueden ser reparados por la célula en una de dos formas: unión del extremo no homólogo y reparación dirigida por homología (recombinación homóloga). En la unión del extremo no homólogo (NHEJ), los rompimientos de cadena doble se reparan por ligación directa de los extremos de rompimiento entre sí. Como tal, no se inserta ningún nuevo material de ácido nucleico en el sitio, aunque se puede perder algo de material de ácido nucleico, dando como resultado una eliminación. En la reparación dirigida por homología, puede usarse un polinucleótido donador con homología con la secuencia de ADN objetivo escindida como molde para la reparación de la secuencia de ADN objetivo escindida, dando como resultado la transferencia de información genética desde el polinucleótido donante al ADN objetivo. Por lo tanto, se puede insertar/copiar nuevo material de ácido nucleico en el sitio. Las modificaciones del ADN objetivo debido a NHEJ y/o la reparación dirigida por homología se pueden usar para la corrección de genes, reemplazo de genes, marcación de genes, inserción de transgenes, eliminación de nucleótidos, rompimiento de genes, mutación genética, etc.

Las proteínas Cas pueden ser de un sistema CRISPR/Cas de tipo II. Por ejemplo, la proteína Cas puede ser una proteína Cas9 o derivarse de una proteína Cas9. Las proteínas Cas9 generalmente comparten cuatro motivos clave con una arquitectura conservada. Los motivos 1, 2 y 4 son motivos similares a RuvC, y el motivo 3 es un motivo HNH. La proteína Cas9 puede ser, por ejemplo, de *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Staphylococcus aureus*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycooides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccii*, *Candidatus desulfurudis*, *Clostridium botulinum*, *Clostridium difficile*, *Fingoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus*, o *Acaryochloris marina*. Ejemplos adicionales de los miembros de la familia Cas9 se describen en el documento WO 2014/131833. La proteína Cas9 de *S. pyogenes* derivada de la misma es una enzima preferida. A la proteína Cas9 de *S. pyogenes* se le asigna al número de acceso Q99ZW2 del SwissProt.

Las proteínas Cas pueden ser proteínas de tipo silvestre (es decir, aquellas que se encuentran en la naturaleza), proteínas Cas modificadas (es decir, variantes de la proteína Cas), o fragmentos de proteínas Cas de tipo silvestre o modificadas. Las proteínas Cas también pueden ser variantes activas o fragmentos de proteínas Cas naturales o modificadas. Las variantes o fragmentos activos pueden comprender al menos el 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de la identidad de secuencia con la proteína de Cas modificada o de tipo silvestre o una porción de la misma, en la que las variantes activas conservan la capacidad para cortar en un sitio de escisión deseado y, por lo tanto, retienen la actividad inductora de muescas o inductora de rompimiento de cadena

doble. Se conocen ensayos para inducir la actividad inductora de muescas o inductora de rompimiento de cadena doble y, en general, medir la actividad global y la especificidad de la proteína Cas en sustratos de ADN que contienen el sitio de escisión.

5 Las proteínas Cas se pueden modificar para aumentar o disminuir la afinidad de unión a ácidos nucleicos, la especificidad de unión a ácidos nucleicos y/o la actividad enzimática. Las proteínas Cas también se pueden modificar para cambiar cualquier otra actividad o propiedad de la proteína, como la estabilidad. Por ejemplo, uno o más dominios nucleasa de la proteína Cas pueden modificarse, eliminarse o inactivarse, o una proteína Cas puede truncarse para eliminar dominios que no son esenciales para la función de la proteína o para optimizar (por ejemplo, mejorar o reducir) la actividad de la proteína Cas.

10 Algunas proteínas Cas comprenden al menos dos dominios nucleasa, tales como dominios DNasa. Por ejemplo, una proteína Cas9 puede comprender un dominio de nucleasa similar a RuvC y un dominio de nucleasa similar a HNH. Los dominios RuvC y HNH pueden cada uno cortar una cadena diferente de ADN de doble cadena para formar un rompimiento de doble cadena en el ADN. Véase, por ejemplo, Jinek et al. (2012) *Science* 337: 816-821.

15 Uno o ambos de los dominios de nucleasa pueden eliminarse o mutarse de forma que ya no sean funcionales o tengan actividad nucleasa reducida. Si uno de los dominios nucleasa se elimina o muta, la proteína Cas resultante (por ejemplo, Cas9) puede denominarse como una nickasa y puede generar un rompimiento de cadena sencilla en una secuencia de reconocimiento de ARN de CRISPR dentro de un ADN de doble cadena, pero no un rompimiento de cadena doble (es decir, puede escindir la cadena complementaria o la cadena no complementaria, pero no ambas). Si los dos dominios nucleasa se eliminan o mutan, la proteína Cas resultante (por ejemplo, Cas9) tendrá una capacidad reducida para escindir ambas cadenas de un ADN de doble cadena. Un ejemplo de una mutación que convierte Cas9 en una nickasa es una mutación D10A (aspartato por alanina en la posición 10 de Cas9) en el dominio RuvC de Cas9 de *S. pyogenes*. Del mismo modo que, H939A (histidina por alanina en la posición del aminoácido 839) o H840A (histidina por alanina en la posición del aminoácido 840) en el dominio HNH de Cas9 de *S. Pyogenes* puede convertir la Cas9 en una nickasa. Otros ejemplos de mutaciones que convierten Cas9 en una nickasa incluyen las mutaciones correspondientes a Cas9 de *S. thermophilus*. Véase, por ejemplo, Sapranaukas et al., (2011) *Nucleic Acids Research* 39: 9275-9282 y el documento WO 2013/141680. Dichas mutaciones pueden generarse usando métodos tales como mutagénesis dirigida al sitio, mutagénesis mediada por PCR o síntesis génica total. Se pueden encontrar ejemplos de otras mutaciones que crean nickasas, por ejemplo, en los documentos WO/2013/176772A1 y WO/2013/142578A1.

20 Las proteínas Cas también pueden ser proteínas de fusión. Por ejemplo, una proteína Cas puede fusionarse a un dominio de escisión, un dominio de modificación epigenética, un dominio de activación transcripcional, o un dominio represor transcripcional, véase el documento WO 2014/089290. Las proteínas Cas también se pueden fusionar con un polipéptido heterólogo que proporciona mayor o menor estabilidad. El dominio fusionado o el polipéptido heterólogo pueden localizarse en el terminal N, el terminal C, o internamente dentro de la proteína Cas.

25 Una proteína Cas se puede fusionar a un polipéptido heterólogo que proporciona la localización subcelular. Dichos péptidos heterólogos incluyen, por ejemplo, una señal de localización nuclear (NLS) tal como la NLS de SV40 para dirigirse al núcleo, una señal de localización mitocondrial para dirigirse a la mitocondria, una señal de retención de ER, y similares. Véase, por ejemplo, Lange et al., (2007) *J. Biol. Chem.* 282: 5101-5105. Dichas señales de localización subcelular pueden localizarse en el terminal N, el terminal C o en cualquier lugar dentro de la proteína Cas. Una NLS puede comprender un tramo de aminoácidos básicos, y puede ser una secuencia monopartita o una secuencia bipartita.

30 Las proteínas Cas también se pueden unir a un dominio de penetración celular. Por ejemplo, el dominio de penetración celular puede derivarse de la proteína TAT del VIH-1, el motivo de penetración celular TLM del virus de la hepatitis B humana, MPG, Pep-1, VP22, un péptido de penetración celular del virus Herpes simple o secuencia de péptido de poliarginina. Véase, por ejemplo, el documento WO 2014/089290. El dominio de penetración celular puede localizarse en el terminal N, el terminal C o en cualquier lugar dentro de la proteína Cas.

35 Las proteínas Cas también pueden comprender un polipéptido heterólogo para facilitar el rastreo o la purificación, tal como una proteína fluorescente, una etiqueta de purificación o una etiqueta de epítipo. Los ejemplos de proteínas fluorescentes incluyen proteínas fluorescentes verdes (por ejemplo, GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Azami Green Monomérica, CopGFP, AceGFP, ZsGreen1), proteínas fluorescentes amarillas (por ejemplo, YFP, eYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), proteínas fluorescentes azules (por ejemplo, eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), proteínas fluorescentes cian (por ejemplo, eCFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), proteínas fluorescentes rojas (mKate, mKate2, mPlum, DsRed monómero, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monómero, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), proteínas fluorescentes naranja (mOrange, mKO, Kusabira-Orange, Kusabira-Orange Monomérica, mTangerine, tdTomato) y cualquier otra proteína fluorescente adecuada. Ejemplos de etiquetas incluyen glutatión S-transferasa (GST), proteína de unión a quitina (CBP), proteína de unión a maltosa, tiorredoxina (TRX), poli (NANP), etiqueta de purificación por afinidad en tándem (TAP), myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, hemaglutinina (HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, histidina (His), proteína portadora de carboxilo de biotina (BCCP) y calmodulina.

En algunas realizaciones, la proteína Cas puede modificarse de modo que la actividad nucleasa resultante se altera.

Ciertas mutaciones en Cas pueden reducir la capacidad de la nucleasa para escindir tanto las cadenas complementarias como las no complementarias del ADN objetivo. Por ejemplo, las proteínas Cas pueden mutarse en posiciones conocidas de manera que la actividad nucleasa se limite a la escisión de la cadena complementaria o la cadena no complementaria. Específicamente, Cas9 que tiene una mutación D10A (aspartato por alanina en la posición de aminoácido 10 de Cas9) puede escindir la cadena complementaria del ADN objetivo, pero tiene una capacidad reducida para escindir la cadena no complementaria del ADN objetivo. En algunas realizaciones, Cas9 que tiene una mutación H840A (histidina por alanina en la posición del aminoácido 840) puede escindir la cadena no complementaria del ADN objetivo, pero tiene una capacidad reducida para escindir la cadena complementaria del ADN objetivo. La actividad nucleasa de Cas9 que tiene una mutación D10A o H840A daría como resultado un rompimiento de una sola cadena (SSB) en lugar de una DSB. Otros residuos pueden mutarse para lograr el mismo efecto (es decir, inactivar una o las otras porciones de nucleasa). Como ejemplos no limitantes, residuos D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, y/o A987 (es decir, sustituidos). Además, pueden ser adecuados aminoácidos sustitutos distintos de alanina. En algunas realizaciones, cuando una nucleasa tiene actividad reducida (por ejemplo, cuando una proteína Cas9 tiene una mutación D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 y/o A987, tal como D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A y/o D986A), la nucleasa aún se puede unir al ADN objetivo de una manera específica del sitio porque todavía se guía a una secuencia de ADN objetivo mediante un ARNg) siempre que conserve la capacidad de interactuar con el ARNg.

En algunas realizaciones, Cas se altera de modo que la nucleasa no escinde ni la cadena complementaria ni la no complementaria del ADN objetivo. Por ejemplo, Cas9 con las mutaciones D10A y H840A tiene una capacidad reducida para escindir tanto las cadenas complementarias como la no complementaria del ADN objetivo. Otros residuos pueden mutarse para lograr el mismo efecto (es decir, inactivar una o las otras porciones de nucleasa). Como ejemplos no limitantes, los residuos D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, y/o pueden sustituirse para eliminar sustancialmente la actividad nucleasa. Además, las mutaciones distintas de las sustituciones de alanina pueden ser adecuadas.

Los términos "sitio objetivo" o "secuencia objetivo" pueden usarse indistintamente e incluyen secuencias de ácido nucleico presentes en un ADN objetivo al que se unirá un segmento de direccionamiento de ADN de un ARNg, con la condición de que existan condiciones suficientes para la unión. Por ejemplo, el sitio objetivo (o la secuencia objetivo) dentro de un ADN objetivo está dirigido por (o está unido por, o está hibridado con, o es complementario a) la proteína Cas o ARNg. Las condiciones de unión de ADN/ARN adecuadas incluyen condiciones fisiológicas normalmente presentes en una célula. Se conocen en la técnica otras condiciones adecuadas de unión ADN/ARN (por ejemplo, condiciones en un sistema sin células) (véase, por ejemplo, Molecular Cloning: A Laboratory Manual, 3ª edición (Sambrook et al., Harbor Laboratory Press 2001)). La cadena del ADN objetivo que es complementaria e hibrida con la proteína Cas o ARNg se denomina como la "cadena complementaria" y la cadena del ADN objetivo que es complementaria a la "cadena complementaria" (y por lo tanto no es complementaria a la proteína Cas o ARNg) se conoce como la "cadena no complementaria" o "cadena plantilla".

La proteína Cas puede escindir el ácido nucleico en un sitio dentro de la secuencia objetivo o fuera de la secuencia objetivo. El "sitio de escisión" incluye la posición de un ácido nucleico en el que una proteína Cas produce un rompimiento de cadena sencilla o un rompimiento de cadena doble. Si la proteína Cas produce un rompimiento de cadena doble, el sitio de escisión puede estar en la misma posición en ambas cadenas del ácido nucleico (produciendo extremos romos) o puede estar en diferentes sitios en cada cadena (produciendo extremos pegajosos o cohesivos). Los extremos pegajosos también se pueden producir mediante el uso de dos proteínas Cas que producen un rompimiento de una sola cadena en los sitios de escisión en cada cadena. La escisión específica del sitio del ADN objetivo por Cas9 puede ocurrir en ubicaciones determinadas por (i) complementariedad de emparejamiento de bases entre el ARNg guía y el ADN objetivo; y (ii) un motivo corto, denominado motivo adyacente protoespaciador (PAM), en el ADN objetivo. Por ejemplo, el sitio de escisión de Cas9 puede ser de aproximadamente 1 a aproximadamente 10 o de aproximadamente 2 a aproximadamente 5 pares de bases (por ejemplo, 3 pares de bases) secuencia arriba de la secuencia de PAM. En algunas realizaciones (por ejemplo, cuando se usa Cas9 de *S. pyogenes*, o una Cas9 estrechamente relacionada), la secuencia de PAM de la cadena no complementaria puede ser 5'-XGG-3', donde X es cualquier nucleótido de ADN y X es inmediatamente 3' de la secuencia objetivo de la cadena no complementaria del ADN objetivo. Como tal, la secuencia de PAM de la cadena complementaria sería 5'-CCY-3', donde Y es cualquier nucleótido de ADN y Y es inmediatamente 5' de la secuencia objetivo de la cadena complementaria del ADN objetivo. En algunas de tales realizaciones, X y Y pueden ser complementarias y el par de bases XY pueden ser cualquier par de bases (por ejemplo, X = C y Y = G; X = G y Y = C; X = A y Y = T, X = T y Y = A).

Las proteínas Cas pueden proporcionarse en cualquier forma. Por ejemplo, una proteína Cas puede proporcionarse en forma de una proteína, tal como una proteína Cas complejada con un ARNg. Alternativamente, se puede proporcionar una proteína Cas en forma de un ácido nucleico que codifica la proteína Cas, tal como un ARN (por ejemplo, ARN mensajero (ARNm)) o ADN. Opcionalmente, el ácido nucleico que codifica la proteína Cas puede optimizarse con un codón para la traducción eficiente en proteína en una célula u organismo particular. Por ejemplo, el ácido nucleico que codifica la proteína Cas puede modificarse para sustituir codones que tienen una mayor frecuencia de uso en una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata, o cualquier otra célula huésped de interés, en comparación con la secuencia de polinucleótido natural. Cuando se introduce un ácido nucleico que codifica la proteína Cas en la célula, la proteína Cas puede expresarse de forma transitoria, condicional o constitutiva en la célula.

Los ácidos nucleicos que codifican las proteínas Cas pueden integrarse de forma estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ácidos nucleicos que codifican las proteínas Cas se pueden unir operativamente a un promotor en una construcción de expresión. Las construcciones de expresión incluyen cualquier construcción de ácido nucleico capaz de dirigir la expresión de un gen u otra secuencia de ácido nucleico de interés (por ejemplo, un gen Cas) y que puede transferir dicha secuencia de ácido nucleico de interés a una célula objetivo. Por ejemplo, el ácido nucleico que codifica la proteína Cas puede estar en el vector de direccionamiento que comprende el inserto de ácido nucleico y/o un vector que comprende el ADN que codifica el ARNg, o puede estar en un vector o un plásmido que está separado del vector objetivo que comprende el inserto de ácido nucleico y/o separado de un vector que comprende el ADN que codifica el ARNg. Los promotores que pueden usarse en una construcción de expresión incluyen, por ejemplo, promotores activos en una célula pluripotente de rata, eucariota, mamífero, mamífero no humano, ser humano, roedor, ratón o hámster. Dichos promotores pueden ser, por ejemplo, promotores condicionales, promotores inducibles, promotores constitutivos o promotores específicos de tejido. Los ejemplos de otros promotores se describen en otra parte del presente documento.

B. ARN guía (ARNg)

Un "ARN guía" o "ARNg" incluye una molécula de ARN que se une a una proteína Cas y dirige la proteína Cas a una ubicación específica dentro de un ADN objetivo. Los ARN guía (ARNg) pueden comprender dos segmentos, un "segmento de dirección a ADN" y un "segmento de unión a proteína". "Segmento" incluye un segmento, sección o región de una molécula, tal como un tramo contiguo de nucleótidos en un ARN. Algunos ARNg comprenden dos moléculas de ARN separadas: un "ARN activador" y un "ARN direccionador". Otros ARNg son una sola molécula de ARN (un solo polinucleótido de ARN), que también se puede llamar un "ARNg de una sola molécula", o un "ARN guía solo" o un "ARNgs". Véase, por ejemplo, los documentos WO/2013/176772A1, WO/2014/065596A1, WO/2014/089290A1, WO/2014/093622A2, WO/2014/099750A2, WO/2013/142578A1, y WO 2014/131833A1. Los términos "ARN de guía" y "ARNg" incluyen tanto ARNg de doble molécula como ARNg de una sola molécula.

Un ejemplo de ARNg de dos moléculas comprende una molécula similar a ARNcr ("ARN de CRISPR" o "ARN direccionador" o "ARNcr" o "repetición de ARNcr") y una molécula correspondiente similar a ARNcrtra ("ARN de CRISPR que actúa en forma trans" o "ARN activador" o "ARNcrtra" o "andamio"). Un ARNcr comprende tanto el segmento de direccionamiento de ADN (de cadena sencilla) del ARNg y un tramo de nucleótidos que forma la mitad del dúplex ARNdc del segmento de unión a proteína del ARNg. Un ARNcrtra correspondiente (ARN activador) comprende un tramo de nucleótidos que forma la otra mitad del dúplex ARNdc del segmento de unión a proteína del ARNg. Un tramo de nucleótidos de un ARNcr es complementario y se hibrida con un tramo de nucleótidos de un ARNcrtra para formar el dúplex de ARNdc del dominio de unión a proteína del ARNg. Como tal, se puede decir que cada ARNcr tiene un ARNcrtra correspondiente. El ARNcr proporciona adicionalmente el segmento de direccionamiento de ADN de cadena sencilla. Por consiguiente, un ARNg comprende una secuencia que hibrida con una secuencia objetivo, y un ARNcrtra.

El ARNcr y el correspondiente ARNcrtra (como un par correspondiente) se hibridan para formar un ARNg. El ARNcr proporciona adicionalmente el segmento de direccionamiento de ADN de cadena sencilla que hibrida con una secuencia de reconocimiento de ARN de CRISPR. Si se usa para modificación dentro de una célula, la secuencia exacta de una molécula de ARNcr o ARNcrtra dada puede diseñarse para que sea específica de la especie en la que se usarán las moléculas de ARN. Véase, por ejemplo, Mali P et al., (2013) Science 2013 15 de febrero; 339 (6121): 823-6; Jinek M et al., Science 2012 17 de agosto; 337 (6096): 816-21; Hwang WY et al., Nat Biotechnol 2013 marzo; 31 (3): 227-9; Jiang W et al., Nat Biotechnol 2013 marzo; 31 (3): 233-9; y, Cong L et al., Science 2013 15 de febrero; 339 (6121): 819-23.

El segmento de direccionamiento de ADN (ARNcr) de un ARNg dado comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN objetivo. El segmento de direccionamiento de ADN de un ARNg interactúa con un ADN objetivo en una forma específica de la secuencia mediante hibridación (es decir, emparejamiento de bases). Como tal, la secuencia de nucleótidos del segmento de direccionamiento de ADN puede variar y determina la ubicación dentro del ADN objetivo con el que interactuarán el ARNg y el ADN objetivo. El segmento de direccionamiento de ADN de un ARNg objetivo puede modificarse para hibridar con cualquier secuencia deseada dentro de un ADN objetivo. Los ARNcr naturales se diferencian dependiendo del sistema Cas9 y del organismo, pero a menudo contienen un segmento de direccionamiento de entre 21 y 72 nucleótidos de longitud, flanqueado por dos repeticiones directas (DR) de una longitud de entre 21 y 46 nucleótidos (véase, por ejemplo, el documento WO2014/131833). En el caso de *S. pyogenes*, las DR tienen 36 nucleótidos de longitud y el segmento de direccionamiento tiene 30 nucleótidos de longitud. La DR localizada en 3' es complementaria y se hibrida con el ARNcrtra correspondiente, que a su vez se une a la proteína Cas9.

El segmento de direccionamiento de ADN puede tener una longitud de aproximadamente 12 nucleótidos a aproximadamente 100 nucleótidos. Por ejemplo, el segmento de direccionamiento de ADN puede tener una longitud de aproximadamente 12 nucleótidos (nt) a aproximadamente 80 nt, de aproximadamente 12 nt a aproximadamente 50 nt, de aproximadamente 12 nt a aproximadamente 40 nt, de aproximadamente 12 nt a aproximadamente 30 nt, de aproximadamente 12 nt a aproximadamente 25 nt, de aproximadamente 12 nt a aproximadamente 20 nt, o de aproximadamente 12 nt a aproximadamente 19 nt. Alternativamente, el segmento de direccionamiento de ADN puede tener una longitud de aproximadamente 19 nt a aproximadamente 20 nt, de aproximadamente 19 nt a aproximadamente 25 nt, de aproximadamente 19 nt a aproximadamente 30 nt, de aproximadamente 19 nt a aproximadamente 35 nt, de

aproximadamente 19 nt a aproximadamente 40 nt, de aproximadamente 19 nt a aproximadamente 45 nt, de aproximadamente 19 nt a aproximadamente 50 nt, de aproximadamente 19 nt a aproximadamente 60 nt, de aproximadamente 19 nt a aproximadamente 70 nt, de aproximadamente 19 nt a aproximadamente 80 nt, de aproximadamente 19 nt a aproximadamente 90 nt, de aproximadamente 19 nt a aproximadamente 100 nt, de aproximadamente 20 nt a aproximadamente 25 nt, de aproximadamente 20 nt a aproximadamente 30 nt, de aproximadamente 20 nt a aproximadamente 35 nt, de aproximadamente 20 nt a aproximadamente 40 nt, de aproximadamente 20 nt a aproximadamente 45 nt, de aproximadamente 20 nt a aproximadamente 50 nt, de aproximadamente 20 nt a aproximadamente 60 nt, de aproximadamente 20 nt a aproximadamente 70 nt, de aproximadamente 20 nt a aproximadamente 80 nt, de aproximadamente 20 nt a aproximadamente 90 nt, o de aproximadamente 20 nt a aproximadamente 100 nt.

La secuencia de nucleótidos del segmento de direccionamiento de ADN que es complementario a una secuencia de nucleótidos (secuencia de reconocimiento de ARN de CRISPR) del ADN objetivo puede tener una longitud de al menos aproximadamente 12 nt. Por ejemplo, la secuencia de direccionamiento a ADN (por ejemplo, la secuencia dentro del segmento de direccionamiento de ADN que es complementaria a una secuencia de reconocimiento de ARN de CRISPR dentro del ADN objetivo) puede tener una longitud de al menos aproximadamente 12 nt, al menos aproximadamente 15 nt, al menos aproximadamente 18 nt, al menos aproximadamente 19 nt, al menos aproximadamente 20 nt, al menos aproximadamente 25 nt, al menos aproximadamente 30 nt, al menos aproximadamente 35 nt, o al menos aproximadamente 40 nt. Alternativamente, la secuencia de direccionamiento a ADN del segmento de direccionamiento de ADN que es complementario a una secuencia objetivo del ADN objetivo puede tener una longitud de aproximadamente 12 nucleótidos (nt) a aproximadamente 80 nt, de aproximadamente 12 nt a aproximadamente 50 nt, de aproximadamente 12 nt a aproximadamente 45 nt, de aproximadamente 12 nt a aproximadamente 40 nt, de aproximadamente 12 nt a aproximadamente 35 nt, de aproximadamente 12 nt a aproximadamente 30 nt, de aproximadamente 12 nt a aproximadamente 25 nt, de aproximadamente 12 nt a aproximadamente 20 nt, de aproximadamente 12 nt a aproximadamente 19 nt, de aproximadamente 19 nt a aproximadamente 20 nt, de aproximadamente 19 nt a aproximadamente 25 nt, de aproximadamente 19 nt a aproximadamente 30 nt, de aproximadamente 19 nt a aproximadamente 35 nt, desde aproximadamente 19 nt a aproximadamente 40 nt, de aproximadamente 19 nt a aproximadamente 45 nt, de aproximadamente 19 nt a aproximadamente 50 nt, de aproximadamente 19 nt a aproximadamente 60 nt, de aproximadamente 20 nt a aproximadamente 25 nt, de aproximadamente 20 nt a aproximadamente 30 nt, de aproximadamente 20 nt a aproximadamente 35 nt, de aproximadamente 20 nt a aproximadamente 40 nt, de aproximadamente 20 nt a aproximadamente 45 nt, de aproximadamente 20 nt a aproximadamente 50 nt, o de aproximadamente 20 nt a aproximadamente 60 nt. La secuencia de nucleótidos (la secuencia de direccionamiento de ADN) del segmento de direccionamiento de ADN que es complementaria a una secuencia de nucleótidos (secuencia objetivo) del ADN objetivo puede tener una longitud de al menos aproximadamente 12 nt. En algunos casos, la secuencia de direccionamiento de ADN puede tener una longitud de al menos aproximadamente 20 nt.

Los ARNcrtra pueden estar en cualquier forma (por ejemplo, ARNcrtra de longitud completa o ARNcrtra parcialmente activos) y de diferentes longitudes. Pueden incluir transcripciones primarias o formas procesadas. Por ejemplo, los ARNcrtra (como parte de un ARN guía solo o como una molécula separada como parte de un ARNg de dos moléculas) pueden comprender o constar de la totalidad o una porción de una secuencia de ARNcrtra de tipo silvestre (por ejemplo, aproximadamente o más de aproximadamente 20, 26, 32, 45, 48, 54, 63, 67, 85 o más nucleótidos de una secuencia de ARNcrtra de tipo silvestre). Los ejemplos de secuencias de ARNcrtra de tipo silvestre de tipo silvestre incluyen versiones de 171 nucleótidos, 89 nucleótidos, 75 nucleótidos y 65 nucleótidos. Véase, por ejemplo, Deltcheva et al., (2011) Nature 471: 602 - 607; WO 2014/093661. Los ejemplos de ARNcrtra en ARN guía solos (ARNgs) incluyen los segmentos de ARNcrtra encontrados en las versiones +48, +54, +67 y +85 de los ARNg, donde "+ n" indica que hasta el nucleótido + n del ARNcrtra de tipo silvestre está incluido en el ARNg. Véase el documento US 8.697.359.

El porcentaje de complementariedad entre la secuencia de direccionamiento de ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN objetivo puede ser de al menos 60% (por ejemplo, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 97%, al menos 98%, al menos 99% o 100%). El porcentaje de complementariedad entre la secuencia de direccionamiento de ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN objetivo es del 100% sobre los siete nucleótidos más 5' contiguos de la secuencia objetivo de la cadena complementaria del ADN objetivo. En ciertas realizaciones, el porcentaje de complementariedad entre la secuencia de direccionamiento de ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN objetivo puede ser de al menos 60% sobre aproximadamente 20 nucleótidos contiguos. Como ejemplo, el porcentaje de complementariedad entre la secuencia de direccionamiento de ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN objetivo es 100% sobre los catorce nucleótidos contiguos en el extremo más 5' de la secuencia de reconocimiento de ARN de CRISPR dentro de la cadena complementaria del ADN objetivo y tan bajo como 0% sobre el resto. En tal caso, la secuencia de direccionamiento de ADN se puede considerar que es de 14 nucleótidos de longitud. Como otro ejemplo, el porcentaje de complementariedad entre la secuencia de direccionamiento de ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN objetivo es 100% sobre los siete nucleótidos contiguos en el extremo más 5' de la secuencia de reconocimiento de ARN de CRISPR dentro de la cadena complementaria del ADN objetivo y tan bajo como 0% sobre el resto. En tal caso, la secuencia de direccionamiento de ADN se puede considerar de 7 nucleótidos de longitud.

Complementariedad de ácidos nucleicos significa que una secuencia de nucleótidos en una cadena de ácido nucleico,

debido a la orientación de sus grupos de nucleobases, se une a través del hidrógeno con otra secuencia en una cadena de ácido nucleico opuesta. Las bases de complementariedad típicamente son, en ADN: A con T y C con G, y, en ARN: C con G, y U con A. La complementariedad puede ser perfecta o sustancial/suficiente. La complementariedad perfecta entre dos ácidos nucleicos significa que los dos ácidos nucleicos pueden formar un dúplex en el que cada base en el dúplex se une a una base complementaria mediante emparejamiento Watson-Crick. Complementariedad "sustancial" o "suficiente" significa que una secuencia en una cadena no es completa y/o perfectamente complementaria a una secuencia en una cadena opuesta, pero que se produce una unión suficiente entre bases en las dos cadenas para formar un complejo híbrido estable en conjunto con las condiciones de hibridación (por ejemplo, concentración de sal y temperatura). Tales condiciones pueden predecirse mediante el uso de las secuencias y cálculos matemáticos estándar para predecir la T_m de cadenas hibridadas, o mediante la determinación empírica de T_m usando métodos rutinarios. T_m se refiere a la temperatura a la que se desnaturaliza 50% de la población de complejos de hibridación formados entre dos cadenas de ácido nucleico. A una temperatura por debajo de la T_m , se favorece la formación de un complejo de hibridación, mientras que, a una temperatura superior a la T_m , se favorece la fusión o la separación de las cadenas en el complejo de hibridación. La T_m puede estimarse para un ácido nucleico que tiene un contenido de G + C conocido en una solución acuosa de NaCl 1 M mediante el uso, por ejemplo, $T_m = 81,5 + 0,41 (\% G + C)$, aunque otros cálculos de T_m conocidos tienen en cuenta las características estructurales del ácido nucleico.

"Condición de hibridación" se refiere al entorno acumulativo en el que una cadena de ácido nucleico se une a una segunda cadena de ácido nucleico mediante interacciones de cadena complementaria y enlaces de hidrógeno para producir un complejo de hibridación. Tales condiciones incluyen los componentes químicos y sus concentraciones (por ejemplo, sales, agentes quelantes, formamida) de una solución acuosa u orgánica que contiene los ácidos nucleicos, y la temperatura de la mezcla. Otros factores, tales como la duración del tiempo de incubación o las dimensiones de la cámara de reacción pueden contribuir al entorno (por ejemplo, Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2ª ed., páginas 1.90-1.91, 9.47- 9.51, 1 1.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)).

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque son posibles desajustes entre las bases. Las condiciones apropiadas para la hibridación entre dos ácidos nucleicos dependen de la longitud de los ácidos nucleicos y del grado de complementación, variables bien conocidas en la técnica. Cuanto mayor es el grado de complementación entre dos secuencias de nucleótidos, mayor es el valor de la temperatura de fusión (T_m) para los híbridos de ácidos nucleicos que tienen esas secuencias. Para las hibridaciones entre ácidos nucleicos con tramos cortos de complementariedad (por ejemplo, complementariedad mayor de 35 o menos, 30 o menos, 25 o menos, 22 o menos, 20 o menos, o 18 o menos nucleótidos) la posición de los desapareamientos se vuelve importante (véase Sambrook et al., citado más arriba, 11.7-11.8). Típicamente, la longitud para un ácido nucleico hibridable es al menos aproximadamente de 10 nucleótidos. Las longitudes mínimas ilustrativas para un ácido nucleico hibridable son: al menos aproximadamente 15 nucleótidos; al menos aproximadamente 20 nucleótidos; al menos aproximadamente 22 nucleótidos; al menos aproximadamente 25 nucleótidos; y al menos aproximadamente 30 nucleótidos). Además, la temperatura y la concentración de sal de la solución de lavado pueden ajustarse según sea necesario de acuerdo con factores tales como la longitud de la región de complementación y el grado de complementación.

La secuencia del polinucleótido no necesita ser 100% complementaria a la de su ácido nucleico objetivo para ser específicamente hibridable. Además, un polinucleótido puede hibridarse sobre uno o más segmentos de manera que los segmentos intervinientes o adyacentes no están implicados en el evento de hibridación (por ejemplo, una estructura de bucle o estructura en horquilla). Un polinucleótido (por ejemplo, ARNg) puede comprender al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 99%, o 100% de complementariedad de secuencia con una región objetivo dentro de la secuencia de ácido nucleico objetivo para la cuál ellos apuntan. Por ejemplo, un ARNg en el que 18 de 20 nucleótidos del ARNg son complementarios a una región objetivo, y por lo tanto hibridarían específicamente, representaría un 90 por ciento de complementariedad. En este ejemplo, los nucleótidos no complementarios restantes pueden estar agrupados o intercalados con nucleótidos complementarios y no necesitan ser contiguos entre sí o con nucleótidos complementarios. El porcentaje de complementariedad entre tramos particulares de secuencias de ácidos nucleicos dentro de ácidos nucleicos se puede determinar rutinariamente usando programas BLAST (herramientas de búsqueda de alineación local básica) y programas PowerBLAST conocidos en la técnica (Altschul et al., *J. Mol. Biol.*, 1990, 215, 403-410; Zhang y Madden, *Genome Res.*, 1997, 7, 649-656) o mediante el uso del programa Gap (Wisconsin Sequence Analysis Package, Versión 8 para Unix, Genetics Computer Group, University Research Park, Madison, WI), usando la configuración predeterminada, que usa el algoritmo de Smith y Waterman (*Adv. Appl. Math.*, 1981, 2, 482-489).

El segmento de unión a proteína de un ARNg objetivo interactúa con una proteína Cas. El ARNg objetivo dirige el polipéptido unido a una secuencia de nucleótidos específica dentro del ADN objetivo a través del segmento de direccionamiento a ADN. El segmento de unión a la proteína de un ARNg objetivo puede comprender dos tramos de nucleótidos que son complementarios entre sí. Los nucleótidos complementarios del segmento de unión a proteínas se hibridan para formar un dúplex de ARN de doble cadena (ARNdc). El segmento de unión a proteínas de un ARNg objetivo interactúa con la proteína Cas, y el ARNg dirige la proteína Cas unida a una secuencia de nucleótidos específica dentro del ADN objetivo a través del segmento de direccionamiento de ADN.

En ciertas realizaciones, un ARNg como se describe en el presente documento comprende dos moléculas de ARN separadas. Cada una de las dos moléculas de ARN de un ARNg objetivo comprende un tramo de nucleótidos que son

complementarios entre sí de modo que los nucleótidos complementarios de las dos moléculas de ARN se hibridan para formar el dúplex de doble cadena (por ejemplo, una horquilla) del segmento de unión a proteína. Un ARNg objetivo puede comprender cualquier par de ARNcr y ARNcrtra correspondiente. En los métodos descritos en este documento, el ARNg se puede usar como un complejo (por ejemplo, complejo de ARNg-Cas) de ARNcr y ARNcrtra o el ARNcr y el correspondiente ARNcrtra se pueden administrar por separado. Por ejemplo, si se usan múltiples ARNg para la reacción de escisión, los ARNcr individuales específicos para cada sitio objetivo pueden administrarse por separado de un ARNcrtra estándar que puede complejarse con cada ARNcr. En dicho método, los ARNcr pueden formar complejos con el ARNcrtra estándar para dirigir una proteína Cas al sitio objetivo.

Los ARN guía pueden incluir modificaciones o secuencias que proporcionan características deseables adicionales (por ejemplo, estabilidad modificada o regulada, direccionamiento subcelular, seguimiento, con una etiqueta fluorescente, un sitio de unión para una proteína o complejo de proteína y similares). Los ejemplos no limitantes de tales modificaciones incluyen, por ejemplo, una tapa 5' (por ejemplo, una tapa de 7-metilguanilato (m7G)); una cola 3' poliadenilada (es decir, una cola 3' poli(A)); una secuencia de riboconmutación (por ejemplo, para permitir la estabilidad regulada y/o accesibilidad regulada por proteínas y/o complejos proteicos); una secuencia de control de estabilidad; una secuencia que forma un dúplex ARNdc (es decir, una horquilla); una modificación o secuencia que dirige el ARN a una ubicación subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos, y similares); una modificación o secuencia que permite el seguimiento (por ejemplo, conjugación directa con una molécula fluorescente, conjugación con una fracción que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, y similares); una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre el ADN, incluyendo activadores transcripcionales, represores transcripcionales, ADN metiltransferasas, ADN desmetilasas, histonas acetiltransferasas, histonas desacetilasas, y similares); y combinaciones de los mismos.

Los ARN guía pueden proporcionarse en cualquier forma. Por ejemplo, el ARNg se puede proporcionar en forma de RNA, como dos moléculas (ARNcr y ARNcrtra por separado) o como una molécula (ARNgs), y opcionalmente en forma de un complejo con una proteína Cas. El ARNg también se puede proporcionar en forma de ADN que codifica el ARN. El ADN que codifica el ARNg puede codificar una molécula de ARN sola (ARNgs) o moléculas de ARN separadas (por ejemplo, ARNcr y ARNcrtra separados). En el último caso, el ADN que codifica el ARNg puede proporcionarse como moléculas de ADN separadas que codifican el ARNcr y el ARNcrtra, respectivamente.

Los ADN que codifican ARNg se pueden integrar de forma estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ADN que codifican ARNg se pueden unir operativamente a un promotor en una construcción de expresión. Por ejemplo, el ADN que codifica el ARNg puede estar en el vector de direccionamiento que comprende el inserto de ácido nucleico y/o un vector que comprende el ácido nucleico que codifica la proteína Cas, o puede estar en un vector o plásmido que se separa del vector de direccionamiento que comprende el inserto de ácido nucleico y/o separado de un vector que comprende el ácido nucleico que codifica la proteína Cas. Tales promotores pueden ser activos, por ejemplo, en una célula pluripotente de rata, eucariota, mamífero, mamífero no humano, humano, roedor, ratón o hámster. Dichos promotores pueden ser, por ejemplo, promotores condicionales, promotores inducibles, promotores constitutivos o promotores específicos de tejido. En algunos casos, el promotor es un promotor de ARN polimerasa III, tal como un promotor U6 humano, un promotor U6 polimerasa III de rata o un promotor U6 de polimerasa III de ratón. Los ejemplos de otros promotores se describen en otra parte del presente documento. Cuando se introduce un ADN que codifica un ARNg en la célula, el ARNg puede expresarse de forma transitoria, condicional o constitutiva en la célula.

Alternativamente, los ARNg pueden prepararse por varios otros métodos. Por ejemplo, los ARNg se pueden preparar mediante transcripción *in vitro* usando, por ejemplo, ARN polimerasa de T7 (véanse, por ejemplo, los documentos WO 2014/089290 y WO 2014/065596). Los ARN guía también pueden ser una molécula producida sintéticamente preparada por síntesis química.

45 C. Secuencias de reconocimiento de ARN de CRISPR

El término "secuencia de reconocimiento de ARN de CRISPR" incluye secuencias de ácido nucleico presentes en un ADN objetivo al que se unirá un segmento de direccionamiento de ADN de un ARNg, siempre y cuando existan condiciones suficientes para la unión. Por ejemplo, las secuencias de reconocimiento de ARN de CRISPR incluyen secuencias para las que se diseña un ARN guía para tener complementariedad, donde la hibridación entre una secuencia de reconocimiento de ARN de CRISPR y una secuencia de direccionamiento de ADN promueve la formación de un complejo de CRISPR. No se requiere necesariamente la complementariedad total, siempre que haya una complementariedad suficiente para causar la hibridación y promueva la formación de un complejo de CRISPR. Las secuencias de reconocimiento de ARN de CRISPR también incluyen sitios de escisión para proteínas Cas, que se describen en más detalle a continuación. Una secuencia de reconocimiento de ARN de CRISPR puede comprender cualquier polinucleótido, que puede localizarse, por ejemplo, en el núcleo o citoplasma de una célula o dentro de un orgánulo de una célula, tal como una mitocondria o cloroplasto.

La secuencia de reconocimiento de ARN de CRISPR dentro de un ADN objetivo puede ser dirigida por (es decir, estar unida a, o hibridar con, o ser complementaria a) una proteína Cas o un ARNg. Las condiciones de unión adecuadas de ADN/ARN incluyen condiciones fisiológicas normalmente presentes en una célula. Se conocen en la técnica otras condiciones adecuadas de unión de ADN/ARN (por ejemplo, condiciones en un sistema sin células) (véase, por ejemplo,

Molecular Cloning: A Laboratory Manual, 3ª edición (Sambrook et al., Harbor Laboratory Press 2001)). La cadena del ADN objetivo que es complementaria e híbrida con la proteína Cas o el ARNg puede denominarse "cadena complementaria" y la cadena del ADN objetivo que es complementaria a la "cadena complementaria" (y por lo tanto no es complementaria a la proteína Cas o ARNg) se puede llamar "cadena no complementaria" o "cadena de plantilla".

5 La proteína Cas puede escindir el ácido nucleico en un sitio dentro o fuera de la secuencia de ácido nucleico presente en el ADN objetivo al que se unirá el segmento de direccionamiento de ADN de un ARNg. El "sitio de escisión" incluye la posición de un ácido nucleico al que una proteína Cas produce un rompimiento de una sola cadena o un rompimiento de cadena doble. Por ejemplo, la formación de un complejo de CRISPR (que comprende un ARNg hibridado con una secuencia de reconocimiento de ARN de CRISPR y complejo con una proteína Cas) puede resultar en la escisión de una o ambas cadenas en o cerca (p. ej., dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 o más pares de bases de) la secuencia de ácido nucleico presente en un ADN objetivo al que se unirá un segmento de direccionamiento de ADN de un ARNg. Si el sitio de escisión está fuera de la secuencia de ácido nucleico a la que se unirá el segmento de direccionamiento de ADN del ARNg, el sitio de escisión todavía se considera dentro de la "secuencia de reconocimiento de ARN de CRISPR". El sitio de escisión puede estar solo en una cadena o en ambas cadenas de un ácido nucleico. 10 Los sitios de escisión pueden estar en la misma posición en ambas cadenas del ácido nucleico (produciendo extremos romos) o pueden estar en sitios diferentes en cada cadena (produciendo extremos escalonados). Los extremos escalonados se pueden producir, por ejemplo, usando dos proteínas Cas, cada una de las cuales produce un rompimiento de una sola cadena en un sitio de escisión diferente en cada cadena, produciendo así un rompimiento de cadena doble. Por ejemplo, una primera nickasa puede crear un rompimiento de cadena sencilla en la primera cadena de ADN de doble cadena (ADNdc), y una segunda nickasa puede crear un rompimiento de cadena sencilla en la segunda cadena de ADNdc de manera que se creen las secuencias sobresalientes. En algunos casos, la secuencia de reconocimiento de ARN de CRISPR de la nickasa en la primera cadena se separa de la secuencia de reconocimiento de ARN de CRISPR de la nickasa en la segunda cadena en al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 o 1.000 pares de bases.

25 La escisión específica del sitio del ADN objetivo por Cas9 puede ocurrir en lugares determinados por (i) complementariedad de apareamiento de bases entre el ARNg y el ADN objetivo y (ii) un motivo corto, llamado motivo adyacente protoespaciador (PAM), en el ADN objetivo. El PAM puede flanquear la secuencia de reconocimiento de ARN de CRISPR. Opcionalmente, la secuencia de reconocimiento de ARN de CRISPR puede estar flanqueada por el PAM. Por ejemplo, el sitio de escisión de Cas9 puede ser de aproximadamente 1 a aproximadamente 10 o de aproximadamente 2 a aproximadamente 5 pares de bases (por ejemplo, 3 pares de bases) secuencia arriba o secuencia abajo de la secuencia PAM. En algunos casos (por ejemplo, cuando se usa Cas9 de *S. piogenes* o una Cas9 estrechamente relacionada), la secuencia del PAM de la cadena no complementaria puede ser 5'-N₁GG-3', donde N₁ es cualquier nucleótido de ADN y es inmediatamente 3' de la secuencia de reconocimiento de ARN de CRISPR de la cadena no complementaria del ADN objetivo. Como tal, la secuencia del PAM de la cadena complementaria sería 5'-CC N₂-3', donde N₂ es cualquier nucleótido de ADN y está inmediatamente 5' de la secuencia de reconocimiento de ARN de CRISPR de la cadena complementaria del ADN objetivo. En algunos casos, N₁ y N₂ pueden ser complementarios y el par de bases N₁-N₂ puede ser cualquier par de bases (p. ej., N₁ = C y N₂ = G; N₁ = G y N₂ = C; N₁ = A y N₂ = T, N₁ = T y N₂ = A).

40 Los ejemplos de secuencias de reconocimiento de ARN de CRISPR incluyen una secuencia de ADN complementaria al segmento de direccionamiento de ADN de un ARNg, o tal secuencia de ADN además de una secuencia de PAM. Por ejemplo, el motivo objetivo puede ser una secuencia de ADN de 20 nucleótidos que precede inmediatamente a un motivo NGG reconocido por una proteína Cas, tal como GN₁₉NGG (SEQ ID NO: 8) o N₂₀NGG (SEQ ID NO: 24) (véase, por ejemplo, el documento WO 2014/165825). La guanina en el extremo 5' puede facilitar la transcripción por la ARN polimerasa en las células. Otros ejemplos de secuencias de reconocimiento de ARN de CRISPR pueden incluir dos nucleótidos de guanina en el extremo 5' (por ejemplo, GGN₂₀NGG, SEQ ID NO: 25) para facilitar la transcripción eficaz mediante polimerasa de T7 *in vitro*. Véase, por ejemplo, el documento WO 2014/065596. Otras secuencias de reconocimiento de ARN de CRISPR pueden tener entre 4-22 nucleótidos de longitud de las SEQ ID NOS: 8, 24 y 25, que incluyen 5' G o GG y el 3' GG o NGG. Aún otras secuencias de reconocimiento de ARN de CRISPR pueden tener entre 14 y 20 nucleótidos de longitud de las SEQ ID NOS: 8, 24 y 25.

50 La secuencia de reconocimiento de ARN de CRISPR puede ser cualquier secuencia de ácido nucleico endógena o exógena a una célula. La secuencia de reconocimiento de ARN de CRISPR puede ser una secuencia que codifica un producto génico (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, una secuencia reguladora) o puede incluir ambas.

55 En una realización, la proteína Cas es una proteína Cas de tipo I. En una realización, la proteína Cas es una proteína Cas de tipo II. En una realización, la proteína Cas de tipo II es Cas9. En una realización, la primera secuencia de ácido nucleico codifica una proteína Cas humana de codón optimizado.

60 En una realización, el ARNg comprende una secuencia de ácido nucleico que codifica un ARNcr y un ARNcrtra. En realizaciones específicas, la proteína Cas es Cas9. En algunas realizaciones, el ARNg comprende (a) el ARN quimérico de la secuencia de ácido nucleico 5'-GUUUUAGAGCUAGAAUAGCAAGUUAAAUAAGGCUAGUCCGUUAUCAA CUUGAAAAGUGGCACCGAGUCGGUGCUUUU-3' (SEQ ID NO: 1); o, (b) el ARN quimérico de la secuencia de ácido nucleico 5'-GUUUUAGAGCUAGAAUAGCAAGUUAAAUAAGGCUAGUCCG-3' (SEQ ID NO: 2). En otra realización el

ARNcr comprende 5'-GUUUUAGAGCUAGAAUAGCAAGUUAAAAU-3' (SEQ ID NO: 3); 5'-GUUUUAGAGCUAGAAUA GCAAGUUAAAAUAG-3' (SEQ ID NO: 4); o 5'-GAGUCCGAGCAGAAGAAGUUUUUA-3' (SEQ ID NO: 5). En aún otras realizaciones, el ARNcrtra comprende, 5'-AAGGCUAGUCCG-3' (SEQ ID NO: 6) o 5'-AAGGCUAGUCCGUUAUCA CUUGAAAAAGUGGCACCGAGUCGUGCUUUU-3' (SEQ ID NO: 7).

5 V. Montaje de polinucleótidos

Los métodos divulgados en la presente pueden ensamblar al menos dos ácidos nucleicos bajo condiciones efectivas para unir las moléculas de ADN para formar una molécula de ADN de cadena doble sustancialmente intacta o sin problemas. Cualquier ácido nucleico de interés que tenga secuencias solapantes se puede ensamblar de acuerdo con los métodos divulgados en este documento. Por ejemplo, cualquier molécula de ADN de interés que tenga secuencias solapantes se puede ensamblar, incluyendo los ADN que son moléculas de ADN clonadas, naturales, los ADN generados sintéticamente, etc. Las moléculas de ADN unidas pueden, si se desea, ser clonadas (por ejemplo, insertarse) en una vector utilizando un método de la invención. El montaje de dos ácidos nucleicos incluye cualquier método de unión de cadenas de dos ácidos nucleicos. Por ejemplo, el montaje incluye la unión de ácidos nucleicos digeridos de manera que las cadenas de cada ácido nucleico se unan a la otra y la extensión, en la que cada cadena sirve como plantilla para la extensión de la otra.

En algunas realizaciones, los ácidos nucleicos se ensamblan con un oligo de unión de tal manera que cada ácido nucleico se ensambla con el oligo de unión en lugar de ensamblarse directamente entre sí. El ensamblaje con un oligo de unión puede posicionar las bases de ácidos nucleicos entre los ácidos nucleicos que se ensamblan y que no forman parte de los ácidos nucleicos que se ensamblan, pero que son parte del oligo de unión. Por lo tanto, los ácidos nucleicos pueden ensamblarse con éxito incluso si quedan bases adicionales entre los ácidos nucleicos. Alternativamente, se puede usar un oligo de unión para un ensamblaje sin problemas, en el que no quedan bases adicionales entre los ácidos nucleicos a ensamblar.

En algunas realizaciones, los ácidos nucleicos pueden prepararse para el ensamblaje por escisión con una proteína Cas, una enzima de restricción (endonucleasa de restricción) (por ejemplo, cualquiera de las diversas endonucleasas de restricción proporcionadas en otra parte de la presente memoria), una meganucleasa (por ejemplo, cualquiera de las diversas meganucleasas proporcionadas en otra parte de la presente memoria), o cualquier combinación de las mismas. Por ejemplo, uno de los ácidos nucleicos a ensamblar se puede escindir con una proteína Cas y otro ácido nucleico a ensamblar se puede escindir con una proteína Cas, una enzima de restricción, una meganucleasa o cualquier combinación de las mismas. Después de la escisión con una nucleasa, el ácido nucleico digerido se puede ensamblar directamente a otro ácido nucleico digerido que tiene secuencias finales superpuestas o se ensambla en un ácido nucleico que no ha sido digerido, pero tiene secuencias finales solapantes. El ácido nucleico digerido también se puede ensamblar a otro ácido nucleico usando un oligo de unión.

En realizaciones que emplean un agente de nucleasa (por ejemplo, una proteína Cas) para producir secuencias terminales solapantes entre dos moléculas de ácido nucleico, se pueden usar métodos combinatorios rápidos para ensamblar los ácidos nucleicos digeridos. Por ejemplo, un primer y un segundo ácido nucleico que tienen extremos solapantes se pueden combinar con una ligasa, exonucleasa, ADN polimerasa y nucleótidos y se pueden incubar a temperatura constante, tal como a 50°C. Específicamente, se podría usar una exonucleasa T5 para eliminar nucleótidos de los extremos 5' del ADNdc produciendo salientes complementarias. Las salientes complementarias de ADN de cadena sencilla se pueden hibridar, la ADN polimerasa se usa para el relleno de espacios, y la tag ADN ligasa se usa para sellar las muescas resultantes a 50°C. Por lo tanto, dos ácidos nucleicos que comparten secuencias finales superpuestas se pueden unir en una molécula sellada covalentemente en una reacción isotérmica de una etapa. Véase, por ejemplo, Gibson, et al., (2009) Nature Methods 6 (5): 343-345. En algunas realizaciones, la purificación de proteinasa K o fenol/cloroformo/alcohol isoamílico (PCI) se usa para eliminar el agente de nucleasa (por ejemplo, proteína Cas) de la mezcla de reacción. En algunas realizaciones, el agente de nucleasa (por ejemplo, proteína Cas) puede eliminarse de la mezcla de reacción mediante purificación en columna a base de gel de sílice.

En ciertas realizaciones, los métodos divulgados en este documento ensamblan un vector con un polinucleótido lineal. En otras realizaciones, los métodos divulgados en la presente memoria ensamblan al menos dos vectores, tales como dos vectores BAC. El término "vector BAC" incluye cualquier cromosoma artificial bacteriano. En realizaciones específicas, el BAC se modifica para contener una región con una secuencia de nucleótidos que se solapa con la secuencia de nucleótidos de la región de un ácido nucleico lineal u otro vector, por ejemplo, otro BAC.

El primer y segundo ácidos nucleicos de cadena sencilla tienen extremos solapantes cuando los extremos respectivos son complementarios entre sí. El primer y segundo ácidos nucleicos de doble cadena tienen extremos solapantes cuando un extremo 5' de una cadena del primer ácido nucleico es complementario al extremo 3' de una cadena del segundo ácido nucleico y viceversa. Por ejemplo, para secuencias finales solapantes de doble cadena, las cadenas de un ácido nucleico pueden tener al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o 100% de identidad con una cadena correspondiente del otro ácido nucleico. En los métodos divulgados en este documento, el extremo 5' de una cadena de una molécula de ADN de doble cadena que se va a ensamblar, comparte secuencias finales superpuestas con el extremo 3' de una cadena de la otra molécula de ADN de cadena doble. La cadena de la región de solapamiento puede hibridarse específicamente con su cadena complementaria cuando las regiones complementarias de las secuencias solapantes se presentan en salientes de

cadena sencilla desde los extremos 5' y 3' de los dos polinucleótidos que se van a ensamblar. En algunas realizaciones, se usa una exonucleasa para eliminar nucleótidos del extremo 5' o 3' para crear secuencias terminales sobresalientes. En algunas realizaciones, la región solapante del primer y/o segundo ácido nucleico no existe en el extremo 5' o 3' hasta después de la digestión con una proteína Cas. Es decir, la región de solapamiento puede ser una región interna que posteriormente se convierte en una secuencia de extremo solapante después de la digestión de los ácidos nucleicos que contienen la región de solapamiento interno con una proteína Cas. La proteína Cas puede escindirse en un sitio objetivo (por ejemplo, sitio de escisión) dentro de la región de solapamiento o fuera de la región de solapamiento.

La longitud de la región de solapamiento es preferiblemente de longitud suficiente de manera que la región se produce solo una vez dentro de cualquiera de los ácidos nucleicos que se ensamblan. De esta manera, se evita que otros polinucleótidos se hibriden con las secuencias finales y el ensamblaje puede ser específico para los ácidos nucleicos objetivo. La longitud de la región de superposición puede variar desde un mínimo de aproximadamente 10 pares de bases (pb) hasta aproximadamente 300 pb o más. En general, es preferible que la longitud del solapamiento sea menor o igual que aproximadamente el tamaño del polinucleótido a combinar, pero no menos de aproximadamente 10 pb y no más de aproximadamente 1000 pb. Para la unión de 2 o 3 polinucleótidos, puede ser suficiente una superposición de aproximadamente 20-30 pb. Para más de 10 fragmentos, una superposición preferida es de aproximadamente 80 pb hasta aproximadamente 300 pb. En una realización, la región de superposición es de una longitud que permite que se genere fácilmente por métodos sintéticos, por ejemplo, aproximadamente 40 pb. En realizaciones específicas, la longitud de la región de superposición puede ser de aproximadamente 20-200 pb. Las superposiciones pueden ser de aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 o 1,000 pb de longitud. En algunas realizaciones, la longitud de la región de superposición es de 20 - 200 pb. En realizaciones específicas de los métodos divulgados en la presente memoria, se pueden ensamblar al menos dos polinucleótidos en donde se genera una región solapante en al menos uno de los polinucleótidos por contacto con un agente de nucleasa (por ejemplo, un complejo de ARNg-Cas). Por ejemplo, la digestión con endonucleasa de un primer polinucleótido puede crear secuencias que se solapan con las secuencias finales de un segundo polinucleótido, en donde las secuencias finales solapantes se ensamblan a continuación.

En los métodos divulgados en este documento, las secuencias solapantes pueden ponerse en contacto con una exonucleasa para exponer secuencias complementarias (por ejemplo, secuencias complementarias de cadena sencilla) entre las secuencias solapantes. La digestión con exonucleasa se lleva a cabo en condiciones que son eficaces para eliminar ("masticar") un número suficiente de nucleótidos para permitir la hibridación específica de las regiones de complementariedad de cadena sencilla expuestas. En general, una porción de la región de superposición o toda la región de superposición se mastica, dejando salientes que comprenden una porción de la región de superposición o toda la región de solapamiento. En algunos métodos, la digestión con exonucleasa puede llevarse a cabo mediante una polimerasa en ausencia de dNTP (por ejemplo, ADN polimerasa T5) mientras que, en otros métodos, la digestión con exonucleasa puede llevarse a cabo mediante una exonucleasa en presencia de dNTP que carece de actividad polimerasa (por ejemplo, exonucleasa III).

Se puede usar cualquiera de una variedad de exodeoxirribonucleasas específicas de cadena doble de 5' a 3' para masticar los extremos de los ácidos nucleicos en los métodos divulgados en este documento. El término "exonucleasa 5'" algunas veces se usa en la presente memoria para referirse a una exodesoxirribonucleasa 5' a 3'. Una exonucleasa "no procesiva", como se usa en el presente documento, es una exonucleasa que degrada un número limitado de nucleótidos (por ejemplo, solo unos pocos) durante cada evento de unión al ADN. La digestión con una exonucleasa 5' produce salientes de cadena sencilla 3' en las moléculas de ADN. Entre otras propiedades que son deseables para una exonucleasa 5' se encuentra que carecen de actividad de exonucleasa 3', genera extremos de fosfato 5', e inicia la degradación de los extremos tanto 5'-fosforilados como no fosforilados. También es deseable que la enzima pueda iniciar la digestión desde el extremo 5' de una molécula, si es un extremo romo, o si tiene un extremo rebajado pequeño 5' o 3'. Las exonucleasas adecuadas serán evidentes para el experto en la materia. Estas incluyen, por ejemplo, la exonucleasa del fago T5 (producto D15 del gen del fago T5), la exonucleasa del fago lambda, RecE del profago Rac, la exonucleasa VIII de *E. coli*, exonucleasa del fago T7 (producto del gen 6 del fago T7) o cualquiera de una variedad de exonucleasa 5' que están implicadas en reacciones de recombinación homóloga. En una realización de la invención, la exonucleasa es exonucleasa T5 o exonucleasa lambda. En otra realización, la exonucleasa es exonucleasa T5. En otra realización, la exonucleasa no es la exonucleasa del fago T7. Los métodos para preparar y usar exonucleasas y otras enzimas empleadas en los métodos de la invención son convencionales; y muchas están disponibles en fuentes comerciales, tales como USB Corporation, 26111 Miles Road, Cleveland, Ohio 44128, o New England Biolabs, Inc. (NEB), 240 County Road, Ipswich, Mass. 01938-2723.

Particularmente, en realizaciones en las que la región de solapamiento es muy larga, puede ser necesario solo masticar una porción de la región (por ejemplo, más de la mitad de la región de solapamiento), siempre que las salientes de cadena sencilla así generadas sean de longitud suficiente y contenido de bases para hibridar específicamente bajo las condiciones de la reacción. El término "que hibrida específicamente" incluye situaciones en las que un par particular de salientes de cadena sencilla se hibridaran preferiblemente (o exclusivamente) entre sí, en lugar de con otras salientes de cadena sencilla (por ejemplo, salientes no complementarios) que están presentes en la reacción mezcla. Por "preferencialmente" se entiende que al menos aproximadamente el 95% de las salientes se hibridará con la saliente complementaria. Una persona experta puede determinar fácilmente la longitud óptima para lograr la hibridación específica de una secuencia de interés en un conjunto dado de condiciones de reacción. En general, las regiones homólogas de solapamiento (los salientes de cadena sencilla o sus complementos) contienen secuencias idénticas. Sin

embargo, pueden usarse secuencias parcialmente idénticas, con la condición de que las salientes de cadena sencilla puedan hibridar específicamente bajo las condiciones de las reacciones.

5 En ciertas realizaciones, el agente de nucleasa (por ejemplo, una proteína Cas) puede crear rompimientos de cadena simple (es decir, "muescas") en el sitio objetivo sin cortar ambas cadenas de ADNdc. Una "nickasa" incluye un agente de nucleasa (por ejemplo, una proteína Cas) que crea muescas en el ADNdc. De esta manera, dos agentes de nucleasa separados (por ejemplo, proteínas Cas) (por ejemplo, Nickasas) específicos para un sitio objetivo en cada cadena de ADNdc pueden crear secuencias sobresalientes complementarias a secuencias sobresalientes en otro ácido nucleico o una región separada en el mismo ácido nucleico. Los extremos sobresalientes creados al poner en contacto un ácido nucleico con dos nickasas específicas para sitios objetivo en ambas cadenas de ADNdc pueden ser extremos sobresalientes de 5' o 3'. Por ejemplo, una primera nickasa puede crear un rompimiento de cadena sencilla en la primera cadena de ADNdc, mientras que una segunda nickasa puede crear un segundo rompimiento de cadena sencilla en la segunda cadena de ADNdc de modo que se creen las secuencias sobresalientes. Los sitios objetivo de cada nickasa que crean el rompimiento de cadena sencilla pueden seleccionarse de manera que las secuencias finales sobresalientes creadas sean complementarias a las secuencias terminales sobresalientes en un segundo ácido nucleico. De acuerdo con esto, los extremos que sobresalen complementarios del primer y segundo ácido nucleico se pueden hibridar mediante los métodos divulgados en este documento. En algunas realizaciones, el sitio objetivo de la nickasa en la primera cadena es diferente del sitio objetivo de la nickasa en la segunda cadena. Diferentes sitios objetivo en cadenas separadas de ADNdc dan como resultado rompimientos de cadena separados por al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 o 1,000 pares de bases.

20 En ciertas realizaciones, el segundo ácido nucleico también se pone en contacto con una primera nickasa que crea una muesca en un primer sitio objetivo en el segundo ácido nucleico y una nickasa que crea una muesca en un segundo sitio objetivo en la segunda molécula de ácido nucleico. Las secuencias extremas sobresalientes creadas por las muescas en dos sitios diferentes en el segundo ácido nucleico pueden ser complementarias a las secuencias terminales sobresalientes creadas por muescas en dos sitios diferentes en el primer ácido nucleico de modo que las secuencias finales sobresalientes complementarias se hibridan.

En algunas realizaciones, la secuencia de ácido nucleico de un gen de interés abarca dos o más BAC. En tales casos, usando los métodos proporcionados en este documento, los agentes de nucleasa diseñados específicamente pueden cortar los dos o más BAC en las ubicaciones deseadas y los fragmentos de ácido nucleico resultantes se unen para formar la secuencia del gen de interés.

30 En algunas realizaciones, los extremos sobresalientes creados por muescas en diferentes sitios objetivo en ambas cadenas de un primer ácido nucleico no son complementarios a los extremos sobresalientes creados por muescas en diferentes sitios objetivo en ambas cadenas de un segundo ácido nucleico. En otras realizaciones, los ácidos nucleicos a ensamblar no tienen extremos complementarios de manera que sea necesario un ácido nucleico separado para ensamblar los extremos no complementarios. Un oligo de unión se puede usar para unir los extremos no complementarios de dos ácidos nucleicos. Un "oligo de unión" incluye brazos complementarios que incluyen un polinucleótido o ácido nucleico que tiene una secuencia complementaria a los extremos de un polinucleótido o ácido nucleico diferente. En algunas realizaciones, un oligo de unión tiene un brazo complementario a un primer ácido nucleico en el extremo 5', una parte central (espaciadora) y un brazo complementario a un segundo ácido nucleico en el extremo 3'. De este modo, los ácidos nucleicos que tienen secuencias terminales no complementarias entre sí pueden ensamblarse hibridando cada ácido nucleico con el mismo oligo de unión después de un tratamiento con exonucleasa. En realizaciones específicas, el oligo de unión tiene un primer brazo complementario a la secuencia del extremo 5' o 3' de un primer ácido nucleico digerido y un segundo brazo complementario a la secuencia 5' o 3' de un segundo ácido nucleico digerido. El oligo de unión puede unir secuencias finales no complementarias que son romas o tienen una secuencia sobresaliente 5' o 3'.

45 La longitud de las secuencias del brazo complementarias del oligo de unión deben ser suficientes para hibridar con los ácidos nucleicos que se ensamblarán después del tratamiento con exonucleasa. Por ejemplo, la longitud de las secuencias del brazo complementario del oligo de unión puede ser de al menos aproximadamente 10, 20, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 110, 120, 130, 140, 150 pb o más. En realizaciones específicas, el brazo complementario es 15-120 pb, 20-100 pb, 30-90 pb, 30-60 pb, o 20-80 pb. En una realización específica, la longitud de las secuencias del brazo complementario del oligo de unión es de 40 pb. Cada brazo complementario de un oligo de unión puede ser de diferentes longitudes. El espaciador del oligo de unión, entre las secuencias finales complementarias a los ácidos nucleicos a ensamblar, puede ser de al menos aproximadamente 20 pb, 30 pb, 35 pb, 40 pb, 45 pb, 50 pb, 55 pb, 60 pb, 65 pb, 70 pb, 75 pb, 80 pb, 90 pb, 100 pb, 250 pb, 500 pb, 750 pb, 1000 pb, 2000 pb, 3000 pb, 4000 pb, 5000 pb, 8000 pb, 10 kb, 15 kb, 20 kb, o más. Por ejemplo, el espaciador de un oligo de unión puede incluir un vector BAC o LTVEC. En algunas realizaciones, el espaciador del oligo de unión puede diseñarse para tener secuencias específicas para detección o secuencias adecuadas para PCR para confirmar el ensamblaje exitoso. En algunas realizaciones, el espaciador del oligo de unión puede diseñarse para introducir uno o más sitios de las enzimas de restricción. En algunas realizaciones, el espacio del oligo de unión puede diseñarse para introducir un gen de resistencia a fármacos o un gen informador. En otras realizaciones, el espaciador puede contener al menos 20 pb de una porción del extremo de un ácido nucleico a ensamblar con el fin de ensamblar sin problemas los ácidos nucleicos. Por ejemplo, para un ensamblaje sin problemas, el espaciador puede ser de aproximadamente 45 bp.

En algunas realizaciones, la relación molar del ácido nucleico con respecto al oligo u oligos de unión puede ser de aproximadamente 1:1 a aproximadamente 1:200. En algunas realizaciones, la relación molar del ácido nucleico con respecto al oligo u oligos de unión es aproximadamente 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:30, 1:40, 1:50, 1:60, 1:70, 1:80, 1:90, 1:100, 1:120, 1:140, 1:160, 1:180 o 1:200. En realizaciones específicas, la relación molar del ácido nucleico con respecto al oligo u oligos de unión puede ser de aproximadamente 1:6 a aproximadamente 1:20. En una realización, la relación molar es aproximadamente 1:6. En otra realización, la relación molar es aproximadamente 1:20.

En realizaciones específicas, se usa un oligo de unión para ensamblar sin problemas al menos dos ácidos nucleicos. Ensamblaje "sin problemas" se refiere al ensamblaje de dos ácidos nucleicos en el que no están presentes bases de ácido nucleico intervinientes entre los extremos adyacentes de los ácidos nucleicos a ensamblar. Por ejemplo, los ácidos nucleicos ensamblados sin problemas no tienen presentes bases de ácidos nucleicos que no sean parte de los ácidos nucleicos que se ensamblarán. Con el fin de ensamblar sin problemas dos ácidos nucleicos, el espaciador de un oligo de unión debe incluir una secuencia de ácido nucleico idéntica a una porción del extremo ya sea del primer o segundo ácido nucleico que se va a ensamblar. Esta porción del extremo se debe eliminar del ácido nucleico antes de ensamblarse con el oligo de unión. Por ejemplo, la porción final puede escindirse mediante un agente de nucleasa (por ejemplo, un complejo de ARNg-Cas) al menos 20 pb desde el extremo del ácido nucleico, tal como al menos 40 pb o al menos 45 pb desde el extremo del ácido nucleico. Alternativamente, la porción del extremo se puede escindir mediante un agente de nucleasa (por ejemplo, un complejo de ARNg-Cas) al menos 2, al menos 4, al menos 6, al menos 8, al menos 10, al menos 12, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 37, al menos 40, al menos 42, al menos 45, al menos 48, al menos 50, al menos 55, al menos 60, al menos 65, al menos 70, al menos 80, al menos 100, al menos 110, al menos 120, al menos 130, al menos 140, al menos 150 bp desde el extremo del ácido nucleico a ensamblar.

En una realización, el oligo de unión puede comprender desde el extremo 5' hasta el extremo 3': aproximadamente una superposición de 15-120 pb con el ácido nucleico 5', aproximadamente 20-50 pb de una región del extremo 3' del ácido nucleico 5', y una superposición de aproximadamente 15-120 pb con el ácido nucleico 3'. En una realización, el oligo de unión puede comprender desde el extremo 5' hasta el extremo 3': aproximadamente una superposición de 15-120 pb con el ácido nucleico 5', aproximadamente 20-50 pb de una región del extremo 5' del ácido nucleico 3', y una superposición de aproximadamente 15-120 pb con el ácido nucleico 3'. Por lo tanto, cuando el oligo de unión se ensambla al primer y segundo ácido nucleico, el espaciador del oligo de unión reconstituye la sección eliminada del ácido nucleico antes del ensamblaje. Véase, Fig. 5 y Fig. 6. El término "reconstituye" incluye el reemplazo de la porción de extremo del ácido nucleico que se escindió para proporcionar un ácido nucleico ensamblado completo cuando se ensambla al oligo de unión. Por ejemplo, la reconstitución del ácido nucleico escindido reemplaza la porción escindida del ácido nucleico con un ácido nucleico incluido en el espaciador del oligo de unión que tiene la secuencia idéntica a la de la porción escindida.

El oligo de unión se puede ensamblar a una primera y una segunda molécula de ácido nucleico simultáneamente o secuencialmente. Cuando se ensamblan simultáneamente, el oligo de unión puede ponerse en contacto con un primer y segundo ácido nucleico en la misma mezcla de reacción de manera que el ácido nucleico ensamblado resultante comprenda el primer ácido nucleico, el oligo de unión y el segundo ácido nucleico. Cuando se ensamblan secuencialmente, el oligo de unión se pone en contacto con el primer ácido nucleico en una reacción de ensamblaje que produce un ácido nucleico ensamblado que comprende el primer ácido nucleico ensamblado con el oligo de unión, pero no el segundo ácido nucleico. Tal ácido nucleico ensamblado puede ponerse en contacto luego con el segundo ácido nucleico en una reacción de ensamblaje separada que produce un ácido nucleico ensamblado que comprende el primer ácido nucleico, el oligo de unión y el segundo ácido nucleico. En otras realizaciones, el oligo de unión se pone en contacto con el segundo ácido nucleico en una reacción de ensamblaje que produce un ácido nucleico ensamblado que comprende el segundo ácido nucleico ensamblado con el oligo de unión, pero no el primer ácido nucleico. Tal ácido nucleico ensamblado se puede poner en contacto entonces con el primer ácido nucleico en una reacción de ensamblaje separada que produce un ácido nucleico ensamblado que comprende el primer ácido nucleico, el oligo de unión y el segundo ácido nucleico.

Cualquier cantidad de oligos de unión se puede usar en los métodos de la presente invención para ensamblar moléculas de ácido nucleico. Por ejemplo, un oligo de unión se puede usar para ensamblar 2 moléculas de ácido nucleico, 2 oligos de unión se pueden usar para ensamblar 3 moléculas de ácido nucleico, 3 oligos de unión se pueden usar para ensamblar 4 moléculas de ácido nucleico, 4 oligos de unión se pueden usar para ensamblar 5 moléculas de ácido nucleico o 5 oligos de unión se pueden usar para ensamblar 6 moléculas de ácido nucleico. El número de oligos de unión puede ser de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más, dependiendo del número de moléculas de ácido nucleico que se van a ensamblar.

En algunas realizaciones, el oligo de unión comprende un ADN de Bloque g. Un "Bloque g" es un fragmento de ADN de doble cadena lineal. El Bloque g puede ser de aproximadamente 50 pb a aproximadamente 2000 pb. El Bloque g puede ser de aproximadamente 50 pb a aproximadamente 100 pb, de aproximadamente 100 pb a aproximadamente 200 pb, de aproximadamente 200 pb a aproximadamente 300 pb, de aproximadamente 300 pb a aproximadamente 400 pb, de aproximadamente 400 pb a aproximadamente 500 pb, de aproximadamente 500 pb a aproximadamente 600 pb, de aproximadamente 600 pb a aproximadamente 800 pb, de aproximadamente 800 pb a aproximadamente 1000 pb, de aproximadamente 1000 pb a aproximadamente 1250 pb, de aproximadamente 1250 pb a aproximadamente 1500 pb, de

aproximadamente 1500 pb a aproximadamente 1750 pb, o desde aproximadamente 1750 pb a aproximadamente 2000 pb.

5 El ensamblaje de dos o más ácidos nucleicos con un Bloque g puede cribarse, por ejemplo, mediante ensayos de PCR descritos en otra parte de este documento (por ejemplo, el Ejemplo 10). En algunos casos, el Bloque g no incluye un casete de selección. Tal método permite la unión rápida de dos o más moléculas de ácido nucleico que pueden cribarse mediante un ensayo de PCR simple. El Bloque g puede comprender cualquier secuencia de ácido nucleico de interés. En algunos casos, el Bloque g puede comprender un sitio objetivo para un agente de nucleasa o un sitio objetivo para cualquiera de las diversas meganucleasas o enzimas de restricción proporcionadas en la presente memoria. En otras realizaciones, un Bloque g puede comprender un casete de selección. En algunas realizaciones, el Bloque g comprende una secuencia de ADN de interés. En una realización, el Bloque g comprende una secuencia de ADN humano.

10 Los ácidos nucleicos que se van a ensamblar o cualquiera de los diversos oligos de unión también pueden comprender un casete de selección o un gen informador. El casete de selección puede comprender una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. El promotor puede ser activo en una célula procariota de interés y/o activo en una célula eucariótica de interés. Dichos promotores pueden ser un promotor inducible, un promotor que es endógeno al gen informador o la célula, un promotor que es heterólogo al gen informador o a la célula, un promotor específico de la célula, un promotor específico del tejido o un promotor específico de la etapa de desarrollo. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (*neo^r*), higromicina B fosfotransferasa (*hyg^r*), puromicina-N-acetiltransferasa (*puro^r*), blasticidina S desaminasa (*bsr^r*), xantina/guanina fosforribosil transferasa (*gpt*), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los mismos. El marcador de selección del vector de direccionamiento puede estar flanqueado por los brazos de homología secuencia arriba y secuencia abajo o puede encontrarse ya sea 5' o 3' con los brazos de homología.

15 En una realización, los ácidos nucleicos a ensamblar o cualquiera de los diversos oligos de unión comprenden un gen informador operativamente unido a un promotor, en el que el gen informador codifica una proteína informadora seleccionada del grupo que consiste en LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína fluorescente verde mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina y una combinación de los mismos. Dichos genes informadores se pueden unir operativamente a un promotor activo en la célula. Dichos promotores pueden ser un promotor inducible, un promotor que es endógeno al gen informador o la célula, un promotor que es heterólogo al gen informador o a la célula, un promotor específico de la célula, una forma de promotor específico del tejido o un promotor específico de la etapa de desarrollo.

20 Después de la hibridación del ADN de cadena sencilla (por ejemplo, salientes producidos por la acción de exonucleasa cuando las moléculas de ADN a unir son ADNdc o salientes producidas por la creación de muescas en diferentes sitios objetivo en cada cadena), los huecos de cadena sencilla dejados por la exonucleasa se rellenan con una ADN polimerasa adecuada, que no desplaza la cadena, y las muescas así formadas se sellan con una ligasa. Una "ADN polimerasa que no desplaza la cadena", como se usa en el presente documento, es una ADN polimerasa que termina la síntesis de ADN cuando encuentra cadenas de ADN que se encuentran en su camino a medida que procede a copiar una molécula de ADNdc o degrada las cadenas de ADN encontradas a medida que avanza mientras llena al mismo tiempo el espacio así creado, generando así una "muesca en movimiento" (traducción de muesca).

25 En algunas realizaciones, las secuencias finales de solapamiento tienen suficiente complementariedad entre las regiones solapantes para hibridar los extremos complementarios de cadena sencilla de cada polinucleótido. Después de hibridación de una cadena sencilla de un primer polinucleótido con la cadena complementaria de un segundo polinucleótido, el extremo 3' del primer polinucleótido puede extenderse con base en la plantilla de la segunda cadena de polinucleótido y el extremo 3' de la segunda cadena de polinucleótido puede extenderse con base en la plantilla de la primera cadena del polinucleótido. Extendiendo el extremo 3' complementario de cada polinucleótido, se pueden ensamblar los polinucleótidos. Después del ensamblaje, las muescas entre el extremo 3' extendido de una cadena de un fragmento y un extremo 5' adyacente de una cadena del otro fragmento se pueden sellar mediante ligación. Más específicamente, el grupo hidroxilo del extremo 3' extendido del primer polinucleótido con el grupo fosfato del extremo 5' del segundo polinucleótido y ligación del grupo hidroxilo del extremo 3' extendido del segundo polinucleótido al grupo fosfato del extremo 5' del primer polinucleótido.

30 La reacción de ligación se puede realizar mediante cualquiera de una variedad de ADN ligasas termoestables adecuadas. Entre las ligasas adecuadas se encuentran, por ejemplo, Taq ligasa, ADN ligasa ampligasa termoestable (Epicenter Biotechnologies), las ligasas termoestables descritas en la patente de EE.UU. No. 6.576.453, la ADN ligasa termoestable TFS de Bioneer, Inc.

35 Una cantidad adecuada de un agente de agrupamiento, tal como PEG, en la mezcla de reacción permite, mejora o facilita el agrupamiento molecular. Sin desear estar limitado por ningún mecanismo en particular, se sugiere que un agente de agrupamiento, que permite el agrupamiento molecular y une y enlaza agua en una solución, permitiendo que los componentes de la solución entren en contacto más cercano entre sí. Por ejemplo, las moléculas de ADN que se recombinarán pueden acercarse más; lo que facilita la hibridación de las salientes de cadena sencilla. También, se sugiere que las enzimas pueden entrar en contacto más cercano con sus sustratos de ADN y pueden estabilizarse

mediante la eliminación de las moléculas de agua. Una cantidad adecuada de un agente de agrupamiento adecuado será evidente para el experto en la materia. Estos incluyen una variedad de macromoléculas bien conocidas, tales como polímeros, por ejemplo, polietilenglicol (PEG); Ficoll, tal como Ficoll 70; dextrano, tal como dextrano 70; o similar. Gran parte de la discusión en esta aplicación está dirigida a PEG. Sin embargo, la discusión también debe aplicarse a otros agentes de agrupamiento adecuados. Un experto en la materia reconocerá cómo implementar cambios de rutina en el método para acomodar el uso de otros agentes de agrupamiento.

Una cantidad adecuada de un agente de agrupamiento, tal como PEG, en la mezcla de reacción permite, mejora o facilita el agrupamiento molecular. Por ejemplo, los agentes de agrupamiento pueden ayudar a que las moléculas de ADN que se recombinan puedan acercarse más; esto facilita por lo tanto la hibridación de las salientes de cadena sencilla. También, se sugiere que las enzimas pueden entrar en contacto más cercano con sus sustratos de ADN y pueden estabilizarse mediante la eliminación de moléculas de agua. Una variedad de agentes de agrupamiento adecuados será evidente para el experto en la materia. Estos incluyen una variedad de macromoléculas bien conocidas, tales como polímeros, por ejemplo, polietilenglicol (PEG); Ficoll, tal como Ficoll 70; dextrano, tal como dextrano 70; o similares. En general, cuando se usa PEG, una concentración de aproximadamente 5% (peso/volumen) es óptima. Sin embargo, la cantidad de PEG puede variar, por ejemplo, de aproximadamente 3 a aproximadamente 7%. Se puede usar cualquier tamaño adecuado de PEG, por ejemplo, que varía desde aproximadamente PEG-200 (por ejemplo, PEG-4000, PEG-6000 o PEG-8000) hasta aproximadamente PEG-20,000, o incluso superior. En los ejemplos de este documento, se usó PEG-8000. El agente de agrupamiento puede, además de mejorar la reacción de hibridación, mejorar la ligación.

Los componentes de la reacción (tales como sales, reguladores, una fuente de energía adecuada (tal como ATP o NAD), pH de la mezcla de reacción, etc.) que están presentes en una mezcla de reacción de ensamblaje pueden no ser óptimos para las enzimas individuales (exonucleasa, polimerasa y ligasa); en vez de eso, sirven como un compromiso que es efectivo para todo el conjunto de reacciones. Por ejemplo, un sistema regulador adecuado identificado por los inventores, a veces denominado en la presente memoria como regulador ISO (ISOthermal) típicamente comprende Tris-Cl 0,1 M, pH 7,5; MgCl₂ 10 mM, 0,2 mM cada uno de dGTP, dATP, dTTP y dCTP, DTT 10 mM, 5% de PEG-8000 y NAD 1 mM.

En los métodos divulgados en la presente memoria, al menos dos ácidos nucleicos se ponen en contacto con una proteína Cas y otras enzimas en condiciones efectivas para ensamblar los ácidos nucleicos para formar una molécula de ADN de doble cadena ensamblada en la que se retiene una sola copia de la región solapamiento. Los métodos descritos pueden usarse para unir cualquier molécula de ADN de interés, incluyendo los ADN que son moléculas de ADN clonadas, naturales, ADN generados sintéticamente, etc. Las moléculas de ADN unidas pueden, si se desea, clonarse en un vector (por ejemplo, usando un método de la invención). En algunas realizaciones, los ácidos nucleicos que se van a ensamblar son de codón optimizado para introducción y expresión en una célula de interés (por ejemplo, una célula de roedor, célula de ratón, célula de rata, célula humana, célula de mamífero, célula microbiana, célula de levadura, etc.).

Las moléculas de ADN de cualquier longitud se pueden unir mediante métodos divulgados en este documento. Por ejemplo, se pueden unir ácidos nucleicos que tienen aproximadamente 100 pb hasta aproximadamente 750 o 1,000, o más. El número de ácidos nucleicos que se pueden ensamblar, en una o varias etapas de ensamblaje según los métodos descritos allí, puede ser al menos aproximadamente 2, 3, 4, 6, 8, 10, 15, 20, 25, 50, 100, 200, 500, 1.000, 5.000 o 10.000 moléculas de ADN, por ejemplo, en el intervalo de aproximadamente 2 hasta aproximadamente 30 ácidos nucleicos. El número de etapas de montaje puede ser de aproximadamente 2, 4, 6, 8, 10 o más. El número de moléculas ensambladas en una sola etapa puede estar en el intervalo de aproximadamente 2 a aproximadamente 10 moléculas. Los métodos de la invención pueden usarse para unir moléculas o casetes de ADN, cada uno de los cuales tiene un tamaño inicial de al menos o no más de aproximadamente 40 pb, 60 pb, 80 pb, 100 pb, 500 pb, 1 kb, 3 kb, 5 kb, 6 kb, 10 kb, 18 kb, 20 kb, 25 kb, 32 kb, 50 kb, 65 kb, 75 kb, 150 kb, 300 kb, 500 kb, 600 kb, 1 Mb o más. Los productos finales ensamblados pueden ser al menos aproximadamente de 500 pb, 1 kb, 3 kb, 5 kb, 6 kb, 10 kb, 18 kb, 20 kb, 25 kb, 32 kb, 50 kb, 65 kb, 75 kb, 150 kb, 300 kb, 500 kb, 600 kb, 1 Mb o mayor, por ejemplo, en el rango de 30 kb a 1 Mb.

En algunas realizaciones, los ácidos nucleicos ensamblados forman un círculo y/o se unen en un vector para formar un círculo. El límite inferior de tamaño para que un ADNdc se circularice es de aproximadamente 200 pares de bases. Por lo tanto, la longitud total de los fragmentos unidos (que incluye, en algunos casos, la longitud del vector) es de al menos aproximadamente 200 pb de longitud. No existe un límite práctico de tamaño superior, y se pueden generar ADN unidos de unos pocos cientos de pares de kilobases, o mayores, mediante los métodos descritos en este documento. Los ácidos nucleicos unidos pueden tomar la forma de un círculo o una molécula lineal.

Los métodos descritos en este documento pueden usarse para ensamblar un fragmento lineal con otro fragmento lineal, un fragmento lineal con una molécula circular de ácido nucleico, una molécula circular de ácido nucleico con otra molécula circular de ácido nucleico o cualquier combinación de ácidos nucleicos lineales y circulares. Un "vector" incluye cualquier molécula de ácido nucleico circular. En ciertas realizaciones, el vector ensamblado mediante los métodos divulgados en la presente memoria es un cromosoma artificial bacteriano (BAC). El vector (por ejemplo, el BAC) puede incluir un ADN humano, un ADN de roedor, un ADN sintético o cualquier combinación de los mismos. Por ejemplo, el BAC puede comprender una secuencia de polinucleótido humano. Cuando se une una mezcla de moléculas de ADN, es preferible que los ADN estén presentes en cantidades aproximadamente equimolares.

El ácido nucleico utilizado para el ensamblaje mediante los métodos divulgados en este documento puede ser un vector de direccionamiento grande. El término "vector de direccionamiento grande" o "LTVEC" incluye vectores que comprenden brazos de homología que corresponden y se derivan de secuencias de ácido nucleico usadas para direccionamiento homóloga en células y/o comprenden ácidos nucleicos insertables que comprenden secuencias de ácido nucleico destinadas a llevar a cabo recombinación homóloga en células. Por ejemplo, el LTVEC hace posible la modificación de loci grandes que no pueden ser acomodados mediante vectores tradicionales de direccionamiento a base de plásmidos debido a sus limitaciones de tamaño. En realizaciones específicas, los brazos de homología y/o el ácido nucleico insertado del LTVEC comprende la secuencia genómica de una célula eucariótica. El tamaño del LTVEC es demasiado grande para permitir el cribado de eventos de direccionamiento mediante ensayos convencionales, por ejemplo, transferencia Southern y PCR de gran intervalo (por ejemplo, 1 kb - 5 kb). Los ejemplos del LTVEC incluyen, pero no se limitan a, vectores derivados de un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o un cromosoma artificial de levadura (YAC). Los ejemplos no limitantes de LTVEC y métodos para su elaboración se describen, por ejemplo, en la patente de los EE.UU. Nos. 6.586.251, 6.596.541, 7.105.348 y el documento WO 2002/036789 (PCT/US01/45375) y el documento US 2013/0137101.

En algunas realizaciones, los casetes pueden insertarse en vectores que pueden eliminarse posteriormente. Se pueden construir diversas formas de casetes para permitir la eliminación en tipos de células o tejidos específicos, en etapas de desarrollo específicas o tras la inducción. Tales casetes pueden emplear un sistema recombinasa en el que el casete está flanqueado en ambos lados por sitios de reconocimiento de recombinasa y puede eliminarse utilizando una recombinasa expresada en el tipo de célula deseado, expresada en la etapa de desarrollo deseada, o expresada o activada tras la inducción. Dichos casetes pueden construirse adicionalmente para incluir una matriz de pares de diferentes sitios de reconocimiento de recombinasa que se colocan de manera que se puedan generar alelos nulos, condicionales o combinación de condicionales/nulos, como se describe en el documento US 2011/0104799. La regulación de los genes de recombinasa se pueden controlar de diversas maneras, tal como unión operativa un gen de recombinasa con un promotor específico de célula, específico de tejido o regulado por desarrollo (u otro elemento regulador), o uniendo operativamente un gen de recombinasa a un 3'-UTR que comprende un sitio de reconocimiento para un miARN que se transcribe solo en tipos de células particulares, tipos de tejidos o etapas de desarrollo. Una recombinasa también se puede regular, por ejemplo, empleando una proteína de fusión que coloca la recombinasa bajo el control de un efector o metabolito (por ejemplo, CreER^{T2}, cuya actividad está controlada positivamente por tamoxifeno), o colocando el gen de recombinasa bajo el control de un promotor inducible (por ejemplo, uno cuya actividad sea controlada por doxiciclina y TetR o variantes de TetR). Se proporcionan ejemplos de diversas formas de casetes y medios para regular genes de recombinasa, por ejemplo, en los documentos US 8.518.392; US 8,354,389; y US 8,697,851, cada uno de los cuales se incorpora por referencia en su totalidad.

Los vectores usados para el ensamblaje como se describe en la presente memoria (por ejemplo, LTVEC) pueden ser de cualquier longitud, incluyendo, pero sin limitarse a, de aproximadamente 20 kb a aproximadamente 400 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a 40 kb, desde de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a 125 kb, de aproximadamente 125 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 175 kb, de aproximadamente 175 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 225 kb, de aproximadamente 225 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 275 kb o de aproximadamente 275 kb a aproximadamente 300 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, de aproximadamente 350 kb a aproximadamente 400 kb, de aproximadamente 350 kb a aproximadamente 550kb. En una realización, el LTVEC es de aproximadamente 100 kb.

Los métodos proporcionados en este documento para ensamblar ácidos nucleicos pueden diseñarse para permitir una eliminación de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 150 kb, o de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500kb a aproximadamente 1Mb, de aproximadamente 1Mb a aproximadamente 1.5Mb, de aproximadamente 1.5Mb a alrededor de 2Mb, de aproximadamente 2Mb a aproximadamente 2.5Mb, o de aproximadamente 2.5Mb a aproximadamente 3Mb.

En otros casos, los métodos proporcionados en este documento están diseñados para permitir la inserción de una secuencia de ácido nucleico exógena que varía de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, desde aproximadamente 300 kb a aproximadamente 350 kb, o desde aproximadamente 350 kb a aproximadamente 400 kb. En una realización, el polinucleótido de inserción es de aproximadamente 130 kb o aproximadamente 155 kb.

Los ácidos nucleicos lineales se pueden ensamblar entre sí o con vectores mediante los métodos divulgados en este

documento. La molécula lineal puede ser un vector que ha sido digerido por una endonucleasa (por ejemplo, proteína Cas) o cualquier ácido nucleico lineal sintético, artificial o de origen natural. En ciertas realizaciones, el ácido nucleico lineal se crea de manera que las secuencias finales se solapan con una región de otro ácido nucleico. Las secuencias finales solapantes de un ácido nucleico lineal se pueden introducir mediante cualquier método conocido en la técnica para generar secuencias de ácido nucleico personalizadas. Por ejemplo, las secuencias finales pueden ser una porción de una molécula producida sintéticamente, se pueden introducir por PCR, o pueden introducirse mediante técnicas de clonación tradicionales.

Ejemplos

Los siguientes ejemplos se presentan para proporcionar a los expertos en la materia una divulgación y descripción completas de cómo hacer y usar la presente invención, y no pretenden limitar el alcance de lo que los inventores consideran su invención ni pretende representar que los experimentos a continuación son todos o los únicos experimentos realizados. Se han realizado esfuerzos para garantizar la precisión con respecto a los números utilizados (por ejemplo, cantidades, temperatura, etc.), pero se deben tener en cuenta algunos errores y desviaciones experimentales. A menos que se indique lo contrario, las partes son partes en peso, el peso molecular es el peso molecular medio ponderado, la temperatura es en grados centígrados y la presión está en o cerca de la atmosférica.

Ejemplo 1: BAC digiere con CAS9 seguido por ensamblaje con un casete de selección

Se diseñaron un ARNcr artificial y un ARNcrtra artificial para dirigir secuencias específicas en la MAID 6177 (LTVEC de 116 kb) para el ensamblaje con un producto de PCR de 3 kb (UB-HYG). El producto de PCR contenía superposiciones de 50 pb con el vector. Primero disolver los ARNcr y ARNcrtra con 100 μM en regulador dúplex (HEPES 30 mM, pH 7,5, acetato de potasio 100 mM). Con el fin de hibridar los ARN, se añaden 10 μL de ARNcr 100 μM y 10 μL de ARNcrtra 100 μM a 80 μL de regulador de hibridación. Se calientan los ARN en un bloque de temperatura a 90°C, luego se retira el bloque de calentamiento y se enfría en el banco. La concentración final de ARN es de aproximadamente 10 μM.

Con el fin de digerir el BAC, se usa ADN de BAC maxiprep limpio y el BAC se digiere de acuerdo con la siguiente mezcla.

	1X
	X μL
ADN de BAC (500 ng)	X μL
BSA (100x)	0,5 μL
ARN	2 μL (1 μL de cada híbrido crtra:ARNcr)
Cas9 (4,5 mg/mL)	1 μL
Regulador 10x	1,5 μL
H ₂ O	hasta 15 μL

Digerir durante 1 hora a 37°C y luego eliminar la sal durante 30 min. El regulador de reacción final contiene: Tris 20 mM 7,5; NaCl 100-150 mM; MgCl₂ 10 mM; DTT 1 mM; EDTA 0,1 mM; 100 μg/mL de BSA; para un volumen final de 15 μL.

Con el fin de ensamblar el BAC e insertarlo, digerir un plásmido o realizar la PCR para crear un inserto. Para las reacciones de PCR, correr una alícuota pequeña en un gel y buscar un solo producto, si el producto tiene una sola banda, entonces realice una limpieza por PCR en lugar de la extracción en gel. Se desea una relación molar de 1:1-1:6 para el BAC:inserto. Usualmente, 50 ng de la inserción purificada funcionarán. La siguiente mezcla de reacción se puede usar:

Digestión de BAC	4 μL
Inserto	1 μL
Mezcla de montaje	15 μL

ES 2 666 179 T3

Agregar el ADN y mezclar sobre hielo o directamente en una máquina de PCR a 50°C. Incubar a 50°C durante 1 hora. Agregar 0,5 µL de proteinasa K (20mg/mL) e incubar a 50°C durante 1 hora. Desalinizar durante 30 minutos y someter a electroporación 8 µL de la reacción en células DH10B. Se pueden correr 10 µL de la digestión de BAC en un gel de campo de pulsos para verificar la eficiencia de la digestión. Usar agua libre de RNasa y reguladores.

- 5 La reacción de ensamblaje se lleva a cabo de la siguiente manera: Regulador Iso-Thermal: 3 mL de Tris-HCL 1M (pH 7,5); 150 µL de MgCl₂ 2M; 60 µL 100 mM de cada uno de: dGTP, dATP, dTTP, dCTP; 300 µL de DTT 1M; 1,5 g de PEG 8000; 300 µL de NAD 100 mM. El regulador Iso-Thermal se almacena en alícuotas de 320 µL a -20°C. La mezcla maestra se prepara de la siguiente manera: 320 µL de regulador Iso-Thermal; 0,64 µL de exonucleasa T5 (concentración del patrón = 10 U/ µL); 20 µL de ADN polimerasa Phusion (concentración del patrón = 2 U/ µL); 160 µL de Taq ADN Ligasa (concentración del patrón = 40 U/µL); 699,36 µL de H₂O; mezclar todo y tomar alícuotas a 15 µL o 30 µL y almacenar a -20°C. Use 15 µL de mezcla maestra (MM) en un volumen total de 20 µL de reacción.

- 10 La secuencia de ARNtracr usada en el ejemplo es: CAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC (SEQ ID NO: 9). Este ARN de CRISPR (ARNcr) contiene: (1) aproximadamente 20 nucleótidos de ARN complementario a la secuencia objetivo y (2) una secuencia de cola (GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO: 10)) que se hibridará con el ARNcrtra.

15 Estas etapas se describen en la Fig. 1.

Ejemplo 2: unir entre sí los dos BAC de superposición: HLA-DQ humanizado + HLA-DR humanizado en el locus MHC II de ratón (H2-A/H2-E)

- 20 Se diseñaron un ARNcr artificial y un ARNcrtra artificial para dirigir secuencias específicas en el BAC de HLA-DQ humanizado para ensamblar con un BAC HLA-DR humanizado. Los vectores que contenían ~70 pb se superponen entre sí creados por escisión de Cas9 en dos sitios en cada vector (véase, Figura 2). Disuelva los ARNcr y ARNcrtra con 100 µM en regulador Hybe. Para hibridar los ARN, agregar 10 µL de ARNcr 100 µM y 10 µL de ARNcrtra 100 µM a 80 µL de regulador de hibridación. Colocar los ARN en un bloque de calor a 90°C, luego retirar el bloque del calentador y enfriar en el banco. La concentración final de ARN es de aproximadamente 10 µM.

- 25 Con el fin de digerir el BAC, se puede usar ADN de BAC maxiprep limpio. Cada BAC se puede digerir individualmente de acuerdo con la siguiente mezcla:

ADN de BAC 2,5 ug	X µL
BSA (100x)	0,5 µL
ARN	4ul (2 µL de cada híbrido crtra:ARNcr)
Cas9 (4,5 mg/mL)	1 µL
Regulador 10x	5 µL
H ₂ O	hasta 50 µL

- 30 Los vectores de BAC se deben digerir a 37°C durante 1 hora y luego inactivados por calor durante 20 min a 65°C. Desalinizar durante 30 min. El ADN digerido se purificó por extracción con fenol/cloroformo/alcohol isoamílico (PCI) y luego se resuspendió en 35 µL de regulador TE.

Con el fin de ensamblar los vectores, utilizar 2,5 µL de los BAC para la reacción de ensamblaje de la siguiente manera:

BAC digeridos	5 µL (total)
Mezcla de montaje	15 µL

- 35 Añadir el ADN y mezclar sobre hielo o directamente en una máquina de PCR a 50°C. Incubar a 50°C durante 1 hora. Desalinizar durante 30 minutos y someter a electroporación 8 µL del ADN ensamblado en células DH10B. Usar agua

libre de RNasa y reguladores.

La secuencia de ARNcrtra usada en el ejemplo es: CAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC (SEQ ID NO: 9). Este ARN de CRISPR (ARNcr) contiene: (1) aproximadamente 20 nucleótidos de ARN complementario a una secuencia objetivo y (2) una secuencia de cola (GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO: 10)) que se hibridará con el ARNcrtra.

Estas etapas se describen en la Fig. 2.

Ejemplo 3: Ensamblaje de 2 fragmentos escindidos de Cas9 de 2 diferentes plásmidos usando enlazadores

Con el fin de construir un vector de direccionamiento, se escindió pMJ8502x con 2 ARNcr idénticos para descartar un fragmento de 400 pb y una cadena principal de Amp de 2283 pb (Figura 7). Se usaron columnas Qiagen para purificar toda la reacción. A continuación, se escindió R6KZenUbiNeo con 2 diferentes ARNcr para separación en resistencia Neo (1086 pb) y cadena principal (5390 pb). Se usaron columnas de Qiagen para purificar toda la reacción. (Figura 7). Reacción de escisión: 1170 ng de ADN, 30 µL de regulador, 4 µL de ARN hibridado (@ 100 µM), 1,7 µL de Cas9 (@ 0,89 ng/µL), H₂O hasta 60 µL. La mezcla se incubó a 37°C durante 1 hora y se purificó en una columna Qiagen antes de eluir en 30 µL de regulador de elución.

Los fragmentos escindidos se ensamblaron luego con dos enlazadores para dar como resultado un montaje sin problemas de acuerdo con la siguiente mezcla de reacción: 0,5 µL de enlazador 1 (5 ng), 0,5 µL de enlazador 2 (5 ng), 2 µL de escisión de Neo (~60 ng), 2 µL escisión de Amp (~60 ng), 15 µL de mezcla maestra de ensamblaje. La mezcla se incubó a 50°C durante 1 hora, y la reacción se dializó frente a H₂O. Se sometieron a electroporación 10 µL de la reacción en células Pir electrocompetentes antes de sembrar en placa sobre placas Carb/Kan. La PCR a través de la unión mostró que 6/8 colonias seleccionadas eran correctas y se confirmaron mediante secuenciación.

Ejemplo 4: Reemplazo de una porción de un BAC con un casete usando enlazadores

Con el fin de construir un vector de direccionamiento de ratón inactivado, se sustituyeron 40 kb de un vector de direccionamiento de BAC objetivo con un casete de selección flanqueado por sitios de reconocimiento de recombinación (Figura 8). Se diseñaron 2 enlazadores para eliminar una región de interés de mBAC e insertar el casete de selección, uno para 5' y uno para 3'. Los enlazadores tenían una superposición de 40 pb con mBAC y una superposición de 40 pb con un casete de selección. Primero, se escindieron 39,5 kb del vector de direccionamiento de 206 kb (mBAC) de acuerdo con la siguiente reacción: 500 µL de reacción (con adición de H₂O): añadir 1 µL de Cas9 (@ 0,89 µg/µL), 2 µL de cada dúplex de ARN (@ 50 µM), 250 µL de regulador, 220 µL (12,5 ng) de maxi prep BAC, e incubar a 37°C durante 1 hora. El ADN digerido se purificó por extracción con fenol/cloroformo/alcohol isoamílico (PCI) y luego se resuspendió en 55 µL de regulador TE. Después de la limpieza con PCI de la escisión de mBAC, se realizó el ensamblaje a 50°C durante 1 hora, y 10 µL de la reacción se sometieron a electroporación en células DH10B (Figura 9). La secuenciación a través de las uniones confirmó el ensamblaje correcto (Figura 10). El enlazador 1 (oligo 1 de unión) es continuo desde la secuencia de mBAC hasta la secuencia del casete (SEQ ID NO: 12). El enlazador 2 (oligo 2 de unión) es continuo desde la secuencia de casete hasta la secuencia de mBAC (SEQ ID NO: 13).

Ejemplo 5: ensamblaje de dos vectores de BAC usando enlazadores (Oligos de unión)

Se utilizó la unión de 2 mBAC mediante el montaje con Cas9/isotérmico para elaborar un vector de direccionamiento que contiene brazos de homología con una región genómica de ratón y sitios de restricción para insertar un gen humano mediante ligación de BAC. Este vector de direccionamiento se usó en una ligación de BAC para elaborar un vector de direccionamiento humanizado. El mBAC se escindió de acuerdo con la siguiente reacción: 12,5 µg de ADN, 2 µL de cada ARN hibridado (@ 50 µM), 10 µL de Cas9 (@ 0,89 µg/µL), 250 µL de regulador, H₂O hasta 500 µL. La mezcla se incubó a 37°C durante una hora; se limpio por extracción con fenol/cloroformo/alcohol isoamílico (PCI); y se resuspendido en 20 µL TE. Los dos BAC de ratón se ensamblaron luego junto con los enlazadores (Figura 11) de acuerdo con la siguiente reacción: 6 µL (2 µg) de escisión de bMQ-208A16, 5,6 µL (2 µg) de escisión de bMQ-50F19, 0,25 µL de cada enlazador (@ 50 µM), 4,3 µL (100 ng) de casete de selección, casete (Ubi-Hyg) 12 µL de mezcla maestra de ensamblaje de alta concentración, 11,35 µL de H₂O. La mezcla de reacción se incubó a 50°C durante 1 hora y se dializó frente a H₂O a 30°C. Se usaron 10 µL o 30 µL de la reacción dializada para transformar las células DH10B. La secuenciación de Sanger confirmó todas las uniones. La secuenciación de Illumina reconfirmó todas las uniones (Figura 12 y SEQ ID NO: 17). El enlazador 1 es continuo desde mBAC hasta el casete (SEQ ID NO: 14). El enlazador 2 no es continuo desde el casete hasta mBAC. Incorpora una secuencia espaciadora humana según el diseño del proyecto. El enlazador 3 no es continuo desde mB2 hasta mB3. Incorpora una secuencia única que se utilizó para la verificación de PCR. Esta área se eliminó cuando se linealizó para electroporación ES (SEQ ID NO: 15).

La Fig. 13 ilustra un ejemplo del uso de 4 oligos de unión (enlazadores) para insertar fragmentos grandes de genes humanos en un mBAC usando cuatro enlazadores y un ensamblaje isotérmico.

Ejemplo 6: Reactivos y mezclas de reacción para escisión y ensamblaje

El ARN de CRISPR (ARNcr) (ordenado como ARNcs) contiene: (1) 20 nucleótidos de ARN que es complementario a un área objetivo para escisión; (2) y una cola que se hibridará con el ARNcrtra: <20 nt de ARNcrispr>

GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO: 10).

ARNcrtra (ordenado como ARNcs): GUUGGAACCAUUCAAAACAGCAUAGCAAGUUAAAUAAGGCUAGUCCGUUUAU
CAACUUGAAAAGUGGGCACCGAGUCGGUGCUUUUUUUU (SEQ ID NO: 11).

5 Todo el ARN se resuspende en 100 μ M en H₂O. Se combinan 2,5 μ L de cada ARNcr y ARNcrtra con 5 μ L de regulador de hibridación (concentraciones finales: Tris 10 mM pH 7.5-8.0, NaCl 50 mM, EDTA 1 mM). La mezcla se incuba luego a 95°C durante 5 minutos y se enfría lentamente a temperatura ambiente durante 1 hora. El regulador de escisión Cas9 2X contiene HEPES 40 mM pH7.5 (final = 20 mM); KCl 300 mM (final = 150 mM); DTT 1 mM (final = 0,5 mM); EDTA 0,2 mM (final = 0,1 mM); MgCl₂ 20 mM (final = 10 mM).

10 Reacción de escisión de Cas9 a gran escala: Añadir en orden a temperatura ambiente: H₂O hasta 500 μ L, 250 μ L de regulador de escisión 2x, 12,5 μ g ADN, 2 μ L de cada ARN (concentración 50 μ M), 10 μ L de Cas9 (concentración 0,89 mg/mL) e incubar a 37°C durante 1 hora.

Esta reacción se puede escalar según sea necesario, por ejemplo: H₂O hasta 50 μ L, 25 μ L de regulador, 125 ng de ADN, 2 μ L de cada ARN (concentración 5 μ M), 1 μ L de Cas9 (concentración 0,89 mg/mL) e incubar a 37°C durante 1 hora.

15 La reacción de ensamblaje se lleva a cabo de la siguiente manera: regulador Iso-Thermal: 3 mL de Tris-HCl 1M (pH 7,5); 150 μ L de MgCl₂ 2M; 60 μ L 100 mM de cada uno de: dGTP, dATP, dTTP, dCTP; 300 μ L de DTT 1M; 1,5 g de PEG 8000; 300 μ L de NAD 100 mM. El regulador Iso-Thermal o se almacena en alícuotas de 320 μ L a -20°C. La mezcla maestra se prepara de la siguiente manera: 320 μ L regulador Iso-Thermal; 0,64 μ L de exonucleasa T5 (concentración del patrón = 10 U/ μ L); 20 μ L de ADN polimerasa Phusion (concentración del patrón = 2 U/ μ L); 160 μ L de Taq ADN Ligasa (concentración del patrón = 40 U/ μ L); 699,36 μ L de H₂O; se mezclan juntos y se toman alícuotas de 15 μ L o 30 μ L y se almacena a -20°C. Se usan 15 μ L de mezcla maestra (MM) en un volumen total de 20 μ L de reacción.

20 Alternativamente, se puede preparar una mezcla maestra de alta concentración (GA MM HC) como sigue: 320 μ L de regulador Iso-Thermal; 0,64 μ L de exonucleasa T5 (concentración del patrón = 10 U/ μ L); 20 μ L de ADN polimerasa Phusion (concentración del patrón = 2 U/ μ L); 160 μ L de Taq ADN Ligasa (concentración del patrón = 40 U/ μ L); se mezclan juntos y se toman alícuotas de 6 μ L o 12 μ L y se almacena -20°C. Usar 6 μ L de la mezcla maestra en un volumen total de 20 μ L de reacción.

30 Para todas las reacciones de ensamblaje, se debe determinar la concentración de ADN (por ejemplo, mediante Nano Drop) y se usa una relación molar 1:6 (vector con respecto al inserto o a los insertos). Para la concentración estándar, se usan 15 μ L de la mezcla maestra de ensamblaje. El ADN y el agua se agregan a un volumen final de 20 μ L en un tubo de PCR de 200 μ L. La reacción se lleva a cabo en un termociclador a 50°C durante 1 hora. La reacción puede almacenarse a -20°C. Para alta concentración, se usan 6 μ L de la mezcla maestra de ensamblaje de alta concentración. El ADN y el agua se agregan hasta un volumen final de 20 μ L en un tubo de PCR de 200 μ L. La reacción se lleva a cabo en un termociclador a 50°C durante 1 hora. La reacción puede almacenarse a -20°C. Una vez completada la reacción, se dializan 10 μ L frente a agua durante 30 minutos y se someten a electroporación en células electrocompetentes apropiadas (por ejemplo, células DH10B o Pir +).

35 Reacción de ensamblaje isotérmico/Cas9: para Cas9 se digieren 2,5 μ g de cada ADN (por ejemplo, ADN de BAC), se digieren 4 μ L de ARN guía/crtra 10 μ M y 5 μ L de proteína Cas9 (0,89 μ g/mL) durante 2 horas a 37°C. La reacción se inactiva por calor a 65°C durante 20 min, se extrae con fenol cloroformo (por ejemplo, para eliminar la proteína Cas9), se lava una vez con etanol al 70% y se resuspende el ADN en 35 μ L de agua. El ensamblaje isotérmico se realiza con 5 μ L de ADN mezclado junto con 15 μ L de la mezcla maestra (MM) como se describe en otra parte de este documento y se incuba a 50°C durante 1 hora. La reacción se desaliniza durante 30 minutos y 8 μ L de la reacción se pueden someter a electroporación en las células.

Ejemplo 7: Ensamblaje isotérmico/Cas9 para insertar secuencia humana en un vector BAC

45 Con el fin de construir un vector de direccionamiento humanizado, se escindió MAID 6236 con un complejo de ARNg-Cas para generar un fragmento escindido con secuencias solapantes. También se escindió VI568 con un complejo de ARNg-Cas para generar secuencias que se superponen con el fragmento de MAID6236. El ensamblaje Cas9/isotérmico se realizó como se describió anteriormente, dando como resultado la inserción del locus humanizado en el vector (VI599). Este proceso se describe en la Fig. 14.

Ejemplo 8: Ensamblaje isotérmico/Cas9 usando un Bloque g sin selección

50 La digestión y el ensamblaje de Cas9 también pueden realizarse sin selección, por ejemplo, utilizando fragmentos de ADN de Bloque g. Con el fin de probar la posibilidad de añadir ADN de doble cadena en un locus sin un casete de selección, se sintetizaron fragmentos de ADN Bloque g y se insertaron en el constructo. Como se describe en la Fig. 15A y B, se diseñó un Cas9/ARNg para dirigirse a dos sitios dentro del locus beta de TCR para eliminar un fragmento de 4,4 kb. Un Bloque g fue diseñado para introducir un sitio de reconocimiento de meganucleasa en la construcción. El Bloque g fue capaz de insertarse en el constructo sin usar un marcador de selección. LA Fig. 15A muestra la inserción de un Bloque g PIScel y la Fig. 15B demuestra la inserción de un Bloque g MauBI.

ES 2 666 179 T3

Las construcciones finales se confirmaron para la inserción exitosa de cada uno de los Bloque g mediante cribados de unión de PCR usando los cebadores indicados en la Tabla 1. El protocolo para los cribados de unión es el siguiente: La reacción de PCR contenía: 1 µL de ADN, 0,5 µL de cebador 1, 0,5 µL de cebador 2, 1 µL de DMSO, 4 µL de dNTP, 2,5 µL de regulador 10x, 0,5 µL de Ex-Taq y 15 µL de agua. La reacción se llevó a cabo en un termociclador a 95°C durante 3 minutos, 95°C durante 30 segundos, 55°C durante 30 segundos durante 25 ciclos, seguido de 72°C durante 30 segundos, y 72°C durante 5 minutos. Las secuencias de unión se confirmaron mediante secuenciación.

Tabla 1: cebadores para el cribado de unión de MAID1715 ya sea con Bloque g PI-Scel o Bloque g MauBI

MAID1715+ Bloque g PI-Scel		
Nombre del cebador	Secuencia	Tamaño de la unión
(m380)5' 302p18 de detección	GGAAAGCCACCCTGTATGCT (SEQ ID NO: 18)	796 bp
302p18(m41) de detección hacia 3'	CTTGGCCAACAGTGGATGG (SEQ ID NO: 19)	
Nombre del cebador Cas9	Secuencia	Secuencia objetivo de ADN
5' de diana 1715	CUAAAUGAUUCUCAUCUGC GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO: 20)	CTAAAATGATTCTCATCTGC(AGG) (SEQ ID NO: 22)
3' de diana 1715	GCUCUCAACUUCACCCUUUC GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO: 21)	GCTCTCAACTTCACCCTTTC(TGG) (SEQ ID NO: 23)
MAID1715+ Bloque g MauBI		
Nombre del cebador	Secuencia	Tamaño de unión
(m380)5' 302p18 de detección	GGAAAGCCACCCTGTATGCT (SEQ ID NO: 18)	759bp
302p18(m41) de detección hacia 3'	CTTGGCCAACAGTGGATGG (SEQ ID NO: 19)	
5' de diana 1715	CUAAAUGAUUCUCAUCUGC GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO: 20)	CTAAAATGATTCTCATCTGC(AGG) (SEQ ID NO: 22)
3' de diana 1715	GCUCUCAACUUCACCCUUUC GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO: 21)	GCTCTCAACTTCACCCTTTC(TGG) (SEQ ID NO: 23)

Ejemplo 9: Ensamblaje isotérmico/Cas9 para insertar secuencia humana en un vector BAC usando Oligos de unión

La Fig. 16 proporciona un ejemplo de humanización directa usando ensamblaje isotérmico/Cas9 y oligos de unión. El fragmento humano y la eliminación del ratón y separación por Cas9 (cada BAC usa 2 ARN de crispr). El fragmento humano y la cadena principal del ratón se unen entre sí en una reacción de ensamblaje de Gibson con 3 conectores (oligos de unión) y un casete de selección.

La Fig. 17 proporciona un ejemplo de humanización indirecta usando un ensamblaje isotérmico/Cas9 y elementos de unión para el ensamblaje en un vector de direccionamiento grande (LTVEC). El fragmento humano en el hBAC es escindido mediante Cas9 usando 2 ARN de crispr. El donante comprende oligos de unión secuencia arriba y secuencia abajo y un casete de selección. Después de la escisión de hBAC por Cas9, el fragmento es "capturado" por ensamblaje de Gibson usando un donante sintético con salientes complementarios incorporados. La construcción del vector de direccionamiento se completa mediante el ensamblaje de Gibson o BHR.

Ejemplo 10: Introducción de una mutación puntual mediante ensamblaje isotérmico/Cas9

La Fig. 18 proporciona un ejemplo de utilización de un ensamblaje isotérmico/Cas9 para introducir una mutación puntual. Un donante se elabora mediante clonación tradicional. Se inserta un casete de selección en un fragmento de ADN sintético que contiene solapamientos de enlazador y la mutación puntual. El mBAC se escinde con Cas9, la secuencia se elimina del mBAC y el mBAC se ensambla mediante Gibson con el donante que da como resultado una construcción (LTVEC) que comprende la mutación puntual y el casete de selección.

Ejemplo 11: Recorte de BAC mediante ensamblaje isotérmico/Cas9

La Fig. 19 proporciona un ejemplo de un recorte de BAC utilizando el método de ensamblaje isotérmico/Cas9. El área que se debe eliminar del LTVEC se recorta con Cas9. En este ejemplo, el recorte de BAC elimina la secuencia de Ori. El Ori se reemplaza en una reacción de ensamblaje de Gibson usando 2 enlazadores (oligos de unión).

Ejemplo 12: Otros métodos para digerir BAC con CAS9 seguido de ensamblaje

Se pueden usar otros métodos en los métodos proporcionados en este documento que incluyen lo siguiente: se hibridaron previamente ARNcrtra sintético o transcrito *in vitro* y ARNcr antes de la reacción calentando a 95°C y enfriando lentamente hasta temperatura ambiente. El ADN del plásmido nativo o linealizado (300 ng (aproximadamente 8 nM)) se incubó durante 60 minutos a 37°C con una proteína Cas9 purificada (50-500 nM) y un dúplex de ARNcrtra: ARNcr (50-500 nM, 1:1) en un regulador de escisión del plásmido Cas9 (HEPES 20 mM, pH 7,5, KCl 150 mM, DTT 0,5 mM, EDTA 0,1 mM) con o sin MgCl₂ 10 mM. Las reacciones se detuvieron con regulador de carga de ADN 5X que contenía EDTA 250 mM, se resolvieron mediante electroforesis en gel de agarosa al 0,8 o al 1% y se visualizaron mediante tinción con bromuro de etidio. Para los ensayos de escisión del mutante Cas9, las reacciones se detuvieron con regulador de carga SDS 5X (glicerol al 30%, SDS al 1,2%, EDTA 250 mM) antes de la carga en el gel de agarosa.

Se diseñaron un ARNcr artificial y un ARNcrtra artificial para dirigir secuencias específicas en MAID 6177 (LTVEC de 116 kb) para el ensamblaje con un producto de PCR de 3 kb (UB-HYG). El producto de PCR contenía superposiciones de 50 pb con el vector. Se usó un montaje isotérmico de una etapa con base en el uso de una exonucleasa 5' a 3' no termoestable aislada que carece de la actividad de exonucleasa 3' de la manera. Se estableció una reacción que contenía lo siguiente: 100 fmol de cada sustrato de ADNdc, 16 µL de regulador ISO 5X, 16 µL de exonucleasa T5 (0,2 U/µL, Epicentre), 8,0 µL de Taq ADN ligasa (40 U/µL, NEB), 1,0 µL ADN polimerasa Phusion^{MR} (2 U/µL, NEB) y agua hasta 80 µL. El regulador ISO 5X (ISOthermal) fue 25% de PEG-8000, Tris-Cl 500 mM, MgCl₂ 50 mM, DTT 50 mM, NAD 5 mM, y 1000 µM de cada dNTP (pH 7,5).

Esto produjo una concentración final de 1,25 fmol/µL de cada ADNdc (o 45 fmol/µL de cada ADNcs) que debía ensamblarse, 5% de PEG-8000, Tris-Cl 100 mM pH 7.5, MgCl₂ 10 mM, DTT 10 mM, 200 mM de cada dNTP, NAD 1 mM, 0.02 U/µL de exonucleasa T5, 4 U/µL de Taq ADN ligasa, y 0,03 U/µL de ADN polimerasa PHUSION.

Los métodos usaron 1,64 µl (0,2 U/µL) de exonucleasa T5 para sustratos que se superponen en 20-80 pb, y para sustratos que tienen superposiciones más grandes (por ejemplo, 200 pb), se usaron 1,6 µL (1 U/µL) de exonucleasa T5. Se usó exonucleasa T5 como una dilución 1:50 (en regulador de almacenamiento de exonucleasa T5) a partir del patrón de enzima concentrada de exonucleasa T5 (Epicentre) de 10 U/µL. La reacción se incubó luego a 50°C durante 15 minutos.

Ejemplo 13: Otros métodos para unir entre sí dos BAC superpuestos

Se pueden usar otros métodos en los métodos proporcionados en este documento que incluyen lo siguiente: se hibridaron previamente ARNcrtra y ARNcr transcritos *in vitro* sintéticos antes de la reacción calentando a 95°C y enfriando lentamente a temperatura ambiente. Se incubó ADN de plásmido nativo o linealizado (300 ng (aproximadamente 8 nM)) durante 60 minutos a 37°C con una proteína Cas9 purificada (50-500 nM) y un dúplex ARNcrtra: ARNcr (50-500 nM, 1:1) en un regulador de escisión del plásmido Cas9 (HEPES 20 mM, pH 7,5, KCl 150 mM, DTT 0,5 mM, EDTA 0,1 mM) con o sin MgCl₂ 10 mM. Las reacciones se detuvieron con regulador de carga de ADN 5X que contenía EDTA 250 mM, se resolvieron mediante electroforesis en gel de agarosa al 0,8 o al 1% y se visualizaron

mediante tinción con bromuro de etidio. Para los ensayos de escisión del mutante Cas9, las reacciones se detuvieron con regulador de carga SDS 5X (glicerol al 30%, SDS al 1,2%, EDTA 250 mM) antes de la carga en el gel de agarosa.

5 Se diseñaron un ARNcr artificial y un ARNcrtra artificial para dirigir secuencias específicas en el BAC HLA-DQ humanizado para el ensamblaje con un BAC HLA-DR humanizado. Los vectores que contenían ~70 pb se superponen entre sí creados por escisión de Cas9 en dos sitios en cada vector (véase, Figura 2). Se usó un montaje isotérmico de una etapa con base en el uso de una exonucleasa 5' a 3' no termoestable aislada que carece de la actividad de exonucleasa 3' como sigue. Se estableció una reacción que contenía aproximadamente lo siguiente: 100 fmol de cada sustrato de ADNdc, 16 µL de regulador ISO 5X, 16 µL de exonucleasa T5 (0,2 U/µL, Epicentre), 8,0 µL de Taq ADN ligasa (40 U/µL, NEB), 1,0 µL de ADN polimerasa Phusion^{MR} (2 U/µL, NEB) y agua hasta 80 µL. El regulador ISO 5x (ISOthermal) fue 25% de PEG-8000, Tris-Cl 500 mM, MgCl₂ 50 mM, DTT 50 mM, NAD 5 mM, y 1000 µM de cada dNTP (pH 7,5).

10 Esto produjo una concentración final de aproximadamente 1,25 fmol/µL de cada ADNdc (o 45 fmol/µL de cada ADNcs) que debía ensamblarse, 5% de PEG-8000, Tris-Cl 100 mM pH 7,5, MgCl₂ 10 mM, DTT 10 mM, 200 mM de cada dNTP, NAD 1 mM, 0,02 U/µL de exonucleasa T5, 4 U/µL de Taq ADN ligasa, y 0,03 U/µL de ADN polimerasa PHUSION.

15 Los métodos usaron 1,64 µL de exonucleasa T5 de 0,2 U/µL para sustratos que se solapan en 20-80 pb, y para sustratos que tienen superposiciones más grandes (por ejemplo, 200 pb), se usaron 1,6 µL de exonucleasa T5 de 1 U/µL. La exonucleasa T5 se usó como una dilución 1:50 (en regulador de almacenamiento de exonucleasa T5) a partir del patrón de enzima concentrada de exonucleasa T5 (Epicentre) de 10 U/µL. La reacción se incubó luego a 50°C durante 15 minutos.

20 Ejemplo 14: Otros métodos para ensamblar un inserto con un vector BAC

Se pueden usar otros métodos en los métodos proporcionados en este documento que incluyen los siguientes: Disolver los ARNcr y ARNcrtra hasta 100 µM en regulador Hybe (regulador 10X: Tris 20 mM 7,5, NaCl 100-150 mM, MgCl₂ 10 mM, DTT 1 mM, EDTA 0,1 mM, 100 µg/mL de BSA). Con el fin de hibridar los ARN, se añadieron 10 µL de ARNcr 100 µM y 10 µL de ARNcrtra 100 µM a 80 µL de regulador de hibridación. Calentar los ARN en un bloque de temperatura a 25 90°C, luego retirar el bloque del calentador y enfriar en la mesa. La concentración final de ARN es de aproximadamente 10 µM.

Con el fin de digerir el BAC, se usa ADN de BAC maxiprep limpio y el BAC se digiere de acuerdo con la siguiente mezcla.

	1X
ADN de BAC 500 ng	X µL
BSA	0,5 µL
ARN	2 µL (1 µL de cada híbrido crtra:ARNcr)
Cas9 (4,5 mg/mL)	1 µL
Regulador 10x	1,5 µL
H ₂ O	hasta 15 µL

30

Digerir durante 1 hora a 37°C y luego eliminar la sal durante 30 min.

Con el fin de ensamblar el BAC y el inserto, digerir un plásmido o realizar una PCR para crear un inserto. Para las reacciones de PCR, correr una pequeña alícuota en un gel y buscar un producto limpio; si el producto no está limpio, realice la limpieza por PCR en lugar de la extracción con gel. Se desea una relación molar de 1:1-1:6 para el BAC: inserto. Usualmente, funcionarán 50 ng del inserto purificado. La siguiente mezcla de reacción se puede usar:

35

ES 2 666 179 T3

Digestión de BAC	4 µL
Inserto	1 µL
Mezcla de montaje	15 µL

5 Agregar el ADN y mezcla sobre hielo o directamente en una máquina de PCR a 50°C. Incubar a 50°C durante 1 hora. Añadir 0,5 µL de proteinasa K (20mg/mL) e incubar a 50°C durante 1 hora. Desalinizar durante 30 minutos y someter a electroporación 8 µL de la reacción en células DH10B. 1Se pueden correr 10 µL de la digestión de BAC en un gel de campo pulsado para verificar la eficiencia de la digestión. Usar agua libre de RNasa y reguladores. El regulador de reacción final contiene: Tris 20 mM 7,5; NaCl 100-150 mM; MgCl₂ 10 mM; DTT 1 mM; EDTA 0,1 mM; 100 µg/mL de BSA; para un volumen final de 15 µL

10 La secuencia de ARNcrtra usada en el ejemplo es: CAAAACAGCAUAGCAAGUUAUUUUUAAGGCUAGUCCGUUAUC (SEQ ID NO: 9). Este ARN de CRISPR (ARNcr) contiene: (1) aproximadamente 20 nucleótidos de ARN complementario con la secuencia objetivo y (2) una secuencia de cola (GUUUUAGAGCUAUGCUGUUUUUG (SEQ ID NO: 10)) que se hibridará con el ARNcrtra.

Listado de secuencias

<110> Schoenherr, Chris

McWhirter, John

15 Momont, Corey

Macdonald, Lynn

Murphy, Andrew

Warshaw, Gregg S.

Rojas, Jose F.

20 Lai, Ka-Man Venus

Valenzuela, David M.

Montagna, Caitlin

<120> Ensamblaje de AND mediado por nucleasa

<130> 057766-461002

25 <140> PCT/US2015/037199

<141> 2015-06-23

<150> US 62/036.983

<151> 2014-08-13

<150> US 62/016.400

30 <151> 2014-06-24

<150> US 62/015.809

<151> 2014-06-23

<160> 25

<170> FastSEQ para Windows Versión 4.0

35 <210> 1

<211> 80

<212> ARN

<213> Secuencia Artificial

<220>

<223> ARNg quimérico sintético

5 <400> 1

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguaaucaac uugaaaaagu 60
ggcaccgagu cggugcuuuu 80

<210> 2

<211> 42

<212> ARN

10 <213> Secuencia Artificial

<220>

<223> ARNg quimérico sintético

<400> 2

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cg 42

15 <210> 3

<211> 30

<212> ARN

<213> Secuencia Artificial

<220>

20 <223> ARNcr sintético

<400> 3

guuuuagagc uagaaauagc aaguuaaaau 30

<210> 4

<211> 33

25 <212> ARN

<213> Secuencia Artificial

<220>

<223> ARNcr sintético

<400> 4

30 guuuuagagc uagaaauagc aaguuaaaau aag 33

<210> 5

<211> 26

<212> ARN

<213> Secuencia Artificial

35 <220>

<223> ARNcr sintético

<400> 5
gaguccgagc agaagaagaa guuuua 26
<210> 6
<211> 12
5 <212> ARN
<213> Secuencia Artificial
<220>
<223> ARNcrtra sintético
<400> 6
10 aagguaguc cg 12
<210> 7
<211> 50
<212> ARN
<213> Secuencia Artificial
15 <220>
<223> ARNcrtra sintético
<400> 7
aagguaguc cguaucaac uugaaaaagu ggcaccgagu cggugcuuuu 50
<210> 8
20 <211> 23
<212> ADN
<213> Secuencia Artificial
<220>
<223> un locus objetivo que se enlaza a un ARN guía (ARNg)
25 <220>
<221> característica nueva
<222> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
<223> n = A, T, C o G
<400> 8
30 gnnnnnnnnn nnnnnnnnn ngg 23
<210> 9
<211> 41
<212> ARN
<213> Secuencia Artificial
35 <220>
<223> ARNcrtra sintético
<400> 9

caaacagca uagcaaguua aaauaaggcu aguccguuuu c 41

<210> 10

<211> 22

<212> ARN

5 <213> Secuencia Artificial

<220>

<223> Región de ARNcr sintético complementaria a ARNcrtra

<400> 10

guuuuagagc uaugcuguuu ug 22

10 <210> 11

<211> 89

<212> ARN

<213> Secuencia Artificial

<220>

15 <223> ARNcrtra sintético

<400> 11

guuggaacca uucaaaacag cauagcaagu uaaaauaagg cuaguccguu aucaacuuga 60
aaaaguggca ccgagucggu gcuuuuuuuu 89

<210> 12

<211> 145

20 <212> ADN

<213> Secuencia Artificial

<220>

<223> Sintética - Confirmación de montaje continuo de mBAC con el casete

<400> 12

ttgtgtgaat ataataatat cagtgttct ttacttccaa aactggacag cgcacaaac 60
atcagaaaca acagtatcag ctctgtccc aactaccatg ggtaccgatt taaatgatcc 120
agtgttcctg cagaggagag attgg 145

25 <210> 13

<211> 205

<212> ADN

<213> Secuencia Artificial

30 <220>

<223> Sintética - Confirmación

<400> 13

cagcccctag ataacttctg ataatgtatg ctatacgaag ttatgctagc tcggtcacac 60
tgctcagcttc ctgtgtttcc taggcatga taagatgcag caaagtttct gcaatgcaca 120
atgaggcagc cgtcgggaata gatttgagaa agtcatgatg atgcaatgtg cacactcttc 180
ctttgtattt atctctatcc accat 205

ES 2 666 179 T3

<210> 14
 <211> 138
 <212> ADN
 <213> Secuencia Artificial
 5 <220>
 <223> Sintética - confirmación del montaje continuo de mBAC al casete
 <400> 14
 actttagggt ttggttggtt tttaaagccc tatttccagt atgtggaat gagccaaccc 60
 aggacagctt ccgctggatc gtggacagct tctatggccg tgcacgtgta cactcgagat 120
 aacttcgtat aatgtatg 138
 <210> 15
 10 <211> 147
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintética
 15 <400> 15
 tccaacgac agcagaacta actgagagga gagcacagta gcggccgcaa attgctttga 60
 gaggtctat aaaaccttag aggctattta aatttaaattg gccggcccga cggccaggcg 120
 gccgccaggc ctaccacta gtcaatt 147
 <210> 16
 <211> 9
 <212> PRT
 20 <213> Desconocida
 <220>
 <223> Sintética
 <400> 16
 Leu Ala Gly Leu Ile Asp Ala Asp Gly
 1 5
 25 <210> 17
 <211> 49631
 <212> ADN
 <213> Secuencia Artificial
 <220>
 30 <223> Sintética
 <220>
 <221> característica nueva
 <222> (22396)...(22533)
 <223> Enlazador 1

- <220>
<221> característica nueva
<222> (22494)...(25426)
<223> Secuencia del Casete
- 5 <220>
<221> característica nueva
<222> (25427)...(25595)
<223> Secuencia del espaciador humano
<220>
- 10 <221> característica nueva
<222> (25596)...(40791)
<223> Secuencia de BMQ-208A16
<220>
<221> característica nueva
- 15 <222> (25387)...(25672)
<223> Enlazador 2
<220>
<221> característica nueva
<222> (40792)...(40858)
- 20 <223> Secuencia adicional única del enlazador 3
<220>
<221> característica nueva
<222> (40752)...(40898)
<223> Enlazador 3
- 25 <220>
<221> característica nueva
<222> (1)...(22395)
<223> bmq-50F19
<220>
- 30 <221> característica nueva
<222> (40899)...(49631)
<223> bmq-50F19
<400> 17

ES 2 666 179 T3

```

gctggagtgt ggtcaggcaa catccccaaa gggatggaga tgccgggacg acacctttag 60
ggaggcagtg gctctggtcc gggattccgg tgctggccat ccctcaccag ccacagcggg 120
tggcgcagga gggatcgccg cgcgcctggg gctagggggc gaactggacc gacttttctt 180
agttcgccta gctgctccga ccgctgccgc gccgagatgt tgaaagcaca ggcgagttct 240
aacttgcgcg ctattctttt cagcgcgggg gaatcggtcg agggccctgc gtggcgctgg 300
cttccaccct cgcggccagg gggcaggcgc gggaggccgg cttcggctcc gtgccctgc 360
aaacttccca agaccttctt tctccccccc acctcaccac ccagttcaat aaaatctacc 420
cttaaaggca gacttgcttt caaatccacg gcacccatta tgtgtttggt gtgaaacgct 480
atcaacattt aaaactctat tgtccaagc gtcccaaadc cctgtaaadc ttccaccagc 540
ctggactcat ttcatctga aaagcctggt tagtttgaat agaaaagcaa tcaggcgcgc 600
ctctcgctct cgttggaatg tcaattaaa tgagatttc tcagagctct ttagcgcgcc 660
aagaagtggg acaaacaggg atatttcagg ctgacaaatg aaagaaatgc tacaatgaag 720
tgggggtggc atgtgcaccc caaactgctt ggagtaccca ctgaaagagt aggtcagggg 780
ttatggtctt acttacgaca gcttatattt ttggggtttc gttgtgttta gggccccccc 840
ttggtgtccc cccccccat gagcccatga cagctccctt ccctattcag ccccggtggg 900
aagtaagggg gccttgaacc agggtagaga ggctacattt agtattaacc tgggagtgtt 960
gacttctccc aggagtaatc cacttgagaa caaaatgcca attgctctgc ccgctgaggt 1020
atcctggaac taccctttaa ggtagcagta cccgtcgcac cgccccctc cccaagggc 1080
ttgccttaaa ttaacctgcc ttcttcaggg acaggggaga gtgtgtaaac gtgtataaca 1140

```

ctgocgaagc	tcaccagccg	ggccctttcg	gccgggtccc	tttgctgtc	tttgaggca	1200
gacttgtgtg	gagatgcccc	caaggggocg	gtggccgtga	agagccatcc	gtcagagtga	1260
gggtgaggac	tccctccctcg	taggctgaga	agagagtatc	ctttcagggg	gaaaaataaa	1320
cacgctgggg	ctttctctgg	ggttcagcct	ccaggaagga	ttatggtatt	gaaggcagga	1380
agctgggatt	gtggccgcca	gcagcatgct	ggcctgtgt	tcccaacacg	gagccttggg	1440
acctaattat	cctgcctagg	aggtcgctca	gcacttttgt	ccactccggg	gaggagctgt	1500
gcagacctgc	tgccgtcact	tctcgcctta	cagaggtttg	aggagggggc	tcctgtgggg	1560
gctgggactt	cgaagaacga	acgttcaagt	tgagtcagcc	tggggcactg	gccatcttcc	1620
tcattcagct	ggagctgagg	tactcctggg	tagtggctag	tagagacagt	gggcccagca	1680
ctctgcttca	agacctactg	ggacctgaga	ttgcaaagt	gctggagagg	ggagtttacc	1740
tgcattctga	aagttccttag	gaaatcaacg	agaatgtttg	tgcactttcc	tttgactggt	1800
atgtagaat	agacaaggaa	ttatcttttg	tgactcttg	ctttaagaag	aaagaagact	1860
tgggggaaca	aaaatccttc	cagccaacta	aaaacactgg	gcacctaact	gctcatatac	1920
ccctggcttt	tgttgttagc	tataccattc	tacctgtgct	taaaaaaaca	accaaacagc	1980
agcttctat	tcccctcttg	gagatggtac	gtcctctctg	ccttagtctc	agtgaaggct	2040
gaaaggaaca	gatttttagga	cggaggttct	ggcagtgtcg	aaatcctgtg	tcataattga	2100
aagcatcaaa	agcgcacggg	attagaattc	ttttctctct	tctctctttt	tcattaaaac	2160
gctcaccat	ccccagtctc	ataaaatggg	catcccagca	tccaaagccc	atggttttgt	2220
gcgatccctt	cctgccattg	gtttcagcag	attctctaaa	gctcgtgcat	tctgactcaa	2280
agattagtca	ctgaagacac	tgaacaaaca	taaagttatt	tgtactgtgg	taagcttttt	2340
tttttgggaa	attctctgct	ttggatctag	taaattgagt	gcccccttgt	aaactgatac	2400
tggaggtttt	agccaatagg	ttagcgtatt	gaaagtccc	aggccaatca	cataccaggg	2460
cagcttgtac	gtatcatcac	cattactaat	aaaatcttga	attattcatc	aagggttcta	2520
tctttacccc	tttgacgtcg	gttgacgata	tttagttagt	atgcctgtac	actgccttgt	2580
agtcagtggg	agggaaattca	ggctttgaat	cccccggttg	gattaaactc	actctttgta	2640
agtggctgct	tggcgaaga	ttgaaatata	cgctgcatt	cgaaaatgaa	ttctgacaa	2700
tgtaaactgg	tgggaatggt	tttgaagcct	tctgagatt	ccttgattct	gttggctccc	2760
tttctttctg	agaaccgttc	tgaagcgagg	acgtgccgct	cagctcagct	gaaatcggt	2820
tctcagagca	gacccttcc	ccagtcagcg	tcttaaagcg	cagctggaat	aagagacgtt	2880
aatgaggctg	gccatgcaa	gcccagcgtt	ttaaactcag	gtttttctgc	agttgcctt	2940
gaaaggaat	aaggtcaagt	tgcttcagca	acctgcagc	tttgatagt	gacggaagg	3000
cacgctgcag	agctgggtgg	ctgggtccta	cagtgatggt	ttatcttgcg	tctctaaaa	3060
gtaagcttaa	aaaaaaaaag	attagcctac	tgcagcttgt	ggactagcct	ggaaacacct	3120
gggacgctga	ggtgaggatg	gaaggctttt	ccgataatga	gaaagaatgt	gtttgcgaat	3180
gtattgagag	gctgagaaat	ggttttatcc	catctgggtt	taagcaagt	ggcactgggg	3240
gaaaaaactg	aatctggctg	aatctctctc	tttcagtggc	agccacagca	gcagcagcag	3300
cagcagtggt	agcaggaaca	gcagtaacaa	cagcaacagc	agcacagccg	cctcagagct	3360
ttggctcctg	agccccctgt	gggtgaagg	cattgcaggt	agcccattgt	ctcagaagaa	3420
gtgtgcagat	gggattaccg	tccacgtgga	gatatggaag	aggaccaggg	attggcactg	3480
tgaccatggt	cagctggggg	cgcttcatct	gcctggctct	ggtcaccatg	gcaaccttgt	3540
ccctggcccc	gccctccttc	agtttagttg	aggataccac	tttagaacca	gaaggttaag	3600
tcatgcgtgc	cattttaagg	gtaccaagt	gttttgggga	tgtgtctggg	ggaagtggtc	3660
tttaagtggg	aggcctgttt	cagccggctg	ccatagagt	agtctctctc	cgcatcatat	3720
cggagcttag	aagggagggt	cttgtctccc	aggcatgagt	ggagtggttt	ggtttctct	3780
gttcttttgt	cttgggtgag	ggaagcagtg	gcagttcttg	tttagccagt	gccttacagc	3840
actctggagg	ggacgtacct	tggcagggtg	actgtggcct	tctgcagttg	ctctctagat	3900
tgagggaaaa	gccttgaatc	acactatctt	ttggctaaag	gaaataggca	gcctctgaaa	3960
gctgactttt	ttttctttt	tccgcattgt	ttaagagaaa	agaaggttct	gaagttgagc	4020
atggagagcc	gtgccatgct	ggatcggttt	ttaagctggt	gtaagctctt	tgtgctttca	4080
cccggcatca	cagagtgggc	aggtttcatg	ttgggaagat	tggaaagtga	atttgccaa	4140
agcttcccc	catctgggga	aaagccagat	ttcactagt	tgtttggctt	tcacacttg	4200
ggtgcaaatg	tgagaagcta	gttgtgagga	ggacgtggct	gaaatccgga	gctgggcaaa	4260
gcgctggctc	ttctcccagg	tccctcagag	acgtggctct	tggccaagcc	tctctccttg	4320
gtgccgcacg	ggaatctgtc	atcaggaggg	aatattggta	ggcgagttat	tttttgagt	4380
gtaatccogag	cgtgacactg	cagatcgacg	cactcatcgc	cacttaatga	acgtgtttgc	4440
tgagggccca	cctgggtccg	gctggcttgg	gagtcogtca	cggtcctgag	tgctggcagg	4500
tcagctgagt	tgctgtggct	atgcacactg	aatcagggtc	ctgattcatc	cagatcatcc	4560
agagggggat	tgtaggaggg	acaggacccc	tcccccaagg	gtgacctcaa	ggagggtctat	4620
gtaccatct	gagaggaggg	cttgagaaat	gggtccccag	taagatccac	ccagacagac	4680
actctccctg	gctttgtgtg	tatgtcgggc	cacacagatg	cctggaaatg	ttataaatta	4740
ccaggtatct	ttggaaagga	aatgaggtag	gagttttgtg	catgaggtgt	gttcaacata	4800
cagcctcacg	tcttttccg	gaaccacctc	tctgtgactt	atcctgtgac	gtcagggaga	4860
gtgtaatctg	caacagtgac	atgttttcaa	aggcttaaat	gtgaggggga	aaggattggg	4920

tttctgaaag tctggtctgc acttctttaa ttttgtaat aattaaaatg gatgtcccc 4980
 taattgccgg ttgtccctgg agtgtgtggc tcagcactaa ctaaggaagc tgagctagga 5040
 tttcctacag cgtgggcttc agaaacagcc cgggttagga aagaattgtc atttttcatt 5100
 tggactctcg gggcagtggt gctgtgagtt gatttcagtt gcagagtata aaatggctct 5160
 ggagggttcc ctggactgca tctaattacc tcagaaaagt tacaagatgt ttgtactcgc 5220
 aaggaggagg caggtggggg agaggaagga cagtgggctg gactccccca aatggctctt 5280
 tgtgtaagaa ccgatatcca acaatgctca cattggtgaa agcagatccc accacctggg 5340
 gacctgtagg tacatgtaag gttagggagg gaggctgaga agtctccgaa gttgtaggtc 5400
 acactttgcc aatgcccctg ggtacacttt gctaggctca gactttgcat gaggttcga 5460
 tcacatatag agtgggaga cgtaagaaa aagaaaagaa aaagaaaaaa ggaaaaaagg 5520
 aaaatgtctc aaggtgtgga gtttcaccag agcaagcttg gaaatgcag agaacccca 5580
 gacccctgat tgggtggatc tctttatcaa tagtactga acagtagtac catccccaga 5640
 tgccctctga ggaccagctc aagagattta gtttcacca gtgacctgga cagaaagcag 5700
 aaagcacagc tcctggcatt gatggtggcc ttggccatcc ccatccccag caagctgggg 5760
 caaagggggg gcacagttct cagtgcagca aacacggtag cctgagatga atgtgtctt 5820
 tggatggagg aggtggtgat gctggatttc ggagggtct gtgctactc tccttctctg 5880
 ttagaccaac attgccactg acatccaggc catcaagcta gaggctaggc tccatgctag 5940
 gctctggtgt ctaatgtgtg catatgtgca tctctccagc cgccatattt gatgcagcca 6000
 ggacttcagc taacactgag ttcagcttct gtctcctgaa gctttaccat ggaaggcatc 6060
 cgtttgctaa tttagaagct cagttagatc aatgtctatt gggccggaca aatagtaatc 6120
 caggaagtcc ctgaaaagag cctgtgcctc actactaagg agcccttttg accctctagg 6180
 gagatgttat gttcagtcac gtatgtctgt gcagtgatg tagccatgca atgtatgtcc 6240
 tcaccccgaa tcctatctctg tccgtgtgtc tctggacact ttctcaagtg gcagcagcag 6300
 gattgggtca agtcagttga cctaagaagg cagtcatctc tgaagattt tcctcggatc 6360
 ttcagaatag aaatgattgt atccagctgg tcatccctgt gacaaaggac aacagatca 6420
 acagttgggg acttcggagg ggtggctccg attctaagta ctgttctgtt gattcaaatc 6480
 ctgaatgttc ccagtgtagt caagcttcat tactgccagt ctggctctt actttcagat 6540
 tccccctgac gttgtcacct gctctgggta ataatgtcat tgttgacatt aagggaatc 6600
 gttacccca gccagtagg agctaaaata aaagggttt cccaaacca aaccctaat 6660
 tactttcca cctctgcta agtgcaagg gacggcctgg ggggtgggtg ggggtgggag 6720
 tgagggagta atttacatgc ctaaaaaaac acccaccatt tcttgggagc tctctgggt 6780
 tgatctgtt ccggattgaa gtgagccagc gaaaacctcg tgagtgtgag gtctacgtg 6840
 agacctgtc aagggttccc cccccctcac caccaccacc acagggtagt tcaagctct 6900
 tgtcagagag tatcctacat gctgtggtct cagccccca catttaattg ccagttaga 6960
 gaagagaaaa gaaatctctg cgtggtgcaa gtggatgatt taacaggagt ctgtgtttc 7020
 ctattatctg cattttttgt tcctcagtgat gactgtgaa atttaagagt tgactgtaac 7080
 ttgatagttt cgctgaggaa caagggctta ttcttggtca attaaacca atgcaggcgc 7140
 atgctgttaa acacacaatg aagtacacat tctttattag aatatagtgt atttcacaat 7200
 tcatgggcaa ggaactgtgt ttaataatc ttctagagca aaaaactggc cagcccagaa 7260
 aattggcatt tatataactc tttctgtctg gcttccactg atctgaatag agcaagttt 7320
 tgtgtgtgtg tgtgtgtgtg tgtgtgtacc tgtctgtcta tgttgcctg 7380
 ttatttcaat cagttaaaag actaaaatac ttaactaaaa aaaaatagcc acccctctc 7440
 agatagcctc taaagacttt atgctgtttt taagccttat ttttaatta ttttaattg 7500
 gggctctgtg aggcctgtgc acatgagtgc aggtccccct ggaggccaga ggtttttgat 7560
 gctagcccgg gacttggctc tggagtgaa tggatcct ctgcaagaac aatagatag 7620
 tacaaaaact gagctatctc tctgctcca atacagttca acatttttc tttttcttt 7680
 ttttttttt ctgttgaggg tgagaattca aacaaccaag cattccaata tgagatgtt 7740
 ccaatctctg atttaatgaa taatcacatg gttatgaaat aactggggtg gtgtttaa 7800
 aggaagggga tgtttaatgt tcacattctc tgtggagtgc gtgcagtag cccgtgctg 7860
 cagtagccct gtgccacact tatagacagt ttggctactt acatagttag ggtggtcatg 7920
 aaaagacaac taagtccctt tcatcaggct ccgtctaac ttttccattt ctgattgaa 7980
 cctgggatc gatccagcag ggtgtctgt ctgggtcagc agctaggagt tattttggg 8040
 gaagggatgc tgcaggctat tttacagata attatgggtt tcctgtgagc aactgtccct 8100
 gtaggggctg gagcaagtga tgattctgtg attaagagca cttcctcctt ttgcagagga 8160
 ccagtggtgg gtccccaca cccacagggg agtactctg attgtcttta tctccagtg 8220
 ctgggggcca gcgcctctga cctactgttg cctgcagtg cccttctcat gtgggctgcc 8280
 ctgcccctta ctgtctttgt ggcttcgtag agaatgatgg gaaaaaattc caagatggt 8340
 gtcccactgg tgaactaagg tgtttagtag tagcttttc ttaaaataca gttggtgagc 8400
 aagatagtg tgacacact taatcccaac actagggagg cagaagcggg tggttctttg 8460
 agattgaggc tagcctgtc tacagagtga gtgccaggac tacacacaca cacacagca 8520
 cacacacacc ccagaaagct tgaagttgta gttttacgaa agtgtattta accgtcagga 8580
 ctaactatga tctttctttt gggctggtag ctgatggtt ggtttttttt ttttttagatg 8640
 ggcactctcc acagcctggc ttgggatttg cctgtagct caggctcgtc tagaacttcc 8700

aatcccccta	cctcaacttc	cactccta	tgccggcat	ccttgaagag	catgtgtctt	8760
gattttctgt	aattttgaaa	aacttggcct	cggattttat	ggcttactta	tctttatgtg	8820
tatctttatg	tgttttgcct	gcatgtatgt	atgtgtacca	tgtatatgtt	tagtgcttgc	8880
agagaccaga	agaggacttg	cggtcccctg	gaatagagt	tatctacgtg	gttttgagct	8940
agcaccagtg	ggtcttcact	tccccctgg	ctctccagcc	cctggattaa	atgtggaatg	9000
tgctgtttgc	ttgcttgaag	ccaccatagg	cagtgcacag	ccttgtggac	tttctacact	9060
ctgagaataa	atgaaagtc	acttgcttgc	tcttggctgg	gtcaagtcag	ggagctaaac	9120
tatcacatac	ctcctcttta	acttcttgc	caactaaaga	atcatgaatc	ccaagccgtt	9180
tctggacaga	gagaattcca	ggttatgggtg	accatgtttt	atgaggatgt	taaaaatagc	9240
tcctaaggag	gatgctgaca	gattcaggaa	ggagaaccgg	gcctcatggt	tatttgggtg	9300
ttatttatgt	agcatgttcc	tgagacatct	caatcctgag	cactaaggaa	gtcaacacat	9360
tgtttcctaa	ccctggaact	tgtttttcac	ttcttatacc	tgacagtta	caaatactgg	9420
ttcccccccc	cccccatgtg	tggccaagtg	ttttaagggt	atctaaccac	gaaaatggcc	9480
aatttgggtg	gctgttatag	atcaaaagga	gatccttggag	actagagatc	tctgtcaagt	9540
ttattctctt	tggaaccctt	tcaagttcac	attgagagct	gacagtggc	tgccctaga	9600
gtcatgtggc	ttgcttcaag	cgcctctcc	cccattccac	ctcaaccctt	tggactcca	9660
ctaagactgt	tgcttagctg	attgtagcag	gtaccttgct	gaatgtgtaa	cctgtataga	9720
ttatgtgttt	cagatttaaa	accactcagg	tctttaaaga	ctaagggatc	tgatccgaca	9780
tttgacttaa	aattttaagt	agaaactaag	taaagttggt	ttgaatagta	tgtgttgtgt	9840
tttctggagg	tacagtctca	taggaaatcg	cccttgggtg	ctgagttga	atgtgctac	9900
tatctacttg	accttagtca	agtgcagataa	cctggttgaa	attccaagat	aatatctgtc	9960
taaattgcac	agattgaata	cacactggac	tgtagttcct	ggcccagtg	aggcgcagg	10020
taagtgtcta	tttctccca	ccccacacct	ttgtcaaact	aaataaaacc	cacatctcaa	10080
agacctaaata	tgatgcttgc	cttgaatct	ataatgataa	atgtcagatt	ttcagacctt	10140
aggccttctt	ttatccaact	ctttttttgc	cctcgggttt	ttgcaagccc	cctggtgttt	10200
agacatgtga	ccctttatct	gcttacagtc	taggtgttca	agggtgactt	tttttttttt	10260
tcttctgtta	aggaagtcaa	ccgtagccac	ccagcacata	gtgagaatat	gtcatggtca	10320
tggttatatt	ttggcaggag	agtctctgt	ttgaggtttt	caaataatcg	atgtaggcca	10380
gtgaagggtg	gtagagaggt	tggtgtgagc	ttggttgggt	gtgttgggtg	tgagcttgg	10440
tggtgtgttt	ggatgtgtta	tagtgtgctg	tgcctgtcc	aagccagtga	agaaccatc	10500
caccattgca	ggtgttgcct	gtctttttgc	atcttctcct	gtaatgccac	catccatttg	10560
cctgcccagg	gagcttagtg	ctgggcttcc	ggtggctgt	atggcaggga	gtccacaga	10620
ctgtgctggg	gtccagacta	gtggaagagc	tggacattca	tgtgcatgg	tcctctaaga	10680
ggggcttgtg	atggcagagg	ctcagggtga	gatcgtgtcc	ttcaactcag	tccttgggct	10740
aatgatggtt	tccatgaaga	caccttagct	cctgctcttt	gctccgtgcc	ttgtgataag	10800
atgctgaagg	tgcagatgct	gagagcgcca	ggcctttatt	aagtgcctgt	aagcggctca	10860
catgtgctag	ggatgttgac	aaattgcccc	ttcccaacaa	acaggcagat	cccaggatcc	10920
catttcatga	ataaaaattt	tgcaattctt	agagatgctg	tggtttccgg	acaccttcac	10980
agtgaccaca	caccaccctt	ttaggtgaac	taattggtgg	aagatggatt	tcacagctca	11040
ttcctccttc	ctcagcaaga	ggatagatat	ttgatggagt	gttaggcacc	cctctgtttt	11100
tttttttttt	tttcaccctt	actttggact	ttaaacttca	gaggacaggc	tggttgttct	11160
tgtttctcct	tcacctcccc	acacctactt	ccttaagtcc	tttgaagaag	agtttcaggc	11220
aataaaaatt	ttctagcact	tatattctgt	agttctggtg	cgatgtagg	agttggtcca	11280
taccctctgc	taccgtgggg	accagtggga	cacagcacag	agtcctagac	gctgacttat	11340
gctgagtcac	tggagaaaag	ctcagaacaa	gaagggccac	cctgctcctg	cctgactgtt	11400
cctcatcgtt	aggtcttctt	ttcctcggat	cctccagacc	ttagcttcat	tgagttgctg	11460
ttttgatagc	atltcaagct	tctcctttca	gcatttctt	ccttttgcaa	caagtgagg	11520
aatcaaaggc	cacctggact	ggactacctg	gaactcctt	aggctgtggt	catgaaaagg	11580
acagggtggt	aggcctttcc	gggaactttt	ttctccagag	attagggact	cacctatctt	11640
ctctccatct	ctatctcctc	ccctctcccc	caggaggaaa	aagaaaagaa	aaaaattcca	11700
agaacgagaa	gtgtggccct	aggggcaaaa	gaagccagga	aatgaagccg	ttttaaaagc	11760
cagcaaagct	cactttgggtg	actttaaaaa	aaaaaaaagt	gacctctggt	cgcagctggg	11820
tatggaggtg	acagtgactg	actaggatac	tgatctttgt	agaggtcatt	tgtgaaatgg	11880
gtggggatgc	tcagagacag	caggtatgaa	gtaaaggcaag	gtcactgctg	aagggaaga	11940
cccaccaca	tcagcttccc	tcagagctgt	acagcctttg	catataacga	ccacttccca	12000
ggctggtaga	gagaagatgg	catctctaga	tgtgcttttc	tagtctcagg	gtaattagtt	12060
cccttttagt	cagtttccca	acttattggc	tgattgggtg	acctagagtc	tcattgagcc	12120
cctgaacgtc	caattcttct	gtctctacct	cctgagtgct	ggaattacag	gcaggacca	12180
ccttagccag	ttcccatctc	atctttgttg	tagaaaagtg	ttcacccttg	aaggggtggc	12240
cagtctgagg	aagctgcacc	gcgctgtagc	ttccccttga	cgtctcttcc	ctggcacttc	12300
actctgatgg	ctttcttggc	tagccatcat	ggaggcaagg	aaatggccag	ggctgagagg	12360
ccagaaaacc	cctgcttctc	ttgggcagag	taatgatgac	ttccctgcct	ggcacagtga	12420
cacacctgtt	cctcgggaag	ccacaatggt	tgggcacctc	gcctggatct	tcctagactc	12480

```

agtggctgag tctgaagggg gccacttttc agatttgctt gctttctgaa agccttcctt 12540
ccaggcaaaag ctgaggtctg tggggcagga gaggaaggtt aaccatggtg ctgccatctt 12600
aatttggaac ttccaagcag atgtggcttt cagtccctct ggatggcatc cccaggcaga 12660
ggcagagagt cctgtgtcca tccgtccgtc cgcccccc aacgcaaac actacagaaa 12720
agtgatcctt ctggctctgg cctacctttc taggtcctgt ggtgtaacc agctgggatg 12780
gtgtggcccc ccgctccaga acgatccctg ccctctcctg aaagcagctt tctgtgaggt 12840
cattgctgtc cagcaacttg cggaccattc ctccagcaga gattcccttc cagcttccat 12900
gcaggcctga gctaactgag ccccagcaac aggatcaaac ccattccaag aggaaggcca 12960
tctgttcctc agcctccagc tgctggccct tcatttgcaa ttggctggga agctttggag 13020
gggtcaggtc ctggggacac atctgcagtc tctgaatggt gttataactg gggctctgct 13080
gagcagagaa agggcaagcc ctttaataaa acttgctgaa caataccccc ccaaggtgt 13140
agagtcagag aagcaggagg cagctttgcc ctttagctaa ctctaacct tggttttgta 13200
gccaggccac ttgaaaacta tttctttatt cagaaagtac ttaccaagcg gagaaggag 13260
gggctgctct gaacaaggaa aatgtcatat aggatttggg catcgatctg ccccttaagg 13320
gaattagagg gcaaatatct ccacttgagt gtatgccatt tattgaatat ttacctcagt 13380
gtcaaagagg tgagcttgtt ccagatgcag cttgtaaaaga gccacaggca gcatgaagtc 13440
ctctcgaact tgccctctgga atgcagttca gccttgggaa caccagccaa tctccctagt 13500
tcattgcaag caggtcccca agctgtagct gctttaggtc ctgtggttct tgggcctgct 13560
tgtagtttgg tggttagggcc cttatttccct gcatccgggg ccctatgata acttagccta 13620
atgctctagg gactttctat agggaccagg ctgtaatcgg gcgtgtgact tcattgagtg 13680
gatgaatggc agttatgcag gtgttgtcca ccttggtttt attgacaggg tctctccctg 13740
actgggaaact taccacatag gttaagctgg ctgggtgagca ggctccctgg agatgtctct 13800
gtctcttcca ctacatctag tttttttgtt ttgtttgttt ttgttttttt ggtttttttt 13860
gtttttgttt tttttgtttt tttttttttt ggtgggttct gaggaattaa agtgatgctt 13920
gcaaggcaag aactttatgg actgagctat ctgtcagcta tctgtcagcc cagccccag 13980
aggtacaata acctgtgggc cctttggctc actggttctc tgaagcaggt attaggcctg 14040
gtctgtatga gacggagcct ccaggacctg cagatgttta gttccacttg agaactttg 14100
aggaatcctc gctcagggaa ggcgtgtata taagatgtga cagatttatt cacttgaaag 14160
aaagccctgg tttggagtca gaggcatgca agtggatatt ctcatggggc catcttaacc 14220
ctctgtctgac tcactctactg acctgttaga atcaggctgt gaccataaaa accaagcccc 14280
aggtggctcc cggctgggtc aatatgtctg cagagcttca cagatagcat ttgccctct 14340
gtgcacagag tgtttccctc cagtgtgctc ctacatcag ggtcagtgag ggacttaaca 14400
gaaagccttg ggttccctct ttgtgccacc gtttgccagt agctggcctt tctggtgtct 14460
cagggacaga gggggccggt cagtacgacc acgttcattt tggacagcag caagcctta 14520
gctttggtct tggacaaaag ggtttctgag ctggcgggtc catcctcagc tgggagccca 14580
atcacagccg ctctttgtgc ctctgagcca caggaggctg accctgggtg gaggctcta 14640
tgcaagataa acctcggta ttgcttcat tttccttctc cccaccttct cagaatgtct 14700
ccacagagaca gttgggtgag aatgaatatg tctcgtgtt ctacgtggat aaaacatagg 14760
ctgtgacatc atgggatgg ggtgacggca tgtgtcataa tgggaaactg gaaactta 14820
agaagagaca tttaggtttt gaaaactgca caggagcctc tcaggtagag aaacagttta 14880
ggtacagggg acagggacag gggacagagg acagacatac cgtctggcta ggcaagccac 14940
catgtgaaatg aacgggggga agaggggaaa ctgggggaat gtggtactcg gtaatgatgt 15000
aaagatttcc tagagagaca ctctattatg gttgggtaca ttccattcag gcctttgctc 15060
ctttaggagc cctatagca ttccttgatg ttgtagctac gaggagcagc aaacctggcc 15120
caaaagagat tcaacagact ttcccagtgg cttttgtctg cctgtggatc cagccctaga 15180
tggcaaggtt tgggactagt gtgtcctaag gagtccctga gacctgggg agcctgtgct 15240
ttctcttgca agtgocctt caggacgag gaggcctgg cctggctggc cagacctcgg 15300
atcacagcgc ctctttgtgc ctctgagcca cgagtgtctg gtactttgac ataacttga 15360
atgccagttt ctacttctg ggtgctatgg aatctaattg ctgagttctc tgggacatgc 15420
tctctcagaa caaaaggttc catttccag ttctgtctca agcaaaagcat caacagctag 15480
gggattttgt tagctgcgca gatttgatct ctctcgcgt cttggtggcc cagtgggaat 15540
ttcagctctg ggagtgtatg aattgagtgc gtatgttgtg accaggcgc tctgtcattt 15600
ggacactatc gtcgcatgac aggatgggg gggagagagg tgcgggtgg taaggagcta 15660
agctgcgcc gctttgagtc taggtaccgg gtgacacaat gattcttagg cccttttgcc 15720
ttttctgat ttttattttc tctgggctc aggcataatt tgtttcaaac tggagggctg 15780
tccacctgt ttctcaaagc caaacctaaa ttacgagggg tgtgcctaaa tatgaaat 15840
gtaatggttc ccatattgaa acatttgcta ccttctagtc ctctccgatg ggcggctga 15900
gccagcccag agtttctggg gctgtccgac tactgcagct gaggtagcta ttggtgggg 15960
tgatgctaac aggaacgtgt ctgaagagat gctccagcta ttggttga acaagagcc 16020
tgggcagcct gctcacctct ctctctccc tagcctcacc atctgcct ccccccccc 16080
cttttttat gcagccgtat ttcttgaggt tgaaaacttc catctttgtc ctgtatgggt 16140
gttggcccc tcctctctt caggatgagt tgtacagagg ccttataagg atgctatcag 16200
gatgtgcaag ttggcacact ggtaagggg aaactttgaa agagttaggag ctgcagcagc 16260

```

ES 2 666 179 T3

cagctctggg atgtogtctt tgtgtctggg gacaaggcta gctaggccgc tcttctctct 16320
 gactccacca aaggaccca ttgtccttaa tatcctttat actgaactct ggtgccagct 16380
 ccattgctgac agtgccatgc aaaaatatgt acaggagagg ctctccaag gtcccagctc 16440
 tgccagggtg caccggtttc taaaagccta ggtggacatt ccagtacat gtgccctgca 16500
 ttctgggtgt ccttgattg aagttacaaa gaaccttca agttctgtac cctgttctat 16560
 ggccagtgac cacagctcac caggccatg gagtggcagg gcatctttat ggctcggagg 16620
 gcagagtggg tcaacccttt gccactcacc tgttatgaac ccagtgtcct gtgactttgc 16680
 agtgacattt ggcagctcga tccccattct ccgtcaagac ttttggcagt cctgtggctt 16740
 tgctgtttat ttgtcttga ttagatggca ctgtctggga gaacgccggg ccattggtatt 16800
 gtccctgtcc cagggttcct gtgcagtct actgggctag aggagtgtg ggaggtgggg 16860
 acagcttagc tgggcagccc cgtccctga caggacatgc ctgtgaaac tgtgccttct 16920
 cctccacctt cctcctcccc tttccctcct gctcctctc tccctcttc tctcatcgt 16980
 ccctgttctt cctgttagct ccctctctg ggtgaatcta ctctgattct gctttgtct 17040
 ttccagaaga atgtgtttt ggatctgatt gtgcctgtg gggagcccc ctaagtgggg 17100
 ctggttgagg taccacctg tatctttaac tcagatcct tagacgtga ctaaagaagt 17160
 cattctgggg acacctaga agtggcttg tgtggtcga ggtgattgt tgccccagag 17220
 gtggttgcta gaagtggct cttctccctg cgtgggtgg aagctgcat gtatctgtg 17280
 ggagacgatt ggccagggca ggacttgac gccatctgt tctctgtttg cagttggggg 17340
 ccatttcaga aaccacagg gaaaagtta taggcaaca tgataaaaag tgacagtctg 17400
 aagtgtctct atcgtgtgct tggcaacta aagcattacc tgaagcagct tctaacttc 17460
 agacgtccta gctgcaacgg gaacccaag atggccatcc tgtgggcgt ggggaagatt 17520
 tcgtttgtgc gcagtgggt gtcttagct cggccccat tacttctga aggtccctt 17580
 tctaggtgta ctccacgaat agcaaggtg cataccctc cccctagct tacaggaag 17640
 taaatacaag ctgtcactag tgacatcagg tgaggctcca cccagaggtt gtgacctact 17700
 tggatctgta gaaggactt gagaaggctc aggaagatt tgcctcagtt tcccttcgc 17760
 ctgggtctga agcccctctc atttctaat ccctattacc tcccagggaa tagtggctt 17820
 aggaatcttt ggaagaaaag agggctcatg gcagggtaac agtcagccac gtgtgogaa 17880
 ttttaagacc agaatctcac tacatagccc aggtggctt tgactgccct cactcagta 17940
 cttagtaggc ggtaaaactc gaagccgat caggctttga acttatgac ttctaacc 18000
 accatgtgcc accatcccc accactgtt atgtttctat tattggatt gatgctgta 18060
 aggaaccctt ttatctttt gtttgtttg tttctgagta tcagagtgt cagctcact 18120
 aaaatagcac caaatgaaa aatgtatgag aacagactt ctgtctgctg caatgtggtt acctgtgg 18180
 agaaacacag ctaagcctcc acacaggaga gcctctggcc actgttgtt ttgtcgcagg 18240
 tagaaacagc tgagcagagc cttcccagaa agtaaacatg tgcctgtt ttgtcagaga 18300
 gtttaggtaa caatgacagt gtatggcca gctcccact atctttcaa gtttccatt 18360
 aattatgaaa aatgtatgag aacagactt ctgtctgctg aaacccctga aagagcatt 18420
 ggtgcctctg ctctgactt ctggaactt ctcccactg tgctgtgag agtgacagag 18480
 gtggaacttg gaagcgtgtg ctccggtaag ccacggcatc agaaatgta aatccagga 18540
 atgttgatat tgctataaaa gagactgtt ggatttcca gggagttcct tgtcctgtt 18600
 caattgtcac gtgtacaca gagcagctt gcagagtgg gcaaggagt gcctgtgtg 18660
 agaggccatc tgagtgggag agacaggtg ggtgtggca gcacagctt tgggtcctt 18720
 gcccttata ggacactatg aggtggttac aatagggagt tgtaacacca caggactct 18780
 aagagcagcc agtgattggg aggagccagt cccgaagcct ggtgaaggat ttaggcacag 18840
 aagagaagcc tttagctctc aagtctccag ggctaggcgg gacaggatg gcatctttc 18900
 agcatgccac ttgggtcca tgttcttagt gcctggctc gtgatgtat tcatgtgtg 18960
 tccatttgca gggagctacc aactgcact gtgtcctgg atgctgttg gtggctttt 19020
 tcttctcacc ccttattat aatcctgct tctcctgtt cttcccctc tacogtatt 19080
 gacctctcc tttctttctg cccttcttt tctgtattc acccaatct cctactcct 19140
 aggatcacca aggaggagt aacattgct tctgtgacg ctgtgacct ctagtgggg 19200
 cctcttgaga gaaggtcact agggagtgt gcattctgcc tatccaagg agatacctt 19260
 gaggaggcct tggcgttagg atggctgat tcatagata cttatcttc tgacgtgct 19320
 gcagatgata ctctatactg tcccaaacag cagtctctt cctgggaaac tagagattt 19380
 ccattttgc ccatgccaac ctggcctcac cattgactga gtgagatgg agccatcag 19440
 tgaagtctt gagattaaa atccagtgt ttctgaagac agtggagcac cacagttata 19500
 gcttgagaac aacggcggat gactgacatt ggttggctt ggaagatcaa gtatacagcc 19560
 ggtggtccc aggcacctcc cgtataatgc cttctgtat gttggtggt ggggatctt 19620
 tggctgagag gctatgcagg gcagagagga aatgagccca gtgtccctgt acccagggca 19680
 gtgtcccttt accaaaatc cagtgtcctg tctacctga gaccctctt cttctgtgt 19740
 cctcacagca tgggtataca gtatggtaga attggtccag catggtccag tagtgcagct 19800
 aaatttcaat gagtcttgg ctttgttga tgttgggtg aggaagggt tctccgtgga 19860
 tgggtgtagc ttaaggctc catcattctt aacattgtac gaatcttgg ttaaaagatg 19920
 ttaagaccag actggcagat ggtatgagac ttaggtcaa atggaacccc ccttcccct 19980
 ccttatttct cttcctcatc cttaaaata tgaaccctt gtttacttg ttgtgctgt 20040

tggtcattat	tctcagtggt	agtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgatgt	20100
gtgagtgcac	gtagacaaat	cagccatgct	gtgtaaagg	cagagtgcag	tctggtagag	20160
tcaattcctt	cttctctctc	taccttttta	agggttctg	ggaaccaaac	ttggctttag	20220
gcaagcatgc	cttctctctc	ggagccatca	ttccagaact	gcctctccct	cattcactta	20280
gccactcagg	tcagctgcct	cttggtttaa	atgggcagg	aaaggcctga	gctgagacct	20340
ttccaactga	attctcaatg	tctttcaaac	ttggttctgt	gtagtgccac	aggggtgtctg	20400
ctacttcttg	gaggagactc	ctatcccctc	ctgcaacaga	agctgaaaca	ccttcgggtga	20460
ggggccaagc	tatcagtggt	tggggcttgt	agaccatgag	atTTTTTTT	tttttcaat	20520
gactctggtc	tgcccgtata	acacaaaagc	agccctagac	aatacatacc	caagtatgta	20580
ttgagatgg	cactgctcca	agaagtcttt	gtttacaaaa	gcaagtggct	gacttgtccc	20640
tcaggccatg	ctttgctggc	tcctgctgcc	cacggggcct	tcgcccaccg	tttccacatg	20700
aacggctacc	tacctgcctc	acccttaaga	ctcccttaca	cacttctctat	tttctctgag	20760
gtttttcttc	actttcattt	gccccactgc	aatggagggt	ccaccagggc	agggatattg	20820
ctaccctcct	gttgcttctc	gagtgtagag	aacaaaagctt	ggcctgtggt	aggtatgcaa	20880
taaacagagg	gcacatgaga	taaacaaagc	cttgaacctt	tacctggctg	tcagttgggt	20940
ttgctttctg	ccctgctttt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	21000
gtgtgggtgt	ggtgggggtt	gtgttccatc	aaacttctgtt	ttcttcccta	tgtgggtttt	21060
acttttctgt	tctgtactgt	ttaacatctg	tgcccctctt	ggctgtgtgc	atttgaagtg	21120
ggggtcccct	gtgagaagcc	tcaggcccct	tgtgttggct	gctgctgcgc	ttcttggacc	21180
agatgtttat	taaatagcag	gactgaaaca	tgaattgact	gtattctagt	cgtgagagaa	21240
tttgttcttt	ggagtgggct	ctgggcagaa	taatcgccct	gtgatgctgc	tgcccagatc	21300
tggaaacctg	ccagtgtggg	gaaggaaagc	ttgtgttttc	caggcttggg	actctgggtg	21360
ccattcagag	ctctcactgt	gggatgaaa	cttatttcat	gagccctcgt	ggccacctct	21420
ggccctgagc	aaggtcagga	gcttcccttc	tctcactttt	tttgggagaa	gctgggaggt	21480
tggatcatag	ttggtttcat	tctgcccctg	cttttagagga	aggcaatgtc	tgcccttctc	21540
gtgtacagca	aagatatcca	gtgtagggat	gggctgtggc	acaatgacct	atcagaactg	21600
agctttctga	tgtgaaggtt	tcctctggaa	gtcaggacac	ccataggcaa	tgtgtctatt	21660
tcagtgtttg	gaggtatagg	gtaggcagat	ggactttaga	gtgggagaga	cccccttagt	21720
ttccagccag	gtgactgatg	cagagtgatg	gatcatggag	ggccatggtt	gacctgggca	21780
tcagaggagg	aactgggcta	aacgggagtg	agagggagga	ccttgtgttc	ataaagaaga	21840
gcaggatgct	tgacggagat	cagggactct	ggggtagtgg	tgggttgggt	ggcaggatgg	21900
atctggctcc	accagtggaa	tgtcgggtag	tagtacaatg	tacttatcca	gtacatgtag	21960
tctatgtgta	tacatggctg	gtttatggta	tagggccatt	aagtgccagt	aattccttac	22020
ttttctttct	ttggacgtta	aaggaccccc	agcatctgtc	atTTTgagga	agatggaatg	22080
tcccagctcg	ccagaacag	atctagctca	gtcctgatcg	ggccccaaaga	gcacataaaa	22140
acaatcaagc	caatagctgc	ctcttcccaa	gtgggtgaaga	gtaattttgt	agatgggtct	22200
gttttgcctc	tgaatttgag	acattttatt	tatatgaaa	agcttgggtc	tgtgagaaca	22260
ggcaaaagtga	aatatgaata	agttagctaa	tcagtgtgag	aacgtgtatg	tacgtgtgca	22320
tgtatcacat	atacagtcac	gctggatggc	tagcttggaa	atcaacttta	cagttttctt	22380
gtggattttt	cttccacttt	agggtttggt	tggtttttaa	agccctattt	ccagtatgtg	22440
gaaatgagcc	aacccaggac	agcttccgct	ggatcgtgga	cagcttctat	ggcctgcagc	22500
gtgtacaactc	gagataaactc	cgtataatgt	atgctatacg	aagttatatg	catggcctcc	22560
gcgcccgggtt	ttggcgcctc	ccgcccgggc	ccccctctc	acggcgagcg	ctgccacgtc	22620
agacgaagg	cgacgagcgc	gtcctgatcc	ttccgcccgg	acgctcagga	cagcggcccg	22680
ctgctcataa	gactcggcct	tagaacccca	gtatcagcag	aaggacattt	taggacggga	22740
cttgggtgac	tctagggcac	tggttttctt	tccagagagc	ggaacaggcg	aggaaaagta	22800
gtcccttctc	ggcgattctg	cggagggatc	tccgtggggc	ggtgaacgcc	gatgattata	22860
taaggacgag	ccgggtgtgg	cacagctagt	tccgtcgcag	ccgggatttg	ggtcgcgggt	22920
cttgtttgtg	gatcgctgtg	atcgtcactt	ggtgagttag	gggctgctgg	gctggcccgg	22980
gctttcgtgg	ccgcccggcc	gctcgggtgg	acggaacggt	gtggagagac	cgccaagggc	23040
tgtagtctgg	gtccgagagc	aaggttgcc	tgaactgggg	gttgggggga	gpcagcaaa	23100
atggcggtcg	ttcccagatc	ttgaaatgaa	gacgcttgtg	aggcgggctg	tgaggtcgtt	23160
gaaacaaggt	ggggggcatg	gtggcgggca	agaacccaag	gtcttgaggc	cttcgcta	23220
gccccgaaagc	tcttattcgg	gtgagatggg	ctggggcacc	atctggggac	cctgacgtga	23280
agtttgtcac	tgactggaga	actcggtttg	togtctgttg	cggggggcgc	agttatggcg	23340
gtgccgttgg	gcagtgcacc	cgtacctttg	ggagcgcgcg	ccctcgtcgt	gtcgtgacgt	23400
caccggtctc	gttggcttat	aatgcagggt	ggggccacct	gcccgtaggt	gtcggtagg	23460
cttttctccg	tcgcaggagc	cagggttcgg	gcctagggtg	ggctctcctg	aatcgacagg	23520
cgccggacct	ctggtgaggg	gagggataag	tgaggcgtca	gtttctttgg	tcggttttat	23580
gtacctatct	tcttaagttag	ctgaaagctcc	ggttttgaac	tatgagctcg	gggttggcga	23640
gtgtgttttg	tgaagttttt	taggcacctt	ttgaaatgta	atcatttggg	tcaaatgta	23700
atTTTcagtg	ttagactagt	aaattgtccg	ctaaattctg	gcccgttttg	gcttttttgt	23760
tagacgtgtt	gacaattaat	catcggcata	gtatatcggc	atagtataat	acgacaaggt	23820

gaggaactaa accatgaaaa agcctgaact caccgcgacg tctgtcgaga agtttctgat 23880
cgaaaagtcc gacagcgtgt ccgacctgat gcagctctcg gagggcgaag aatctcgtgc 23940
tttcagcttc gatgtaggag ggcgtggata tgtcctgcgg gtaaatagct gcgccgatgg 24000
tttctacaaa gatcgttatg tttatcggca ctttgcacatg gccgcgctcc cgattccgga 24060
agtgcctgac attggggaat tcagcgagag cctgacctat tgcactctcc gccgtgcaca 24120
gggtgtcacg ttgcaagacc tgacctgaaac cgaactgccc gctgttctgc agccggctgc 24180
ggaggccatg gatgcgattg ctgcgggcca tcttagccag acgagcgggt tcggccatt 24240
cggaccgcaa ggaatcggtc aatacactac atggcgtgat ttcatatgcg cgattgctga 24300
tccccatgtg tatcactggc aaactgtgat ggacgacacc gtcagtgcgt ccgtcgcgca 24360
ggctctcgat gagctgatgc tttgggcca ggactgcccc gaagtccggc acctcgtgca 24420
cgcggatttc ggctccaaca atgtcctgac ggacaatggc cgcataacag cggtcattga 24480
ctggagcgag gcgatgttcg gggattccca ataccgaggt gccacaatct tcttctggag 24540
gccgtggttg gcttgtatgg agcagcagac gcgctacttc gagcggaggc atccggaagct 24600
tgcaggatcg ccgcggtcc gggcgtatat gctccgcat ggtcttgacc aactctatca 24660
gagcttggtt gacggcaatt tcgatgatgc agcttggggc caggtcgat gcgacgcaat 24720
cgtccgatcc ggagccggga ctgtcggcg tacacaaatc gcccgcaaaa gcgcggccgt 24780
ctggaccgat ggctgtgtag aagtactcgc cyatagtga aaccgacgcc ccagcactcg 24840
tccgagggca aaggaatagg gggatccgct gtaagtctgc agaaattgat gatctattaa 24900
acaataaaga tgtccactaa aatggaagtt tttcctgtca tactttgtta agaaggtga 24960
gaacagagta cctacatttt gaatggaagg attggagcta cgggggtggg ggtgggtgg 25020
gattagataa atgcctgctc tttactgaag gctctttact attgctttat gataatgttt 25080
catagttgga tatcataatt taacaagca aaaccaaatt aaggccagc tcattctctc 25140
cactcatgat ctatagatct atagatctct cgtgggatca ttgtttttct cttgattccc 25200
actttgtggt tctaagtact gtggtttcca aatgtgtcag tttcatagcc tgaagaacga 25260
gatcagcagc ctctgttcca catacacttc attctcagta ttgttttgcc aagttctaatt 25320
tccatcagac ctgcacctgc agcccctagc ccgggataac ttcgtataat gtatgctata 25380
cgaagttatg ctagtaacta taacggctcct aaggtagcga gctagcccac cttgccttga 25440
gaatggctgt cgccttttgg ttcccttgggt tgtgctatga tgcgtcagtc tgggtgtgcta 25500
actctatggc ctgcttatct gttcctcctc ctgtgatctg caatctagcg cctggaagag 25560
aaaaggagat tacagcttcc ccagactacc tggagatagc tatttactgc ataggggtct 25620
tcttaatcgc ctgcatgggt gtgacagtca tcttttgccg aatgaagacc acgaccaaga 25680
agccagactt cagcagccag ccagctgtgc acaagctgac caagcgcac cccctcggga 25740
gacaggtaac agaaagtaga taaagattt gaagaaattt actcccctcc cccaccagc 25800
cagctcctgg atcttcttcc ctctgatttt cccctaact tctgtgagct ccagaactgc 25860
aggcaattct aatctgccac tgtgtggagg ttcagtcagc ggttgggact aaagagcatt 25920
aagtcacaat gctgctctgg gcttggtagg ctggctctgg ttttaaagga caagagtgtg 25980
aagactggag ctgccagtg ggatgggagc aggaggccat gccctctgcg cccctcaagc 26040
tcacggctcc tttgggagaa caagcatttg gtctggctcc attgcttctg tatgaggcca 26100
gatgttcggg ttcaagtttt acccttcata ggaaagagag ttaattttc tttgatttac 26160
tatttttaag agagatcaga aacagaggat ggaggtatac ctgaactaat gcttgcataa 26220
aagtggctctg tgatgtcttc taaactgggt tttggctgat tttgtctggt ttttaaacg 26280
ctgtatgcgt atagtttatt gttacaggtt tggtcagggg ttcagtgata ggatgattgt 26340
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtttgt 26400
atthtagtta taagtacatg tgtgcaggtg cctgtggaga ccaaagaga gtgtcaggtc 26460
cactagtgtc agagtttatc agcataggtg gtaggattg aactctggtc ttctagaagt 26520
gcagcaagca ctctttaacc actgagccag ttctccagcc cccagatacg atgattgcta 26580
tgtagaacag ggagaaaatt acctttaacc ttgagcttga tctttgatgg ctggctttgg 26640
gggaggttaag gcaatagaac ctccctgtg ccataaaaaca aagccctca aaggtggata 26700
aggaaaaaat gcttgacttc tgtacttgct cctggattcc cctagctgac catgtgtggg 26760
tgtaaatctt tatgataaga ttcggaactt gattctgata agattgtcac tattttttt 26820
aaattagcaa tggaaatgaa caacctggcc tgtgctatgg ggaggtgcat cttagtgttt 26880
gttaaaactg catattcatt agtttcaacc ctagaaattc ttatttagta cttcttgaat 26940
ggatctgtaa gactctgcat tttaaacact ttctcgggtg atactgtgta atacctaag 27000
aatctctggg ttcaacccaa ccctgccttt cctgggccct ttctgtggac aaggtgggaa 27060
ctagcaggtc agtagtggtc tggacacagg gccttggctg ttctcaacct agcttcacac 27120
tcagaggtga gcaggtcctt gtcaacgtcc ttgaagcctc gttccaaca ggtgtttctt 27180
gacaggggtc aagatataat ttggagcata tgtcagatgc agcctttggc cagctgttga 27240
atgtggagt caaaaggctc agttgggttc cttttaatcc tgagaatgct gtgctacttt 27300
gagtgcacc actgtcattt gtgggacct agaagctaga tgggtgctgaa gttaaagttg 27360
gttgacctgaa tgagttgctg gaagagtct taataaaact cttacctggc tagatagtgt 27420
taaggcttca ggctgaatgg ccaactcctt ggccactcct ttggctactc cttcacagcc 27480
tctcctgatc tcttagccct gggccattct taacatctgg actctggtct agggagtta 27540
agtaaagga gcaatgtcct gtctcatttg tttttataat agagaaaaga agtaaaatcc 27600

ataagttgag gtagataggc cattgaccct taattatrtc atcatttaaa aaactgatgt 27660
 gtgtgtgtgt gtgtttatac atgtgtgctg gagctcctgg aaagctggaa gagagcactg 27720
 gatccctcag agctggaggg ccggtagttg tgagctctct gatgtggtgc tgggactaga 27780
 actcgggtcc tctgcaagag cagcaagcgc tctaatacag tgagccatct atctctccag 27840
 cctgcatcac ttttaagaa aactctttct atttctccat tttccatttc catttccatc 27900
 tttttacatt tataatattac atttatataat ttatgtagct tgggcacgtg tgcgtttgtg 27960
 tgggggcatg cacatggcat agcaagggtta gtgaaggctc ttttgaatg gtggccaggg 28020
 gacaacttac atgggagcca gcacctttct ctaccacgtg attcccagga atcaaacagg 28080
 tcaggttcag tcaggtcatc gtatctgatg accgattttg ttgactccat cgcttttaaa 28140
 gaaaaaaaaga attaacacct attacagcgc tcttcctttt gcttcatgtg aaaaagacag 28200
 agggcctgga gttcccaggc acatggattc agcatgtcct cctttctggt tgtccaactg 28260
 agtttcttca ttttctgtcc acctaagctg tccattttgt ttgtttttat attcctgtg 28320
 tgaccggagg gaaaagttgt tttttttttt tttcatttac ctccctttct tcttgtattc 28380
 attgttattt actgagtgc aagtttcttt tagtgcattg gcctaaatca ggactcttgg 28440
 gctgggagtg tggctcagtg ctggcatgct cgctagctt gtccaagcct cagtatcacc 28500
 aaagaaaata attaagccag tttcgtgtca gagaaagcct gccatttgc cactgttgc 28560
 aaggttagtg aaggtccttt tgaaatgttt cttcatgctg acgctggata acaaatgtgt 28620
 gaggcccagg ctctctgcat gaggaagcct ctgggagata aatgggttga aaaggactg 28680
 ataatacccc agcatttctc agaagtcatg gggagatag gtactaactg cctcttcca 28740
 aaggatttcc caaagcttag gccactggga ggaggaggag gaggaaaagg aggagaagg 28800
 ggaggggaga tgcttatcat gactctggat aaagaggggt ttgggctgta gctggaggcc 28860
 tgcagatagg ttaatgacag agtgaattcc tcagggatgc caagcatgcc ttacctggcg 28920
 acagatgagc ctgtaatcag atgtctggag gacgggtgct ccaggcact aagaggctag 28980
 ccctttattt gtgtaggccc aagcttctat atgatgcagc atccatgcc tggccttgc 29040
 ccaggacggc gaggagcgc atagcctctc tccattcact ccatcttgt cttgtttta 29100
 gaacgagaaa agttggtttg tttattcatg ctgttttttt ccatgtgcac aagcgcgtg 29160
 tcggaaagtg tgtaggtgtg tacagaaagt gtgctgaggc caaatgataa ccttgggtgt 29220
 cattcctcag gtgccgtcca ctcttactc tttgtgtgtg tatgtgtgtg tgtgtgtgtg 29280
 tgtgtgtgtg ctaagtttct cacctgcttg cctgaactag ccaagtaagc taggttatct 29340
 ggtctgtgag ccccagggtc ccaattgctc ctctctctcc tctctctgt taggattcca 29400
 agtgtcggtc ccaagcctg actcttcttt tttttctga gacagggttc ctctgttag 29460
 ccctggctgc cctggaactc actctgtaga ccaggctgg cttgaactca gaaatccgc 29520
 tgcctctgcc tccaagtgtc gggatttaag gcatggacca ccaccgccg gcttttttt 29580
 tttttttttt tttttttttt taatatagct cctagggatt gctctcaggc caaggaaggc 29640
 aggcgttttg atccttcttt ctctttagg tttgcttccc tgcctgaac ttgtttaaa 29700
 caggcatttc actttaaaaa ggtagggtct tttttttgt ggggtggagg ggtgggggt 29760
 gttttctgta tcaataaat tctttatagt ctttctagta aacattaatt ttgggagaca 29820
 ttgtgcttgg agtaagatac gcaacttttt ggtgggacag cctggtaggc agcctgtggg 29880
 atctctaagg aggagtcatc tctctcacc aaggttagga ctgggcactt tgtaagcgtc 29940
 tgcgcacttg cctctacttc ttggtaccta tggttaaatg gcaatagtca gtctagagaa 30000
 gggcaccttg tgaccaact ggaccatcag tggctactgg gcagtggctt ttgtgtactt 30060
 ctgagtccaa gtggaaagat ttgccttctg tgatttccac aagtccctg ttggggaggc 30120
 gggccgtatg tgagtgcaga gcgggggtga ggaagccttg ttctgtggag tgcttgtttg 30180
 tggaggagct ttctgggta ggttcagctt ctttctggag ccagaagttt gcttagggca 30240
 agatggagat ccatctgtct gtgtccagat gactgcatac cctaccgat cccccagtct 30300
 cacacaggac tgtagttagt ttgttcccag cctcagccat tgacatgggt agctgagaaa 30360
 accagagagc aatttcataa tgcgttttag acccatgggt attcagggtg ggctgggggg 30420
 aacactaat tccagagctg ttctcagggc aatgtattcg tggcttaga gatatgaaa 30480
 ctcagtgaat gtgagtgctg actgcttagc atcccagcac cgtgacctgg aatctccatc 30540
 gtacgaggtg tagtgcattc agagttgcag tgtaccggtt ctgggaaaca tttgggcagc 30600
 tggatagttg tggatgacct gactggagtc ttctgtttcc tagggatcga tcgcttctct 30660
 tgtcccagct ttggcctgtc ttttctctca gccctgaaa gacatgctgc cttggctgag 30720
 atccacccta gacttttctg gatgagctat aagtaggttc agaacacctg agtcaggtag 30780
 ttttactgtt gtgacagggc attcaagagt ccagaggagg tagaagctgt ctaaggggca 30840
 gtgtgagcaa ttacctagat tttgttatgg aaggaaaaac aaaacaaaac aaaacaaaac 30900
 caaaaactc cactcccaga aactctctga agcttgggtg ggtgcagggt tttctgtgtg 30960
 ccatagaggt gtgtgggct agacttaaga tagaacacac tggcctctg ttctgatgtg 31020
 gaaggctcca tctgctgctt gggagtgcga ggtgtctca agtctgctgt agtccaaggg 31080
 catgtgtcaa ttctcaggaa taaagacaaa cttgactcac ccttccccgt actgtctttg 31140
 cttccgctgt cgctgttgtc tgtgaggtcc cctctgaatg ttccagcttca tccagcataa 31200
 agggagacgg ctatgacttg gtggctcttt aaaagaaaaa ggggagaaaa cccacttctt 31260
 ccgttaatct cccatagtga ccgtggaaat atatgaaaag cacatttagt taaaagcttg 31320
 atttatggca cgtggtaaag agatccggc atgtgaaggct gccgaattgg agactgtgaa 31380

gagtgTgcgg	ctttctaaaa	accgcctgcc	aagatTTggg	gtggggaatg	ggggTggggc	31440
ggagcaacag	tttactacag	tgttagcgtt	tattgtttat	aagtgaactt	ctaacagtgg	31500
gatgttttta	agtgcgttga	aagggaact	ccaaaatgga	agtttctaga	ttaaggattg	31560
agaactatct	gaggagggaa	gttataagta	caagagaaag	agaagaaagg	aagtctgtaa	31620
tacagtggTg	tgaggaacct	tccaaggtgg	gcggtggggc	cacaattcag	agggaaggag	31680
ccctgaaaa	gccaggctcc	tccagggacc	cctgctgggg	atTTtgccaa	gccctccaga	31740
caggtTgcct	ttctgaggag	aggcgagtga	agagaaagc	agtcAtgctt	tatagcccca	31800
gagaggatTT	taaaagtata	gtaaaacgca	tggaggtaga	attaagatgg	acctctgtag	31860
acagggagag	cagagtgtat	gctcagagac	ttTgctatt	ttcttaccct	ttccactct	31920
gggtgttttt	tacaaaggta	ttttccaggc	ttgtacattg	aacctgaatc	tgcactgtgt	31980
attgaacaaa	attcccacac	atgaaggcag	ttttacattt	tgataccaat	gtgcagcaac	32040
gactgccaag	gttttttttt	tttttcttcc	gtattagttt	agtttttttt	tttttttttt	32100
tctccccgtt	ttccattttg	aaaatgtTgc	cctTaaaacc	ttgtggaggT	gctctgtTgt	32160
ggggTgggta	tgcgTatggg	aaactTgcac	cccaggcctg	Tgctgtgcat	tctgtTtggg	32220
tcaaaggTcc	tccacagagt	agTtgatgTc	agactggatg	gTaaatctct	ctgtTttgag	32280
gtAACcccta	agTcatgTtc	accagcgTga	cctTgctgct	ctatggtTtt	cttcttctcc	32340
tctaattcct	acattaaaaa	tatatatagt	cTtgcttact	ggaactccag	gctatcctgg	32400
ctggcagTtt	agggtcccat	ttTgtaaatc	agactcgcaa	ttcaggtgta	Tgccatctaa	32460
aatcagaaca	aactcacctt	gtagagcaga	ctggTgagct	atggctgtcc	cagctcagca	32520
ataagcactt	gatgctgtct	tcaTtctgtc	ctgctaactc	Tgagaccacc	Tgagactcac	32580
atagaccccc	ggaatctgac	cTtgactTca	cgTtaccatt	gaccaggatg	tagcctgcca	32640
gggcatcTtg	gccctgggtg	atcaccaggt	cacacattga	aggatgcgga	aacatcacia	32700
aacagcctgg	ggtggggggg	acaaaaaaga	agtgccatcg	ggcgtctTgc	tagTttctaa	32760
actgaagtct	gcataattca	acctgtgccc	ttctTtctct	gctgttcata	tttattttat	32820
tccaaatgct	atTTtggtca	aagaaagaat	gtctactaaa	acacaaagga	aacacaagac	32880
cagggtaaata	aaatctatat	gatgtagaaa	gttctagaat	aagacctgtt	tcctaccttg	32940
ctccctattc	ttgatctctc	actctctctc	cgaaggTgac	caTgctaaa	tccttagata	33000
tctttccaga	aaacatttcc	TgctTtgctt	cccAagtctt	gatctctctc	cccaaagggg	33060
acctggctca	aacccttaga	tatccttcca	gaaaatgcct	gtggtcacia	cccatcctgt	33120
aagcctctat	gtgctgagta	ctgactccca	aggacaggcc	acagaagctg	cgatgtgcca	33180
ctagcctctg	gccattacca	tcattcagaa	ctgtggTctt	ctgagatttc	tcagatccc	33240
ctcctcactg	gcttagcac	acagtggTtc	ctaacaacta	agctaggaac	tttaggtccc	33300
agtgatgca	aggcaagctg	atgatggccc	tataaagagt	atcctggcta	cacacagtct	33360
ctgtTggctc	ttTgctccct	ggggTctgtg	ttgtctcatt	actgggcaga	cttttacttg	33420
ttTggctgta	gcttcttgcc	tctgattatc	Tggtgtgaat	ttttactata	tttctactgg	33480
gagatgattt	ttgcctattt	gtgtggaaa	actgccagaa	agatcttaa	aattaaaaaa	33540
aattacatgc	cttttgcaag	cataaactgt	gagcctgatt	cagaatgagt	caggTgggtg	33600
gttccacaga	agcactatgg	accagctcca	ttccagaatc	ttctgagtcc	cttgtctgta	33660
gatggagctc	acgatgtTtt	Tgtggccagt	ggaaaatgga	catcttgatg	ttgtcaggaa	33720
acttctggTt	tctgatgcag	cctgctcacc	acagTtaggc	Tggacaccat	gcccagcagTg	33780
gaaggggctt	gggagttatc	ttttgtcctg	ctgggatgga	atgcctattc	Tggaacaagg	33840
caagtgggtt	ctagaggcac	tgcgTgtTtc	cctgctcacc	ttccccTgct	Tgcttctgcg	33900
ttTgccttag	agattgggat	cctTgaatgt	atggctctct	attacagaat	taccaggttc	33960
cttctctctc	ttcttctttt	tttttaatta	aaaaaaaagc	atcaattttt	gTtgtggcac	34020
aaggagtaaa	Tgtcctgtct	gcatagtata	atgtatatac	agcttcttct	Tgggtacggg	34080
Tgagatggct	caatggacaa	aggcacttac	gctgatgacc	aagcctgacg	ctctgtagTc	34140
aatcttcaga	gcctatgtgg	taggaaaaga	gaactgacc	tcagaagTtg	tcctctgacc	34200
tccacactga	tatgcacaca	aacacatgca	cacagataca	ttttttttca	tttaaaaaga	34260
aaatcacctt	tctccttccc	aaaagatact	tagaaggTtc	agaaaagtcc	ttatgtgtat	34320
tttaaaataat	aagatttcat	atcaaaaatt	gcttactgat	tttaacattt	ctttgtgggg	34380
ttttttttct	ttTgaggggg	ggaggatagg	gtctctggga	ttgagctcag	Tgggtagcca	34440
aggataacca	tcaTgactt	taatactgca	aacacttttc	ttcaattcta	ttaagggtag	34500
ttgggtttcc	aaagagcaga	agggcttgcc	aatgggacag	tcagTcctgg	gaacaacata	34560
ggacctTggg	ttcctctgat	gagagtctag	gatccacatg	ggagagtTcc	ttTggcttta	34620
tctTtgccag	ctggattgag	gagTttgtat	actcagcagg	ggattgtcac	ccatgtggga	34680
gctggaaagc	Tggtgtgctt	gctgagTggc	tctTgtctaa	cctcacacc	atgtctccgg	34740
gaccaaagcc	tccgtTgtgg	tctgagTtga	aaagcagTatc	cagcagccca	ccatcacacc	34800
aagattgtgt	agTcatacc	aggcAcaggc	ttTgtgtggg	ctctgggtat	atTTtctttc	34860
gcagaaatca	gccaaaggaga	gacggTgtgt	ttcagagata	gacactgggt	ctgacacagt	34920
ctgctataca	tcaaggcaaa	ctTggTgaag	cctgtgtgTc	Tgctgggtga	gagaggacc	34980
ttcccgtgtg	gctctgagTg	aaagtatctt	ttccttaacc	ctTggtctcc	Tgtattcact	35040
gctctgcttt	ctgaagctaa	agTgacaaga	gtcagcccat	tttactata	Tggtctgggc	35100
atcatcaagt	ttcagaagga	ctggggagag	atggagaata	gcctccccgt	gcctggaact	35160

ctggatttct tgaataaaa acctttgagt taccagaatg ccctttccct gtgtcttagt 35220
taggatttta ttgctgcaaa gagacaacac aatgtaactt aaaaaaatta tttatttggt 35280
ttatgtatat gagtgcacca tgactctctt cagagacact agaagagggc atcagatccc 35340
attacagatg gttgtgagcc accatgtggt tgcctgggaat tgaactgagg acttctcaaa 35400
gagcagttgg tgctcttaac tactgagtca tctctccagc ccccagtgca actcttataa 35460
agaaacacac ttaattgggg cttgcttaca gtttcagagg tttagttcat tattgtcatg 35520
gtggaaagca tggcagcttc ctggcagaca cagtgtgga gagagaaga gctgagagt 35580
ctacatcttg atccacaggc agcagaaggg gattgtgtgc catactcttt gaggttgag 35640
caaagaaaac ctcaaagccc gccccacag tgagaaactc cctccaacia ggccacatgt 35700
tctctagcaa ggccacacct cctaataagcg cctatgggccc aggtattcaa accaccacac 35760
catacatatc ttacagctct tccttgaga tctttcttta tactttggag gcaatggcag 35820
cacggatgac ctcaactgtt agatgtttgt gaatccctcc ctgctgactt gattttggat 35880
gtgtttttat tttatggtgc tggacattgt acatgagaca agcatcctgt aattgagccc 35940
agcctttgag ttagtgatct ataggctgag caaaaaacta taatgaagtc agtagagtct 36000
gtctgcacat tcttaagtgg ctgtcttaaa acaattaagg taaggggctg gagagatgct 36060
tcctcggtca agagcactgg ctgctcttcc agaggacctg gggtcagttc ccagaccca 36120
tatggcagct cacaactgtc tatacctcca gttccagctc gacatcctca catagacata 36180
catgcaggca aaacaccaat gtacattaaa aaaaacacct aatttttaa aagttcagat 36240
gaaaagaaga aatactatga ttaacttctc agaaacattt ctatttgtaa acttgacctc 36300
ccaaggtcaa ggatcctgtg acttctcatt tttgcccctg tattttgttg ttgtgtgtgt 36360
ttttgtttgt ttgtttgtg ttttgtttgt tgtttagttt agtttctcgt tgtttgtttt 36420
gtcctttcct ggttccttcc cctttctttg taagcactcc tgctctggct gggcccagc 36480
tcacttccag cctcctctga tggagccagc attacatctg ctgttttgca tttgtatac 36540
aggtttcggc cgagtcagc tcctccatga actccaacac cccgctggtg aggataaca 36600
cgcgtctgtc ctcaacagcg gacaccccga tgcctagcagg ggtctccgag tatgagttgc 36660
cagaggatcc aaagtgggaa ttccccagag ataagtaagt actctccctc tgggagggtc 36720
gttgtctgca cctcctggga ctgagcgcag gtcttggttg tgggagtctc cacctgtgtc 36780
ttggtaatca gggacctgtg tcttggaat cagggacctt cgaactgtaa actgtaact 36840
gtaactgca gcaagatggt gcaattaaca gagctgctgg tgcacagggg aggctaccag 36900
cctgtgccct tgaggtggaa gaccaacctt agctctggga agtgaggatc ctggaaggct 36960
ggcagcttcc ttcttgtagg attagcgtct aaacagcttg agagtaacag aaggtggaaa 37020
aattggctct ttctgcatca aagacacagg aatacgtcc cagcttgctt gaagacaact 37080
cgtctgctca tcttgacatt ttttcagtgt cttcctaaga ttgttagtga tatgtttaac 37140
acacacagcg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg taticagaga 37200
ttcagaaaaca caaagacaag ctttcccaag tctgggtatg gccagctcta gttgaaagt 37260
gaaggaagct taggccccct tggagcttcg tgtcacacta tccaggtat tttggttccc 37320
aacttagaga tctttacata tatcctaate cagggccgag gaactgtctg tgtagtcatc 37380
ttatgtcagt ggaaagaatt tccagtttct ttatatactg cagtgaaaag agcattcatt 37440
cattccttca ttcatcccg gaatgtttag tggacatttt tgggtgagg tatgaaaga 37500
aacacaaaaga atggcttctt tacctataga attttgagaa aggaaggtat gtttctcttg 37560
ccccttgcca gcctcctagc tgggtgtgat tatagaaata gctgggttcg tgtgcatt 37620
cctatgtcta caaggctgga ctggggagtt gtgctcatat gctaaaaact tgcagcttcc 37680
gggttgagcc tgtgcttttg gccacctgtg acaccacagt caacagtgtg agtctgtgtt 37740
gcccagcagc tcccaccg ggtacccaa gtgtggaaaa tctgagcgt atgcatttcc 37800
aagcagtggt tagcacaat gtaggtggag cacattccca atgaatgaag gctatttaga 37860
aatggttat taggttgag gcaggagtt cacaagaga ggtttgttat ggttgtagta 37920
acaggggcaa taggacataa attagccatt ttccaacgca aatgtttatt ttctgccaag 37980
atgttcaatt taattttagt tcttgactgg aaagggggcg ccttcagcag agcaagtgct 38040
ggtatccatt ttcttttcc tttttgact tttaaaagcg ttgctgctt ttgcaaatg 38100
ctgtttactg aggacagctc aatactggc tgtactggc ttatggtttt ttggttttg 38160
tttttttaa aacaaaacia aacaaaacia aacaaaacac cccgaaaagc atactcaggc 38220
tggagaggtat tctctctggt ctaaagcaga cactgcctg taaaggacca gagttcgggt 38280
cccagcacc atgatgtgta gctcacacc tctgtaacta cagatctggg gaatctgtct 38340
gggctcctta ggtgcctgtg ctacatgccc catacctct cccagaaac acacatacac 38400
ataatcaaga ataaattgaa aaaaaataaa aataaaaaaac acactcctaa gtattaattc 38460
caaagacttc cctgttccct tggcttctgg aacatctaaa ataattgcag gtcatttgtc 38520
tgtgtgtgat aaaacttaca tgcttagaaa tgtaacttgt gctgttttct atttttttt 38580
ttcctgttt cctttgggta gtgataagga atcctaaact tatgtcaaaa aggtatcgtg 38640
cctgatttct agaagttttt cttaatgaga caccgataaat tatttgaaac gtgctgaga 38700
ttcctacct gcaactgggc aatcgatgta accataaaat ctaccggtat tgaataatag 38760
tgatttgaga gttgccactt tacagggaca gaaaataaga acagactttc actttttttt 38820
tcacctctgc gacattttaa attataaata ttaattggc tcaagaccaa aagctcccta 38880
tggctggcat gcagggagcc tgaccaccgc gctagcaagg acaccttcca taaagaaaaa 38940

gaaaataaat	cgagaggaca	aatgtgaaat	ttaatagtcc	ctccaacagt	aattgacgtt	39000
ctggaaaaac	atcactaaga	aaatagcctg	cgtgtgtatc	ggaggctcat	tggttccata	39060
tgcatgcctc	tggaagattt	ttatatttag	ttctggaatt	tcctccctcg	tgcccctcgg	39120
ccagactcgc	ggtgtgctaa	tcccgtattt	acacatttag	gctgacgctg	ggcaaaccct	39180
tgggggaagg	ttgcttcggg	caagtagtca	tggctgaagc	agtgggaatc	gataaagaca	39240
aaccaagga	ggcggtcacc	gtggcagtga	agatgttgaa	aggtgagtgg	gcggatgggc	39300
ggtcggggag	gagagggtct	tatcaggagc	gagcgttcct	tttgtgacat	gtgaactctg	39360
cagggacgtg	gggtcagaga	gcacatactt	gacctggcgg	ttgagggggt	tttcaggata	39420
aatgagcaaa	tgagatggag	gatttacctt	gagctgtgtg	tacttaaaaa	gaaaagccag	39480
tttagcagca	agttgtagct	tgctgggctg	aaccggctct	taactcctta	gaaaagggct	39540
ccgattctct	tcttttctgt	gtgttcatgg	gtttagaag	tttagggggt	ttatttagct	39600
ggttaaattt	tggaccaga	cttttaacat	acaataaag	agaggtaggt	gttggagtgg	39660
caactggaga	cagaatgtca	aaatgtggat	tcaaagagtc	gcttagaagc	caaaaaggag	39720
caaacaattg	gaactgatgc	agaatcccag	ggacatgtaa	acaataatgc	cacgctataa	39780
atgcccgctt	tggtcttttc	ttttcttttc	ttttcttttc	tttttttttt	ttttgaggga	39840
ggggagggga	ggggatctgg	gaatttatcc	acaaccttcc	taacacagct	tgatgatgac	39900
gccaagggag	cttaaatgct	tttcaactat	taacttatcc	ttgcatgggt	attcttttat	39960
cgaagagata	aagggaagg	tcacattata	aatcctgttg	ttggggaatc	tcagaaggga	40020
gaaaggagcc	atgttcaatg	ttccctggc	ttgtgggag	agaagtctgt	cccgggcctg	40080
tgggatgtgg	catgttctca	ggagtccgac	cttttctctc	tttgatagga	cacttaccac	40140
atccctccct	gatgcagaca	acaagggcc	aggacatggt	tcatthttgtc	agthtttagtt	40200
attgacctga	gactcccagt	gaaatctggg	atgttccctt	ctttgggagc	tgataccagg	40260
aaggagatag	caagtatcgg	ggcaccaggc	cgagggcagc	ccttgggtacc	tactggaagc	40320
tgtgggttgg	gaaggatcag	gcatcact	gctttccaca	gaacctctgg	ttttgagatc	40380
cctggagcta	gtgcaaaagg	gaggtttagg	ggttggccct	tccctttaaag	caagatcacc	40440
caccatcctt	ttcatcgtgg	tcagaggaca	tgctttttca	acattctttg	tgacagccag	40500
aggatggctg	aggtgtaagg	aagacaagtg	tactgagcca	tgtgtctgtc	catagtcctc	40560
tcttccctct	tctctgtatt	ggtcaggata	gatttttggg	tacctgtgcc	tctatttcat	40620
ttttaaccct	tttgcttttc	ttttagctca	gatttttctt	ttctaagtat	ttctgtattg	40680
aattagctta	gtgacagaac	acttgcgtgg	tgtgcacatg	gtactggggt	tgatccttag	40740
cattacaaga	atcctaacga	cagcagaact	aactgagagg	agagcacagt	agcggccgca	40800
aattgctttg	agagctcta	taaaacctta	gaggctattt	aaattttaat	ggccggcccc	40860
acggccaggc	ggccgccagg	cctaccact	agtcaattcg	ggaggatcga	aacggcagat	40920
cgaaaaaac	agtacataca	gaaggagaca	tgaacatgaa	catcaaaaaa	attgtaaaac	40980
aagccacagt	tctgactttt	acgactgcac	ttctggcagg	aggagcgact	caagccttcg	41040
cgaaagaaaa	taaccaaaaa	gcatacaaa	aaacgtacgg	cgctctctcat	attacacgcc	41100
atgatatgct	gcagatccct	aaacagcagc	aaaacgaaaa	ataccaagtg	cctcaattcg	41160
atcaatcaac	gattaaaaat	attgagtctg	caaaaggact	tgatgtgtgg	gacagctggc	41220
cgctgcaaaa	cgctgacgga	acagttagctg	aatacaacgg	ctatcacggt	gtgthttgctc	41280
ttgccccgag	cccgaagagc	gctgatgaca	catcaatcta	catgthttat	caaaaggctcg	41340
gcgacaactc	aatcgacagc	tggaaaaacg	cgggccgtgt	ctttaaagac	agcgataagt	41400
tcgacgccaa	cgatccgatc	ctgaaagatc	agacgcaaga	atgggtccggt	tctgcaacct	41460
ttacatctga	cggaaaaatc	cgthttattct	acactgacta	ttccggtaaa	cattacggca	41520
aacaaagcct	gacaacagcg	caggtaaatg	tgtcaaaatc	tgatgacaca	ctcaaaatca	41580
acggagtgga	agatcacaac	acgatttttg	acgggagcgg	aaaaacatat	cagaacgtht	41640
agcagthttat	cgatgaaggc	aattatacat	ccggcgacaa	ccatacgctg	agagaccctc	41700
actacgthtga	agacaaaggc	cataaatacc	ttgtattcga	agccaacacg	ggaacagaaa	41760
acggatacca	aggcgaagaa	tctthattta	acaaagcgtg	ctacggcggc	ggcacgaaact	41820
tcttccgtaa	agaaagccag	aagcttcagc	agagcgctaa	aaaacgcgat	gctgagthtag	41880
cgaacggcg	cctcggtatc	atagagthta	ataatgatta	cacattgaaa	aaagthaatga	41940
agccgctgat	cacttcaaac	acggthactg	atgaaatcga	gcgcgcgaat	gthttcaaaa	42000
tgaacggcaa	atggthactt	ttcactgatt	cacgcgtht	aaaaatgacg	atcgatggta	42060
ttaactcaaa	cgatathttac	atgctthggt	atgtatcaaa	ctctthtaacc	ggccctthaca	42120
agccgctgaa	caaaacaggg	ctthgtctgc	aaatggthct	tgatccaaac	gatgtgacat	42180
tcacttactc	tcacttcgca	gtgcccgaag	ccaaaggcaa	caatgtgtht	atcacaagct	42240
acatgacaaa	cagaggtht	ttcgaggata	aaaaggcaac	atthtgccca	agctthcttaa	42300
tgaacatcaa	aggcaataaa	acatccgtht	tcaaaaacag	catcctggag	caaggacagc	42360
tgacagthcaa	ctaataacag	caaaaagaaa	atgccgatac	ttcattggca	thttcttht	42420
thtctcaaca	agatgthgaa	ttgactagth	gthtagatcca	caggacgggt	gtgthcgcca	42480
tgatcgcgta	gtcgatagth	gthcctaaag	gcgaagcgag	caggactggg	cgccggccaa	42540
agccgthcgga	cagthgctccg	agaacggthg	cgcatagaaa	thgcatcaac	gcataatagcg	42600
ctagcagcac	gcatagthga	ctggcgatgc	tgtcggaatg	gacgatatcc	cgcaagaggc	42660
ccggcagthac	cggcataaac	aagcctatgc	ctacagcatc	caggthgacg	gtgccgagga	42720 "

tgacgatgag	cgcattgtta	gatttcatac	acggtgcctg	actgcgtag	caatttaact	42780
gtgataaact	accgcattaa	agcttatoga	tgataagctg	tcaaacatga	gaattgatcc	42840
ggaaccctta	atataacttc	gtataatgta	tgctatacga	agttattagg	tccctcgact	42900
atagggtcac	cgtcgcacagc	gacacacttg	catcggatgc	agcccggtta	acgtgccggc	42960
acggcctggg	taaccaggta	ttttgtccac	ataaccgtgc	gcaaaatggt	gtggataagc	43020
aggacacagc	agcaatccac	agcaggcata	caaccgcaca	ccgaggttac	tccgtctcac	43080
aggttacgac	gacatgtcaa	tacttgccct	tgacaggcat	tgatggaatc	gtagtctcac	43140
gctgatagtc	tgatcgacaa	tacaagtggg	accgtgggtc	cagaccgata	atcagaccga	43200
caacacgagt	gggatcgtgg	tcccagacta	ataatcagac	cgacgatacg	agtgggaccg	43260
tggtcccaga	ctaataatca	gaccgacgat	acgagtggga	ccgtggttcc	agactaataa	43320
tcagaccgac	gatacgagtg	ggaccgtggt	cccagactaa	taatcagacc	gacgatacga	43380
gtgggaccat	ggtcccagac	taataatcag	accgacgata	cgagtgggac	cgtggtccca	43440
gtctgattat	cagaccgacg	atacgagtg	gaccgtggtc	ccagactaat	aatcagaccg	43500
acgatacag	tgggaccgtg	gtcccagact	aataatcaga	ccgacgatac	gagtgggacc	43560
gtggtcccag	tctgattatc	agaccgacga	tacaagtgga	acagtgggcc	cagagagaat	43620
atcaggcca	gttatgcttt	ctggcctgta	acaaggaca	ttaagtaaag	acagataaac	43680
gtagactaaa	acgtggtcgc	atcaggggtc	tggtttttca	agttccttaa	gaatggcctc	43740
aattttctct	atacactcag	ttggaacacg	agacctgtcc	aggttaagca	ccatttttatc	43800
gcccttatac	aatactgtcg	ctccaggagc	aaactgatgt	cgtgagctta	aactagttct	43860
tgatgcagat	gacgttttaa	gcacagaagt	taaaagagt	ataacttctt	cagcttcaaa	43920
tatcacccca	gcttttttct	gctcatgaag	gttagatgcc	tgctgcttaa	gtaattcctc	43980
tttatctgta	aaggcttttt	gaagtgcac	acctgaccgg	gcagatagtt	caccgggggtg	44040
agaaaaaaga	gcaacaactg	atttaggcaa	tttggcggtg	ttgatacagc	gggtaataat	44100
cttacgtgaa	atattttccg	catcagccag	cgagaaaata	tttccagcaa	atcattctctg	44160
caatcggctt	gcataacgct	gaccacgttc	ataagcactt	gttgggcgat	aatcgttacc	44220
caatctggat	aatgcagcca	tctgctcatc	atccagctcg	ccaaccagaa	cacgataatc	44280
actttcggta	agtgcagcag	ctttacgacg	gcgactccca	tccgcaat	ctatgacacc	44340
agatactctt	cgaccgaacg	ccggtgtctg	ttgaccagtc	agtagaaaag	aagggatgag	44400
atcatccagt	gcgtcctcag	taagcagctc	ctggtcacgt	tcattacctg	accataccgg	44460
agaggtcttc	tcaacactat	caccocggag	cacttcaaga	gtaaacttca	catcccagacc	44520
acatacagc	aaagtaatgg	cattaccgcg	agccattact	cctacgcgcg	caattaacga	44580
atccaccatc	gggacagctg	gtgtcgataa	cgaagtatct	tcaaccggtt	gagatttgag	44640
cgtatgtttt	ggaataacag	gcgcacgctt	cattatctaa	tctcccagcg	tggttaaatc	44700
agacgatcga	aaatttcatt	gcagacaggt	tcccaaatag	aaagagcatt	tctccaggca	44760
ccagttgaag	agcgttgatc	aatggcctgt	tcaaaaacag	ttctcatccg	gatctgacct	44820
ttaccaactt	catccgtttc	acgtacaaca	ttttttagaa	ccatgcttcc	ccaggcatcc	44880
cgaatttgct	cctccatcca	cggggactga	gagccattac	tattgctgta	tttggttaagc	44940
aaaatacgt	catcaggctc	gaacccttta	agatcaacgt	tcttgagcag	atcacgaagc	45000
atatacga	actgcagtgc	ggaggtgtag	tcaacaact	cagcaggcgt	gggaacaatc	45060
agcacatcag	cagcacatac	gacattaatc	gtgccgatac	ccaggttagg	cgcgctgtca	45120
ataactatga	catcatagtc	atgagcaaca	gtttcaatgg	ccagtcggag	catcaggtgt	45180
ggatcgggtg	gcagtttacc	ttcatcaaat	ttgccatta	actcagttc	aatacgggtc	45240
agagccagac	aggaaggaat	aatgtcaagc	cccggccagc	aagtgggctt	tattgcataa	45300
gtgacatcgt	ccttttcccc	aagatagaaa	ggcaggagag	tgtcttctgc	atgaatatga	45360
agatctggta	cccattccgtg	atacattgag	gctgttccct	gggggtcggt	accttccacg	45420
agcaaaacac	gtagcccctt	cagagccaga	tcctgagcaa	gatgaacaga	aactgaggtt	45480
ttgtaaacgc	cacctttatg	ggcagcaacc	ccgatcaccg	gtggaaatac	gtcttcagca	45540
cgtcgaatc	gcgtaccaa	cacatcacgc	atatgattaa	tttgttcaat	tgtataacca	45600
acacggttct	caacccttcc	tccaatttcc	atatccgggt	gcggtagtcg	ccctgctttc	45660
tccgcatctc	tgatagcctg	agaagaaacc	ccaactaaat	ccgctgcttc	acctattctc	45720
cagcgcgggg	ttattttcct	cgcttccggg	ctgtcatcat	taaactgtgc	aatggcgata	45780
gccttcgtca	tttcatgacc	agcgtttatg	cactggttaa	gtgtttccat	gagtttcatt	45840
ctgaacatcc	tttaatcatt	gctttgctgt	tttttattaa	atcttgcaat	ttactgcaaa	45900
gcaacaacaa	aatcgcgaa	tcatacaaaa	accgcaaagt	tgttttaa	aagagcaaca	45960
ctacaaaagg	agataagaag	agcacatacc	tcagtcactt	attatcacta	gcgctcgccg	46020
cagccgtgta	accgagcata	gcgagcgaac	tggcgaggaa	gcaaaaga	actgttctgt	46080
cagatagctc	ttacgctcag	cgcaagaaga	aatatccacc	gtgggaaaaa	ctccaggtag	46140
aggtacacac	tgggatagcc	aattcagagt	aataaactgt	gataatcaac	cctcatcaat	46200
gatgacgaac	gaaccccga	tatcaggtca	catgacgaag	ggaaagagaa	ggaaatcaac	46260
tgtgacaaac	tgccctcaa	tttggcttcc	ttaaaaatta	cagttcaaaa	agtatgagaa	46320
aatccatgca	ggctgaagga	aacagcaaaa	ctgtgacaaa	ttaccctcag	taggtcagaa	46380
caaatgtgac	gaaccaccct	caaatctgtg	acagataacc	ctcagactat	cctgtcgtca	46440
tggaaagtgat	atcgcggaag	gaaaatacga	tatgagtcgt	ctggcggcct	ttctttttct	46500

caatgtatga gaggcgcatt ggagttctgc tghtgatctc attaacacag acctgcagga 46560
 agcggcggcg gaagtcaggc atacgctggt aactttgagg cagctggtaa cgctctatga 46620
 tccagtcgat tttcagagag acgatgcctg agccatccgg cttacgatac tgacacaggg 46680
 attcgtataa acgcatggca tacggattgg tgatttcttt tgtttcacta agccgaaact 46740
 gcgtaaaccg gttctgtaac ccgataaaga agggaatgag atatgggttg atatgtacac 46800
 tgtaaagccc tctggatgga ctgtgocgac gtttgataaa ccaaggaaaa gattcatagc 46860
 ctttttcatc gccggcatcc tcttcagggc gataaaaaac cacttccttc cccgcgaaac 46920
 tcttcaatgc ctgccgtata tccttactgg cttccgcaga ggtcaatccg aatatttcag 46980
 catathtagc aacatggatc tcgcagatac cgatcatgtc ctgtaggggtg ccatcagatt 47040
 ttctgatctg gtcaacgaac agatacagca tacgtttttg atcccgggag agactatatg 47100
 ccgcctcagt gaggtcgttt gactggacga ttccgagggt atttttacgt ttcttgtgat 47160
 tgataaccgc tgtttccgcc atgacagatc catgtgaagt gtgacaagt tttagattgt 47220
 cacactaaat aaaaaagagt caataagcag ggataacttt gtgaaaaaac agcttcttct 47280
 gagggcaatt tgtcacaggg ttaagggcaa tttgtcacag acaggactgt catttgaggg 47340
 tgatttgtca cactgaaagg gcaatttgtc acaacacctt ctctagaacc agcatggata 47400
 aaggcctaca aggcgctcta aaaaagaaga tctaaaaact ataaaaaaaa taattataaa 47460
 aatatccccg tggataagtg gataacccca agggaaagtt tttcaggcat cgtgtgtaag 47520
 cagaatataa aagtgtctgt ccctgggtgt tccctcgtca ctcgagggtc tcgccctgtc 47580
 gctcaactgc ggcgagcact actggctgta aaaggacaga ccacatcatg gttctgtgtt 47640
 cattaggttg ttctgtccat tgctgacata atccgctcca cttcaacgta acaccgcacg 47700
 aagatttcta ttgttcctga aggcataatc aaatcgtttt cgttaccgct tgcaggcatc 47760
 atgacagaac actacttctt ataaacgcta cacaggctcc tgagattaat aatgcggatc 47820
 tctacgataa tgggagattt tcccgactgt ttcgttcgct tctcagtgga taacagccag 47880
 cttctctgtt taacagacaa aaacagcata tccactcagt tccacatttc catataaagg 47940
 ccaaggcatt tattctcagg ataattgttt cagcatcgca accgcatcag actccggcat 48000
 cgcaaacctg acccgggtgcc gggcagccac atccagcgca aaaaccttcg ttagacttc 48060
 cgttgaactg atggacttat gtcccacag gctttgcaga actttcagcg gtataccggc 48120
 atacagcatg tgcacgcat aggaatggcg gaacgtatgt ggtgtgaccg gaacagagaa 48180
 cgtcacaccg tcagcagcag cggcggcaac cgctcccca atccaggctc tgaccgttct 48240
 gtccgtcact tcccagatcc gcgctttctc tgccttctc gtgcgacggt tacgccgctc 48300
 catgagctta tcgcgaataa atacctgtga cggagatca cttcgcagaa taaataaatc 48360
 ctgggtgtccc tgttgatacc gggaaagcct gggccaactt ttggcgaaaa tgagacgttg 48420
 atccgacagt aagaggttcc aactttcacc ataataaat aagatcacta ccgggcgat 48480
 tttttgagtt atcgagattt tcaggagcta aggaagctaa aatggagaaa aaaatcactg 48540
 gatataccac cgttgatata tcccattggc atcgtaaaga acattttgag gcatttcagt 48600
 cagttgctca atgtacctat aaccagaccg ttcagctgga tattacggcc tttttaaaga 48660
 ccgtaaagaa aaataagcac aagttttatc cggcctttat tcacattctt gcccgctga 48720
 tgaatgctca tccggagttc cgtatggcaa tgaaagacgg tgagctggtg atatgggata 48780
 gtgttcaccc ttgttacacc gttttccatg agcaaacctga aacgttttca tcgctctgga 48840
 gtgaaatacca cgacgatttc cggcagtttc tacacatata ttcgcaagat gtggcgtgtt 48900
 acggtgaaaa cctggcctat ttccctaaag ggtttattga gaatatgtt ttctctcag 48960
 ccaatccctg ggtgagtttc accagttttg atttaaacgt ggccaatatg gacaacttct 49020
 tcgccccctg tttcaccatg ggcaaatatt atacgcaagg cgacaagggt ctgatgccgc 49080
 tggcgattca ggttcatcat gccgtttgtg atggcttcca tgtcggcaga atgcttaatg 49140
 aattacaaca gtactgcatg gactggcagg gcggggcgta atttttttaa ggagttatt 49200
 ggtgccctta aacgcctggt tgctacgcct gaataagtga taataagcgg atgaatggca 49260
 gaaattcgat gataagctgt caaacatgag aattggctga cggcgcgcca aagcttgcac 49320
 gcctgcagcc gcgtaacctg gcaaaaatcgg ttacggttga gtaataaatg gatgccctgc 49380
 gtaagcgggg cacatttcat tacctctttc tccgcacccg acatagataa taacttcgta 49440
 tagtatacat tatacgaagt tatctagtag acttaattaa ggatcgatcc ggcgcgcca 49500
 tagtcatgcc ccgcgccac cgggaaggagc tgactgggtt gaaggctctc aagggcatcg 49560
 gtcgagcttg acattgtagg actatattgc tctaataaat ttgcggccgc taatacgact 49620
 cactataggg a

<210> 18

<211> 20

<212> ADN

5 <213> Secuencia Artificial

<220>
 <223> Sintética
 <400> 18
 ggaaagccac cctgtatgct 20
 5 <210> 19
 <211> 19
 <212> ADN
 <213> Secuencia Artificial
 <220>
 10 <223> Sintética
 <400> 19
 cttggccaac agtggatgg 19
 <210> 20
 <211> 42
 15 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Sintética
 <400> 20
 20 cuaaaaugau ucucaucugc guuuuagagc uaugcuguuu ug 42
 <210> 21
 <211> 42
 <212> ARN
 <213> Secuencia Artificial
 25 <220>
 <223> Sintética
 <400> 21
 gcucucaacu ucacccuuuc guuuuagagc uaugcuguuu ug 42
 <210> 22
 30 <211> 23
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintética
 35 <400> 22
 ctaaatgat tctcatctgc agg 23
 <210> 23

- <211> 23
<212> ADN
<213> Secuencia Artificial
<220>
5 <223> Sintética
<400> 23
gctctcaact tcacccttc tgg 23
<210> 24
<211> 23
10 <212> ADN
<213> Secuencia Artificial
<220>
<223> Sintética- un locus objetivo que se enlaza a un ARN guía (ARNg)
<220>
15 <221> característica nueva
<222> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
<223> n = A, T, C o G
<400> 24
nnnnnnnnnn nnnnnnnnn ngg 23
20 <210> 25
<211> 25
<212> ADN
<213> Secuencia Artificial
<220>
25 <223> Sintética- un locus objetivo que se enlaza a un ARN guía (ARNg)
<220>
<221> característica nueva
<222> 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
<223> n = A, T, C o G
30 <400> 25
ggnnnnnnnn nnnnnnnnn nnngg 25

REIVINDICACIONES

1. Un método para ensamblar al menos dos ácidos nucleicos, que comprende:
- (a) poner en contacto un primer ácido nucleico con un primer agente de nucleasa, en el que el primer agente de nucleasa comprende una proteína Cas y un ARN guía (ARNg) (complejo de ARNg-Cas), una nucleasa de dedo de zinc o una nucleasa efectora similar al activador de transcripción (TALEN), en el que el primer agente de nucleasa escinde el primer ácido nucleico en un primer sitio objetivo para producir un primer ácido nucleico digerido con una secuencia final solapante compartida por un segundo ácido nucleico;
- (b) poner en contacto el primer ácido nucleico digerido y el segundo ácido nucleico digerido con una exonucleasa para exponer secuencias complementarias entre el primer ácido nucleico digerido y el segundo ácido nucleico; y
- (c) ensamblar los dos fragmentos de ácido nucleico generados a partir de la etapa (b).
2. El método de la reivindicación 1, en el que la etapa (c) comprende:
- (i) hibridar las secuencias complementarias expuestas;
- (ii) extender los extremos 3' de las secuencias complementarias hibridadas; y
- (iii) ligar el primer y el segundo ácidos nucleicos.
3. El método de la reivindicación 1 o 2, en el que al menos dos ácidos nucleicos son de doble cadena; y/o la secuencia final solapada varía de 20 pb a 200 pb de longitud.
4. El método de cualquiera de las reivindicaciones 1-3, en el que la etapa (a) comprende además poner en contacto el segundo ácido nucleico con un segundo agente de nucleasa, en el que el segundo agente de nucleasa escinde al segundo ácido nucleico en un segundo sitio objetivo para producir un segundo ácido nucleico digerido con la secuencia final solapada, y
- en el que el segundo ácido nucleico de la etapa (b) es el segundo ácido nucleico digerido.
5. El método de cualquiera de las reivindicaciones 1-4, en el que el primer agente de nucleasa comprende la proteína Cas y el ARNg, en el que la proteína Cas es una proteína Cas9, el ARNg comprende una secuencia de ácido nucleico que codifica un ARN de repeticiones palindrómicas cortas agrupadas y regularmente intercaladas (CRISPR) (ARNcr) y un ARN de CRISPR que se activa en forma trans (ARNcrtra), y el primer sitio objetivo está flanqueado inmediatamente por una secuencia de un motivo adyacente protoespaciador (PAM), opcionalmente en el que la proteína Cas9 comprende un dominio RuvC y un dominio HNH, al menos uno de los cuales carece de actividad endonucleasa.
6. El método de cualquiera de las reivindicaciones 1-5, en el que el primer ácido nucleico, el segundo ácido nucleico o ambos ácidos nucleicos se derivan de un cromosoma artificial bacteriano.
7. El método de la reivindicación 6, en el que el cromosoma artificial bacteriano comprende un ADN humano, un ADN de roedor, un ADN sintético o una combinación de los mismos.
8. Un método para ensamblar dos o más ácidos nucleicos, que comprende:
- (a) poner en contacto un primer ácido nucleico con al menos un agente de nucleasa, en el que al menos un agente de nucleasa comprende una proteína Cas y un ARN guía (ARNg) (complejo de ARNg-Cas), una nucleasa de dedo de zinc o una nucleasa efectora similar al activador de transcripción (TALEN), en el que al menos un agente de nucleasa escinde al primer ácido nucleico en un primer sitio objetivo para generar un primer ácido nucleico digerido;
- (b) poner en contacto el primer ácido nucleico digerido con un segundo ácido nucleico, un oligo de unión y una exonucleasa,
- en el que el oligo de unión comprende:
- (i) una primera secuencia complementaria que es complementaria al primer ácido nucleico digerido;
- (ii) un espaciador; y
- (iii) una segunda secuencia complementaria que es complementaria al segundo ácido nucleico,
- en el que la exonucleasa expone la primera y segunda secuencias complementarias; y
- (c) ensamblar el oligo de unión con el primer ácido nucleico digerido y el segundo ácido nucleico.
9. El método de la reivindicación 8, en el que el ensamblaje en la etapa (c) comprende:

- (i) hibridar la primera secuencia complementaria del oligo de unión con el primer ácido nucleico digerido y la segunda secuencia complementaria del oligo de unión con el segundo ácido nucleico; y
- (ii) ligar el oligo de unión al primer ácido nucleico digerido y al segundo ácido nucleico.
- 5 10. El método de la reivindicación 9, en el que la etapa (i) comprende adicionalmente extender el extremo 3' del primer ácido nucleico digerido y/o el segundo ácido nucleico.
11. El método de cualquiera de las reivindicaciones 8-10, en el que la primera secuencia complementaria del oligo de unión está entre 15 y 120 bases complementarias, y la segunda secuencia complementaria del oligo de unión está entre 15 y 120 bases complementarias; y/o el espaciador del oligo de unión comprende ácidos nucleicos no complementarios.
- 10 12. El método de cualquiera de las reivindicaciones 8-11, en el que los dos o más ácidos nucleicos son de doble cadena.
13. El método de cualquiera de las reivindicaciones 8-12, en el que el primer ácido nucleico digerido se ensambla en forma continua con el segundo ácido nucleico.
- 15 14. El método de la reivindicación 13, en el que al menos un agente de nucleasa está diseñado para escindir un fragmento de al menos 20 pb del extremo del primer ácido nucleico en el que se producirá el ensamblaje en forma continua,
- en el que el espaciador del oligo de unión comprende una secuencia idéntica al fragmento de al menos 20 pb, en el que no están presentes bases de ácido nucleico entre la primera secuencia complementaria y el fragmento de al menos 20 pb, y no están presentes bases de ácido nucleico entre la segunda secuencia complementaria y el fragmento de al menos 20 pb,
- 20 de modo que el ensamblaje del primer ácido nucleico digerido con el oligo de unión y el segundo ácido nucleico reconstituye el fragmento de al menos 20 pb y ensambla en forma continua el primer ácido nucleico y el segundo ácido nucleico,
- opcionalmente en el que el fragmento de al menos 20 pb es de doble cadena.
- 25 15. El método de la reivindicación 13 o 14, en el que el oligo de unión comprende un fragmento lineal de ADN de doble cadena, opcionalmente en el que el fragmento lineal de ADN de doble cadena no comprende un casete de selección.
16. El método de cualquiera de las reivindicaciones 13-15, en el que el oligo de unión es de aproximadamente 50 pb a aproximadamente 400 pb, opcionalmente en el que el oligo de unión es de aproximadamente 100 pb a aproximadamente 300 pb.
- 30 17. El método de cualquiera de las reivindicaciones 8-16, en el que el espaciador está entre aproximadamente 20 pb y aproximadamente 120 pb; y/o el oligo de unión se ensambla al primer ácido nucleico y el segundo ácido nucleico en la misma reacción o el oligo de unión se ensambla con el primer ácido nucleico y el segundo ácido nucleico secuencialmente.
18. El método de cualquiera de las reivindicaciones 8 a 17, en el que la etapa (a) comprende, además:
- 35 (i) poner en contacto el segundo ácido nucleico con un segundo agente de nucleasa, en el que el segundo agente de nucleasa escinde al segundo ácido nucleico para producir un segundo ácido nucleico digerido que comprende una secuencia de nucleótidos que es complementaria a la segunda secuencia complementaria del oligo de unión, en el que el primer ácido nucleico digerido se ensambla con el segundo ácido nucleico digerido; o
- 40 (ii) poner en contacto el segundo ácido nucleico con una enzima de restricción o meganucleasa, en el que la enzima de restricción o la meganucleasa escinde el segundo ácido nucleico para producir un segundo ácido nucleico digerido que comprende una secuencia de nucleótidos que es complementaria a la segunda secuencia complementaria en el oligo de unión en el que el primer ácido nucleico digerido se ensambla al segundo ácido nucleico digerido.
19. El método de cualquiera de las reivindicaciones 8-18, en el que al menos un agente de nucleasa comprende la proteína Cas y el ARNg, en el que la proteína Cas es una proteína Cas9, en el que el ARNg comprende una secuencia de ácido nucleico que codifica un ARN de repeticiones palindrómicas cortas agrupadas y regularmente intercaladas (CRISPR) (ARNcr) y un ARN de CRISPR que se activa en forma trans (ARNcrtra), y en el que el primer sitio objetivo está flanqueado inmediatamente por una secuencia de un motivo adyacente protoespaciador (PAM), opcionalmente en el que la proteína Cas9 comprende un dominio RuvC y un dominio HNH, al menos uno de los cuales carece de actividad endonucleasa.
- 45 20. El método de cualquiera de las reivindicaciones 8-19, en el que el primer ácido nucleico, el segundo ácido nucleico o ambos ácidos nucleicos se derivan de un cromosoma artificial bacteriano; y/o el primer ácido nucleico, el segundo ácido nucleico o ambos ácidos nucleicos comprenden un ADN humano, un ADN de roedor, un ADN sintético o una combinación de los mismos.
- 50

21. El método de cualquiera de las reivindicaciones 1-20, en el que el primer ácido nucleico, el segundo ácido nucleico o ambos ácidos nucleicos son de al menos 10 kb.

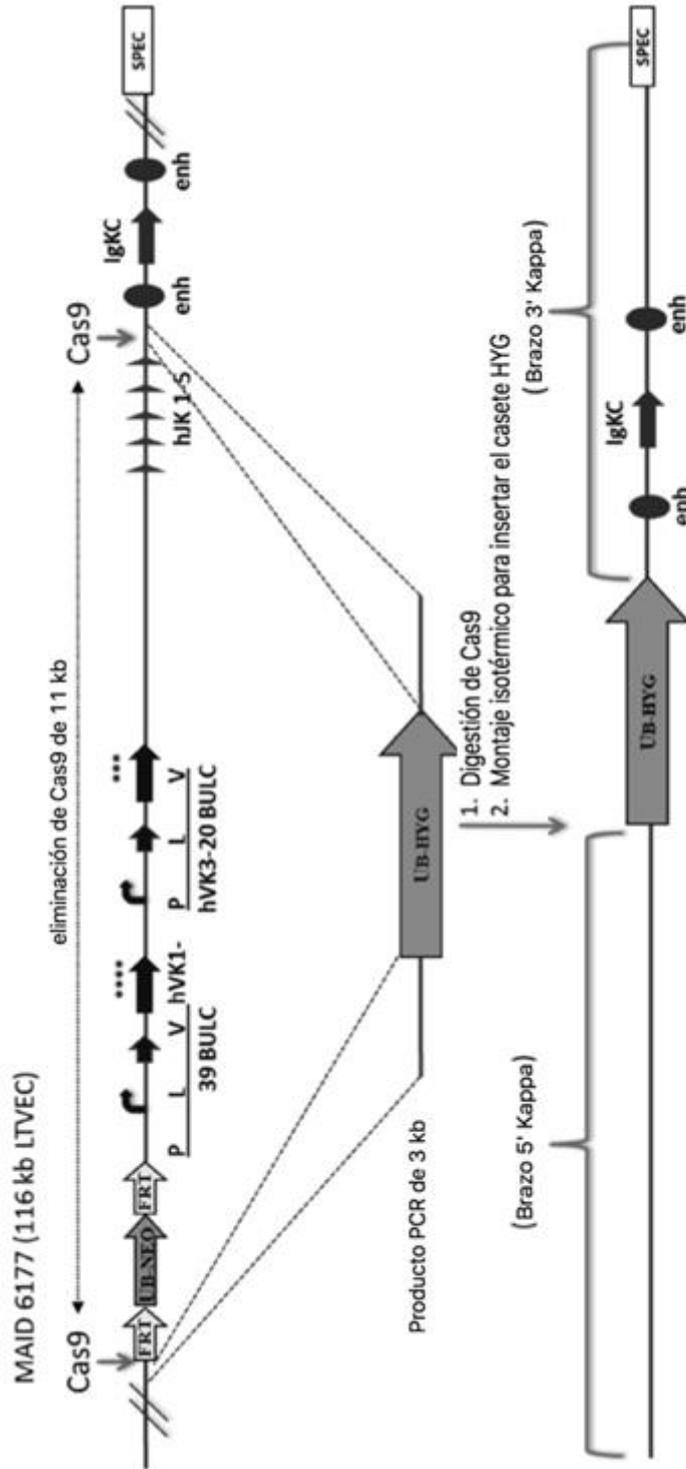
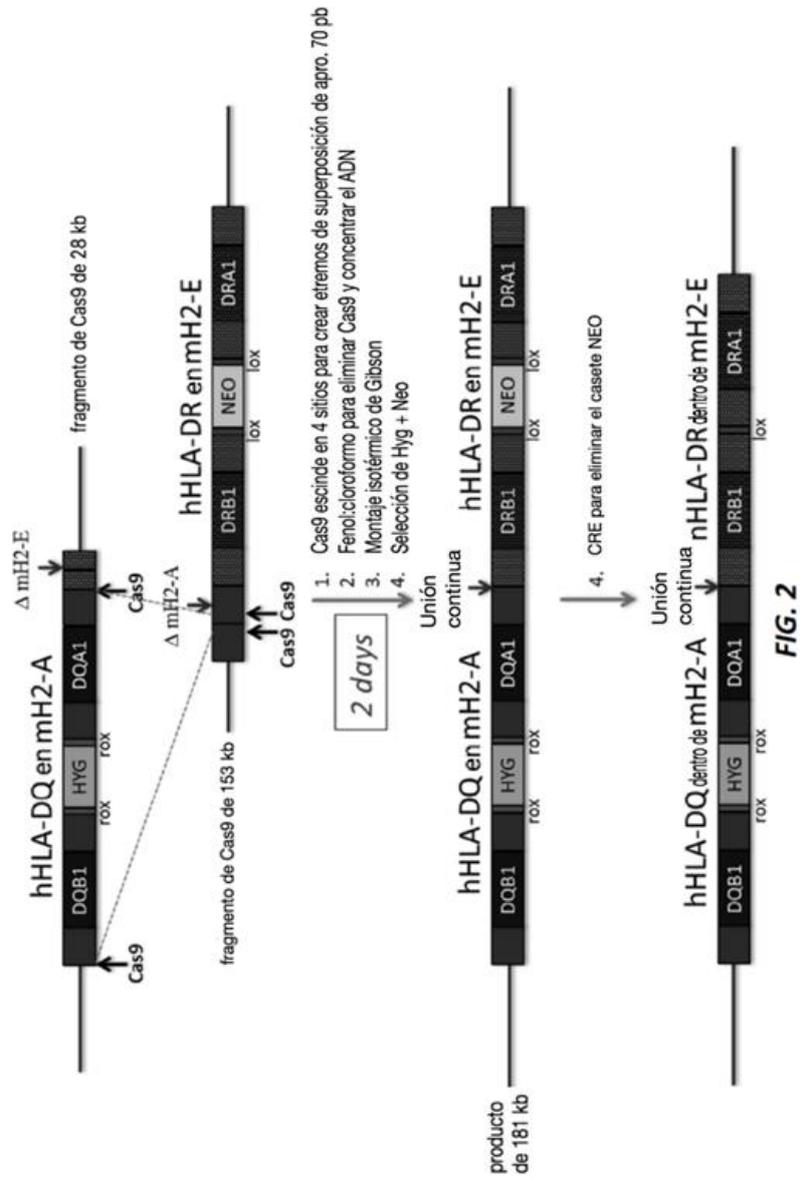


FIG. 1



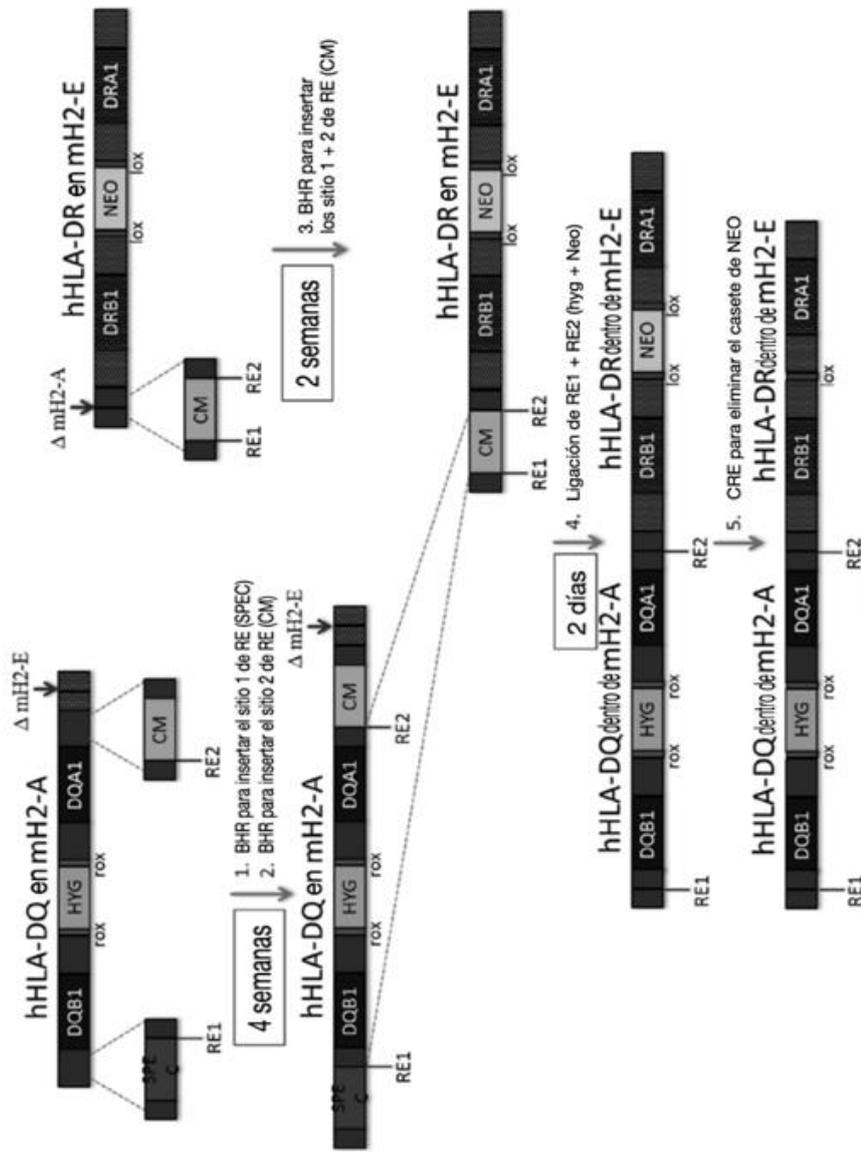


FIG. 3

Eficiencias de clonación para el método de ensamblaje de Cas9/Gibson

Construc ^{to}	Cas9 usada	Método para remover Cas9	# total de colonias	# de clones correctos (%)
6177(BDLC) + HYG	Cas9-6xHis	Proteinasa K	3	3 (100)
HLA-DQ + HLA-DR	6xHis-MBP-Cas9	Proteinasa K, fenol:cloroformo	1	1 (100)
		SDS, inact. por calor, fenol:cloroformo	1	1 (100)
		Inact. por calor, fenol:cloroformo	16	16 (100)

Tiempo requerido para las etapas de clonación de BAC §

Método	Tiempo
BHR	~ 1 semana
Ligación de BHR + BAC	> 2 semanas
Ligación de 2BHR + BAC	~ 5 semanas
Montaje de Cas9/Gibson	2 días
§ partiendo de ADN maxiprep de BAC terminando con colonias de E. coli	
BHR: recombinación homóloga bacteriana	

FIG. 4

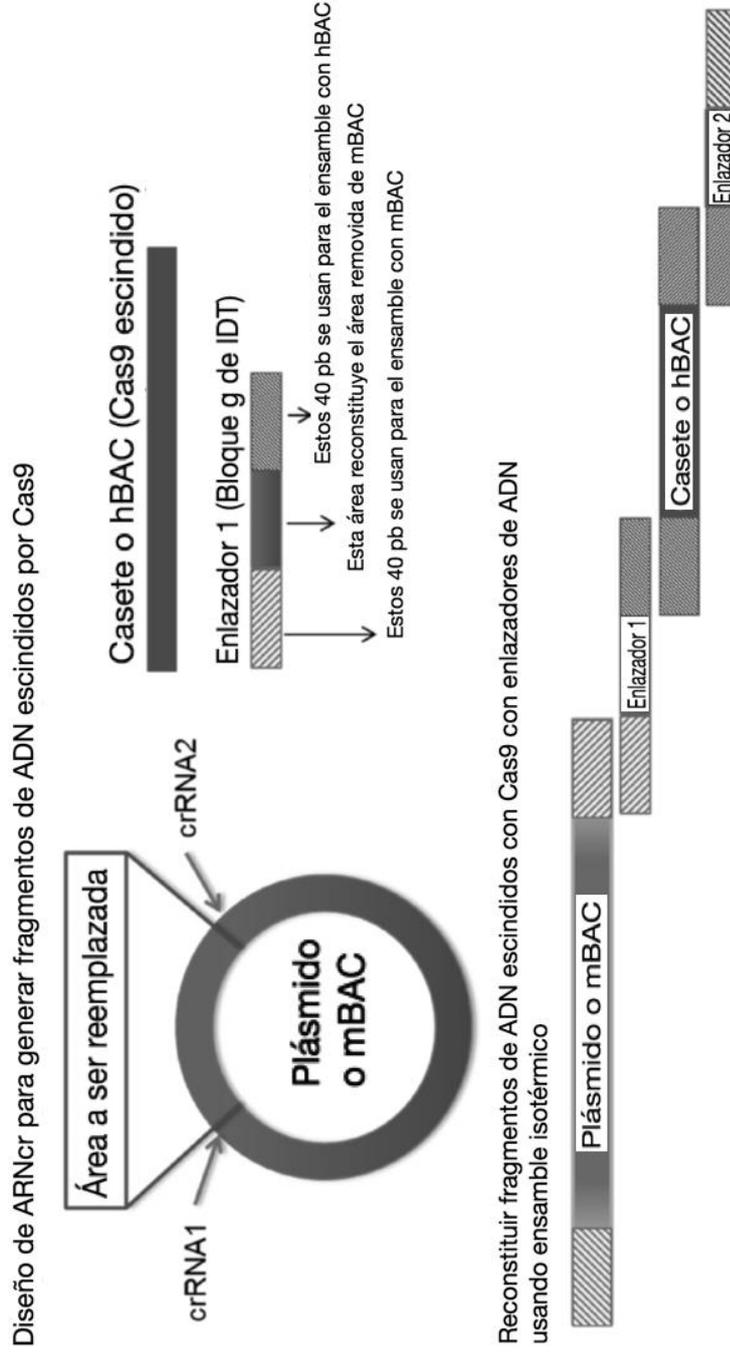
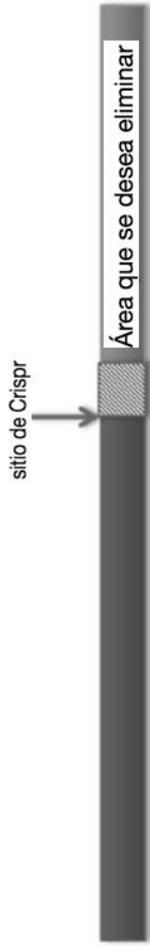


FIG. 5

Diseño de ARNcr para dirigir 5' secuencia arriba del área de interés:



La porción eliminada se reconstituirá mediante un oligo de unión durante el ensamble

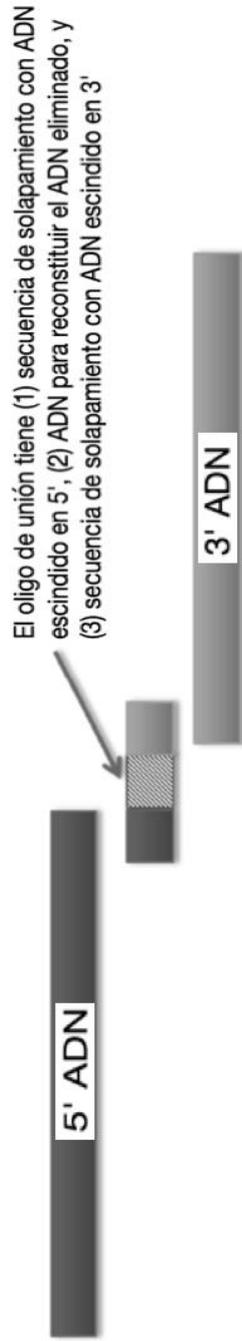
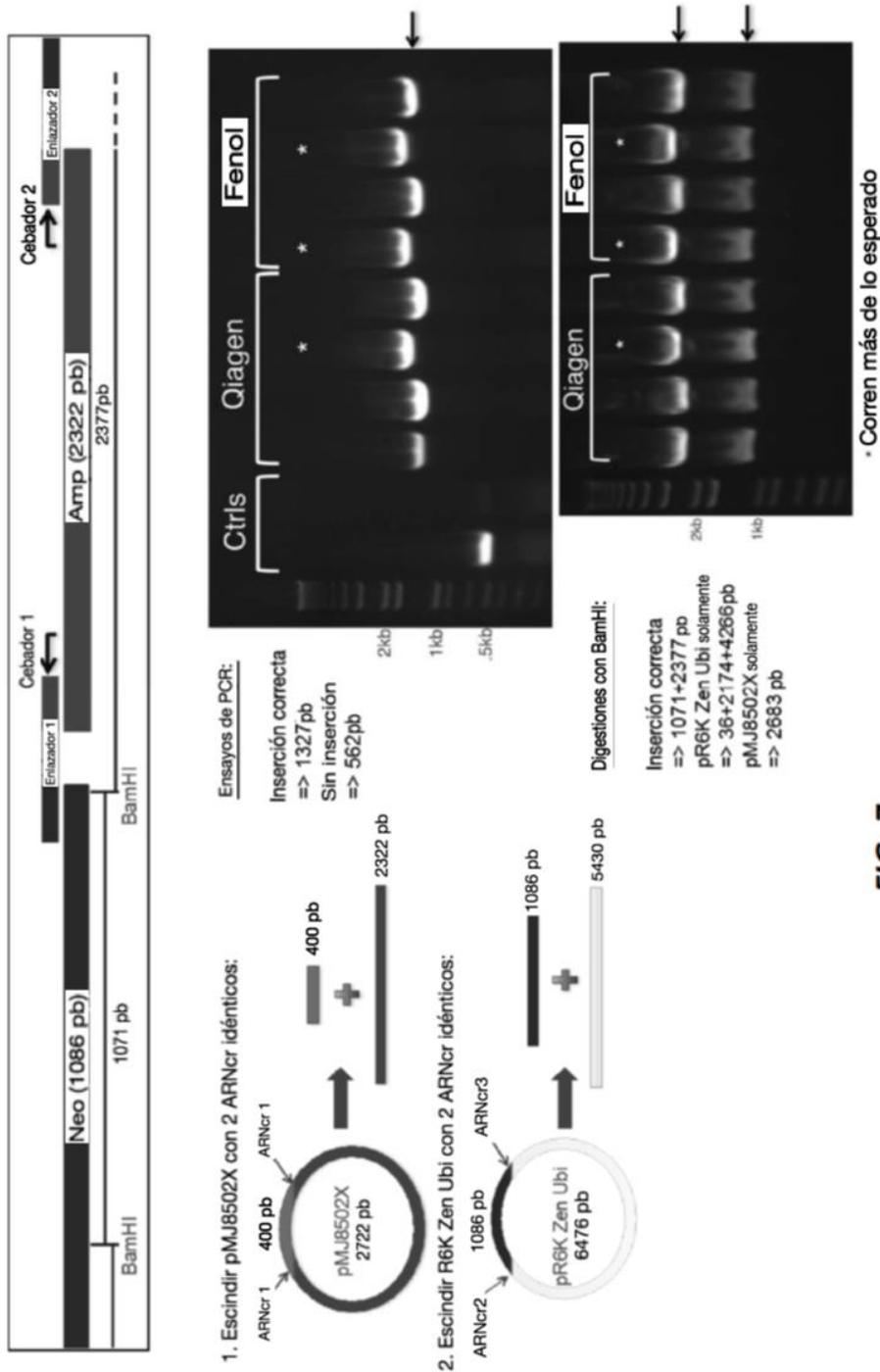


FIG. 6



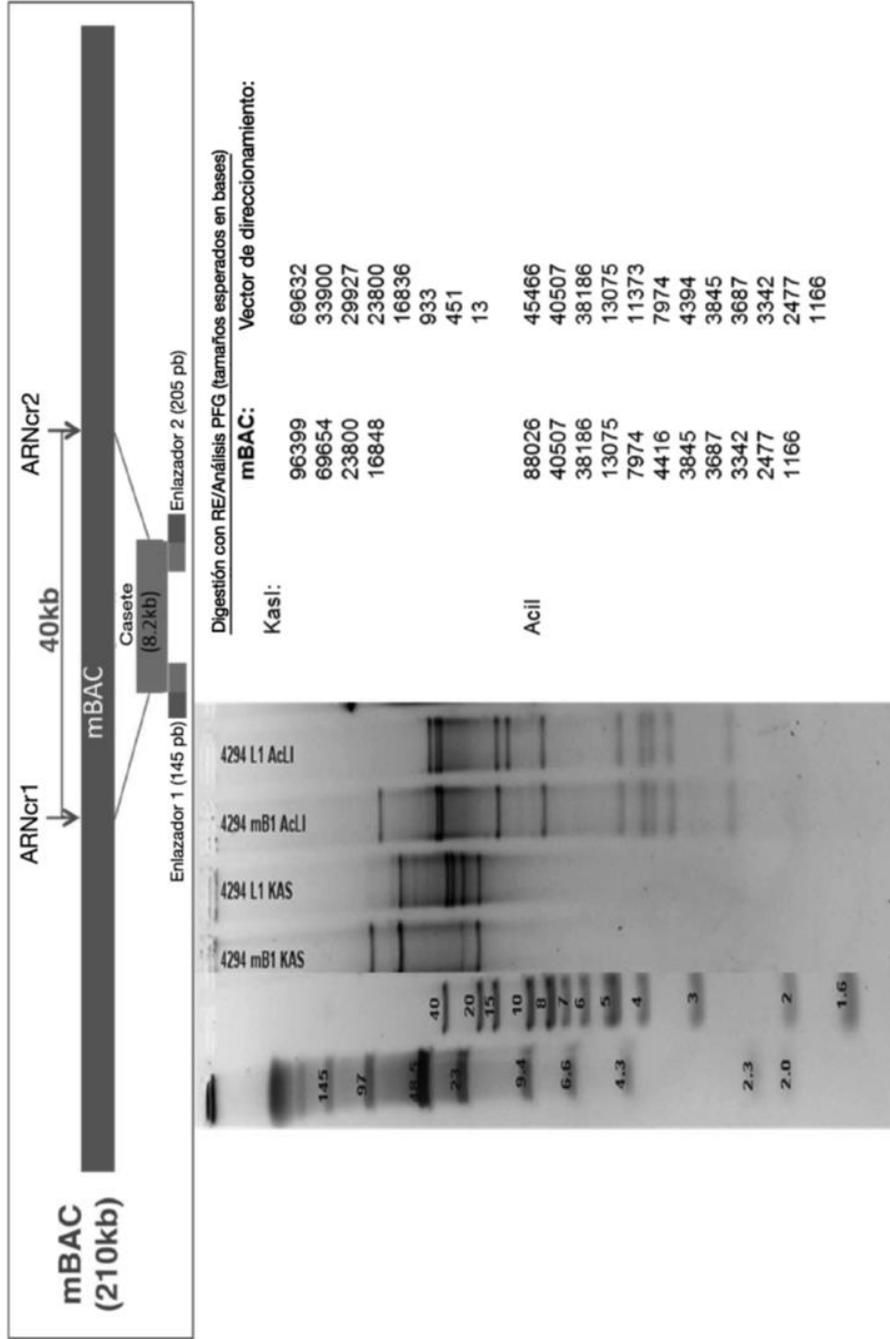


FIG. 8

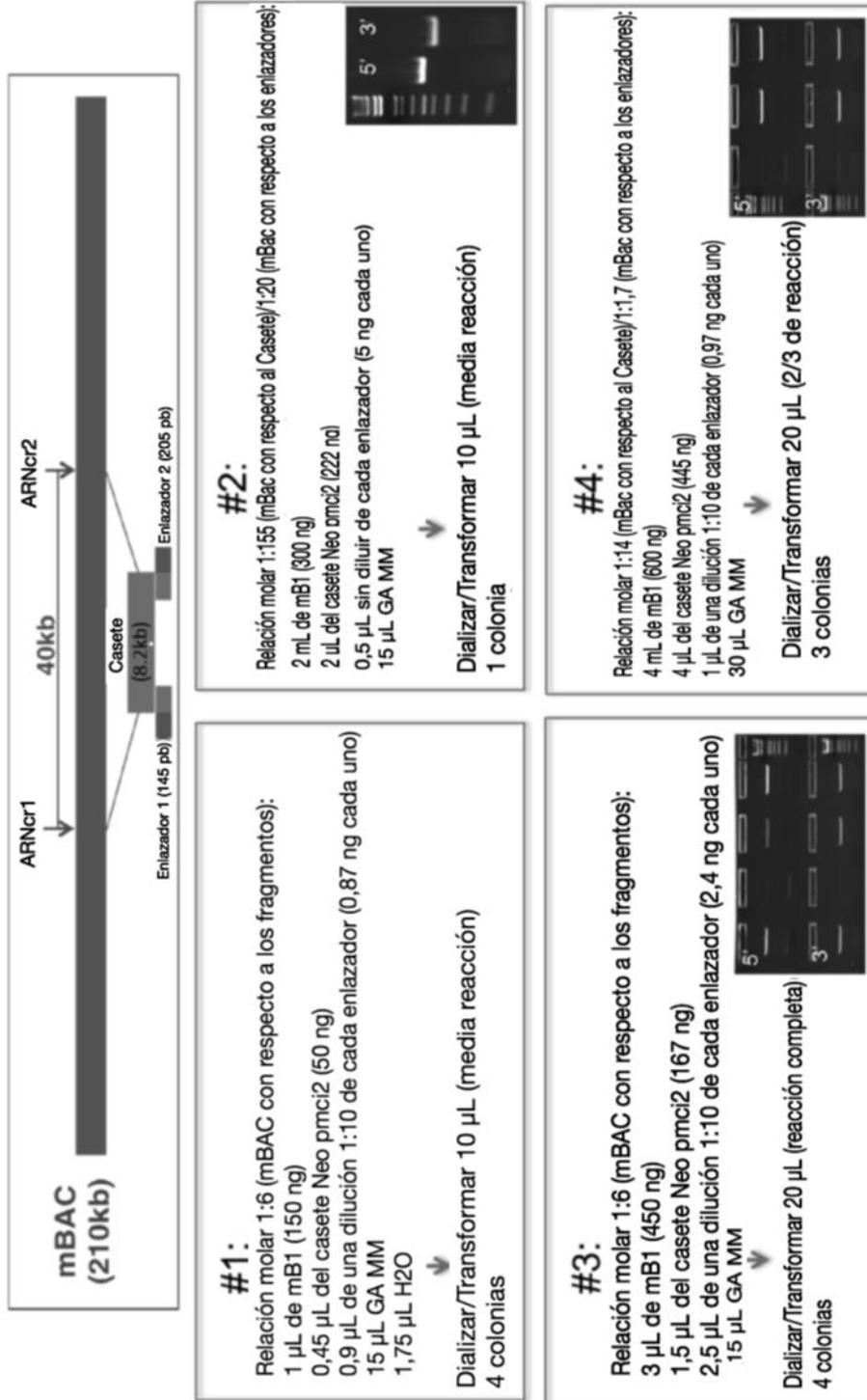
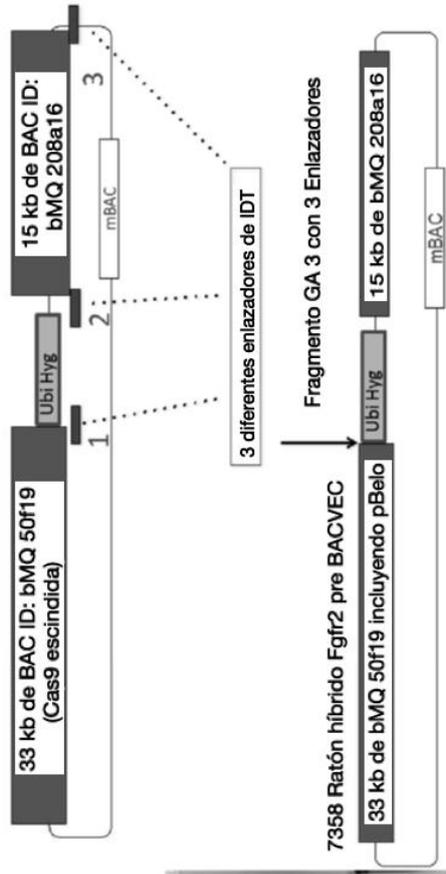


FIG. 9



- L2 y L3 coinciden con el LTVEC predicho
- PVU: 73,1; 11,4; 8,7; 6 y 2,6 kb
- Zra 51,5; 28,5; 10,8; 7,4; 2,3; 1,2 kb

Análisis PFG de mFgfr2 preBACVEC

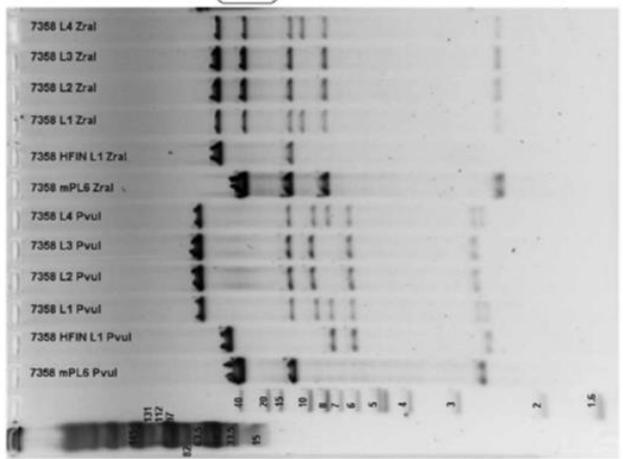


FIG. 11

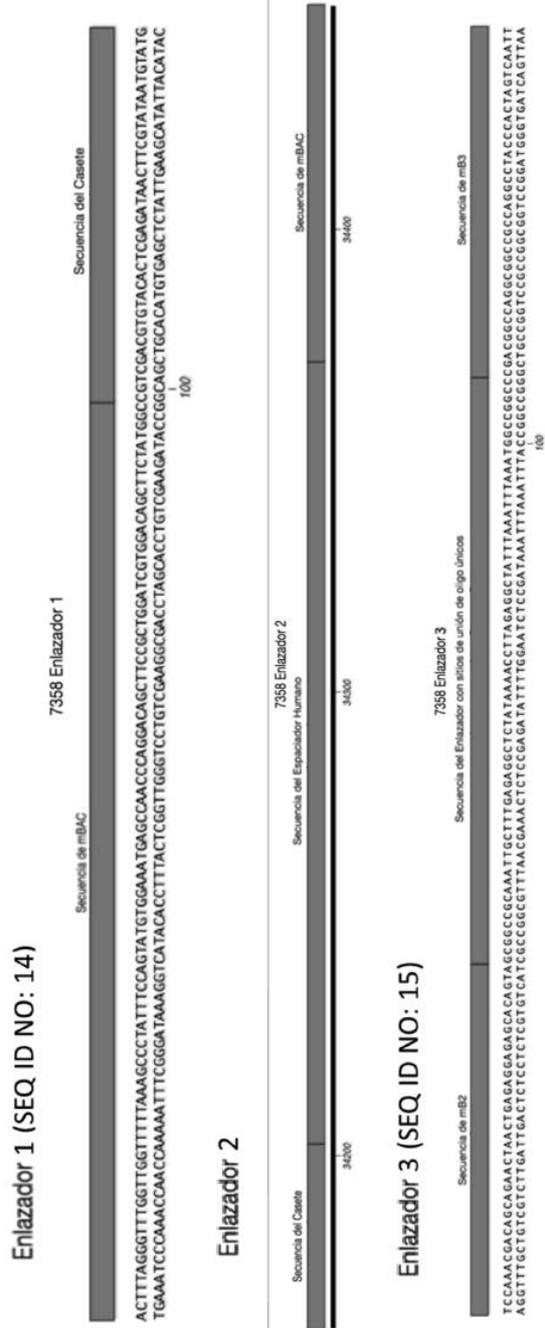


FIG. 12

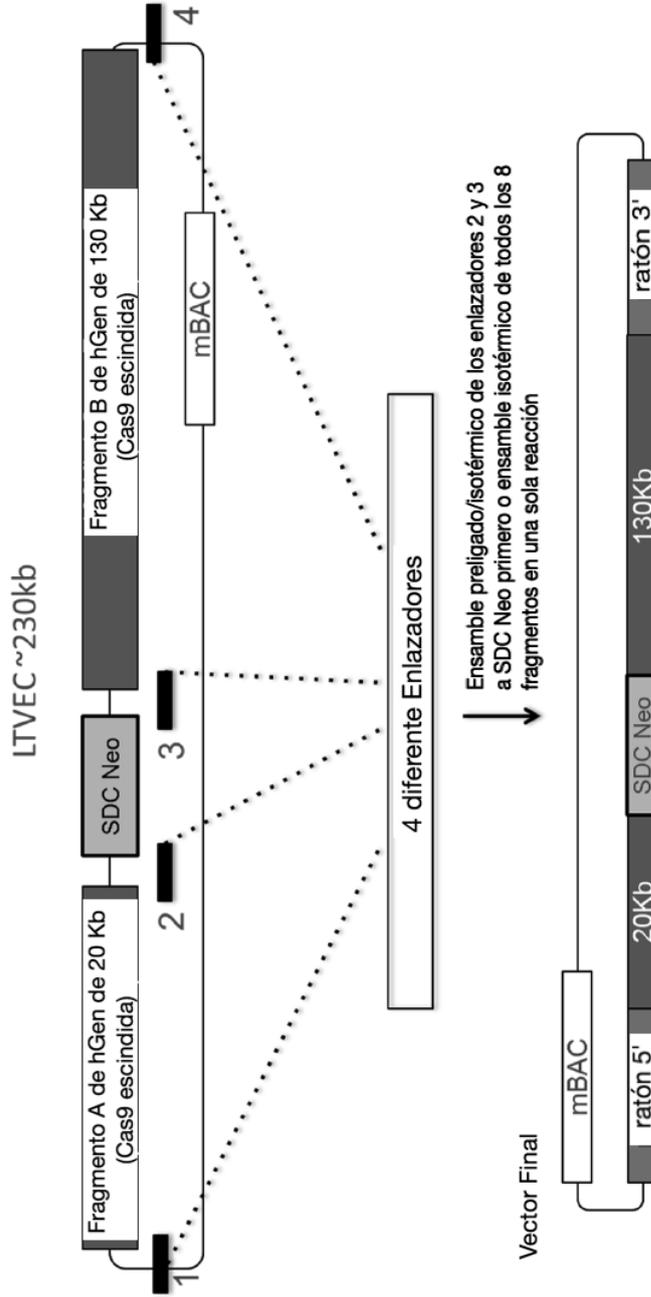


FIG. 13

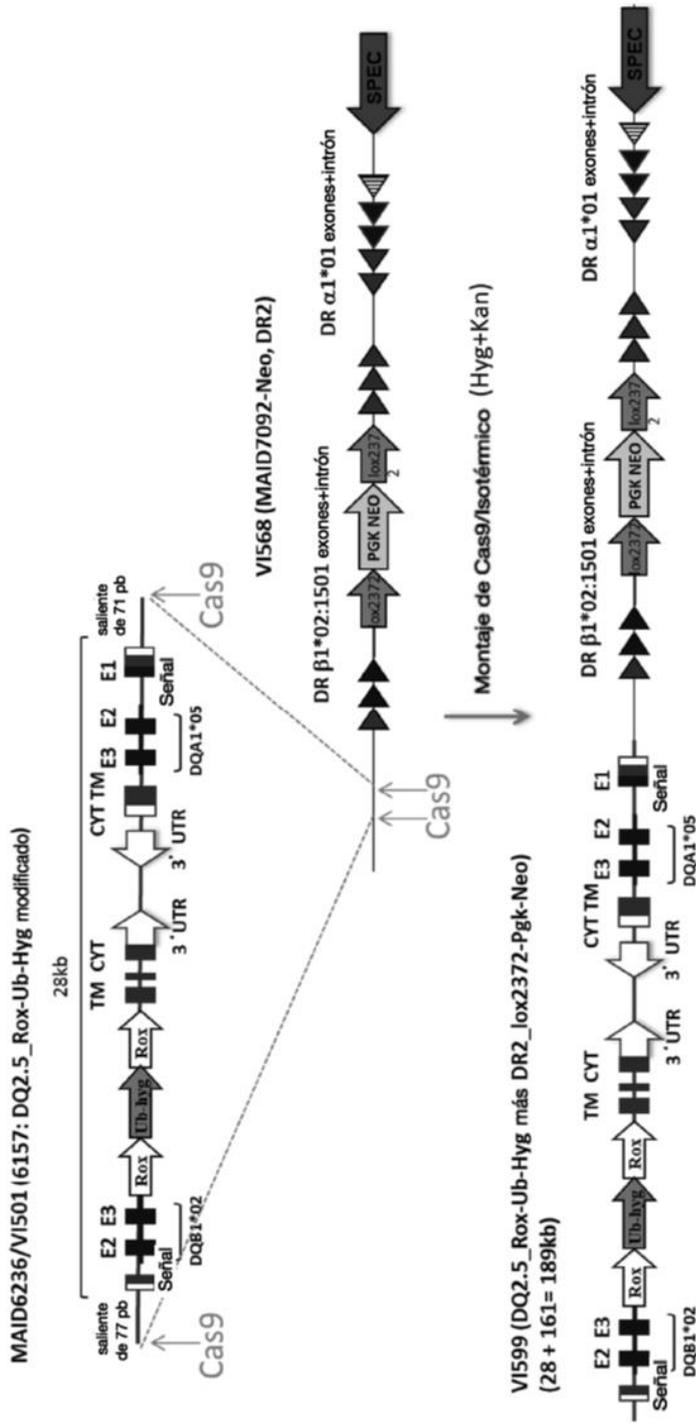


FIG. 14

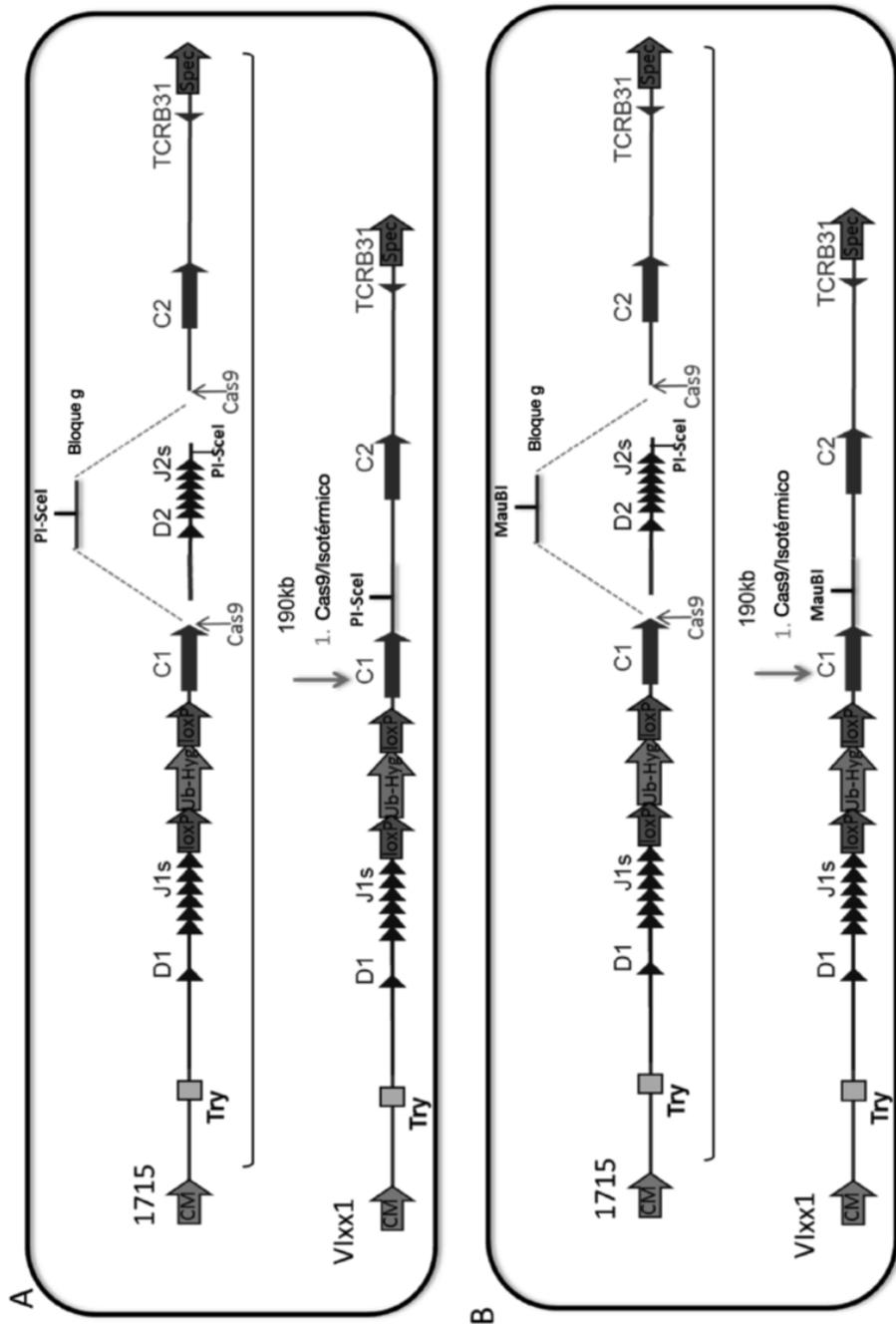


FIG. 15

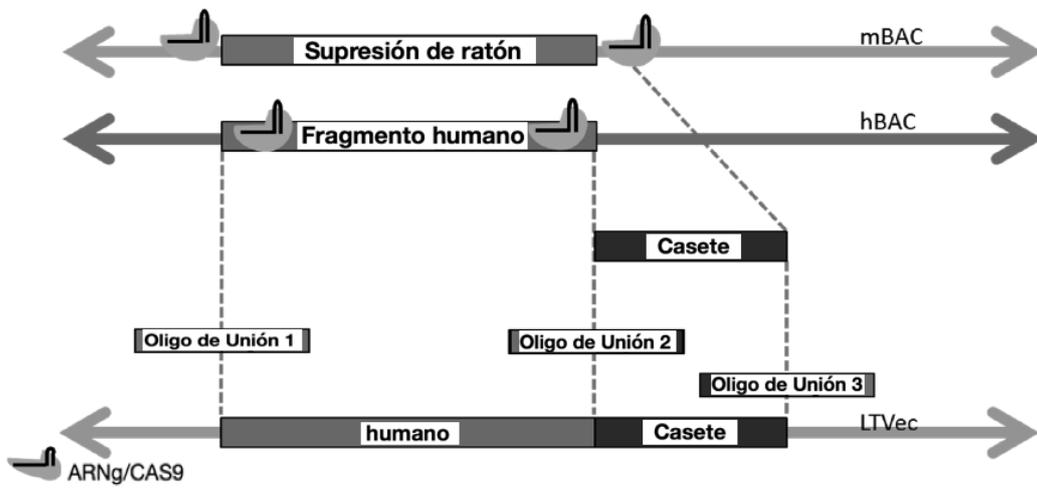
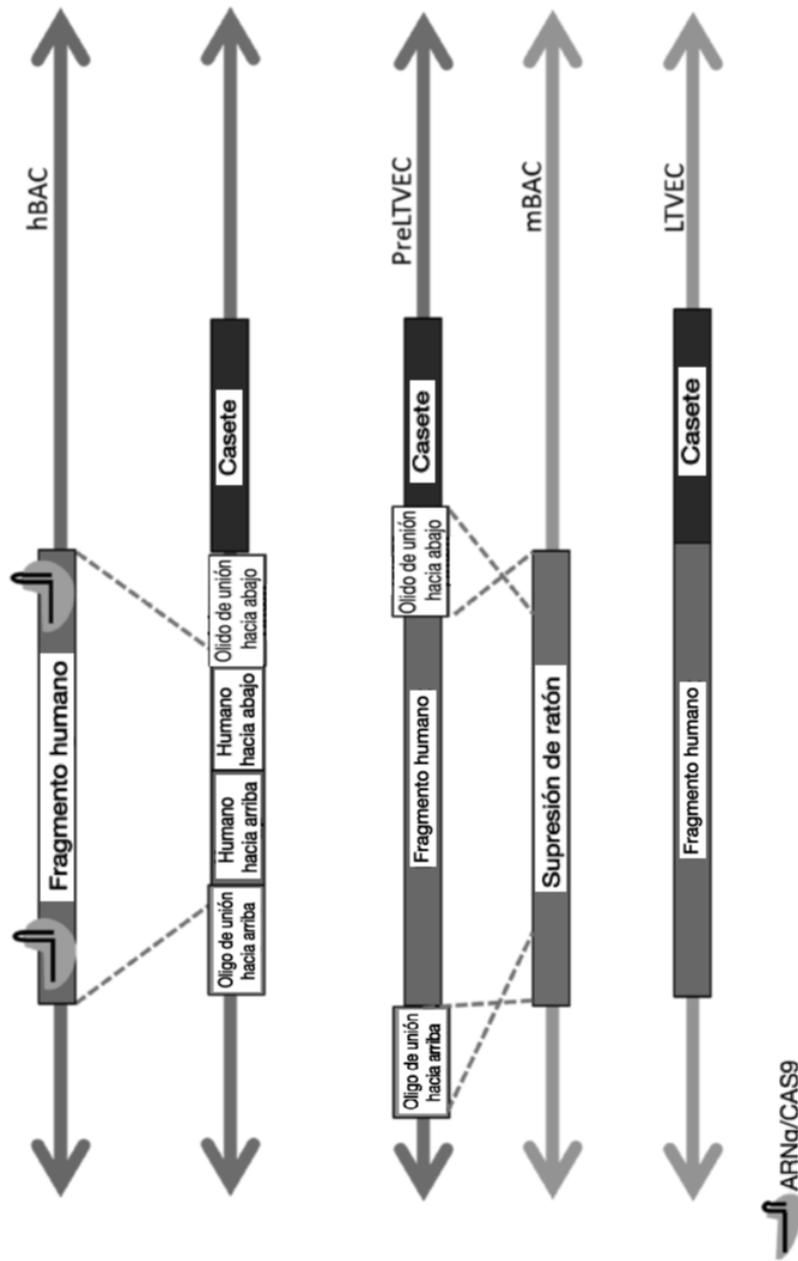


FIG. 16



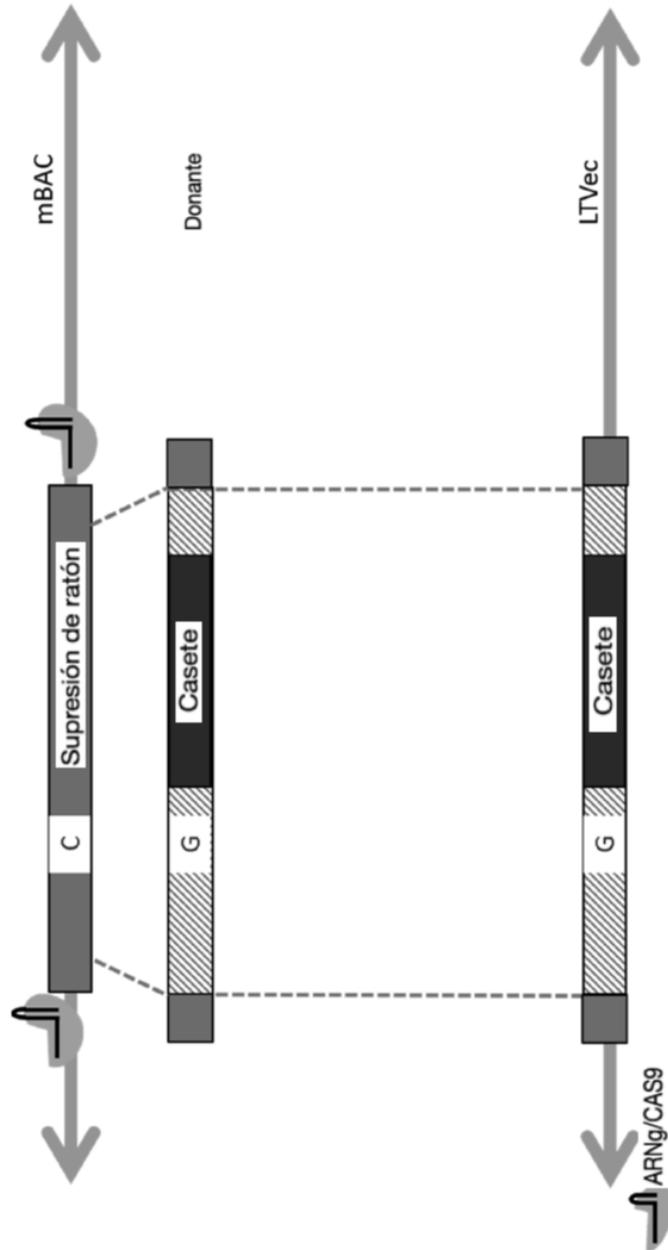


FIG. 18

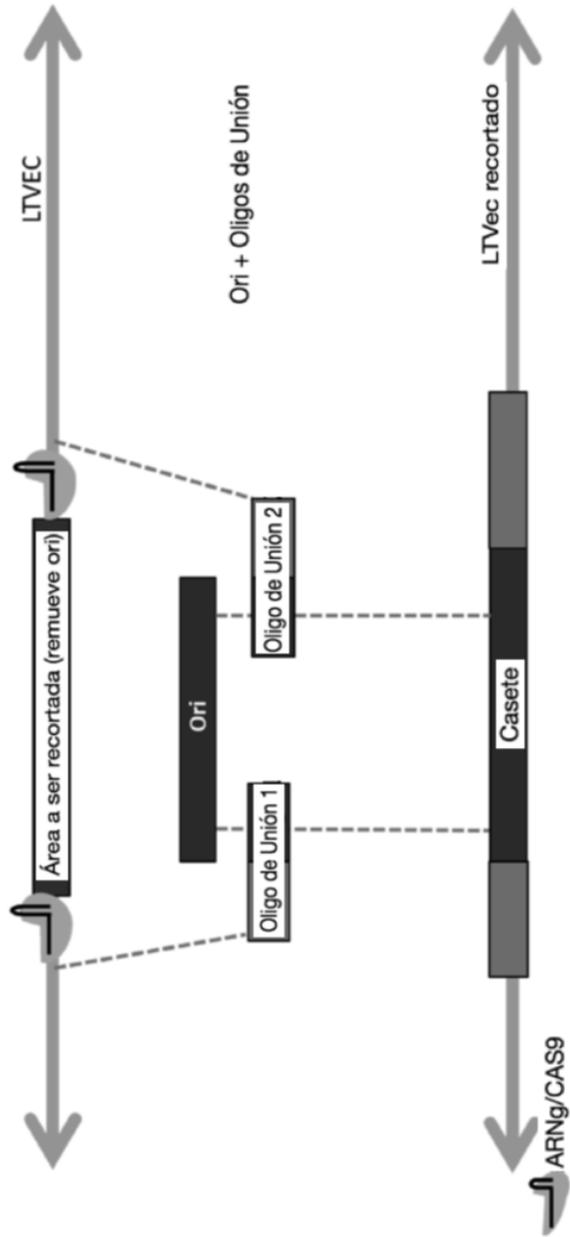


FIG. 19