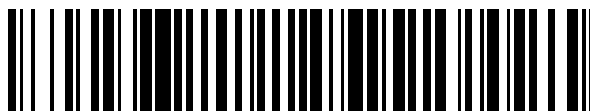


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 902 210**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/66 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.08.2017 PCT/CN2017/000483**

87 Fecha y número de publicación internacional: **27.09.2018 WO18170614**

96 Fecha de presentación y número de la solicitud europea: **02.08.2017 E 17901979 (9)**

97 Fecha y número de publicación de la concesión europea: **06.10.2021 EP 3604524**

54 Título: **Nueva técnica para la clonación directa de fragmentos genómicos grandes y el ensamblaje multimolecular de ADN**

30 Prioridad:

23.03.2017 CN 201710177676

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

25.03.2022

73 Titular/es:

**SHANDONG UNIVERSITY (50.0%)
Binhailu 72 Aoshanwei Street Jimo District
Qingdao, Shandong 266237, CN y
TECHNISCHE UNIVERSITÄT DRESDEN (50.0%)**

72 Inventor/es:

**ZHANG, YOUMING;
WANG, HAILONG;
FU, JUN y
STEWART, ADRIAN FRANCIS**

74 Agente/Representante:

**INGENIAS CREACIONES, SIGNOS E
INVENCIONES, SLP**

ES 2 902 210 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

Nueva técnica para la clonación directa de fragmentos genómicos grandes y el ensamblaje multimolecular de ADN

- 5 La divulgación se refiere a un método de clonación y ensamblaje de ácidos nucleicos, y más particularmente a un método de clonación y ensamblaje de ácido desoxirribonucleico (ADN).

La clonación de ADN es una técnica central utilizada en biología molecular y biotecnología para estudios de la función genética. Con el progreso de la tecnología de secuenciación de ADN y la disminución de los costes de secuenciación, la secuenciación de todo el genoma es cada vez más fácil. Los investigadores han descubierto que existen abundantes recursos sin explotar en el genoma. Los fragmentos cortos de ADN se obtienen fácilmente mediante PCR o síntesis química, pero el método convencional de clonación de fragmentos de ADN de más de 10 kb depende de la construcción y selección de bibliotecas de ADN. Los métodos convencionales de construcción y selección de bibliotecas son complicados, consumen mucho tiempo y son laboriosos, y los fragmentos de ADN diana a menudo se encuentran en varias regiones diferentes. Los investigadores subclonan los fragmentos, eliminan las secuencias redundantes o las integran en una ruta biosintética completa para investigar la función de los genes, por lo que los métodos convencionales de construcción y selección de bibliotecas ya no pueden satisfacer las necesidades de la época, y los investigadores necesitan con urgencia métodos genómicos simples, eficientes y rápidos. Técnicas de clonación y modificación de ADN. Con el progreso de las técnicas utilizadas en biología de síntesis, también se pueden ensamblar fragmentos de ADN grandes de más de 10 kb a partir de fragmentos pequeños seleccionados usando ensamblaje de Gibson *in vitro*¹ o ensamblador de ADN *in vivo*². Sin embargo, en el proceso de ensamblaje de ADN anterior, los fragmentos pequeños solo se preparan mediante PCR o síntesis química, por lo que es fácil introducir mutaciones aleatorias que causen inconvenientes en estudios posteriores de función de los genes. Los métodos anteriores requieren fragmentos de ADN que tengan una alta pureza y concentración, y no se pueden usar para ensamblar directamente un fragmento de ADN diana a partir de una mezcla de fragmentos de ADN genómico enzimático.

Una secuencia de ADN específica de ADN genómico se clona directamente en el vector, al que se hace referencia en el presente documento como clonación directa. Las proteínas recombinantes RecE y RecT del fago Rac son capaces de mediar eficazmente la recombinación homóloga entre moléculas de ADN lineal en las células de *E. coli*, también conocida como recombinación lineal-lineal. RecE es una exonucleasa 5' a 3', y RecT es una proteína de hibridación de ADN monocatenario (ADNmc), y la interacción proteína-proteína entre RecE y RecT es necesaria para la recombinación lineal-lineal, en particular, la eficiencia de la recombinación lineal-lineal de la acción combinada de RecET es 1000 veces mayor que la del efecto único. La técnica de clonación directa de RecET puede capturar directamente fragmentos de ADN genómico de más de 10 kb en un vector de expresión y también puede ensamblar 2-5 fragmentos de ADN. Se pueden encontrar sitios de enzimas de restricción adecuados en el genoma para liberar el fragmento de ADN diana, y luego el ADN genómico y el vector lineal digeridos con enzimas de restricción se electroporan en células de *E. coli* que expresan la RecET recombinasa. El fragmento de ADN diana y el vector lineal en las células se recombinan de forma homóloga mediante cajas homólogas en ambos extremos para formar un plásmido circular y, finalmente, las colonias que contienen la molécula de ADN recombinante se pueden obtener mediante selección de antibióticos y análisis de restricción. Después de preparar el vector lineal y el ADN genómico, solo se necesitan 3 días para completar la clonación directa del fragmento de ADN diana. En la actualidad, esta técnica se ha utilizado ampliamente en los estudios de clonación y expresión heteróloga de rutas de biosíntesis de productos naturales bacterianos. Por ejemplo, diez grupos³ de genes de policétido sintasas (PKS) de 10-52 kb/polipéptido sintetasa no ribosomal (NRPS) derivados de *Photorhabdus luminescens*, grupos⁴ de genes de sevadicina de 12 kb derivados de *Paenibacillus larvae*, grupos⁵ de genes de *siringolina* de 19 kb derivados de *Pseudomonas syringae*, grupos⁶ de genes de glidobactina de 25 kb derivados de *Burkholderia* DSM7029, grupos⁷ de genes de colibactina de 50 kb derivados de *E. coli* Nissle 1917.

Las ventajas de la clonación directa impulsan muchas de las investigaciones relevantes en el campo en comparación con la construcción y selección de bibliotecas. Larionov et al., establecieron una técnica⁸ de clonación de TAR que se basa en un sistema de recombinación de levadura, y luego combinan TAR con la escisión de Cas9 para aumentar la eficiencia de la clonación de TAR para una aplicación^{9,10} más fácil. La tecnología de direccionamiento de segmentos cromosómicos (CATCH) asistida por Cas9 establecida recientemente empalma grandes fragmentos de ADN en BAC¹¹ usando la escisión *in vitro* de Cas9 y ensamblaje de Gibson. Actualmente, este método solo se aplica a genomas de procariontes, y se requiere una selección previa de las colonias mediante PCR antes de la digestión con enzimas de restricción del ADN recombinante. De acuerdo con el conocimiento de los inventores, todos los métodos de clonación directa conocidos actualmente, incluidos los dos métodos anteriores, están limitados a la manipulación por parte de algunos expertos.

Aunque la clonación directa de RecET es conveniente para clonar fragmentos de ADN por debajo de 50 kb del genoma bacteriano^{3-6,12}, es difícil clonar fragmentos de ADN más grandes del genoma bacteriano o clonar fragmentos de ADN del genoma de mamífero. La razón es que la técnica de clonación directa de RecET se basa en la expresión intracelular de la RecET recombinasa, y la recombinación homóloga ocurre solo cuando el vector de clonación y el fragmento de ADN diana entran simultáneamente en las células de *E. coli* y se reúnen, lo que conduce a los siguientes defectos: (1) Limita el tamaño de los fragmentos de ADN genómico. Por ejemplo, no es posible clonar los fragmentos de ADN

mayores de 50 kb del genoma bacteriano o clonar los fragmentos del genoma mayores de 10 kb de mamíferos (tal como ratones y humanos). Esto se debe a que la concentración de un determinado fragmento de ADN en el genoma es muy baja, lo que da como resultado una pequeña probabilidad de que el fragmento de ADN diana y el vector de clonación entren en las células al mismo tiempo durante una transformación. Además, debido a la influencia de la fuerza de corte física durante la preparación del ADN genómico, cuanto más grande es el fragmento de ADN diana, menor es la concentración en el genoma, lo que da como resultado de que la probabilidad de que el fragmento de ADN diana entre a las células junto con el vector de clonación sea menor. Los genomas de ratones (2800 Mb) y humanos (3200 Mb) son casi tres órdenes de magnitud más grandes que el genoma bacteriano (5 Mb), por lo que la concentración de fragmentos de ADN genómico del mismo tamaño es mucho menor en el genoma de los mamíferos que en las bacterias. Por lo tanto, la probabilidad de que el fragmento de ADN diana entre en células de *E. coli* simultáneamente con el vector de clonación es mucho menor cuando se clona directamente un fragmento de ADN genómico de mamífero. (2) Afectada por el número de fragmentos de ADN, durante el ensamblaje de múltiples fragmentos, la eficiencia del ensamblaje se reduce drásticamente cuando el número de fragmentos de ADN supera los cuatro. En la actualidad, la técnica de clonación directa de RecET solo ensambla hasta cinco fragmentos de ADN. Para el ensamblaje de múltiples fragmentos, cuanto mayor sea el número de fragmentos de ADN, menor será la posibilidad de que entren en las células juntos y menor será la eficiencia del ensamblaje. Por lo tanto, el campo necesita particularmente una técnica de clonación directa fácil de operar y ampliamente utilizada que no esté limitada por el tamaño del fragmento de ADN resultante y la complejidad del genoma.

Para resolver el problema técnico de clonar directamente grandes fragmentos de ADN del genoma, los inventores primero escinden el ADN genómico y el vector de clonación con exonucleasa *in vitro*, luego el producto de reacción *in vitro* se recombinó homológamente en presencia de RecET recombinasa para establecer la técnica de clonación ExoCET. La técnica ExoCET puede clonar fragmentos de ADN de más de 100 kb directamente del genoma bacteriano y clonar fragmentos de ADN de más de 50 kb de células de mamíferos y sangre humana. La técnica ExoCET también es capaz de ensamblar de manera eficiente al menos veinte fragmentos de ADN para formar un plásmido completo.

Al igual que la técnica de clonación directa de RecET, ExoCET tiene una ventaja comparativa sobre la PCR en que tiene alta fidelidad y no destruye el haplotipo de ADN, y el ADN diana se clona directamente en un vector plásmido para facilitar los estudios de expresión. Además, ExoCET es más eficiente que el ensamblaje de Gibson porque Gibson se basa en moléculas de ADN circulares producidas por ensamblaje *in vitro*. Además, debido a la ciclación del propio portador vacío, el ensamblaje de Gibson puede producir un trasfondo muy serio en el proceso de clonación directa, que puede ser la razón por la que Jiang et al.,¹¹ realizan una preselección por PCR antes de identificar el ADN recombinante; también es la razón por la que Zhou et al.,¹³ necesita eliminar el fragmento de ADN genómico de 20 kb o menor generado por digestión de restricción y electroforesis en gel de agarosa cuando se utiliza ensamblaje de Gibson para clonar el grupo de genes de conglobatina del genoma de *Streptomyces globatus*.

ExoCET también se puede usar para clonar directamente fragmentos de ADN de genomas de mamíferos tales como sangre, líneas celulares asociadas a enfermedades, etc., para facilitar los estudios de haplotipos de polimorfismos de un solo nucleótido (SNP) y para construir rápidamente un vector de direccionamiento isogénico de haplotipos (HIT) para el direccionamiento de células madre humanas mediado por nucleasa. La importancia de la reprogramación de células madre humanas aisladas de pacientes, sangre del cordón umbilical o células somáticas en la investigación biomédica ha recibido cada vez más atención, y la investigación sobre la modificación precisa del genoma de las células madre también ha recibido una atención generalizada. Modificar el genoma humano es más desafiante que modificar el genoma de los ratones de laboratorio debido a la compleja diversidad genética humana. La importancia de la isogenicidad (similitud de secuencia) para la recombinación homóloga se hizo evidente hace muchos años cuando la gente usaba células madre embrionarias de ratón para el direccionamiento de genes¹⁴. Sin embargo, no se ha estudiado bien el efecto del desajuste de secuencias sobre la recombinación homóloga. Por ejemplo, ¿cuánta eficiencia de recombinación reduce un solo desajuste (un SNP o una inserción/eliminación)? ¿Cómo afecta la distancia entre el sitio no coincidente y el sitio de recombinación la eficacia de la recombinación? ¿Cómo afectan los múltiples desajustes a la eficiencia de la recombinación? Estos problemas anteriores aún no se han aclarado. En cualquier caso, el uso de secuencias idénticas en el direccionamiento de genes es claramente muy recomendable, por lo que ExoCET es una técnica eficaz para obtener rápidamente cajas de homología ideales. A diferencia del método de amplificación de cajas de homología del genoma mediante PCR, ExoCET no está limitado por el tamaño de los fragmentos, no introduce mutaciones y es capaz de mantener un haplotipo de ADN. Además, los extremos de las cajas de homología también se pueden seleccionar de acuerdo con las formas de genotipificación (tales como transferencia Southern o PCR de ligación), optimizando la longitud de las cajas de homología. ExoCET, por lo tanto, ofrece ventajas para la cirugía genómica individualizada, especialmente cuando se combina con CRISPR/Cas9¹⁵. ExoCET también se puede utilizar como el método más confiable para la genotipificación de un genoma modificado, mientras que la transferencia Southern y la PCR de ligación y tienen señales positivas falsas.

Fu, J et al. Nature Biotech (2012) 30 (5): 440-446; el documento US 2015/072381; y Kolodner R et al., Mol Microbiol (1994) 11 (1): 23-30 divulgan métodos basados en la recombinación homóloga de RecET.

ExoCET es capaz de capturar selectivamente grandes fragmentos de ADN de genomas complejos, lo que lo convierte en un medio de diagnóstico de enfermedades y pruebas patológicas. Por ejemplo, capturar secuencias de ADN directamente de medicamentos individualizados o aislar virus de ADN de muestras de pacientes. ExoCET se utilizará

ampliamente en la investigación de genómica funcional y genómica comparativa, especialmente para la clonación directa de vías de síntesis de procariotas o el ensamblaje de múltiples moléculas de ADN para biología de síntesis.

La invención está definida por las reivindicaciones adjuntas.

En un aspecto, la divulgación proporciona un método de recombinación homóloga, comprendiendo el método tratar dos o más moléculas de ácido nucleico diana con una primera exonucleasa, recombinando las dos o más moléculas de ácido nucleico diana después de ser tratadas en presencia de una segunda exonucleasa y una proteína de hibridación, cuando las moléculas de ácido nucleico diana recombinadas comparten al menos una secuencia homóloga.

En otro aspecto, la divulgación proporciona un método de recombinación homóloga, el método comprende tratar una primera molécula de ácido nucleico y una segunda molécula de ácido nucleico con una primera exonucleasa, recombinando la primera molécula de ácido nucleico y la segunda molécula de ácido nucleico después de ser tratadas en presencia de una segunda exonucleasa y una proteína de hibridación, en la que la primera molécula de ácido nucleico y la segunda molécula de ácido nucleico comparten al menos una secuencia homóloga.

En otro aspecto más, la divulgación proporciona un método para ensamblar una molécula de ácido nucleico lineal, comprendiendo el método tratar dos o más moléculas de ácido nucleico con una primera exonucleasa, recombinando las dos o más moléculas de ácido nucleico después de ser tratadas en presencia de una segunda exonucleasa y una proteína de hibridación, en la que cada molécula de ácido nucleico comparte al menos una secuencia homóloga con una molécula de ácido nucleico adyacente en un producto de ensamblaje resultante.

Además, la divulgación proporciona un método de clonación de ADN genómico, el método comprende tratar una mezcla de fragmentos de ADN genómico y un vector de clonación lineal con una primera exonucleasa, recombinando la mezcla de fragmentos de ADN genómico y el vector de clonación lineal después de ser tratado en presencia de una segunda exonucleasa y una proteína de hibridación, en la que el vector de clonación lineal comparte al menos una secuencia homóloga con el fragmento de ADN diana de la mezcla de fragmentos de ADN genómico.

Al menos una secuencia homóloga puede estar ubicada dentro o en un extremo de las dos o más moléculas de ácido nucleico diana, particularmente, al menos una secuencia homóloga en un extremo de una molécula de ácido nucleico diana, y más particularmente, todas las secuencias homólogas en un extremo de las moléculas de ácido nucleico diana.

Al menos una secuencia homóloga tiene una longitud de al menos 6, al menos 10, al menos 20, al menos 25, al menos 30, al menos 40, al menos 50, al menos 60, al menos 70, al menos 80 nucleótidos, particularmente 25, 40 u 80 nucleótidos, y más particularmente 80 nucleótidos.

La primera exonucleasa puede ser una exonucleasa 5' a 3' o una exonucleasa 3' a 5', particularmente ADN polimerasa T4, fragmento Klenow de ADN polimerasa I, exonucleasa T5, exonucleasa T7 y más particularmente ADN polimerasa T4 o T5 exonucleasa.

El tratamiento en presencia de la primera exonucleasa comprende la unión *in vitro*. La unión *in vitro* puede unir dos o más moléculas de ácido nucleico diana o la primera molécula de ácido nucleico a la segunda molécula de ácido nucleico, o unir el vector de clonación lineal tratado al fragmento de ADN diana de una mezcla de fragmentos de ADN genómico.

El tratamiento en presencia de la primera exonucleasa comprende digestión e hibridación con enzimas; la digestión enzimática de diferentes moléculas de ácido nucleico se puede realizar por separado o simultáneamente, tal como en una mezcla de muestras.

El tratamiento en presencia de la primera exonucleasa comprende además añadir una ADN polimerasa, dNTP y ADN ligasa.

El tratamiento en presencia de la primera exonucleasa comprende además la adición de una ADN polimerasa que tiene actividad exonucleasa 3' a 5'.

El tratamiento en presencia de la primera exonucleasa excluye la adición de dNTP.

El tratamiento en presencia de la primera exonucleasa comprende el tratamiento con ADN polimerasa T4 o el ensamblaje de Gibson.

La segunda exonucleasa es RecE y, en particular, la RecE es un producto de expresión recombinante.

La proteína de hibridación monocatenaria incluye RecA, RAD51, Redβ, RecT, Pluβ o RAD52, y particularmente, la proteína de hibridación es RecT, más particularmente, la RecT es un producto de expresión recombinante.

La proteína de hibridación es RecT, particularmente, la RecT es un producto de expresión recombinante.

La recombinación homóloga se lleva a cabo *in vitro* o en una célula huésped.

5 La célula huésped puede ser una célula de levadura, particularmente la célula de levadura es una célula de *Saccharomyces cerevisiae*; o una célula bacteriana, particularmente la célula bacteriana es *Bacillus subtilis* o *Escherichia coli*.

La célula huésped expresa una exonucleasa, particularmente una segunda exonucleasa y una proteína de hibridación.

10 La célula huésped expresa una exonucleasa, una proteína de hibridación y Red γ . En particular, la célula huésped expresa además RecA, más particularmente, la célula huésped expresa RecE, RecT, Red γ y RecA.

15 La célula huésped es una célula de *E. coli* que expresa RecE y/o RecT de longitud completa, particularmente, la célula huésped es una célula de *E. coli* que expresa RecE, RecT y Red γ de longitud completa, más particularmente la célula huésped es una célula de *E. coli* que expresa RecE, RecT, Red γ y RecA de longitud completa.

La célula huésped es una célula de *E. coli* que expresa RecE y RecT truncadas.

20 La célula huésped es una célula de *E. coli* que expresa Red α y Red β .

La célula huésped expresa una exonucleasa en un vector plásmido y/o cromosoma, particularmente, la segunda exonucleasa, la proteína de hibridación, Red γ y/o RecA, particularmente, que son expresadas por un vector plásmido, y más particularmente, expresadas por un vector plásmido y un cromosoma simultáneamente.

25 La molécula de ácido nucleico diana o el fragmento de ADN diana es un segmento de ADN lineal seleccionado de un fragmento de ADN digerido por endonucleasa, un fragmento de ADN amplificado por PCR, un fragmento de ADN genómico, un miembro de la biblioteca de ADNc, un fragmento derivado de cromosomas artificiales bacterianos (BAC) y un fragmento de vectores de clonación.

30 La endonucleasa puede ser una enzima de restricción o una endonucleasa programable, tal como Cas9.

El número de moléculas de ácido nucleico diana o fragmentos de ADN es 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más.

35 La molécula de ácido nucleico diana comprende una secuencia de 0,5 kb o más (por ejemplo, 1 kb o más, 2,5 kb o más, 4 kb o más, 5 kb o más, 7,5 kb o más, 10 kb o más, 15 kb o más, 20 kb o más, 25 kb o más, 40 kb o más, 50 kb o más, 75 kb o más o 100 kb o más).

40 Las dos o más moléculas de ácido nucleico diana, la primera molécula de ácido nucleico diana y la segunda molécula de ácido nucleico diana o fragmentos de ADN diana comprenden uno o más fragmentos de ADN amplificados por PCR, fragmentos de ADN genómico, miembros de la biblioteca de ADNc y/o un fragmento derivado de BAC.

La primera exonucleasa tiene actividad exonucleasa de 3' a 5' y, en particular, es ADN polimerasa T4.

45 El tratamiento con una exonucleasa se lleva a cabo *in vitro* en ausencia de dNTP.

La segunda exonucleasa es una RecE de longitud completa.

50 La proteína de hibridación es RecT.

La recombinación homóloga se lleva a cabo en una célula huésped bacteriana que expresa RecE y RecT de longitud completa, particularmente en *Escherichia coli*.

55 Las dos o más moléculas de ácido nucleico diana comprenden uno o más fragmentos de ADN amplificados por PCR, fragmentos de ADN genómico, miembros de la biblioteca de ADNc y/o un fragmento derivado de BAC, plásmido lineal y/o fragmento de vector de clonación, y particularmente, tres o más plásmidos lineales y/o fragmentos de vector de clonación.

60 La primera exonucleasa comprende un ensamblaje de Gibson.

La segunda exonucleasa es RecE de longitud completa.

La proteína de hibridación es RecT.

65 La recombinación homóloga se lleva a cabo en una célula huésped bacteriana que expresa RecE y RecT de longitud

completa, particularmente en *E. coli*.

Las dos o más moléculas de ácido nucleico diana incluyen tres o más fragmentos de ADN amplificados por PCR, fragmentos de ADN genómico, miembros de la biblioteca de ADNc y/o un fragmento derivado de BAC, plásmidos lineales y/o fragmentos de vector de clonación, particularmente tres o más plásmidos lineales y/o fragmentos de vector de clonación.

La primera exonucleasa comprende un ensamblaje de Gibson.

La segunda exonucleasa es RecE de longitud completa.

La proteína de hibridación es RecT.

La recombinación homóloga se lleva a cabo en una célula huésped bacteriana que expresa RecE y RecT de longitud completa, particularmente en *E. coli*.

También se proporciona un kit que comprende la primera exonucleasa y la segunda exonucleasa descritas en el método anterior, o un ácido nucleico que codifica la primera exonucleasa y la segunda exonucleasa descritas en el método anterior.

Un kit comprende la primera exonucleasa y la segunda exonucleasa descritas en el método anterior, o un ácido nucleico que codifica la primera exonucleasa y la segunda exonucleasa. En particular, el kit comprende además una célula huésped que expresa la segunda exonucleasa, en particular, la célula huésped expresa una exonucleasa, una proteína de hibridación y Redy, en particular, la célula huésped también expresa RecA, más particularmente, la célula huésped expresa RecE, RecT, Redy y RecA, la célula huésped puede ser una célula de levadura, particularmente la célula de levadura es una célula de *Saccharomyces cerevisiae*; o una célula bacteriana, particularmente la célula bacteriana es *Bacillus subtilis* o *Escherichia coli*. La célula huésped expresa una exonucleasa, una proteína de hibridación, Redy y/o RecA en un vector plásmido y/o cromosoma simultáneamente, particularmente, expresada por un vector plásmido, más particularmente, expresado por un vector plásmido o cromosoma, además particularmente, el kit puede comprender además uno o más vectores lineales preparados previamente.

La primera exonucleasa es una ADN polimerasa que tiene actividad exonucleasa 3' a 5', tal como ADN polimerasa T4, fragmentos Klenow de ADN polimerasa I, exonucleasa T5 o exonucleasa T7, la segunda exonucleasa es RecE de longitud completa.

El kit comprende además una célula huésped que expresa una segunda exonucleasa, particularmente, la célula huésped comprende un ácido nucleico que codifica RecE, RecT, Redy y RecA de longitud completa.

El kit comprende además uno o más vectores lineales preparados previamente.

Uso del método o kit antes mencionado en la construcción de un vector de direccionamiento.

Uso del método o kit antes mencionado en la genotipificación de células de mamífero.

Uso del método o kit antes mencionado en la síntesis de ADN.

La Figura 1 muestra la acción concertada del ensamblaje *in vitro* y RecE/RecT de longitud completa mejora la eficacia de la clonación directa. (a): diagrama esquemático de la clonación directa del grupo de genes *lux* de 14 kb del genoma de *P. phosphoreum* ANT-2200. El vector p15A-cm y el fragmento de ADN genómico diana tienen la misma secuencia al final. (b): Los brazos de homología más largos aumentan la eficiencia de clonación de ExoCET. Los vectores lineales con cajas de homología de 25 pb, 40 pb u 80 pb se mezclaron con ADN genómico, respectivamente, y la reacción se llevó a cabo durante 20 minutos a 25 °C con 0,02 U μl^{-1} de T4pol antes del hibridación y conversión a GB05-dir de *E. coli* inducida por arabinosa. Barras de error, DE; n = 3. (c): Optimización de la concentración de T4pol. El vector lineal con cajas de homología de 80 pb y el ADN genómico se trataron de la misma manera que en (b), excepto por la diferente concentración de T4pol. (d): Efecto del tiempo de incubación de T4pol sobre la eficacia de la clonación. Al igual que con (c), se utilizaron 0,02 U μl^{-1} de T4pol, pero con diferentes tiempos de incubación. (e): Un mayor número de copias de ETgA aumenta la eficiencia de clonación de ExoCET. Al igual que con (d), el tiempo de incubación fue de 1 h, y luego los productos ensamblados *in vitro* se electroporaron por separado en *E. coli* GB05-dir inducida por arabinosa (con una copia de ETgA en el cromosoma), que contenía GB2005 (con 5 copias de ETgA en el plásmido pSC101) de pSC101-BAD-ETgA-tet o que contienen GB05-dir (con 6 copias de ETgA) de pSC101-BAD-ETgA-tet. (f): ExoCET aumenta la eficiencia de la clonación directa. Al igual que con (e), se usó GB05-dir (ExoCET) que contenía pSC101-BAD-ETgA-tet, o no se usó tratamiento con T4pol (ETgA) o no se usó inducción de arabinosa (T4pol).

La Figura 2 muestra el efecto de diferentes exonucleasas sobre la clonación directa del grupo de genes *lux*. El vector p15A-cm y el ADN genómico de *P. phosphoreum* se trataron con exonucleasa, se hibridaron y luego se electroporaron en *E. coli* GB05-dir inducida por arabinosa. (a): Detección inicial de diferentes exonucleasas. (b-d): Optimización de

las concentraciones de Kle, T5exo y T7exo. (e): Comparación de la eficiencia de clonación después de 20 min de digestión con T4pol, Kle, T5exo y T7exo a concentraciones óptimas. (f): Optimización de la temperatura y el tiempo de digestión de T4pol (0,02 U μl^{-1}). Barras de error, DE; n = 3.

5 La Figura 3 muestra el efecto de la velocidad de hibridación sobre la clonación directa de los grupos de genes *lux*. Barras de error, DE; n = 3. El vector p15A-cm y el ADN genómico de *P. phosphoreum* se digirieron con 0,02 U μl^{-1} de T4pol y luego se hibridaron en diferentes métodos (A, B, C) y finalmente se electroporaron en *E. coli* GB05-dir inducida por arabinosa.

10 La Figura 4 es un diagrama de flujo de la preparación del vector de clonación lineal y la técnica de clonación directa ExoCET. (a): El vector p15A-cm se preparó mediante amplificación por PCR y los cebadores empleados llevan una caja de homología de 80 nucleótidos. (b): La estrategia estándar para la clonación directa de ExoCET. El producto ensamblado *in vitro* se transformó en GB05-dir inducido por arabinosa que contenía pSC101-BAD-ETgA-tet, y se obtuvo el recombinante correcto mediante selección de antibióticos.

15 Figura 5 (a): La posición de las cajas de homología de 80 pb entre los fragmentos de ADN genómico p15A-cm y *lux* de 14 kb: (A) Ambas cajas de homología estaban muy al extremo; (B, C) uno estaba ubicado a 1 kb del extremo y el otro estaba muy al extremo; (D) Ambos estaban ubicados a 1 kb del extremo. Las condiciones de reacción fueron las mismas que en (1f). (b): El número de colonias se obtuvo mediante la combinación de ETgA, T4pol o ExoCET con las cuatro cajas de homología anteriores. (c): Utilizando ExoCET y cajas de homología terminal en GB2005, se obtuvo la eficiencia de clonación directa del grupo de genes *lux* de 14 kb mediante las diferentes combinaciones de proteínas recombinantes expresadas por el plásmido pSC101: ETg no expresa RecA; Eg, no expresa RecA y RecT expresos; Tg no expresa RecA y RecE; vector vacío pSC101-tet. Barras de error, DE; n = 3.

25 La Figura 6 muestra la clonación directa del grupo de genes de salinomicina de 106 kb utilizando el ADN genómico de *Streptomyces albus* digerido con EcoRV o Cas9. (a): Bajo la acción de ExoCET, el grupo de genes de salinomicina se clonó a partir de ADN genómico digerido con EcoRV o Cas9 gARN2/Cas9-gARN7 y luego se insertó en el vector pBeloBAC11. La caja de homología (azul) se insertó primero en el vector BAC y el vector BAC se linealizó mediante digestión con BamHI para que fuera un vector de clonación directa. La longitud de la caja de homología se marcó al final del fragmento de ADN genómico. (b): Análisis de restricción con PvuII de ADN recombinante, los clones correctos se marcaron con flechas.

35 La Figura 7 muestra una comparación de la eficiencia del ensamblaje de ExoCET y Gibson. (a): Un diagrama esquemático de la clonación directa de un fragmento de ADN de 45 kb que contiene el gen *Wnt4* del genoma del ratón usando las cajas de homología terminal. (b): Número de colonias obtenidas por los métodos ExoCET y Gibson. El ADN genómico de ratón digerido con p15A-cm y SwaI se trató con ExoCET y Gibson, respectivamente, y luego se transformó en el GB05 que contiene pSC101-BAD-ETgA-tet-dir, ya sea por inducción con arabinosa (ExoCET o Gibson + ETgA) o sin inducción (T4pol o Gibson). (c): Un diagrama esquemático del ensamblaje de múltiples fragmentos de ADN en plásmidos mediante T4pol, ExoCET, Gibson y Gibson + ETgA. El intervalo de tamaño del fragmento de ADN fue de 1,0 kb a 5,4 kb, el intervalo de tamaño del plásmido p5A ensamblado fue de 29,8 kb a 54,9 kb y los plásmidos fueron resistentes al cloranfenicol (cm). (d): El número de clones y la tasa correcta obtenida por el experimento de ensamblaje de múltiples fragmentos. El producto ensamblado *in vitro* se transformó en GB05-dir que contiene pSC101-BAD-ETgA-tet, ya sea por inducción con arabinosa (ExoCET o Gibson + ETgA) o sin inducción (T4pol o Gibson).

45 La Figura 8 muestra la construcción de vectores HIT (dirección isogénica de haplotipo) para DPY30 usando ADN genómico de mamífero. (a): diagrama esquemático de la clonación del codón de terminación DPY30 utilizando ADN genómico humano digerido con SpeI. Una vez completada la clonación directa, el extremo terminal C de DPY30 se marcó con el elemento mVenus^{16,17} de la recombinación de Red $\alpha\beta$. (b): Empleando el ADN genómico aislado de sangre humana como plantilla, el ADN recombinante clonado directamente por ExoCET se sometió a análisis de digestión con EcoRI. (c): Análisis de digestión con EcoRI del ADN recombinante obtenido por clonación directa a través de ExoCET y ADN genómico aislado de células 293T de riñón embrionario humano. (d): El análisis de digestión PvuII del ADN recombinante obtenido después de insertar el elemento mVenus mediante recombinación con Red $\alpha\beta$. Todos los clones obtenidos fueron correctos, siendo el carril 11 el control. El clon correcto estaba marcado con flechas.

55 La Figura 9 muestra la construcción de vectores HIT (dirección isogénica de haplotipo) para Dpy30 usando ADN genómico de ratón. (a): Un diagrama esquemático de la clonación del codón de terminación Dpy30 utilizando ADN genómico de ratón digerido con BamHI+KpnI. Una vez clonado directamente, el extremo terminal C de DPY30 se marcó con un casete mVenus^{16,17} usando recombinación de Red $\alpha\beta$. (b): Análisis de digestión con EcoRI para el ADN recombinante obtenido por clonación directa a través de ExoCET y ADN genómico aislado de células B16 de melanoma de ratón. (c): Análisis de digestión NheI del ADN recombinante obtenido insertando el elemento mVenus de la ingeniería recombinante Red $\alpha\beta$. Todos los clones obtenidos fueron correctos, siendo el carril 11 el control. El clon correcto estaba marcado con flechas.

65 La Figura 10 muestra el genotipado de células de mamíferos usando ExoCET. (a): Un diagrama esquemático de genotipado usando ExoCET. Los sitios de restricción se ubicaron secuencia arriba y secuencia abajo del elemento objetivo, respectivamente. (b): Genotipado de células madre embrionarias de ratón dirigidas por Kmt2d-AID-neo con

resistencia a kanamicina usando ExoCET. Se liberó del genoma un fragmento de ADN que contenía un elemento de direccionamiento utilizando Sspl y SspI. La clonación de ExoCET se realizó usando 10 µg de ADN genómico digerido por restricción y vector p15A-cm amplificado por PCR. Los fragmentos de dirección y los fragmentos de tipo silvestre clonados en el vector p15A pueden separarse mediante digestión con enzimas de restricción y doble raya.

La Figura 11 muestra el análisis de restricción con EcoRV + PstI de colonias resistentes a cloranfenicol obtenidas por genotipado de ExoCET de células madre embrionarias de ratón dirigidas a Klf4-Venus neo.

La ingeniería recombinante de ADN es una técnica de ingeniería genética para modificar moléculas de ADN en células de *E. coli*, que está mediada por recombinación homóloga de las proteínas sin/exo del fago (principalmente Redα y Redβ)¹⁸⁻²². La ingeniería recombinante de ADN se descubrió por primera vez en la cepa *sbcA* de *E. coli* (represor recBC), que tiene una actividad que media eficazmente la recombinación homóloga entre moléculas de ADN con cajas de homología²³. La cepa *sbcA* fue descubierta en un experimento clásico de AJ Clark en busca de una vía de recombinación homóloga en *E. coli*. Usó la cepa recBC, que es muy sensible al daño del ADN, para detectar su inhibidor, y encontró cepas mutantes de *sbcA* con actividades de expresión de RecE y RecT²⁴⁻²⁶. Estudios posteriores han demostrado que RecE y RecT se expresan mediante el fago Rac integrado en los cromosomas, que funcionan de manera idéntica al fago Redα y Redβ²⁷, y solo 280 aminoácidos en el extremo terminal C de la proteína RecE se expresan en la cepa mutante *sbcA*²⁸⁻³⁰. La RecE truncada es similar a Redα (266 aminoácidos) y es una exonucleasa³¹ de 5' a 3'. RecT es similar a Redβ y es una proteína de hibridación monocatenaria (SSAP)³². RecE/RecT y Redα/Redβ pertenecen al par de proteínas syn/exo exonucleasa 5' a 3' /SSAP^{21,33}, y es necesario un potencial de interacción proteína-proteína específico entre cada par de proteínas para la recombinación homóloga de ADN bicatenario^{29,34,35}. La recombinación homóloga mediada por Redα/Redβ ocurre principalmente en la horquilla de replicación y requiere replicación simultánea. Aunque los inventores descubrieron inicialmente técnicas de ingeniería recombinante a través de RecE/RecT truncada, los inventores utilizaron posteriormente Redα/Redβ para modificar moléculas de ADN porque estas últimas eran más eficientes^{16,38-42}. No obstante, los inventores están estudiando constantemente las características de RecE/RecT, y encontraron que los 600 residuos de aminoácidos en el extremo terminal N de RecE cambian su actividad de recombinación de dependiente de la replicación a independiente de la replicación³. Por lo tanto, dos moléculas de ADN lineales pueden formar un plásmido circular mediante recombinación homóloga eficiente a través de una caja homóloga muy corta. En comparación con la ingeniería recombinante Redα/Redβ, este mecanismo de recombinación lineal-lineal tiene diferentes aplicaciones, tal como la clonación directa de grandes fragmentos de ADN del genoma^{3-6,12} o la realización de ensamblajes de múltiples fragmentos de ADN^{15,43}.

La divulgación proporciona un método de recombinación homóloga (recombinación lineal-lineal) entre dos o más moléculas de ácido nucleico lineal diana que comparten al menos una secuencia homóloga. El método comprende la mezcla de moléculas de ácido nucleico lineal diana tratadas con una primera exonucleasa; luego, las moléculas de ácido nucleico lineal diana tratadas se someten a recombinación homóloga en presencia de una segunda exonucleasa y una proteína de hibridación. La segunda exonucleasa puede ser RecE, y la secuencia de aminoácidos de RecE de longitud completa de *E. coli* K12 se divulga en el documento WO2011/154927. O la segunda exonucleasa también puede ser RecE truncada, y las formas truncadas de RecE incluyen la proteína RecE que consta de los aminoácidos 588-866, 595-866, 602-866 o 606-866²⁹.

La recombinación homóloga está mediada por la segunda exonucleasa y la proteína de hibridación. En algunas realizaciones, la proteína de hibridación usada en los métodos de la divulgación es la proteína hibridada divulgada en el documento WO2011/154927. En particular, la proteína de hibridación es RecT o un fragmento de la misma (derivado del fago Rac). Más particularmente, la proteína de hibridación es la RecT de longitud completa y la segunda exonucleasa es la RecE de longitud completa. Sin embargo, se puede usar cualquier otra proteína de hibridación adecuada siempre que la proteína de hibridación interactúe con la exonucleasa usada. En el documento WO 02/062988 se proporcionan ejemplos de otras proteínas de hibridación adecuadas. La recombinación lineal-lineal puede ocurrir en las células huésped que carecen de expresión de RecT, tales como la cepa GB2005 de *E. coli*, posiblemente debido a la presencia de ciertas actividades endógenas similares a RecT. Sin embargo, la eficiencia de la recombinación lineal-lineal mediada por RecE de longitud completa aumenta significativamente en presencia de RecT.

Los métodos de la divulgación pueden verse afectados total o parcialmente en una célula huésped. Las células huésped adecuadas incluyen células de muchas especies, incluidos parásitos, procariotas y eucariotas, pero bacterias tales como bacterias Gram negativas son los huéspedes preferidos. Más particularmente, la célula huésped es una célula bacteriana entérica tal como células de *Salmonella*, *Klebsiella*, *Bacillus*, *Neisseria*, *Photobacterium* o *Escherichia Coli* (el método de la divulgación juega un papel eficaz en todas las cepas de *E. coli* que se han probado). Una célula huésped preferida es *E. coli* K12. Sin embargo, cabe señalar que los métodos de la divulgación son igualmente aplicables a células u organismos eucariotas, tales como células de hongos, levaduras, plantas o animales. Se ha demostrado que este sistema es funcional en las células madre embrionarias de ratón y es razonable especular que también es funcional en otras células eucariotas. La célula huésped es típicamente una célula huésped aislada, pero pueden ser células huésped no aisladas.

La célula huésped de la divulgación comprende un ácido nucleico que codifica una exonucleasa (particularmente RecE de longitud completa), una proteína de hibridación (particularmente RecT) y Redγ. En algunas realizaciones, la célula

huésped comprende además un ácido nucleico que codifica RecA. Particularmente, la célula huésped expresa RecE, RecT y Redy, y opcionalmente RecA. Más particularmente, la célula huésped expresa RecE, RecT, Redy y RecA.

La exonucleasa, la proteína de hibridación, Redy y/o RecA de la divulgación pueden ser un producto de expresión recombinante de un ADN extraño en una célula huésped, por ejemplo, expresado por un vector transformado en una célula huésped. Un ejemplo de un vector adecuado es el plásmido pSC101, aunque también se pueden usar otros vectores adecuados. Puede usarse cualquier promotor adecuado para impulsar la expresión de estas proteínas. Sin embargo, en el caso de expresar RecE, se prefiere un promotor inducible tal como un promotor inducible por arabinosa (P_{BAD}) o un promotor inducible por ramnosa (P_{RhaSR}). Estos promotores son bien conocidos en la técnica.

La célula huésped de la divulgación expresa una exonucleasa, una proteína de hibridación, Redy y/o RecA por los promotores inducibles en un vector plásmido o un cromosoma. En particular, la exonucleasa, la proteína de hibridación, Redy y/o RecA se expresan en la célula huésped mediante un vector plásmido. Más particularmente, la exonucleasa, la proteína de hibridación, Redy y/o RecA se expresan simultáneamente en la célula huésped por el vector plásmido y el cromosoma.

El genoma de la célula huésped de *E. coli* K12 consiste en una copia endógena del gen *recE* de longitud completa y el gen *recT*, que están presentes en el fago Rac que se ha integrado en el genoma del huésped. Sin embargo, dado que el gen es silencioso, la expresión de RecE de longitud completa no puede ocurrir naturalmente a partir del gen integrado. Por lo tanto, en realizaciones en las que la exonucleasa 5' a 3' se expresa mediante ADN exógeno, el método se puede llevar a cabo en ausencia de un gen *recE* endógeno.

También se proporcionan células huésped transformadas con la molécula de ácido nucleico codificante de la exonucleasa anterior. En particular, la exonucleasa se expresa mediante la molécula de ácido nucleico y, por lo tanto, la divulgación también proporciona la célula huésped que expresa la exonucleasa enumerada en el método de la divulgación. La exonucleasa se expresa particularmente bajo el control de un promotor inducible, tal como un promotor inducible por arabinosa (P_{BAD}) o un promotor inducible por ramnosa (P_{RhaSR}).

En las realizaciones anteriores, los métodos de la divulgación pueden verse afectados total o parcialmente *in vitro*. Por ejemplo, se puede usar exonucleasa 5' a 3' purificada y proteína de hibridación (particularmente proteínas RecE y RecT purificadas), o se puede usar un extracto de células de *E. coli* que expresan la exonucleasa 5' a 3' y la proteína de hibridación. Cuando el método se lleva a cabo *in vitro*, es ventajoso pretratar la primera y la segunda moléculas de ácido nucleico diana lineales para exponer extremos homólogos monocatenarios.

La recombinación lineal-lineal requiere que se comparta al menos una secuencia homóloga entre las moléculas de ácido nucleico lineal diana en las que se produce la recombinación homóloga. En algunas realizaciones, la primera molécula de ácido nucleico diana comparte una secuencia homóloga con la segunda molécula de ácido nucleico diana para realizar la recombinación lineal-lineal entre la primera y la segunda moléculas de ácido nucleico diana, para producir un producto lineal. En las realizaciones en las que se produce una recombinación lineal-lineal entre el primer y el segundo ácidos nucleicos lineales y uno o más ácidos nucleicos lineales adicionales para formar un producto lineal, cada ácido nucleico lineal comparte una secuencia homóloga con el ácido nucleico lineal que forma su vecino en los productos lineales de la reacción de recombinación lineal-lineal. En realizaciones en las que se produce una recombinación lineal entre el primer y segundo ácidos nucleicos lineales y una o más moléculas de ácido nucleico diana lineales adicionales, para formar un producto cíclico, cada ácido nucleico lineal comparte una secuencia homóloga con el ácido nucleico lineal que forma su vecino en el producto cíclico de la reacción de recombinación lineal-lineal. En algunas realizaciones, la primera molécula de ácido nucleico diana y la segunda molécula de ácido nucleico diana comparten dos secuencias homólogas para realizar una recombinación lineal-lineal entre la primera y la segunda moléculas de ácido nucleico diana, para formar una molécula cíclica. Los expertos en la técnica saben cómo diseñar secuencias homólogas para formar moléculas lineales o cíclicas.

En particular, al menos una caja de homología está al final de cada fragmento lineal. Cuando las cajas de homología están al final de cada fragmento lineal y las diferentes cajas de homología están en el otro extremo, estas secuencias homólogas o 'cajas homólogas' producen la configuración óptima, y la construcción de estas cajas de homología permite la recombinación para generar un anillo. La recombinación lineal puede ocurrir cuando la caja de homología no está al final, pero la eficiencia se reduce. Por lo tanto, en una realización preferida, al menos una secuencia homóloga se localiza en el extremo más externo de uno o ambos extremos de la molécula de ácido nucleico diana. En algunas realizaciones, al menos una secuencia homóloga es interna a la determinada molécula de ácido nucleico diana.

Las secuencias homólogas de la divulgación son de al menos 4, al menos 6, al menos 10, al menos 20, al menos 30, al menos 40, al menos 50, al menos 60, al menos 70, al menos 80, al menos 90, al menos 100 nucleótidos de longitud. Por ejemplo, en algunas realizaciones, las secuencias homólogas son de 4-6, 6-9, 6-30, 6-100, 10-20, 20-29, 20-40, 20-50, 10-100, 25-30, 25-40, 25-50, 30-40, 30-50, 40-50, 40-80 o más de 80 nucleótidos. La eficacia de la recombinación homóloga aumenta con la longitud de las cajas de homología utilizadas, por lo que se pueden utilizar cajas de homología más largas.

"Homólogo" entre dos moléculas de ácido nucleico significa que cuando las secuencias de dos moléculas de ácido nucleico están alineadas, hay muchos residuos de nucleótidos que son idénticos en la misma posición en la secuencia. El grado de homología es fácil de calcular.

(Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Parte 1, Griffin, A.M., y Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; y Sequence Analysis Primer, Gribskov, M. y Devereux, J., eds., M Stockton Press, New York, 1991).

En algunas realizaciones, los métodos de la divulgación comprenden unir una pluralidad de moléculas de ácido nucleico lineales para formar una molécula de ácido nucleico circular, tal como un plásmido circular. Cada molécula de ácido nucleico diana comparte al menos una secuencia homóloga con una molécula de ácido nucleico diana que forma su vecina en el producto cíclico resultante y se somete a recombinación lineal-lineal de acuerdo con los métodos de la divulgación. El número de moléculas de ácido nucleico diana es 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más.

En algunas realizaciones, al menos una de las moléculas de ácido nucleico lineal diana comprende un marcador de selección para permitir la selección del recombinante correcto. En la divulgación se puede utilizar cualquier marcador de selección adecuado. En algunas realizaciones, el marcador de selección es un gen de resistencia a antibióticos, por ejemplo, un gen de resistencia al cloranfenicol, ampicilina, kanamicina o blasticidina.

La molécula de ácido nucleico lineal diana puede derivarse de cualquier fuente adecuada. Por ejemplo, se pueden incluir secuencias de ácido nucleico de eucariotas o procariotas. En algunas realizaciones, la primera molécula de ácido nucleico lineal diana es ADN genómico. Normalmente, el ADN genómico es un fragmento de ADN genómico. El ADN genómico consta particularmente de una secuencia diana. En algunas realizaciones, se puede obtener un fragmento de ADN genómico mediante escisión o digestión de ADN genómico (por ejemplo, usando una enzima de restricción) para obtener una secuencia diana completa que lo contiene. En algunas realizaciones, la primera molécula de ácido nucleico lineal diana (tal como un fragmento de ADN genómico, un miembro de la biblioteca de ADNc o un fragmento derivado de BAC) comprende una secuencia diana de 2 kb o más (por ejemplo, 2,5 kb o más, 4-kb o más, 5-kb o más, 7,5-kb o más, 10 kb o más, 15 kb o más, 20 kb o más, 25 kb o más, 40 kb o más, 50 kb o más, 75 kb o más o 100 kb o más). En particular, la secuencia diana es la región completa entre las cajas de homología en cada extremo de la primera molécula de ácido nucleico lineal diana. Por ejemplo, un grupo de genes que codifica una vía de metabolitos secundarios o una vía de síntesis de ácidos grasos. En algunas realizaciones, los métodos de la divulgación se pueden usar para clonar directamente una región de ADN de un genoma animal humano o no humano. Por ejemplo, terapias regenerativas para la investigación de la salud o para la corrección por direccionamiento génico. Por ejemplo, en algunas realizaciones, la primera molécula de ácido nucleico diana comprende o consiste en un fragmento de ADN genómico de un animal humano o no humano. El fragmento de ADN genómico puede comprender una secuencia diana, tal como un gen que comprende una mutación, en la que mutación da como resultado una enfermedad o afección y la modificación de la mutación a una secuencia de tipo silvestre puede tratar o prevenir la enfermedad o afección. En realizaciones en las que la primera molécula de ácido nucleico diana es un fragmento de ADN genómico, la segunda molécula de ácido nucleico diana es particularmente un vector de clonación lineal.

En realizaciones en las que la primera molécula de ácido nucleico diana es un fragmento de ADN genómico, el método comprende generar una primera molécula de ácido nucleico diana digiriendo o escindiendo el ADN genómico para obtener un fragmento de ADN genómico lineal que comprende la secuencia diana, luego, la primera exonucleasa se usa para tratar la mezcla del fragmento de ADN genómico y el vector de clonación lineal, procesando las etapas de escisión de la molécula de ácido nucleico diana e hibridación para ligar la molécula de ácido nucleico diana, y luego se transfiere la mezcla de las moléculas de ácido nucleico tratadas a las células huésped. La segunda molécula de ácido nucleico diana comprende particularmente un marcador de selección.

En una realización, los métodos de la divulgación comprenden la etapa de ligación de moléculas de ADN *in vitro*.

El proceso de ligación *in vitro* comprende digestión con exonucleasa seguida de hibridación.

La exonucleasa es la polimerasa T4.

El proceso de ligación *in vitro* comprende el ensamblaje de Gibson.

El proceso de ligación *in vitro* comprende la síntesis de ADN mediante la ADN polimerasa con o sin exonucleasa seguida de hibridación.

El proceso de ligación *in vitro* comprende la hibridación mediante una proteína de hibridación monocatenaria, tal como RecA/RAD51, Red β , RecT, Plu β o RAD52.

Las células huésped utilizadas para la recombinación homóloga son células de *E. coli*.

Las células huésped para recombinación homóloga son células de *E. coli* que expresan RecE y/o RecT de longitud completa.

5 Las células huésped para recombinación homóloga son células de *E. coli* que expresan RecE, RecT y/o RedY de longitud completa.

Las células huésped para recombinación homóloga son células de *E. coli* que expresan RecE, RecT y/o RedY truncados.

10 Las células huésped para recombinación homóloga son cualquier célula huésped bacteriana que exprese RecE y/o RecT de longitud completa.

Las células huésped para recombinación homóloga son células de *E. coli* que expresan Red α , Red β y/o RedY.

15 La célula huésped para recombinación homóloga es una célula de *Saccharomyces cerevisiae*.

Se proporcionan kits para usar en la divulgación. En algunas realizaciones, los kits comprenden un ácido nucleico que codifica una exonucleasa como se describe en el presente documento. En algunas realizaciones, el kit comprende una exonucleasa como se describe en el presente documento. En particular, la primera exonucleasa es la ADN polimerasa T4 (T4pol), el fragmento Klenow de la ADN polimerasa I (Kle), la ADN polimerasa T7 (T7pol), la exonucleasa III (ExoIII), la ADN polimerasa Phusion (Phu), la exonucleasa T5 (T5exo), exonucleasa T7 (T7exo) y la exonucleasa lambda (λ exo), y la segunda exonucleasa es RecE de longitud completa. Más particularmente, los kits comprenden una célula huésped como se describe en el presente documento. Por ejemplo, en algunas realizaciones, las células huésped en un kit comprenden un ácido nucleico que codifica RecE, RecT, RedY y RecA de longitud completa descritos en el presente documento bajo el control de un promotor inducible. Los kits también pueden incluir uno o más vectores de clonación lineal preparados previamente.

Otra aplicación preferida de la divulgación se refiere al ensamblaje de moléculas de ácido nucleico lineales en biología de síntesis, particularmente ADN lineal. Por lo tanto, en algunas realizaciones, la primera y la segunda moléculas de ácido nucleico diana son lineales, y el método comprende además poner en contacto la primera y la segunda moléculas de ácido nucleico diana con una o más de otras moléculas de ácido nucleico diana lineales (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, al menos 10, al menos 20 de otras moléculas de ácido nucleico diana) en presencia de una exonucleasa 5' a 3' y una proteína de hibridación, para producir un producto lineal o circular. En una realización preferida, la recombinación homóloga entre la primera y la segunda moléculas de ácido nucleico diana y una o más de otras moléculas de ácido nucleico diana da como resultado la producción de genes, operones, cromosomas o genomas completos. El ensamblaje biológico de síntesis de ácidos nucleicos de ADN se ha utilizado para generar genes, operones, cromosomas, o se ha utilizado recientemente para generar genomas completos. En una realización de la divulgación, la combinación de la primera exonucleasa y la segunda exonucleasa aumenta significativamente la eficiencia de ensamblaje de la molécula de ácido nucleico lineal, la divulgación será un método preferido para el ensamblaje de ADN biológico de sintético en el comercio y la investigación.

Otra aplicación preferida de la divulgación es construir un vector de direccionamiento de haplotipo isogénico, que puede clonar directamente un fragmento de ADN de 5 a 10 kb del genoma de mamífero usando el método de la divulgación como una caja de homología isogénica, y estos fragmentos de ADN son no solo los genes idénticos sino que también mantiene un haplotipo polimórfico, por lo que los inventores lo denominan vector de direccionamiento de haplotipo isogénico, es decir, el llamado vector de direccionamiento de haplotipo isogénico (HIT). El marcador de selección y otros elementos funcionales se insertan luego en el vector de HIT mediante ingeniería recombinante para obtener un vector para el direccionamiento. Otra aplicación preferida de la divulgación es la genotipificación de células de mamífero. El fragmento de ADN que contiene el elemento de direccionamiento completo se clona a partir del genoma de las posibles células madre embrionarias diana mediante el método de la divulgación, y el plásmido recombinante obtenido mediante la clonación se somete a análisis de restricción y secuenciación de ADN, y la célula se determina con éxito de acuerdo con el resultado.

Ejemplos

55 Materiales y método

Cepa y plásmido

60 La *E. coli* GB2005 se derivó de DH10B eliminando *fhuA*, *ybcC* y *recET*^{3,16,44}. GB05-dir se derivó de GB2005 integrando el operón P_{BAD}-ETgA (*recE*, *recT*, *redY* y *recA* de longitud completa bajo el promotor P_{BAD} inducible por arabinosa)³ en el locus *ybcC*. GB08-red se derivó de GB2005 integrando el operón P_{BAD}-gbaA (*redY*, *red β* , *red α* y *recA* bajo el promotor P_{BAD} inducible por arabinosa)¹⁶ en el locus *ybcC*. pSC101-BAD-ETgA-tet³ transmite resistencia a la tetraciclina y porta el operón ETgA de longitud completa P_{BAD} y un origen de replicación pSC101 sensible a la temperatura que se replica a 30 °C pero no a 37 °C, por lo que se puede eliminar fácilmente del huésped por cambio de temperatura en ausencia de selección.

Preparación y digestión de ADN genómico.

Se cultivaron *Photobacterium phosphoreum* ANT-2200 y *Photorhabdus luminescens* DSM15139 Gramnegativos durante la noche en 50 ml de medio. Después de la centrifugación, las células se resuspendieron completamente en 8 ml de Tris-Cl 10 mM (pH 8,0). Se añadieron quinientos microlitros de 20 mg ml⁻¹ de proteinasa K y 1 ml de SDS al 10% y se incubaron a 50 °C durante 2 h hasta que la solución se volvió transparente. El ADN genómico se recuperó del lisado mediante extracción con fenol-cloroformo-alcohol isoamílico (25:24:1, pH 8,0) y precipitación con etanol. El ADN se disolvió en Tris-Cl 10 mM (pH 8,0) y se digirió con BamHI + KpnI para la clonación del grupo de genes *lux* de 14 kb.

Se cultivó *Streptomyces albus* DSM41398 Grampositiva en 50 ml de caldo de soja tríptico a 30 °C durante 2 días. El ADN genómico se aisló de acuerdo con el método descrito en la referencia (10) con una ligera modificación. Después de la centrifugación, las células se resuspendieron completamente en 8 ml de tampón SET (NaCl 75 mM, EDTA 25 mM, Tris 20 mM, pH 8,0) y se añadieron 10 mg de lisozima. Después de la incubación a 37 °C durante 1 h, se añadieron 500 µl de 20 mg ml⁻¹ de proteinasa K y 1 ml de SDS al 10% y se incubaron a 50 °C durante 2 h hasta que la solución se volvió transparente. Se añadieron al lisado tres mililitros y medio de NaCl 5 M. El ADN genómico se recuperó del lisado mediante extracción con fenol-cloroformo-alcohol isoamílico (25:24:1, pH 8,0) y precipitación con etanol. El ADN se disolvió en Tris-Cl 10 mM (pH 8,0).

Se purificó ADN genómico a partir de células B16 de melanoma de ratón, células 293T de riñón embrionario humano y sangre humana utilizando los kits de ADN de cultivo celular y sangre de Qiagen de acuerdo con las instrucciones del fabricante, excepto que se recuperó el ADN del lisado tratado con proteinasa K mediante extracción con fenol-cloroformo-alcohol isoamílico (25:24:1, pH 8,0) y precipitación con etanol. El ADN se disolvió en Tris-Cl 10 mM (pH 8,0). El ADN genómico digerido por restricción se extrajo con fenol-cloroformo-alcohol isoamílico (25:24:1, pH 8,0) y se precipitó con etanol. El ADN se disolvió en Tris-Cl 10 mM (pH 8,0). Se utilizaron puntas de pipeta cortadas en los extremos para evitar el corte del ADN genómico.

El ADN genómico de *P. luminescens* DSM15139 se digirió con XbaI para la clonación *plu3535-plu3532* y XbaI + XmaI para la clonación de *plu2670*. El ADN genómico de *S. albus* se digirió con EcoRV o complejos de Cas9-gARN para la clonación del grupo de genes de salinomicina. El ADN genómico de ratón se digirió con HpaI para la clonación de *Prkar1a*, BamHI + KpnI para la clonación de *Dpy30* y Swal para la clonación de *Wnt4* o *Lmbr11-Tuba1a*. El ADN genómico humano se digirió con SpeI para la clonación de *DPY30*, NdeI + BstZ17I para la clonación de *IGFLR1-LIN37*, BstZ17I para la clonación de *IGFLR1-ARHGAP33* y NdeI para la clonación de *ZBTB32-LIN37*. El ADN genómico digerido se extrajo con fenol-cloroformo-alcohol isoamílico (25:24:1, pH 8,0) y se precipitó con etanol. El ADN se disolvió en ddH₂O y se concentró hasta 1 µg µl⁻¹. Se utilizaron puntas de pipeta cortadas en los extremos para evitar el corte del ADN genómico. Se utilizaron diez microgramos de ADN genómico digerido para la clonación de ExoCET.

Digestión con Cas9 del ADN genómico de *S. albus*

La proteína Cas9 de *S. pyogenes* se adquirió a través de New England Biolab. La digestión con Cas9 del ADN genómico de *S. albus* se llevó a cabo en un sistema de reacción de 800 µl que contenía 80 µl de tampón de reacción Cas9 ×10 (NEB), 80 µg de ADN genómico, 40 µg de gARN-2, 40 µg de gARN-7 y 20 µg de Cas9. Dado que la eficiencia de escisión de Cas9 se vio severamente afectada por la pureza del sustrato de ADN, en este experimento, el ADN genómico de *S. albus* necesitó ser extraído tres veces con fenol-cloroformo-alcohol isoamílico (25:24:1, pH 8,0) para asegurar la eficiencia de la escisión de Cas9. Después de que el producto resultante se incubó a 37 °C durante 6 h, se agregaron 100 µg de RNasa A (Thermo Scientific) y después de la incubación a 37 °C durante 1 h, se agregaron 100 µg de proteinasa K (Roche) y se continuó la incubación a 50 °C durante 1 h. Luego, el ADN genómico se extrajo una vez con fenol-cloroformo-alcohol isoamílico (25: 24:1, pH 8,0), y después de la precipitación con etanol, se disolvió en una cantidad apropiada de ddH₂O hasta una concentración final de aproximadamente 1 µg µl⁻¹. Finalmente, se utilizaron 10 µg de ADN genómico escindido por la proteína Cas9 para el experimento de clonación del método de la divulgación.

Preparación del vector de clonación lineal

Usando el plásmido p15A-Pamp-luxABECD (Genebridge) como plantilla, el vector p15A-cm se amplificó por PCR con ADN Polimerasa PrimeSTAR Max (Takara), los cebadores usados (Tabla 1) constan de 80 cajas de homología de nucleótidos y fueron purificados por PAGE. El producto de PCR eliminó la interferencia de los cebadores en experimentos posteriores mediante recuperación en gel. El kit utilizado fue el kit de extracción en gel QIAquick (Qiagen). Finalmente, el ADN se eluyó con ddH₂O a una concentración de aproximadamente 200 ng/µl, y se usaron 200 ng para el experimento de clonación ExoCET.

El vector pBeloBAC11 usado para clonar el grupo de genes de salinomicina y el vector pBAC2015 usado para clonar *plu3535-3532* se construyeron de acuerdo con las referencias ^{43,46}. Los vectores BAC se linealizaron con BamHI para exponer ambos brazos de homología y se extrajeron con fenol-cloroformo-alcohol isoamílico (25:24:1, pH 8,0) y se precipitaron con isopropanol. El ADN se disolvió en ddH₂O y se concentró a 1 µg/µl. Se utilizó un microgramo de vectores BAC lineales para la clonación de ExoCET.

Tabla 1. Oligonucleótidos amplificados por el vector lineal p15A-cm

Gen	Cebador	Secuencias (5'-3')
<i>plu2670</i>	plu2670-1	ctgggaaaactacggtgagcatatgtttaccacttctcagagaaccgtgaatattgcg ttaaaccgatgaactgcccggtTACGCCCCGCCCTGCCACTC (SEQ. ID. NO. 1)
	plu2670-2	aaatccatgattaccaaccactcccaaagatattataaaatctcaacatgctgattg atcctttggaatgggctaaatctagaGGTAACGAATCAGACAATTGAC G (SEQ. ID. NO. 2)
<i>Prkar1a</i> de ratón	mPrkar1a-1	agatgacgcccctcactctgcataccaattccacataaccactctcaaattataat gctcacagaccctctaagggtTACGCCCCGCCCTGCCACTC (SEQ. ID. NO. 3)
	mPrkar1a-2	catagtaccattgttatagagtacttagctatgcctcagagtaatgctcttgatctagg tgctttcttttactgttGGTAACGAATCAGACAATTGACG (SEQ. ID. NO. 4)
<i>Dpy30</i> de ratón	mDpy30-1	ggattttataagtgaagggtgacattgtagtattccttaccctgtctggttctcatacgc attaactcataatccttggtaccTACGCCCCGCCCTGCCACTC (SEQ. ID. NO. 5)
	mDpy30-2	tgccctttcagaagacctagggttgattcctagcaccactgatgctaacaaccagac ataatgccagttctaggggatccGGTAACGAATCAGACAATTGACG (SEQ. ID. NO. 6)
<i>Wnt4</i> de ratón	mWnt4-1	gtttaactcccttaacacacacacacacatgcacacgcacatgcacacgcgcac tcacaccacaaaatcacacaatttTACGCCCCGCCCTGCCACTC (SEQ. ID. NO. 7)
	mWnt4-2	ctgtgctcagagctttgagtgcctacatataacaggacacaaactcttcttactttctg attgccacatggtccatttGGTAACGAATCAGACAATTGACG (SEQ. ID. NO. 8)
<i>Lmbr1l-Tuba1a</i> de ratón	mLT-1	ctatgccataggaatttgaggagaaattaggcttactggtccattaattattccaacgga tgcagtcagaatgtatattTACGCCCCGCCCTGCCACTC (SEQ. ID. NO. 9)
	mLT-2	ttagagcagacagtgctcttaaccgttgagccatctctccagtccttgacacactgat tttaagttgacattaaaatttGGTAACGAATCAGACAATTGACG (SEQ. ID. NO. 10)
<i>DPY30</i> humano	hDpy30-1	ccaccacgccaggctaatttttagcttaataaccctattacctctcagggtgtgtctaaatttt ctttgatggaaaaactagtTACGCCCCGCCCTGCCACTC (SEQ. ID. NO. 11)
	hDpy30-2	agttatgtcagtttgctattataatgtttattatttcttgaataaaaaagaatatgcacct aataaacgttgactagtGGTAACGAATCAGACAATTGACG (SEQ. ID. NO. 12)
<i>IGFLR1 -LIN37</i> humano	hIL-1	Agtggccctagaggggtcggttaaggagccaagggcctatgaaggatgctcacac acctcccatcccacttcccagacacatatgTACGCCCCGCCCTGCC ACTC (SEQ. ID. NO. 13)
	hIL-2	

Gen	Cebador	Secuencias (5'-3')
		gtgctgtccaagcttggtcatctggggttgctgggcttaacacccaataaagaacttt gctgactactaagcccagtaGGTAACGAATCAGACAATTGACG (SEQ. ID. NO. 14)
<i>IGFLR1 ARHGA</i> <i>P33 humano</i>	hIA-1	acggccggcgactgctcctcagtgaggaggcgctcactcaatatccctgcagtggcg gccgcccattgatcaaacggtaTTACGCCCCGCCCTGCCACTC (SEQ. ID. NO. 15)
	hIA-2	gtgctgtccaagcttggtcatctggggttgctgggcttaacacccaataaagaacttt gctgactactaagcccagtaGGTAACGAATCAGACAATTGACG (SEQ. ID. NO. 16)
<i>ZBTB32 -LIN37</i> <i>humano</i>	hZL-1	cctagagggtcgtaaggagccaaggcgctatgaaggatgctcacacacctccc atccccacctccagacacatatgTTACGCCCCGCCCTGCCACTC (SEQ. ID. NO. 17)
	hZL-2	aagaactagactagaagttggaaaacagggttcactgcatcccttcacagtgaatg cccttctctcagcactcatatgGGTAACGAATCAGACAATTGACG (SEQ. ID. NO. 18)

Las letras minúsculas en la Tabla 1 representan la secuencia de caja homóloga.

Preparación de productos de PCR para ensamblaje de múltiples fragmentos

5 Los fragmentos de ADN con homologías de 40 pb usados en el experimento de ensamblaje de múltiples piezas se amplificaron por PCR usando ADN genómico de *P. luminescens* DSM15139 como plantilla y ADN polimerasa PrimeSTAR Max (Takara) de acuerdo con las instrucciones del fabricante. Los productos de PCR se extrajeron de geles de agarosa después de la electroforesis y se purificaron usando el kit de extracción en gel QIAquick (Qiagen) de acuerdo con las instrucciones del fabricante, excepto que el ADN se eluyó de la columna con ddH₂O y se concentró hasta 200 ng µl⁻¹. Se usaron 250 ng de cada fragmento para el ensamblaje de ADN.

Preparación del casete de ADN mVenus-PGK-neo

15 El casete mVenus-PGK-neo se amplificó a partir de pR6K-2Ty1-2PreS-mVenus-Biotin-PGK-em7-neo¹⁷ con PCR usando la prueba de lectura d ADN polimerasa PrimeSTAR Max (Takara) de acuerdo con las instrucciones del fabricante. Los cebadores se enumeran en la Tabla 1. Los productos de PCR se purificaron con el kit de purificación de PCR QIAquick (Qiagen) de acuerdo con las instrucciones del fabricante, excepto que el ADN se eluyó de la columna con ddH₂O y se concentró hasta 100 ng µl⁻¹. Se utilizaron doscientos nanogramos del casete para la ingeniería recombinante.

Montaje *in vitro*

25 Se ensamblaron diez microgramos de ADN genómico y 200 ng de vector lineal p15A-cm de 2,2 kb (1 µg de vector BAC lineal de 8 kb) en reacciones de 20 µl que consisten en 2 µl de tampón NEB 2.1 y 0,13 µl de 3 U µl⁻¹ de T4pol (NEB, catálogo No. M0203). Las reacciones de ensamblaje se prepararon en tubos de PCR de 0,2 ml y se ciclaron en un termociclador de la siguiente manera: 25 °C durante 1 h, 75 °C durante 20 min, 50 °C durante 30 min, luego se mantuvo a 4 °C. Para el ensamblaje de múltiples piezas, se agregaron 250 ng de cada fragmento y se utilizó un tiempo de masticación de 20 min. Los productos de ensamblaje *in vitro* se desalaron a temperatura ambiente durante 30 minutos mediante diálisis por gota contra ddH₂O utilizando filtros de membrana Millipore (Merck-Millipore, catálogo No. VSWP01300) antes de la electroporación. Todos los experimentos se realizaron por triplicado.

35 Las reacciones de ensamblaje con otras exonucleasas se ciclaron como sigue: T5exo: 50 °C durante 30 min, luego se mantuvo a 4 °C; T7exo: 25 °C durante 20 min, 50 °C durante 30 min, luego se mantuvo a 4 °C; Kle, T7pol y lexo: 25 °C durante 20 min, 75 °C durante 20 min, 50 °C durante 30 min, luego se mantuvo a 4 °C; ExoIII: 37 °C durante 20 min, 75 °C durante 20 min, 50 °C durante 30 min, luego se mantuvo a 4 °C; Phu: 37 °C durante 20 min, 50 °C durante 30 min, luego se mantuvo a 4 °C. El ensamblaje de Gibson se realizó a 50 °C durante 30 min con mezcla maestra de ensamblaje de Gibson (NEB, cat. E2611), luego se mantuvo a 4 °C.

Preparación de células de *E. coli* electrocompetentes

Se cultivó *E. coli* GB05-dir que contenía el plásmido pSC101-BAD-ETgA-tet (DO600 = 3-4) durante la noche a 30 °C en LB suplementado con 4 µg/ml de tetraciclina. Se transfirieron 40 µl de cultivo durante la noche (DO600 = 3-4) a 1,4 ml de LB suplementado con los antibióticos apropiados, luego la mezcla se colocó en un termomezclador Eppendorf a 30 °C y se incubó a 950 rpm durante 2 h (DO600 = 0,35-0,4). Se añadieron 35 µl de L-arabinosa al 10% (p/v, en ddH₂O) para inducir la expresión de la enzima recombinante (ETgA o gbaA) y se continuó la incubación durante 40 min a 37 °C (DO600 = 0,7 a 0,8). Las células se recogieron por centrifugación a 9.400 g durante 30 segundos a 2 °C.

El sobrenadante se descartó y el sedimento celular se suspendió en 1 ml de ddH₂O enfriado con hielo. Las células se recogieron por centrifugación a 9.400 g durante 30 segundos a 2 °C. El sobrenadante se descartó y el sedimento celular se suspendió en 1 ml de ddH₂O enfriado con hielo. Las células se centrifugaron repetidamente, se resuspendieron y se centrifugaron de nuevo, y las células se suspendieron con 20 µl de ddH₂O enfriado con hielo. Luego se agregaron 5 µl de producto ensamblado *in vitro* desalado, mientras que en el experimento de inserción del elemento mVenus-PGK-neo, se agregó la mezcla de 200 ng de plásmido y 200 ng de producto de PCR. La mezcla de células y ADN se transfirió a cubetas de 1 mm y se sometió a electroporación con un electroporador Eppendorf 2510 a un voltaje de 1350 V, una capacitancia de 10 µF y una resistencia de 600 Ω. Se añadió 1 ml de LB a la taza de descarga eléctrica, se lavaron las células y se transfirieron a un tubo de 1,5 ml con orificios, y luego se colocaron en el termomezclador Eppendorf a 950 rpm durante 1 h a 37 °C. Finalmente, se extendió una cantidad apropiada de la solución bacteriana a una placa LB suplementada con un antibiótico adecuado (15 µg/ml de cloranfenicol o 15 µg/ml de kanamicina) y se incubó a 37 °C durante la noche.

Ejemplo 1

La acción concertada del ensamblaje *in vitro* y RecE/RecT de longitud completa mejora la eficacia de la clonación directa.

Los inventores probaron una serie de exonucleasas y métodos de hibridación mediante clonación directa de un grupo de genes *lux* de 14 kb de *Photobacterium phosphoreum* ANT-2200⁴⁷ (Figura 1A). En este experimento de clonación, se mezclan 10 µg de ADN genómico de ANT-2200 digerido usando BamHI y KpnI con 200 ng de vector lineal p15A-cm de 2,2 kb. El vector lineal tiene una secuencia en ambos extremos que es idéntica a los extremos (caja homóloga) del fragmento genómico que contiene el grupo de genes *lux* producido por digestión de restricción. El vector lineal se recombina de forma homóloga con el fragmento de ADN diana a través de las cajas homólogas en ambos extremos, para formar un plásmido circular final. El vector de clonación y el ADN genómico digerido con BamHI + KpnI se trataron primero *in vitro* con diferentes enzimas exonucleasas, y luego el producto de reacción después del tratamiento *in vitro* se transformó en *E. coli* que expresaba la recombinasa RecE/RecT. Las exonucleasas ensayadas incluyen: ADN polimerasa T4 (T4pol), fragmento Klenow de ADN polimerasa I (Kle), ADN polimerasa T7 (T7pol), exonucleasa III (ExoIII); ADN polimerasa Phusion (ADN polimerasa Phusion; Phu), exonucleasa T5 (T5exo); Exonucleasa T7 (T7exo) y exonucleasa lambda (lexo) (Figura 2A). Los resultados mostraron que T4pol, Kle, T5exo y T7exo mejoran significativamente la eficacia de la clonación directa (Figura 2A-E). La velocidad de hibridación después de la digestión con exonucleasa no tiene ningún efecto sobre la eficacia de la clonación, por lo que los inventores eligen una velocidad de enfriamiento predeterminada (2 °C s⁻¹) en la máquina de PCR Eppendorf MC nexus (Figura 3). Se ensayó el efecto de la concentración, la temperatura de reacción y el tiempo de reacción de T4pol sobre la eficacia de la clonación directa (Figura 1C, d y Figura 2F). Se comparó la eficacia de la clonación directa usando RecE/RecT de longitud completa sola, hibridación *in vitro* de T4pol sola y combinando la hibridación *in vitro* de T4pol con RecE/RecT de longitud completa (Figura 1F). Cuando se usa RecE/RecT solo, el grupo de genes *lux* de 14 kb se clona directamente del cromosoma de la bacteria luminiscente al vector p15A con una alta precisión (427), pero la clonación directa ensambla los productos de reacción después del ensamblaje *in vitro* de T4pol en una *E. coli* de ingeniería estándar, es mucho más eficiente (427 vs 4.880). Los resultados mostraron que: (1) el sistema de reparación de ADN endógeno de *E. coli* puede cerrar hábilmente la cadena principal del plásmido; (2) T4pol es altamente eficaz en ensamblaje *in vitro* (4.880); (3) la técnica ExoCET que combina el ensamblaje *in vitro* de T4pol con la recombinación intracelular RecE/RecT es más eficiente que cualquiera de las dos técnicas por sí sola (32.500).

Ejemplo 2

Efecto de las cajas de homología sobre la eficiencia de la clonación

Cuanto más larga sea la caja de homología, mayor será la eficacia de clonación (Figura 1 B). Los inventores colocaron la caja de homología de 80 pb en un extremo del vector de clonación en una posición de 1 kb dentro del grupo de genes, o ambos extremos en una posición de 1 kb dentro del grupo de genes, y luego compararon la eficiencia de ExoCET con la eficiencia de T4pol y RecET (Figura 5b). Cuando ambas cajas homólogas del vector estaban al final, los inventores obtuvieron los mismos resultados que antes. Sin embargo, cuando se colocaron una o dos cajas de homología dentro de 1 kb del grupo de genes, la eficacia de clonación del tratamiento con T4pol solo fue muy baja, lo que indica que la hibridación después del tratamiento con exonucleasa T4pol dependía del emparejamiento complementario de la secuencia de ADN terminal. Vale la pena señalar que la eficacia de la recombinación de RecET tenía poco que ver con la posición de la caja de homología. Los resultados experimentales obtenidos cuando una caja

de homología estaba al final y la otra caja de homología estaba dentro, fueron estimulantes. Cuando solo había una caja de homología terminal entre el vector y el fragmento de ADN genómico, la eficiencia de ExoCET fue 12 veces mayor que la de RecET, lo que indicó que el tratamiento *in vitro* con T4pol permitió que dos moléculas de ADN se ligaran de manera eficiente mediante la hibridación de un extremo antes de la electroporación y la recombinación de RecET. Los datos anteriores indicaron que T4pol solo podía actuar sobre la caja de homología terminal, y la recombinación de la caja de homología interna requería la acción de RecET. Por lo tanto, los inventores habían llegado a la conclusión de que la principal contribución de T4pol a ExoCET era aumentar significativamente su eficacia de cotransformación hibridando un extremo de dos moléculas de ADN lineales. RecET podría utilizarse para facilitar la reorganización en el otro extremo. En el experimento de clonación del grupo de genes *lux* de 14 kb, cuando ambas cajas de homología estaban ubicadas en el extremo del fragmento genómico digerido por la diana, la eficiencia de clonación de ExoCET fue de 6 a 8 veces mayor que la de T4pol (Figura 1F y 5B), lo que indicó que la mayoría de los productos ensamblados *in vitro* tenían solo un extremo terminal unido (> 85%), y el plásmido recombinante producido por RecET después de la cotransformación del producto cíclico ensamblado *in vitro* y dos moléculas de ADN lineales contribuyeron poco a la alta eficiencia de ExoCET. Al diseñar un experimento de clonación directa con ExoCET, la posición de la posición de la caja de homología fue muy importante para la eficiencia de la clonación. Para lograr la mayor eficacia de clonación, ambas cajas de homología se colocaron al final del fragmento de ADN diana. De hecho, para ejercer el efecto de la exonucleasa de T4pol y el apareamiento *in vitro*, se colocó al menos una caja de homología en el extremo. Sin embargo, podría colocarse otra caja de homología dentro del fragmento de ADN diana porque RecE/RecT podría localizarlo y recombinarlo. Esto fue muy ventajoso para el uso de la clonación directa para construir vectores de expresión, ya que una de las cajas de homología podría colocarse al final del extremo 3' del fragmento diana, y podría colocarse el extremo 5' del gen diana directamente bajo el control del promotor y los sitios de unión del ribosoma usando cajas de homología internas.

Ejemplo 3

RecE y RecT eran necesarios para ExoCET

Dado que RecT es una proteína de hibridación de ADN monocatenario, los inventores plantearon la hipótesis de que RecT puede hibridar la región de ADN monocatenario producida por T4pol (exonucleasa 3'), por lo que RecE puede no ser necesaria en el sistema de técnica ExoCET. Para verificar esta conjetura, los inventores transformaron el sustrato de ADN tratado con T4pol en células de *E. coli* que expresan RecT y Redy (pSC101-Tg) y no expresan RecE, y no encontraron interacción entre RecT y T4pol. Por lo tanto, tanto RecE como RecT fueron necesarios para ExoCET (Figura 5C). RecA tuvo una cierta mejora en la eficiencia de la clonación directa.

Ejemplo 4

Verificación de la clonación directa de grandes fragmentos de ADN

Con el fin de verificar la superioridad de la técnica ExoCET, los inventores la utilizaron para realizar algunos experimentos que anteriormente eran difíciles de realizar con la técnica RecET. Hay dos grandes grupos de genes en el genoma de *Photobacterium luminescens*: plu3535-3532 de 37,5 kb y plu2670 de 52,6 kb. Fue muy difícil para los inventores clonar directamente estos dos grupos de genes con la técnica RecET, y las eficiencias fueron sólo 2/12 y 0/48³, respectivamente. Mientras que los inventores que utilizaron la técnica ExoCET lograron tasas correctas de 10/12 y 11/17, respectivamente (Tabla 2).

Tabla 2. Grandes fragmentos de ADN genómico clonados directamente de bacterias, células de mamíferos y sangre humana utilizando ExoCET

Gen diana	Fuente	Genoma (Mb)	Enzima de restricción genómica	Tamaño (kb)	Vector	Número de colonias (/ml)	Número correcto/ Número de detección
<i>plu3535-3532</i>	(<i>Photobacterium luminescens</i>) DSM15139	5,69	XbaI	38	pBAC2 015	1815 ± 132	10/12
<i>plu2670</i>	(<i>Photobacterium luminescens</i>) DSM15139	5,69	XbaI + X mal	53	p15A	1152 ± 211	11/17
Grupo de genes de salinomicina	(<i>Streptomyces albus</i>) DSM41398	8,38	EcoRV	106	pBeloB AC11	425 ± 91	2/24
Grupo de genes de salinomicina	(<i>Streptomyces albus</i>)	8,38	Cas9	106	pBeloB AC11	260 ± 14	1/24

Gen diana	Fuente	Genoma (Mb)	Enzima de restricción genómica	Tamaño (kb)	Vector	Número de colonias (/ml)	Número correcto/ Número de detección
	DSM41398						
<i>Prkar1a</i>	Células B16 de melanoma de ratón	2800,06	HpaI	8	p15A	205 ± 17	10/12
<i>Dpy30</i>	Células B16 de melanoma de ratón	2800,06	BamHI + KpnI	8,7	p15A	273 ± 18	9/12
<i>Wnt4</i>	Células B16 de melanoma de ratón	2800,06	Swal	45	p15A	76 ± 16	8/25
<i>Lmbr1l-Tuba1a</i>	Células B16 de melanoma de ratón	2800,06	Swal	53	p15A	52 ± 6	1/12
<i>DPY30</i>	Células 293T de riñón embrionario humano	3221,49	SpeI	9,1	p15A	40 ± 10	17/24
<i>DPY30</i>	Sangre humana	3221,49	SpeI	9,1	p15A	45 ± 2	5/24
<i>IGFLR1-LIN37</i>	Sangre humana	3221,49	NdeI + BstZ17I	14	p15A	320 ± 67	9/48
<i>IGFLR1-ARHGAP3 3</i>	Sangre humana	3221,49	BstZ17I	41	p15A	275 ± 76	5/48
<i>ZBTB32-LIN37</i>	Sangre humana	3221,49	NdeI	45	p15A	115 ± 35	2/48
<i>Oct4-Venus</i>	Célula madre embrionaria R1 de ratón	2800,06	EcoRV + PacI	9,6	p15A	34 ± 1	9/36
<i>Nanog-Cherry</i>	Célula madre embrionaria R1 de ratón	2800,06	NdeI	13	p15A	49 ± 12	17/54
<i>Gata2-Venus</i>	Célula madre embrionaria R1 de ratón	2800,06	BstZ17I	16,8	p15A	212 ± 27	5/45
<i>MII4(1)</i>	Célula madre embrionaria R1 de ratón	2800,06	SspI + SpeI	17,1	p15A	127 ± 38	7+3/24
<i>MII4(2)</i>	Célula madre embrionaria R1 de ratón					323 ± 65	2+2/36
<i>MII4(3)</i>	Célula madre embrionaria R1 de ratón					142 ± 27	6+9/72
<i>MII4(4)</i>	Célula madre embrionaria R1 de ratón					483 ± 91	3+5/36

Además, anteriormente el inventor falló en el intento de clonar directamente el grupo de genes de salinomicina de 106 kb del genoma de *Streptomyces albidans*. Entonces, el inventor tuvo que dividirlo en tres fragmentos y luego realizar la clonación paso a paso, y luego integrarlo en un grupo de genes completo⁴³. Sin embargo, a través de ExoCET, los inventores pudieron clonar directamente el grupo de genes de salinomicina de 106 kb en el vector BAC utilizando un

vector BAC con una caja de homología y ADN genómico digerido con EcoRV, y obtuvieron la tasa correcta de 2/24 (Tabla 2 y Figura 6). Dado que había un sitio de restricción EcoRV a cada lado del grupo de genes de salinomicina de 106 kb, los inventores pudieron liberarlo del cromosoma y clonarlo en el vector. Al clonar fragmentos de ADN grandes, a veces era difícil encontrar sitios de restricción adecuados en ambos lados del fragmento de ADN diana, pero el uso de nucleasas programables podría eliminar la restricción en los sitios de restricción, especialmente es una endonucleasa mediada por ARN-Cas9^{48, 49}. Para probar esta idea, los inventores usaron Cas9 para liberar el grupo de genes de salinomicina de 106 kb del cromosoma donde estaba muy cerca del EcoRV, y luego clonaron el mismo fragmento de ADN de 106 kb en el vector BAC usando el mismo vector BAC. Finalmente se obtuvo una eficacia de clonación similar (Tabla 2 y Figura 6). Por lo tanto, ExoCET tuvo un rendimiento significativamente superior en la clonación directa de grandes fragmentos de ADN en comparación con RecET.

A continuación, los inventores probaron si la eficacia de ExoCET podría cumplir el requisito de clonar directamente grandes fragmentos de ADN del genoma de mamífero. Los inventores usaron SwaI para liberar un fragmento de 45 kb que contenía el gen *Wnt4* del genoma del ratón (Figura 7A), y el inventor obtuvo la tasa correcta de 8/25 a través de ExoCET (Figura 7B). Los inventores también probaron el uso del ensamblaje de Gibson¹ para clonar este fragmento de ADN. El ensamblaje de Gibson usa T5exo, ADN polimerasa Phusion y ADN ligasa Taq para ensamblar moléculas de ADN con cajas de homología entre sí. Los inventores obtuvieron un gran número de colonias (181.000 y 257.000) transformando el producto de ADN ensamblado de Gibson en *E. coli* GB05-dir inducida y no inducida por arabinosa que contiene pSC101-BAD-ETgA-tet. Y se detectaron 60 colonias y no se obtuvo ningún clon correcto (Figura 7B), y todos eran vectores vacíos de p15A de autocircularización.

Ejemplo 5

Ensamblaje de fragmentos de ADN usando ExoCET

Gibson era un método para el ensamblaje de ADN de múltiples fragmentos, por lo que los inventores compararon ExoCET y Gibson a través de algunos experimentos de ensamblaje de múltiples fragmentos de ADN (Figura 7C). Estos fragmentos de ADN se amplificaron mediante PCR y tenían una caja de homología de 40 pb al final. La eficiencia del ensamblaje de ExoCET y Gibson fue buena en los experimentos de ensamblaje de 7 fragmentos y 10 fragmentos. Cuando los productos ensamblados *in vitro* de Gibson se transformaron en células de *E. coli* que expresaban RecE, RecT, Red γ y RecA (Gibson y ETgA), la eficiencia y precisión del ensamblaje mejoraron significativamente (Figura 7D). ExoCET no pudo ensamblar más de 13 fragmentos de ADN, Gibson no pudo ensamblar más de 16 fragmentos de ADN, mientras que Gibson y ETgA pudieron ensamblar al menos 20 fragmentos de ADN en un plásmido de 54,9 kb. Por lo tanto, combinando el ensamblaje *in vitro* con la recombinación de RecET, las ventajas del ensamblaje de ADN eran obvias.

Ejemplo 6

Construcción del vector de direccionamiento haplotipo isogénico usando ExoCET

ExoCET también podría usarse para clonar directamente fragmentos de ADN de genomas de mamíferos que incluyen sangre, líneas celulares asociadas a enfermedades, etc. para facilitar los estudios de haplotipos de SNP y para construir rápidamente vectores de direccionamiento singenéticos de haplotipos (HIT) para el direccionamiento de células madre humanas mediado por nucleasas. La importancia de las células madre humanas aisladas de los pacientes, la reprogramación de la sangre del cordón umbilical o de las células somáticas en la investigación biomédica ha recibido cada vez más atención. La investigación sobre la modificación precisa del genoma de las células madre también había recibido una amplia atención. Transformar el genoma humano fue más desafiante que estructurar el genoma de ratones experimentales porque la diversidad genética humana era compleja. La importancia de la isogenicidad (similitud de secuencia) para la recombinación homóloga se hizo evidente hace muchos años cuando la gente usaba células madre embrionarias de ratón para el direccionamiento de genes¹⁴.

El efecto de la recombinación homóloga de secuencia no coincidente no se había estudiado bien. Por ejemplo, ¿cuánta eficiencia de recombinación podría reducir una sola falta de coincidencia (una SNP o una inserción/eliminación)? ¿Cómo afectó la distancia del sitio de recombinación no coincidente a la eficiencia de la recombinación? ¿Cómo afectaron las múltiples faltas de coincidencia a la eficiencia de la recombinación? Estas cuestiones aún no se han aclarado. En cualquier caso, el uso de secuencias idénticas en el direccionamiento de genes fue claramente muy recomendable, por lo que ExoCET fue una forma eficaz de obtener rápidamente cajas de homología ideales. A diferencia del método de amplificación de cajas de homología del genoma mediante PCR, ExoCET no estaba limitado por el tamaño del fragmento, no introducía mutaciones y era capaz de mantener un haplotipo de ADN. Además, los extremos de las cajas de homología también podrían seleccionarse de acuerdo con la forma de genotipado (tal como transferencia Southern o PCR de ligación), por lo que podría optimizarse la longitud de las cajas de homología. Por lo tanto, ExoCET ofrece ventajas para la cirugía genómica individualizada, especialmente cuando se combina con CRISPR/Cas9¹⁵.

Los inventores utilizaron ExoCET para construir vectores de direccionamiento isogénicos para diseñar genomas de mamíferos. Dada la experiencia⁵⁰ en la investigación de células madre embrionarias de ratón, los inventores intentaron

clonar un fragmento de ADN de 5 a 10 kb directamente del genoma humano o de ratón como una caja de homología isogénica (Figura 8A y Figura 9A). Es de destacar que estos fragmentos de ADN no solo eran del mismo gen, sino que también mantienen el haplotipo polimórfico, por lo que los inventores lo llamaron vector HIT (dirección isogénica de haplotipo). Los inventores clonaron directamente fragmentos de ADN de 8-9 kb de los genomas humanos (líneas celulares cultivadas *in vitro* y sangre humana) y de ratón (líneas celulares cultivadas *in vitro*) mediante ExoCET (Figura 8B y Figura 9B). A continuación, se insertaron marcadores de selección y otros elementos funcionales¹⁶ en el vector HIT mediante ingeniería recombinante con Redαβ (Figura 8C y Figura 9C).

Ejemplo 7

Genotipado de células de mamíferos usando ExoCET

ExoCET también podría usarse como el método más confiable para genotipificación de un genoma modificado, mientras que la PCR de ligación y transferencia Southern podrían producir señales positivas falsas. Dado que la PCR de largo alcance era propensa a señales positivas falsas en estudios de genotipado de mamíferos, los inventores previamente querían confirmar las células madre embrionarias de ratón dirigidas neo-Kmt2d-AID cribadas por PCR de segmento largo por transferencia Southern. Sin embargo, los inventores nunca consiguieron una buena prueba. Por lo tanto, los inventores clonaron un fragmento de ADN que contenía todo el elemento de direccionamiento del genoma de cuatro posibles células madre embrionarias de ratón dirigidas por Kmt2d-AID-neo usando el método ExoCET mostrado en la Figura 10 A. El plásmido recombinante clonado por ExoCET se sometió a análisis de restricción y secuenciación de ADN, y los resultados mostraron que las cuatro células fueron direccionadas con éxito y fueron direccionadas en forma única (Figura 10B). Además, los inventores han utilizado con éxito ExoCET para revalidar las células madre embrionarias de ratón diana Oct4-Venus-neo, Nanog-Cherry-neo, Gata2-Venus-neo y Set1b-TC-neo que se obtuvieron previamente (Tabla 3). Estas células diana se verificaron previamente mediante transferencia Southern. Estos resultados indicaron que no había restricción de sitio en la genotipificación de ExoCET. Los inventores anteriormente no pudieron determinar si una célula madre embrionaria de ratón dirigida a Klf4-Venus-neo fue dirigida con éxito, porque la PCR de largo alcance y la transferencia Southern no habían obtenido una señal exacta. Los inventores no clonaron un fragmento de ADN que tuviera resistencia a la kanamicina en la región correspondiente del genoma usando ExoCET (Tabla 3). El análisis de restricción del plásmido resistente al cloranfenicol clonado encontró que el 50% de ellos contenía la secuencia de ADN de tipo silvestre (Figura 11). Por lo tanto, los inventores sabían exactamente que esta célula no estaba correctamente direccionada.

Tabla 3. Datos experimentales de genotipado ExoCET

Célula madre embrionaria de ratón (cantidad de ADN transformado)	Enzima de restricción genómica	Tamaño (kb)	Número de colonias en placas de cloranfenicol (/ml)	Placas de kanamicina y cloranfenicol/doble raya (Km/Cm)	Tasa correcta de colonias en la antiplaca doble de kanamicina + cloranfenicol*
Oct4-Venus-neo #7 (R1) (10 µg ADN genómico + 250 ng vector)	EcoRV + PacI	9,6	34 ± 1	9/36	9/9
Nanog-Cherry-neo #18 (R1) (7,7 µg ADN genómico + 400 ng vector)	NdeI	13	49 ± 12	17/54	7/17**
Gata2-Venus-neo #19 (GM8) (1,5 µg ADN genómico + 500 ng vector)	BstZ17I	16,8	212 ± 27	5/45	5/5
Set1 b-TC-neo #4 (R1) (10 µg ADN genómico + 250 ng vector)	Asel	24	49 ± 8	1/36	1/1
Klf4-Venus-neo #9 (R1) (10 µg de ADN genómico + 250 ng vector)	AflII	10,2	18 ± 4	0/36	-

* en la Tabla 3 significa después del análisis de restricción.

** en la Tabla 3 significa que los 10 restantes fueron recombinación intramolecular (que contienen 11 repeticiones directas mayores de 40 pb en la secuencia diana clonada)

Tabla 4. La cantidad de ADN genómico de células madre embrionarias de ratón dirigida a Oct4-Venus-neo necesaria para optimizar el genotipado de ExoCET

Se mezclaron 500 ng de ADN genómico digerido con EcoRV y PacI con la cantidad de vector p15A-cm que se muestra a continuación		
Vector	Número de colonias en placa de doble resistencia a kanamicina y cloranfenicol (/ml)	Número correcto/número de detección*
500 ng	12	3/3, 100%
1000 ng	10	3/3, 100%
2000 ng	8	3/3, 100%
Se mezclaron 1000 ng de ADN genómico digerido con EcoRV y PacI con la cantidad de vector p15A-cm que se muestra a continuación		
Vector	Número de colonias en placa de doble resistencia a kanamicina y cloranfenicol (/ml)	Número correcto/número de detección*
500 ng	10	3/3, 100%
1000 ng	10	3/3, 100%
2000 ng	10	3/3, 100%
* en la Tabla 4 significa después del análisis de restricción.		

La genotipificación de ExoCET no produjo una señal positiva falsa en comparación con la PCR de largo alcance. En comparación con la transferencia Southern, la genotipificación de ExoCET es más simple y no requirió un cribado engorroso de las sondas de hibridación. En la genotipificación de ExoCET, los sitios de enzimas de restricción para la liberación de elementos de direccionamiento intactos estaban fácilmente disponibles y, en el caso de genomas bien preparados, los resultados de la genotipificación se obtuvieron en tres días. Más importante aún, ExoCET nunca produjo una señal positiva falsa. Dado que el elemento de direccionamiento tenía un marcador de selección, siempre que 500 ng de ADN genómico de la enzima de restricción fueran suficientes para obtener una mejor eficacia de clonación (Tabla 4). Para aumentar el rendimiento de la genotipificación de ExoCET, se pueden utilizar células cultivadas en placas de 96 pocillos.

Ejemplo 8

Técnica de clonación ExoCET aplicada a muestras metagenómicas

El análisis funcional de los resultados de la secuenciación del genoma completo requiere un método simple y rápido de construcción del vector de expresión. De acuerdo con el método de la divulgación, los inventores pueden clonar un fragmento de ADN de hasta 50 kb a partir de un genoma de $3,0 \times 10^9$ pb. Con este fin, los inventores diluyeron 1 ng de ADN genómico de *P. phosphoreum* en 10 µg de ADN genómico de *Bacillus subtilis* para imitar la metagenómica. El experimento clonó con éxito el grupo de genes *lux* de 14 kb por ExoCET y obtuvo una eficiencia considerable (Tabla 5). Las muestras ambientales usualmente contenían más de 10^4 especies⁵¹⁻⁵³, por lo que los resultados pueden motivar a los inventores a aplicar la técnica de clonación ExoCET a muestras metagenómicas.

Tabla 5. El grupo de genes *lux* de 14 kb se clonó directamente del genoma de *P. phosphoreum* diluido utilizando ExoCET

<i>P. phosphoreum</i> (ng) (BamHI + KpnI)	<i>B. subtilis</i> (µg) (BamHI)	Vectores	Número de colonias (/ml)	Número correcto/ Número de detección
10	10	p15A-cm	200 ± 2	7/12
5	10	p15A-cm	142 ± 22	5/12
2	10	p15A-cm	102 ± 8	2/12
1	10	p15A-cm	104 ± 18	2/24

Referencias:

- Gibson, D.G. et al. Enzymatic assembly of ADN molecules up to several hundred kilobases. Nat. Methods 6, 343-345. (2009).
- Shao, Z. & Zhao, H. ADN assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 37, e16 (2009).

3. Fu, J. et al. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol* 30, 440-446 (2012).
4. Tang, Y. et al. Heterologous expression of an orphan NRPS gene cluster from *Paenibacillus* larvae in *Escherichia coli* revealed production of sevadicin. *J. Biotechnol.* 194, 112-114 (2015).
5. Bian, X. et al. Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in *E. coli* through Red/ET recombineering. *ChemBioChem* 13, 1946-1952 (2012).
6. Bian, X. et al. Heterologous production of glidobactins/luminmycins in *Escherichia coli* Nissle containing the glidobactin biosynthetic gene cluster from *Burkholderia* DSM7029. *Chembiochem* 15, 2221-2224 (2014).
7. Bian, X. et al. *In vivo* evidence for a prodrug activation mechanism during colibactin maturation. *Chembiochem* 14, 1194-1197 (2013).
8. Larionov, V. et al. Specific cloning of human ADN as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. USA* 93, 491-496 (1996).
9. Lee, N.C., Larionov, V. & Kouprina, N. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Res.* 43, e55 (2015).
10. Kouprina, N. & Larionov, V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* (2016).
11. Jiang, W. et al. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.* 6, 8101 (2015).
12. Bian, X., Plaza, A., Zhang, Y. & Müller, R. Luminmycins A-C, cryptic natural products from *Photobacterium* *luminescens* identified by heterologous expression in *Escherichia coli*. *J. Nat. Prod.* 75, 1652-1655 (2012).
13. Zhou, Y. et al. Iterative mechanism of macrodiolide formation in the anticancer compound conglobatin. *Chem. Biol.* 22, 745-754 (2015).
14. te Riele, H., Maandag, E.R. & Berns, A. Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic ADN constructs. *Proc. Natl. Acad. Sci. USA* 89, 5128-5132 (1992).
15. Baker, O. et al. RAC-tagging: Recombineering And Cas9-assisted targeting for protein tagging and conditional analyses. *Sci. Rep.* 6, 25529 (2016).
16. Fu, J., Teucher, M., Anastassiadis, K., Skarnes, W. & Stewart, A.F. A recombineering pipeline to make conditional targeting constructs. *Methods Enzymol.* 477, 125-144 (2010).
17. Hofemeister, H. et al. Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. *Methods* 53, 437-452 (2011).
18. Copeland, N.G., Jenkins, N.A. & Court, D.L. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* 2, 769-779 (2001).
19. Muirers, J.P., Zhang, Y. & Stewart, A.F. Techniques: Recombinogenic engineering--new options for cloning and manipulating ADN. *Trends Biochem. Sci.* 26, 325-331 (2001).
20. Narayanan, K. & Chen, Q. Bacterial artificial chromosome mutagenesis using recombineering. *J. Biomed. Biotechnol.* 2011, 971296 (2011).
21. Murphy, K.C. Phage recombinases and their applications. *Adv. Virus Res.* 83, 367-414 (2012).
22. Murphy, K.C. lambda Recombination and Recombineering. *EcoSal Plus* 7 (2016).
23. Zhang, Y., Buchholz, F., Muirers, J.P. & Stewart, A.F. A new logic for ADN engineering using recombination in *Escherichia coli*. *Nat. Genet.* 20, 123-128 (1998).
24. Clark, A.J. et al. Genes of the RecE and RecF pathways of conjugational recombination in *Escherichia coli*. *Cold Spring Harb. Symp. Quant. Biol.* 49, 453-462 (1984).
25. Hall, S.D., Kane, M.F. & Kolodner, R.D. Identification and characterization of the *Escherichia coli* RecT protein, a protein encoded by the recE region that promotes renaturation of homologous single-stranded ADN. *J. Bacteriol.* 175, 277-287 (1993).
26. Clark, A.J., Satin, L. & Chu, C.C. Transcription of the *Escherichia coli* recE gene from a promoter in Tn5 and IS50. *J. Bacteriol.* 176, 7024-7031 (1994).
27. Kuzminov, A. Recombinational repair of ADN damage in *Escherichia coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.* 63, 751-813 (1999).
28. Chu, C.C., Templin, A. & Clark, A.J. Suppression of a frameshift mutation in the recE gene of *Escherichia coli* K-12 occurs by gene fusion. *J. Bacteriol.* 171, 2101-2109 (1989).
29. Muirers, J.P., Zhang, Y., Buchholz, F. & Stewart, A.F. RecE/RecT and Reda/Redb initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev.* 14, 1971-1982 (2000).
30. Chang, H.W. & Jullin, D.A. Structure and function of the *Escherichia coli* RecE protein, a member of the RecB nuclease domain family. *J. Biol. Chem.* 276, 46004-46010 (2001).
31. Zhang, J., Xing, X., Herr, A.B. & Bell, C.E. Crystal structure of *E. coli* RecE protein reveals a toroidal tetramer for processing double-stranded ADN breaks. *Structure* 17, 690-702 (2009).
32. Zhang, Y., Muirers, J.P., Rientjes, J. & Stewart, A.F. Phage annealing proteins promote oligonucleotide-directed mutagenesis in *Escherichia coli* and mouse ES cells. *BMC Mol. Biol.* 4, 1 (2003).
33. Szczepanska, A.K. Bacteriophage-encoded functions engaged in initiation of homologous recombination events. *Crit. Rev. Microbiol.* 35, 197-220 (2009).
34. Kolodner, R., Hall, S.D. & Luisi-DeLuca, C. Homologous pairing proteins encoded by the *Escherichia coli* recE and recT genes. *Mol Microbiol* 11, 23-30 (1994).
35. Iyer, L.M., Koonin, E.V. & Aravind, L. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC genomics* 3, 8 (2002).

36. Maresca, M. et al. Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol. Biol.* 11, 54 (2010).
37. Poteete, A.R. Involvement of Escherichia coli ADN replication proteins in phage Lambda Red-mediated homologous recombination. *PloS one* 8, e67440 (2013).
- 5 38. Muirers, J.P., Zhang, Y., Testa, G. & Stewart, A.F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* 27, 1555-1557 (1999).
39. Zhang, Y., Muirers, J.P., Testa, G. & Stewart, A.F. ADN cloning by homologous recombination in Escherichia coli. *Nat. Biotechnol.* 18, 1314-1317 (2000).
40. Testa, G. et al. Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nat. Biotechnol.* 21, 443-447 (2003).
- 10 41. Sarov, M. et al. A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. *Nat. Methods* 3, 839-844 (2006).
42. Bird, A.W. et al. High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. *Nat. Methods* 9, 103-109 (2012).
- 15 43. Yin, J. et al. Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from Streptomyces albus DSM41398 in Streptomyces coelicolor A3(2). *Sci. Rep.* 5, 15081 (2015).
44. Wang, H. et al. Improved seamless mutagenesis by recombineering using ccdB for counterselection. *Nucleic Acids Res* 42, e37 (2014).
45. Pospiech, A. & Neumann, B. A versatile quick-prep of genomic ADN from Gram-positive bacteria. *Trends Genet.* 11, 217-218 (1995).
- 20 46. Wang, H. et al. RecET direct cloning and Redalphabeta recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* 11, 1175-1190 (2016).
47. Zhang, S.D. et al. Genome Sequence of Luminous Piezophile Photobacterium phosphoreum ANT-2200. *Genome Announc.* 2, e0009614 (2014).
- 25 48. Jinek, M. et al. A programmable dual-RNA-guided ADN endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012).
49. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific ADN cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* 109, E2579-2586 (2012).
50. Skarnes, W.C. et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* 474, 337-342 (2011).
- 30 51. Torsvik, V., Goksoyr, J. & Daae, F.L. High diversity in ADN of soil bacteria. *Appl. Environ. Microbiol.* 56, 782-787 (1990).
52. Rappe, M.S. & Giovannoni, S.J. The uncultured microbial majority. *Annu. Rev. Microbiol.* 57, 369-394 (2003).
53. Charlop-Powers, Z., Milshteyn, A. & Brady, S.F. Metagenomic small molecule discovery methods. *Curr. Opin. Microbiol.* 19, 70-75 (2014).
- 35

REIVINDICACIONES

1. Un método de recombinación homóloga, que comprende:

- 5 tratar dos o más moléculas de ácido nucleico diana con una primera exonucleasa y recombinar las dos o más moléculas de ácido nucleico diana en presencia de una segunda exonucleasa y una proteína de hibridación en una célula huésped, en el que las moléculas de ácido nucleico diana recombinadas comparten al menos una secuencia idéntica, en la que al menos una secuencia idéntica está dentro o en un extremo de cada una de las dos o más moléculas de ácido nucleico diana,
- 10 en el que dicho tratamiento en presencia de la primera exonucleasa comprende la unión *in vitro*; comprendiendo dicha unión *in vitro* unir las dos o más moléculas de ácido nucleico diana,
- y en el que dicho tratamiento en presencia de la primera exonucleasa comprende digestión enzimática e hibridación;
- y en el que la segunda exonucleasa es RecE y la proteína de hibridación es RecT.
- 15 2. El método de la reivindicación 1, caracterizado porque al menos una secuencia idéntica comprende al menos 6 nucleótidos.
3. El método de la reivindicación 1, caracterizado porque la primera exonucleasa se selecciona del grupo que consiste en ADN polimerasa T4, fragmento Klenow de ADN polimerasa I, exonucleasa T5 y exonucleasa T7.
- 20 4. El método de la reivindicación 1, caracterizado porque el tratamiento de dos o más moléculas de ácido nucleico diana comprende además la adición de una ADN polimerasa, dNTP y una ADN ligasa.
5. El método de la reivindicación 1, caracterizado porque la célula huésped es una célula de levadura o una célula bacteriana y expresa una exonucleasa y una proteína de hibridación.
- 25 6. El método de la reivindicación 5, caracterizado porque la célula huésped es una célula de *E. coli* que expresa RecE y RecT de longitud completa, o la célula huésped expresa la exonucleasa y la proteína de hibridación en un vector plásmido y/o un cromosoma.
- 30 7. El método de la reivindicación 1, caracterizado porque las dos o más moléculas de ácido nucleico diana son un segmento de ADN lineal seleccionado de un fragmento de ADN digerido por endonucleasa, un fragmento de ADN amplificado por PCR, un fragmento de ADN genómico, un miembro de la biblioteca de ADNc, un fragmento derivado de cromosomas artificiales bacterianos (BAC) y un fragmento de vectores de clonación.
- 35 8. Un método para ensamblar una molécula de ácido nucleico, que comprende: tratar dos o más moléculas de ácido nucleico con una primera exonucleasa y recombinar las dos o más moléculas de ácido nucleico en presencia de una segunda exonucleasa y una proteína de hibridación, en el que cada molécula de ácido nucleico comparte al menos una secuencia idéntica con una molécula de ácido nucleico adyacente en un producto de ensamblaje resultante, y al
- 40 menos una secuencia idéntica está dentro o en un extremo de cada una de las dos o más moléculas de ácido nucleico, en el que dicho tratamiento en presencia de la primera exonucleasa comprende la unión *in vitro*; comprendiendo dicha unión *in vitro* unir las dos o más moléculas de ácido nucleico, y en el que dicho tratamiento en presencia de la primera exonucleasa comprende digestión enzimática e hibridación; y en el que la segunda exonucleasa es RecE y la proteína de hibridación es RecT.
- 45 9. El método de la reivindicación 8, caracterizado porque al menos una secuencia idéntica comprende al menos 6 nucleótidos.
10. El método de la reivindicación 8, caracterizado porque la primera exonucleasa se selecciona del grupo que consiste en ADN polimerasa T4, fragmento Klenow de ADN polimerasa I, exonucleasa T5 y exonucleasa T7.
- 50 11. El método de la reivindicación 8, caracterizado porque el tratamiento de dos o más moléculas de ácido nucleico comprende además la adición de una ADN polimerasa, dNTP y una ADN ligasa.
12. El método de la reivindicación 8, caracterizado porque la célula huésped es una célula de levadura o una célula bacteriana y expresa una exonucleasa y una proteína de hibridación.
- 55 13. El método de la reivindicación 12, caracterizado porque la célula huésped es una célula de *E. coli* que expresa RecE y RecT de longitud completa, o la célula huésped expresa la exonucleasa y la proteína de hibridación en un vector plásmido y/o un cromosoma.
- 60 14. El método de la reivindicación 8, caracterizado porque las dos o más moléculas de ácido nucleico son un segmento de ADN lineal seleccionado de un fragmento de ADN digerido por endonucleasa, un fragmento de ADN amplificado por PCR, un fragmento de ADN genómico, un miembro de la biblioteca de ADNc, un fragmento derivado de cromosomas artificiales bacterianos (BAC) y un fragmento de vectores de clonación.
- 65

15. Un método de clonación de un ADN genómico, que comprende:

tratar un vector de clonación lineal y una mezcla de fragmentos de ADN genómico con una primera exonucleasa, y recombinar el vector de clonación lineal y un fragmento de ADN diana de la mezcla de fragmentos de ADN genómico, caracterizado porque el vector de clonación lineal comparte al menos una secuencia idéntica con el fragmento de ADN diana de la mezcla de fragmentos de ADN genómico dentro o en un extremo de cada vector de clonación lineal y fragmento de ADN diana, en el que dicho tratamiento en presencia de la primera exonucleasa comprende la unión *in vitro*; comprendiendo dicha unión *in vitro* la unión de dos o más moléculas de ácido nucleico, y en el que dicho tratamiento en presencia de la primera exonucleasa comprende digestión enzimática e hibridación; y en el que la segunda exonucleasa es RecE y la proteína de hibridación es RecT.

16. El método de la reivindicación 15, caracterizado porque al menos una secuencia idéntica comprende al menos 6 nucleótidos.

17. El método de la reivindicación 15, caracterizado porque la primera exonucleasa se selecciona del grupo que consiste en ADN polimerasa de T4, fragmento Klenow de ADN polimerasa I, exonucleasa de T5 y exonucleasa de T7.

18. El método de la reivindicación 15, caracterizado porque el tratamiento de un vector de clonación lineal y una mezcla de fragmentos de ADN genómico comprende además la adición de una ADN polimerasa, dNTP y una ADN ligasa.

19. El método de la reivindicación 15, caracterizado porque la célula huésped es una célula de levadura o una célula bacteriana y expresa una exonucleasa y una proteína de hibridación.

20. El método de la reivindicación 19, caracterizado porque la célula huésped es una célula de *E. coli* que expresa RecE y RecT de longitud completa, o la célula huésped expresa la exonucleasa y la proteína de hibridación en un vector plásmido y/o un cromosoma.

21. El método de la reivindicación 15, caracterizado porque el fragmento de ADN diana es un segmento de ADN lineal seleccionado de un fragmento de ADN digerido por endonucleasa, un fragmento de ADN amplificado por PCR, un fragmento de ADN genómico, un miembro de la biblioteca de ADNC, un fragmento derivado de cromosomas artificiales bacterianos (BAC) y un fragmento de vectores de clonación.

22. Uso de un método de cualquiera de las reivindicaciones 1-7 en el ensamblaje de un fragmento de ADN, construcción de un vector de direccionamiento o genotipificación de células de mamífero.

23. Uso de un método de cualquiera de las reivindicaciones 8-14 en el ensamblaje de un fragmento de ADN.

24. Uso de un método de cualquiera de las reivindicaciones 15-21 en la construcción de un vector de direccionamiento o genotipificación de células de mamífero.

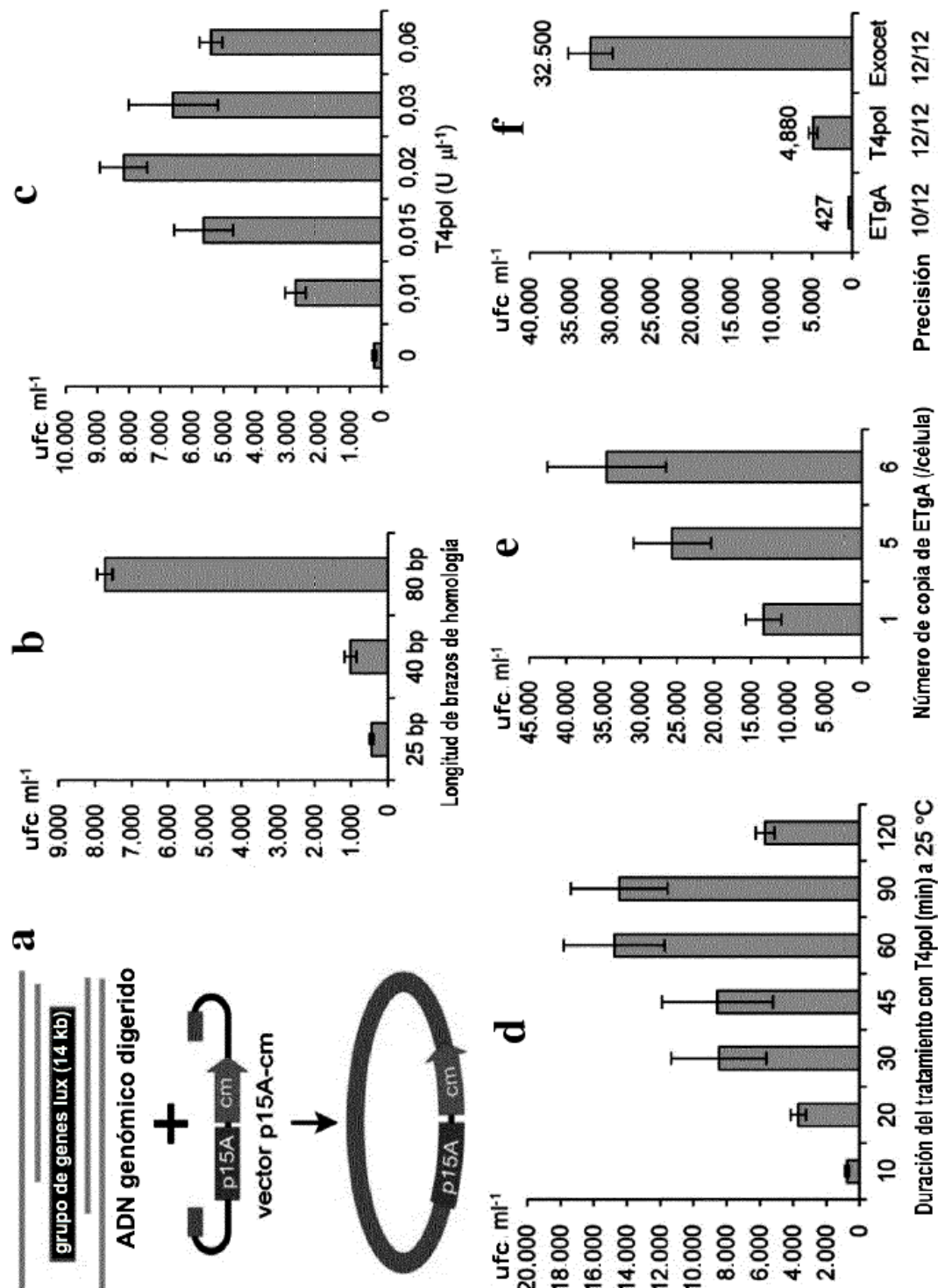


FIG. 1

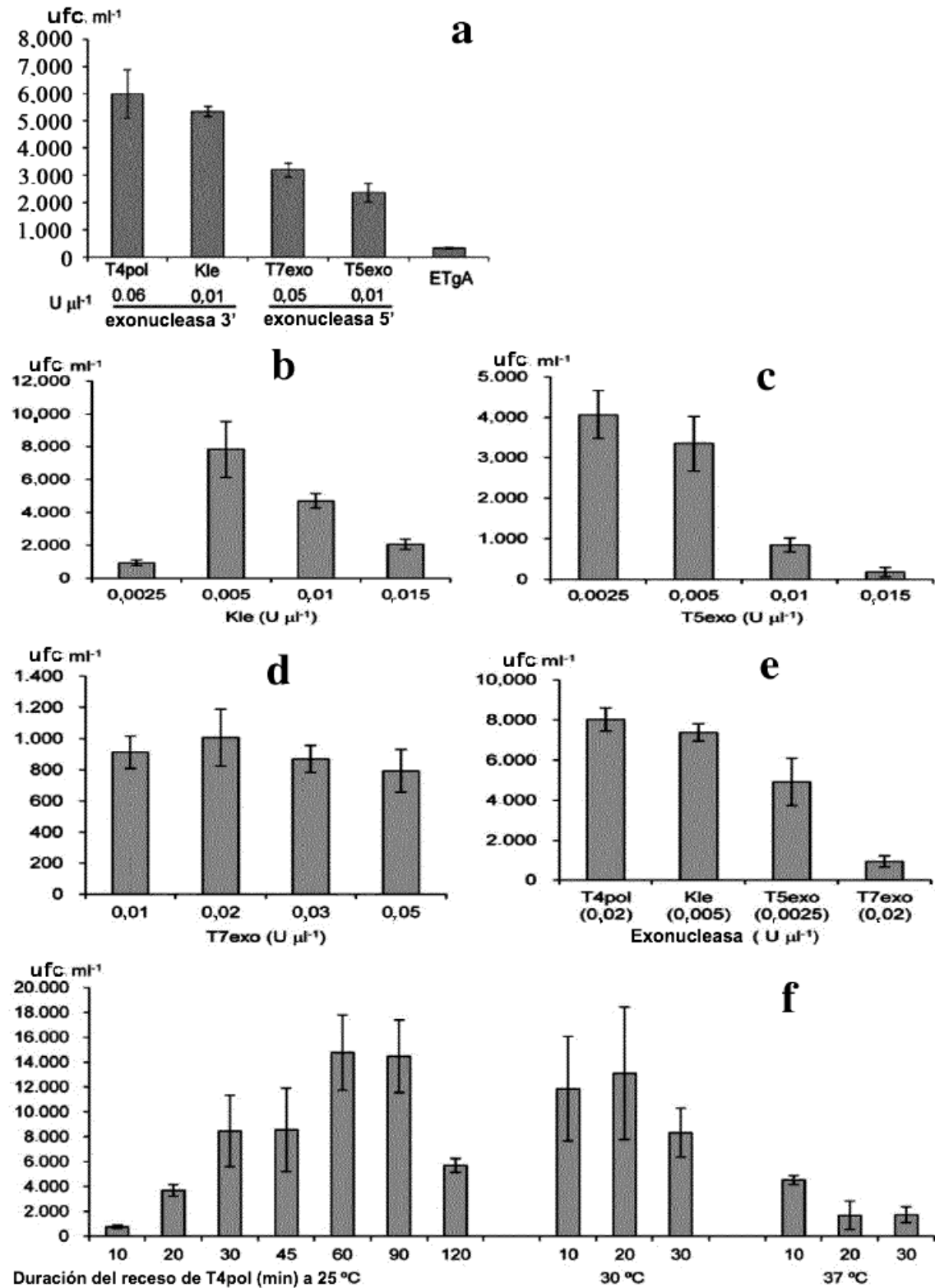
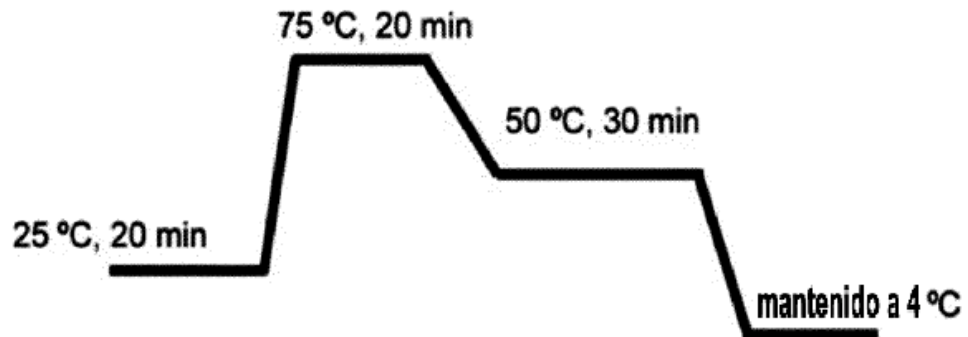


FIG. 2



Velocidad de enfriamiento:

A: 2 °C seg^{-1} (predeterminado)

B: 0,1 °C seg^{-1}

C: 75 °C a 25 °C en 40 min

25 °C a 4 °C: 0,1 °C seg^{-1}

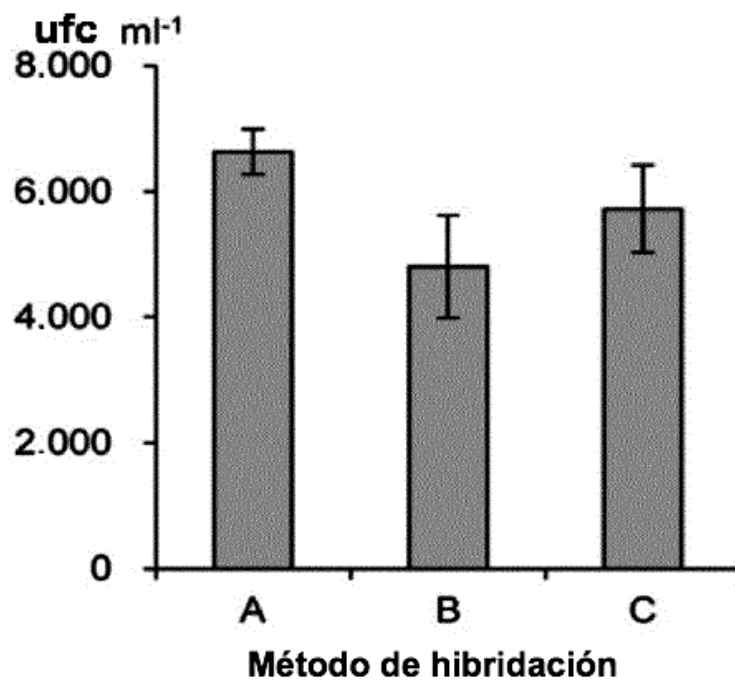


FIG. 3

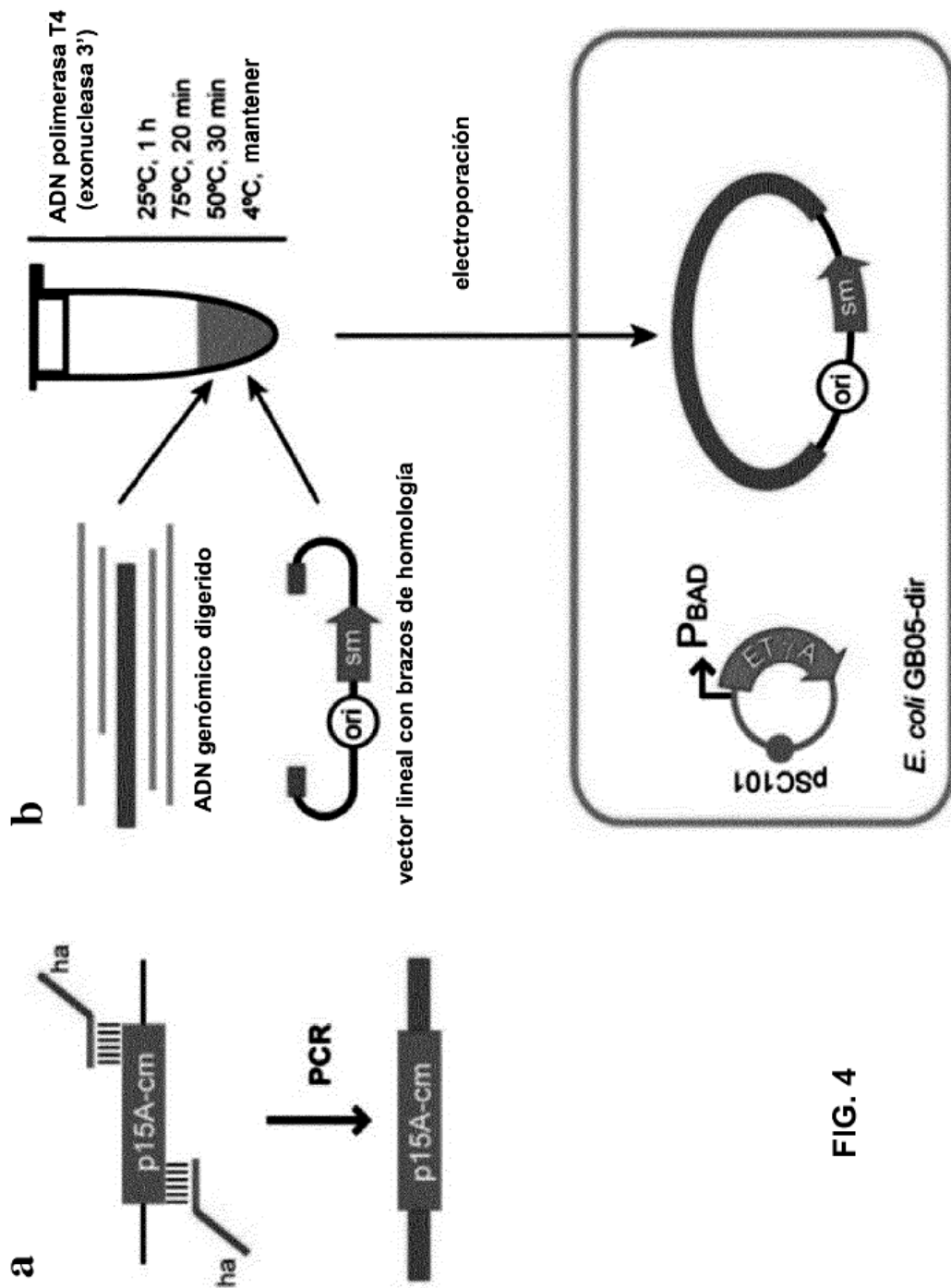


FIG. 4

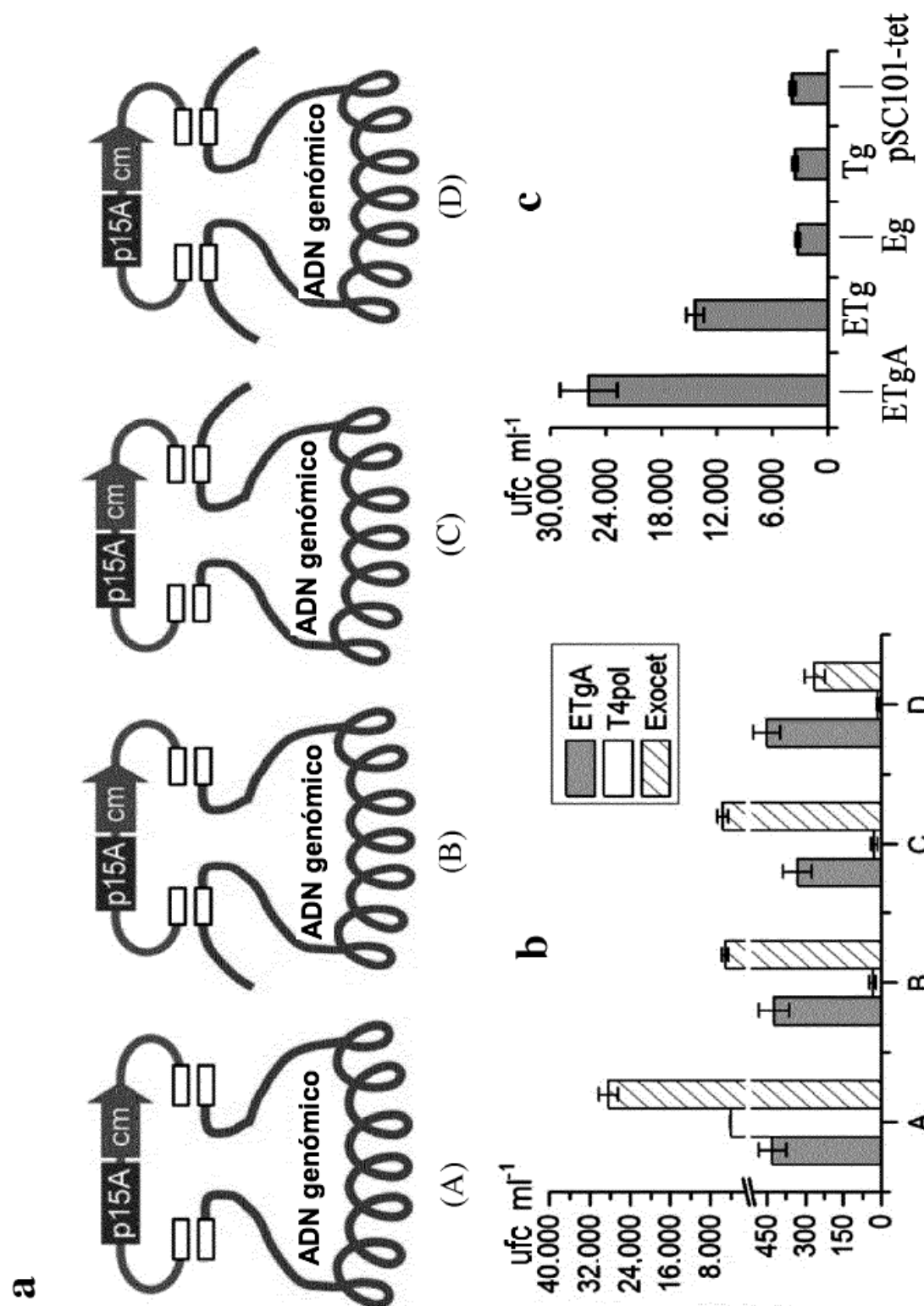


FIG. 5

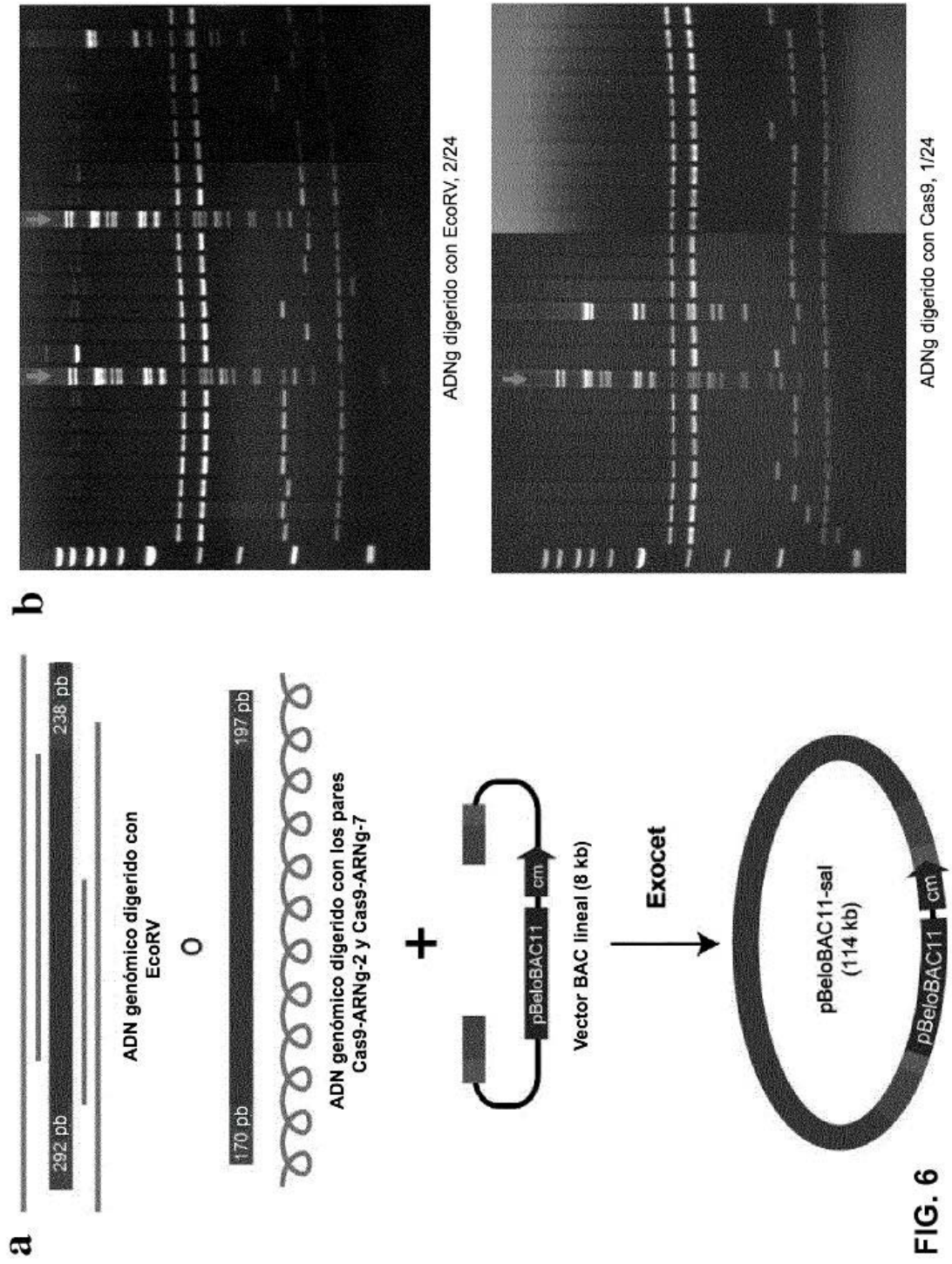


FIG. 6

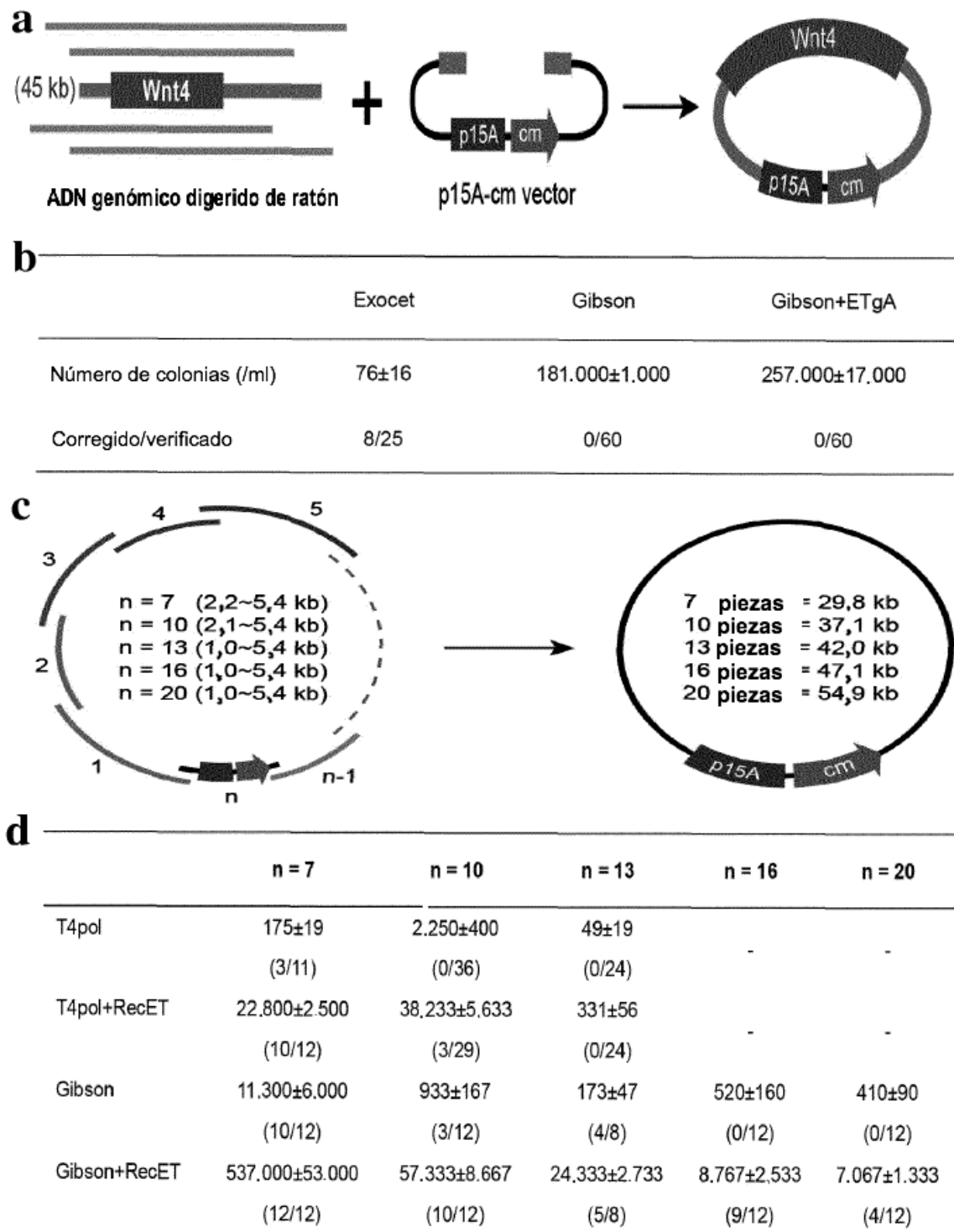
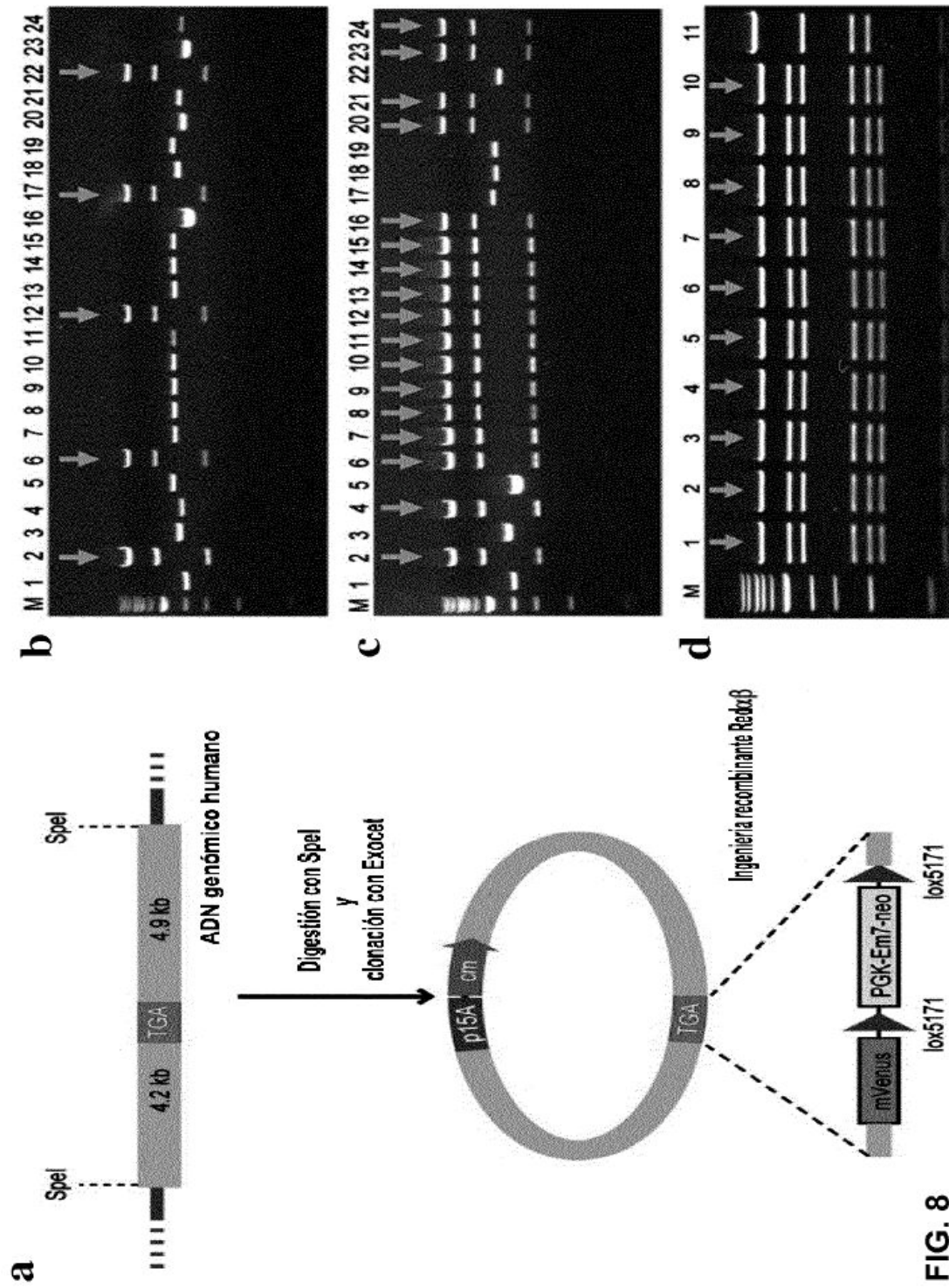


FIG. 7



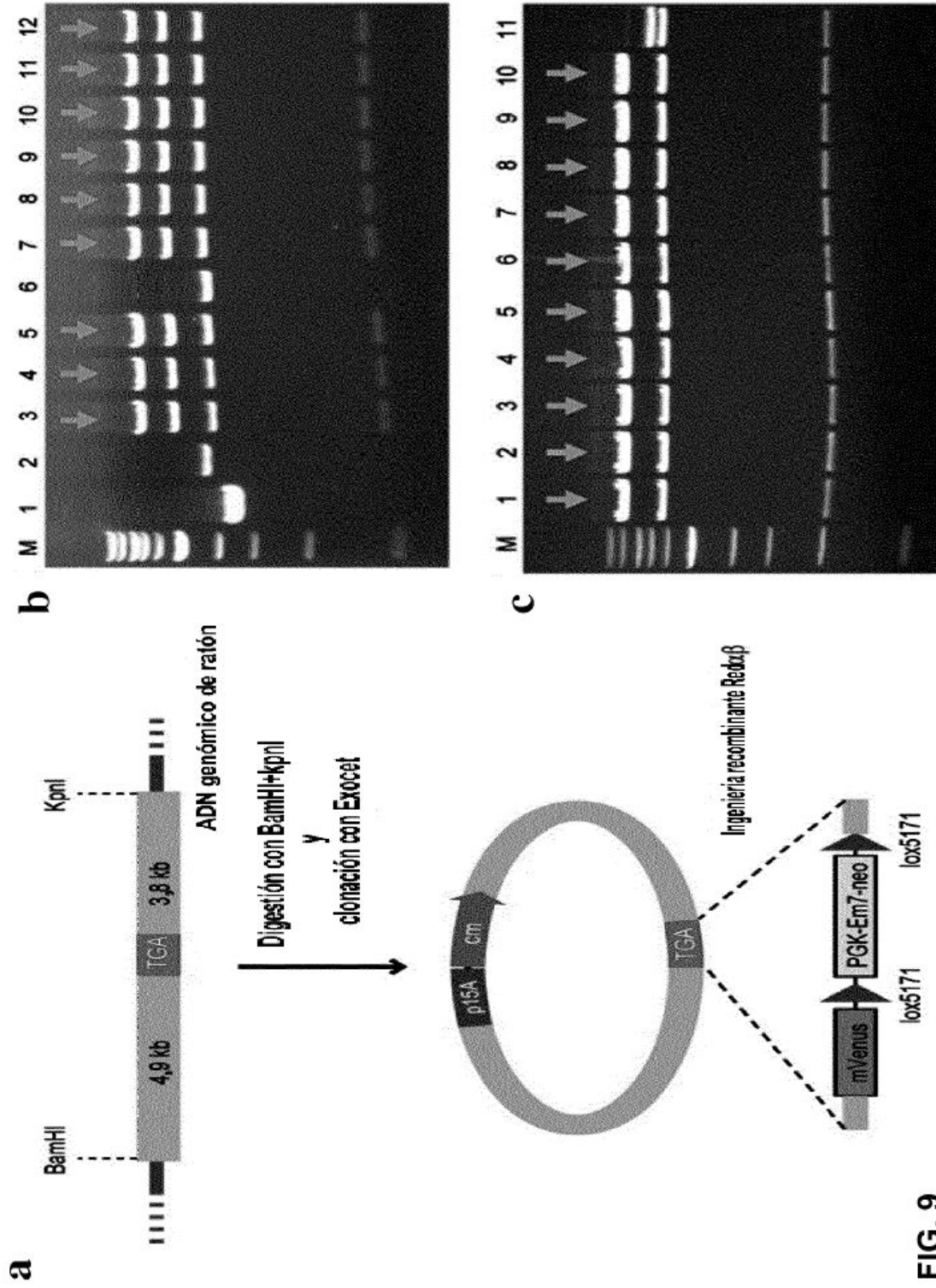


FIG. 9

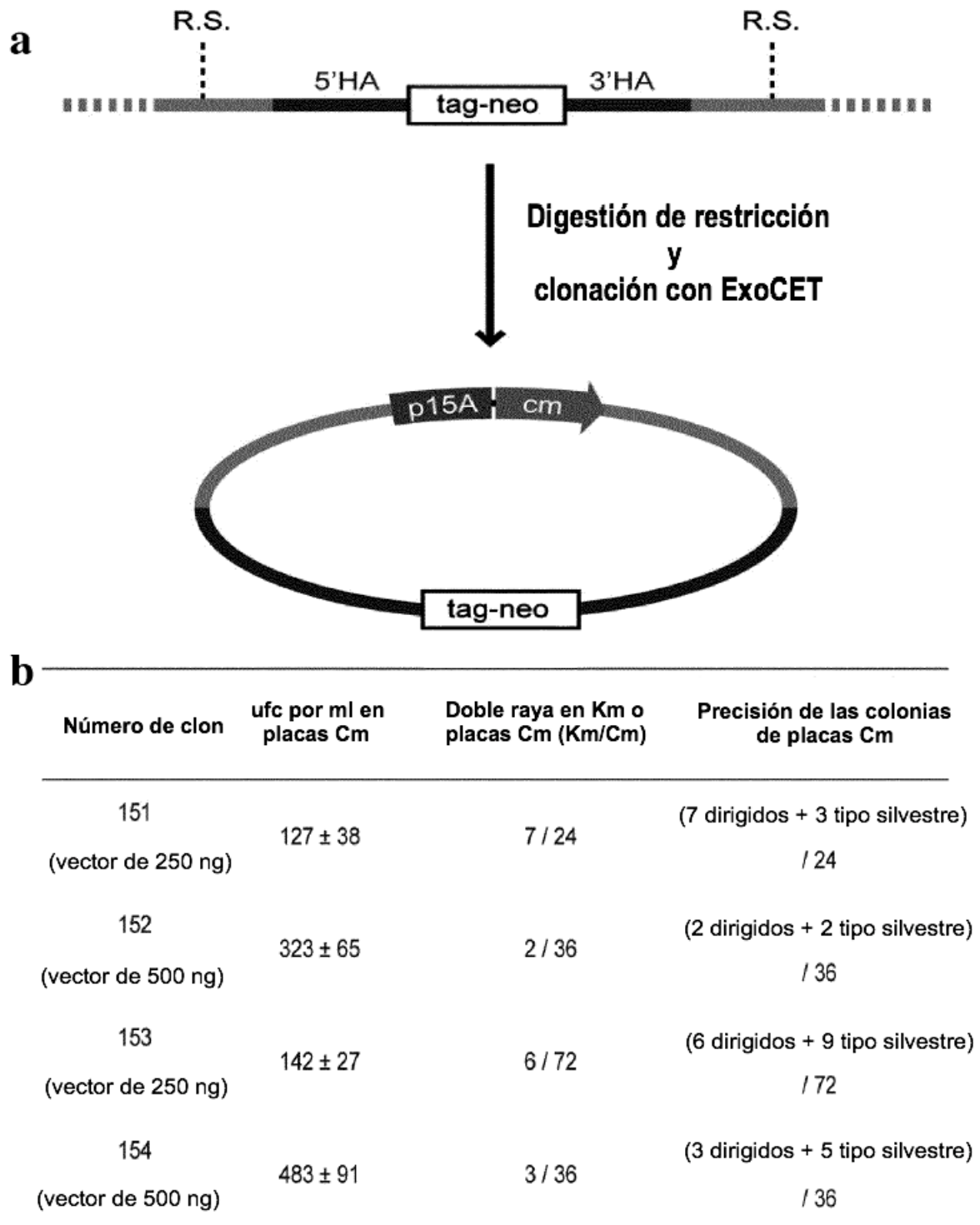


FIG. 10

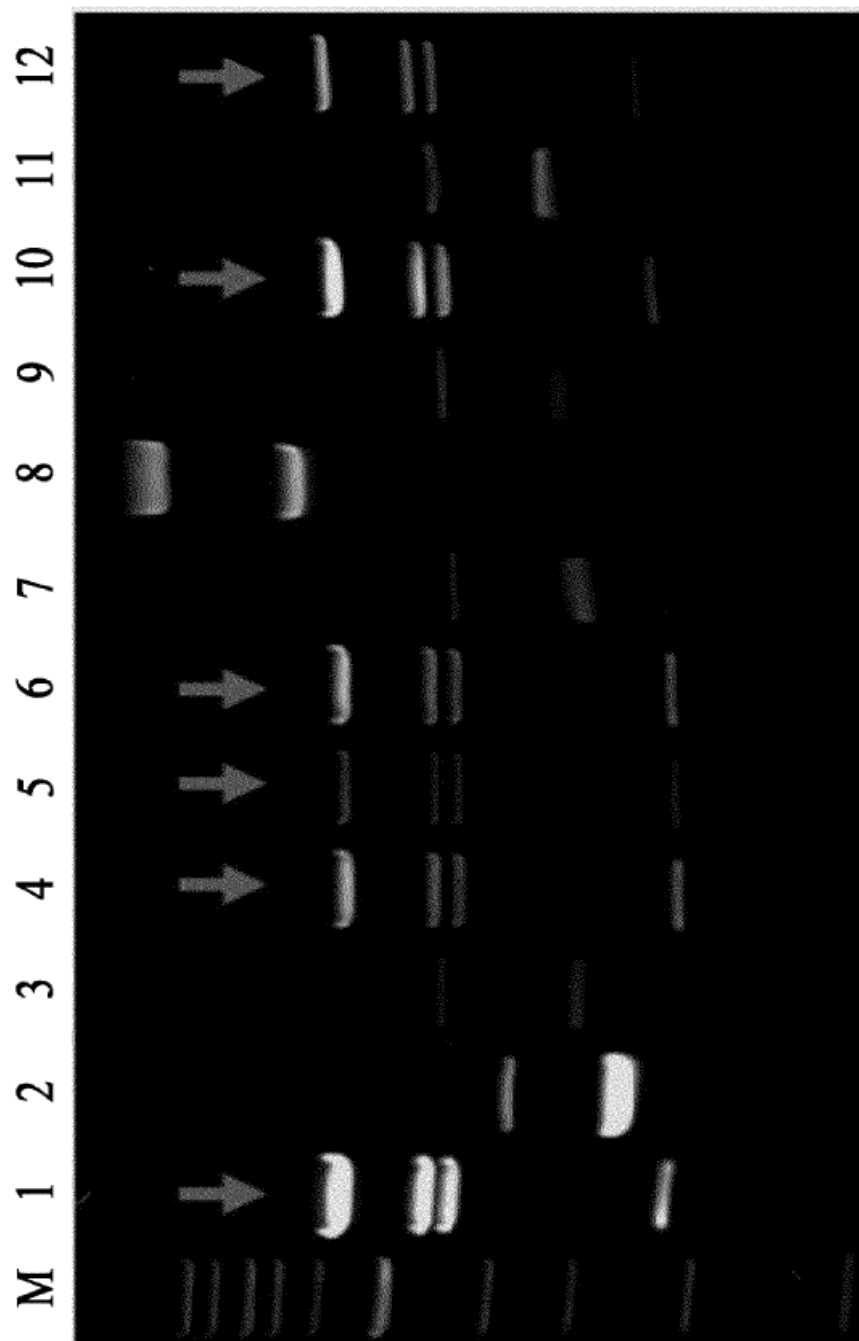


FIG. 11