

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 891 087**

51 Int. Cl.:

**C12N 9/12** (2006.01)

**C12N 15/52** (2006.01)

**C12N 15/62** (2006.01)

**C12N 15/90** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.10.2016 PCT/US2016/055824**

87 Fecha y número de publicación internacional: **13.04.2017 WO17062668**

96 Fecha de presentación y número de la solicitud europea: **06.10.2016 E 16854354 (4)**

97 Fecha y número de publicación de la concesión europea: **18.08.2021 EP 3359671**

54 Título: **Vectores de ADN, trasposones y transposasas para la modificación del genoma eucariota**

30 Prioridad:

**08.10.2015 US 201562239109 P**

**21.04.2016 US 201662325872 P**

**11.08.2016 US 201662373422 P**

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

**26.01.2022**

73 Titular/es:

**DNA TWOPOINTO INC. (100.0%)**

**37950 Central Court**

**Newark, CA 94560, US**

72 Inventor/es:

**MINSHULL, JEREMY;**

**WELCH, MARK;**

**GOVINDRAJAN, SRIDHAR;**

**LEE, MAGGIE;**

**CAVES, KATE y**

**NESS, JON**

74 Agente/Representante:

**SÁNCHEZ SILVA, Jesús Eladio**

**Observaciones:**

**Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes**

ES 2 891 087 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

Vectores de ADN, trasposones y transposasas para la modificación del genoma eucariota

## 5 2. Antecedentes de la invención

La eficacia con la que un primer polinucleótido puede efectuar la integración de ADN heterólogo en el genoma de una célula diana depende de la configuración de los elementos de secuencia dentro del polinucleótido. Los niveles de expresión de genes codificados por el ADN heterólogo integrado también dependen de la configuración de los elementos de secuencia dentro del ADN heterólogo integrado. La eficiencia de la integración, el tamaño de la secuencia de ADN heteróloga que se puede integrar, el número de copias de la secuencia de ADN heteróloga que se integran en cada genoma y el tipo de loci genómico donde ocurre la integración a menudo se pueden mejorar aún más al colocar el ADN heterólogo en un transposón.

Los transposones comprenden dos extremos que son reconocidos por una transposasa. La transposasa actúa sobre el transposón removiéndolo de una molécula de ADN e integrándolo en otra. El ADN entre los dos extremos del transposón se transpone por la transposasa junto con los extremos del transposón. El ADN heterólogo flanqueado por un par de extremos del transposón, de modo que sea reconocido y transpuesto por una transposasa, se le denomina en la presente descripción transposón sintético. La introducción de un transposón sintético y una transposasa correspondiente en el núcleo de una célula eucariota puede dar como resultado la transposición del transposón al genoma de la célula. Los transposones y transposasas más activos (hiperactivos) dan como resultado una mayor frecuencia de transposición, lo que conduce a una mayor fracción de células cuyos genomas contienen una copia integrada del transposón y/o células cuyos genomas contienen un mayor número de copias integradas del transposón. Estos resultados son útiles porque aumentan la eficiencia de transformación y porque pueden aumentar los niveles de expresión del ADN heterólogo integrado. Por tanto, existe una necesidad en la técnica de transposasas y transposones hiperactivos.

La transposición mediante una transposasa similar a piggyBac es perfectamente reversible. El transposón se integra inicialmente en una secuencia diana de integración en una molécula de ADN receptora, durante la cual la secuencia diana se duplica en cada extremo de las repeticiones terminales invertidas del transposón (ITR). La transposición posterior elimina el transposón y restaura el ADN del receptor a su secuencia anterior, con la duplicación de la secuencia diana y el transposón eliminado. Sin embargo, esto no es suficiente para eliminar un transposón de un genoma en donde se ha integrado, ya que es muy probable que el transposón se elimine de la primera secuencia diana de integración pero se integre en una segunda secuencia diana de integración en el genoma. Las transposasas que son deficientes para la función de integración, por otro lado, pueden escindir el transposón de la primera secuencia diana, pero no podrán integrarse en una segunda secuencia diana. Las transposasas de integración deficiente son, por tanto, útiles para revertir la integración genómica de un transposón.

## 40 3. Resumen de la invención

La expresión génica heteróloga a partir de construcciones de polinucleótidos que se integran de manera estable en el genoma de una célula diana se puede mejorar colocando el polinucleótido de expresión entre un par de extremos de un transposón: elementos de secuencia que son reconocidos y transpuestos por transposasas. Las secuencias de ADN insertadas entre un par de extremos de transposones pueden escindirse mediante una transposasa de una molécula de ADN e insertarse en una segunda molécula de ADN. Se describen dos nuevos sistemas de transposón-transposasa de similar a piggyBac que no se derivan de la oruga de la col *Trichoplusia ni*; uno se deriva del gusano de seda *Bombyx mori* y el otro se deriva de la rana *Xenopus tropicalis*. Cada uno de estos comprende secuencias que funcionan como extremos de transposones y que pueden usarse junto con una transposasa correspondiente que reconoce y actúa sobre esos extremos de transposones, como sistemas de transferencia de genes para introducir de manera estable ácidos nucleicos en el ADN de una célula. También se describen variantes de transposasa hiperactivas y deficientes en integración.

La invención proporciona, como se especifica en las reivindicaciones, una transposasa que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO: 407 o 415, en la que la transposasa no tiene una secuencia de origen natural, que comprende al menos una sustitución de aminoácido, con respecto a la SEQ ID NO: 407, como se muestra en la Tabla 4, columna G o H o en la Tabla 11, columna H. La invención también proporciona, como se especifica en las reivindicaciones, un polinucleótido que codifica una transposasa como se especifica en las reivindicaciones. La invención también proporciona un método para crear una célula transgénica que comprende (i) introducir en una célula eucariota (a) la transposasa de la reivindicación 1, en donde la transposasa i) se proporciona como una proteína, que se introduce en la célula, o ii) proporcionado como un polinucleótido, opcionalmente ARNm, que codifica la transposasa, que se introduce en la célula y expresa la transposasa, y (b) un polinucleótido que comprende un transposón que comprende extremos del transposón de un transposón similar a piggyBac que flanquea un polinucleótido heterólogo, en donde un extremo del transposón comprende al menos 16 bases contiguas de SEQ ID NO: 25 y el otro extremo del transposón comprende al menos 16 bases contiguas de SEQ ID NO: 31 e (ii) identificar una célula en la que el transposón está incorporado en el genoma de la célula eucariota; en donde el método no implica modificar la identidad genética de la línea germinal de ningún ser humano. La invención

también proporciona la transposasa de la reivindicación 1 para su uso en terapia génica, en donde no se modifica la identidad genética de la línea germinal de ningún ser humano. El tema descrito en la presente descripción pero no cubierto por las reivindicaciones puede ser útil para comprender la invención. La especificación proporciona secuencias de *transposasas hiperactivas de Xenopus* que son al menos un 90 % idénticas a la SEQ ID NO: 61, y posiciones y sustituciones de aminoácidos que pueden introducirse para mejorar la actividad de la transposasa o para mantener la función de la transposasa. La especificación también proporciona secuencias de extremos del transposón que comprenden al menos 16 bases contiguas de SEQ ID NO: 7 y al menos 16 bases contiguas de SEQ ID NO: 16, y las repeticiones terminales invertidas de SEQ ID NO: 19. Estas secuencias, cuando se colocan a ambos lados de un polinucleótido heterólogo, crean un *transposón de Xenopus* sintético que se puede escindir de un polinucleótido mediante transposasas de *Xenopus*. El transposón sintético puede integrarse en un genoma diana mediante una transposasa de *Xenopus*.

La invención proporciona secuencias de *transposasas de Bombyx* hiperactivas que son al menos un 90 % idénticas a SEQ ID NO: 415, y posiciones y sustituciones de aminoácidos que pueden introducirse para mejorar la actividad de la transposasa o para mantener la función de la transposasa. La especificación también proporciona secuencias de extremos del transposón que comprenden al menos 16 bases contiguas de SEQ ID NO: 25 y al menos 16 bases contiguas de SEQ ID NO: 31, y repeticiones terminales invertidas que son al menos 87 % idénticas a SEQ ID NO: 33. Estas secuencias, cuando se colocan a ambos lados de un polinucleótido heterólogo, crean un *transposón de Bombyx* sintético que puede escindirse de un polinucleótido mediante transposasas *Bombyx*. El transposón sintético puede integrarse en un genoma diana mediante una transposasa de *Bombyx*.

La especificación proporciona métodos para integrar un polinucleótido heterólogo en el genoma de una célula diana, mediante la introducción de un *transposón Xenopus* y una *transposasa Xenopus*, o una *transposasa de Bombyx* y un *transposón de Bombyx*, en una célula diana. La transposasa puede introducirse como proteína o como polinucleótido que codifica la transposasa y puede expresarse en la célula diana.

La especificación también proporciona configuraciones de vector, incluidas configuraciones de transposón, que son particularmente ventajosas para la expresión de genes en sistemas de mamíferos.

Los transposones y transposasas de la presente invención pueden usarse en métodos, por ejemplo, pero sin limitarse a, expresión génica heteróloga, terapia génica, mutagénesis por inserción o descubrimiento de genes.

#### 4. Breve descripción de las composiciones de las construcciones y procedimientos experimentales

##### 4.1 Composiciones de las construcciones

La SEQ ID NO: 39 contiene un promotor débil (el promotor de fosfoglicerato cinasa murina (PGK), SEQ ID NO: 937), unido operativamente a un único marco de lectura abierto que codifica DasherGFP acoplado traduccionalmente a través de una secuencia CHYSEL a puromicina N-acetil transferasa, seguido por la señal de poliadenilación de la beta globina humana.

SEQ ID NO: 40 comprende un promotor débil (el promotor de fosfoglicerato cinasa murina (PGK), SEQ ID NO: 937), unido operativamente a un marco de lectura abierto que codifica puromicina N-acetil transferasa, seguido de la señal de poliadenilación de beta globina humana. La SEQ ID NO: 40 también comprende el promotor EF1a unido operativamente a un gen que codifica DasherGFP seguido de los elementos potenciadores de la expresión SEQ ID NO. 866 y la secuencia de poliadenilación de globina de conejo.

##### 4.2 Procedimientos experimentales

###### 4.2.1 Transfección y selección de CHO-K1

Células CHO-K1 (de ATCC) se cultivaron en F12-K (de ATCC) + 10 % de FBS (de ATCC) + 1 % de penicilina-estreptomomicina (de ATCC) a 37 °C, 5 % de CO<sub>2</sub> al 80 % de confluencia. Se sembraron 500 000 células en placas de cultivo de tejidos de 24 pocillos y se incubaron a 37 °C, 5 % de CO<sub>2</sub> durante 24 horas antes de la transfección. Las transfecciones se realizaron por triplicado. Cada transfección usó un total de 500-1000 µg de ADN con el reactivo Roche Extreme Gene 9 (relación 2:1) según el protocolo del fabricante. Se añadió medio con puromicina 50 µg/ml 72 horas después de la transfección. La selección con puromicina se llevó a cabo durante 72 horas, después de lo cual se eliminó la puromicina. Las células se cultivaron durante 14 días después de la selección con puromicina con dos pases y cambios de medio. Las células se recogieron raspando y midiendo en un lector de placas fluorimétrico.

###### 4.2.2 Transfección y selección de CHO-S

Células CHO-S (de la ATCC) se cultivaron en medio de expresión CHOgro (de Mirus) a 37 °C, 5 % de CO<sub>2</sub> y se sembraron a 2 x 10<sup>6</sup> células/ml. Se transfectó 1 ml de células con 1 µg de ácido nucleico total. Las transfecciones se realizaron por duplicado. Cada transfección usó el reactivo Mirus Transit-Pro y Mirus TransIT-mRNA según el

protocolo del fabricante. Se añadió medio con puromicina 72 horas después de la transfección. La selección con puromicina se llevó a cabo durante el número de días indicado, con un cambio de medio completo en medio fresco que contenía puromicina después de 5 días.

#### 5 4.2.3 Preparación de ARNm

El ARNm que codifica las transposasas se preparó mediante *transcripción in vitro* mediante el uso de ARN polimerasa de T7. El ARNm comprendía una secuencia 5' SEQ ID NO: 699 que precede a la secuencia que codifica el marco de lectura abierto, y una secuencia 3' SEQ ID NO: 700 que sigue al codón de terminación al final del marco de lectura abierto. El ARNm tenía un análogo de caperuza anti-inversa (3'-O-Me-m<sup>7</sup> G(5')ppp(5')G, y estaba completamente sustituido con pseudouridina y 5-metil-citosina.

### 5. Descripción detallada de la invención

#### 15 5.1 Definiciones

El uso de las formas singulares "un", "una" y "el" incluye referencias en plural a menos que el contexto lo indique claramente de cualquier otra manera. Así, por ejemplo, la referencia a "un polinucleótido" incluye una pluralidad de polinucleótidos, la referencia a "un sustrato" incluye una pluralidad de tales sustratos, la referencia a "una variante" incluye una pluralidad de variantes y similares.

Términos tales como "conectado", "adjunto", "enlazado" y "conjugado" se usan indistintamente en la presente descripción y abarcan conexión, unión, enlace o conjugación directa o indirecta, a menos que el contexto lo indique claramente de cualquier otra manera. Cuando se indica un intervalo de valores, debe entenderse que cada valor entero intermedio, y cada fracción del mismo, entre los límites superior e inferior enumerados de ese intervalo también se describe específicamente, junto con cada subintervalo entre dichos valores. Los límites superior e inferior de cualquier intervalo se pueden incluir o excluir de forma independiente del intervalo, y cada intervalo en el que se incluyen uno o ambos límites o ambos límites también se incluye en la invención. Cuando un valor que se está discutiendo tiene límites inherentes, por ejemplo, cuando un componente puede estar presente en una concentración de 0 a 100 %, o cuando el pH de una solución acuosa puede variar de 1 a 14 esos límites inherentes se describen específicamente. Cuando se indica explícitamente un valor, debe entenderse que los valores que son aproximadamente la misma cantidad o cantidad que el valor indicado también están dentro del alcance de la descripción. Cuando se describe una combinación, cada subcombinación de los elementos de esa combinación también se describe específicamente y está dentro del alcance de la descripción. A la inversa, cuando se describen individualmente diferentes elementos o grupos de elementos, también se describen combinaciones de los mismos. Cuando se describe que cualquier elemento de una invención tiene una pluralidad de alternativas, también se describen ejemplos de esa invención en donde cada alternativa se excluye individualmente o en cualquier combinación con las otras alternativas; más de un elemento de una invención puede tener tales exclusiones, y por la presente se describen todas las combinaciones de elementos que tienen tales exclusiones.

A menos que se defina de cualquier otra manera en la presente descripción, todos los términos técnicos y científicos usados en la presente descripción tienen el mismo significado que entiende habitualmente un experto en la técnica a la que pertenece esta invención. Singleton, y otros, Dictionary of Microbiology and Molecular Biology, 2<sup>da</sup> Edición, John Wiley and Sons, Nueva York (1994), y Hale & Marham, The Harper Collins Dictionary of Biology, Harper Perennial, NY, 1991, proporcionan a un experto un diccionario general de muchos de los términos que se usan en esta descripción. Se describen métodos y materiales preferidos en relación con la práctica o ensayo de la invención. A menos que se indique de cualquier otra manera, los ácidos nucleicos se escriben de izquierda a derecha en orientación 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en orientación amino a carboxilo, respectivamente. Los términos definidos inmediatamente a continuación se definen de forma más completa por referencia a la especificación en su conjunto.

La "configuración" de un polinucleótido significa los elementos de secuencia funcional dentro del polinucleótido y el orden y dirección de esos elementos.

Los términos "transposón correspondiente" y "transposasa correspondiente" se usan para indicar una relación de actividad entre una transposasa y un transposón. Una transposasa transpone su transposón correspondiente. Muchas transposasas pueden corresponder con un solo transposón, por ejemplo, todas las SEQ ID NO: 52-402 son transposasas correspondientes para el transposón SEQ ID NO: 44). Un transposón es transpuesto por su correspondiente transposasa. Muchos transposones pueden corresponder con una sola transposasa, por ejemplo, los transposones que se muestran en las filas 4-21 de la Tabla 5 son todos transposones correspondientes para la transposasa SEQ ID NO: 48.

El término "marcador de selección inversa" significa una secuencia de polinucleótidos que confiere una desventaja selectiva a una célula huésped. Ejemplos de marcadores de selección inversa incluyen *sacB*, *rpsL*, *tetAR*, *pheS*, *thyA*, *gata-1*, *ccdB*, *kid* y *barnasa* (Bernard, 1995, Journal/Gene, 162:159-160; Bernard y otros, 1994, Journal/Gene, 148: 71-74; Gababt y otros, 1997, Journal/Biotechniques, 23: 938-941; Gababt y otros, 1998, Journal/Gene, 207: 87-

92; Gababt y otros, 2000, Journal/Biotechniques, 28: 784-788; Galvao y de Lorenzo, 2005, Journal/Appl Environ Microbiol, 71: 883-892; Hartzog y otros, 2005, Journal/yeat, 22: 789-798; Knipfer y otros, 1997, Journal/Plasmid, 37: 129-140; Reytrat y otros, 1998, Journal/Infect Immun, 66: 4011-4017; Soderholm y otros, 2001, Journal/Biotechniques, 31: 306-310, 312; Tamura y otros, 2005, Journal/Appl Environ Microbiol, 71: 587-590; Yazynin y otros, 1999, Journal/FEBS Lett, 452: 351-354). Los marcadores de selección inversa a menudo confieren su desventaja selectiva en contextos específicos. Por ejemplo, pueden conferir sensibilidad a compuestos que se pueden agregar al entorno de la célula huésped, o pueden matar a un huésped con un genotipo pero no a un huésped con un genotipo diferente. Las condiciones que no confieren una desventaja selectiva a una célula que lleva un marcador de selección se describen como "permissivas". Las condiciones que confieren una desventaja selectiva a una célula que lleva un marcador de selección inversa se describen como "restrictivas".

El término "elemento de acoplamiento" o "elemento de acoplamiento traduccional" se refiere a una secuencia de ADN que permite asociar la expresión de un primer polipéptido a la expresión de un segundo polipéptido. Los sitios internos de entrada a ribosomas (elementos IRES) y los elementos hidrolasa que actúan en cis (elementos CHYSEL) son ejemplos de elementos de acoplamiento.

Los términos "secuencia de ADN", "secuencia de ARN" o "secuencia de polinucleótidos" significan una secuencia de ácido nucleico contigua. La secuencia puede ser un oligonucleótido de 2 a 20 nucleótidos de longitud hasta una secuencia genómica de longitud completa de miles o cientos de miles de pares de bases.

El término "construcción de expresión" significa cualquier polinucleótido diseñado para transcribir un ARN. Por ejemplo, una construcción que contiene al menos un promotor que está o puede estar unido operativamente a un gen, región codificante o secuencia polinucleotídica corriente abajo (por ejemplo, un ADNc o fragmento de ADN genómico que codifica un polipéptido o proteína, o una molécula efectora de ARN, por ejemplo, un ARN antisentido, ARN formador de tríplex, ribozima, un ligando de ARN de alta afinidad seleccionado artificialmente (aptámero), un ARN bicatenario, por ejemplo, una molécula de ARN que comprende un ARNbc de tallo-bucle u horquilla, o un ARNbc de dos dedos o de varios dedos o un microARN, o cualquier ARN). Un "vector de expresión" es un polinucleótido que comprende un promotor que puede unirse operativamente a un segundo polinucleótido. La transfección o transformación de la construcción de expresión en una célula receptora permite que la célula exprese una molécula, polipéptido o proteína efectora de ARN codificada por la construcción de expresión. Una construcción de expresión puede ser un plásmido, virus, virus recombinante modificado genéticamente o un cromosoma artificial derivado de, por ejemplo, un bacteriófago, adenovirus, virus adenoasociado, retrovirus, lentivirus, poxvirus o herpesvirus. Dichos vectores de expresión pueden incluir secuencias de bacterias, virus o fagos. Dichos vectores incluyen vectores cromosómicos, episómicos y derivados de virus, por ejemplo, vectores derivados de plásmidos bacterianos, bacteriófagos, episomas de levadura, elementos cromosómicos de levadura y virus, vectores derivados de combinaciones de los mismos, tales como los derivados de elementos genéticos de plásmidos y bacteriófagos, cósmidos y fagémidos. Una construcción de expresión se puede replicar en una célula viva o se puede preparar sintéticamente. Para los propósitos de esta solicitud, los términos "construcción de expresión", "vector de expresión", "vector" y "plásmido" se usan indistintamente para demostrar la aplicación de la invención en un sentido ilustrativo general, y no pretenden limitarse a un tipo particular de construcción de expresión.

El término "polipéptido de expresión" significa un polipéptido codificado por un gen en una construcción de expresión.

El término "sistema de expresión" se refiere a cualquier sistema biológico *in vivo* o *in vitro* que se usa para producir uno o más productos génicos codificados por un polinucleótido.

Un "sistema de transferencia de genes" comprende un vector o vector de transferencia de genes, o un polinucleótido que comprende el gen a transferir que se clona en un vector (un "polinucleótido de transferencia de genes" o "construcción de transferencia de genes"). Un sistema de transferencia de genes también puede comprender otras características para facilitar el proceso de transferencia de genes. Por ejemplo, un sistema de transferencia de genes puede comprender un vector y una mezcla de empaquetamiento de lípidos o virus para permitir que un primer polinucleótido entre en una célula, o puede comprender un polinucleótido que incluye un transposón y una segunda secuencia de polinucleótidos que codifica una transposasa correspondiente para mejorar la integración productiva genómica del transposón. Las transposasas y transposones de un sistema de transferencia de genes pueden estar en la misma molécula de ácido nucleico o en diferentes moléculas de ácido nucleico. La transposasa de un sistema de transferencia de genes puede proporcionarse como un polinucleótido o como un polipéptido.

Dos elementos son heterólogos entre sí si no están asociados naturalmente. Por ejemplo, una secuencia de ácido nucleico que codifica una proteína unida a un promotor heterólogo significa un promotor distinto del que dirige naturalmente la expresión de la proteína. Un ácido nucleico heterólogo flanqueado por extremos del transposón o ITR significa un ácido nucleico heterólogo que no está flanqueado naturalmente por esos extremos del transposón o ITR, tal como un ácido nucleico que codifica un polipéptido distinto de una transposasa, que incluyen una cadena ligera o pesada de anticuerpo. Un ácido nucleico es heterólogo para una célula si no se encuentra naturalmente en la célula o si se encuentra naturalmente en la célula pero en una ubicación diferente (por ejemplo, una ubicación episómica o genómica diferente) que la ubicación descrita.

5 El término "huésped" significa cualquier organismo procariota o eucariota que puede ser receptor de un ácido nucleico. Un "huésped", como se usa el término en la presente descripción, incluye organismos procariotas o eucariotas que pueden modificarse genéticamente. Para ejemplos de tales huéspedes, véase Maniatis y otros, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982). Como se usa en la presente descripción, los términos "huésped", "célula huésped", "sistema huésped" y "huésped de expresión" se pueden usar indistintamente.

10 Una transposasa "hiperactiva" es una transposasa que es más activa que la transposasa de origen natural de la que se deriva. Por tanto, las transposasas "hiperactivas" no son secuencias de origen natural. Las transposasas hiperactivas de *Xenopus* son aquellas que son más activas que la SEQ ID NO: 48. Las transposasas de *Bombyx* hiperactivas son aquellas que son más activas que la SEQ ID NO: 407.

15 "Integración defectuosa" significa una transposasa que puede escindir su transposón correspondiente, pero que integra el transposón escindido en el genoma del huésped a una frecuencia más baja que una transposasa de origen natural correspondiente. Las transposasas de *Xenopus* con integración defectuosa son deficientes con respecto a la SEQ ID NO: 48. Las transposasas de *Bombyx* con integración defectuosa son deficientes con respecto a la SEQ ID NO: 407.

20 Un "IRES" o "sitio interno de entrada del ribosoma" significa una secuencia especializada que promueve directamente la unión de ribosoma, independientemente de la estructura de la caperuza.

25 Un polipéptido o polinucleótido "aislado" significa un polipéptido o polinucleótido que ha sido eliminado de su entorno natural, producido mediante el uso de técnicas recombinantes o sintetizado química o enzimáticamente. Los polipéptidos o polinucleótidos de esta invención (como se especifica en las reivindicaciones) pueden purificarse, es decir, esencialmente libres de cualquier otro polipéptido o polinucleótido y productos celulares asociados u otras impurezas.

30 Los términos "nucleósido" y "nucleótido" incluyen aquellas porciones que contienen no solo las bases de purina y pirimidina conocidas, sino también otras bases heterocíclicas que se han modificado. Dichas modificaciones incluyen purinas o pirimidinas metiladas, purinas o pirimidinas aciladas u otros heterociclos. Los nucleósidos o nucleótidos modificados también pueden incluir modificaciones en la porción azúcar, por ejemplo, cuando uno o más de los grupos hidroxilo se reemplazan con halógeno, grupos alifáticos o se funcionalizan como éteres, aminas o similares. El término "unidad nucleotídica" pretende abarcar nucleósidos y nucleótidos.

35 "Marco abierto de lectura" u ORF significa una porción de un polinucleótido que, cuando se traduce en aminoácidos, no contiene codones de terminación. El código genético lee las secuencias de ADN en grupos de tres pares de bases, lo que significa que una molécula de ADN de doble cadena puede leerse en cualquiera de los seis marcos de lectura posibles: tres en la dirección directa y tres en la reversa. Normalmente, un ORF también incluye un codón de iniciación en el que puede comenzar la traducción.

40 El término "unido operativamente" se refiere al enlace funcional entre dos secuencias de manera que una secuencia modifica el comportamiento de la otra. Por ejemplo, un primer polinucleótido que comprende una secuencia de control de la expresión del ácido nucleico (tal como un promotor, secuencia IRES, potenciador o arreglo de sitios de unión a factor de transcripción) y un segundo polinucleótido están unidos operativamente si el primer polinucleótido afecta la transcripción y/o traducción del segundo polinucleótido. De manera similar, una primera secuencia de aminoácidos que comprende una señal de secreción o una señal de localización subcelular y una segunda secuencia de aminoácidos están unidas operativamente si la primera secuencia de aminoácidos hace que la segunda secuencia de aminoácidos sea secretada o localizada en una ubicación subcelular.

50 El término "saliente" o "ADN saliente" se refiere a la porción monocatenaria en el extremo de una molécula de ADN bicatenaria. Los salientes complementarios son aquellos que se emparejarán entre sí.

55 Una "transposasa similar a piggyBac" significa una transposasa con al menos un 20 % de identidad de secuencia identificada mediante el algoritmo TBLASTN con la transposasa piggyBac de *Trichoplusia ni* (SEQ ID NO: 698), y como se describe con más detalle en Sakar, A. y otros, (2003). Mol. Gen. Genomics 270: 173-180. "Molecular evolutionary analysis of the widespread piggyBac transposon family and related 'domesticated' species", y además se caracteriza por un motivo DDD similar a DDE, con residuos de aspartato en las posiciones correspondientes a D268, D346 y D447 de la transposasa de *Trichoplusia ni* piggyBac en alineamiento máxima. Las transposasas similares a PiggyBac también se caracterizan por su capacidad para escindir sus transposones con precisión con una alta frecuencia. Un "transposón similar a piggyBac" significa un transposón que tiene extremos del transposón que son iguales o al menos 80 % y preferentemente al menos 90, 95, 96, 97, 98 o 99 % idénticos a los extremos del transposón de un transposón de origen natural que codifica una transposasa similar a piggyBac. Un transposón similar a piggyBac incluye una secuencia de repetición terminal invertida (ITR) de aproximadamente 12-16 bases en cada extremo, y está flanqueado en cada lado por una secuencia de 4 bases correspondiente a la secuencia objetivo de integración que se duplica en la integración del transposón (el objetivo Duplicación de sitio o duplicación

de secuencia de destino o TSD). Los transposones y transposasas similares a PiggyBac se encuentran de forma natural en una amplia gama de organismos, incluidos *Argyrogramma agnate* (GU477713), *Anopheles gambiae* (XP\_312615; XP\_320414; XP\_310729), *Aphis gossypii* (GU329918), *Acyrtosiphon pisum* (XP\_001948139), *Agrotis ypsilon* (GU477714), *Bombyx mori* (BAD11135), *Ciona intestinalis* (XP\_002123602), *Chilo suppressalis* (JX294476), *Drosophila melanogaster* (AAL39784), *Daphnia pulex* (AAM76342), *Helicoverpa armigera* (ABS18391), *Homo sapiens* (NP\_689808), *Heliothis virescens* (ABD76335), *Macdunnoughia crassisigna* (EU287451), *Macaca fascicularis* (AB179012), *Mus musculus* (NP\_741958), *Pectinophora gossypiella* (GU270322), *Rattus norvegicus* (XP\_220453), *Tribolium castaneum* (XP\_001814566), *Trichoplusia ni* (AAA87375) y *Xenopus tropicalis* (BAF82026), aunque la actividad de transposición no se ha descrito para casi ninguno de estos.

Los términos "polinucleótido", "oligonucleótido", "ácido nucleico" y "molécula de ácido nucleico" y "gen" se usan indistintamente para referirse a una forma polimérica de nucleótidos de cualquier longitud, y pueden comprender ribonucleótidos, desoxirribonucleótidos, análogos de los mismos o mezclas de los mismos. Este término se refiere solo a la estructura primaria de la molécula. Por tanto, el término incluye ácido desoxirribonucleico ("ADN") de cadena triple, doble y simple, así como ácido ribonucleico ("ARN") de cadena triple, doble y simple. También incluye formas modificadas, por ejemplo, por alquilación y/o por protección, y no modificadas del polinucleótido. Más particularmente, los términos "polinucleótido", "oligonucleótido", "ácido nucleico" y "molécula de ácido nucleico" incluyen polidesoxirribonucleótidos (que contienen 2-desoxi-D-ribosa), polirribonucleótidos (que contienen D-ribosa), que incluyen ARNt, ARNr, ARNh, ARNip y ARNm, ya sea empalmado o no empalmado, cualquier otro tipo de polinucleótido que sea un N- o C-glucósido de una base purina o pirimidina, y otros polímeros que contengan cadenas principales no nucleotídicas, por ejemplo, poliamida (por ejemplo, ácidos nucleicos peptídicos ("PNA")) y polimorfolino (disponible comercialmente de Anti-Virals, Inc., Corvallis, Oreg., como Neugene) polímeros, y otros polímeros de ácidos nucleicos específicos de secuencia sintética siempre que los polímeros contengan bases nucleotídicas en una configuración que permita el apareamiento de bases y apilamiento de bases, como se encuentra en el ADN y el ARN. No existe una distinción intencionada en la longitud entre los términos "polinucleótido", "oligonucleótido", "ácido nucleico" y "molécula de ácido nucleico", y estos términos se usan indistintamente en la presente descripción. Estos términos se refieren únicamente a la estructura primaria de la molécula. Así, estos términos incluyen, por ejemplo, 3'-desoxi-2', 5'-ADN, oligodesoxirribonucleótidos N3' P5' fosforamidatos, ARN 2'-O-alkil-sustituido, ADN bicatenario y monocatenario, así como ARN bicatenario y monocatenario e híbridos de los mismos, que incluyen, por ejemplo, híbridos entre ADN y ARN o entre PNA y ADN o ARN, y también incluyen tipos conocidos de modificaciones, por ejemplo, etiquetas, alquilación, "caperuzas", sustitución de uno o más de los nucleótidos con un análogo, modificaciones internucleotídicas como, por ejemplo, aquellos con enlaces no cargados (por ejemplo, metilfosfonatos, fosfotriésteres, fosforamidatos, carbamatos o similares) con enlaces cargados negativamente (por ejemplo, fosforotioatos, fosforoditioatos o similares), y con enlaces cargados positivamente (por ejemplo, aminoalquilfosforamidatos, aminoalquilfosfotriésteres), aquellos que contienen porciones salientes, tales como, por ejemplo, proteínas (que incluyen enzimas (por ejemplo, nucleasas), toxinas, anticuerpos, péptidos señal, poli-L-lisina o similares), aquellos con intercaladores (por ejemplo, acridina, psoraleno o similares), aquellos que contienen quelatos (de, por ejemplo, metales, metales radiactivos, boro, metales oxidativos o similares), aquellos que contienen alquilantes, aquellos con enlaces modificados (por ejemplo, ácidos nucleicos alfa anoméricos o similares), así como formas no modificadas del polinucleótido u oligonucleótido.

Un "promotor" significa una secuencia de ácido nucleico suficiente para dirigir la transcripción de una molécula de ácido nucleico unida operativamente. También se incluyen en esta definición aquellos elementos de control de la transcripción (por ejemplo, potenciadores) que son suficientes para hacer controlable la expresión génica dependiente del promotor de una manera específica de tipo celular, específica de tejido o específica temporal, o que son inducibles por señales o agentes; tales elementos, pueden estar dentro de la región 3' de un gen o dentro de un intrón. Convenientemente, un promotor está unido operativamente a una secuencia de ácido nucleico, por ejemplo, un ADNc o una secuencia génica, o una secuencia codificante de ARN efectora, de tal manera que permita la expresión de la secuencia de ácido nucleico, o se proporciona un promotor en un casete de expresión en donde se puede insertar convenientemente una secuencia de ácido nucleico seleccionada a transcribir.

El término "marcador de selección" significa un segmento de polinucleótido que permite seleccionar a favor o en contra una molécula o una célula que lo contiene, a menudo en condiciones particulares. Estos marcadores pueden codificar una actividad, tal como, pero sin limitarse a, la producción de ARN, péptido o proteína, o pueden proporcionar un sitio de unión para ARN, péptidos, proteínas, compuestos o composiciones inorgánicas y orgánicas. Los ejemplos de marcadores de selección incluyen, pero no se limitan a: (1) segmentos de ADN que codifican productos que proporcionan resistencia contra compuestos de otra manera tóxicos (por ejemplo, antibióticos); (2) segmentos de ADN que codifican productos que de otro modo faltan en la célula receptora (por ejemplo, genes de ARNt, marcadores auxotróficos); (3) segmentos de ADN que codifican productos que suprimen la actividad de un producto génico; (4) segmentos de ADN que codifican productos que pueden identificarse fácilmente (por ejemplo, marcadores fenotípicos tales como beta-galactosidasa, proteína verde fluorescente (GFP) y proteínas de la superficie celular); (5) segmentos de ADN que se unen a productos que de otro modo serían perjudiciales para la supervivencia y/o función celular; (6) segmentos de ADN que inhiben de otro modo la actividad de cualquiera de los segmentos de ADN descritos en los números 1-5 anteriores (por ejemplo, oligonucleótidos antisentido); (7) segmentos de ADN que se unen a productos que modifican un sustrato (por ejemplo, endonucleasas de restricción); (8) segmentos de ADN que se pueden usar para aislar una molécula deseada (por ejemplo, sitios de unión a

proteínas específicas); (9) segmentos de ADN que codifican una secuencia de nucleótidos específica que de otro modo puede ser no funcional (por ejemplo, para la amplificación por PCR de subpoblaciones de moléculas); y/o (10) segmentos de ADN, que cuando están ausentes, confieren directa o indirectamente sensibilidad a compuestos particulares.

5 La identidad de la secuencia se puede determinar alineando secuencias mediante el uso de algoritmos, como BESTFIT, FASTA y TFASTA en Wisconsin Genetics Software Package versión 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.), mediante el uso de parámetros predeterminados para las brechas, o por inspección y el mejor alineamiento (es decir, que da como resultado el mayor porcentaje de similitud de secuencia en una ventana de comparación). El porcentaje de identidad de secuencia se calcula comparando dos secuencias alineadas de manera óptima en una ventana de comparación, determinando el número de posiciones en las que ocurren los residuos idénticos en ambas secuencias para producir el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones coincidentes y no coincidentes sin contar las brechas en la ventana de comparación (es decir, el tamaño de la ventana), y multiplicar el resultado por 100 para producir el porcentaje de identidad de secuencia. A menos que se indique de cualquier otra manera, la ventana de comparación entre dos secuencias se define por la longitud total de la más corta de las dos secuencias.

20 Un "ácido nucleico diana" es un ácido nucleico en donde se insertará un transposón. Tal diana puede ser parte de un cromosoma, episoma o vector.

Una "secuencia diana de integración" o "secuencia diana" o "sitio diana" para una transposasa es un sitio o secuencia en una molécula de ADN diana en donde una transposasa puede insertar un transposón. La transposasa piggyBac de *Trichoplusia ni* inserta su transposón en la secuencia diana 5'-TTAA-3'. Las transposasas similares a PiggyBac transponen sus transposones mediante el uso de un mecanismo de cortar y pegar, lo que da como resultado la duplicación de su secuencia diana de 4 pares de bases al insertarla en una molécula de ADN. Por tanto, la secuencia diana se encuentra a cada lado de un transposón de similar a piggyBac integrado.

30 El término "traducción" se refiere al proceso mediante el cual un polipéptido es sintetizado por un ribosoma "leyendo" la secuencia de un polinucleótido.

Una "transposasa" es un polipéptido que cataliza la escisión de un transposón correspondiente de un polinucleótido donante, por ejemplo un vector, y (siempre que la transposasa no sea deficiente en integración) la posterior integración del transposón en un ácido nucleico diana. Una " *transposasa de Bombyx* " significa una transposasa con al menos un 80 % de identidad de secuencia con la SEQ ID NO. 407, incluidas las variantes hiperactivas de la SEQ ID NO. 407, que pueden transponer un transposón correspondiente. Una " *transposasa de Xenopus* " significa una transposasa con al menos un 80 % de identidad de secuencia con la SEQ ID NO. 48, incluidas las variantes hiperactivas de la SEQ ID NO. 48, que pueden, cuando se fusionan con una secuencia de localización nuclear heteróloga, transponer un transposón correspondiente.

40 El término "transposición" se usa en la presente descripción para describir la acción de una transposasa al escindir un transposón de un polinucleótido y luego integrarlo, ya sea en un sitio diferente en el mismo polinucleótido, o en un segundo polinucleótido.

45 El término "transposón" significa un polinucleótido que puede escindirse de un primer polinucleótido, por ejemplo, un vector, e integrarse en una segunda posición en el mismo polinucleótido, o en un segundo polinucleótido, por ejemplo, el ADN genómico o extracromosómico de una célula, por la acción de una transposasa de acción trans correspondiente. Un transposón comprende un primer extremo del transposón y un segundo extremo del transposón que son secuencias de polinucleótidos reconocidas y transpuestas por una transposasa. Un transposón normalmente comprende además una primera secuencia de polinucleótidos entre los dos extremos del transposón, de manera que la primera secuencia de polinucleótidos se transpone junto con los dos extremos del transposón por la acción de la transposasa. Los transposones naturales comprenden frecuentemente ADN que codifica una transposasa que actúa sobre el transposón. Los transposones de la presente invención son "transposones sintéticos" que comprenden una secuencia polinucleotídica heteróloga que es transponible en virtud de su yuxtaposición entre dos extremos del transposón.

55 El término "extremo del transposón" significa las secuencias de nucleótidos que actúan en cis que son suficientes para el reconocimiento y la transposición por una transposasa correspondiente. Los extremos del transposón de los transposones similar a piggyBac comprenden repeticiones perfectas o imperfectas, de modo que las repeticiones respectivas en los dos extremos del transposón son complementos inversos entre sí. Estos se conocen como repeticiones terminales invertidas (ITR) o repeticiones invertidas terminales (TIR). Un extremo del transposón puede incluir o no una secuencia adicional próxima al ITR que promueve o aumenta la transposición.

65 El término "vector" o "vector de ADN" o "vector de transferencia de genes" se refiere a un polinucleótido que se usa para realizar una función "portadora" de otro polinucleótido. Por ejemplo, los vectores se usan a menudo para permitir que un polinucleótido se propague dentro de una célula viva, o para permitir que un polinucleótido sea empaquetado para su liberación en una célula, o para permitir que un polinucleótido se integre en el ADN genómico

de una célula. Un vector puede comprender además elementos funcionales adicionales, por ejemplo, puede comprender un transposón.

## 5.2 Descripción

### 5.2.1 Integración genómica

La expresión de un gen a partir de un polinucleótido heterólogo en una célula huésped eucariota puede mejorarse si el polinucleótido heterólogo se integra en el genoma de la célula huésped. La integración de un polinucleótido en el genoma de una célula huésped también lo hace generalmente heredable de manera estable, sometiéndolo a los mismos mecanismos que aseguran la replicación y división del ADN genómico. Esta heredabilidad estable es deseable para lograr una expresión buena y consistente durante largos períodos de crecimiento. Para la fabricación de biomoléculas, particularmente para aplicaciones terapéuticas, la estabilidad del huésped y la consistencia de los niveles de expresión también son importantes para fines regulatorios. Se describen así las células con vectores de transferencia de genes, incluidos los vectores de transferencia de genes basados en transposones, integrados en sus genomas.

Los polinucleótidos heterólogos pueden integrarse de manera más eficiente en un genoma diana si forman parte de un transposón, por ejemplo, para que puedan integrarse mediante una transposasa. Un beneficio particular de un transposón es que se integra todo el polinucleótido entre los ITR del transposón. Esto contrasta con la integración aleatoria, donde un polinucleótido introducido en una célula eucariota a menudo se fragmenta al azar en la célula, y solo partes del polinucleótido se incorporan al genoma diana, generalmente a baja frecuencia. Se ha demostrado que el transposón piggyBac de la oruga de la col *Trichoplusia ni* se transpone por su transposasa en células de muchos organismos (véase por ejemplo Keith y otros, (2008) BMC Molecular Biology 9:72 "Analysis of the piggyBac transposase reveals a functional nuclear targeting signal in the 94 c-terminal residues"). Los polinucleótidos heterólogos incorporados en transposones similares a piggyBac pueden integrarse en células eucariotas, incluidas células animales, células fúngicas o células vegetales. Las células animales preferidas pueden ser de vertebrados o invertebrados. Las células de vertebrados preferidas incluyen células de mamíferos que incluyen roedores tales como ratas, ratones y hámsteres; ungulados, como vacas, cabras u ovejas; y cerdos. Las células de vertebrados preferidas también incluyen células de tejidos humanos y células madre humanas. Los tipos de células diana incluyen linfocitos, hepatocitos, células neurales, células musculares, células sanguíneas, células madre embrionarias, células madre somáticas, células hematopoyéticas, así como embriones, cigotos y espermatozoides no humanos (algunos de los cuales están abiertos para ser manipulados en un entorno *in vitro*). Las células preferidas pueden ser células pluripotentes (células cuyos descendientes pueden diferenciarse en varios tipos de células restringidas, como células madre hematopoyéticas u otras células madre) o células totipotentes no humanas (es decir, una célula cuyos descendientes pueden convertirse en cualquier tipo de célula en un organismo, por ejemplo, células madre embrionarias). Las células de cultivo preferidas son células de ovario de hámster chino (CHO) o células de riñón embrionario humano (HEK293). Las células fúngicas preferidas son células de levadura que incluyen *Saccharomyces cerevisiae* y *Pichia pastoris*. Las células vegetales preferidas son las algas, por ejemplo, *Chlorella*, tabaco, maíz y arroz (Nishizawa-Yokoi y otros, (2014) Planta J. 77: 454-63 "Precise marker excision system using an animal derived piggyBac transposon in plants").

Los sistemas de transferencia de genes preferidos comprenden un transposón en combinación con una proteína transposasa correspondiente que transpone el transposón, o un ácido nucleico que codifica la proteína transposasa correspondiente y es expresable en la célula diana.

Se puede introducir una proteína transposasa en una célula como proteína o como ácido nucleico que codifica la transposasa, por ejemplo como ácido ribonucleico, que incluyen ARNm o cualquier polinucleótido reconocido por la maquinaria de traducción de una célula; como ADN, por ejemplo como ADN extracromosómico que incluye ADN episómico; como ADN plasmídico o como ácido nucleico viral. Además, el ácido nucleico que codifica la proteína transposasa se puede transfectar en una célula como un vector de ácido nucleico, tal como un plásmido, o como un vector de expresión génica, que incluyen un vector viral. El ácido nucleico puede ser circular o lineal. El ADN que codifica la proteína transposasa puede insertarse de manera estable en el genoma de la célula o en un vector para expresión constitutiva o inducible. Cuando la proteína transposasa se transfecta en la célula o se inserta en el vector como ADN, la secuencia que codifica la transposasa está preferentemente unida operativamente a un promotor heterólogo. Hay una variedad de promotores que podrían usarse, que incluyen promotores constitutivos, promotores específicos de tejido, promotores inducibles y similares. Se contemplan expresamente todas las secuencias de ADN o ARN que codifican las proteínas transposasa de *Bombyx* o *Xenopus*. Alternativamente, la transposasa puede introducirse en la célula directamente como proteína, por ejemplo, mediante el uso de péptidos que penetran en la célula (por ejemplo, como se describe en Ramsey y Flynn (2015) Pharmacol. Ther. 154: 78-86 "Cell-penetrating peptides transport therapeutics into cells"); mediante el uso de pequeñas moléculas que incluyen sales y propanobetaina (por ejemplo, como se describe en Astolfo y otros, (2015) Cell 161: 674-690); o electroporación (por ejemplo, como se describe en Morgan y Day (1995) Methods in Molecular Biology 48: 63-"The introduction of proteins into mammalian cells by electroporation").

Es posible insertar el transposón en el ADN de una célula mediante recombinación no homóloga a través de una variedad de mecanismos reproducibles e incluso sin la actividad de una transposasa. Los transposones descritos en la presente descripción se pueden usar para la transferencia de genes independientemente de los mecanismos por los que se transfieren los genes.

5

### 5.2.2 Transposones similares a piggybac derivados de *xenopus*

Los transposones de ADN natural se someten a un sistema de replicación de "cortar y pegar" en el que el transposón se escinde de una primera molécula de ADN y se inserta en una segunda molécula de ADN. Los transposones de ADN se caracterizan por repeticiones terminales invertidas (ITR) y son movilizados por una transposasa codificada por elementos. El sistema transposón/transposasa piggyBac es particularmente útil debido a la precisión con la que el transposón se integra y se escinde (ver, por ejemplo, "Fraser, M. J. (2001) The TTA-Specific Family of Transposable Elements: Identification, Functional Characterization, and Utility for Transformation of Insects. Insect Transgenesis: Methods and Applications. AM Handler y AA James. Boca Raton, Fla., CRC Press: 249-268"; y "US 20070204356 A1: PiggyBac constructs in vertebrates" y referencias en el mismo).

15

Se han encontrado muchas secuencias con similitud de secuencia con la transposasa piggyBac de *Trichoplusia ni* en los genomas de especies filogenéticamente distintas, desde hongos hasta mamíferos, pero se ha demostrado que muy pocas poseen actividad transposasa (ver, por ejemplo, Wu M, y otros (2011 Genetica 139: 149-54. "Cloning and characterization of piggyBac-like elements in lepidopteran insects", y las referencias dentro del mismo).

20

Se ha identificado actividad de escisión en transposasas Txb de *Xenopus* (Hikosaka y otros, Mol. Biol. Evol., 24 (12): 2648-2656, 2007), pero los autores no informaron pruebas de la integración de la diana extirpada en el genoma. Este informe sugirió que tales transposasas carecen de actividad de integración. Sin embargo, sorprendentemente hemos encontrado que las transposasas identificadas originalmente en el genoma de *Xenopus tropicalis* (SEQ ID NO. 48 y 49) son transposicionalmente activas en células de mamíferos cuando se fusionan con una señal de localización nuclear heteróloga. En ausencia de una señal de localización nuclear fusionada, las transposasas de *Xenopus* de origen natural son esencialmente inactivas para la integración genómica (véase el Ejemplo 6.1.1 y la Tabla 1). Nuestro descubrimiento revela por qué Hikosaka no pudo ver la integración: los experimentos realizados por Hikosaka y otros implicó la transfección de una diana de ADN y el ADN que codifica una transposasa en células de mamífero. La transposasa, producida en el citoplasma, podría actuar sobre el ADN transfectado en el citoplasma para escindir el transposón. Sin embargo, no se detectaría actividad de integración genómica si la transposasa, que carecía de NLS, permaneciera citoplasmática.

25

30

Aquí hemos identificado extremos de transposones que incluyen ITR que, cuando se agregan a los extremos de una secuencia de polinucleótidos heteróloga, crean un transposón de *Xenopus* sintético que se integra de manera eficiente en el ADN genómico mediante una transposasa de *Xenopus*. Se añade una secuencia diana izquierda seguida de una secuencia terminal del transposón izquierdo que comprende una secuencia seleccionada de SEQ ID NO: 1-8 en un lado de un polinucleótido heterólogo. Se añade al otro lado del polinucleótido heterólogo una secuencia del extremo derecho del transposón que comprende una secuencia seleccionada de SEQ ID NO: 12-16, y seguida de una secuencia diana derecha. El polinucleótido resultante es un transposón de *Xenopus* sintético y se transpone eficientemente mediante transposasas seleccionadas de SEQ ID NO: 48 o 49, fusionadas a una señal de localización nuclear heteróloga. Consulte las Tablas 1-3 y los Ejemplos 6.1.1, 6.1.2.1 y 6.1.2.2.

35

40

Las transposasas de *Xenopus* reconocen los transposones sintéticos de *Xenopus*. Estas escindieron el transposón de una primera molécula de ADN, cortando el ADN en la secuencia diana en el extremo izquierdo de un extremo del transposón y la secuencia diana en el extremo derecho del segundo extremo del transposón, volviendo a unir los extremos cortados de la primera molécula de ADN para dejar una sola copia de la secuencia diana. La secuencia del transposón escindida, incluido cualquier ADN heterólogo que se encuentre entre los extremos del transposón, es integrada por la transposasa en una secuencia diana de una segunda molécula de ADN, como el genoma de una célula diana.

50

Estos extremos del transposón izquierdo y derecho de *Xenopus* comparten una secuencia repetida casi perfectamente de 14 pb con orientación invertida en los dos extremos: (5'- CCYTTTBMCTGCCA: SEQ ID NO: 19) adyacente a la secuencia diana. Aquí y en otros lugares, cuando las repeticiones invertidas están definidas por una secuencia que incluye un nucleótido definido por un código de ambigüedad, la identidad de ese nucleótido puede seleccionarse independientemente en las dos repeticiones. La conservación casi perfecta de esta secuencia ITR de 14 pb en ambos extremos del transposón *Xenopus* nos permite identificarlo como el transposón ITR. Los transposones que comprenden un polinucleótido heterólogo insertado entre dos extremos del transposón, cada uno de los cuales comprende la SEQ ID NO: 19 en orientaciones invertidas en los dos extremos del transposón, y flanqueados por una secuencia diana, pueden transponerse de una molécula de ADN a otra, mediante sus correspondientes transposasas de *Xenopus*. Las transposasas de *Xenopus* de origen natural (SEQ ID NO: 48 y 49) deben fusionarse con una señal de localización nuclear heteróloga para efectuar esta transposición.

60

Las versiones truncadas y modificadas de los extremos derecho e izquierdo del transposón de origen natural funcionarán como parte de un transposón *Xenopus* sintético. Por ejemplo, como se muestra en el Ejemplo 6.1.2.2 y

65

las Tablas 2 y 3, un extremo izquierdo del transposón que consta de una secuencia diana seguida de una secuencia seleccionada de SEQ ID NO: 4-7, y un extremo derecho del transposón que consta de una secuencia seleccionada de SEQ ID NO: 13-16 seguida de una secuencia diana contiene todas las secuencias necesarias para la transposición de ADN por una transposasa de *Xenopus* fusionada a una señal de localización nuclear heteróloga. Observamos que las diferencias de secuencia se toleran dentro de los extremos del transposón truncado además de las redundancias observadas en las secuencias de ITR. Por ejemplo, la SEQ ID NO: 7 del extremo izquierdo del transposón consta de la SEQ ID NO: 9 además de la ITR, mientras que la SEQ ID NO: 5 del extremo izquierdo del transposón consta de la SEQ ID NO: 10 además de la ITR. De forma similar, la SEQ ID NO: 16 del extremo derecho del transposón consta de la SEQ ID NO: 17 además de la ITR, mientras que la SEQ ID NO: 13 del extremo derecho del transposón consta de la SEQ ID NO: 18 además de la ITR.

Un transposón de *Xenopus* puede comprender un polinucleótido heterólogo flanqueado por dos extremos del transposón, en el que un extremo del transposón comprende una secuencia que es al menos un 90 % idéntica o al menos un 95 % idéntica o al menos un 99 % idéntica a la SEQ ID NO: 7 y un extremo del transposón comprende una secuencia que es al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a SEQ ID NO: 16.

Un transposón de *Xenopus* puede comprender un polinucleótido heterólogo flanqueado por dos extremos del transposón, en donde un extremo del transposón comprende al menos 14 o al menos 16 o al menos 18 o al menos 20 o al menos 25 bases contiguas de SEQ ID NO: 7 y un extremo del transposón comprende al menos 14, o al menos 16, o al menos 18, o al menos 20 bases contiguas de SEQ ID NO: 16.

Un transposón de *Xenopus* puede comprender un polinucleótido heterólogo flanqueado por dos extremos del transposón en donde cada extremo del transposón comprende la secuencia 5'-CCYTTTBMCTGCCA-3' (SEQ ID NO: 19) invertida en orientación en los dos extremos del transposón. Un extremo de este transposón *Xenopus* puede comprender además al menos 14, o al menos 16, o al menos 18, o al menos 20 bases contiguas de SEQ ID NO: 9 y el otro extremo puede comprender además al menos 14 o al menos 16 o al menos 18 o al menos 20 o al menos 25 bases contiguas de SEQ ID NO: 17.

Los transposones de *Xenopus* son transponibles mediante transposasas de *Xenopus*, por ejemplo, mediante al menos un polipéptido seleccionado de SEQ ID NO: 48, 49 o 52-402 y fusionado a una señal de localización nuclear heteróloga. La operatividad de un transposón de *Xenopus* puede demostrarse mediante la capacidad de una transposasa que tiene la secuencia de aminoácidos de SEQ ID NO: 61 fusionada a un NLS heterólogo para transponer el transposón.

Se describen las células cuyos genomas contienen un transposón de *Xenopus*. La célula puede ser cualquier célula eucariota.

### 5.2.3 Transposones similar a piggybac derivados de *bombyx*

Se identificó un transposón del genoma de *Bombyx mori* con los extremos funcionales del transposón contenidos en la SEQ ID NO: 23 y la SEQ ID NO: 29. Una transposasa que puede reconocer y transponer un transposón que comprende estos extremos del transposón es la SEQ ID NO: 407. Las repeticiones terminales invertidas (ITR) en los extremos del transposón natural que comprende las SEQ ID NO: 23 y 29 no estaban flanqueadas por la secuencia diana canónica 5'-TTAA-3' normalmente observada para transposones con una identidad de secuencia significativa con piggyBac de *Trichoplusia ni*; estos estaban flanqueados por secuencias 5'-TTAT-3' adyacentes a las ITR.

Aquí hemos identificado los extremos de los transposones, incluidos los ITR, que se pueden agregar a los extremos de una secuencia de polinucleótidos heteróloga para efectuar la integración eficiente del polinucleótido en el ADN genómico mediante la acción de una transposasa de *Bombyx*. Se añade una secuencia diana izquierda seguida de una secuencia terminal del transposón izquierdo que comprende una secuencia seleccionada de piggyBac: 23-27 en un lado de un polinucleótido heterólogo. Se añade al otro lado del polinucleótido heterólogo una secuencia del extremo derecho del transposón que comprende una secuencia seleccionada de piggyBac: 29-32, seguida de una secuencia diana derecha. El polinucleótido resultante es un transposón de *Bombyx* sintético y se transpone eficientemente mediante la transposasa SEQ ID NO: 407, esté o no fusionada a una señal de localización nuclear heteróloga. Consulte las Tablas 1 y 2 y los Ejemplos 6.1.1 y 6.1.2.1.

Las transposasas de *Bombyx* reconocen los transposones de *Bombyx* sintéticos. Estas escindieron el transposón de una primera molécula de ADN, cortando el ADN en la secuencia diana en el extremo izquierdo de un extremo del transposón y la secuencia diana en el extremo derecho del segundo extremo del transposón, volviendo a unir los extremos cortados de la primera molécula de ADN para dejar una sola copia de la secuencia diana. La secuencia del transposón escindida, incluido cualquier ADN heterólogo que se encuentre entre los extremos del transposón, se integra en una secuencia diana de una segunda molécula de ADN, como el genoma de una célula diana.

Los extremos del transposón de *Bombyx* izquierdo y derecho comparten una secuencia repetida de 16 pb en sus extremos (5'-CCCGGCGAGCATGAGG-3' : SEQ ID NO: 33) invertidos en orientación en los dos extremos inmediatamente adyacentes a la secuencia diana. Es decir, el extremo izquierdo del transposón comienza con la

secuencia 5'-CCCGGCGAGCATGAGG-3' (SEQ ID NO: 33), y el extremo derecho del transposón con el complemento inverso de esta secuencia: 5'-CCTCATGCTCGCCGGG-3' (SEQ ID NO: 34). La perfecta conservación de esta secuencia de 16 pb en ambos extremos del transposón nos permitió identificarlo como el transposón ITR.

5 La redundancia observada para el transposón tipo piggy-Bac de *Xenopus* descrito en la Sección 5.2.2 sugiere que esta secuencia no es completamente inmutable, pero puede aceptar uno o dos o tres cambios de nucleótidos del consenso (como se describe para SEQ ID NO: 19), que proporcionan ITR de *Bombyx* funcionales con una identidad de secuencia del 93 %, 87 % u 81 % con la SEQ ID NO: 33 (o (SEQ ID NO: 34) respectivamente. Un transposón de *Bombyx* puede comprender un polinucleótido heterólogo insertado entre un extremo izquierdo del transposón y  
10 derecho, en donde cada extremo del transposón comprende una secuencia al menos 81 % idéntica o al menos 87 % idéntica o al menos 93 % idéntica a la secuencia 5'-CCCGGCGAGCATGAGG-3' (SEQ ID NO: 33) en un extremo, una secuencia al menos 81 % idéntica o al menos 87 % idéntica o al menos 93 % idéntica a la secuencia 5'-CCTCATGCTCGCCGGG-3' (SEQ ID NO: 34) en el otro extremo.

15 Las versiones truncadas y modificadas de los extremos derecho e izquierdo del transposón también funcionan como parte de un transposón de *Bombyx* sintético. Por ejemplo, como se muestra en el Ejemplo 6.1.2.1 y la Tabla 2, una secuencia diana seguida de un extremo izquierdo del transposón que comprende una secuencia seleccionada de SEQ ID NO: 23-25, y un extremo derecho del transposón que comprende SEQ ID NO: 29 o 31, seguido de una secuencia diana, contiene todas las secuencias necesarias para la transposición de una transposasa de *Bombyx*.

20 Un transposón de *Bombyx* puede comprender un polinucleótido heterólogo flanqueado por dos extremos del transposón en donde un extremo del transposón comprende una secuencia que es al menos 90 % o al menos 95 % idéntica o al menos 99 % idéntica a SEQ ID NO: 25 y un extremo del transposón que comprende una secuencia que es al menos 90 % idéntica o al menos 95 % o al menos 99 % idéntica a SEQ ID NO: 31.

25 Un transposón de *Bombyx* puede comprender un polinucleótido heterólogo flanqueado por dos extremos del transposón, en el que un extremo del transposón comprende al menos 14 o al menos 16 o al menos 18 o al menos 20 bases contiguas de la SEQ ID NO: 25 y un extremo del transposón que comprende al menos 14 o al menos 16 o al menos 18 o al menos 20 bases contiguas de SEQ ID NO: 31.

30 Un transposón de *Bombyx* puede comprender un polinucleótido heterólogo flanqueado por dos extremos del transposón en el que cada extremo del transposón comprende una secuencia que es al menos un 81 % idéntica o al menos un 87 % idéntica o al menos un 93 % idéntica a la secuencia 5'-CCCGGCGAGCATGAGG-3' (SEQ ID NO: 33) con orientación invertida en los dos extremos del transposón. Un extremo de este transposón de *Bombyx* puede comprender además al menos 14, o al menos 16, o al menos 18, o al menos 20 bases contiguas de SEQ ID NO: 27 y el otro extremo puede comprender además al menos 14 o al menos 16 o al menos 18 o al menos 20 bases contiguas de SEQ ID NO: 32.

35 Los transposones de *Bombyx* son transponibles mediante transposasas de *Bombyx*, por ejemplo, mediante al menos un polipéptido seleccionado de SEQ ID NO: 407 o 412-697, opcionalmente fusionado a una señal de localización nuclear heteróloga. La operatividad de un transposón de *Bombyx* puede demostrarse mediante la capacidad de una transposasa que tiene la secuencia de aminoácidos de SEQ ID NO: 415 fusionada a una NLS heteróloga para transponer el transposón.

45 Se describen las células cuyos genomas contienen un transposón de *Bombyx*. La célula puede ser cualquier célula eucariota.

#### 5.2.4 Secuencias diana de transposón modificadas

50 Habiendo observado que los transposones naturales de *Bombyx* y *Xenopus* estaban flanqueados por diferentes secuencias diana (5'-TTAT-3' y 5'TTAA-3' respectivamente), intentamos modificar las secuencias diana de transposones similares a piggyBac cambiando la secuencia adyacente al ITR. Se espera que esto cambie los salientes 5' del transposón extirpado (Mitra y otros, 2008. EMBO J. 27: 1097-1109 "piggyBac can bypass DNA synthesis during cut and paste transposition"). Creamos un transposón piggyBac-TTAT uniendo una secuencia diana  
55 5'-TTAT-3' al extremo izquierdo del transposón piggyBac SEQ ID NO 37 y colocándolo en un lado de la construcción indicadora SEQ ID NO 39, y uniendo el extremo derecho del transposón piggyBac SEQ ID NO 38 seguido de la secuencia diana 5'-TTAT-3' al otro lado. Observamos que *in vivo* en células de mamíferos, el transposón TTAT piggyBac fue integrado por la transposasa piggyBac (SEQ ID NO. 698) para dar expresión de la proteína codificada en el transposón a niveles comparables a los del transposón piggyBac de TTAA (ver Sección 6.1.2 y comparar las filas 24 y 26 de la Tabla 2).

60 Hicimos un cambio similar de la secuencia diana 5'-TTAA-3' a 5'-TTAT-3' para el transposón *Xenopus*. Nuevamente observamos que *in vivo* en células de mamíferos, el transposón TTAT *Xenopus* fue integrado por una transposasa *Xenopus* fusionada a una señal de localización nuclear heteróloga, para dar expresión de la proteína codificada en el transposón a niveles comparables a los del transposón TTAA *Xenopus* integrado por la misma transposasa (consulte la Sección 6.1.2 y compare las filas 14 y 22 de la Tabla 2). Por tanto, una transposasa de *Xenopus* es

eficiente en la transposición de transposones con diferentes secuencias diana, incluidas las secuencias diana 5'-TTAT-3' y 5'-TTAA-3'.

Finalmente, también hicimos el cambio inverso para el transposón de *Bombyx*, cambiando su secuencia diana a TTAA. Observamos que *in vivo* en células de mamíferos, el transposón TTAA *Bombyx* fue integrado por una transposasa de *Bombyx*, para dar expresión de la proteína codificada en el transposón a niveles comparables a los del transposón TTAT *Bombyx* integrado por la misma transposasa (ver Sección 6.1. 2 y compare las filas 3 y 11 de la Tabla 2). Por tanto, una transposasa de *Bombyx* es eficiente en la transposición de transposones con diferentes secuencias diana, incluidas las secuencias diana 5'-TTAT-3' y 5'-TTAA-3'.

En todos los casos de transposones similares a piggyBac que probamos (*Trichoplusia ni*, *Bombyx* y *Xenopus*), las transposasas escindieron sus transposones precisamente del ADN en donde estaban presentes originalmente, dejando una sola copia del 5'-TTAA-3' o secuencia diana 5'-TTAT-3' que estaba inicialmente presente adyacente a cada uno de los ITR del transposón. La escisión precisa de todos estos transposones por sus transposasas es consistente con el mecanismo de cortar y pegar descrito para *Trichoplusia ni* piggyBac.

La transposasa de *Bombyx* SEQ ID NO. 407 comparte una identidad de secuencia del 36 % con la transposasa piggyBac de *Trichoplusia ni*; las transposasas de *Xenopus* SEQ ID NO. 48 y 49 comparten solo el 23 % de identidad de secuencia con la transposasa piggyBac de *Trichoplusia ni*; las transposasas de *Xenopus* SEQ ID NO: 48 y 49 comparten solo el 22 % de identidad de secuencia con la transposasa de *Bombyx* SEQ ID NO: 407. Las 3 de estas transposasas son capaces de transponer eficientemente sus transposones cuando la secuencia diana en el transposón se cambia entre 5'-TTAA-3' y 5'-TTAT-3' o viceversa. Estos datos proporcionan evidencia de que la secuencia diana para cualquier transposón similar a piggyBac se puede cambiar de 5'-TTAA-3' a 5'-TTAT-3' simplemente cambiando la secuencia diana que flanquea los ITR del transposón. Se puede crear un transposón con secuencias diana modificadas para transposasas activas con al menos un 23 % de identidad de secuencia con la transposasa piggyBac de *Trichoplusia ni* (SEQ ID NO: 698), o un 22 % de identidad de secuencia con la transposasa de *Bombyx* SEQ ID NO: 407, o 22 % de identidad de secuencia con transposasas de *Xenopus* SEQ ID NO: 48 o 49, según se identificó mediante el uso del algoritmo TBLASTN, tomando extremos funcionales del transposón izquierdo y derecho y cambiando las secuencias diana adyacentes a los ITR de 5'-TTAA-3' a 5'-TTAT-3'.

La integración eficiente en las secuencias diana 5'-TTAT-3'/ 5'-ATAA-3' puede ser ventajosa, porque 5'-TTAT-3' es un complemento inverso de 5'-ATAA-3' que es parte de la señal canónica de poliA de mamífero 5'-aATAAa-3'. Por tanto, el sitio de inserción 5'-TTAT-3' dirigido al transposón dirigido a TTAT se produce en casi todas las señales poliA. Las señales de poliA están asociadas con regiones transcripcionalmente activas del cromosoma. Por tanto, es probable que los transposones que se insertan en los sitios 5'-TTAT-3', incluidos los transposones *Bombyx* y los transposones *Xenopus* y piggyBac modificados, produzcan niveles de expresión más altos de los genes que portan que los transposones que se insertan en los sitios 5'-TTAA-3'. Este efecto puede volverse más pronunciado con el tiempo, ya que las regiones transcripcionalmente activas de forma natural pueden ser más resistentes a los efectos silenciadores.

Otras secuencias diana usables para los transposones piggyBac 5'-CTAA-3', 5'-TTAG-3', 5'-ATAA-3', 5'-TCAA-3', 5'-AGTT-3', 5'-ATTA-3', 5'-GTTA-3', 5'-TTGA-3', 5'-TTTA-3', 5'-TTAC-3', 5'-ACTA-3', 5'-AGGG-3', 5'-CTAG-3', 5'-GTAA-3', 5'-AGGT-3', 5'-ATCA-3', , 5'-CTCC-3', 5'-TAAA-3', 5'-TCTC-3', 5'-TGAA-3', 5'-AAAT-3', 5'-AATC-3', 5'-ACAA-3', 5'-ACAT-3', 5'-ACTC-3', 5'-AGTG-3', 5'-ATAG-3', 5'-CAAA-3', 5'-CACA-3', 5'-CATA-3', 5'-CCAG-3', 5'-CCCA-3', 5'-CGTA-3', 5'-CTGA-3', 5'-GTCC-3', 5'-TAAG-3', 5'-TCTA-3', 5'-TGAG-3', 5'-TGTT-3', 5'-TTCA-3', 5'-TTCT-3' y 5'-TTTT-3' (Li y otros, 2013. Proc. Natl. Acad. Sci vol. 110, no. 6, E478-487). Esto sugiere que se puede crear un transposón sintético similar a piggyBac mediante el uso de una repetición de una de estas secuencias en lugar de la secuencia diana natural 5'-TTAA-3' o 5'-TTAT-3' que flanquea los ITR del transposón. Por ejemplo, un transposón de *Bombyx* comprende una primera secuencia diana usable, secuencia ITR SEQ ID NO. : 33, un polinucleótido heterólogo, una segunda secuencia ITR SEQ ID NO: 33 invertida en orientación con respecto a la primera, y una segunda secuencia diana usable, donde la primera y la segunda secuencias diana usables son preferentemente las mismas. Un transposón de *Xenopus* comprende una primera secuencia diana usable, secuencia ITR SEQ ID NO: 19, un polinucleótido heterólogo, una segunda secuencia ITR SEQ ID NO. 19 invertida en orientación con respecto a la primera, y una segunda secuencia diana usable, donde la primera y las segundas secuencias diana usables son preferentemente las mismas. Se describen las células cuyos genomas contienen transposones *Xenopus* o *Bombyx*.

### 5.2.5 Sistemas de selección para modificar transposasas similares a piggybac

Dos propiedades de las transposasas que son de particular interés para las modificaciones genómicas son su capacidad para integrar un polinucleótido en un genoma diana y su capacidad para escindir con precisión un polinucleótido de un genoma diana. Ambos pueden seleccionarse con un sistema adecuado.

Un sistema para seleccionar como el primer paso de transposición, que es la escisión de un transposón de un primer polinucleótido, comprende los siguientes componentes: (i) un primer polinucleótido que codifica un primer marcador de selección unido operativamente a secuencias que hacen que se exprese en un huésped de selección y (ii) un primer transposón que comprende extremos del transposón reconocidos por la primera transposasa. El primer

transposón está presente e interrumpe la secuencia de codificación del primer marcador de selección, de manera que el primer marcador de selección no está activo. El primer transposón se coloca en el primer marcador de selección de manera que la escisión precisa del primer transposón hace que se reconstituya el primer marcador de selección. Las células hospederas que contienen el primer polinucleótido, cromosómicamente o extracromosómicamente, pueden usarse para detectar transposasas que pueden escindir el primer transposón.

Si el primer transposón comprende un segundo marcador de selección, operativamente ligado a secuencias que hacen que el segundo marcador de selección sea expresable en el huésped de selección, la transposición del segundo marcador de selección en el genoma de la célula huésped producirá un genoma que comprende el primer y segundo marcadores de selección activos. Por tanto, la célula crecerá en condiciones selectivas para ambos marcadores. El segundo marcador de selección, como el primer marcador de selección, puede ser un gen que codifica un gen de resistencia a antibióticos, o un marcador auxotrófico, o cualquier otro marcador de selección.

Si el primer transposón comprende un primer marcador de contraselección, operativamente ligado a secuencias que hacen que el primer marcador de contraselección sea expresable en el huésped de selección, la transposición del primer marcador de contraselección en el genoma de la célula huésped producirá una célula con un primer marcador de selección activo y un primer marcador de contraselección activo. Por lo tanto, la célula morirá en condiciones restrictivas para el primer marcador de contraselección.

Estos dos esquemas de selección se pueden combinar mediante el uso de un segundo marcador de selección que también es un primer marcador de contraselección. Los ejemplos de tales marcadores incluyen genes marcadores auxotróficos en las rutas sintéticas de uracilo o triptófano. Estos genes pueden seleccionarse cultivando células en ausencia del nutriente, en este caso uracilo o triptófano, respectivamente. Los genes biosintéticos también pueden actuar como marcadores de contraselección si permiten que una célula incorpore un análogo tóxico en lugar de un precursor metabólico genuino en sus moléculas. Los genes en la vía biosintética del uracilo pueden convertir el compuesto no tóxico ácido 5-fluoroorótico en 5-fluorouracilo tóxico, por lo que el crecimiento de células con ácido 5-fluoroorótico es restrictivo para una vía funcional del uracilo. De manera similar, el ácido 5-fluoroantranílico se convierte mediante la ruta de síntesis de triptófano en el tóxico 5-fluorotriptófano, por lo que el crecimiento de células con ácido 5-fluoroantranílico es restrictivo para una ruta funcional del triptófano. Las células huésped que contienen un primer polinucleótido que comprende un primer marcador de selección interrumpido por un transposón que comprende un gen de uracilo o triptófano, pueden usarse para cribar simultáneamente para transposasas hiperactivas y deficientes en integración. Por ejemplo, un polinucleótido expresable en la célula huésped que codifica una primera transposasa o una primera biblioteca de transposasas, tal como una biblioteca de mutagénesis de saturación de sitio para una o más posiciones de aminoácidos, se introduce en las células huésped que contienen el primer polinucleótido. Estas células se dividen en dos grupos. El primer grupo se cultiva en condiciones que son selectivas para el primer marcador de selección y restrictivas para el primer marcador de contraselección. A continuación, los genes que codifican las transposasas se aíslan de las células huésped que adquirieron la capacidad de crecer, y los genes de transposasas de este primer grupo de células pueden analizarse para identificar cambios de aminoácidos que mejoran la actividad de escisión pero no la actividad de integración. El segundo grupo se cultiva en condiciones que son selectivas para el primer marcador de selección y para el segundo marcador de selección. Los genes que codifican las transposasas se aíslan luego de las células hospederas que adquirieron la capacidad de crecer, y los genes de transposasas de este segundo grupo de células pueden analizarse para identificar cambios de aminoácidos que mejoran la actividad de transposasas completas.

Estos sistemas de selección pueden usarse para identificar transposasas con actividades modificadas mediante el cribado de bibliotecas de transposasas variantes. Un tipo de biblioteca es un conjunto de polinucleótidos que codifican todas las posibles sustituciones de aminoácidos en una primera posición de aminoácido en la transposasa. Una biblioteca de mutagénesis por saturación de sitio en una sola posición codifica veinte polipéptidos diferentes, incluido uno que es la secuencia de transposasa natural. Para una transposasa que tiene 600 aminoácidos de longitud, todas las posibles sustituciones de un solo aminoácido están presentes en 600 de tales bibliotecas de mutagénesis de saturación de sitio, una para cada posición. Estas bibliotecas se pueden probar mediante el uso de un sistema de selección de transposasa para identificar sustituciones activas en cada posición.

Las mutaciones favorables individuales se pueden combinar de diversas formas, por ejemplo, mediante "alternado de ADN" o mediante los métodos descritos en Patente de EE. UU. 8,635,029 B2. Se puede obtener una transposasa con actividad modificada, ya sea para la actividad en una nueva secuencia diana que incluye una secuencia diana 5'-TTAT-3', o una mayor actividad en una secuencia diana existente mediante el uso de variaciones del esquema de selección descrito anteriormente con un transposón correspondiente.

La actividad de las transposasas también puede aumentar mediante la fusión de la señal de localización nuclear (NLS) en el extremo N, el extremo C, en ambos extremos N y C o en las regiones internas de la proteína transposasa, siempre que se mantenga la actividad de la transposasa. Una señal o secuencia de localización nuclear (NLS) es una secuencia de aminoácidos que "marca" o facilita la interacción de una proteína, ya sea directa o indirectamente con proteínas de transporte nuclear para su importación al núcleo celular. Las señales de localización nuclear (NLS) usadas pueden incluir, pero no se limitan a, secuencias de NLS consenso, secuencias de NLS virales, secuencias de NLS celulares y combinaciones de las mismas.

Las transposasas también pueden fusionarse con otros dominios funcionales de proteínas. Dichos dominios funcionales de proteínas pueden incluir dominios de unión a ADN, regiones de bisagra flexibles que pueden facilitar una o más fusiones de dominios y combinaciones de los mismos. Pueden realizarse fusiones en el extremo N-terminal, el extremo C-terminal o en las regiones internas de la proteína transposasa siempre que se mantenga la actividad transposasa. Los dominios de unión al ADN usados pueden incluir un dominio de hélice-giro-hélice, un dominio de dedo de zinc, un dominio de zipper de leucina o un dominio de hélice-blucle-hélice. Los dominios de unión a ADN específicos usados pueden incluir un dominio de unión a ADN de Gal4, un dominio de unión a ADN LexA o un dominio de unión a ADN Zif268. Las regiones de bisagra flexibles usadas pueden incluir enlazadores de glicina/serina y variantes de los mismos.

Puede usarse un proceso comparable para aumentar la capacidad de transposición de los extremos del transposón mediante una transposasa. En este caso, el transposón puede comprender un primer marcador de selección activo. Los extremos del transposón se pueden seleccionar de cualquier transposón similar a piggyBac. La secuencia de uno o ambos extremos del transposón puede someterse a cambios de secuencia predeterminados o aleatorios, que incluyen cambios en la secuencia diana, la ITR o en otras partes de los extremos del transposón. A continuación, el transposón puede introducirse en una primera célula que contiene un polinucleótido diana que comprende un segundo marcador de selección activo y una transposasa activa. Si la transposasa puede transponer el transposón, alguna fracción de los transposones se transpondrá al polinucleótido diana. El polinucleótido diana se purifica de la primera célula y se introduce en una segunda célula que se somete a condiciones restrictivas para las que requiere el primer marcador de selección y el segundo marcador de selección para sobrevivir. El transposón puede recuperarse, por ejemplo, secuenciando fuera del transposón para identificar la secuencia flanqueante y luego amplificando el transposón mediante el uso de PCR. El proceso puede realizarse en grupos de variantes: un transposón más activo creará polinucleótidos diana que contienen ambos marcadores de selección con mayor frecuencia y, por lo tanto, estará más representado en la población. En este proceso, el transposón puede estar presente opcionalmente como una interrupción reversible en un marcador de selección como se describe para la pantalla de actividad de la transposasa. Sin embargo, esto no es necesario para la pantalla de actividad de transposones, ya que los transposones transpuestos se detectan directamente.

#### 5.2.6. Transposasas de *xenopus* modificadas

Sometimos la transposasa de *Xenopus* SEQ ID NO: 48 a mutagénesis de saturación como se describe en el ejemplo 6.3.1.1, e identificamos 1793 (16,0 %) sustituciones de aminoácidos que estaban asociadas con una mayor actividad de transposición (una medida compuesta de integración y escisión) y 1074 (9,6 %) sustituciones de aminoácidos que se asociaron con una mayor actividad de escisión, de un total de 11.172 posibles sustituciones (19 posibles sustituciones en cada uno de los 588 aminoácidos excluyendo la invariante metionina N-terminal). Las dos clases de sustituciones tenían cierta superposición, pero no eran idénticas ni una clase contenía completamente a la otra. Estas sustituciones beneficiosas se muestran en las columnas C y D de la Tabla 4.

Se encontró que un número similar de sustituciones eran esencialmente neutrales en cuanto al efecto sobre la actividad de transposición o escisión: es decir, estaban presentes aproximadamente con la misma frecuencia en bibliotecas no seleccionadas y posteriores a la selección. Por tanto, las transposasas de *Xenopus* aceptan fácilmente muchas sustituciones de aminoácidos sin detrimento funcional significativo.

Por tanto, se pueden crear transposasas de *Xenopus* que no son secuencias de origen natural (por ejemplo, no SEQ ID NO: 48 o 49), pero que son al menos 99 % idénticas, o al menos 98 % idénticas, o al menos 97 % idénticas, o al menos 96 % idénticas, o al menos 95 % idénticas, o al menos 90 % idénticas, o al menos 84 % idénticas a SEQ ID NO:48. Dichas variantes pueden retener la actividad parcial de la transposasa de SEQ ID NO: 48 (según lo determinado por una o ambas actividades de transposición y/o escisión), pueden ser funcionalmente equivalentes a la transposasa de SEQ ID NO: 48 en una o ambas de transposición y escisión, o puede tener una actividad mejorada con respecto a la transposasa de SEQ ID NO: 48 en transposición, actividad de escisión o ambas. Dichas variantes pueden incluir mutaciones que se muestra en la presente descripción para aumentar la transposición y/o escisión, mutaciones que se muestra en la presente descripción como neutrales en cuanto a transposición y/o escisión, y mutaciones perjudiciales para la transposición y/o integración o cualquier combinación de tales mutaciones. Las variantes preferidas incluyen mutaciones que se ha demostrado que son neutras o que mejoran la transposición y/o la escisión. Algunas de estas variantes carecen de mutaciones que se ha demostrado que son perjudiciales para la transposición y/o la escisión. Algunas de estas variantes incluyen sólo mutaciones que se ha demostrado que potencian la transposición, sólo mutaciones que se ha demostrado que potencian la escisión, o mutaciones que se ha demostrado que potencian tanto la transposición como la escisión.

Actividad mejorada significa actividad (por ejemplo, actividad de transposición o escisión) que es mayor más allá del error experimental que la de una transposasa de referencia de la que se derivó una variante. La actividad puede ser mayor en un factor de, por ejemplo, 1,5, 2, 5, 10, 20, 50 o 100 veces la transposasa de referencia. La actividad mejorada puede encontrarse dentro de un intervalo de, por ejemplo, 2-100 veces, 2-50 veces, 5-50 veces o 2-10 veces de la transposasa de referencia. Aquí y en otros lugares las actividades se pueden medir como se demuestra en los ejemplos.

Equivalencia funcional significa que una transposasa variante puede mediar en la transposición y/o escisión del mismo transposón con una eficiencia comparable (dentro del error experimental) a una transposasa de referencia. Más de 80 secuencias representativas de transposasas variantes de *Xenopus* con frecuencias de transposición comparables a la transposasa de *Xenopus* de origen natural SEQ ID NO. 48 son SEQ ID NO: 325-402.

Además, se pueden crear secuencias variantes de SEQ ID NO.48 combinando dos, tres o cuatro, o cinco o más sustituciones seleccionadas de la Tabla 4, columna C o D. Combinando sustituciones beneficiosas, por ejemplo, las que se muestran en la columna C o D de la Tabla. 4 puede resultar en variantes hiperactivas de SEQ ID NO.48. Dichas variantes pueden crearse en una biblioteca, por ejemplo, mediante mezcla de ADN, y luego identificarse mediante selección mediante el uso de un esquema como se describe en la Sección 5.2.5 o en el Ejemplo 6.3.1. Alternativamente, los métodos descritos en Patente de EE. UU. 8,635,029 se puede usar para diseñar, sintetizar y probar un pequeño número de variantes que incorporan sustituciones de aminoácidos para obtener transposasas con actividades de integración o escisión mejoradas.

Se prepararon variantes de la transposasa de *Xenopus* que son hiperactivas para la integración en células de levadura y de mamífero como se describe en el Ejemplo 6.3.1.1. Identificamos al menos 25 transposasas de *Xenopus* (SEQ ID NO: 52-76) con frecuencias de transposición aproximadamente al menos 50 veces mayores que las de la transposasa de *Xenopus* de origen natural SEQ ID NO: 48. Identificamos más de 130 transposasas de *Xenopus* (SEQ ID NO: 77-210) con frecuencias de transposición entre aproximadamente 10 veces mayores y 50 veces mayores que las de la transposasa de *Xenopus* de origen natural SEQ ID NO: 48. Identificamos más de 100 transposasas de *Xenopus* (SEQ ID NO: 211-324) con frecuencias de transposición entre aproximadamente 2 veces mayores y 10 veces mayores que las de la transposasa de *Xenopus* de origen natural SEQ ID NO: 48. Estas transposasas comprendían una o más de las sustituciones (relativas a SEQ ID NO: 48) enumeradas en la Tabla 4, columnas C y D. Las transposasas hiperactivas preferidas de *Xenopus* comprendían una o más de las sustituciones (relativas a SEQ ID NO: 48) enumeradas en la Tabla 11 columna C. Las transposasas hiperactivas de *Xenopus* preferidas incluyen polipéptidos que comprenden una de las SEQ ID NO: 52-402; algunas transposasas hiperactivas pueden comprender además una secuencia de localización nuclear heteróloga.

Las transposasas hiperactivas de *Xenopus* preferidas comprenden una secuencia de aminoácidos, distinta de una proteína de origen natural (por ejemplo, una transposasa cuya secuencia de aminoácidos comprende SEQ ID NO: 48 o 49), que es al menos 85 % idéntica o al menos 90 % idéntica o al menos 95 % idéntica, o al menos 99 % idéntica a la secuencia de aminoácidos de cualquiera de las SEQ ID NO: 51-406, incluida la SEQ ID NO: 61. Algunas transposasas hiperactivas preferidas comprenden una secuencia de aminoácidos, distinta de una proteína natural, que es al menos un 85 % idéntica o al menos un 90 % idéntica o al menos un 95 % idéntica, o al menos un 99 % idéntica a la secuencia de aminoácidos de la SEQ ID NO: 61 y que comprende al menos una sustitución de aminoácido (con respecto a SEQ ID NO: 48) que se muestra en la Tabla 4 columna C, Tabla 4 columna D o Tabla 11 columna C. Las transposasas hiperactivas preferidas de *Xenopus* incluyen polipéptidos que comprenden una sustitución de aminoácido en una posición seleccionada entre los aminoácidos 6, 7, 16, 19, 20, 21, 22, 23, 24, 26, 28, 31, 34, 67, 73, 76, 77, 88, 91, 141, 145, 146, 148, 150, 157, 162, 179, 182, 189, 192, 193, 196, 198, 200, 210, 212, 218, 248, 263, 270, 294, 297, 308, 310, 333, 336, 354, 357, 358, 359, 377, 423, 426, 428, 438, 447, 447, 450, 462, 469, 472, 498, 502, 517, 520, 523, 533, 534, 576, 577, 582, 583 o 587 (con respecto a SEQ ID NO: 48). Las transposasas hiperactivas de *Xenopus* preferidas incluyen polipéptidos que comprenden una sustitución de aminoácido, con respecto a la SEQ ID NO: 48, seleccionados entre Y6C, S7G, M16S, S19G, S20Q, S20G, S20D, E21D, E22Q, F23T, F23P, S24Y, S26V, S28Q, V31K, A34E, L67A, G73H, A76V, D77N, P88A, N91D, Y141Q, Y141A, N145E, N145V, P146T, P146V, P146K, P148T, P148H, Y150G, Y150S, Y150C, L2C, H157Y, H157Y T189G, L192H, S193N, S193K, V196I, S198G, T200W, L210H, F212N, N218E, A248N, L263M, Q270L, S294T, T297M, S308R, L310R, L333M, D3H35, C435, L735H, L333F, D3 P426K, K428R, S438A, T447G, T447A, L450V, A462H, A462Q, I469V, I472L, Q498M, L502V, E517I, P520D, P520G, N523S, I533E, D534A, F576E, I533E, D534A, Y583F, L587I o cualquier combinación de los mismos que incluya al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o todas estas mutaciones.

Se prepararon variantes de la transposasa de *Xenopus* que son activas para la escisión pero deficientes en la integración en células de levadura y de mamífero como se describe en el Ejemplo 6.3.1.1. Las secuencias de transposasa de *Xenopus* de integración deficiente preferidas incluyen SEQ ID NO: 51 y 403-406; estas transposasas de *Xenopus* deficientes en integración pueden comprender además una secuencia de localización nuclear heteróloga. Las transposasas de *Xenopus* de integración deficiente preferidas comprenden una secuencia de aminoácidos, distinta de una proteína de origen natural, que es 90 % idéntica a la secuencia de aminoácidos de SEQ ID NO. 405. Algunas transposasas de integración deficiente preferidas comprenden una secuencia de aminoácidos, distinta de una proteína de origen natural, que comprende una sustitución de aminoácido (con respecto a SEQ ID NO.48) en la que el Asn en la posición del aminoácido 218 se reemplaza con un residuo Glu o un Asp (N218D o N218E).

Se describen métodos para crear células transgénicas mediante el uso de transposasas hiperactivas de *Xenopus*. Un método para crear una célula transgénica comprende (i) introducir en una célula eucariota una transposasa hiperactiva de *Xenopus* (como una proteína o como un polinucleótido que codifica la transposasa) y un transposón

de *Xenopus* correspondiente; (ii) identificar una célula en la que se incorpora un transposón de *Xenopus* en el genoma de la célula eucariota. La identificación de la célula en la que se incorpora un transposón de *Xenopus* en el genoma de la célula eucariota puede comprender seleccionar la célula eucariota para un marcador de selección codificado en el transposón de *Xenopus*. El marcador de selección puede ser cualquier polipéptido seleccionable, incluido cualquiera de los descritos en la presente descripción.

#### 5.2.7. Transposasas de *bombyx* modificadas

Sometimos la transposasa de *Bombyx* SEQ ID NO.407 a mutagénesis de saturación como se describe en el Ejemplo 6.3.1, e identificamos 1176 sustituciones de aminoácidos (10,1 %) que estaban asociadas con una mayor actividad de transposición y 1044 (9,0 %) sustituciones de aminoácidos que estaban asociadas con aumento de la actividad de escisión, de un total de 11.571 posibles sustituciones (19 posibles sustituciones en cada uno de los 609 aminoácidos excluyendo la invariante metionina N-terminal). Las dos clases de sustituciones tenían cierta superposición, pero no eran idénticas ni una clase contenía completamente a la otra. Estas sustituciones beneficiosas se muestran en la Tabla 4, columnas G y H.

Se encontró que un número similar de sustituciones eran esencialmente neutrales en cuanto a la actividad de transposición o escisión: es decir, estaban presentes aproximadamente con la misma frecuencia en bibliotecas no seleccionadas y de post-selección. Por tanto, las transposasas de *Bombyx* aceptan fácilmente muchas sustituciones de aminoácidos sin detrimento funcional significativo.

Por tanto, se pueden crear transposasas que no son secuencias de origen natural, por ejemplo, no *Bombyx* transposasa SEQ ID NO. 407, pero que son al menos 99 % idénticas, o al menos 98 % idénticas, o al menos 97 % idénticas, o al menos 96 % idéntica, o al menos 95 % idéntica, o al menos 90 % idéntica, o al menos 84 % idéntica a SEQ ID NO. 407 (pero no comprenden SEQ ID NO: 407 per se).

Tales variantes pueden retener la actividad parcial de la transposasa de SEQ ID NO: 407 (transposición y/o actividad de escisión), pueden ser funcionalmente equivalentes a la transposasa de SEQ ID NO: 407 en cualquiera o ambas de transposición y actividad de escisión, o pueden tener actividad mejorada con respecto a la transposasa de SEQ ID NO: 407 en transposición, actividad de escisión o ambas. Dichas variantes pueden incluir mutaciones que se muestra en la presente descripción para aumentar la transposición y/o escisión, mutaciones que se muestra en la presente descripción como neutrales en cuanto a transposición y/o escisión, y mutaciones perjudiciales para la transposición y/o integración o cualquier combinación de tales mutaciones. Las variantes preferidas incluyen mutaciones que se ha demostrado que son neutras o que potencian la transposición y/o la escisión. Algunas de estas variantes carecen de mutaciones que se ha demostrado que son perjudiciales para la transposición y/o la escisión. Algunas de estas variantes incluyen solo mutaciones que han demostrado mejorar la transposición, solo mutaciones que han demostrado mejorar la escisión o mutaciones que han demostrado mejorar tanto la transposición como la escisión.

Actividad mejorada significa actividad que es mayor más allá del error experimental de la de una transposasa de referencia de la que se derivó una variante. La actividad puede ser mayor en un factor de, por ejemplo, 1,5, 2, 5, 10, 20, 50 o 100 veces la transposasa de referencia. La actividad mejorada puede encontrarse dentro de un intervalo de, por ejemplo, 2-100 veces, 2-50 veces, 5-50 veces o 2-10 veces de la transposasa de referencia. Aquí y en otros lugares las actividades se pueden medir como se demuestra en los ejemplos.

Más de 60 secuencias representativas de variantes de transposasas de *Bombyx* con frecuencias de transposición comparables a la transposasa de *Bombyx* de origen natural SEQ ID NO.407 son SEQ ID NO: 634-697.

Además, se pueden crear secuencias variantes de SEQ ID NO: 407 combinando dos, tres o cuatro, o cinco o más sustituciones que se muestran en la Tabla 4, columnas G y H. Combinando sustituciones beneficiosas, por ejemplo, las que se muestran en la columna G o H de La Tabla 4 puede dar como resultado variantes hiperactivas de SEQ ID NO: 407. Dichas variantes pueden crearse en una biblioteca, por ejemplo, mediante mezcla de ADN, y luego identificarse mediante selección mediante el uso de un esquema como se describe en la Sección 5.2.5 o en el Ejemplo 6.3.1.

Las variantes de la transposasa de *Bombyx* que son hiperactivas para la integración en células de levadura y de mamífero se prepararon como se describe en el Ejemplo 6.3.2.1. Se obtuvieron muchas transposasas hiperactivas. Identificamos al menos 20 transposasas de *Bombyx* (SEQ ID NO: 412-431) con frecuencias de transposición aproximadamente al menos 50 veces mayores que las de la transposasa de *Bombyx* de origen natural SEQ ID NO: 407. Identificamos más de 90 transposasas de *Bombyx* (SEQ ID NO: 432-524) con frecuencias de transposición entre aproximadamente 10 veces mayores y 50 veces mayores que las de la transposasa de *Bombyx* de origen natural SEQ ID NO: 407. Identificamos más de 100 transposasas de *Bombyx* (SEQ ID NO: 525-633) con frecuencias de transposición entre aproximadamente 2 veces mayores y 10 veces mayores que las de la transposasa de *Bombyx* de origen natural SEQ ID NO: 407. Estas transposasas comprendían una o más de las sustituciones (relativas a SEQ ID NO: 407) enumeradas en la Tabla 4, columnas G y H. Las transposasas de *Bombyx* hiperactivas preferidas comprenden una o más de las sustituciones (relativas a SEQ ID NO: 407) enumeradas en la Tabla 4 columnas G y H o columna H de la Tabla 11. Las transposasas de *Bombyx* hiperactivas preferidas incluyen

polipéptidos que comprenden una de las SEQ ID NO: 412-524; estas transposasas hiperactivas pueden comprender además una secuencia de localización nuclear heteróloga. Las transposasas hiperactivas preferidas comprenden una secuencia de aminoácidos, distinta de una proteína natural, que es al menos un 85 % idéntica o al menos un 90 % idéntica o al menos un 95 % idéntica, o al menos un 99 % idéntica a la secuencia de aminoácidos de SEQ ID NO: 415. Las transposasas de *Bombyx* hiperactivas preferidas incluyen polipéptidos que comprenden una sustitución de aminoácido en una posición seleccionada entre 92, 93, 96, 97, 165, 178, 189, 196, 200, 201, 211, 215, 235, 238, 246, 253, 258, 261, 263, 271, 303, 321, 324, 330, 373, 389, 399, 402, 403, 404, 448, 473, 484, 507, 523, 527, 528, 543, 549, 550, 557, 601, 605, 607, 609 o 610 (con respecto a SEQ ID NO: 407). Las transposasas de *Bombyx* hiperactivas preferidas incluyen polipéptidos que comprenden una sustitución de aminoácido, con respecto a la SEQ ID NO: 407, seleccionados de Q92A, V93L, V93M, P96G, F97H, F97C, H165E, H165W, E178S, E178H, C189P, A196G, L200I, A201 L211A, W215Y, G219S, Q235Y, Q235G, Q238L, K246I, K253V, M258V, F261L, S263K, C271S, N303R, F321W, F321D, V324K, V324H, A330V, L372K, V38N, L937L D404S, D404M, N441R, G448W, E449A, V469T, C473Q, R484K, T507C, G523A, I527M, Y528K, Y543I, E549A, K550M, P557S, E601V, E605H, E605W, S609H o cualquier combinación de los mismos. Algunas combinaciones incluyen al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o todas estas mutaciones. Algunas transposasas hiperactivas preferidas comprenden una secuencia de aminoácidos, distinta de una proteína natural, que es al menos un 85 % idéntica o al menos un 90 % idéntica o al menos un 95 % idéntica, o al menos un 99 % idéntica a la secuencia de aminoácidos de la SEQ ID NO: 415 y que comprende al menos una sustitución de aminoácido (relativa a SEQ ID NO: 407) mostrada en la Tabla 4 columna F, Tabla 4, G o Tabla 11 columna H.

Las variantes de la transposasa de *Bombyx* que son activas para la escisión pero deficientes en la integración en células de levadura y de mamífero se prepararon como se describe en el Ejemplo 6.3.2.1. Las secuencias de transposasa de integración deficiente preferidas comprenden una de SEQ ID NO: 409-411; estas transposasas deficientes en integración pueden comprender además una secuencia de localización nuclear heteróloga. Las transposasas de integración deficiente preferidas comprenden una secuencia de aminoácidos, distinta de una proteína de origen natural, que es al menos un 90 % idéntica a la secuencia de aminoácidos de SEQ ID NO: 411.

Los métodos para crear células transgénicas mediante el uso de transposasas *Bombyx* hiperactivas se especifican en las reivindicaciones. Un método para crear una célula transgénica comprende (i) introducir en una célula eucariota una transposasa de *Bombyx* hiperactiva (como una proteína o como un polinucleótido que codifica la transposasa) y un transposón de *Bombyx* correspondiente; (ii) identificar una célula en la que se incorpora un transposón de *Bombyx* en el genoma de la célula eucariota. La identificación de la célula en la que se incorpora un transposón de *Bombyx* en el genoma de la célula eucariota puede comprender seleccionar la célula eucariota para un marcador de selección codificado en el transposón de *Bombyx*. El marcador de selección puede ser cualquier polipéptido seleccionable, incluido cualquiera de los descritos en la presente descripción.

#### 5.2.8 Sistemas de transferencia de genes

Los sistemas de transferencia de genes comprenden un polinucleótido que se transferirá a una célula huésped. El sistema de transferencia de genes puede comprender cualquiera de los transposones o transposasas descritos en la presente descripción, o puede comprender uno o más polinucleótidos que tienen otras características que facilitan la transferencia de genes eficiente sin la necesidad de una transposasa o transposón.

Cuando hay múltiples componentes de un sistema de transferencia de genes, por ejemplo el uno o más polinucleótidos que comprenden genes para la expresión en la célula diana y que opcionalmente comprenden extremos del transposón, y una transposasa (que puede proporcionarse como una proteína o codificada por un ácido nucleico), estos componentes se pueden transfectar en una célula al mismo tiempo o secuencialmente. Por ejemplo, una proteína transposasa o su ácido nucleico codificante puede transfectarse en una célula antes, simultáneamente con o después de la transfección de un transposón correspondiente. Además, la administración de al menos un componente del sistema de transferencia de genes puede ocurrir repetidamente, por ejemplo, mediante la administración de al menos dos dosis de este componente.

Las proteínas transposasa de *Bombyx* o *Xenopus* pueden estar codificadas por polinucleótidos que incluyen ARN o ADN. Las moléculas de ARN preferibles incluyen aquellas con sustituciones apropiadas para reducir los efectos de toxicidad en la célula, por ejemplo, sustitución de uridina por pseudouridina y sustitución de citosina por 5-metil citosina. De manera similar, el ácido nucleico que codifica la proteína transposasa o el transposón de esta invención se puede transfectar en la célula como un fragmento lineal o como un fragmento circularizado, preferentemente como un plásmido o como ADN viral recombinante.

Los componentes del sistema de transferencia de genes se pueden transfectar preferentemente en una o más células mediante técnicas como bombardeo de partículas, electroporación, microinyección, combinación de los componentes con vesículas que contienen lípidos, como vesículas lipídicas catiónicas, reactivos de condensación de ADN (por ejemplo, fosfato cálcico, polilisina o polietilenimina) e insertar los componentes (es decir, los ácidos nucleicos del mismo en un vector viral y poner en contacto el vector viral con la célula). Cuando se usa un vector viral, el vector viral puede incluir cualquiera de una variedad de vectores virales conocidos en la técnica, que incluyen los vectores virales seleccionados del grupo que consiste en un vector retroviral, un vector de adenovirus o

un vector viral adenoasociado. El sistema de transferencia de genes puede formularse de una manera adecuada como se conoce en la técnica, o como una composición farmacéutica o kit.

#### 5.2.9 Elementos de secuencia en sistemas de transferencia de genes

La expresión de genes a partir de un polinucleótido de transferencia de genes integrado en el genoma de una célula huésped suele estar fuertemente influenciada por el entorno de cromatina en el que se integra. Los polinucleótidos que se integran en la eucromatina tienen niveles de expresión más altos que los que se integran en la heterocromatina o que se silencian después de su integración. El silenciamiento de un polinucleótido heterólogo puede reducirse si comprende un elemento de control de cromatina. Por tanto, es ventajoso que los sistemas de transferencia de genes comprendan elementos de control de la cromatina tales como secuencias que eviten la propagación de la heterocromatina (aislantes). Por ejemplo, es ventajoso que un polinucleótido de transferencia de genes que se integrará en un genoma del huésped comprenda una secuencia que sea al menos un 95 % idéntica a una secuencia seleccionada entre las SEQ ID NO: 869-876 y SEQ ID NO: 866. Los polinucleótidos de transferencia de genes ventajosos comprenden una secuencia aislante que es al menos un 95 % idéntica a una secuencia seleccionada de una de las SEQ ID NO: 859-865, también pueden comprender elementos de apertura de cromatina de acción ubicua (UCOE) o elementos estabilizantes y antirrepresores (STAR), para aumentar la expresión estable a largo plazo del polinucleótido de transferencia de genes integrado.

En algunos casos, es ventajoso que un polinucleótido de transferencia de genes comprenda dos aislantes, uno a cada lado del polinucleótido heterólogo que contiene las secuencias a expresar. Los aislantes pueden ser iguales o pueden ser diferentes. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden una secuencia que es al menos un 95 % idéntica a una secuencia seleccionada entre SEQ ID NO: 864 o SEQ ID NO: 865 y una secuencia que es al menos un 95 % idéntica a una secuencia seleccionada de una de SEQ ID NO: 859-865. Los aislantes también protegen los elementos de control de expresión entre sí. Por ejemplo, cuando un polinucleótido de transferencia de genes comprende genes que codifican dos marcos de lectura abiertos, cada uno unido operativamente a un promotor diferente, un promotor puede reducir la expresión del otro en un fenómeno conocido como interferencia transcripcional. La intercalación de una secuencia aislante que es al menos un 95 % idéntica a una secuencia seleccionada de una de las SEQ ID NO: 859-865 entre las dos unidades transcripcionales puede reducir esta interferencia y aumentar la expresión de uno o ambos promotores.

En casos preferidos, un vector de transferencia de genes comprende elementos de expresión capaces de impulsar altos niveles de expresión génica. En las células eucariotas, la expresión génica está regulada por varias clases diferentes de elementos, incluidos potenciadores, promotores, intrones, elementos de exportación de ARN, secuencias de poliadenilación y terminadores transcripcionales.

Los polinucleótidos de transferencia de genes particularmente ventajosos para la transferencia de genes para expresión en células de mamíferos comprenden un potenciador para los genes tempranos inmediatos 1, 2 y 3 del citomegalovirus (CMV) de células humanas o murinas (por ejemplo, secuencias al menos un 95 % idénticas a cualquiera de SEQ ID NO: 877-889), un potenciador del potenciador de la proteína tardía principal adenoviral (por ejemplo, secuencias al menos un 95 % idénticas a SEQ ID NO: 890), o un potenciador de SV40 (por ejemplo secuencias al menos un 95% idénticas a SEQ ID NO: 891).

Los polinucleótidos de transferencia de genes particularmente ventajosos para la transferencia de genes para expresión en células de mamíferos comprenden un promotor EFla de cualquier especie de mamífero o ave, incluidos humanos, ratas, ratones, pollos y hámster chino (por ejemplo, secuencias al menos un 95 % idénticas a cualquiera de las SEQ ID NO: 892-910); un promotor de los genes tempranos inmediatos 1, 2 y 3 del citomegalovirus (CMV) de células humanas o murinas (por ejemplo, secuencias al menos un 95 % idénticas a cualquiera de SEQ ID NO: 911-920); un promotor del factor de alargamiento eucariota 2 (EEF2) de cualquier especie de mamífero o ave, incluidos humanos, ratas, ratones, pollos y hámsteres chinos (por ejemplo, secuencias idénticas al menos en un 95 % a cualquiera de las SEQ ID NO: 921-928); un promotor GAPDH de cualquier especie de mamífero o levadura (por ejemplo, secuencias al menos un 95 % idénticas a cualquiera de las SEQ ID NO: 936 y 949-951), un promotor de actina de cualquier especie de mamífero o ave, incluidos humanos, ratas, ratones, pollos y Hámster chino (por ejemplo, secuencias idénticas al menos en un 95 % a cualquiera de SEQ ID NO: 929-935); un promotor PGK de cualquier especie de mamífero o ave, incluidos humanos, ratas, ratones, pollos y hámsteres chinos (por ejemplo, secuencias al menos un 95 % idénticas a cualquiera de las SEQ ID NO: 937-940), o un promotor de ubiquitina (por ejemplo, secuencias en al menos un 95 % idéntico a SEQ ID NO: 941).

Los polinucleótidos de transferencia de genes particularmente ventajosos para la transferencia de genes para expresión en células de mamíferos comprenden un intrón de los genes tempranos inmediatos 1, 2 y 3 del citomegalovirus (CMV) de células humanas o murinas (por ejemplo, secuencias al menos un 95 % idénticas a cualquiera de SEQ ID NO: 958-965), un intrón de EFla de cualquier especie de mamífero o ave, incluidos humanos, ratas, ratones, pollos y hámsteres chinos, (por ejemplo, secuencias al menos un 95 % idénticas a cualquiera de SEQ ID NO: 970-976), un intrón de EEF2 de cualquier especie de mamífero o ave, incluidos humanos, ratas, ratones, pollos y hámsteres chinos (por ejemplo, secuencias idénticas al menos en un 95 % a cualquiera de las SEQ ID NO: 989-996); un intrón de actina de cualquier especie de mamífero o ave, incluidos humanos, ratas, ratones, pollos y

hámsteres chinos (por ejemplo, secuencias al menos 95 % idénticas a cualquiera de SEQ ID NO: 977-985), un intrón GAPDH de cualquier mamífero o ave especies que incluyen humano, rata, ratón, pollo y hámster chino (por ejemplo, secuencias idénticas al menos en un 95 % a cualquiera de las SEQ ID NO: 986 o 987); un intrón que comprende el potenciador de la proteína tardía principal adenoviral, por ejemplo, secuencias al menos 95 % idénticas a SEQ ID NO: 988) o un intrón híbrido / sintético (por ejemplo, secuencias al menos 95 % idénticas a cualquiera de SEQ ID NO: 966-969).

Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden combinaciones de promotores e intrones en los que un promotor de un gen se combina con un intrón para un gen diferente, es decir, el intrón es heterólogo del promotor. Por ejemplo, un promotor de CMV temprano inmediato de ratón (por ejemplo, SEQ ID NO: 916-920) o de ser humano (por ejemplo, SEQ ID NO: 912-915) es seguido ventajosamente por un intrón de EFla (p. Ej., SEQ ID NO: 970 -976) o un intrón de EEF2 (por ejemplo, SEQ ID NO: 989-996).

Los polinucleótidos de transferencia de genes particularmente ventajosos para la transferencia de genes para expresión en células de mamíferos comprenden uno o más de un potenciador de la expresión que mejora la exportación de ARN desde el núcleo, como el elemento regulador postranscripcional de la hepatitis de la marmota (WPRE) o el regulador postranscripcional del virus de la hepatitis B elemento (HPRE) (por ejemplo, secuencias al menos 95 % idénticas a cualquiera de SEQ ID NO: 867 o 868) y elementos tales como secuencias de región de unión de andamio (SAR) (por ejemplo, secuencias al menos 95 % idénticas a cualquiera de SEQ ID NO: 869-876). Estos elementos potenciadores de la expresión son particularmente ventajosos cuando se colocan en el 3' de una secuencia a expresar. Hemos determinado que las secuencias de SAR SEQ ID NO: 869-871 mejoran la expresión de un marco de lectura abierto más cuando están dentro de la transcripción que cuando están después de la señal de poliadenilación. Esto es inesperado, ya que el papel propuesto de los SAR es unir las secuencias de ADN genómico al andamio nuclear. Las SEQ ID NO: 869-871 de SAR son particularmente beneficiosas para la expresión de un polipéptido cuando se combinan con el elemento regulador postranscripcional HPRE de origen natural SEQ ID NO: 868, por ejemplo, como en SEQ ID NO: 866. Son igualmente beneficiosos cuando se combinan con una variante modificada del elemento regulador postranscripcional HPRE SEQ ID NO: 867, que hicimos mediante la introducción de un par de mutaciones coincidentes para eliminar un sitio de restricción BfuAI sin alterar la estructura de ARN tallo-bucle del elemento, por ejemplo, como en SEQ ID NO: 1100. Probamos los efectos potenciadores de la expresión de SEQ ID NO: 1100, comparando la expresión de un gen que codifica DasherGFP de polinucleótidos que comprendían SEQ ID NO: 866 o SEQ ID NO: 1100 o ningún elemento potenciador de la expresión adicional entre el gen DasherGFP y el secuencia poliA de globina de conejo. Los polinucleótidos se integraron en el genoma de las células CHO y se midió la expresión de DasherGFP. Las SEQ ID NO: 866 y SEQ ID NO: 1100 produjeron niveles de expresión de Dasher GFP que eran al menos 110 % o al menos 120 % o al menos 200 % o al menos 500 % de la expresión conseguida sin ningún elemento. Los polinucleótidos de transferencia de genes ventajosos comprenden una secuencia que es al menos 95 % idéntica o al menos 98 % idéntica o al menos 99 % idéntica o al menos 99,5 % idéntica a SEQ ID NO: 866, o una secuencia que es SEQ ID NO: 866 o SEQ ID NO: 1100. Estos son particularmente beneficiosos cuando se combinan además con una secuencia señal de poliadenilación fuerte, por ejemplo, la señal del gen de la beta globina de conejo, por ejemplo, como en SEQ ID NO: 1101-2. Los efectos de estos elementos pueden mejorarse aún más cuando se combinan con una secuencia aislante. Se proporcionan combinaciones particularmente ventajosas como SEQ ID NO: 820-858. Un polinucleótido de transferencia de genes ventajoso comprende una secuencia que es al menos un 90 % idéntica o al menos un 95 % idéntica o al menos un 99 % idéntica a cualquiera de las SEQ ID NO: 820-858. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden un transposón *Xenopus* o *Bombyx* que comprende una secuencia que es al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a una secuencia seleccionada de SEQ ID NO: 820-858.

Los polinucleótidos de transferencia de genes particularmente ventajosos para la transferencia de un primer y un segundo gen para la coexpresión en células de mamíferos comprenden un promotor y opcionalmente un potenciador e intrones unidos operativamente al primer gen, y un elemento de acoplamiento traduccional tal como un IRES que enlaza operativamente la expresión de un segundo gen al primero. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden una secuencia IRES seleccionada de SEQ ID NO: 1050-1094.

La expresión de dos genes a partir de un único polinucleótido también se puede lograr enlazando operativamente la expresión de cada gen a un promotor separado, cada uno de los cuales puede opcionalmente enlazarse operativamente a potenciadores e intrones como se describió anteriormente. A menudo es ventajoso colocar un aislante genético como el núcleo HS4 o el núcleo D4Z4, entre los dos promotores, por ejemplo, después de la secuencia de poliadenilación unida operativamente al gen que codifica el primer polipéptido y antes del promotor unido operativamente al gen que codifica el segundo polipéptido. Vea el Ejemplo 6.2.1 y la Tabla 7, y compare la fila 12 con la fila 13, la fila 14 con la fila 15, la fila 16 con la fila 17, la fila 18 con la fila 19, la fila 20 con la fila 21 y la fila 22 con la fila 23. En cada caso, la expresión del primer polipéptido, el segundo polipéptido o ambos polipéptidos se incrementó por la presencia de una secuencia aislante intercalada entre los dos promotores.

Las combinaciones particularmente ventajosas de promotores para la expresión de dos polipéptidos incluyen configuraciones en las que un polipéptido se expresa unido operativamente al promotor EF1a o al promotor CMV y el segundo polipéptido se expresa unido operativamente al promotor CMV, el promotor GAPDH, el promotor EF1a o el promotor CMV o el promotor de actina. Las combinaciones específicas de señales de poliadenilación,

terminadores, potenciadores, promotores, intrones, 5'UTR y secuencias aislantes que funcionan bien cuando se colocan después de un gen que codifica un primer polipéptido y preceden a un gen que codifica un segundo polipéptido (es decir, en un polinucleótido espaciador) incluyen SEQ ID NO: 998-1049. Los polinucleótidos de transferencia de genes particularmente ventajosos para la transferencia de un primer y un segundo gen para la coexpresión en células de mamíferos comprenden una secuencia al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a una secuencia seleccionada de SEQ ID NO: 998-1049.

#### 5.2.10 Aumento de la expresión por selección

Pueden obtenerse altos niveles de expresión a partir de genes codificados en polinucleótidos de transferencia de genes que están integrados en regiones del genoma que son transcripcionalmente activas, o que están integradas en el genoma en múltiples copias, o que están presentes extracromosómicamente en múltiples copias.

La expresión de un primer polipéptido de expresión codificado en un polinucleótido de transferencia de genes (el "polipéptido de expresión") puede aumentarse si el polinucleótido de transferencia de genes también comprende una secuencia que codifica un polipéptido seleccionable. A menudo es ventajoso unir operativamente el gen que codifica el polipéptido seleccionable a elementos de control de la expresión que dan como resultado niveles bajos de expresión del polipéptido seleccionable del polinucleótido de transferencia de genes y/o usar condiciones que proporcionen una selección más rigurosa. En estas condiciones, para que la célula de expresión produzca niveles suficientes del polipéptido seleccionable codificado en el polinucleótido de transferencia de genes para sobrevivir a las condiciones de selección, el polinucleótido de transferencia de genes debe estar presente en una ubicación favorable en el genoma de la célula para niveles altos de expresión, o debe estar presente un número suficientemente alto de copias del polinucleótido de transferencia de genes, de modo que estos factores compensen los bajos niveles de expresión que se pueden lograr debido a los elementos de control de la expresión.

El polipéptido de expresión y el polipéptido seleccionable pueden incluirse en el mismo polinucleótido de transferencia de genes, pero enlazados operativamente a diferentes promotores. En este caso, se pueden lograr niveles bajos de expresión del marcador de selección mediante el uso de un promotor constitutivo débilmente activo como el promotor de fosfoglicerocinasa (PGK) (por ejemplo, SEQ ID NO: 937-940), el promotor de timidina cinasa del virus del herpes simple (HSV-TK) (por ejemplo, SEQ ID NO: 943), el promotor MC1 (por ejemplo, SEQ ID NO: 944), el promotor de ubiquitina (por ejemplo, SEQ ID NO: 941). Se pueden construir deliberadamente otros promotores débilmente activos, por ejemplo, un promotor atenuado por truncamiento, como un promotor SV40 truncado (por ejemplo, SEQ ID NO: 945 o 946), un promotor HSV-TK truncado (por ejemplo, SEQ ID NO: 942), o un promotor atenuado por la inserción de un 5'UTR desfavorable para la expresión entre un promotor y el gen que codifica el polipéptido seleccionable, por ejemplo SEQ ID NO: 956 o 957. Los ejemplos de promotores atenuados incluyen un promotor PGK atenuado (SEQ ID NO: 947) y un promotor HSV-TK atenuado (SEQ ID NO: 948). Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden una secuencia que es al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a cualquiera de SEQ ID NO: 937-948, operativamente enlazada a un gen que codifica un marcador de selección.

Los niveles de expresión de un marcador de selección también pueden reducirse ventajosamente mediante otros mecanismos tales como la inserción del intrón del antígeno t pequeño de SV40 después del gen del marcador de selección. El intrón t pequeño de SV40 acepta sitios de empalme 5' aberrantes y puede conducir a eliminaciones dentro del gen precedente en una fracción de los ARNm empalmados, reduciendo así la expresión del marcador de selección. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden el intrón SEQ ID NO: 997, unido operativamente a un gen que codifica un marcador de selección. Para que este mecanismo de atenuación sea eficiente, es preferible que el gen que codifica el marcador de selección comprenda un donante de intrón fuerte dentro de su región codificante. La glutamina sintasa SEQ ID NO: 703 puede estar codificada por la secuencia SEQ ID NO: 704 que comprende un donante de intrón fuerte. La puromicina acetiltransferasa SEQ ID NO: 715 puede estar codificada por la secuencia SEQ ID NO: 716 que comprende un donante de intrón fuerte. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden una secuencia al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a cualquiera de SEQ ID NO: 704 o SEQ ID NO: 716, y SEQ ID NO: 997.

Los niveles de expresión de un marcador de selección también pueden reducirse ventajosamente mediante otros mecanismos tales como la inserción de un 5'-UTR inhibidor dentro del transcrito, por ejemplo, SEQ ID NO: 956 o 957. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden un promotor unido operativamente a un gen que codifica un marcador de selección, en el que se intercala una secuencia que es al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a SEQ ID NO: 956 o 957, entre el promotor y el marcador de selección.

La Tabla 13 muestra la transposición de transposones que comprenden un marcador de selección de puromicina unido operativamente a un promotor PGK relativamente fuerte SEQ ID NO: 937; Tabla 13 filas 2-4), o a un promotor HSV-TK más débil (SEQ ID NO: 942; Tabla 13 filas 5-9). La expresión de los transposones en los que la puromicina acetil transferasa estaba unida operativamente al promotor más débil fue sustancialmente mayor que la de los transposones en los que la puromicina acetil transferasa estaba unida operativamente al promotor más fuerte. Sin

embargo, esta alta expresión requirió cotransfección del transposón con una transposasa. Al unir operativamente el marcador de selección a elementos que dan como resultado una expresión débil, se seleccionan células que incorporan múltiples copias del transposón o en las que el transposón está integrado en una ubicación genómica favorable para una alta expresión. El uso de un sistema de transferencia de genes que comprende un transposón y una transposasa correspondiente, en particular un transposón *Xenopus* y una transposasa *Xenopus* hiperactiva o un transposón de *Bombyx* y una transposasa de *Bombyx* hiperactiva aumenta la probabilidad de que se produzcan células con múltiples copias del transposón, o en las que el transposón está integrado en una ubicación genómica favorable para una alta expresión, como se muestra en los Ejemplos 6.3.1.2 y 6.3.2.2. Los sistemas de transferencia de genes que comprenden un transposón y una transposasa correspondiente son por tanto particularmente ventajosos cuando el transposón comprende un marcador de selección unido operativamente a promotores débiles, y cuando la transposasa es una transposasa hiperactiva. Los transposones particularmente ventajosos comprenden marcadores de selección unidos operativamente a un promotor con al menos 90 % de identidad o al menos 95 % de identidad o al menos 99 % de identidad a una secuencia seleccionada de SEQ ID NO: 942-948. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden secuencias con al menos un 90 % de identidad o al menos un 95 % de identidad o al menos un 99 % de identidad con una secuencia seleccionada de SEQ ID NO: 719-749.

Otra forma de seleccionar altos niveles de expresión de un primer polipéptido de expresión es acoplar traduccionalmente el gen que codifica un marcador de selección y el primer polipéptido de expresión usando un IRES. Preferentemente, el IRES da como resultado una expresión mucho más alta del primer polipéptido de expresión que el marcador de selección. Muchas actividades IRES nuevas se muestran en la Tabla 9 y se describen en el Ejemplo 6.4.1. En estos ejemplos, el primer polipéptido de expresión es una proteína verde fluorescente y el polipéptido seleccionable es una proteína roja fluorescente. Cada tabla también muestra el nivel de expresión del primer polipéptido de expresión en una construcción que carece del IRES y del gen del polipéptido seleccionable. Los elementos IRES particularmente deseables son aquellos que tienen una alta relación de expresión entre el primer polipéptido de expresión y el polipéptido seleccionable, y que también tienen niveles de expresión del primer polipéptido de expresión que están cerca de los niveles de expresión obtenidos en ausencia del IRES y un gen para el polipéptido seleccionable. En la Tabla 9, se puede ver que IRES SEQ ID NO: 1089, 1078, 1080, 1086, 1076, 1075, 1081, 1077, 1088, 1079, 1091, 1066, 1094, 1093, 1072, 1068, 1071 tienen niveles de expresión del primer polipéptido de expresión que son al menos el 50 % de los niveles de expresión obtenidos en ausencia del IRES y un segundo marco de lectura abierto en células CHO, e IRES SEQ ID NO: 1084, 1079, 1073, 1085, 1082, 1074, 1080, 1066 tienen niveles de expresión del primer polipéptido de expresión que son al menos el 50 % de los niveles de expresión obtenidos en ausencia del IRES y un segundo marco de lectura abierto en células HEK. IRES SEQ ID NO: 1091, 1070, 1069, 1090, 1094, 1077, 1067, 1089, 1068, 1078, 1066, 1072, 1093, 1092, 1079, 1080, 1081, 1052, 1074, 1085, 1076, 1088, 1075 y 1086 todos expresan el segundo ORF al 15 % o menos que el nivel del primer ORF en las células CHO. IRES SEQ ID NO: 1091, 1067, 1094, 1070, 1089, 1092, 1090, 1069, 1078, 1074, 1077, 1085, 1084, 1053, 1096 y 1073 todos expresan el segundo ORF al 20 % o menos que el nivel del primer ORF en células HEK.

Por lo tanto, estos elementos IRES son particularmente ventajosos, cuando se usan para unir la expresión de un primer polipéptido de expresión a la expresión de un gen que codifica un marcador de selección en un polinucleótido de transferencia de genes, donde el polinucleótido de transferencia de genes comprende un gen que codifica un primer polipéptido de expresión en el lado 5' del IRES y un gen que codifica un marcador de selección en el lado 3' del IRES. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden marcadores de selección unidos operativamente a un IRES seleccionado de SEQ ID NO: 1052, 1053, 1066, 1068, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1084, 1085, 1086, 1088, 1089, 1090, 1091, 1093 o 1094.

Los polipéptidos seleccionables comunes confieren resistencia de las células eucariotas a antibióticos como la neomicina (resistencia conferida por una aminoglucósido 3'-fosfotransferasa, por ejemplo, SEQ ID NO: 709-712), puromicina (resistencia conferida por puromicina acetiltransferasa, por ejemplo, SEQ ID NO: 713-716), blasticidina (resistencia conferida por una blasticidina acetiltransferasa y una blasticidina desaminasa), higromicina B (resistencia conferida por higromicina B fosfotransferasa, por ejemplo, SEQ ID NO: 717-718 y zeocina (resistencia conferida por la proteína de unión, por ejemplo, SEQ ID NO: 702). Otros polipéptidos seleccionables incluyen aquellos que son fluorescentes (tales como GFP, RFP, etc.) y, por lo tanto, pueden seleccionarse, por ejemplo, usando citometría de flujo. Otros polipéptidos seleccionables incluyen proteínas transmembrana que pueden unirse a una segunda molécula (proteína o molécula pequeña) que puede marcarse con fluorescencia de modo que la presencia de la proteína transmembrana puede seleccionarse, por ejemplo, mediante el uso de citometría de flujo.

Por ejemplo, la glutamina sintasa (GS, por ejemplo las SEQ ID NO: 703 y 705) se usa como marcador de selección que permite la selección mediante el metabolismo de la glutamina. La glutamina sintasa es la enzima responsable de la biosíntesis de glutamina a partir del glutamato y el amoníaco, y es un componente crucial de la única vía de formación de glutamina en una célula de mamífero. En ausencia de glutamina en el medio de crecimiento, la enzima GS es esencial para la sobrevivencia de las células de mamíferos en cultivo. Algunas líneas celulares, por ejemplo las células de mieloma de ratón, no expresan suficiente enzima GS para sobrevivir sin adición de glutamina. En estas células, un gen de GS transfectado puede funcionar como un marcador de selección al permitir el crecimiento en un medio libre de glutamina. En otras líneas celulares, por ejemplo, las células de ovario de hámster chino (CHO)

expresan suficiente GS para sobrevivir sin glutamina añadida de forma exógena. Estas líneas celulares pueden manipularse mediante técnicas de edición del genoma que incluyen CRISPR/Cas9 para reducir o eliminar la actividad de la enzima GS. En todos estos casos, los inhibidores de GS como la metionina sulfoximina (MSX) se pueden usar para inhibir la actividad de GS endógena de una célula. Los protocolos de selección incluyen introducir una construcción que comprende secuencias que codifican un primer polipéptido y un marcador de selección de glutamina sintasa, y luego tratar la célula con inhibidores de glutamina sintasa como metionina sulfoximina. Mientras más altos sean los niveles de metionina sulfoximina que se usan, mayor será el nivel de expresión de glutamina sintasa que se requiere para permitir que la célula sintetice suficiente glutamina para sobrevivir. Algunas de estas células también mostrarán una expresión aumentada del primer polipéptido.

Preferentemente, el gen GS está operativamente ligado a un promotor débil u otros elementos de secuencia que atenúan la expresión como se describe en la presente descripción, de modo que los altos niveles de expresión solo pueden ocurrir si están presentes muchas copias del polinucleótido de transferencia de genes, o si están integradas en una posición en el genoma donde ocurren altos niveles de expresión.

Un segundo sistema para aumentar la expresión por selección usa la enzima dihidrofolato reductasa (DHFR, por ejemplo SEQ ID NO: 707 o 708) que se requiere para catalizar la reducción de 5,6-dihidrofolato (DHF) a 5,6,7,8-tetrahidrofolato (THF) y se usa como marcador de selección. Algunas líneas celulares no expresan suficiente DHFR para sobrevivir sin THF añadido. En estas células, un gen de DHFR transfectado puede funcionar como un marcador de selección al permitir el crecimiento en un medio libre de THF. Pueden producirse líneas celulares deficientes en DHFR, por ejemplo, células de ovario de hámster chino (CHO) mediante técnicas de edición del genoma que incluyen CRISPR/Cas9 para reducir o eliminar la actividad de la enzima DHFR endógena. La DHFR confiere resistencia al metotrexato (MTX). La DHFR puede inhibirse por niveles más altos de metotrexato. Los protocolos de selección incluyen introducir una construcción que comprende secuencias que codifican un primer polipéptido y un marcador de selección de DHFR en una célula con o sin un gen de DHFR endógeno, y luego tratar la célula con inhibidores de DHFR tales como metotrexato. Mientras más altos sean los niveles de metotrexato que se usan, mayor será el nivel de expresión de DHFR que se requiere para permitir que la célula sintetice suficiente DHFR para sobrevivir. Algunas de estas células también mostrarán una expresión aumentada del primer polipéptido. Preferentemente, el gen DHFR está operativamente ligado a un promotor débil u otros elementos de secuencia que atenúan la expresión como se describió anteriormente, de modo que los altos niveles de expresión solo pueden ocurrir si están presentes muchas copias del polinucleótido de transferencia de genes, o si están integradas en una posición en el genoma donde ocurren altos niveles de expresión.

La combinación de marcador de selección y sus elementos de control unidos operativamente afectan profundamente la expresión que se puede obtener a partir de un polinucleótido de transferencia de genes. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden una secuencia que es al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a una secuencia seleccionada de SEQ ID NO: 719-749. Preferentemente, estas secuencias están flanqueadas por un par de extremos del transposón.

Como se muestra en la Tabla 15 y se describe en el Ejemplo 6.3.2.2, la combinación de marcador de selección y sus elementos de control enlazados operativamente afectan profundamente la expresión que puede obtenerse de un segundo promotor en un polinucleótido de transferencia de genes. Estos efectos también están influenciados por la presencia de secuencias aislantes en el polinucleótido de transferencia de genes. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden una secuencia que es al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a una secuencia seleccionada de SEQ ID NO: 751-819. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden un transposón *Xenopus* o *Bombyx* que comprende una secuencia que es al menos 90 % idéntica o al menos 95 % idéntica o al menos 99 % idéntica a una secuencia seleccionada de SEQ ID NO: 751-819.

El uso de transposones y transposasas junto con tales marcadores de selección que se requieren para el metabolismo celular normal tiene varias ventajas sobre las construcciones sin transposones. Una es que el enlace entre la expresión del primer polipéptido y el marcador de selección es mejor para los transposones, porque una transposasa integrará la secuencia completa que se encuentra entre los dos extremos del transposón en el genoma. Por el contrario, cuando se introduce ADN heterólogo en el núcleo de una célula eucariota, por ejemplo, una célula de mamífero, se rompe gradualmente en fragmentos aleatorios que pueden integrarse en el genoma de la célula o degradarse. Por tanto, si una construcción que comprende secuencias que codifican un primer polipéptido y un marcador de selección requerido para el metabolismo celular normal se introduce en una población de células, algunas células integrarán las secuencias que codifican el marcador de selección pero no las que codifican el primer polipéptido y viceversa. La selección de células que expresan altos niveles de marcador de selección está, por tanto, solo algo correlacionada con células que también expresan altos niveles del primer polipéptido. Por el contrario, debido a que la transposasa integra todas las secuencias entre los extremos del transposón, es muy probable que las células que expresan altos niveles de marcador de selección también expresen altos niveles del primer polipéptido.

Una segunda ventaja de los transposones y transposasas es que son mucho más eficientes en la integración de secuencias de ADN en el genoma. Por tanto, es probable que una fracción mucho mayor de la población celular

reciba una o más copias de la construcción en sus genomas, por lo que habrá una probabilidad correspondientemente mayor de una buena expresión estable tanto del marcador de selección como del primer polipéptido.

- 5 Por tanto, puede usarse un transposón que comprende una secuencia que codifica un primer polipéptido y un marcador de selección que puede ser inhibido por un inhibidor de molécula pequeño para obtener células que expresan niveles elevados del primer polipéptido. El primer polipéptido puede ser parte de un anticuerpo. Los marcadores de selección preferidos son la glutamina sintasa y la DHFR.
- 10 Puede seleccionarse un mayor número de transposones integrados mediante el uso de marcadores de selección requeridos para el metabolismo celular normal como DHFR o glutamina sintasa.

### 5.3 Kits

- 15 La presente descripción también presenta kits que comprenden una transposasa de *Bombyx* como una proteína o codificada por un ácido nucleico y/o un transposón de *Bombyx*; o un sistema de transferencia de genes como se describe en la presente descripción que comprende una transposasa de *Bombyx* como una proteína o codificada por un ácido nucleico como se describe en la presente descripción, en combinación con un transposón de *Bombyx*; opcionalmente junto con un portador, adyuvante o vehículo farmacéuticamente aceptable, y opcionalmente con instrucciones de uso. Cualquiera de los componentes del kit puede administrarse y/o transfectarse en células en un orden posterior o en paralelo, por ejemplo una proteína transposasa de *Bombyx* o su ácido nucleico codificante se puede administrar y/o transfectar en una célula como se definió anteriormente antes, simultáneamente con o después de la administración y/o transfección de un transposón de *Bombyx*. Alternativamente, un transposón de *Bombyx* se puede transfectar en una célula como se definió anteriormente antes, simultáneamente con o después de la transfección de una proteína transposasa de *Bombyx* o su ácido nucleico codificante. Si se transfecta en paralelo, preferentemente ambos componentes se proporcionan en una formulación separada y/o se mezclan entre sí directamente antes del suministro para evitar la transposición antes de la transfección. Además, el suministro y/o transfección de al menos un componente del kit puede ocurrir en un modo escalonado en el tiempo, por ejemplo, suministrando múltiples dosis de este componente.

- 30 Además, la presente descripción también presenta kits que comprenden una transposasa de *Xenopus* como una proteína o codificada por un ácido nucleico y/o un transposón de *Xenopus*; o un sistema de transferencia de genes como se describe en la presente descripción que comprende una transposasa de *Xenopus* como una proteína o codificada por un ácido nucleico como se describe en la presente descripción, en combinación con un transposón de *Xenopus*; opcionalmente junto con un portador, adyuvante o vehículo farmacéuticamente aceptable, y opcionalmente con instrucciones de uso. Cualquiera de los componentes del kit puede administrarse y/o transfectarse en células en un orden posterior o en paralelo, por ejemplo una proteína transposasa de *Xenopus* o su ácido nucleico codificante puede administrarse y/o transfectarse en una célula como se definió anteriormente antes, simultáneamente con o después de la administración y/o transfección de un transposón de *Xenopus*. Alternativamente, un transposón de *Xenopus* se puede transfectar en una célula como se definió anteriormente antes, simultáneamente con o después de la transfección de una proteína transposasa de *Xenopus* o su ácido nucleico codificante. Si se transfecta en paralelo, preferentemente ambos componentes se proporcionan en una formulación separada y/o se mezclan entre sí directamente antes de la administración para evitar la transposición antes de la transfección. Además, el suministro y/o transfección de al menos un componente del kit puede ocurrir en un modo escalonado en el tiempo, por ejemplo, suministrando múltiples dosis de este componente.

## 6. Ejemplos

- 50 Los siguientes ejemplos ilustran los métodos, composiciones y kits descritos en la presente descripción y no deben interpretarse como limitantes de ninguna manera. Varios equivalentes resultarán evidentes a partir de los siguientes ejemplos.

### 6.1 Transposasas

#### 55 6.1.1 Transposasas derivadas de *xenopus* y *bombyx*

- Unir un par de extremos del transposón en los extremos de un polinucleótido heterólogo puede crear un transposón sintético que puede integrarse en un genoma diana mediante una transposasa. La Tabla 1 muestra las configuraciones de 4 transposones sintéticos diferentes creados al unir el extremo del transposón cuya SEQ ID NO. se da en la columna A a un lado de la construcción indicadora de SEQ ID NO: 39, uniendo el extremo del transposón cuya SEQ ID NO se da en la columna B al otro lado de la construcción indicadora y flanqueando ambos por la secuencia diana dada en la columna C. Estos transposones luego se transfectaron en células CHO-K1 junto con genes que codifican transposasas cuya SEQ ID NO se muestra en la columna F, operativamente enlazada a la del promotor de CMV. La cantidad de cada ADN en cada transfección se muestra en las columnas E (transposón) y H (transposasa) de la Tabla 1.

Se transfectaron células CHO-K1 y se seleccionaron con puromicina como se describe en la Sección 4.2.1. La fluorescencia se midió a Ex/Em de 488/518 nm, y es una medida de expresión del ORF que codifica el indicador fluorescente a partir de transposones integrados de manera estable, la fluorescencia de 3 transfecciones independientes se muestra en la Tabla 1, columnas JL.

Las filas 3, 5, 10 y 15 de la Tabla 1 muestran los resultados de tres transposones diferentes transfectados en células CHO sin ninguna transposasa cotransfectada. En cada caso, hubo pocas o ninguna célula viva que sobrevivió a la selección con puromicina (columna I), y ninguna fluorescencia de Dasher GFP (columnas J, K y L), lo que indica que los transposones no se habían integrado o no se habían integrado en una forma que permitió la expresión posterior de los genes codificados en los transposones.

La Tabla 1, filas 3 y 4, compara la fluorescencia obtenida de un transposón con extremos tomados del transposón piggyBac de la oruga de la col *Trichoplusia ni* (SEQ ID NO: 35 y 36), ya sea transfectado solo (fila 3) o cotransfectado con un plásmido que lleva un gen que codifica la transposasa piggyBac hiperactiva (SEQ ID NO698) unida operativamente al promotor de CMV (fila 4). La cotransfección con el gen de la transposasa aumentó la viabilidad celular para dar un 100 % de confluencia y la señal fluorescente aumentó desde el fondo a -660 unidades.

La Tabla 1 filas 5-14 compara la fluorescencia obtenida de un transposón con extremos SEQ ID NO: 1 y 11, ya sea transfectado solo (fila 5) o cotransfectado con un gen que codifica la transposasa de *Xenopus* SEQ ID NO: 49 solo (fila 7) o fusionado a una señal de localización nuclear heteróloga (fila 6), o cotransfectado con un gen que codifica la transposasa de *Xenopus* SEQ ID NO: 48 solo (fila 9) o fusionado a una señal de localización nuclear heteróloga (fila 8); o un transposón con extremos SEQ ID NO: 3 y 12, ya sea transfectado solo (fila 10) o cotransfectado con un gen que codifica la transposasa de *Xenopus* SEQ ID NO: 49 solo (fila 12) o fusionado a una señal de localización nuclear heteróloga (fila 11), o cotransfectado con un gen que codifica la transposasa de *Xenopus* SEQ ID NO: 48 solo (fila 14) o fusionado a una señal de localización nuclear heteróloga (fila 13). La cotransfección con transposasa fusionada a una señal de localización nuclear aumentó la viabilidad celular para dar un 100 % de confluencia y la señal fluorescente aumentó desde el fondo a -1000 unidades. En ausencia de la señal de localización nuclear, las células viables y los niveles de expresión eran inferiores al 10 % de los valores obtenidos con las transposasas fusionadas a señales de localización nuclear heterólogas. Por tanto, se requieren señales de localización nuclear heterólogas para las transposasas de *Xenopus* de origen natural (por ejemplo, SEQ ID NO: 48 y 49) para integrar eficientemente los transposones en el núcleo de las células de mamíferos de una manera que permita la expresión posterior de los genes codificados en los transposones.

Los datos de la Tabla 1 muestran que, cuando se fusionan con una señal de localización nuclear heteróloga, las transposasas de *Xenopus* SEQ ID NO: 48 y 49 son activas en la transposición de transposones de *Xenopus* sintéticos al genoma de una célula de mamífero. Cada uno de estos extremos del transposón contiene un ITR con la secuencia 5'-CCYTTTBMCTGCCA -3'(SEQ ID NO: 19), donde los ITR se encuentran en los dos extremos en una orientación invertida entre sí. También muestra que la fusión de estas transposasas a una señal de localización nuclear heteróloga es más activa en este ensayo que la transposasa piggyBac hiperactiva derivada de la oruga de la col *Trichoplusia ni*.

La Tabla 1 filas 15-19 compara la fluorescencia obtenida de un transposón con extremos SEQ ID NO: 23 y 29, ya sea transfectado solo (fila 15) o cotransfectado con un gen que codifica la transposasa de *Bombyx* SEQ ID NO: 750 solo (fila 17) o fusionado a una señal de localización nuclear heteróloga (fila 16), o cotransfectado con un gen que codifica la transposasa de *Bombyx* SEQ ID NO: 407 solo (fila 19) o fusionado a una señal de localización nuclear heteróloga (fila 18). La cotransfección con transposasa SEQ ID NO: 407, se fusionó o no a una señal de localización nuclear aumentó la viabilidad celular para dar un 100 % de confluencia, y la señal fluorescente aumentó desde el fondo a -1000 unidades. La cotransfección con transposasa SEQ ID NO: 750, se fusionó o no a una señal de localización nuclear dio como resultado células viables y los niveles de expresión fueron menos del 1 % de los valores obtenidos con la transposasa SEQ ID NO: 407. Por tanto, la transposasa SEQ ID NO: 407 es activa en la transposición de transposones *Bombyx* sintéticos al genoma de una célula de mamífero.

### 6.1.2 Transposones derivados de *xenopus* y *bombyx*

#### 6.1.2.1 Extremos del transposón de *Bombyx*

Los extremos de los transposones de origen natural se modificaron mediante truncamiento o cambiando las secuencias diana. Estos extremos del transposón se unieron luego a los extremos de un polinucleótido heterólogo para crear transposones sintéticos que pueden integrarse en un genoma diana mediante una transposasa. La Tabla 2 muestra las configuraciones de 12 transposones sintéticos diferentes creados al unir el extremo del transposón cuya SEQ ID NO se da en la columna A a un lado de la construcción indicadora SEQ ID NO: 39, uniendo el extremo del transposón cuya SEQ ID NO se da en la columna B al otro lado de la construcción indicadora y flanqueando ambos por la secuencia diana dada en la columna C. Estos transposones se transfectaron luego en células CHO-K1 junto con genes que codifican transposasas cuya SEQ ID NO se muestra en la columna G, opcionalmente fusionadas a una señal de localización nuclear heteróloga (como se muestra en la columna H) y operativamente enlazada al promotor de CMV. La cantidad de cada ADN en cada transfección se muestra en las columnas F (transposón) e I (transposasa) de la Tabla 2. La transfección y la selección fueron las descritas en la Sección 4.2.1.

Las células se recolectaron raspando y midiendo en un lector de placas fluorimétrico, la fluorescencia de 3 transfecciones independientes se muestra en la Tabla 2, columnas JL. La fluorescencia se midió a Ex/Em de 488/518 nm, y es una medida de expresión del ORF que codifica el indicador fluorescente DasherGFP a partir de transposones integrados de manera estable.

Los extremos del transposón izquierdo y derecho de *Bombyx* podrían truncarse desde el extremo proximal (que es el extremo más alejado del ITR) mientras se conserva la función del transposón. La Tabla 2 filas 2-9 muestra que la expresión de polinucleótidos heterólogos insertados en el genoma de CHO se mejoró mediante cotransfección con una construcción que codifica la transposasa de *Bombyx* SEQ ID NO: 407 en donde los polinucleótidos heterólogos comprendían un extremo del transposón izquierdo de una secuencia diana seguida de SEQ ID NO: 23, 24 o 25, y un extremo derecho del transposón de SEQ ID NO: 29 o 31 seguido de una secuencia diana.

La prueba realizada aquí muestra que estos extremos comprenden todas las secuencias necesarias para crear un transposón de *Bombyx* que se puede integrar en el genoma de una célula diana. Sin embargo, se ha demostrado previamente para el transposón piggyBac de oruga de la col que se requieren secuencias más largas para la transformación de los genomas diana que para la escisión del transposón por la transposasa, o para la transposición entre plásmidos, como se describe en Li y otros, (2005) *Insect Mol. Biol.* 14: 17-30. "piggyBac internal sequences are necessary for efficient transformation of target genomes." y Li y otros, (2001) *Mol Genet Genomics* 266: 190-8. "The minimum internal and external sequence requirements for transposition of the eukaryotic transformation vector piggyBac." Inferimos que las secuencias más cortas del transposón de *Bombyx* también serán competentes para la escisión o para la transposición entre plásmidos. Las secuencias importantes para la escisión del transposón piggyBac de la oruga de la col son las repeticiones terminales y las repeticiones internas en cada extremo. El transposón de *Bombyx* comprende varias repeticiones internas que probablemente realizan funciones análogas. La SEQ ID NO: 25 del extremo izquierdo de *Bombyx* comprende la SEQ ID NO: 1103 y una copia invertida de esta SEQ ID NO: 1104; también comprende la SEQ ID NO: 1105 y una copia invertida de esta SEQ ID NO: 1106; también comprende dos palíndromos ricos en AT SEQ ID NO: 1107 y SEQ ID NO: 1108. La SEQ ID NO: 31 del extremo derecho de *Bombyx* comprende dos copias de la secuencia rica en AT SEQ ID NO: 1110. La SEQ ID NO: 31 del extremo derecho de *Bombyx* también comprende una copia de la SEQ ID NO: 1106, que se encuentra repetida en ambas orientaciones en la SEQ ID NO: 25 del extremo izquierdo. La SEQ ID NO: 25 del extremo izquierdo de *Bombyx* y la SEQ ID NO: 31 del extremo derecho también comprenden cada una copia de la SEQ ID NO: 1109. Un transposón de *Bombyx* puede comprender un extremo izquierdo que comprende 1 o 2 o 3 o 4 o 5 o 6 o 7 secuencias seleccionadas de SEQ ID NO: 1103-1110. Un transposón de *Bombyx* puede comprender un extremo derecho que comprende 1 o 2 o 3 secuencias seleccionadas de SEQ ID NO: 1106 y 1109-1110.

También descubrimos que podíamos cambiar la secuencia diana 5'-TTAT-3' que flanquea el transposón basado en *Bombyx* a 5'-TTAA-3' y aun así obtener una señal DasherGFP alta dependiente de transposasa (compare las filas 3 y 11 en la Tabla 2). Por tanto, una transposasa de *Bombyx* es eficiente en la transposición de transposones con diferentes secuencias diana, incluidas las secuencias diana 5'-TTAT-3' y 5'-TTAA-3'.

#### 6.1.2.2 Extremos del transposón *Xenopus*

Las Tablas 2 y 3 también muestran la expresión de transposones de *Xenopus* con extremos truncados o secuencias diana modificadas. Ambas tablas muestran las configuraciones de transposones sintéticos creados al unir el extremo del transposón cuya SEQ ID NO. se da en la columna A a un lado de la construcción indicadora SEQ ID NO: 39, uniendo el extremo del transposón cuya SEQ ID NO. se da en la columna B para el otro lado de la construcción indicadora y flanqueando ambos por la secuencia diana dada en la columna C. Estos transposones se transfectaron luego en células CHO-K1 junto con genes que codifican transposasas cuya SEQ ID NO. se muestra en la columna G, opcionalmente fusionados a una señal de localización nuclear heteróloga (como se muestra en la columna H) y operativamente enlazada al promotor de CMV. La cantidad de cada ADN en cada transfección se muestra en las columnas F (transposón) e I (transposasa) de las Tablas 2 y 3. La transfección y la selección fueron las descritas en la Sección 4.2.1.

Las células se recogieron raspando y midiendo en un lector de placas fluorimétrico, la fluorescencia de 3 transfecciones independientes se muestra en las Tablas 2 y 3 columnas JL. La fluorescencia se midió a Ex/Em de 488/518 nm y es una medida de la expresión del ORF que codifica el indicador fluorescente DasherGFP a partir de transposones integrados de manera estable.

La Tabla 2 filas 13-20 y la Tabla 3 filas 2-11 muestran que la expresión de polinucleótidos heterólogos insertados en el genoma de CHO se mejoró mediante cotransfección con una construcción que codifica la transposasa de *Xenopus* SEQ ID NO: 48 fusionada a una secuencia de localización nuclear, en la que la Los polinucleótidos heterólogos comprendían los extremos de un transposón izquierdo de una secuencia diana seguido de la SEQ ID NO: 1 o 3-7 y un extremo derecho del transposón de la SEQ ID NO: 11-13 o 15-16 seguido de una secuencia diana.

La prueba realizada aquí muestra que estos extremos comprenden todas las secuencias necesarias para crear un transposón de *Xenopus* que se puede integrar en el genoma de una célula diana. Sin embargo, se ha demostrado

previamente para el transposón piggyBac de oruga de la col que se requieren secuencias más largas para la transformación de los genomas diana que para la escisión del transposón por la transposasa, o para la transposición entre plásmidos, como se describe en Li y otros (2005) *Insect Mol. Biol.* 14: 17-30. "piggyBac internal sequences are necessary for efficient transformation of target genomes." y Li y otros, (2001) *Mol Genet Genomics* 266: 190-8. "The minimum internal and external sequence requirements for transposition of the eukaryotic transformation vector piggyBac." Inferimos que las secuencias más cortas del transposón de *Xenopus* también serán competentes para la escisión o para la transposición entre plásmidos.

También descubrimos que podíamos cambiar una secuencia objetivo 5'-TTAA-3' que flanqueaba el transposón basado en *Xenopus* a una secuencia objetivo 5'-TTAT-3' y aun así obtener una señal DasherGFP alta dependiente de transposasa (comparar filas 14 y 22 en la Tabla 2). Los ITR para estos transposones eran adyacentes a la secuencia diana izquierda, secuencia SEQ ID NO: 20 (5'-CCCTTGCCTGCCA-3'), y adyacentes a la secuencia diana derecha, secuencia SEQ ID NO: 21 (5'-TGGCAGTGAAGGG-3').

Por tanto, una transposasa de *Xenopus* es eficiente en la transposición de transposones con diferentes secuencias diana, incluidas las secuencias diana 5'-TTAT-3' y 5'-TTAA-3'.

También descubrimos que podíamos cambiar la secuencia diana 5'-TTAA-3' que flanquea el transposón basado en *Trichoplusia ni* a 5'-TTAT-3' y aun así obtener una señal DasherGFP alta dependiente de la transposasa (compare las filas 24 y 26 en la Tabla 2). Por tanto, la transposasa piggyBac de *Trichoplusia ni* es eficiente en la transposición de transposones con diferentes secuencias diana, incluidas las secuencias diana 5'-TTAT-3' y 5'-TTAA-3'.

### 6.1.3 Integración de transposon mediante transposasa proporcionada como ARNm

#### 6.1.3.1 ARNm de Transposasa de *Xenopus*

Puede proporcionarse una transposasa como una proteína o como un polinucleótido que codifica la transposasa; el polinucleótido codificante puede ser ADN o ARN expresable. Se preparó ARNm que codifica la transposasa de *Xenopus* SEQ ID NO: 48 fusionado a una NLS heteróloga como se describe en la Sección 4.2.3.

Los extremos del transposón se unieron a los extremos de un polinucleótido heterólogo para crear transposones sintéticos: los extremos del transposón cuyas SEQ ID NO. se dan en las columnas A y B de la Tabla 5 se unieron a cada lado de la construcción indicadora SEQ ID NO: 39, y flanqueados por Secuencias diana 5'-TTAA-3'. Estos transposones se transfectaron en células CHO-K1 junto con polinucleótidos que codifican transposasas cuya SEQ ID NO. se muestra en la columna E de la Tabla 5. Si el polinucleótido era ADN, el gen que codificaba la transposasa estaba operativamente ligado al promotor indicado en la columna G, y la cantidad de ADN del gen de la transposasa por transfección se indica en la columna H. Si el polinucleótido se proporcionó como ARNm, la cantidad de ARN por la transfección se indica en la columna I. La cantidad de cada ADN de transposón en cada transfección se muestra en la columna D de la Tabla 5. La transfección y selección fue como se describe en la Sección 4.2.1.

Las células se recogieron raspando y midiendo en un lector de placas fluorimétrico. La fluorescencia, mostrada en las columnas JL, se midió a Ex/Em de 488/518 nm, y es una medida de expresión del ORF que codifica el indicador fluorescente DasherGFP a partir de transposones integrados de manera estable.

La Tabla 5 muestra que la cotransfección de un transposón con extremos del transposón de *Xenopus* (cada uno de los cuales comprende la SEQ ID NO: 19) junto con el ARNm que codifica una transposasa de *Xenopus* fusionada a una NLS heteróloga, dio como resultado aumentos de hasta 50 veces en la expresión con respecto a las células transfectadas con transposón solo, y comparable a la potenciación de la expresión obtenida cuando la transposasa se proporcionó codificada en el ADN (compare las filas 9-12 con las filas 5-8 y la fila 4. También compare las filas 18-21 con las filas 14-17 y la fila 13). Por tanto, se puede proporcionar una transposasa de *Xenopus* como ARNm que se puede traducir en la célula diana.

#### 6.1.3.2 ARNm de Transposasa *Bombyx*

Se realizó un experimento similar al descrito en 6.1.3.1 con ARNm que codifica la transposasa 407 de *Bombyx* fusionado a un NLS heterólogo, también preparado por transcripción *in vitro* como se describe en la Sección 4.2.3.

Los extremos de los transposones se unieron a los extremos de un polinucleótido heterólogo para crear transposones sintéticos que pueden integrarse en un genoma diana mediante una transposasa. La Tabla 6 muestra las configuraciones de 3 transposones sintéticos diferentes. Los transposones tenían una secuencia diana 5'-TTAA-3' seguida de un extremo del transposón cuya SEQ ID NO. se da en la columna A, seguido de la construcción indicadora cuya SEQ ID NO. se da en la columna B, seguido del extremo del transposón cuya SEQ ID NO. se proporciona en la columna C, seguido de una secuencia diana 5'-TTAA-3'.

Estos transposones se transfectaron en células CHO-K1 junto con polinucleótidos que codifican transposasas cuya SEQ ID NO. se muestra en la columna F de la Tabla 6. Si el polinucleótido era ADN, el gen que codifica la

transposasa estaba operativamente ligado al promotor de CMV, y la cantidad de ADN del gen de la transposasa por transfección se indica en la columna H. Si el polinucleótido se proporcionó como ARNm, se indica la cantidad de ARN por transfección en la columna I. La cantidad de cada ADN de transposón en cada transfección se muestra en la columna E de la Tabla 6. La transfección y la selección fueron las descritas en la Sección 4.2.1. Las células se recogieron raspando y midiendo en un lector de placas fluorimétrico. La fluorescencia, que se muestra en las columnas J-L, se midió a Ex/Em de 488/518 nm, y es una medida de expresión del marco de lectura abierto (ORF) que codifica el indicador fluorescente DasherGFP a partir de transposones integrados de manera estable.

La Tabla 6 muestra que la cotransfección de un transposón con extremos del transposón de *Bombyx* (cada uno de los cuales comprende la SEQ ID NO: 33) junto con el ARNm que codifica una transposasa de *Bombyx* fusionada a una NLS heteróloga, dio como resultado aumentos de hasta 100 veces en la expresión con respecto a las células transfectadas con transposón solo, y comparable a la mejora de la expresión obtenida cuando la transposasa se proporcionó codificada en el ADN (compare las filas 5-8 con las filas 9-11 y la fila 4. También compare las filas 13-16 con las filas 17-19 y la fila 12). Por tanto, se puede proporcionar una transposasa de *Bombyx* como ARNm que se puede traducir en la célula diana.

#### 6.1.3.3 ARNm de transposasa hiperactiva de *Xenopus*

Es ventajoso proporcionar una transposasa como ARN expresable, ya que esto evita cualquier posibilidad de que el gen de la transposasa pueda integrarse en el genoma diana. Se prepararon ARN mensajero que codificaban transposasas hiperactivas de *Xenopus*, SEQ ID NO: 168, 189 y 175, fusionadas a una NLS heteróloga mediante transcripción *in vitro* mediante el uso de ARN polimerasa T7 como se describe en la Sección 4.2.3.

Un transposón comprendía secuencias diana 5'-TTAA-3', secuencias terminales de transposón SEQ ID NO: 2 y 12, y un potenciador de CMV y un promotor de CMV operativamente ligado a un gen que codifica DasherGFP. El transposón también comprendía el promotor de fosfoglicerato cinasa (PGK) murino, SEQ ID NO: 937, unido operativamente a un gen que codifica puromicina N-acetil transferasa.

El transposón se transfectó en células en suspensión CHO-S junto con polinucleótidos que codifican transposasas cuya SEQ ID NO. se muestra en la columna C de la Tabla 12. La cantidad de ARNm de transposasa por transfección se indica en la columna F. La cantidad de ADN de cada transposón en cada transfección se muestra en la columna E de la Tabla 12.

Las células CHO-S se transfectaron como se describe en la Sección 4.2.2. La selección con puromicina (50 µg/ml) se llevó a cabo durante 10 días y las células se cultivaron durante 5 días después de la selección con puromicina con dos pases y cambios de medio. Las células se recogieron pipeteando directamente en una placa fluorimétrica y se midieron en un lector de placas fluorimétricas. La fluorescencia se midió a Ex/Em de 488/518 nm, y es una medida de expresión del ORF que codifica el indicador fluorescente DasherGFP a partir de transposones integrados de manera estable.

La Tabla 12 muestra que no hubo diferencia significativa en la fluorescencia entre las muestras a las 72 horas después de la transfección. Sin embargo, después de 10 días de tratamiento con puromicina, la fluorescencia de todos los transposones que se habían cotransfectado con un ARNm de transposasa hiperactiva era entre 5 y 30 veces más brillante que la de los transposones cotransfectados con ARNm que codifica la transposasa de *Xenopus* natural SEQ ID NO: 48. Después de 5 días de recuperación, todas las cotransfecciones de transposasa hiperactivas todavía superaban a la SEQ ID NO: 48. Por tanto, las transposasas hiperactivas identificadas mediante el uso de un cribado funcional en *Saccharomyces cerevisiae* conducen a tiempos de recuperación reducidos y expresión aumentada de transposones en células de mamíferos, y las transposasas hiperactivas de *Xenopus* pueden proporcionarse como ARNm que puede traducirse en la célula diana.

#### 6.1.3.4 ARNm de transposasa hiperactiva de *Xenopus* y *Bombyx*

El ARN mensajero que codifica las transposasas hiperactivas de *Xenopus* SEQ ID NO: 175 y 189, fusionadas a un NLS heterólogo, o las transposasas *Bombyx* SEQ ID NO: 407 y 1098 se prepararon mediante transcripción *in vitro* mediante el uso de T7 RNA polimerasa como se describe en la Sección 4.2.3.

Los transposones comprendían secuencias terminales de transposones con las SEQ ID NO. dadas en las columnas B y C de la Tabla 13. Los transposones comprendían además un potenciador de CMV y un promotor de CMV operativamente ligado a un gen que codifica DasherGFP, y un gen que codifica puromicina acetil transferasa operativamente ligado a un promotor con la SEQ ID NO. mostrada en la columna F. Los transposones (750 ng) se cotransfectaron con 250 ng ARNm que codifica una transposasa con SEQ ID NO. mostrada en la columna G, fusionado a una señal de localización nuclear heteróloga.

Se transfectaron células CHO-S (de ATCC) como se describe en la Sección 4.2.2. La selección con puromicina (10 µg/ml) se llevó a cabo durante 10 días, con un cambio de medio completo en medio fresco que contenía puromicina después de 5 días. Después de 10 días, las células se transfirieron a medio fresco que contenía 25 µg/ml de

puromicina durante 4 días. Las células se cultivaron durante 5 días después de la selección con puromicina con dos pases y cambios de medio. Las células se recogieron pipeteando directamente en una placa fluorimétrica y se midieron en un lector de placas fluorimétricas. La fluorescencia se midió a Ex/Em de 488/518 nm, y es una medida de expresión del ORF que codifica el indicador fluorescente a partir de transposones integrados de manera estable, las mediciones por triplicado se muestran en las columnas HJ de la Tabla 13.

La Tabla 13 muestra que los ARNm que codifican las transposasas hiperactivas de *Xenopus* o *Bombyx* son activos en sus respectivos transposones.

## 6.2 Construcciones de polinucleótidos para la expresión de múltiples proteínas

### 6.2.1 Transposones con configuraciones de doble promotor o ires (1)

Pueden usarse transposones y transposasas para integrar eficientemente construcciones de polinucleótidos que codifican dos o más polipéptidos en el genoma de una célula. Los transposones son útiles para este propósito porque una transposasa generalmente integrará todo el ADN entre los dos extremos del transposón durante el proceso de transposición. Esto significa que es más probable que se conserven los elementos de secuencia que están configurados para lograr una relación específica de expresión entre los diferentes genes codificados que si se insertan fragmentos aleatorios del polinucleótido en el genoma de la célula diana. También significa que todos los genes codificados se integrarán en cada evento de integración, lo que es útil con construcciones polinucleotídicas más grandes.

La Tabla 7 muestra las configuraciones de un conjunto de transposones que comprenden secuencias diana 5'-TTAA-3' y extremos del transposón SEQ ID NO: 2 y 12. Los transposones comprendían genes que codifican DasherGFP y CayenneRFP, excepto para las filas 1 y 2 en las que solo estaba presente una proteína fluorescente (columnas C y D). Cuando ambos genes de proteínas fluorescentes estaban presentes, el gen DasherGFP siempre aparecía primero. Todos los transposones también comprendían un gen de resistencia a puromicina unido operativamente al promotor PGK y transcrito en la dirección opuesta a los genes de las proteínas fluorescentes. Todos los transposones también comprendían un par de aislantes HS4, uno adyacente a cada extremo del transposón. En algunas configuraciones (filas 8-23) cada uno de los dos genes estaba unido operativamente a promotores separados y señales de poliadenilación, en algunas configuraciones los genes estaban unido operativamente a un solo promotor que precede al primer gen y a una única señal poliA después del segundo gen, donde los dos genes estaban unidos operativamente por una secuencia IRES o una secuencia CHYSEL. Los transposones comprendían la SEQ ID NO. proporcionada en la columna A entre el primer y el segundo marco de lectura abierto. Los elementos reguladores asociados con cada gen se muestran en las columnas E-L de la Tabla 7. El número 1 en los encabezados indica los promotores que preceden o las señales poliA que siguen al primer gen, el número 2 en los encabezados indica los promotores que preceden o las señales poliA que siguen al segundo gen. Todas estas secuencias comprendían además la secuencia potenciadora de la expresión SEQ ID NO: 866 secuencia que precede a la última señal de poliadenilación. Los transposones se transfectaron en células CHO, solos o junto con un gen que codifica una transposasa (SEQ ID NO: 48) fusionada a una señal de localización nuclear, unida operativamente al promotor de CMV. Las células CHO se transfectaron y seleccionaron como se describe en la Sección 4.2.1. La fluorescencia representa la expresión de los ORF que codifican el indicador fluorescente DasherGFP a partir de transposones integrados de manera estable medidos a Ex/Em de 488/518 nm y CayenneRFP se midió a Ex/Em de 525/580 nm. La fluorescencia media de las transfecciones independientes por triplicado se muestra en la Tabla 7, columnas M y N, respectivamente.

La columna O en la Tabla 7 muestra la relación de fluorescencia roja a verde obtenida para cada transposón. Debido a que las dos proteínas fluorescentes emiten fluorescencia con diferentes intensidades, esta no es una medida de las relaciones de concentración de las dos proteínas. La fila 3 muestra la fluorescencia obtenida cuando las dos proteínas se acoplan mediante una secuencia CHYSEL. Esto produce cantidades cercanas a equimolares de las dos proteínas y da una relación de fluorescencia de rojo a verde de 0,21. La columna P muestra las relaciones de rojo a verde de la columna O, normalizadas al dividir por 0,21, para obtener la relación en la que se expresan las dos proteínas. La columna Q muestra la expresión de DasherGFP con respecto a la expresión cuando DasherGFP es la única proteína fluorescente codificada.

El acoplamiento de la traducción de dos marcos de lectura abiertos a través de IRES SEQ ID NO: 1062-1064 produjo niveles muy altos de expresión de la primera proteína codificada (al menos el 70 % de la cantidad de proteína obtenida de una construcción monocistrónica, compare la fila 2 con las filas 4 -7). Estas construcciones también produjeron la segunda proteína codificada entre 0,33 y 0,48 los niveles de la primera. Las proporciones de 1:0,33 a 1:0,48 son muy buenas para la producción de cadenas ligeras y pesadas de anticuerpos, respectivamente. Por tanto, los IRES SEQ ID NO: 1062-1064 son componentes preferidos de construcciones de polinucleótidos para la producción eficiente de anticuerpos.

Otra forma de obtener la expresión de dos genes a partir de un polinucleótido es unir operativamente cada gen a un promotor separado. Las filas 8 a 23 de la Tabla 7 muestran una variedad de configuraciones diferentes para los elementos reguladores asociados con el segundo marco de lectura abierto. En general, la expresión de la primera

proteína codificada fue menor en estas construcciones que en las construcciones acopladas a IRES (columna Q). Sin embargo, se obtuvo una mayor variedad de expresiones del segundo gen codificado (columna P). La expresión de un gen y, a menudo, ambos genes aumentaba sustancialmente si se intercalaba un aislante del núcleo HS4 (SEQ ID NO: 865) entre la secuencia de poliadenilación unida operativamente al primer marco de lectura abierto y el promotor unido operativamente al segundo marco de lectura abierto (compare las filas 14 y 15, las filas 16 y 17, las filas 18 y 19, las filas 20 y 21, las filas 22 y 23 y las filas 12 y 13). La ubicación de una secuencia aislante, como la secuencia aislante del núcleo HS4, entre una secuencia de poliadenilación unida operativamente a un primer marco de lectura abierto, y un promotor unido operativamente a un segundo marco de lectura abierto, es por tanto una configuración preferida para expresar dos genes del mismo polinucleótido.

Las configuraciones de vector ventajosas para la expresión de dos polipéptidos incluyen aquellas en las que un gen que codifica un primer polipéptido está operativamente ligado a elementos de control, incluido un intrón EF1 de una especie, y un gen que codifica un segundo polipéptido está operativamente ligado a elementos de control que incluyen un intrón EF1 de una segunda especie. Las configuraciones de vector ventajosas para la expresión de dos polipéptidos incluyen aquellas en las que un gen que codifica un primer polipéptido está operativamente ligado a elementos de control, incluido un promotor EF1 de una especie, y un gen que codifica un segundo polipéptido está operativamente ligado a elementos de control que incluyen un promotor EF1 de una segunda especie. Las configuraciones de vector ventajosas para la expresión de dos polipéptidos incluyen aquellas en las que un gen que codifica un primer polipéptido está operativamente ligado a elementos de control que incluyen un promotor de CMV humano, y un gen que codifica un segundo polipéptido está operativamente ligado a elementos de control que incluyen un promotor de CMV murino.

Las configuraciones de vector ventajosas para la expresión de dos polipéptidos incluyen aquellas en las que un gen que codifica un primer polipéptido está unido operativamente a elementos de control que incluyen una secuencia que es al menos 95 % idéntica a una secuencia seleccionada de SEQ ID NO: 1015, 1019, 1022, 1026, 1027, 1028, 1029 o 1099.

#### 6.2.2 Transposones con configuraciones de doble promotor o ires (2)

Las configuraciones de un conjunto de transposones que comprenden genes que codifican DasherGFP y/o CayenneRFP se indican en la Tabla 8 (columnas D y E). Cuando ambos genes de proteínas fluorescentes estaban presentes, el gen DasherGFP siempre aparecía primero. Todos los transposones también comprendían un gen de resistencia a puromicina unido operativamente al promotor PGK y transcrito en la dirección opuesta a los genes de las proteínas fluorescentes. Dos de los transposones (Transposones 188209 y 188219, columna B) también comprendían un par de aislantes HS4, uno adyacente a cada extremo del transposón. En algunas configuraciones, cada uno de los dos genes estaba unido operativamente a promotores separados y señales de poliadenilación, en algunas configuraciones los genes estaban unidos operativamente a un solo promotor que precede al primer gen y a una única señal poliA después del segundo gen, donde los dos genes están unido operativamente por una secuencia IRES o una secuencia CHYSEL. El número 1 indica los promotores que preceden o las señales poliA que siguen al primer gen, el número 2 indica los promotores que preceden o las señales poliA que siguen al segundo gen. Todas estas secuencias comprendían además la secuencia potenciadora de la expresión SEQ ID NO: 866 que precede a la última señal de poliadenilación. Los transposones se transfectaron en células CHO más o menos un gen que codifica una transposasa (SEQ ID NO: 48) fusionada a una señal de localización nuclear, operativamente unida al promotor CMV (columna O).

Las células CHO-K1 se transfectaron y seleccionaron como se describe en la Sección 4.2.1. La fluorescencia representa la expresión de los ORF que codifican el indicador fluorescente DasherGFP a partir de transposones integrados de manera estable medidos a Ex/Em de 488/518 nm y CayenneRFP se midió a Ex/Em de 525/580 nm (Tabla 8 columnas P-U)

La cotransfección de transposones con el vector que codifica la transposasa aumentó la expresión de ambas proteínas codificadas por el transposón entre 4 y casi 20 veces con respecto a las transfecciones con el transposón solo. La expresión de dos polipéptidos de un transposón se mejora mediante la actividad de una transposasa. Los transposones de mejor expresión comprendían los aislantes HS4 flanqueantes y la expresión de los dos polipéptidos acoplados por un IRES. Por tanto, estas son configuraciones preferidas para transposones para la expresión de dos polipéptidos.

#### 6.3 Transposasas modificadas

##### 6.3.1 Transposasas hiperactivas de *xenopus*

###### 6.3.1.1 Identificación de transposasas hiperactivas de *Xenopus*

Para identificar las mutaciones de la transposasa de *Xenopus* que condujeron a una mayor actividad de transposición o una mayor actividad de escisión, con respecto a la secuencia de transposasa de origen natural SEQ ID NO: 48, primero creamos bibliotecas que juntas contenían todos los posibles cambios de aminoácidos

individuales de SEQ ID NO: 48, fusionado a una secuencia de localización nuclear heteróloga. Para ello, se amplificó mediante PCR un gen que codifica la transposasa de *Xenopus* fusionada a una secuencia de localización nuclear heteróloga (SEQ ID NO: 50), mediante el uso de cebadores redundante para incorporar todos los aminoácidos posibles en una única posición. Esta biblioteca de "saturación de sitio" se clonó en un vector que comprendía un marcador de selección de leucina; los genes que codifican los mutantes de la transposasa se unieron operativamente al promotor Gal-1 de *Saccharomyces cerevisiae*. Se creó una biblioteca para cada uno de los 589 aminoácidos de la transposasa. Cada biblioteca se secuenció a lo largo de la posición mutagenizada para garantizar que no se hubieran introducido mutaciones no deseadas y que el codón diana estuviera efectivamente mutado.

A continuación, se agruparon las bibliotecas clonadas, cada grupo comprendía bibliotecas para seis posiciones de aminoácidos adyacentes, es decir, los aminoácidos 1-6, 7-12, 13-18, etc. Cada grupo de bibliotecas se transformó luego en células de *Saccharomyces cerevisiae* que llevaban un grupo genómico integrado en un casete de selección de transposones de *Xenopus* y sembrado en medio mínimo completo que carece de leucina para seleccionar los transformantes que portan el marcador de selección de leucina.

El casete de selección de transposones de *Xenopus* (SEQ ID NO: 44) comprendía el promotor URA3 de *Saccharomyces cerevisiae* seguido de un gen que codifica la primera parte de la proteína URA3 de *Saccharomyces cerevisiae*. La proteína URA3 se interrumpió en una secuencia de TTAA dentro de su región codificante por un inserto de transposón que comprende el extremo izquierdo del transposón de *Xenopus* SEQ ID NO: 2, el promotor de TEF de *Ashbya gossypii*, un marco de lectura abierto que codifica la proteína TRP1 de *Saccharomyces cerevisiae*, el terminador de TEF de *Ashbya gossypii* y extremo derecho del transposón de *Xenopus* SEQ ID NO: 12. En el otro lado del inserto de transposón había una secuencia de ADN que codifica el resto de la proteína URA3 de *Saccharomyces cerevisiae*, seguida del terminador URA3 de *Saccharomyces cerevisiae*. El transposón se insertó de manera que la secuencia de TTAA estuviera presente en ambos extremos, y la eliminación del transposón para dejar una única copia de este TTAA daría como resultado un gen completo y funcional que codifica la proteína URA3 de *Saccharomyces cerevisiae*.

Dos días después de la siembra en placa, se recolectaron células de *Saccharomyces cerevisiae* transformadas añadiendo 5 ml de agua estéril a cada placa y raspando suavemente para resuspender las células. Las células se combinaron en grupos que representan 60 aminoácidos adyacentes, es decir, 1-60, 61-120, 121-180, etc. Se midió la  $A_{600}$  de cada grupo y se usó para estimar la concentración de células vivas. Se preparó ADN plasmídico a partir de una porción de cada grupo para secuenciar. Esto fue para determinar la frecuencia de cada cambio de aminoácido en la biblioteca nueva (no seleccionada). A continuación, las células se seleccionaron bajo 3 regímenes diferentes.

Selección 1: se transfirieron  $2 \times 10^8$  células de cada grupo a un medio mínimo menos leucina que contenía galactosa al 2 % y se cultivaron durante 4 horas a 30 °C. A continuación, las células se sembraron en medio mínimo completo que carecía de uracilo, triptófano y leucina. Dos días después de la siembra en placa, se recolectaron células de *Saccharomyces cerevisiae* transformadas añadiendo 5 ml de agua estéril a cada placa y raspando suavemente para resuspender las células. Se preparó ADN plasmídico para determinar la frecuencia de cada cambio de aminoácido en la biblioteca seleccionada.

Selección 2: Se transfirieron  $2 \times 10^8$  células de cada grupo a un medio mínimo menos leucina que contenía galactosa al 2 % y se cultivaron durante 20 horas a 30 °C. A continuación, las células se sembraron en medio mínimo completo que carecía de uracilo, triptófano y leucina. Dos días después de la siembra en placa, se recolectaron células de *Saccharomyces cerevisiae* transformadas añadiendo 5 ml de agua estéril a cada placa y raspando suavemente para resuspender las células. Se preparó ADN plasmídico para determinar la frecuencia de cada cambio de aminoácido en la biblioteca seleccionada.

Selección 3: Se transfirieron  $2 \times 10^8$  células de cada grupo a un medio mínimo menos leucina que contenía galactosa al 2 % y se cultivaron durante 20 horas a 30 °C. A continuación, las células se sembraron en placas sobre medio mínimo completo que carecía de uracilo y leucina y que contenían 0,5 g/L de ácido 5-fluoroantranílico. Cinco días después de la siembra en placa, se recolectaron células de *Saccharomyces cerevisiae* transformadas añadiendo 5 ml de agua estéril a cada placa y raspando suavemente para resuspender las células. Se preparó ADN plasmídico para determinar la frecuencia de cada cambio de aminoácido en la biblioteca seleccionada.

Las selecciones 1 y 2 identificaron células en las que la transposasa había escindido con precisión el transposón del gen uracilo (por lo que las células son URA+) y se reintegraron en otro sitio del genoma (por lo que las células eran TRP+). La selección 3 identificó células en las que la transposasa había escindido con precisión el transposón del gen uracilo (por lo que las células eran URA+) pero no se habían reintegrado en otro sitio del genoma (por lo que las células eran TRP- y resistentes al ácido 5-fluoroantranílico).

Los genes de transposasa mutados de la biblioteca nueva y cada una de las bibliotecas seleccionadas se secuenciaron mediante el uso de un Illumina HiSeq. Las frecuencias de mutación de la biblioteca nueva se compararon con las frecuencias de las bibliotecas seleccionadas. Las mutaciones que estaban más representadas en las bibliotecas seleccionadas en las condiciones 1 o 2, en comparación con la biblioteca sin tratamiento previo, fueron aquellas que aumentan la actividad de transposición de la transposasa. Las mutaciones que estaban más

representadas en la biblioteca seleccionada en la condición 3, en comparación con la biblioteca sin tratamiento previo, fueron aquellas que aumentan la actividad de escisión de la transposasa. La operabilidad de una transposasa de *Xenopus* se puede demostrar por la capacidad de la transposasa, cuando se fusiona con un NLS heterólogo, para escindir el transposón de la SEQ ID NO: 44 y, excepto en el caso de una transposasa de integración deficiente, para integrar el transposón en el ADN genómico de una célula diana.

La Tabla 4 muestra las sustituciones de aminoácidos que se representaron al menos 2 veces más frecuentemente en la biblioteca seleccionada que en la biblioteca sin tratamiento previo. Los datos se procesaron de la siguiente manera. Teniendo en cuenta cada cambio de aminoácido de forma independiente, cualquier sustitución de aminoácido que se produjera en la biblioteca sin tratamiento previo menos de una vez por cada 200 sustituciones observadas en esa posición, se descartó de consideración adicional. Cualquier sustitución de aminoácido que se haya observado menos de 100 veces en la biblioteca seleccionada se descartó de una consideración adicional. A continuación, se calculó la frecuencia de cada sustitución en cada biblioteca seleccionada con respecto a la frecuencia con la que se produjo la sustitución en la biblioteca nueva. Las sustituciones que ocurrieron al menos con el doble de frecuencia en una biblioteca seleccionada para la transposición, en comparación con su frecuencia en la biblioteca nueva, se muestran en la columna C de la Tabla 4. Las sustituciones que ocurrieron al menos el doble de frecuencia en una biblioteca seleccionada para la escisión, en comparación con su frecuencia en la biblioteca nueva se muestran en la columna D de la Tabla 4.

Algunas de las sustituciones de aminoácidos mostradas en la Tabla 4 se seleccionaron para incorporarlas en transposasas variantes de *Xenopus* para crear variantes hiperactivas. Inicialmente, se creó un conjunto de 95 variantes de transposasa SEQ ID NO: 48 seleccionando 57 de las sustituciones que se muestran en la Tabla 4 e incorporando 3 de estas en cada una de las 95 variantes, de modo que se maximice el número de posibles pares y cada sustitución ocurre 5 veces en el conjunto de variantes. Las transposasas se clonaron en un vector que comprende un marcador de selección de leucina, de modo que las variantes de transposasas se unieron operativamente al promotor Gal-1 de *Saccharomyces cerevisiae*. A continuación, cada una de estas variantes se transformó individualmente en una cepa de *Saccharomyces cerevisiae* que portaba una copia integrada cromosómicamente de SEQ ID NO: 44, como se describió anteriormente. Las variantes se indujeron con galactosa, se cultivaron durante 4 horas, luego se sembraron alícuotas (a) en medios que carecían de leucina, uracilo y triptófano (para contar la integración), (b) en medios que carecían de leucina y uracilo (para contar la escisión) y (c) en medios que carecen de leucina (para contar el total de células vivas). Dos días después, se contaron las colonias para determinar la frecuencia de transposición (= número de células en medio -leu-ura-trp dividido por el número de células en medio -leu) y escisión (= número de células en medio -leu-ura dividido por el número de células en -leu media).

Las frecuencias de transposición se modelaron como se describe en Patente de EE. UU. 8,635,029 y se calcularon los valores promedios y las desviaciones estándar de las ponderaciones de regresión para cada sustitución. Se diseñaron conjuntos posteriores de variantes que incorporan más de 3 sustituciones con respecto a la secuencia de SEQ ID NO: 48. Estas variantes combinaron dos o más sustituciones con ponderaciones de regresión mayores que una desviación estándar por encima de cero. Las variantes también comprendían opcionalmente una o más sustituciones seleccionadas de la columna C o D en la Tabla 4. Se probaron nuevas variantes como se describió anteriormente para medir las frecuencias de transposición y/o escisión para las nuevas variantes de transposasas. Las ponderaciones de regresión y las desviaciones estándar para sustituciones con un efecto positivo en la actividad de transposición se muestran en la Tabla 11, columnas D y E. Las frecuencias de transposición para algunas transposasas hiperactivas de *Xenopus* se muestran en la Tabla 14, columnas A y B. Las frecuencias se midieron para la escisión del transposón indicador de SEQ ID NO: 44 y la integración de ese transposón en el genoma de *Saccharomyces cerevisiae*, y se expresan con respecto a las frecuencias de transposición medidas para la secuencia de origen natural SEQ ID NO: 48 en condiciones idénticas.

Se encontró que las transposasas con SEQ ID NO: 51 y 403-406 tenían frecuencias de escisión que eran al menos 10 veces más altas que sus frecuencias de integración. Las transposasas en las que el aminoácido en la posición 218 se cambió de Asn a Asp o Glu también mostraron frecuencias de escisión mucho más altas que de integración. Por tanto, estas transposasas deficientes en integración son útiles para eliminar transposones integrados de un genoma del huésped.

#### 6.3.1.2 Transposasas hiperactivas de *Xenopus* en células de mamíferos

Se probó la capacidad de varias transposasas hiperactivas de *Xenopus* para integrar tres configuraciones de transposones diferentes en el genoma de CHO. Las configuraciones de transposones se muestran en la Tabla 16.

Los transposones (750 ng) se cotransfectaron con polinucleótidos que codifican las transposasas de *Xenopus* fusionadas con una señal de localización nuclear heteróloga. Los ácidos nucleicos de transposones y transposasas se transfectaron en 1 ml de células CHO adaptadas a la suspensión. Las células se cultivaron durante 72 horas después de la transfección y luego se diluyeron a 250 000 células por ml en 40 µg/ml de puromicina durante 8 días. Se eliminó la puromicina y se cultivaron las células durante 7 días más. La expresión de Dasher GFP se midió en un fluorímetro de placa (excitación a 485 nm y emisión medida a 515 nm).

Esta selección es muy estricta: el gen de la puromicina acetil transferasa está operativamente ligado con un promotor débil y las células se diluyeron a niveles bajos en niveles altos de puromicina. En todos los casos en estas condiciones estrictas, la ausencia de transposasa (Tabla 16 filas 3, 6 y 9) o cotransfección de transposones de *Xenopus* de origen natural y *Xenopus* transposasa SEQ ID NO: 48 (Tabla 16 filas 1, 4 y 7), resultó en una muerte celular esencialmente completa y niveles muy bajos de expresión de DasherGFP. En contraste, la cotransfección de transposones de *Xenopus* con transposasas hiperactivas de *Xenopus* SEQ ID NO: 57, 58 y 61 dio como resultado grupos con altos niveles de expresión de DasherGFP (Tabla 16 filas 2, 5 y 8). Por tanto, es ventajoso cotransfectar células de mamífero con transposones de *Xenopus* y transposasas hiperactivas de *Xenopus*, incluidas las SEQ ID NO: 57, 58 y 61.

### 6.3.2 Transposasas hiperactivas de *bombyx*

#### 6.3.2.1 Identificación de transposasas de *Bombyx* hiperactivas

Para identificar las mutaciones de la transposasa de *Bombyx* que condujeron a una mayor actividad de transposición o una mayor actividad de escisión, con respecto a la secuencia de origen natural SEQ ID NO: 407, primero creamos bibliotecas que juntas contenían todos los posibles cambios de aminoácidos individuales de SEQ ID NO: 407, fusionados a una secuencia de localización nuclear heteróloga. Para hacer esto, se mutó, clonó y secuenció un gen que codificaba la transposasa de *Bombyx* fusionada a una secuencia de localización nuclear heteróloga (SEQ ID NO: 408) como se describe para la transposasa de *Xenopus* en la Sección 6.3.2.1.

Las bibliotecas clonadas se combinaron y se transformaron en células de *Saccharomyces cerevisiae* que llevaban un casete de selección de transposón de *Bombyx* integrado genómicamente como se describe para la transposasa de *Xenopus* en la Sección 6.3.1.

El casete de selección del transposón de *Bombyx* (SEQ ID NO: 47) fue como se describió para el casete *Xenopus* SEQ ID NO: 44 en la Sección 6.3.1, excepto que los extremos del transposón *Xenopus* SEQ ID NO. 2 y 12 fueron reemplazados por las secuencias del extremo del transposón de *Bombyx* SEQ ID NO. 22 y 30 respectivamente. La operabilidad de una transposasa de *Bombyx* se puede demostrar por la capacidad de la transposasa, cuando se fusiona con un NLS heterólogo, para escindir el transposón de la SEQ ID NO: 47 y, excepto en el caso de una transposasa de integración deficiente, para integrar la transposón en el ADN genómico de una célula diana.

Las bibliotecas de mutantes de transposasa de *Bombyx* se seleccionaron, secuenciaron y procesaron, como se describe para las bibliotecas de transposasa de *Xenopus* en la sección 6.3.1. Las sustituciones que ocurrieron al menos con el doble de frecuencia en una biblioteca de *Bombyx* seleccionada para transposición, en comparación con su frecuencia en la biblioteca sin tratamiento previo, se muestran en la Tabla 4, columna G. Sustituciones que ocurrieron al menos el doble de frecuencia en una biblioteca de *Bombyx* seleccionada para la escisión, en comparación con su frecuencia en la biblioteca ingenua se muestra en la Tabla 4, columna H.

Algunas de las sustituciones de aminoácidos que se muestran en la Tabla 4 se seleccionaron para su incorporación en la transposasa de *Bombyx* SEQ ID NO: 407 para crear variantes hiperactivas, como se describe para la transposasa de *Xenopus* en la Sección 6.3.1 pero mediante el uso de una cepa de *Saccharomyces cerevisiae* que lleva una copia integrada cromosómicamente de SEQ ID NO: 47 en lugar de SEQ ID NO: 44.

Se diseñaron conjuntos posteriores de variantes que incorporan más de 3 sustituciones con respecto a la secuencia de SEQ ID NO: 407. Estas variantes combinaron dos o más sustituciones con pesos de regresión mayores que una desviación estándar por encima de cero. Las variantes también comprendieron opcionalmente una o más sustituciones seleccionadas de la columna G o H en la Tabla 4. Se probaron nuevas variantes como se describió anteriormente para medir las frecuencias de transposición y/o escisión para las nuevas variantes de transposasas. Las ponderaciones de regresión y las desviaciones estándar para sustituciones con un efecto positivo en la actividad de transposición se muestran en la Tabla 11, columnas I y J. Las frecuencias de transposición para diferentes transposasas de *Bombyx* hiperactivas se muestran en la Tabla 14, columnas C y D. Las frecuencias se midieron para la escisión del transposón de indicador SEQ ID NO: 47 y la integración de ese transposón en el genoma de *Saccharomyces cerevisiae*, y se expresan con respecto a las frecuencias para la secuencia de origen natural SEQ ID NO: 407 en condiciones idénticas.

#### 6.3.2.2 Transposasas de *Bombyx* hiperactivas en células de mamíferos

Se probó la capacidad de varias transposasas *Bombyx* hiperactivas para integrar cuatro configuraciones de transposones diferentes en el genoma de CHO. Las configuraciones de transposones se muestran en la Tabla 15. El transposón 194094 comprendía un promotor PGK (SEQ ID NO: 937) unido operativamente a un gen de puromicina acetil transferasa y un promotor CMV unido operativamente a un gen que codifica Dasher GFP. El transposón 240671 era el mismo que el de 194094, excepto que las secuencias de los extremos del transposón eran diferentes, como se muestra en la Tabla 15. El transposón 246143 era el mismo que el 240671, excepto que el promotor PGK se reemplazó con el promotor HSV-TK SEQ ID NO: 942. El transposón 246170 era similar al 246143, pero tenía el

promotor EF1a unido operativamente al gen que codifica Dasher GFP, también está flanqueado por secuencias aislantes (aislante HS4 SEQ ID NO: 864 en un lado y el aislante D4Z4 SEQ ID NO: 860 en el otro).

Los transposones (750 ng) se cotransfectaron con polinucleótidos que codifican una transposasa con la SEQ ID NO. mostrada en la columna K, fusionados a una señal de localización nuclear heteróloga. Los ácidos nucleicos de transposones y transposasas se transfectaron en 1 ml de células CHO adaptadas a la suspensión. Las células se cultivaron durante 72 horas después de la transfección y luego se diluyeron hasta 1 000 000 de células por ml en 40 µg/ml de puromicina durante 7 días. Se eliminó la puromicina y se cultivaron las células durante 7 días más. La expresión de Dasher GFP se midió en un fluorímetro de placa (excitación a 485 nm y emisión medida a 515 nm) (Tabla 15 columnas OQ). Se realizó una estimación del número de células vivas midiendo la absorbancia a 600 nm ( $A_{600}$ ) (Tabla 15, columnas LN).

En todos los casos, la ausencia de transposasa dio como resultado niveles muy bajos de expresión de DasherGFP y un  $A_{600}$  muy bajo que indica una falta de expresión de puromicina acetil transferasa y supervivencia celular (filas 8, 16, 24 y 32). Todas las transposasas dieron como resultado niveles comparables de supervivencia celular para las células cotransfectadas con los transposones 194094 y 240671 (compare las columnas L, M y N de la Tabla 15 para las filas 1-7 y 9-15). Sin embargo, las transposasas hiperactivas dieron como resultado niveles significativamente mayores de expresión de DasherGFP (compare las columnas O, P y Q de la Tabla 15 para las filas 1-7 y 9-15). Por tanto, es ventajoso cotransfectar células de mamífero con transposones *Bombyx* y transposasas *Bombyx* hiperactivas, que incluyen SEQ ID NO: 1098, 412, 457 y 415-417.

Las células transfectadas con el transposón 246143 murieron todas en las condiciones de selección usadas, independientemente de que transposasa se cotransfectó (Tabla 15, columnas L, M, N, O, P y Q para las filas 17-24). Sin embargo, las células transfectadas con el transposón 246170 y cotransfectadas con transposasas *Bombyx* hiperactivas, SEQ ID NO: 1098, 412 y 415-417, dieron como resultado células con fluorescencia Dasher GFP. No sobrevivió ninguna célula cuando este transposón se cotransfectó con la transposasa de *Bombyx* de origen natural (SEQ ID NO: 407). Las transposasas de *Bombyx* hiperactivas, SEQ ID NO: 412 y 415 fueron particularmente ventajosas en combinación con esta configuración de transposón.

La configuración del transposón, el rigor de la selección y la actividad de la transposasa son interdependientes para determinar el nivel de expresión que resulta del transposón integrado posteriormente. El promotor que está operativamente ligado a un polipéptido de expresión (en este ejemplo DasherGFP) también puede modificar la fuerza del promotor que está operativamente ligado al marcador de selección y. Como se describe en la Sección 5.2.10, un promotor fuerte operativamente ligado al marcador de resistencia (como en el transposón 240671) proporcionará la selección menos rigurosa, mientras que un promotor débil operativamente ligado al marcador de resistencia (como en el transposón 246143) proporcionará una mayor selección rigurosa, particularmente en combinación con un promotor interferente unido operativamente con el polipéptido de expresión.

Aquí se muestra el beneficio de una selección más estricta junto con una transposasa hiperactiva. La transposasa de *Bombyx* hiperactiva SEQ ID NO: 412 y 415 produjeron cada una un conjunto de células con una expresión sustancialmente más alta del polipéptido de expresión del transposón 246170 que la que se logró con los transposones con el promotor más fuerte asociado con el marcador de selección (compare la fila 9, columna O con la fila 25 columna O y fila 26 columna Q). Además, la productividad de las células del transposón seleccionado más estrictamente (expresión dividida por el número de células vivas, que es aproximadamente proporcional a la fluorescencia dividida por  $A_{600}$ ) es aproximadamente 10 veces mayor que para el transposón seleccionado menos estrictamente.

Aunque las frecuencias de integración relativas de las transposasas hiperactivas mostradas en la Tabla 14 dan una comparación cuantitativa de la actividad de la transposasa, el aumento de la actividad de la transposasa por sí solo no es suficiente para garantizar el aumento de la expresión resultante de los transposones integrados en el genoma de una célula diana. Como se muestra en este ejemplo, la configuración del transposón y el rigor de la selección son factores que influyen en la expresión de un polinucleótido de expresión. En particular, el gen que codifica el marcador de selección, el promotor (y otros elementos reguladores) ligados operativamente al marcador de selección, el promotor ligado operativamente al gen que codifica el polipéptido de expresión y cualquier elemento aislante presente son determinantes importantes de la expresión de un polinucleótido de transferencia de genes. Los polinucleótidos de transferencia de genes particularmente ventajosos comprenden una secuencia que es al menos un 95 % idéntica a una secuencia seleccionada de SEQ ID NO: 751-819. Preferentemente, estas secuencias están dentro de un transposón.

Los datos que se muestran en la Tabla 15 muestran solo la fluorescencia promedio dentro de un grupo de células. Estos conjuntos se derivaron de muchas células transfectadas de forma independiente. Cada uno de estos dará lugar a un patrón de integración de transposones diferente (número de transposones integrados y posición de cada uno de estos transposones dentro del genoma de la célula diana). Las líneas individuales se pueden aislar de un grupo de células como este, y algunas de estas a menudo tienen productividades sustancialmente más altas que el grupo.

## 6.4 elementos ires

### 6.4.1 Niveles de expresión de dos polipéptidos mediante el uso de elementos ires en células HEK293 Y CHO

- 5 Un sistema de transferencia de genes que comprende genes que codifican dos polipéptidos puede unir operativamente ambos polipéptidos al mismo promotor, por ejemplo usando un IRES.

10 La Tabla 9 muestra los niveles de expresión observados en células HEK y CHO para dos polipéptidos diferentes (en este caso, dos proteínas fluorescentes diferentes, DasherGFP y CayenneRFP) codificadas en un único polinucleótido de transferencia de genes. Los genes de las dos proteínas diferentes se unieron operativamente a un único potenciador, promotor, señal de poliadenilación y, opcionalmente, un intrón. La expresión de los dos genes se unió operativamente mediante un elemento IRES, como se indica en la columna A, el orden de los elementos es DasherGFP-IRES-CayenneRFP.

15 Se cultivaron células HEK 293a (de ATCC) en EMEM (de ATCC) + 10 % de FBS (de ATCC) + 1 % de penicilina-estreptomicina (de ATCC) a 37 °C, 5 % de CO<sub>2</sub> a 80 % de confluencia, 1E+05 células se sembraron en placas de cultivo de tejidos de 24 pocillos y se incubaron a 37 °C, 5 % de CO<sub>2</sub> durante 24 horas antes de la transfección, las transfecciones se establecieron por triplicado. Cada transfección usó 0,5 µg de ADN con Lipofectamina 2000 según el protocolo del fabricante. Las células se recolectaron 72 horas después de la transfección. Células CHO-K1 (de ATCC) se cultivaron en F12-K (de ATCC) + 10 % de FBS (de ATCC) + 1 % de penicilina-estreptomicina (de ATCC) a 37 °C, 5 % de CO<sub>2</sub> al 80 % de confluencia. Las células 5E+05 se sembraron en placas de cultivo de tejidos de 24 pocillos y se incubaron a 37 °C, 5 % de CO<sub>2</sub> durante 24 horas antes de la transfección, las transfecciones se establecieron por triplicado. Cada transfección usó 0,5 µg de ADN con Lipofectamina 2000 según el protocolo del fabricante. Las células se recolectaron 72 horas después de la transfección. Se midió la fluorescencia de los dos ORF que codifican los indicadores fluorescentes DasherGFP y CayenneRFP a Ex/Em de 488/518 nm para DasherGFP y Ex/Em de 525/580 nm para CayenneRFP.

30 Un polinucleótido de transferencia de genes que comprende las dos proteínas acopladas traduccionalmente por una secuencia CHYSEL expresa las dos proteínas en una relación equimolar y se usó para normalizar las diferentes intensidades fluorescentes de las proteínas. La Tabla 9 muestra que se pueden usar diferentes elementos IRES para obtener diferentes relaciones de expresión entre dos polinucleótidos diferentes. El uso de elementos IRES es particularmente ventajoso para la expresión de polipéptidos cuando la relación de expresión es importante a nivel de células individuales, por ejemplo en la expresión de anticuerpos donde la cadena ligera puede realizar una función de chaperonina para la cadena pesada. A veces es ventajoso expresar una relación lo más grande posible entre dos polipéptidos, por ejemplo, en el caso de que un polipéptido sea un marcador de selección.

40 Hemos identificado elementos IRES que muestran diferentes niveles de actividad como se ve en los niveles de expresión variables para los dos marcos de lectura abiertos (ORF) unidos por un elemento IRES que se muestra en la Tabla 9. Una elección de elementos IRES con diferentes actividades permite usar el elemento IRES apropiado para controlar los niveles de expresión relativos de dos ORF. Hemos demostrado el uso de un elemento IRES que une dos transcripciones operativamente enlazadas a un promotor. Se contempla expresamente el uso de dos o más elementos IRES que unen tres o más ORF. Se contemplan expresamente las construcciones de expresión con dos o más elementos IRES seleccionados de manera que los niveles de expresión de dos o más ORF se modulan selectivamente. Los elementos IRES identificados funcionan bien en vectores de integración transitorios y estables en las dos líneas celulares probadas, células de riñón embrionario humano (HEK293) y células de ovario de hámster chino (CHO). Los casos preferidos de un polinucleótido de transferencia de genes incluyen todos los elementos IRES que se muestran en la Tabla 9.

## 6.5 Actividad transposasa en levadura

50

### 6.5.1 Transposones en *pichia pastoris*

55 Para integrar un polinucleótido en el genoma de *Pichia pastoris*, generalmente es necesario linealizar una construcción de transferencia de genes antes de la transformación. Es ventajoso que los extremos de la construcción de transferencia de genes lineal sean homólogos a las secuencias vecinas en el *genoma de Pichia pastoris*, de modo que la construcción pueda integrarse en el cromosoma mediante recombinación homóloga. Tales construcciones de transferencia de genes generalmente comprenden un gen que codifica un marcador de selección (por ejemplo, resistencia a zeocina (por ejemplo, SEQ ID NO: 702), norsoterisina (por ejemplo, SEQ ID NO: 701) o geneticina (por ejemplo, SEQ ID NO: 706). Pueden obtenerse altos niveles de expresión exponiendo las células a altos niveles del agente de selección correspondiente, lo que da como resultado la amplificación del gen. La amplificación generalmente se logra mediante la duplicación en tándem del gen, que es una disposición inherentemente inestable. Debido a que los transposones se integran casi al azar en todo el genoma objetivo, ofrecen la ventaja de una alta expresión resultante de múltiples copias insertadas, al tiempo que mejoran la estabilidad porque las copias se distribuyen por todo el genoma.

65

Se construyeron tres transposones para modificar el genoma de *Pichia pastoris* y comprendían los extremos del transposón SEQ ID NO: 2 y 12 que flanqueaban un polinucleótido heterólogo. El polinucleótido heterólogo comprendía un promotor AOX (SEQ ID NO: 953) unido operativamente a un gen que codifica Dasher GFP (SEQ ID NO: 42), y un promotor ILV5 (SEQ ID NO: 955) unido operativamente a un gen que codifica resistencia a zeocina (SEQ ID NO: 702). Uno de estos transposones (251587) se transportó en un plásmido que comprendía el promotor GAP SEQ ID NO: 949 unido operativamente a un gen que codifica la transposasa de *Xenopus* SEQ ID NO: 118; se transportó un segundo transposón (251588) en un plásmido que comprendía el promotor de TEF SEQ ID NO: 954 unido operativamente a un gen que codifica la transposasa de *Xenopus* SEQ ID NO: 118; se transportó un tercer transposón (251589) en un plásmido sin transposasa. Todas las transposasas eran parte de la parte no transponible del plásmido.

Los tres transposones se transformaron como ADN circular superenrollado en células de *Pichia pastoris* competentes mediante electroporación (mediante el uso de un Bio-Rad *E. coli* Pulser en cubetas con un espacio de 0,2 cm y 1,5 kV). Además, el transposón 251589 (cuyo plásmido carecía por completo de transposasa) se electroporó después de la linealización con Pmel, que corta dentro del promotor AOX. Después de la electroporación, las células se cultivaron en un medio no selectivo (900 µl de caldo YPD más sorbitol 1 M) durante 5 o 24 horas a 30 °C, antes de sembrar 100 µl de cultivo en 200 µg/ml de zeocina y se incubaron las placas a 30 °C durante 48 horas. Se contó el número de colonias resistentes a zeocina en cada placa y se muestra en la Tabla 10 (columnas E y F).

Sin linealización, se formaron muy pocas colonias en ausencia de una transposasa (filas 9-11). Por el contrario, la linealización antes de la electroporación dio como resultado aproximadamente 1000 colonias de 100 µl de cultivo (fila 12). De manera similar, la expresión de la transposasa de *Xenopus* SEQ ID NO: 118, transcrita del promotor GAP (filas 3-5) o del promotor TEF (filas 6-8) dio como resultado decenas a cientos de colonias. Por tanto, los transposones y transposasas de *Xenopus* son útiles para integrar construcciones de transferencia de genes en el genoma de la levadura *Pichia pastoris*.

Breve descripción de las tablas

Tabla 1. Integración de transposones catalizada por transposasas modificadas con o sin señales de localización nuclear heteróloga.

Los transposones y transposasas se transfectaron en células CHO-K1 y se seleccionaron como se describe en el Ejemplo 6.1.1. La fluorescencia se midió raspando las células y colocándolas en un fluorímetro. Las lecturas fluorescentes obtenidas en transfecciones triplicadas independientes se muestran en las columnas JL. Las columnas A, B y F se refieren a las SEQ ID NO.

Tabla 2. Integración de transposones con extremos de transposones modificados.

Los transposones y transposasas se transfectaron en células CHO-K1 y se seleccionaron como se describe en el Ejemplo 6.1.2.1. La fluorescencia se midió raspando las células y colocándolas en un fluorímetro. Las lecturas fluorescentes obtenidas en transfecciones triplicadas independientes se muestran en las columnas JL. Las columnas A, B y G se refieren a las SEQ ID NO.

Tabla 3. Integración de transposones con extremos de transposones modificados.

Los transposones y transposasas se transfectaron en células CHO-K1 y se seleccionaron como se describe en el Ejemplo 6.1.2.2. La fluorescencia se midió raspando las células y colocándolas en un fluorímetro. Las lecturas fluorescentes obtenidas en transfecciones triplicadas independientes se muestran en las columnas JL. Las columnas A, B y G se refieren a las SEQ ID NO.

Tabla 4. Sustituciones en transposasas asociadas a hiperactividad.

Se produjeron, seleccionaron y secuenciaron transposasas mutantes de *Xenopus* y *Bombyx* como se describe en los Ejemplos 6.3.1.1 y 6.3.2.1. Las posiciones relativas a la transposasa de *Xenopus* SEQ ID NO: 48 se muestran en la columna A; el aminoácido de origen natural está en la columna B; las sustituciones que ocurrieron al menos dos veces más frecuentemente en una biblioteca de *Xenopus* seleccionada para transposición, en comparación con su frecuencia en la biblioteca sin tratamiento previo, se muestran en la columna C; las sustituciones que ocurrieron al menos dos veces más frecuentemente en una biblioteca de *Xenopus* seleccionada para la escisión, en comparación con su frecuencia en la biblioteca sin tratamiento previo, se muestran en la columna D. Las posiciones relativas a la transposasa de *Bombyx* SEQ ID NO: 407 se muestran en la columna E; el aminoácido de origen natural está en la columna F; las sustituciones que ocurrieron al menos dos veces más frecuentemente en una biblioteca de *Bombyx* seleccionada para transposición, en comparación con su frecuencia en la biblioteca sin tratamiento previo, se muestran en la columna G; las sustituciones que ocurrieron al menos dos veces más frecuentemente en una biblioteca de *Bombyx* seleccionada para la escisión, en comparación con su frecuencia en la biblioteca nueva, se

muestran en la columna H. Las posiciones en las dos transposasas que comparten una línea en la tabla no corresponden a una alineación de secuencia entre las dos proteínas.

Tabla 5. Integración de transposones mediante el uso de ARNm de transposasa.

Los transposones y transposasas se transfectaron en células CHO-K1 y se seleccionaron como se describe en el Ejemplo 6.1.3.1. La fluorescencia se midió raspando las células y colocándolas en un fluorímetro. Las lecturas fluorescentes obtenidas en transfecciones triplicadas independientes se muestran en las columnas JL. Las columnas A, B y E se refieren a las SEQ ID NO.

Tabla 6. Integración de transposones mediante el uso de ARNm de transposasa.

Los transposones y transposasas se transfectaron en células CHO-K1 y se seleccionaron como se describe en el Ejemplo 6.1.3.3. La fluorescencia se midió raspando las células y colocándolas en un fluorímetro. Las lecturas fluorescentes obtenidas en transfecciones triplicadas independientes se muestran en las columnas JL. Las columnas A, B, C y F se refieren a las SEQ ID NO.

Tabla 7. Integración de transposones derivados de *Xenopus* para la expresión de dos polipéptidos.

Los transposones comprendían secuencias diana 5'-TTAA-3', secuencias terminales de transposones SEQ ID NO: 2 y SEQ ID NO: 12, y el promotor EFla y el intrón operativamente ligado a un gen que codifica DasherGFP (filas 2-23) o CayenneRFP (fila 1). Para las filas 3-7, los vectores comprendían además un gen que codifica CayenneRFP unido operativamente a los elementos de control de la expresión mediante una secuencia de acoplamiento traduccional (la SEQ ID NO se proporciona en la columna A). Para las filas 8-23, los vectores comprendían además un gen que codifica Cayenne RFP unido operativamente a un segundo potenciador (columna I), un segundo promotor (columna J), un segundo intrón (columna K) y una segunda señal de poliadenilación (columna L). Opcionalmente, se intercaló una secuencia aislante entre la primera señal de poliadenilación y el segundo potenciador (columna H). Los transposones comprendían una secuencia cuya SEQ ID NO. se da en la columna A entre los dos ORF. Los transposones se transfectaron en transfecciones triplicadas independientes en células CHO junto con un gen que codifica una transposasa (SEQ ID NO. 48) fusionado a una señal de localización nuclear heteróloga. Se seleccionaron las células y se midió la expresión de las proteínas fluorescentes (las columnas M y N muestran los promedios de 3 medidas para cada proteína fluorescente) como se describe en el Ejemplo 6.2.1.

Tabla 8. Integración de *transposones derivados de Xenopus* para la expresión de dos polipéptidos.

Los transposones comprendían secuencias diana 5'-TTAA-3', secuencias terminales de transposones SEQ ID NO: 2 y 12, y un potenciador (columna F), promotor (columna G) e intrón (columna H) unidos operativamente a un gen que codifica DasherGFP. Para las filas 3-6, los vectores comprendían además un gen que codifica CayenneRFP unido operativamente a los elementos de control de la expresión mediante una secuencia de acoplamiento traduccional (secuencias identificadas en la columna A). Para las filas 7-18, los vectores comprendían además un gen que codifica Cayenne RFP unido operativamente a un segundo potenciador (columna K), un segundo promotor (columna L) y un segundo intrón (columna M). Las señales de poliadenilación se vincularon al primer (columna I) y al segundo (columna N) marcos de lectura abiertos. Opcionalmente, se intercaló una secuencia aislante entre la primera señal de poliadenilación y el segundo potenciador (columna J). Los transposones comprendían una secuencia cuya SEQ ID NO. se proporciona en la columna A entre los dos ORF. Los transposones se transfectaron en células CHO, opcionalmente (como se indica en la columna O) junto con un gen que codifica una transposasa (SEQ ID NO: 48) fusionado a una señal de localización nuclear heteróloga; se seleccionaron las células y se midió la expresión de las proteínas fluorescentes (columnas P-U) como se describe en el Ejemplo 6.2.2. Las filas 1-2 y 19-20 muestran la transfección de construcciones que codifican solo GFP (filas 1-2 o RFP (filas 19-20)). Las filas 21 y 22 muestran la cotransfección de las construcciones mostradas en las filas 1 y 19. Los detalles se dan en la Sección 6.2.2.

Tabla 9. Expresión de sistemas de transferencia de genes que comprenden genes que codifican dos polipéptidos unidos por elementos de acoplamiento traduccional IRES (5).

Los polinucleótidos de transferencia de genes comprendían un potenciador, un promotor, un intrón y una señal de poliadenilación unida operativamente a un gen que codifica DasherGFP, un elemento IRES y un gen que codifica CayenneRFP. Las SEQ ID NO. de los elementos IRES se dan en la columna A). Los vectores se transfectaron en células HEK o CHO y se midió la expresión de las proteínas fluorescentes como se describe en el Ejemplo 6.4.1. La fluorescencia relativa de las dos proteínas se muestra en la columna B (HEK) o F (CHO). El nivel de expresión relativo de CayenneRFP a DasherGFP se calculó corrigiendo los niveles de fluorescencia relativa de las dos proteínas (CayenneRFP solo produce 0,3 veces la señal de DasherGFP para el mismo nivel de proteína). Esta es la eficiencia IRES que se muestra en la columna C (HEK) o G (CHO). El nivel de expresión de DasherGFP en la construcción IRES se comparó con la expresión de DasherGFP de una construcción que carece de IRES y CayenneRFP y se muestra como % GFP mostrado en la columna D (HEK) o H (CHO). El número de experimentos independientes que miden la expresión de cada IRES en cada sistema se muestra en la columna E (HEK) o I (CHO).

Tabla 10. Transposones de *Xenopus* en *Pichia pastoris*

Se construyeron tres transposones (columna B) para modificar el genoma de *Pichia pastoris*, y comprendían secuencias diana 5'-TTAA-3' y SEQ ID NO: 2 y 12 del extremo del transposón que flanquean un polinucleótido heterólogo como se describe en la Sección 6.5.1. Los plásmidos de las filas 3-8 también comprendían un promotor (cuya SEQ ID NO. se proporciona en la columna D) operativamente enlazado a un gen que codifica la transposasa de *Xenopus* SEQ ID NO: 118 en una porción no transponible del plásmido. Se transformaron diferentes cantidades (columna E) de los tres transposones en células de *Pichia pastoris* competentes, se cultivaron y se colocaron en placas como se describe en la Sección 6.5.1. Se contó el número de colonias resistentes a zeocina en cada placa después de que las células crecieran en medio no selectivo (900 µl de caldo YPD más sorbitol 1 M) durante 5 (columna E) o 24 horas (columna F) a 30 °C.

Tabla 11. Sustituciones que confieren hiperactividad a la transposasa de *Xenopus*.

Se crearon variantes de transposasa, se midieron las frecuencias de transposición y se modeló el efecto de las sustituciones de aminoácidos sobre las frecuencias de transposición como se describe en las Secciones 6.3.1 y 6.3.2. La columna A muestra la posición de las sustituciones en la numeración de la transposasa de *Xenopus* desde el comienzo de la SEQ ID NO: 48, la columna B muestra la identidad de ese aminoácido en la SEQ ID NO: 48, la columna C muestra la identidad de una sustitución de aminoácido que confiere hiperactividad en la transposasa, la columna D muestra el peso de regresión medio de esa sustitución y la columna E muestra la desviación estándar de la ponderación de regresión. La columna F muestra la posición de las sustituciones en la numeración de la transposasa de *Bombyx* desde el comienzo de la SEQ ID NO: 407, la columna G muestra la identidad de ese aminoácido en la SEQ ID NO: 407, la columna H muestra la identidad de una sustitución de aminoácido que confiere hiperactividad en la transposasa, la columna I muestra la ponderación de regresión medio de esa sustitución y la columna J muestra la desviación estándar del peso de regresión.

Tabla 12. Transposasa hiperactiva de *Xenopus* activa como ARNm en células CHO.

Los transposones se cotransfectaron en células CHO con ADN plasmídico (fila 2) o ARNm (filas 3-14) que codificaba una transposasa de *Xenopus* (con SEQ ID NO. mostrada en la columna C) fusionada a una señal de localización nuclear heteróloga. Las cantidades de transposón y ácido nucleico de transposasa transfectadas en 1 ml de células CHO se muestran en las columnas E y F. La transfección, el crecimiento y la selección fueron como se describe en la Sección 6.1.3.3. La expresión de Dasher GFP se midió en un fluorímetro de placa. La fila 15 muestra un control sin ADN. Las columnas H e I muestran fluorescencia de muestras duplicadas 3 días después de la transfección, pero antes de la selección, las columnas J y K muestran fluorescencia de muestras duplicadas inmediatamente después de una selección de 10 días, las columnas L y M muestran fluorescencia de muestras duplicadas 5 días después de la selección.

Tabla 13. Transposasas hiperactivas de *Xenopus* y *Bombyx* activas como ARNm en células CHO.

Los transposones comprendían secuencias terminales de transposones con SEQ ID NO. dadas en las columnas B y C flanqueadas por secuencias diana indicadas en la columna D. Los transposones además comprendían un potenciador de CMV y un promotor de CMV operativamente ligado a un gen que codifica DasherGFP, y un gen que codifica puromicina acetil transferasa unido operativamente a un promotor con SEQ ID NO. mostrado en la columna F. Se cotransfectaron transposones (750 ng) con 250 ng de ARNm que codifica una transposasa con SEQ ID NO. mostrado en la columna G, fusionados a una señal de localización nuclear heteróloga. La transfección, el crecimiento y la selección fueron como se describe en la Sección 6.1.3.4. La expresión de Dasher GFP se midió en un fluorímetro de placa. La fila 16 muestra un control sin ADN. Las columnas H, I y J muestran fluorescencia de muestras por triplicado.

Tabla 14. Frecuencias de transposición relativas para transposasas hiperactivas.

Se crearon variantes de transposasa y se midieron las frecuencias de transposición en *Saccharomyces cerevisiae* como se describe en la Sección 6.3.1 y 6.3.2. La columna A muestra la SEQ ID NO. de transposasas hiperactivas de *Xenopus*, la columna B muestra la frecuencia de transposición de la transposasa hiperactiva en *Saccharomyces cerevisiae*, con respecto a la frecuencia de la secuencia de origen natural SEQ ID NO: 48 en condiciones idénticas. La columna C muestra la SEQ ID NO. de transposasas de *Bombyx* hiperactivas, la columna D muestra la frecuencia de transposición de la transposasa hiperactiva en *Saccharomyces cerevisiae*, con respecto a la frecuencia para la secuencia de origen natural SEQ ID NO: 407 en condiciones idénticas.

Tabla 15. Transposasas de *Bombyx* hiperactivas activas en las células CHO.

Los transposones comprendían secuencias terminales de transposones con SEQ ID NO. dadas en las columnas B y C, flanqueadas por secuencias diana dadas en la columna D. Los transposones además comprendían un promotor (columna F) operativamente ligado a un gen que codifica DasherGFP, y un gen que codifica puromicina acetil transferasa operativamente enlazado a un promotor con la SEQ ID NO. que se muestra en la columna E. Los

5 transposones (750 ng) se cotransfectaron con 250 ng de ARNm (columna I) o ADN (columna J) que codifica una transposasa con la SEQ ID NO. que se muestra en la columna K, fusionada a una señal de localización nuclear heteróloga. Las columnas L, M y N muestran la absorbancia a 600 nm para muestras por triplicado. Las columnas O, P y Q muestran la fluorescencia DasherGFP de las muestras correspondientes. Los detalles experimentales se dan en la Sección 6.3.2.2.

Tabla 16. Transposasas hiperactivas de *Xenopus* activas en células CHO.

10 Los transposones comprendían secuencias aislantes con las SEQ ID NO. mostradas en las columnas D y E, dentro de las secuencias del extremo del transposón SEQ ID NO: 2 y 12, flanqueadas por secuencias diana 5'-TTAA-3'. Los transposones comprendían además un promotor (columna C) operativamente ligado a un gen que codifica DasherGFP, y un gen que codifica puromicina acetil transferasa operativamente ligado a un promotor con la SEQ ID NO. que se muestra en la columna B. Los transposones (750 ng) se cotransfectaron con 250 ng ADN que codifica una transposasa con SEQ ID NO. que se muestra en la columna F, fusionado a una señal de localización nuclear heteróloga. Las columnas G, H e I muestran la fluorescencia DasherGFP de las transfecciones independientes por triplicado de las muestras correspondientes. Los detalles experimentales se dan en la Sección 6.3.1.2.

#### 7. Referencias

20 Las realizaciones específicas descritas en la presente descripción se ofrecen únicamente a modo de ejemplo.

TABLAS 1

1	A	B	C	D	E	F	G	H	I	J	K	L
2	SEC. izquierda*	SEC. derecha*	SEC Int	sistema	adn	SEC transposasa*	nls	adn (ng)	Confluencia (0-100 %)	GFP1	GFP2	GFP3
3	35	36	5'-TTAA-3'	piggyBac	500	ninguno	N/A	0	0	5	5	2
4	35	36	5'-TTAA-3'	piggyBac	500	698	no	160	100	707	677	659
5	1	11	5'-TTAA-3'	Xenopus	500	ninguno	N/A	0	0	4	3	3
6	1	11	5'-TTAA-3'	Xenopus	500	49	sí	160	100	886	890	779
7	1	11	5'-TTAA-3'	Xenopus	500	49	no	160	15	104	105	109
8	1	11	5'-TTAA-3'	Xenopus	500	48	sí	160	100	828	904	803
9	1	11	5'-TTAA-3'	Xenopus	500	48	no	160	5	47	45	55
10	3	12	5'-TTAA-3'	Xenopus	500	ninguno	N/A	0	0	5	5	6
11	3	12	5'-TTAA-3'	Xenopus	500	49	sí	160	100	918	858	820
12	3	12	5'-TTAA-3'	Xenopus	500	49	no	160	10	27	25	26
13	3	12	5'-TTAA-3'	Xenopus	500	48	sí	160	100	953	933	921
14	3	12	5'-TTAA-3'	Xenopus	500	48	no	160	10	73	76	65
15	23	29	5'-TTAT-3'	Bombyx	500	ninguno	N/A	0	0	2	4	4
16	23	29	5'-TTAT-3'	Bombyx	500	750	sí	160	0	11	8	8
17	23	29	5'-TTAT-3'	Bombyx	500	750	no	160	0	2	4	4
18	23	29	5'-TTAT-3'	Bombyx	500	407	sí	160	100	1042	1089	1099
19	23	29	5'-TTAT-3'	Bombyx	500	407	no	160	100	972	1046	960

\*SEQ ID NO.

**TABLA 2**

	A	B	C	D	E	F	G	H	I	J	K	L
1	SEC izquierda*	SEC derecha*	SEC Int	transposón	sistema	adn	SEC transposasa*	nls	adn	GFP1	GFP2	GFP3
2	23	29	5'-TTAT-3'	192465	Bombyx	500 no		N/A	0	6	6	6
3	23	29	5'-TTAT-3'	192465	Bombyx	500 407		si	160	817	788	705
4	24	29	5'-TTAT-3'	214228	Bombyx	500 no		N/A	0	5	4	4
5	24	29	5'-TTAT-3'	214228	Bombyx	500 407		si	160	600	591	602
6	25	29	5'-TTAT-3'	214229	Bombyx	500 no		N/A	0	5	5	4
7	25	29	5'-TTAT-3'	214229	Bombyx	500 407		si	160	103	119	127
8	23	31	5'-TTAT-3'	214230	Bombyx	500 no		N/A	0	5	4	5
9	23	31	5'-TTAT-3'	214230	Bombyx	500 407		si	160	13	12	11
10	22	30	5'-TTAA-3'	214404	Bombyx	500 no		N/A	0	5	5	5
11	22	30	5'-TTAA-3'	214404	Bombyx	500 407		si	160	1035	994	983
12	22	30	5'-TTAA-3'	214404	Bombyx	500 750		si	160	4	4	3
13	3	12	5'-TTAA-3'	192462	Xenopus	500 no		N/A	0	4	3	4
14	3	12	5'-TTAA-3'	192462	Xenopus	500 48		si	160	1048	994	977
15	4	12	5'-TTAA-3'	214231	Xenopus	500 no		N/A	0	4	5	4
16	4	12	5'-TTAA-3'	214231	Xenopus	500 48		si	160	1346	1278	1269
17	5	12	5'-TTAA-3'	217099	Xenopus	500 no		N/A	0	4	3	3
18	5	12	5'-TTAA-3'	217099	Xenopus	500 48		si	160	964	872	901
19	3	13	5'-TTAA-3'	214233	Xenopus	500 no		N/A	0	5	4	6
20	3	13	5'-TTAA-3'	214233	Xenopus	500 48		si	160	1075	1014	1035
21	8	14	5'-TTAT-3'	214406	Xenopus	500 no		N/A	0	4	3	4
22	8	14	5'-TTAT-3'	214406	Xenopus	500 48		si	160	1205	1083	1058
23	35	36	5'-TTAA-3'	136214	piggyBac	500 no		N/A	0	4	4	5
24	35	36	5'-TTAA-3'	136214	piggyBac	500 698		no	160	610	558	577
25	37	38	5'-TTAT-3'	214405	piggyBac	500 no		N/A	0	4	3	3
26	37	38	5'-TTAT-3'	214405	piggyBac	500 698		no	160	485	464	451
27	N/A	N/A	N/A	ninguno	negativo	0 no		N/A	0	4	5	5

\*SEQ ID NO:

TABLA 3

	A	B	C	D	E	F	G	H	I	J	K	L
1	SEC izquierda*	SEC derecha*	SEC Int	indicador	transposón	adn	SEC transposasa*	nls	adn	GFP1	GFP2	GFP3
2	1	11	5'-TTAA-3'	RC3	223949	500	no	N/A	0	3	4	3
3	1	11	5'-TTAA-3'	RC3	223949	500	48	sí	160	1157	1169	1095
4	6	11	5'-TTAA-3'	RC3	223950	500	no	N/A	0	3	3	4
5	6	11	5'-TTAA-3'	RC3	223950	500	48	sí	160	788	930	887
6	7	11	5'-TTAA-3'	RC3	223953	500	no	N/A	0	3	4	3
7	7	11	5'-TTAA-3'	RC3	223953	500	48	sí	160	965	957	918
8	1	15	5'-TTAA-3'	RC3	223951	500	no	N/A	0	4	4	4
9	1	15	5'-TTAA-3'	RC3	223951	500	48	sí	160	1118	1136	1139
10	1	16	5'-TTAA-3'	RC3	223952	500	no	N/A	0	4	4	3
11	1	16	5'-TTAA-3'	RC3	223952	500	48	sí	160	849	889	863
12	35	36	5'-TTAA-3'	RC1	136214	500	no	N/A	0	3	3	4
13	35	36	5'-TTAA-3'	RC1	136214	500	698	no	160	536	555	572
14	N/A	N/A	N/A	ninguno	ninguno	0	no	N/A	0	3	3	3

SEQ ID NO:

TABLA 4

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
<b>Posición</b>	<b>PESO</b>	<b>Transposición</b>	<b>Extirpación</b>	<b>Posición</b>	<b>PESO</b>	<b>Transposición</b>	<b>Extirpación</b>
2	A	RHL	GKRPWIMNTEH	4	E	T	N/A
3	K	VM	QLGWCA	9	R	N/A	P
4	R	MCPK	AIEM	12	A	T	PYQH
5	F	CNQR	CTHPVYEGN	13	M	P	GQN
6	Y	LHVICGASF	RPDLNHVIC	14	L	T	-n/a-
7	S	GV	DG	15	E	G	-n/a-
8	A	-n/a-	ED	20	D	G	HWPQFGKV
9	E	-n/a-	WD	21	Y	-n/a-	K
10	E	-n/a-	AIVNPDM	23	D	-n/a-	HPKS
11	A	DV	CD	24	E	C	YPT
12	A	-n/a-	YARDA	25	S	TKML	YGKQL
13	A	P	TPK	26	S	R	FMWPYCTKA
14	H	-n/a-	PTG	27	S	N	NCYAGHPLMT
15	C	GI	AGVDRLYI	28	E	-n/a-	AYCNIHPL
16	M	ENDSQTA	ELHFINDS	30	E	-n/a-	HGW
17	A	SYVLMT	ESYVQL	32	D	KQW	INQGWFSYL
18	S	CYMLQGPAAWHK	ICYMVLQG	33	H	S	EGDK
19	S	CVLFKEDGNAMPYRT	CVLFKQKEDG	36	E	Q	K
20	S	GMLVHWACQDFN	RGMLVHWACQD	37	H	-n/a-	WSDIQG
21	E	NWFGQLDAPTSYV	NWFGQLDM	39	V	-n/a-	A
22	E	CHRLKSGMVQAYW	CHRLKSDGMVQT	41	Y	-n/a-	SM
23	F	QADWKTVMNPHECR	QADYWKTYMNP	42	D	-n/a-	V
24	S	LWHVPIFKYDCNQ	LWHGAVPIFKYDCN	43	T	-n/a-	YKIW

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>					
A	B	C	D	E	F	G	H	
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación	
25	G	N	N	44	E	CPAQ	ATNQHMDFIVL	
26	S	FHV	FQHYWV	45	E	SM	KPCF	
27	D	LV	L	46	E	P	APY	
28	S	K	YCMMLHTQ	47	R	-n/a-	KM	
29	E	L	LK	48	I	A	-n/a-	
31	V	-n/a-	LTIQK	49	D	K	Y	
32	P	-n/a-	SAV	50	S	-n/a-	A	
33	P	-n/a-	VHS	55	S	-n/a-	A	
34	A	-n/a-	LE	58	R	M	-n/a-	
35	S	-n/a-	ELM	62	A	K	RWCV	
36	E	-n/a-	SVD	63	N	T	ETAW	
37	S	-n/a-	DC	64	A	P	HTMV	
38	D	-n/a-	FNA	65	I	GQ	-n/a-	
39	S	-n/a-	VT	66	I	E	PWFT	
40	S	-n/a-	T	67	A	-n/a-	DS	
41	T	-n/a-	M	68	N	GHQ	SPGVC	
42	E	KN	KNPAS	69	E	WQHTMPDLV	QPHTDVC	
43	E	P	QCWKGGA	70	S	-n/a-	D	
44	S	MEQ	WLM	71	D	KLMCNV	KMNLY	
45	W	-n/a-	LCVSPK	72	S	FT	ETPDH	
46	C	EQTLHP	MI	73	D	-n/a-	AHYECV	
47	S	NC	FAN	74	P	-n/a-	VL	
48	S	VA	VKWA	75	D	-n/a-	HAW	
49	S	G	YLPVKTME	76	D	AQGWVSC	SAH	

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>			H	
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
50	T	DR	ANSK	77	D	-n/a-	TY
51	V	QKYM	TFH	78	L	-n/a-	I
52	S	A	VAKPF	79	P	KG	-n/a-
53	A	YV	EQYK	81	S	-n/a-	Q
54	L	VPCE	IANV	83	V	-n/a-	WQDF
55	E	HPK	S	84	R	NKHYW	ENHYTW
56	E	V	LPYGQW	85	Q	EMKHN	TEKFLH
57	P	VHTQ	KVSA	87	A	MNKHYIC	MNF
58	M	IVPAKFL	DIRVHN	88	S	IY	MV
59	E	YAH	MDN	89	A	-n/a-	KMYC
60	V	-n/a-	EQ	90	S	-n/a-	AD
62	E	SCWVITLQFKG	-n/a-	91	R	-n/a-	A
63	D	TQP	-n/a-	92	Q	EAPNIYHFRDMWCGLV	ADYNRGMFWHTVCPL
64	V	IMQSKHFLTC	-n/a-	93	V	PKMFWL	AIPQWFM
65	D	MVPLKE	-n/a-	94	S	EKTHCL	KEIHYC
66	D	GEAFW	-n/a-	95		EAQTKNIMHDFLC	ATD
67	L	ATMVCHEY	-n/a-	96	P	ATMRGVEQ	EARCV
68	E	SMYAPNVLQHD	-n/a-	97	F	QKHTCWVEPDARG	NADKTHRGYC
69	D	RAPMLHVSWS	-n/a-	98	Y	Q	QF
70	Q	CLTNSGH	-n/a-	99	T	N	DA
71	E	PYMRWLF	-n/a-	102	D	-n/a-	W
72	A	EMTYQIGVFNKLCR	-n/a-	103	G	-n/a-	QM
73	G	HNKFVDSWL	-n/a-	107	Y	-n/a-	M
74	D	T	-n/a-	108	K	-n/a-	S

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>			
A	B	C	D	E	F	G
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición
75	R	WCLMQ	-n/a-	117	L	-n/a-
76	A	LREIV	-n/a-	122	I	-n/a-
77	D	QYLT	-n/a-	128	Q	-n/a-
78	A	QVGRC	-n/a-	132	I	-n/a-
79	A	FVR	-n/a-	135	D	-n/a-
80	A	LY	-n/a-	137	S	-n/a-
81	G	STKV	-n/a-	139		-n/a-
82	G	SLQWE	-n/a-	140	Y	-n/a-
83	E	FCHRDVN	-n/a-	145	I	A
84	P	SFGNVW	-n/a-	149	S	EHPQADT
85	A	MCR	-n/a-	150	D	PEQ
86	W	G	-n/a-	152	L	GR
87	G	L	-n/a-	153	Q	-n/a-
88	P	AENHDL	-n/a-	154	E	Q
89	P	HM	-n/a-	157	T	Q
90	C	KDGNWVQTML	-n/a-	160	N	H
91	N	RALHV	-n/a-	161	S	N
92	F	YRGA	-n/a-	162	S	QWE
93	P	K	-n/a-	164	R	-n/a-
95	E	QVNL	-n/a-	165	H	EGQTMVL
96	I	TW	-n/a-	166	R	CV
97	P	HV	-n/a-	167	Q	-n/a-
98	P	R	-n/a-	168	T	YSWCNMGFALV
99	F	SY	-n/a-	169	K	HPSWGCMV

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>			
A	B	C	D	E	F	G
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición
100	T	VL	-n/a-	170 T		HNG
101	T	GFSVL	-n/a-	171 A		TG
103	P	RV	-n/a-	172 A		-n/a-
104	G	E	-n/a-	173 E		QPA
105	V	F	-n/a-	174 N		-n/a-
106	K	RGME	-n/a-	175 S		NKG
107	V	SR	-n/a-	176 S		HT
108	D	TH	-n/a-	177 A		-n/a-
109	T	IV	-n/a-	178 E		SHYFCAQGV
111	N	F	-n/a-	179 T		Q
114	P	V	-n/a-	180 S		-n/a-
115	I	Q	-n/a-	182 Y		-n/a-
116	N	DQAF	-n/a-	183 M		K
117	F	L	-n/a-	184 Q		H
118	F	CSLQ	-n/a-	185 E		-n/a-
119	Q	HKS	-n/a-	186 T		I
122	M	V	VC	187 T		D
123	T	L	H	188 L		I
124	E	N	SPQ	189 C		DYIWTMFPQ
125	A	VNITPKLGS	VQD	194 L		AMVSTYC
126	I	CVLS	-n/a-	195 I		FMV
127	L	FMC	V	196 A		G
128	Q	K	IE	198 L		EQWTMI
129	D	NI	EQL	200 L		IFCM
						H
						Extirpación
						GQ
						NYK
						CQ
						HQCLM
						R
						KG
						HKT
						IYCWVYVG
						YCDHLPWQSAVG
						HRK
						YRKV
						H
						-n/a-
						YPG
						K
						-n/a-
						D
						TG
						IKQTV
						C
						MV
						G
						WT
						YICMF

(continuación)

Transposasa de <i>Xenopus</i>				Transposasa de <i>Bombyx</i>			
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
130	M	W	-n/a-	201	A	QLM	MQ
132	L	NFTHEMYQ	KNFTH	203	L	VDGECTMA	* YTCMA
133	Y	T	FH	204	I	FACMTGV	DCMNTG
136	V	MITH	FMDRN	205	K	HR	H
137	Y	HAFNR	HAFQSLN	206	S	A	-n/a-
138	A	G	-n/a-	207	N	GA	G
139	E	S	AITVN	209	Q	E	YT
140	Q	RN	TR	210	S	NC	-n/a-
141	Y	IMQSEWVFACKLHR	IMQSEWVF	211	L	GMCTVA	CTV
142	L	VFANQMIRKGYHW	VFATNQCMIRK	212	K	C	-n/a-
143	T	-n/a-	AYV	213	D	E	E
144	Q	RLMEGFDATV	RNLHPSECG	214	L	IM	IM
145	N	CMAQIFGDEVHWY	CLRMAQSIFGDEV	215	W	Y	Y
146	P	VTW	CQLYKVNFE	216	R	KA	K
147	L	PQGKVTMFRI	PWQHG	217	T	VAIPCQM	IFDQCAKV
148	P	MRVFT	MRVCFTQH	219	G	SAC	CHAQ
149	R	-n/a-	LQGP	220	T	-n/a-	C
150	Y	WAGFHSVCMNDEQK	WAGFHSV	222	V	T	A
151	A	GS	REGCS	223	D	E	NS
154	H	-n/a-	CL	224	I	V	V
155	A	QM	-n/a-	227	T	NI	N
157	H	YFT	SWY	228	T	-n/a-	C
158	P	VE	VGS	229	M	F	-n/a-
159	T	-n/a-	PR	234	F	-n/a-	Y

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
160	D	-n/a-	YWC	235	Q	CNHGWYATEM	AHCEWMTFG
161	I	AVLQ	AVLYHK	237	L	I	IV
162	A	LVCKT	GMSLIYVCQ	238	Q	CMHVL	MTHIL
163	E	-n/a-	KGD	239	N	GSA	G
164	M	-n/a-	E	240	N	CHMA	WSCAMH
165	K	-n/a-	RTFC	302	P	K	-n/a-
166	R	-n/a-	A	303	N	CRG	ADSHERKLQ
167	F	-n/a-	R	304	K	-n/a-	Y
168	V	LTIM	L	305	P	H	-n/a-
169	G	-n/a-	D	306	A	QC	-n/a-
170	L	-n/a-	D	307	K	R	-n/a-
171	T	A	P	308	Y	V	-n/a-
172	L	I	SAR	310	I	WML	L
173	A	LSG	LMSIG	311	K	L	-n/a-
174	M	ATQ	WASG	312	I	FCALTVG M	CAMLV
175	G	-n/a-	APC	313	L	FQIEHCYMV	IHQM
176	L	-n/a-	DM	314	A	DQET	T
177	I	RVA	RLV	315	L	IVM	M
178	K	-n/a-	RG	316	V	IA	TC
179	A	TKSVR	TK	317	D	C	C
180	N	-n/a-	TSQ	318	A	TLECV	CV
181	S	-n/a-	A	319	K	CGNHMALQVSDT	SDICATQV
182	L	VIQTWR	SVI	320	N	ALVRDTQCS	RATGHCMVLK
184	S	-n/a-	Y	321	F	HRNYWDGEMKAQ	NHKMYW

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
185	Y	-n/a-	T	322	Y	F	FM
187	D	GI	LMQNGFH	323	V	ILTM	MIALT
188	T	RQSMHIV	R <sup>C</sup>	324	V	NACILTKYHFSQM	GIYHFAMTQLK
189	T	CNLKQVAVYGF	CNLKHQVAVYGG	325	N	-n/a-	HCK
190	T	C	NW	326	L	GCA	AMC
191	V	-n/a-	AELMQI	327	E	NQCHDWFLA	NHMQT
192	L	VCHM	VICH	328	V	TIMP	TAL
193	S	PTRKQGNFH	PTRKQGYDN	330	A	KVP	SPECTLV
194	I	VP	LHRGCV	331	G	-n/a-	A
195	P	G	SGR	332	K	-n/a-	CQ
196	V	LSWAF	MI	333	Q	PTMH	SM
197	F	-n/a-	SML	334	P	H	-n/a-
198	S	R	AK	335	S	HTYKMAQCQLV	NPKYMAECHTQV
199	A	HGNCKRQWSM	HGNCIKR	336	G	PVS	-n/a-
200	T	CIMLNWVQYH	CRIMLNWV	337	P	WEHAMNDKQ	DGCKMALV
201	M	-n/a-	C	339	A	G	-n/a-
202	S	A	PA	340	V	G	-n/a-
203	R	-n/a-	V	341	S	NCPA	-n/a-
204	N	-n/a-	PT	342	N	Q	-n/a-
205	R	-n/a-	L	343	R	SKG	-n/a-
206	Y	-n/a-	P	344	P	GNSA	GSNA
207	Q	-n/a-	T	345	F	STAQGC	PAKMC
208	L	Q	PG	346	E	IQN	-n/a-
209	L	IM	IMA	347	V	L	-n/a-

(continuación)

Transposasa de Xeropus			Transposasa de Bombyx					
A	B	C	D	E	F	G	H	
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación	
210	L	H	A	349	E	TD	TG	
211	R	TCQASK	TCQASK	352	I	V	-n/a-	
212	F	YNM	CAYN	353	Q	NET	-n/a-	
213	L	-n/a-	PM	355	V	F	-n/a-	
214	H	-n/a-	NYMQASE	356	A	R	-n/a-	
215	F	WE	Q	357	R	W	-n/a-	
216	N	-n/a-	Q	359	H	-n/a-	GC	
217	N	E	Q	361	N	TCQM	VCM	
218	N	VRTC	VRGIPDE	362	V	-n/a-	L	
219	A	WGEV	DTLQWIMY	365	D	YKT	T	
220	T	EI	ADEL	367	W	YF	-n/a-	
221	A	M	VC	368	F	-n/a-	Y	
222	V	TILK	QPTILSK	369	T	SA	-n/a-	
223	P	-n/a-	TS	370	G	HQ	-n/a-	
224	P	QDS	QMDVR*EK	371	Y	-n/a-	P	
225	D	KYL	KEPGRMAN	372	E	-n/a-	PT	
226	Q	R	AP	373	L	VIST	TI	
227	P	VDASTNF	VHGDAES	374	M	G	-n/a-	
228	G	-n/a-	HTRQ	375	L	C	-n/a-	
229	H	LP	VD	376	H	YAK	Q	
230	D	-n/a-	Q	379	N	GA	-n/a-	
231	R	P	-n/a-	380	E	WC	T	
233	H	PV	FPW	381	Y	-n/a-	HL	
234	K	-n/a-	ALCDVE	382	R	NK	K	

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>					
A	B	C	D	E	F	G	H	
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación	
235	L	IV	IV	385	S	-n/a-	-n/a-	
236	R	-n/a-	Q	386	V	TICL	TLIC	
237	P	Q	-n/a-	387	G	-n/a-	S	
238	L	-n/a-	VDN	388	T	-n/a-	V	
239	I	V	L	389	V	IMTL	MAL	
240	D	N	HVR	391	K	-n/a-	IMTPL	
242	L	IAE*FRS	NWIAE	392	N	RFV	HFVQ	
243	S	GTLQ	GT	394	R	HKT	PMTA	
244	E	RNHLMQ	RFVDNHLM	395	Q	PFECA	HSYPA	
245	R	QITECP	QKNI	398	E	-n/a-	QA	
246	F	SRL	SR	399	S	NEKHDYGQRTAV	KMQG	
247	A	RECSQHV	RECGSNQ	400	F	GCPWLYM	YWM	
248	A	SLHCNIQY	SLMHCDN	401	I	VCKT	WK	
249	V	TPIAY	TPMID	402	R	YKDFGNEMSQTCLV	SEQFK	
250	Y	PHT	PHT	403	T	WAVFLYNGCISMQK	NFGICEQVL	
251	T	ISKVLMQD	ISNYKVR	404	D	ISENHCMGAQLPV	WMESFANLGVQP	
252	P	L	L	405	R	NTL	G	
253	C	RTLHNGDQVM	RTLHNGDQ	406	Q	FG	ICE	
254	Q	MVL	RH	407	P	KTIQMV	K	
255	N	-n/a-	A	408	N	FIAEMSDYHCQVWL	IAPEKLHV	
256	I	VC	V	409	S	HYNIDFTC	QDNT	
257	C	VYR	V	410	S	TU	TC	
258	I	DRH	DRH	411	V	EQHDS	-n/a-	
259	D	TR	TR	412	F	AW	-n/a-	

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>					
A	B	C	D	E	F	G	H	
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación	
260	E	V	V	414	F	W	-n/a-	
261	S	A	A	415	Q	N	AN	
262	L	SA	SA	416	K	-n/a-	S	
263	L	VAMRD	VAM	418	I	C	-n/a-	
264	L	VPDKMR	VS	419	T	FICS	-n/a-	
265	F	YK	EHYW	420	L	M	-n/a-	
266	K	GRA	GR	424	A	-n/a-	D	
267	G	PL	P	426	K	-n/a-	T	
268	R	A	ACHYQK	428	N	S	H	
269	L	SIVCQ	SIVC	430	V	-n/a-	D	
270	Q	VKACPLIEGYNTW	VHKACPL	432	V	Y	THMC	
271	F	VPT	VP	433	V	-n/a-	L	
272	R	KILSVC	K	434	M	Q	A	
273	Q	MVE	MVITN	436	T	-n/a-	S	
274	Y	HI	H	440	D	SCMLV	SIKCAQ	
275	I	PLM	PL	441	N	FRMGCDL	GFAVLW	
276	P	IRAWCL	IR	442	S	YKFVL	KGFCYWV	
277	S	EAK	EA	443	I	EFV	AYK	
278	K	AL	A	444	D	QIMV	MA	
279	R	YQKVGS	Y	445	E	PYHCGKMQL	CGMKPNLTW	
280	A	S	S	446	S	-n/a-	EAMDYCPLWG	
281	R	LY	LYK	447	T	-n/a-	QS	
282	Y	LQGCVHSNT	LQGCVH	448	G	WYHCTV	NWQ	
283	G	YI	YAI	449	E	APTL	HGTCIL	

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>					
A	B	C	D	E	F	G	H	
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación	
284	I	QVGLF	QV	450	K	-n/a-	T	
285	K	I	I	451	Q	VENDSRYHGFCITPMWL	ENTRCSMYAW	
286	F	LT	L	452	K	IFVL	-n/a-	
287	Y	QKSF	QKSF	454	E	-n/a-	C	
288	K	TADFLC	T	455	M	QCLV	PGCVI	
289	L	CTRGYVE	CTR	456	I	ACMLTV	V	
290	C	TVQ	T	457	T	CGA	A	
291	E	VD	VCND	458	F	ADC	A	
292	S	RVA	R	461	S	KGEDYA	KTL	
293	S	NDHTWAK	NDHTWAK	464	A	S	T	
294	S	RNGT	RCNGT	466	V	TC	C	
295	G	TDSL	TDS	468	V	QMT	CT	
296	Y	HF	HF	469	V	TAHCL	HACT	
297	T	CPVMLD	CPVML	471	E	Q	-n/a-	
298	S	EVMKGLNCQA	EVMKGLNCQA	472	L	KQM	-n/a-	
299	Y	HKCREGAN	HKCREGA	473	C	GSQT	IGSTM	
300	F	YM	VCIYM	474	A	CQMGT	CTV	
301	L	-n/a-	-n/a-	475	N	-n/a-	S	
302	I	-n/a-	V	477	N	-n/a-	D	
304	E	H	DHSQC	483	K	R	N	
305	G	-n/a-	E	484	R	-n/a-	HK	
306	K	-n/a-	NL	485	W	FYTDKEQMV	-n/a-	
307	D	-n/a-	F	486	P	-n/a-	EMA	
308	S	-n/a-	RG	488	T	KV	V	

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
309	K	-n/a-	GCHMLQE	489	L	YIV	CTV
310	L	R	IRV	491	Y	V	-n/a-
311	D	-n/a-	FHYWSNRILC	492	G	A	A
312	P	-n/a-	C	493	V	HQWMIL	IML
313	P	SMLY	VSKMHE	496	M	D	L
314	G	ASIHIL	NQM	499	I	DHWTCEMALV	CWV
315	C	-n/a-	SR	502	C	SYML	-n/a-
316	P	N	RDA	503	I	MALQF	QFL
317	P	DLK	DNFMHCGVALKE	505	Y	-n/a-	Q
318	D	RL	NTAFKQHRCWEM	507	T	RDSGKIMECAL	DIMECAVL
319	L	CVF	CIDVAM	509	K	H	-n/a-
320	T	CGSNKHMV	CGSRNKQ	510	N	KGA	QGK
321	V	I	INT	511	V	KA	CTEKA
322	S	-n/a-	TIC	512	T	MA	MCA
323	G	-n/a-	A	513	I	V	M
324	K	G	SR	514	K	-n/a-	P
325	I	-n/a-	L	515	R	K	-n/a-
326	V	-n/a-	WT	516	T	-n/a-	s
327	W	-n/a-	M	517	E	DA	NKAQ
328	E	SHKVVWFQLT	SYHHRKVWADFMCQ	521	S	HKQGE	HCGETK
329	L	-n/a-	GM	523	G	QTAMSC	TMSICLA
330	I	AV	M	524	L	KM	HIYM
331	S	AGQP	AKWNRG	525	S	Q	CNDTQ
332	P	Q	KGD	527	I	MV	M



(continuación)

Transposasa de <i>Xeropus</i>			Transposasa de <i>Bombyx</i>					
A	B	C	D	E	F	G	H	
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación	
360	T	P	PLYSP	559	H	KSC	SIWK	
361	P	-n/a-	QS	560	V	FPI	HYKIP	
362	A	-n/a-	P	561	N	P	QGA	
363	C	-n/a-	W	562	V	Y	ISM	
364	G	-n/a-	D	563	P	ITKE	DE	
366	I	-n/a-	L	564	G	L	QPCF	
367	N	-n/a-	TGR	565	R	K	-n/a-	
368	R	-n/a-	TAWKP	566	Y	M	-n/a-	
369	D	-n/a-	RVAQSLML	567	V	IH	N	
371	K	-n/a-	A	570	Q	F	N	
372	G	-n/a-	FY	571	D	SFVQM	NSMTAV	
373	L	-n/a-	MI	573	P	K	MT	
375	R	-n/a-	SQKAVPT	574	Y	V	A	
376	A	-n/a-	TCLVEIMGK	575	K	-n/a-	H	
377	L	-n/a-	VI	576	K	W	I	
378	L	-n/a-	ITMCKYV	581	K	H	-n/a-	
379	D	-n/a-	A	583	S	M	-n/a-	
380	K	-n/a-	LAFTVE	585	N	-n/a-	SKGL	
381	K	-n/a-	LPVRITN	586	A	E	NH	
382	L	-n/a-	VC	588	A	GRF	-n/a-	
383	N	-n/a-	LIDFPVESHGAK	593	M	-n/a-	I	
384	R	-n/a-	C	594	E	C	-n/a-	
385	G	-n/a-	NWAMCHYSKQ	597	K	-n/a-	W	
387	T	-n/a-	RE	598	F	M	-n/a-	

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
388	Y	-n/a-	FV	599	L	Y	-n/a-
389	A	-n/a-	YF	601	E	V	FQW
390	L	-n/a-	V	602	N	GRQHTEDS	GTQMEH
392	K	-n/a-	YMWC	603	C	D	-n/a-
393	N	-n/a-	EA	604	A	I	DTSI
394	E	-n/a-	F	605	E	RWKMPYCHAQSV	PYAMRWHQVIGK
397	A	-n/a-	SIFLCVM	606	L	VQYAEKGCKNHM	EQCYANWVMK
399	K	-n/a-	ASH	607	D	VYCNWTAHQELKG	HQTYWCANLEKG
400	F	-n/a-	C	608	S	EDRM	QWRCV
401	F	-n/a-	YD	609	S	RWHVQGTKN	WHTGNYKV
402	D	-n/a-	G	610	L	TIKAWDQSFN	DSI
405	N	-n/a-	VD				
406	L	-n/a-	GTM				
409	L	-n/a-	I				
422	R	G	QLKWSM				
423	V	NPTFHCS	GARLNPTFHC				
424	G	CNSL	CKQYPWNTHS				
425	E	-n/a-	SAQCAPGH				
426	P	LKYF	TWLVCSQHKYN				
428	K	RQ	NTRF				
429	N	GPYM	GWPYEHRAMS				
430	K	R	QPR				
431	P	LQ	LTC				
432	L	TMF	HTSQMN				

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extracción	Posición	PESO	Transposición	Extracción
434	S	A	A				
435	K	LTR	YMHISVLA				
436	E	QAMLY	WHCQAFML				
438	S	QA	QMA				
439	K	LMR	HALMCR				
440	Y	FLT	WQFLH				
442	G	-n/a-	W				
443	G	V	V				
444	V	-n/a-	C				
446	R	LMK	HTLMK				
447	T	SAC	QSNAGC				
450	L	MVA	MIVE				
451	Q	LMF	ALMF				
452	H	-n/a-	AS				
455	N	D	E				
457	T	-n/a-	C				
458	R	-n/a-	*				
460	T	S	A				
461	R	YQKT	Y				
462	A	MTYFKRQHE	MTYNFKRQH				
464	Y	-n/a-	QW				
465	K	VHM	VHT				
467	V	TCA	KTC				
468	G	SF	CST				

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
469	I	V	NV				
470	Y	H	-n/a-				
471	L	VM	FCVM				
472	I	VLLW	VMLFW				
473	Q	-n/a-	D				
474	M	A	IQAT				
475	A	STG	-n/a-				
476	L	M	VNFMCC				
477	R	AQ	HV				
478	N	-n/a-	K				
479	S	L	L				
480	Y	HF	H				
482	V	L	L				
483	Y	T	T				
484	K	G	GEAFVS				
485	A	CQV	C				
486	A	EHCV	R				
487	V	NTR	NCMW				
488	P	EHKQFM	ELND				
489	G	-n/a-	YFQ				
490	P	GTHAKL	GTCIMH				
491	K	QV/GCLM	QVIGW				
492	L	-n/a-	VQ				
493	S	ATPI	GA				

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
494	Y	F	M				
495	Y	FL	MF				
496	K	H	VQ				
497	Y	T	-n/a-				
498	Q	CM	VLGHTCEM				
499	L	HACVQTRNW	HGACVQKTR				
500	Q	ECRHA	ETCRFMVH				
501	I	LMVS	TLM				
502	L	IMVG	FIMV				
503	P	HENCASLQ	HENCV				
504	A	NMVIPWDQLTKGFHYS	NMVIPWDQLTK				
505	L	MC	M				
506	L	MIC	HQMI				
507	F	VVHMK	IV				
508	G	QTYR	ES				
509	G	NLRMKQHIPCFA	TNWLRMKQ				
510	V	MCAN	HKM				
511	E	TMILP	TYQMF				
512	E	SYMKVARLTI	SGP				
513	Q	YFVNISKW	YFVPMAE				
514	T	QVHFMRP	QWNVHFMGR				
515	V	FTRAL	KFHS				
516	P	-n/a-	LM				
517	E	MVAKL	MGVASI				

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extrpación	Posición	PESO	Transposición	Extrpación
518	M	SHLFTA	IRSHLWFVG				
519	P	WR	FWMND				
520	P	WRMFLFQVGDKY	WERMLTFQVGDK				
521	S	AHCVW	TKFAHG				
522	D	AR	VNEAFH				
523	N	WAGSPM	WQALKGDSFC				
524	V	PMA	PH				
525	A	QLIR	QMLNI				
527	L	VHMRANFW	SYVHTCMRAIQ				
528	I	LKVFQHNTG	RLKVFAMQHY				
529	G	-n/a-	MAHDLVWC				
530	K	QG	MVQR				
531	H	RP	-n/a-				
532	F	CMVQ	CIHYRNIMKVTA				
533	I	VF	MVTSFGE				
534	D	EQLRVCMNAGF	THEQLKRVCMNSA				
535	T	SRALV	SRCAFLGVHKINMW				
536	L	Q	DMIQHSRKEVF				
537	P	-n/a-	NF				
538	P	-n/a-	TAFGVYKW				
539	T	SNL	SMKQI				
540	P	KH	KEVRMIFN				
541	G	-n/a-	K				
542	K	-n/a-	QNYHT				

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
543	Q	-n/a-	T				
544	R	-n/a-	FD				
545	P	-n/a-	T				
546	Q	T	TAVYN				
547	K	-n/a-	T				
548	G	-n/a-	DE				
549	C	-n/a-	Sí				
550	K	N	NPCSFLT				
551	V	-n/a-	AYM				
552	C	-n/a-	H				
553	R	K	NMTVHK				
554	K	VT	VMACLGYFIPE				
555	R	HV	ELKHVGDIFMTN				
556	G	SCN	VMFDSKCQA				
557	I	KFSV	RNHQLKPCFG				
558	R	G	GL				
559	R	H	HIVLTGKEYSWMFQN				
560	D	G	TRLHSV MAGNC				
561	T	V	SVAQI				
562	R	-n/a-	VK				
563	Y	-n/a-	AFNSGR				
564	Y	TV	GMFTNQ				
565	C	-n/a-	V				
566	P	VHKQ	VHGASMIK				

(continuación)

Transposasa de <i>Xenopus</i>			Transposasa de <i>Bombyx</i>				
A	B	C	D	E	F	G	H
Posición	PESO	Transposición	Extirpación	Posición	PESO	Transposición	Extirpación
567	K	ML	MLQVT				
568	C	-n/a-	W				
569	P	YT	VLYSEMTF				
570	R	F	VLMTYK				
571	N	VDMK	FVWTDMYK				
572	P	-n/a-	VNQF				
573	G	-n/a-	RCA				
574	L	M	TMIP				
575	C	H	H				
576	F	LKVDWMCR	LKQAVYDWNMGCIER				
577	K	LGDRHYI	EVLGDRNHYI				
578	P	SENTQVM	SCKENGID				
579	C	-n/a-	Y				
580	F	M	EAH				
581	E	IWRSGVHAC	ILWRFQSGDTVM				
582	I	VKRMG	NEVKAQ				
583	Y	L	CFDQ				
584	H	-n/a-	L				
585	T	G	QHANCY				
586	Q	LCRYHFENKGAW	LVTDMCRYHFENKGA				
587	L	FDRIPNESYMQGWKT	FDRIPNESYMQGWKT				
588	H	RK	SMWRGE				
589	Y	VCKMIEDQR	SVCFHKNWGPMI				

TABLA 5

	A	B	C	D	E	F	G	H	I	J	K	L
1	SEC izquierda *	SEC derecha *	transposón	ADN (ng)	SEC transposasa *	nls	promotor de transposasa	ADN (ng)	ARN (ng)	GFP1	GFP2	GFP3
2	35	36	136214	500	no	-	N/A	0	0	4	5	6
3	35	36	136214	500	698	n/a-	N/A	160	0	468	458	476
4	1	11	202970	500	no	N/A	N/A	0	0	6	6	5
5	1	11	202970	500	48	sí	CMV	160	0	1079	1086	1137
6	1	11	202970	500	48	sí	PGK	160	0	248	269	244
7	1	11	202970	500	48	sí	UBB	160	0	179	188	196
8	1	11	202970	500	48	sí	SV40	160	0	1305	1247	1293
9	1	11	202970	500	48	sí	N/A	0	125	74	78	78
10	1	11	202970	500	48	sí	N/A	0	250	262	252	249
11	1	11	202970	500	48	sí	N/A	0	500	328	347	342
12	1	11	202970	500	48	sí	N/A	0	1000	45	40	45
13	3	12	192462	500	no	N/A	N/A	0	0	5	6	6
14	3	12	192462	500	48	sí	CMV	160	0	820	873	915
15	3	12	192462	500	48	sí	PGK	160	0	44	42	40
16	3	12	192462	500	48	sí	UBB	160	0	29	31	31
17	3	12	192462	500	48	sí	SV40	160	0	1535	1523	1537
18	3	12	192462	500	48	sí	N/A	0	125	13	14	15
19	3	12	192462	500	48	sí	N/A	0	250	97	113	115
20	3	12	192462	500	48	sí	N/A	0	500	283	271	277
21	3	12	192462	500	48	sí	N/A	0	1000	31	31	30
	N/A	N/A	N/A	0	no	N/A	N/A	0	0	2	5	4

\*SEQ ID NO:

TABLA 6

	A	B	C	D	E	F	G	H	I	J	K	L
1	SEC izquierda *	SEC indicador*	SEC derecha *	transposón	ADN (ng)	SEC transposasa *	nls	ADN (ng)	ARN (ng)	GFP1	GFP2	GFP3
2	35	39	36	136214	500	no	N/A	0	0	7	7	9
3	35	39	36	136214	500	698	no	160	0	57	59	60
4	23	39	29	192465	500	407	sí	0	0	6	6	5
5	23	39	29	192465	500	407	sí	0	125	37	35	35
6	23	39	29	192465	500	407	sí	0	250	786	783	792
7	23	39	29	192465	500	407	sí	0	500	903	908	934
8	23	39	29	192465	500	407	sí	0	1000	184	198	225
9	23	39	29	192465	500	407	sí	125	0	311	322	336
10	23	39	29	192465	500	407	sí	250	0	254	260	272
11	23	39	29	192465	500	407	sí	500	0	174	176	193
12	23	40	29	194093	500	no	N/A	0	0	884	911	936
13	23	40	29	194093	500	407	sí	0	125	2861	2533	2830
14	23	40	29	194093	500	407	sí	0	250	4123	3907	4074
15	23	40	29	194093	500	407	sí	0	500	5668	5564	5554
16	23	40	29	194093	500	407	sí	0	1000	7387	7062	7355
17	23	40	29	194093	500	407	sí	125	0	7863	7281	7000
18	23	40	29	194093	500	407	sí	250	0	7684	8043	8335
19	23	40	29	194093	500	407	sí	500	0	8201	7826	7684
20	N/A	N/A	N/A	N/A	0	no	N/A	0	0	4	4	4

SEQ ID NO:

Tabla 7

Fila	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	SEC enlazador *	Gen	GFP	RFP	Promotor 1	Intrón 1	poliA1	aislante intergénico	Potenciador 2	Promotor 2	Intrón 2	pA2	GFP	RFP	R/G	ORF2/ORF1	D %
1	N/A	188550	no	sí	EF1a	EF1a	globina (conejo)	N/A	N/A	N/A	N/A	N/A	4	1564	N/A	N/A	0,00
2	N/A	181650	sí	no	EF1a	EF1a	globina (conejo)	N/A	N/A	N/A	N/A	N/A	8542	2	N/A	N/A	1,00
3	N/A	146674	sí	sí	EF1a	EF1a	N/A	N/A	N/A	N/A	N/A	globina (conejo)	1508	322	0,21	1,00	0,18
4	1051	188209	sí	sí	EF1a	EF1a	N/A	N/A	N/A	N/A	N/A	globina (conejo)	9964	741	0,07	0,33	1,17
5	1062	206694	sí	sí	EF1a	EF1a	N/A	N/A	N/A	N/A	N/A	globina (conejo)	6248	604	0,10	0,48	0,73
6	1063	206695	sí	sí	EF1a	EF1a	N/A	N/A	N/A	N/A	N/A	globina (conejo)	6206	529	0,09	0,43	0,73
7	1064	206696	sí	sí	EF1a	EF1a	N/A	N/A	N/A	N/A	N/A	globina (conejo)	6280	586	0,09	0,43	0,74
8	1011	203906	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	CMV	EF1a	no	globina (conejo)	1830	117	0,06	0,29	0,21
9	1023	203907	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	CMV	p-actina	no	globina (conejo)	1982	97	0,05	0,24	0,23
10	1025	203909	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	no	EF1a	no	globina (conejo)	2714	80	0,03	0,14	0,32
11	998	203910	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	CMV	CMV	CMVc	globina (conejo)	1613	565	0,35	1,67	0,19
12	1002	203914	sí	sí	EF1a	EF1a	HSV- TK/gastrina	2xHS4c	CMV	GAPDH	GAPDH	globina (conejo)	2432	688	0,28	1,33	0,29
13	1000	203912	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	CMV	GAPDH	eMLP	globina (conejo)	2150	316	0,15	0,71	0,25

(Continuación)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
SEC enlazador *	Gen	GFP	RFP	Promotor 1	Intrón 1	poliA1	aislante intergénico 2	Potenciador 2	Promotor 2	Intrón 2	pA2	GFP	RFP	R/G	ORF2/ORF1	D %
14	203913	sí	sí	EF1a	EF1a	HSV- TK/gastrina	2xHS4c	no	EF1a	EF1a	globina (conejo)	2853	1504	0,53	2,52	0,33
15	203908	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	no	EF1a	EF1a	globina (conejo)	2795	252	0,09	0,43	0,33
16	203915	sí	sí	EF1a	EF1a	HSV- TK/gastrina	2xHS4c	CMV	CMV	no	globina (conejo)	2505	142	0,06	0,29	0,29
17	203911	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	CMV	CMV	no	globina (conejo)	1012	118	0,12	0,57	0,12
18	203916	sí	sí	EF1a	EF1a	HSV- TK/gastrina	2xHS4c	no	EF1a	eMLP	globina (conejo)	3430	537	0,16	0,76	0,40
19	203917	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	no	EF1a	eMLP	globina (conejo)	2390	185	0,08	0,38	0,28
20	203918	sí	sí	EF1a	EF1a	HSV- TK/gastrina	2xHS4c	CMV	GAPDH	CMVc	globina (conejo)	1903	533	0,28	1,33	0,22
21	203919	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	CMV	GAPDH	CMVc	globina (conejo)	2169	310	0,14	0,67	0,25
22	203920	sí	sí	EF1a	EF1a	HSV- TK/gastrina	2xHS4c	CMV	GAPDH	no	globina (conejo)	2046	410	0,20	0,95	0,24
23	207390	sí	sí	EF1a	EF1a	HSV- TK/gastrina	no	CMV	GAPDH	no	globina (conejo)	2087	226	0,11	0,52	0,24

\*SEQ ID NO:



(continuación)

Fila	SEC Enlazador *	Transposición										Transposasa			Expresión de GFP			Expresión de RFP				
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
		Transposición 1	Transposición 2	GFP	RFP	Potenciador 1	Promotor 1	Intrón 1	poliA1	Aislante intergénico	Potenciador 2	Promotor 2	Intrón 2	pA2			1	2	3	1	2	3
13	1001	198836	N/A	sí	sí	CMV	CMV	ninguno	HSV-TK	Núcleo HS4 2x	ninguno	EF1a	EF1a	globina (conejo)	no		38	39	36	104	94	98
14	1001	198836	N/A	sí	sí	CMV	CMV	ninguno	HSV-TK	Núcleo HS4 2x	ninguno	EF1a	EF1a	globina (conejo)	si		320	374	449	1102	1245	1471
15	1002	198837	N/A	sí	sí	CMV	CMV	ninguno	HSV-TK	Núcleo HS4 2x	CMV	GAPDH	GAPDH	globina (conejo)	no		67	55	55	58	45	42
16	1002	198837	N/A	sí	sí	CMV	CMV	ninguno	HSV-TK	Núcleo HS4 2x	CMV	GAPDH	GAPDH	globina (conejo)	si		396	470	411	418	483	425
17	1003	198838	N/A	sí	sí	CMV	CMV	ninguno	HSV-TK	Núcleo HS4 2x	CMV	CMV	ninguno	globina (conejo)	no		25	27	22	11	13	10
18	1003	198838	N/A	sí	sí	CMV	CMV	ninguno	HSV-TK	Núcleo HS4 2x	CMV	CMV	ninguno	globina (conejo)	si		280	260	245	122	118	104
19	N/A	200967	N/A	no	sí	CMV	CMV	ninguno	globina (conejo)	N/A	N/A	N/A	N/A	N/A	no		5	5	4	4	10	11
20	N/A	200967	N/A	no	sí	CMV	CMV	ninguno	globina (conejo)	N/A	N/A	N/A	N/A	N/A	si		5	6	6	375	389	392
21	N/A	187151	200967	sí	sí	CMV	CMV	ninguno	globina (conejo)	N/A	N/A	N/A	N/A	N/A	no		34	33	35	7	8	8
22	N/A	187151	200967	sí	sí	CMV	CMV	ninguno	globina (conejo)	N/A	N/A	N/A	N/A	N/A	si		546	583	628	186	196	197

\*SEQ ID NO.

TABLA 9

	A	B	C	D	E	F	G	H	I
Fila	SEC IRES*	HEK RFP/GFP	HEK Eficiencia IRES	HEK % GFP	HEK # mediciones	CHO RFP/GFP	CHO Eficiencia IRES	CHO % GFP	CHO # mediciones
1	1050	0,07	0,22	0,39	3	0,08	0,28	0,15	4
2	1051	0,12	0,41	0,44	4	0,11	0,38	0,29	4
3	1052	0,10	0,35	0,25	4	0,04	0,13	0,47	5
4	1053	0,05	0,17	0,39	3	0,05	0,17	0,25	4
5	1065	0,08	0,27	0,26	2	0,09	0,30	0,14	2
6	1066	0,07	0,23	0,51	4	0,02	0,08	0,62	4
7	1067	0,00	0,01	0,31	2	0,01	0,05	0,21	2
8	1068	0,08	0,25	0,48	3	0,02	0,06	0,58	4
9	1069	0,03	0,08	0,21	1	0,01	0,02	0,39	1
10	1070	0,01	0,04	0,15	1	0,00	0,01	0,26	1
11	1071	0,13	0,45	0,46	3	0,09	0,28	0,54	3
12	1072	0,07	0,23	0,16	2	0,02	0,08	0,58	3
13	1073	0,06	0,19	0,58	2	0,07	0,24	0,25	3
14	1074	0,03	0,10	0,54	1	0,04	0,14	0,38	2
15	1075	0,12	0,39	0,37	2	0,05	0,15	0,72	3
16	1076	0,11	0,35	0,32	1	0,04	0,15	0,79	2
17	1077	0,03	0,11	0,33	1	0,01	0,05	0,66	2
18	1078	0,03	0,10	0,32	1	0,02	0,06	0,85	1
19	1079	0,07	0,22	0,61	2	0,03	0,12	0,64	1
20	1080	0,07	0,22	0,51	2	0,04	0,12	0,84	1
21	1081	0,11	0,35	0,32	1	0,04	0,12	0,67	1
22	1082	0,06	0,22	0,56	2	0,08	0,25	0,35	2
23	1083	0,08	0,27	0,40	2	0,10	0,33	0,25	2

(continuación)

Fila	A	B		C		D		E		F		G		H		I	
	SEC IRES*	HEK RFP/GFP	HEK Eficiencia IRES	HEK % GFP	HEK # mediciones	CHO RFP/GFP	CHO Eficiencia IRES	CHO % GFP	CHO # mediciones	CHO RFP/GFP	CHO Eficiencia IRES	CHO % GFP	CHO # mediciones	CHO RFP/GFP	CHO Eficiencia IRES	CHO % GFP	CHO # mediciones
24	1084	0,05	0,16	0,66	1	0,06	0,21	0,47	1	0,06	0,21	0,47	1	0,06	0,21	0,47	1
25	1085	0,04	0,13	0,57	1	0,04	0,14	0,30	1	0,04	0,14	0,30	1	0,04	0,14	0,30	1
26	1086	0,11	0,35	0,30	2	0,05	0,15	0,82	3	0,05	0,15	0,82	3	0,05	0,15	0,82	3
27	1087	0,08	0,27	0,42	2	0,09	0,28	0,16	2	0,09	0,28	0,16	2	0,09	0,28	0,16	2
28	1088	0,11	0,36	0,41	2	0,04	0,15	0,66	3	0,04	0,15	0,66	3	0,04	0,15	0,66	3
29	1089	0,02	0,06	0,38	1	0,02	0,05	0,85	1	0,02	0,05	0,85	1	0,02	0,05	0,85	1
30	1090	0,02	0,07	0,33	1	0,01	0,03	0,41	1	0,01	0,03	0,41	1	0,01	0,03	0,41	1
31	1091	0,00	0,01	0,21	1	0,00	0,01	0,63	1	0,00	0,01	0,63	1	0,00	0,01	0,63	1
32	1092	0,02	0,07	0,26	1	0,03	0,11	0,35	1	0,03	0,11	0,35	1	0,03	0,11	0,35	1
33	1093	0,07	0,23	0,25	3	0,03	0,10	0,58	4	0,03	0,10	0,58	4	0,03	0,10	0,58	4
34	1094	0,00	0,01	0,17	2	0,01	0,04	0,60	3	0,01	0,04	0,60	3	0,01	0,04	0,60	3
35	1096	0,06	0,18	0,29	1	ND	ND	ND	0	ND	ND	ND	0	ND	ND	ND	0

\*SEQ ID NO.

Tabla 10

	A	B	C	D	E	F	
5	<b>1</b>	<b>Construcción</b>	<b>configuración de plásmido</b>	<b>SEC Promotor de transposasa*</b>	<b>ADN (ug)</b>	<b>crecimiento horas</b>	<b>5 24 horas</b>
	2	N/A	N/A	N/A	0	0	0
	3	251587	circular	949	0,2	0	48
10	4	251587	circular	949	1	11	93
	5	251587	circular	949	2	26	276
	6	251588	circular	954	0,2	13	58
	7	251588	circular	954	1	60	221
15	8	251588	circular	954	2	137	456
	9	251589	circular	ninguno	0,2	2	0
	10	251589	circular	ninguno	1	0	1
	11	251589	circular	ninguno	2	1	6
20	12	251589	lineal	ninguno	1	661	~1000

\*SEQ ID NO:

25

Tabla 11

	A	B	C	D	E	F	G	H	I	J
	<b>Posición de Xenopus</b>	<b>De</b>	<b>Para</b>	<b>Peso</b>	<b>Peso estándar</b>	<b>Posición de Bombyx</b>	<b>De</b>	<b>Para</b>	<b>Peso</b>	<b>Peso estándar</b>
30	6	Y	C	0,09	0,03	85	Q	E	-0,01	0,03
	7	S	G	0,25	0,05	92	Q	A	0,09	0,03
	9	E	D	0,00	0,01	92	Q	L	-0,06	0,08
35	16	M	S	0,23	0,05	92	Q	N	-0,04	0,02
	18	S	G	-0,03	0,05	93	V	L	0,35	0,08
	19	S	G	0,05	0,02	93	V	M	0,20	0,09
40	20	S	D	0,20	0,02	96	P	G	0,07	0,02
	20	S	G	0,26	0,03	97	F	C	0,03	0,03
	20	S	Q	0,40	0,05	97	F	H	0,18	0,03
	21	E	D	0,38	0,07	165	H	E	0,28	0,07
45	22	E	Q	0,17	0,05	165	H	W	0,27	0,07
	23	F	P	0,25	0,07	178	E	H	0,13	0,06
	23	F	T	0,37	0,10	178	E	S	0,29	0,04
50	24	S	Y	0,17	0,05	189	C	P	0,12	0,08
	26	S	V	0,10	0,05	196	A	G	0,48	0,02
	28	S	Q	0,10	0,03	200	L	F	-0,10	0,08
	31	V	K	0,04	0,02	200	L	I	0,46	0,05
55	34	A	E	0,03	0,02	200	L	M	0,01	0,02
	67	L	A	0,10	0,04	201	A	Q	0,22	0,10
	73	G	H	0,29	0,06	203	L	T	-0,03	0,11
	76	A	V	0,15	0,04	207	N	G	-0,01	0,07
60	77	D	N	0,11	0,02	211	L	A	0,20	0,03
	88	P	A	0,05	0,02	215	W	Y	0,19	0,03
	91	N	D	0,14	0,06	217	T	A	-0,05	0,02
65	141	Y	A	0,14	0,03	217	T	K	0,00	0,08

ES 2 891 087 T3

(Continuación)

	A	B	C	D	E	F	G	H	I	J
	Posición de Xenopus	De	Para	Peso	Peso estándar	Posición de Bombyx	De	Para	Peso	Peso estándar
5	141	Y	Q	0,33	0,04	219	G	A	-0,04	0,04
	145	N	E	0,03	0,02	219	G	S	0,02	0,03
10	145	N	V	0,02	0,03	235	Q	G	0,13	0,08
	146	P	K	0,10	0,03	235	Q	N	-0,06	0,09
	146	P	T	0,11	0,04	235	Q	Y	0,33	0,08
	146	P	V	0,11	0,03	238	Q	L	0,51	0,08
15	148	P	H	0,03	0,02	242	R	Q	-0,06	0,06
	148	P	T	0,42	0,04	246	K	I	0,24	0,05
	150	Y	C	0,10	0,05	253	K	V	0,32	0,10
20	150	Y	G	0,25	0,05	258	M	V	0,18	0,06
	150	Y	S	0,21	0,04	261	F	L	0,15	0,05
	157	H	Y	0,37	0,06	263	S	K	0,28	0,07
25	162	A	C	0,18	0,06	271	C	S	0,36	0,04
	179	A	K	0,36	0,04	303	N	R	0,11	0,07
	182	L	I	0,27	0,06	312	I	V	-0,02	0,08
	182	L	V	0,16	0,08	321	F	D	0,12	0,06
30	189	T	G	0,04	0,03	321	F	W	0,18	0,08
	192	L	H	0,01	0,02	323	V	T	0,01	0,02
	193	S	K	0,03	0,05	324	V	H	0,28	0,07
	193	S	N	0,03	0,03	324	V	K	0,32	0,08
35	196	V	I	0,03	0,02	330	A	V	0,34	0,09
	198	S	G	0,26	0,04	333	Q	M	0,00	0,04
	200	T	W	0,02	0,02	337	P	A	-0,02	0,03
40	202	S	A	-0,01	0,06	368	F	Y	-0,08	
	210	L	H	0,15	0,05	373	L	C	0,25	0,06
	212	F	N	0,17	0,09	373	L	V	0,10	0,04
	218	N	E	0,11	0,06	389	V	L	0,15	0,05
45	248	A	N	0,50	0,05	394	R	T	-0,01	0,11
	263	L	M	0,35	0,06	395	Q	P	-0,11	0,10
	270	Q	L	0,07	0,03	399	S	N	0,07	0,02
50	294	S	T	0,23	0,06	402	R	K	0,11	0,06
	297	T	M	0,18	0,07	403	T	L	0,09	0,04
	304	E	Q	-0,02	0,03	404	D	I	-0,02	0,01
	308	S	R	0,05	0,03	404	D	M	0,10	0,07
55	310	L	R	0,26	0,07	404	D	Q	0,35	0,07
	333	L	M	0,14	0,09	404	D	S	0,27	0,07
	336	Q	M	0,02	0,05	408	N	H	-0,03	0,03
60	354	A	H	0,12	0,03	409	S	N	-0,07	0,08
	357	C	V	0,31	0,06	441	N	R	0,02	0,08
	358	L	F	0,08	0,04	448	G	W	0,09	0,05
65	359	D	N	0,28	0,09	449	E	A	0,04	0,05

ES 2 891 087 T3

(Continuación)

	A	B	C	D	E	F	G	H	I	J
	Posición de Xenopus	De	Para	Peso	Peso estándar	Posición de Bombyx	De	Para	Peso	Peso estándar
5										
10	377	L	I	0,10	0,08	469	V	T	0,02	0,03
	423	V	H	0,25	0,06	472	L	M	-0,06	0,07
	426	P	K	0,21	0,07	473	C	Q	0,30	0,04
	428	K	R	0,04	0,04	484	R	K	0,15	0,10
15	434	S	A	-0,06	0,09	507	T	C	0,17	0,03
	438	S	A	0,08	0,05	523	G	A	0,10	0,03
	447	T	A	0,20	0,05	527	I	M	0,05	0,11
20	447	T	C	-0,01	0,04	528	Y	K	0,80	0,08
	447	T	G	0,34	0,07	543	Y	I	0,20	0,06
	450	L	V	0,08	0,05	549	E	A	0,18	0,02
25	462	A	H	0,67	0,03	550	K	M	0,28	0,07
	462	A	Q	0,37	0,04	556	S	V	-0,04	0,07
	467	V	C	-0,04	0,04	557	P	S	0,22	0,06
	469	I	V	0,21	0,05	559	H	K	-0,04	0,06
30	472	I	L	0,01	0,06	560	V	F	-0,01	0,02
	476	L	M	-0,02	0,05	561	N	P	-0,04	0,05
	488	P	E	0,00	0,05	562	V	Y	-0,08	0,05
	498	Q	M	0,17	0,09	567	V	H	0,00	0,05
35	502	L	V	0,31	0,07	567	V	I	0,02	0,05
	517	E	I	0,05	0,02	583	S	M	-0,02	0,05
	520	P	D	0,35	0,05	601	E	V	0,31	0,06
40	520	P	G	0,09	0,07	605	E	C	-0,11	0,09
	520	P	K	0,00	0,03	605	E	H	0,28	0,05
	521	S	G	0,00	0,05	605	E	M	-0,06	0,06
	523	N	S	0,34	0,05	605	E	W	0,05	0,05
45	533	I	E	0,02	0,07	607	D	C	-0,05	0,02
	534	D	A	0,17	0,04	607	D	H	0,04	0,03
	576	F	E	0,12	0,05	607	D	K	-0,02	0,01
50	576	F	R	0,42	0,06	607	D	N	-0,02	0,04
	577	K	I	0,26	0,03	609	S	H	0,25	0,03
	582	I	R	0,01	0,07	609	S	V	-0,02	0,01
	583	Y	F	0,06	0,07	610	L	I	0,19	0,03
55	587	L	W	0,03	0,07					
	587	L	Y	0,35	0,06					

60

65

Tabla 12

	A	B	C	D	E	F	G	H	I	J	K	L	M
	ARN	ADN	SEC	Transposasa*	Transposón	Transposasa	Relación ADN:ARN	Transitorio	Transitorio	Selección	Selección	Recuperación	Recuperación
1													
2	no	sí	48	CMV sin aislantes	N/A	ninguno	N/A	1150	629	172	157	233	143
3	sí	no	48	CMV sin aislantes	750 ng	250 ng	3:1	735	976	351	916	4211	4229
4	sí	no	48	CMV sin aislantes	660 ng	330 ng	2:1	516	509	228	184	1505	822
5	sí	no	48	CMV sin aislantes	500 ng	500 ng	1:1	436	351	146	139	134	118
6	sí	no	168	CMV sin aislantes	750 ng	250 ng	3:1	1006	476	1342	2053	4229	6040
7	sí	no	168	CMV sin aislantes	660 ng	330 ng	2:1	842	770	1918	4350	5936	5709
8	sí	no	168	CMV sin aislantes	500 ng	500 ng	1:1	548	542	2263	1284	5162	4927
9	sí	no	189	CMV sin aislantes	750 ng	250 ng	3:1	1107	420	2073	1072	5883	5323
10	sí	no	189	CMV sin aislantes	660 ng	330 ng	2:1	837	654	1119	1796	5126	6111
11	sí	no	189	CMV sin aislantes	500 ng	500 ng	1:1	664	680	3935	2853	6218	4647
12	sí	no	175	CMV sin aislantes	750 ng	250 ng	3:1	872	468	3442	3012	5676	7511
13	sí	no	175	CMV sin aislantes	660 ng	330 ng	2:1	928	605	2479	2233	5616	5173
14	sí	no	175	CMV sin aislantes	500 ng	500 ng	1:1	644	508	3832	1840	5276	5344
15	no	no	ninguno	ninguno	0	0	N/A	236	280	143	140	143	122

\*SEQ ID NO:

Tabla 13

	A	B	C	D	E	F	G	H	I	J
1	transposón	SEC izquierda*	SEC derecha*	SEC Int	sistema	SEC promotor puro*	SEC transposasa*	GFP1	GFP2	GFP3
2	187151	2	12	5'-TTAA-3'	Xenopus	937	175	875	63	979
3	187151	2	12	5'-TTAA-3'	Xenopus	937	189	909	957	135
4	187151	2	12	5'-TTAA-3'	Xenopus	937	ninguno	236	84	84
5	241555	1095	11	5'-TTAA-3'	Xenopus	942	189	2594	91	3168
6	241555	1095	11	5'-TTAA-3'	Xenopus	942	175	2934	3746	4365
7	241555	1095	11	5'-TTAA-3'	Xenopus	942	ninguno	94	93	102
8	246143	2	12	5'-TTAA-3'	Xenopus	942	175	2445	2361	2324
9	246143	2	12	5'-TTAA-3'	Xenopus	942	ninguno	66	68	63
10	194094	23	29	5'-TTAT-3'	Bombyx	937	407	426	710	630
11	194094	23	29	5'-TTAT-3'	Bombyx	937	1098	708	89	741
12	194094	23	29	5'-TTAT-3'	Bombyx	937	ninguno	88	92	94
13	240671	22	30	5'-TTAA-3'	Bombyx	937	407	641	89	89
14	240671	22	30	5'-TTAA-3'	Bombyx	937	1098	664	808	681
15	240671	22	30	5'-TTAA-3'	Bombyx	937	ninguno	379	94	94
16	ninguno	N/A	N/A	N/A	ninguno	N/A	N/A	87	91	87

SEQ ID NO:

Tabla 14

	A	B	C	D
	SEC Xenopus *	hiperactividad	SEC Bombyx *	hiperactividad
5	228	9	654	1,3
	244	7	639	1,9
	247	6	634	2,0
10	252	6	619	3
	268	5	614	3
	51	0,4	1097	4
	64	80	596	4
15	56	126	595	4
	57	122	588	5
	124	25	557	7
20	52	414	518	11
	58	116	517	11
	54	127	508	12
	73	58	491	15
25	71	64	488	15
	65	79	457	31
	63	91	449	35
	62	95	417	94
30	61	99	416	97
	59	112	415	107
	168	15	414	122
35	189	13	413	130
	175	15	412	164
	118	22		
	211	10		
40	216	9		
*SEQ ID NO:				

45

50

55

60

65

TABLA 15

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
transposón	SEC izquierda*	SEC derecha*	SEC Int	SEC Promotor Puro*	Promotor GFP	SEC Aislante L*	SEC Aislante R*	ARN	ADN	SEC Transposasa Bombyx*	A600	A600	A600	GFP	GFP	GFP
1	194094	29	5'-TTAT-3'	937	CMV	ninguno	ninguno	si	no	1098	0,44	0,34	0,42	1947	1547	1876
2	194094	29	5'-TTAT-3'	937	CMV	ninguno	ninguno	no	si	415	0,34	0,27	0,32	1455	1240	1231
3	194094	29	5'-TTAT-3'	937	CMV	ninguno	ninguno	no	si	457	0,28	0,34	0,30	1107	1152	1213
4	194094	29	5'-TTAT-3'	937	CMV	ninguno	ninguno	no	si	417	0,30	0,30	0,34	1061	950	1098
5	194094	29	5'-TTAT-3'	937	CMV	ninguno	ninguno	no	si	412	0,31	0,33	0,35	860	1049	1143
6	194094	29	5'-TTAT-3'	937	CMV	ninguno	ninguno	no	si	416	0,32	0,29	0,35	1016	910	1004
7	194094	29	5'-TTAT-3'	937	CMV	ninguno	ninguno	no	si	407	0,31	0,30	0,30	943	800	866
8	194094	29	5'-TTAT-3'	937	CMV	ninguno	ninguno	no	no	N/A	0,02	0,04	0,04	150	171	167
9	240671	30	5'-TTAA-3'	937	CMV	ninguno	ninguno	si	no	1098	0,48	0,44	0,26	2177	1757	1016
10	240671	30	5'-TTAA-3'	937	CMV	ninguno	ninguno	no	si	415	0,34	0,30	0,35	1525	1480	1514
11	240671	30	5'-TTAA-3'	937	CMV	ninguno	ninguno	no	si	457	0,29	0,34	0,31	1257	1191	1144
12	240671	30	5'-TTAA-3'	937	CMV	ninguno	ninguno	no	si	412	0,34	0,29	0,28	1001	1032	897
13	240671	30	5'-TTAA-3'	937	CMV	ninguno	ninguno	no	si	416	0,27	0,33	0,29	917	874	953
14	240671	30	5'-TTAA-3'	937	CMV	ninguno	ninguno	no	si	407	0,32	0,26	0,27	1006	784	885

(continuación)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
transposón	SEC izquierda*	SEC derecha*	SEC Int	SEC Promotor Puro*	Promotor GFP	SEC Aislante L*	SEC Aislante R*	ARN	ADN	SEC Transposasa Bombyx*	A600	A600	A600	GFP	GFP	GFP
15 240671	22	30	5'-TTAA-3'	937	CMV	ninguno	ninguno	no	sí	417	0,27	0,25	0,23	800	859	777
16 240671	22	30	5'-TTAA-3'	937	CMV	ninguno	ninguno	no	no	N/A	0,03	0,17	0,06	178	261	168
17 246143	22	30	5'-TTAA-3'	942	CMV	ninguno	ninguno	no	sí	415	0,00	0,00	0,00	102	109	142
18 246143	22	30	5'-TTAA-3'	942	CMV	ninguno	ninguno	no	sí	412	0,00	0,00	0,00	114	103	107
19 246143	22	30	5'-TTAA-3'	942	CMV	ninguno	ninguno	no	sí	416	0,00	-0,01	0,00	109	102	106
20 246143	22	30	5'-TTAA-3'	942	CMV	ninguno	ninguno	no	sí	417	0,00	0,00	0,00	106	100	98
21 246143	22	30	5'-TTAA-3'	942	CMV	ninguno	ninguno	no	sí	407	0,00	-0,01	0,00	105	98	101
22 246143	22	30	5'-TTAA-3'	942	CMV	ninguno	ninguno	sí	no	1098	-0,01	0,00	0,00	99	104	96
23 246143	22	30	5'-TTAA-3'	942	CMV	ninguno	ninguno	no	sí	457	0,00	0,00	0,00	97	101	100
24 246143	22	30	5'-TTAA-3'	942	CMV	ninguno	ninguno	no	no	N/A	0,00	0,00	0,00	109	105	104
25 246170	22	30	5'-TTAA-3'	942	EF1a	864	860	no	sí	415	0,18	0,00	0,04	5477	162	1559
26 246170	22	30	5'-TTAA-3'	942	EF1a	864	860	no	sí	412	0,03	0,04	0,06	1148	1589	3145
27 246170	22	30	5'-TTAA-3'	942	EF1a	864	860	no	sí	417	0,02	0,01	0,00	637	683	203
28 246170	22	30	5'-TTAA-3'	942	EF1a	864	860	no	sí	416	0,00	0,02	0,00	146	652	217

(continuación)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
	transposón	SEC izquierda*	SEC derecha*	SEC Int	SEC Promotor Puro*	Promotor GFP	SEC Aislante L*	SEC Aislante R*	ARN	ADN	SEC Transposasa Bombyx*	A600	A600	A600	GFP	GFP	
29	246170	22	30	5'-TTAA-3'	942	EF1a	864	860	sí	no	1098	0,00	0,01	0,00	237	286	118
30	246170	22	30	5'-TTAA-3'	942	EF1a	864	860	no	sí	457	0,00	0,00	0,00	106	122	115
31	246170	22	30	5'-TTAA-3'	942	EF1a	864	860	no	sí	407	0,00	0,00	0,00	108	101	113
32	246170	22	30	5'-TTAA-3'	942	EF1a	864	860	no	no	N/A	0,00	0,00	0,00	108	128	114

\*SEQ ID NO.

TABLA 16

A	B	C	D	E	F	G	H	I
transposición	SEC promotor puro*	Promotor GFP	SEC Aislante L*	SEC Aislante R*	SEC Transposasa Uribo*	GFP	GFP	GFP
1 246143	942	CMV	ninguno	ninguno	48	94	94	112
2 246143	942	CMV	ninguno	ninguno	58	99	2600	111
3 246143	942	CMV	ninguno	ninguno	ninguno	107	94	98
4 246170	942	EF1a	864	860	48	95	93	108
5 246170	942	EF1a	864	860	61	4075	113	94
6 246170	942	EF1a	864	860	ninguno	114	95	100
7 261961	948	EF1a	864	864	48	96	97	112
8 261961	948	EF1a	864	864	57	128	2008	490
9 261961	948	EF1a	864	864	ninguno	86	104	94

\*SEQ ID NO.

**REIVINDICACIONES**

- 5 1. Una transposasa que tiene al menos un 90 % de identidad de secuencia con la SEQ ID NO:407 o 415, en donde la transposasa no tiene una secuencia de origen natural, que comprende al menos una sustitución de aminoácido, con respecto a la SEQ ID NO:407, como se muestra en la Tabla 4, columna G o H o la Tabla 11, columna H.
- 10 2. La transposasa de la reivindicación 1 que tiene una combinación de sustituciones con respecto a la SEQ ID NO:407 mostrada en la Tabla 4, columnas G y H y actividad mejorada con respecto a la SEQ ID NO:407.
- 15 3. La transposasa de la reivindicación 1 que tiene una actividad mejorada con respecto a la SEQ ID NO:407, por ejemplo, mayor en 1,5, 2, 5, 10, 20, 50 o 100 veces.
- 20 4. La transposasa de la reivindicación 1, que cuando se fusiona con una secuencia de localización nuclear heteróloga puede (i) escindir un transposón de SEQ ID NO:47, o (ii) transponer un transposón de SEQ ID NO:47, al escindir el transposón e integrarlo en el ADN genómico de una célula diana.
- 25 5. La transposasa de la reivindicación 1, que comprende al menos una sustitución de aminoácido, con respecto a la SEQ ID NO:407, como se muestra en la Tabla 11, columna H.
- 30 6. La transposasa de la reivindicación 1, que comprende más de una sustitución de aminoácido, con respecto a la SEQ ID NO:407, como se muestra en la Tabla 11, columna H.
- 35 7. La transposasa de la reivindicación 1, que comprende al menos una sustitución de aminoácido en una de las siguientes posiciones (numeradas de acuerdo con la SEQ ID NO:407): 92, 93, 96, 97, 165, 178, 189, 196, 200, 201, 211, 215, 235, 238, 246, 253, 258, 261, 263, 271, 303, 321, 324, 330, 373, 389, 399, 402, 403, 404, 448, 473, 484, 507, 523, 527, 528, 543, 549, 550, 557, 601, 605, 607, 609 o 610.
- 40 8. La transposasa de la reivindicación 1, que comprende al menos una de las siguientes sustituciones de aminoácidos, con respecto a la SEQ ID NO:407: Q92A, V93L, V93M, P96G, F97H, F97C, H165E, H165W, E178S, E178H, C189P, A196G, L200I, A201Q, L211A, W215Y, G219S, Q235Y, Q235G, Q238L, K246I, K253V, M258V, F261L, S263K, C271S, N303R, F321W, F321D, V324K, V324H, A330V, L373C, L373V, V389L, S399N, R402K, T403L, D404Q, D404S, D404M, N441R, G448W, E449A, V469T, C473Q, R484K, T507C, G523A, I527M, Y528K, Y543I, E549A, K550M, P557S, E601V, E605H, E605W, D607H, S609H o L610I.
- 45 9. La transposasa de la reivindicación 1, en donde la transposasa comprende una secuencia seleccionada de una de las SEQ ID NO:409-697.
- 50 10. La transposasa de la reivindicación 1, que comprende además una secuencia de localización nuclear heteróloga.
- 55 11. Un polinucleótido que codifica la transposasa de cualquier reivindicación anterior.
- 60 12. El polinucleótido de la reivindicación 11, en donde el polinucleótido es una molécula de ARNm.
13. El polinucleótido de la reivindicación 11, en donde el polinucleótido comprende un promotor unido operativamente a un segmento del polinucleótido que codifica la transposasa.
14. Un método para crear una célula transgénica que comprende (i) introducir en una célula eucariota (a) la transposasa de la reivindicación 1, en donde la transposasa i) se proporciona como una proteína, que se introduce en la célula, o ii) se proporciona como un polinucleótido, opcionalmente ARNm, que codifica la transposasa, que se introduce en la célula y expresa la transposasa, y (b) un polinucleótido que comprende un transposón que comprende extremos del transposón de un transposón similar a piggyBac que flanquea un polinucleótido heterólogo, en donde un extremo del transposón comprende al menos 16 bases contiguas de SEQ ID NO:25 y el otro extremo del transposón comprende al menos 16 bases contiguas de SEQ ID NO:31 e (ii) identificar una célula en la que el transposón está incorporado en el genoma de la célula eucariota; en donde el método no implica modificar la identidad genética de la línea germinal de ningún ser humano.
15. La transposasa de la reivindicación 1 para su uso en terapia génica, en donde no se modifica la identidad genética de la línea germinal de ningún ser humano.